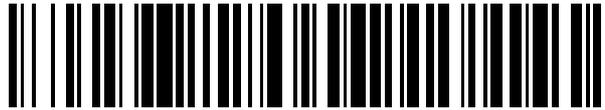


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 796 493**

51 Int. Cl.:

**G10L 19/16** (2013.01)  
**G10L 19/02** (2013.01)  
**G10L 19/20** (2013.01)  
**G10L 19/09** (2013.01)  
**G10L 25/90** (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **27.08.2008 E 08015123 (6)**

97 Fecha y número de publicación de la concesión europea: **06.05.2020 EP 2104096**

54 Título: **Aparato y método para convertir una señal de audio en una representación parametrizada, aparato y método para modificar una representación parametrizada, aparato y método para sintetizar una representación parametrizada de una señal de audio**

30 Prioridad:

**20.03.2008 US 38300 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**27.11.2020**

73 Titular/es:

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

72 Inventor/es:

**DISCH, SASCHA**

74 Agente/Representante:

**ARIZTI ACHA, Monica**

**ES 2 796 493 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

5 Aparato y método para convertir una señal de audio en una representación parametrizada, aparato y método para modificar una representación parametrizada, aparato y método para sintetizar una representación parametrizada de una señal de audio

La presente invención se relaciona con la codificación de audio y, en particular, con esquemas de codificación de audio parametrizados, que se aplican en vocodificadores.

10 Una clase de vocodificadores es los vocodificadores de fase. Un tutorial sobre vocodificadores de fase es la publicación "The Phase Vocoder: A tutorial", Mark Dolson, Computer Music Journal, volumen 10, n.º 4, páginas 14 a 27, 1986. Una publicación adicional es "New phase vocoder techniques for pitch-shifting, harmonizing and other exotic effects", L. Laroche y M. Dolson, proceedings 1999, IEEE workshop on applications of signal processing to audio and acoustics, New Paltz, Nueva York, 17 a 20 de octubre, 1999, páginas 91 a 94.

15 Las figuras 5 a 6 ilustran diferentes implementaciones y aplicaciones para un vocodificador de fase. La figura 5 ilustra una implementación del banco de filtro de un vocodificador de fase, en el que se proporciona una señal de audio en una entrada 500, y en donde, en una salida 510, se obtiene una señal de audio sintetizada. Específicamente, cada canal del banco de filtro ilustrado en la figura 5 comprende un filtro 501 de paso de banda y un oscilador 502 conectado posteriormente. Las señales de salida de todos los osciladores 502 de todos los canales se combinan a través de un combinador 503, que se ilustra como un sumador. En la salida del combinador 503, se obtiene la señal de salida 510.

20 Cada filtro 501 se implementa para proporcionar, de una parte, una señal de amplitud  $A(t)$ , y de otra parte, la señal de frecuencia  $f(t)$ . La señal de amplitud y la señal de frecuencia son señales de tiempo. La señal de amplitud ilustra un desarrollo de la amplitud dentro de una banda de filtro a lo largo del tiempo y la señal de frecuencia ilustra el desarrollo de la frecuencia de una señal de salida del filtro a lo largo del tiempo.

25 En la figura 6 se ilustra una implementación esquemática de un filtro 501. La señal de entrada se envía en dos rutas paralelas. En una ruta, la señal se multiplica por una onda senoidal con una amplitud de 1,0 y una frecuencia igual a la frecuencia central del filtro de paso de banda como se ilustra en 551. En la otra ruta, la señal se multiplica por una onda coseno de la misma amplitud y frecuencia como se ilustra en 551. Así, las dos rutas paralelas son idénticas excepto por la fase de la forma de onda de multiplicación. Después, en cada ruta, el resultado de la multiplicación se carga en un filtro 553 de paso bajo. La operación de multiplicación en si misma también se conoce como una modulación de anillo simple. Multiplicar cualquier señal mediante una onda senoidal (o coseno) de frecuencia constante tiene el efecto de cambiar simultáneamente todos los componentes de frecuencia en la señal original al sumar y restar la frecuencia de la onda senoidal. Si este resultado se pasa ahora a través de un filtro de paso bajo adecuado, solo permanecerá la porción de baja frecuencia. Esta secuencia de operaciones también se conoce como heterodinización. Esta heterodinización se realiza en cada una de las dos rutas paralelas, pero debido a que una ruta se heterodina con una onda senoidal, mientras la otra ruta utiliza una onda coseno, las señales heterodinadas resultantes en las dos rutas están fuera de fase en  $90^\circ$ . El filtro 553 de paso bajo superior, por lo tanto, proporciona una señal 554 cuadrada y el filtro 553 más bajo proporciona una señal en fase. Estas dos señales, que también se conocen como señales I y Q, se remiten a un transformador 556 coordinado, que genera una representación de magnitud/fase de la representación rectangular.

30 La señal de amplitud se hace salir en 557 y corresponde a  $A(t)$  de la figura 5. La señal de fase se introduce en un desempaquetador 558 de fase. En la salida del elemento 558 no hay un valor de fase entre  $0$  y  $360^\circ$  sino que hay un valor de fase, que aumenta en una forma lineal. Este valor de fase "desempaquetado" se introduce en un convertidor 559 de fase/frecuencia que puede, por ejemplo, ser implementado como un dispositivo de diferencia de fase que sustrae una fase en un momento de tiempo anterior desde la fase en un momento de tiempo actual con el fin de obtener el valor de frecuencia para el momento de tiempo actual.

Este valor de frecuencia se agrega a un valor de frecuencia constante  $f_i$  del canal de filtro  $i$ , con el fin de obtener un valor de frecuencia que varía con el tiempo en una salida 560.

35 El valor de frecuencia en la salida 560 tiene una porción DC  $f_i$  y una porción cambiante, que también se conoce como la "fluctuación de frecuencia", mediante la cual se desvía una frecuencia de corriente de la señal en el canal de filtro de la frecuencia central  $f_i$ .

40 Así, el vocodificador de fase como se ilustra en la figura 5 y la figura 6 proporciona una separación de información espectral e información de tiempo. La información espectral está comprendida en la ubicación del canal del banco de filtro específico en la frecuencia  $f_i$ , y la información de tiempo está en la fluctuación de frecuencia y en la magnitud a lo largo del tiempo.

Otra descripción del vocodificador de fase es la interpretación de transformación de Fourier. Consiste en una sucesión

de transformaciones de Fourier sobrepuestas tomadas sobre ventanas de tiempo de duración finita. En la interpretación de la transformación de Fourier, la atención se enfoca en los valores de fase y magnitud para todas las diferentes bandas de filtro o intervalos de frecuencia en un único punto de tiempo. Aunque en la interpretación del banco de filtro, puede observarse la resíntesis como un ejemplo clásico de síntesis aditiva con los controles de frecuencia y amplitud que varían con el tiempo para cada oscilador, la síntesis, en la implementación de Fourier, se logra al convertir de nuevo a la forma real e imaginaria y solapar-agregar las transformaciones de Fourier inversas sucesivas. En la interpretación de Fourier, el número de bandas de filtro en el vocodificador de fase es el número de puntos de frecuencia en la transformación de Fourier. De manera similar, la separación igual en la frecuencia de los filtros individuales se puede reconocer como la característica fundamental de la transformación de Fourier. De otra parte, la forma de las bandas de paso de filtro, es decir, la pendiente del corte en los bordes de bandas se determina mediante la forma de la función de la ventana que se aplica antes de calcular la transformación. Para una forma particular característica, por ejemplo, la ventana Hamming, la pendiente del corte de filtro aumenta en proporción directa a la duración de la ventana.

Es útil observar que las dos interpretaciones diferentes del análisis de vocodificador de fase aplican solo a la implementación del banco de filtros de paso de banda. La operación mediante la que las salidas de estos filtros se expresan como frecuencias y amplitudes que varían con el tiempo es la misma para ambas implementaciones. La meta básica de este vocodificador de fase es separar la información temporal de la información espectral. La estrategia operativa es dividir la señal en un número de bandas espectrales y caracterizar la señal que varía con el tiempo en cada banda.

Dos operaciones básicas son particularmente significativas. Estas operaciones son la transposición de altura y el cambio de escala de tiempo. Siempre es posible frenar un sonido grabado simplemente al reproducirlo a una menor velocidad de muestra. Esto es análogo a reproducir una grabación en cinta a una menor velocidad de reproducción. Pero, esta clase de expansión de tiempo sin pista reduce simultáneamente el tono mediante el mismo factor que la expansión de tiempo. Frenar la evolución temporal de un sonido sin alterar su altura requiere una separación explícita de la información espectral y temporal. Como se observó anteriormente, esto es precisamente lo que el vocodificador de fase intenta hacer. Extendiendo las señales de frecuencia y amplitud varían con el tiempo  $A(t)$  y  $f(t)$  para la figura 5a no cambia la frecuencia de los osciladores individuales del todo, pero lo hace reduciendo la evolución temporal: del sonido compuesto. El resultado es un sonido expandido en el tiempo con la altura original. La vista de la transformación de Fourier de cambio de escala en tiempo es tal que, con el fin de expandir en el tiempo un sonido, la FFT inversa puede simplemente separarse adicionalmente del análisis FFT. Como resultado, ocurren más lentamente cambios espectrales en el sonido sintetizado que en el original en esta aplicación, y la fase se reescala de manera precisa por el mismo factor mediante el cual el sonido se expande en el tiempo.

La otra aplicación es la transposición de altura. Debido a que el vocodificador de fase puede utilizarse para cambiar la evolución temporal de un sonido sin cambiar su tono, también debe ser posible hacerlo inverso, es decir, cambiar su altura sin cambiar la duración. Esto se hace o bien al cambiar la escala de tiempo utilizando el factor de cambio de altura deseado y luego reproducir los sonidos resultantes a una velocidad de muestra equivocada o bien disminuir la resolución mediante un factor deseado y reproducirlo a una velocidad sin cambio. Por ejemplo, para aumentar la altura mediante un octavo, el sonido se expande primero en el tiempo mediante un factor de 2 y la expansión en el tiempo se reproduce luego a dos veces la velocidad de muestra original.

El vocodificador (o 'VODER') fue inventado por Dudley como un dispositivo sintetizador operado manualmente para generar voz humana [2]. Algún tiempo considerable después del principio de su operación se extendió hacia los denominados vocodificadores de fase [3][4]. El vocodificador de fase opera en el solapamiento del espectro DFT de tiempo corto y por lo tanto sobre un conjunto de filtros de sub-banda con frecuencias centrales fijas. El vocodificador ha tenido una amplia aceptación como un principio subyacente para manipular archivos de audio. Por ejemplo, efectos de audio como transposición de altura y extensión del tiempo se logran fácilmente mediante un vocodificador [5]. Desde entonces, se han publicado muchas modificaciones y mejoras a esta tecnología. Específicamente las restricciones de tener filtros de análisis de frecuencia fijo caen al agregar un mapeo derivado de una frecuencia fundamental ( $f_0$ ), por ejemplo en el vocodificador 'DERECHO' [6]. Sin embargo, se mantiene el caso de uso prevalente, que es el procesamiento/codificación de voz.

Otra área de interés para la comunidad del procesamiento de audio ha sido la descomposición de señales de voz en componentes modulados. Cada componente consiste en un portador, una parte de modulación de amplitud (AM) y una parte de modulación de frecuencia (FM) de alguna clase. Una forma de señal adaptativa de tal descomposición se publicó por ejemplo en [7] que sugiere el uso de un conjunto de filtros de paso de banda adaptativos de la señal. En [8] se presentó un enfoque que utiliza información AM en combinación con un codificador paramétrico `sinusoidales más ruido`. Otro método de descomposición se publicó en [9] que utiliza la denominada estrategia `FAME`: aquí, las señales de voz se han descompuesto en cuatro bandas utilizando filtros de paso de banda con el fin de extraer posteriormente su contenido AM y FM. Las publicaciones más recientes también están dirigidas a reproducir señales de audio a partir de solo información AM (envolturas de sub-banda) y sugiere métodos iterativos para recuperar la información de fase asociada que contiene predominantemente el FM [10].

Nuestro método presentado en el presente documento está dirigido al procesamiento de señales de audio generales, por lo tanto también incluye la música. Es similar a un vocodificador de fase pero modificado con el fin de realizar una señal perceptivamente dependiente motivada por la descomposición de sub-banda en un conjunto de frecuencias del portador de sub-banda con señales AM y FM asociadas. Destacamos que esta descomposición es perceptivamente significativa y que sus elementos se pueden interpretar en una forma directa hacia adelante, así como todas las clases de procesamiento de modulación en los componentes de la descomposición llega a ser factible.

Para lograr la meta establecida anteriormente, basándose en la observación de que existen señales perceptivamente similares. Una señal de paso de banda de tono de banda suficientemente estrecha se representa bien perceptivamente mediante un portador sinusoidal en su posición "centro de gravedad" (COG) espectral y su envoltura Hilbert. Esto tiene raíz en el hecho de que ambas señales evocan aproximadamente el mismo movimiento de la membrana basilar en la oreja humana [11]. Un ejemplo simple para ilustrar esto es el complejo de dos tonos (1) con frecuencias  $f_1$  y  $f_2$  suficientemente cercanas una de la otra de tal manera que ellas se fusionan perceptivamente en un componente (sobre) modulado,

$$s_i(t) = \text{sen}(2\pi f_1 t) + \text{sen}(2\pi f_2 t) \quad (1)$$

Una señal que consiste en un portador sinusoidal en una frecuencia igual al COG espectral del  $S_i$  y que tiene la misma envoltura de amplitud absoluta como el  $S_i$  es  $S_m$  según (2).

$$s_m(t) = 2 \text{sen}\left(2\pi \frac{f_1 + f_2}{2} t\right) \cdot \left| \cos\left(2\pi \frac{|f_1 - f_2|}{2} t\right) \right| \quad (2)$$

En la figura 9b se representan (gráfico superior y central) la señal de tiempo y la envoltura Hilbert de ambas señales. Observe que el salto de fase de  $\pi$  en la primera señal en ceros de la envoltura se opone a la segunda señal. La figura 9a visualiza los gráficos de densidad espectral de potencia de las dos señales (gráficos superior y central).

Aunque estas señales son considerablemente diferentes en su contenido espectral sus pistas preceptuales predominantes, la frecuencia media representada por el COG, y la envoltura de amplitud son similares. Esto los hace sustitutos perceptivamente mutuos con respecto a una región espectral limitada por banda centrada en el COG como se representa en la figura 9a y la figura 9b (gráfico inferiores). El mismo principio sigue siendo válido para señales más complicadas.

Generalmente, los sistemas de análisis/síntesis de modulación que descomponen una señal de banda ancha en un conjunto de componentes cada uno comprendiendo un portador, información de modulación de frecuencia y de modulación de amplitud que tienen muchos grados de libertad ya que, en general, esta tarea es un problema mal planteado. Los métodos que modifican las envolturas de magnitud de sub-banda de los espectros de audio complejos y los recombinan posteriormente con sus fases no modificadas para resíntesis dan como resultado artefactos, ya que, estos procedimientos no prestan atención al receptor final del sonido, es decir, el oído humano.

Adicionalmente, al aplicar FFT muy largos, es decir, ventanas muy largas con el fin de obtener una resolución de frecuencia fina reduce concurrentemente la resolución del tiempo. De otra parte las señales transitorias no requerirían una resolución de alta frecuencia, pero requerirían una alta resolución de tiempo, ya que, en un cierto momento de tiempo las señales de paso de banda exhiben fuerte correlación mutua, que también se conoce como la "coherencia vertical". En esta terminología, uno imagina un gráfico de tiempo-espectrograma en donde en el eje horizontal, se utiliza la variable de tiempo y en donde en el eje vertical, se utiliza la variable de frecuencia. Las señales transitorias de procesamiento con una muy alta resolución de frecuencia por lo tanto, darán como resultado una baja resolución de tiempo, que, al mismo tiempo significa una pérdida casi completa de coherencia vertical. De nuevo, el receptor final del sonido, es decir, el oído humano no se considera en tal modelo.

La publicación [22] da a conocer una metodología de análisis para extraer parámetros sinusoidales precisos de señales de audio. El método combina la estimación de parámetros de vocodificadores modificados con algoritmos de detección de pico utilizados actualmente en el moldeamiento sinusoidal. El sistema procesa la entrada cuadro a cuadro, busca picos similares al modelo de análisis sinusoidal pero también selecciona dinámicamente canales de vocodificadores a través de los cuales se procesan los picos borrosos en el dominio FFT. De esta forma, las trayectorias de los componentes sinusoidales de frecuencia cambiante dentro de un cuadro se pueden parametrizar en forma precisa. En una etapa de análisis espectral, se identifican picos y valles en la magnitud del FFT. En un aislamiento pico, el espectro se fija en cero al exterior del pico de interés y se retienen las versiones de frecuencia

positiva y negativa del pico. A continuación, se calcula la transformación Hilbert de este espectro y, posteriormente, se calcula el IFFT del espectro original y el espectro Hilbert transformado para obtener dos señales de dominio de tiempo, que están 90° fuera de fase entre sí. Se utilizan las señales para obtener la señal analítica utilizada en el análisis del vocodificador. Pueden detectarse picos espurios y se modelarán después como ruido o se excluirán del modelo.

De nuevo, el criterio perceptivo tal como un ancho de banda variante del oído humano sobre el espectro, es decir, tal como un ancho de banda pequeño en la parte inferior del espectro y un mayor ancho de banda en la parte superior del espectro no se consideran. Adicionalmente, una característica significativa del oído humano es que, como se discute en relación con la figura 9a, 9b y 9c el oído humano combina tonos sinusoidales dentro de un ancho de banda que corresponde al ancho de banda crítico del oído humano de tal manera que un ser humano no escucha dos tonos estables que tienen una pequeña diferencia de frecuencia pero percibe un tono que tiene una amplitud variante, en donde la frecuencia de este tono se posiciona entre las frecuencias de los tonos originales. Este efecto aumenta más y más cuando el ancho de banda crítico del oído humano aumenta.

Adicionalmente, el posicionamiento de las bandas críticas en el espectro no es constante, pero es dependiente de la señal. Los psicoacústicos han encontrado que el oído humano selecciona dinámicamente las frecuencias centrales de las bandas críticas dependiendo del espectro. Cuando, por ejemplo, el oído humano percibe un tono fuerte, entonces una banda crítica se centra alrededor de este tono fuerte. Cuando, después, se percibe un tono fuerte en una frecuencia diferente, entonces el oído humano posiciona una banda crítica alrededor de esta frecuencia diferente de tal manera que la percepción humana no solo se adapta por la señal a lo largo del tiempo sino que también filtra teniendo una alta resolución espectral en la porción de baja frecuencia y teniendo una resolución espectral baja, es decir, un ancho de banda alto en la parte superior del espectro.

La publicación "Speech analysis and synthesis using an AM-FM modulation model", A. Potamianos, P. Maragos, Speech Communication 28 (1999) 195-209 da a conocer un modelo de modulación AM-FM para análisis, síntesis y codificación de voz. Un sistema de análisis-síntesis de modulación AM-FM extrae tres o cuatro bandas de formantes que varían con el tiempo del espectro filtrando la señal de voz a lo largo de las pistas de formantes. Las pistas de formantes se obtienen a partir del algoritmo de rastreo de formantes de desmodulación de multibandas. La filtración se realiza por un banco de filtros Gabor con frecuencias centrales que varían con el tiempo que siguen las pistas de formantes. Las señales de resonancia se desmodulan a envoltura de amplitud y señales de frecuencia instantánea. Las señales de información se diezman. Las señales de información diezgadas se modelan y se codifican. Para sintetizar las señales de voz, la fase se obtiene respecto a la ejecución solidaria de la frecuencia instantánea y las bandas de formantes se reconstruyen a partir de las señales de fase y amplitud. La señal de voz sintética es la suma de las bandas de formantes reconstruidas.

Es el objeto de la presente invención proporcionar un concepto mejorado para parametrizar una señal de audio.

Este objeto se logra mediante un aparato para convertir una señal de audio según la reivindicación 1, un método para convertir una señal de audio según la reivindicación 9 o un programa de ordenador según la reivindicación 10.

La presente invención se basa en el hallazgo de que el ancho de banda variable de las bandas críticas se puede utilizar ventajosamente para diferentes propósitos. Un propósito es mejorar la eficiencia al utilizar la baja resolución del oído humano. En este contexto, la presente invención busca no calcular los datos cuando no se requieren datos con el fin de mejorar la eficiencia.

La segunda ventaja, sin embargo, es que, en la región, en donde se requiere una alta resolución, se calculan los datos necesarios con el fin de mejorar la calidad de una señal parametrizada y, de nuevo, una señal re-sintetizada.

La ventaja principal, sin embargo, es el hecho de que, este tipo de descomposición de señal proporciona un manejo para la manipulación de la señal en una forma directa, intuitiva y perceptivamente adaptada, por ejemplo, para abordar directamente propiedades como rugosidad, altura, etc.

Para este fin, se realiza un análisis adaptativo de señal de la señal de audio y, basado en los resultados del análisis, se estima una pluralidad de filtros de paso de banda en una forma de señal adaptativa. Específicamente, los anchos de banda de los filtros de paso de banda no son constantes, pero dependen de la frecuencia central del filtro de paso de banda. Por lo tanto, la presente invención permite variar las frecuencias de filtro de paso de banda y, adicionalmente, variar los anchos de banda del filtro de paso de banda, de tal manera que, se obtienen para cada señal de paso de banda perceptivamente correcta, una modulación de amplitud y una modulación de frecuencia junto con una frecuencia central de corriente, que aproximadamente es la frecuencia central de paso de banda calculada. A modo de ejemplo, el valor de frecuencia de la frecuencia central en una banda representa el centro de gravedad (COG) de la energía dentro de esta banda con el fin de modelar el oído humano tanto como sea posible. Así, un valor de frecuencia de una frecuencia central de un filtro de paso de banda no se selecciona necesariamente para estar sobre un tono específico en la banda, pero la frecuencia central de un filtro de paso de banda puede descansar

fácilmente sobre un valor de frecuencia, en donde no existe un pico en el espectro FFT.

La información de modulación de frecuencia se obtiene al mezclar la señal de paso de banda con la frecuencia central determinada. Así, aunque la frecuencia central se ha determinado con una baja resolución de tiempo debido a la determinación basada en FFT (basada en espectro), se guarda la información de tiempo instantánea en la modulación de frecuencia. Sin embargo, la separación de la variación de tiempo largo en una frecuencia portadora y la variación de tiempo corto en la información de modulación de frecuencia junto con la modulación de amplitud permite la representación parametrizada similar a vocodificador en un sentido perceptivamente correcto.

5 Así, la presente invención es ventajosa porque se satisface la condición de que la información extraída es perceptivamente significativa e interpretable en el sentido de que el procesamiento de modulación aplicado sobre la información de modulación debe producir resultados perceptivamente suaves evitando artefactos indeseados introducidos por las limitaciones de la representación de la modulación en sí misma.

15 Otra ventaja de la presente invención es que la información de portador extraída sola ya prevé una reconstrucción de "bosquejo" representativa y agradable perceptivamente pero áspera de la señal de audio y cualquier aplicación sucesiva de una información relacionada con AM y FM debe refinar esta representación hacia la transparencia y detalles completos, lo que significa que el concepto inventivo permite la escalabilidad completa desde una baja capa de escala que se basa en la reconstrucción del "bosquejo" que utiliza solo la información de portador extraída, que ya es perceptivamente agradable, hasta una alta calidad que utiliza capas de escala mayores adicionales que tienen la información relacionada con AM y FM en resolución de precisión/tiempo incrementada.

25 Una ventaja de la presente invención, de una parte, es que es altamente deseable para el desarrollo de nuevos efectos de audio y de otra parte como un elemento fundamental para algoritmos de compresión de audio eficientes futuros. Aunque, en el pasado, siempre ha habido una distinción entre los métodos de codificación paramétricos y la codificación en forma de ondas, esta distinción se puede obviar por la presente invención en un alto grado. Aunque los métodos de codificación de forma de onda cambian de escala fácilmente hasta transparencia dado que está disponible la velocidad de bits necesaria, esquemas de codificación paramétricos, tal como los esquemas CELP o ACELP se someten a las limitaciones de los modelos de fuentes subyacentes, y aunque la velocidad de bits se aumenta más y más en estos codificadores, ellos no pueden alcanzar transparencia. Sin embargo, los métodos paramétricos ofrecen usualmente un amplio rango de posibilidades de manipulación, que se pueden explotar para una aplicación de efectos de audio, aunque la codificación en forma de ondas se limita estrictamente a la mejor reproducción posible de la señal original.

35 La presente invención obviará este espacio al permitir una transición sin problemas entre ambos enfoques.

Posteriormente, las realizaciones de la presente invención se discuten en el contexto de los dibujos adjuntos, en los que:

40 la figura 1 es una representación esquemática de una realización de un aparato o método para convertir una señal de audio;

la figura 1b es una representación esquemática de otra realización preferida;

45 la figura 2a es un diagrama de flujo que ilustra un operación de procesamiento en el contexto de la realización de la figura 1a;

la figura 2b es un diagrama de flujo que ilustra el proceso de operación para generar la pluralidad de señales de paso de banda en un ejemplo;

50 la figura 2c ilustra una segmentación espectral adaptativa de señal basada en el cálculo COG y las restricciones perceptivas;

la figura 2d ilustra un diagrama de flujo que ilustra el proceso realizado en el contexto de la realización de la figura 1b;

55 la figura 3a ilustra una representación esquemática de una realización de un concepto para modificar la representación parametrizada;

la figura 3b ilustra un ejemplo del concepto ilustrado en la figura 3a;

60 la figura 3c ilustra una representación esquemática para explicar una descomposición de una información AM en una información de estructura fina y áspera;

la figura 3d ilustra un escenario de compresión basado en la realización de la figura 3c;

- la figura 4a ilustra una representación esquemática del concepto de síntesis;
- 5 la figura 4b ilustra un ejemplo del concepto de la figura 4a;
- la figura 4c ilustra una representación de un solapamiento de la señal de audio con dominio del tiempo procesada, un flujo de bits de la señal de audio y un procedimiento de solapamiento/suma para la síntesis de información de modulación;
- 10 la figura 4d ilustra un diagrama de flujo de un ejemplo para sintetizar una señal de audio que utiliza una representación parametrizada;
- la figura 5 ilustra una estructura de vocodificador de análisis/síntesis de la técnica anterior;
- 15 la figura 6 ilustra la implementación de filtro de la técnica anterior de la figura 5;
- la figura 7a ilustra un espectrograma de un elemento de música original;
- la figura 7b ilustra un espectrograma de solo los portadores sintetizados;
- 20 la figura 7c ilustra un espectrograma de portadores refinados mediante AM y FM áspero;
- la figura 7d ilustra un espectrograma de portadores refinados mediante AM y FM áspero;
- 25 la figura 7e ilustra un espectrograma de los portadores y síntesis después de AM y FM no procesado;
- la figura 8 ilustra un resultado de una prueba de calidad de audio subjetiva;
- la figura 9a ilustra una densidad espectral de potencia de una señal de 2 tonos, una señal de múltiples tonos y una señal de múltiples tonos limitada adecuadamente a la banda;
- 30 la figura 9b ilustra una forma de onda y envoltura de una señal de dos tonos, una señal de múltiples tonos y una señal de múltiples tonos limitada a banda apropiadamente; y
- 35 la figura 9c ilustra ecuaciones para generar dos señales equivalentes perceptivamente en un sentido del paso de banda.
- La figura 1 ilustra un aparato para convertir una señal 100 de audio en una representación 180 parametrizada. El aparato comprende un analizador 102 de señal para analizar una porción de la señal de audio para obtener un resultado de análisis 104. El resultado de análisis se ingresa en un estimador 106 de paso de banda para estimar la información sobre una pluralidad de filtros de paso de banda para la porción de señal de audio basada en el resultado de análisis de señal. Así, la información 108 en la pluralidad de filtros de paso de banda se calcula en una forma de señal adaptativa.
- 40
- 45 Específicamente, la información 108 en la pluralidad de filtros de paso de banda comprende información en una forma de filtro. La forma de filtro puede incluir un ancho de banda de un filtro de paso de banda y/o una frecuencia central del filtro de paso de banda para la porción de la señal de audio, y/o una forma espectral de una función de transferencia de magnitud en una forma paramétrica o en una forma no paramétrica. De manera importante, el ancho de banda de un filtro de paso de banda no es constante sobre el rango de frecuencia completo, pero depende de la frecuencia central del filtro de paso de banda. Preferiblemente, la dependencia es tal que el ancho de banda aumenta a frecuencias centrales mayores y se reduce a frecuencias centrales menores. Aún más preferiblemente, el ancho de banda de un filtro de paso de banda se determina en una escala completa y perceptivamente correcta, tal como la escala de Bark, de tal manera que el ancho de banda de un filtro de paso de banda siempre es dependiente del ancho de banda realizado actualmente por el oído humano para una cierta frecuencia central determinada adaptativamente
- 50
- 55 por señal.
- Para este fin, se prefiere que el analizador 102 de señal realice un análisis espectral de una porción de señal de la señal de audio y, particularmente, analice la distribución de potencia en el espectro para hallar regiones que tengan una concentración de potencia, debido a que tales regiones se determinan por el oído humano también cuando reciben
- 60 y procesan adicionalmente el sonido.
- El aparato de la invención comprende adicionalmente un estimador 110 de modulación para estimar una modulación 112 de amplitud o una modulación 114 de frecuencia para cada banda de la pluralidad de filtros de paso de banda para la porción de la señal de audio. Para este fin, el estimador 110 de modulación utiliza la información sobre la

pluralidad de filtros 108 de paso de banda como se discutirá adelante.

El aparato de la invención de la figura 1a comprende adicionalmente una interfaz 116 de salida para transmitir, almacenar o modificar la información en la modulación 112 de amplitud, la información de la modulación 114 de frecuencia o la información sobre la pluralidad de filtros 108 de paso de banda, que pueden comprender información de forma de filtro tal como los valores de las frecuencias centrales de los filtros de paso de banda para esta porción/bloque específico de la señal de audio u otra información como se discutió anteriormente. La salida es una representación 180 parametrizada tal como se ilustra en la figura 1a.

La figura 1b ilustra una realización preferida del estimador 110 de modulación y el analizador de señal 102 de la figura 1a y el estimador 106 de paso de banda de la figura 1a combinado en una única unidad, que se denomina "estimación de frecuencia portadora" en la figura 1b. El estimador 110 de modulación preferiblemente comprende un filtro 110a de paso de banda, que proporciona una señal de paso de banda. Esta ingresa en un convertidor 110b de señal analítico. La salida del bloque 110b es útil para calcular la información AM y la información FM. Para calcular la información AM, la magnitud de la señal analítica se calcula mediante el bloque 110c. La salida del bloque de señal analítica 110b ingresa a un multiplicador 110d, que recibe, en su otra entrada, una señal osciladora de un oscilador 110e, que se controla mediante la frecuencia portadora actual  $f_c$  del paso de banda 110a. Luego, la fase de salida de multiplicador se determina en el bloque 110f. La fase instantánea se diferencia en el bloque 110g con el fin de obtener finalmente la información FM.

Así, la descomposición en señales portadoras y sus componentes de modulaciones asociadas se ilustra en la figura 1b.

En la fotografía se muestra el flujo de señal para la extracción de un componente. Todos los otros componentes se obtienen de forma similar. La extracción se lleva a cabo preferiblemente sobre una base bloque a bloque utilizando un tamaño de bloque de  $N = 2^{14}$  en una frecuencia de muestreo de 48 KHz y solapamiento de  $\frac{3}{4}$  que corresponde aproximadamente a un intervalo de tiempo de 340 ms y un progreso de 85 ms. Observe que también se pueden utilizar otros factores de solapamiento o tamaños de bloque. Esto consiste en un filtro de paso de banda adaptativo de señal que se centra en un COG local [12] en el espectro DFT de la señal. Se estiman candidatos COG locales al buscar transiciones positivas a negativas en la función CogPos definida en (3). Un procedimiento de post-selección asegura que las posiciones COG estimadas finales son aproximadamente equidistantes a una escala perceptiva.

$$\begin{aligned}
 CogPos(k, m) &= \frac{nom(k, m)}{denom(k, m)} \\
 nom(k, m) &= \alpha \sum_{i=-B(k)/2}^{+B(k)/2} \left( iw(i) |X(k+i, m)|^2 \right) + (1-\alpha) nom(k, m-1) \\
 denom(k, m) &= \alpha \sum_{i=-B(k)/2}^{+B(k)/2} \left( w(i) |X(k+i, m)|^2 \right) + (1-\alpha) denom(k, m-1) \\
 \alpha &= \frac{1}{\tau F_s}; i \in \mathbb{Z}
 \end{aligned} \quad (3)$$

Para cada índice de coeficiente espectral  $k$  se produce el desfase relativo hacia el centro de gravedad local en la región espectral que está cubierta por una ventana deslizante suave  $w$ . El ancho  $B(k)$  de la ventana sigue una escala perceptiva, por ejemplo, la escala de Bark.  $X(k, m)$  es el coeficiente espectral  $k$  en el bloque de tiempo  $m$ . Adicionalmente, se hace un suavizado temporal recursivo de primer orden con la constante de tiempo  $\tau$ .

Alternativamente se puede concebir las funciones que calculan el valor de centro de gravedad, que pueden ser iterativas o no iterativas. Una función no iterativa por ejemplo incluye agregar valores de energía para diferentes porciones de una banda y comparar los resultados de la operación de adición con porciones diferentes.

El COG local corresponde a la frecuencia "media" que es percibida por un oyente humano debido a la contribución espectral en esa región de frecuencia. Para ver esta relación observe la equivalencia del COG y "frecuencia instantánea de intensidad media ponderada" (IWAIF) como se deriva en [12]. La ventana de estimación COG y el ancho de banda de transición del filtro resultante se seleccionan con respecto a la resolución del oído humano ("bandas críticas"). Aquí, un ancho de banda de aproximadamente 0,5 Bark se encontró empíricamente que es un valor bueno para todas las clases de elementos de prueba (voz, música, ambiente). Adicionalmente, esta elección está soportada por la bibliografía [13].

Posteriormente, se obtiene la señal analítica utilizando la transformación Hilbert de la señal filtrada de paso de banda y se heterodina mediante la frecuencia COG estimada. Finalmente la señal se descompone adicionalmente en su envoltura de amplitud y su pista de frecuencia instantánea (IF) que produce las señales AM y FM deseadas. Observe

que el uso de las señales de paso de banda centradas en las posiciones COG locales corresponden al paradigma de "regiones de influencia" de un vocodificador de fase tradicional. Ambos métodos conservan la envoltura temporal de una señal de paso de banda. La primera intrínsecamente y la última al asegurar coherencia de fase espectral local.

5 Se debe tener cuidado de que el conjunto de filtros resultante de una parte cubre el espectro sin problemas y de otra parte los filtros adyacentes no se solapan mucho debido a que esto resultará en efectos de pulsación indeseados después de la síntesis de los componentes (modificados). Esto implica algún compromiso con respecto al ancho de banda de los filtros que siguen a una escala perceptiva pero, al mismo tiempo, tienen que proporcionar cubrimiento espectral sin problemas. Así el diseño de filtro adaptativo de señal y la estimación de frecuencia de portador llegan a ser partes cruciales para la significancia perceptiva de los componentes de descomposición y así tienen una fuerte influencia en la calidad de la señal resintetizada. Un ejemplo de tal segmentación compensada se muestra en la figura 2c.

15 La figura 2a ilustra un proceso preferido para convertir una señal de audio en una representación parametrizada como se ilustra en la figura 2b. En una primera etapa 120, se forman bloques de muestras de audio. Para este fin, se utiliza preferiblemente una función de ventana. Sin embargo, el uso de una función de ventana no es necesario en ningún caso. A continuación, en la etapa 121, se desarrolla la conversión espectral en un espectro de resolución de alta frecuencia 121. Luego, en la etapa 122, la función de centro de gravedad se calcula preferiblemente utilizando la ecuación (3). Este cálculo se realizará en el analizador 102 de señal y los cruces por cero determinados posteriormente serán el resultado 104 del análisis proporcionado desde el analizador 102 de señal de la figura 1a al estimador 106 de paso de banda de la figura 1a.

25 Como se puede ver a partir de la ecuación (3), la función de centro de gravedad se calcula con base en diferentes anchos de banda. Específicamente, el ancho de banda  $B(k)$ , que se utiliza en el cálculo numerador  $\text{num}(k,m)$  y el denominador  $(k,m)$  en la ecuación (3) es dependiente de la frecuencia. El índice de frecuencia  $k$ , por lo tanto, determina el valor de  $B$  y, aún más preferiblemente, el valor de  $B$  aumenta a un índice de frecuencia aumentado  $k$ . Por lo tanto, como es claro en la ecuación (3) para  $\text{num}(k,m)$ , una "ventana que tiene el ancho de ventana  $B$  en el dominio espectral se centra alrededor de un cierto valor de frecuencia  $k$ , en donde  $i$  corre de  $-B(k)/2$  a  $+B(k)/2$ .

30 Este índice  $i$ , que se multiplica a una ventana  $w(i)$  en el término  $\text{num}$  asegura que el valor de potencia espectral  $X^2$  (en donde  $X$  es una amplitud espectral) a la izquierda del valor de frecuencia actual  $k$  entra en la operación de suma con un signo negativo, aunque los valores espectrales cuadrados a la derecha del índice de frecuencia  $k$  entran en la operación de suma con el signo positivo. Naturalmente, esta función puede ser diferente, de tal manera que, por ejemplo, la mitad superior entra con un signo negativo y la mitad inferior entra con un signo positivo. La función  $B(k)$  asegura que tiene lugar un cálculo perceptivamente correcto de un centro de gravedad, y esta función se determina preferiblemente, por ejemplo como se ilustra en la figura 2c, en donde se ilustra una segmentación espectral perceptivamente correcta.

40 En una implementación alternativa, los valores espectrales  $X(k)$  se transforman en un dominio logarítmico antes de calcular la función de centro de gravedad. A continuación, el valor  $B$  en el término para el numerador y el denominador en la ecuación (3) es independiente de la frecuencia (escala logarítmica). En el presente documento, la dependencia perceptivamente correcta ya se incluye en los valores espectrales de  $X$ , que están, en este ejemplo, presentes en la escala logarítmica. Naturalmente, un ancho de banda igual en una escala logarítmica corresponde a un ancho de banda incrementado con respecto a la frecuencia central en una escala no logarítmica.

45 Tan pronto como cruza por el cero y, específicamente, se calculan las transiciones de positivo a negativo en la etapa 122, se realiza el procedimiento de selección posterior en la etapa 124. Aquí, los valores de frecuencia en los cruces por cero se modifican con base en el criterio perceptivo. Esta modificación tiene varias restricciones, que son, que se cubre preferiblemente el espectro completo y preferiblemente no se permiten todos los espectros. Adicionalmente, las frecuencias centrales de los filtros de paso de banda se posicionan en los cruces por cero de la función de centro de gravedad en la medida de lo posible y, preferiblemente, el posicionamiento de las frecuencias centrales en la porción más baja del espectro se favorece con respecto al posicionamiento en la porción mayor del espectro. Esto significa que la segmentación espectral adaptativa de señal trata de seguir los resultados de centro de gravedad de la etapa 122 en la porción inferior del espectro más estrechamente y cuando, basado en esta determinación, el centro de gravedad en la porción mayor del espectro no coincide con las frecuencias centrales de paso de banda, se acepta este desfase.

50 Tan pronto como los valores de frecuencia central y las amplitudes correspondientes de los filtros de paso de banda se determinan, se filtra el bloque de señal 126 de audio con el banco de filtro que tiene filtros de paso de banda con anchos de banda variantes en los valores de frecuencia modificados como se obtienen mediante la etapa 124. Así, con respecto al ejemplo en la figura 2c, un banco de filtro como se ilustra en la segmentación espectral adaptativa de señal se aplica al calcular coeficientes de filtro y configurar estos coeficientes de filtro, y el banco de filtro se utiliza posteriormente para filtrar la porción de la señal de audio que se ha utilizado para calcular estas segmentaciones espectrales.

Esta filtración se realiza preferiblemente con un banco de filtro o una transformación de frecuencia de tiempo tal como una ventana DFT e IDFT ponderada espectral posterior, en donde se ilustra un filtro de paso de banda sencillo en 110a y los filtros de paso de banda para los otros componentes 101 forman el banco de filtro junto con el filtro de paso de banda 110a. Basado en las señales de sub-banda  $\tilde{x}$ , la información AM y la información FM, es decir, 112, 114 se calculan en la etapa 128 y salen junto con la frecuencia portadora para cada paso de banda como la representación parametrizada del bloque de valores de muestreo de audio.

A continuación, se completa el cálculo para un bloque y en la etapa 130, se aplica un valor; de avance o progreso en el dominio de tiempo en una forma sobrepuesta con el fin de obtener el siguiente bloque de muestras de audio como se indica por 120 en la figura 2a.

Este procedimiento se ilustra en la figura 4c. Se ilustra la señal de audio con dominio del tiempo en la parte superior en donde se ilustran siete porciones a modo de ejemplo, comprendiendo cada porción preferiblemente el mismo número de muestras de audio. Cada bloque consiste en N muestras. El primer bloque 1 consiste en las primeras cuatro porciones adyacentes 1, 2, 3 y 4. El siguiente bloque 2 consiste en las porciones de señal 2, 3, 4, 5, el tercer bloque, es decir, el bloque 3 comprende las porciones de señal 3, 4, 5, 6 y el cuarto bloque, es decir, el bloque 4 comprende las porciones de señal posteriores 4, 5, 6 y 7 como se ilustra. En el flujo de bits, la etapa 128 de la figura 2a genera una representación parametrizada para cada bloque, es decir, para el bloque 1, bloque 2, bloque 3, bloque 4 o una parte seleccionada del bloque, preferiblemente la porción central  $N/2$  ya que las porciones externas pueden contener un filtro que timbra o el balanceo espontáneo característico de una ventana de transformación que se diseña en consecuencia. Preferiblemente, la representación parametrizada para cada bloque se transmite en un flujo de bits en una forma secuencial. En el ejemplo ilustrado en el gráfico superior de la figura 4c, se forma una operación de solapamiento de 4 veces. Alternativamente, se puede realizar un solapamiento de dos veces también de tal manera que el valor de progreso o el valor de avance aplicado en la etapa 130 tiene dos porciones en la figura 4c en lugar de una porción. Básicamente no es necesario una operación de solapamiento para todos pero se prefiere con el fin de evitar artefactos que bloqueen y con el fin de permitir ventajosamente una operación de atenuación cruzada de bloque a bloque, que, según una realización preferida de la presente invención, no se realiza en el dominio de tiempo sino que se realiza en el dominio AM./FM como se ilustra en la figura 4c, y como se describe después con respecto a la figura 4a y 4b.

La figura 2b ilustra una implementación general del procedimiento específico en la figura 2a con respecto a la ecuación (3). Este procedimiento en la figura 2b se desarrolla parcialmente en el analizador de señal y el estimador de paso de banda. En la etapa 132, se analiza una porción de la señal de audio con respecto a la distribución de potencia espectral. La etapa 132 puede implicar una transformación de tiempo/frecuencia. En una etapa 134, los valores de frecuencia estimados para las concentraciones de potencia local en el espectro se adaptan para obtener una segmentación espectral perceptivamente correcta tal como la segmentación espectral en la figura 2c, que tiene anchos de banda perceptivamente motivados de los filtros de paso de banda diferentes y que no tiene ningún orificio en el espectro. En la etapa 135, la porción de la señal de audio se filtra con la segmentación espectral determinada que utiliza el banco de filtro o un método de transformación, en donde un ejemplo para una implementación para banco de filtro se da en la figura 1b por un canal que tiene un paso 110a de banda y que corresponde a filtros de paso de banda para los demás componentes 101 en la figura 1b. El resultado de la etapa 135 es una pluralidad de señales de paso de banda para las bandas que tienen un aumento de ancho de banda a frecuencias mayores. Luego, en la etapa 136, cada señal de paso de banda se procesa separadamente utilizando los elementos 110a a 110g en la realización preferida.

Posteriormente, se discutirá la figura 2d, en la que se ilustra una secuencia preferida de etapas para procesar en forma separada cada señal de paso de banda. En una etapa 138, se configura un filtro de paso de banda utilizando el valor de frecuencia central calculada y utilizando un ancho de banda como se determina por la segmentación espectral como se obtiene en la etapa 134 de la figura 2b. Esta etapa utiliza información de filtro de paso de banda y también se puede utilizar para generar información de filtros de paso de banda a la interfaz 116 de salida en la figura 1a. En la etapa 139, la señal de audio se filtra utilizando el filtro de paso de banda establecido en la etapa 138. En la etapa 140, se forma una señal analítica de la señal de paso de banda. Aquí, se puede aplicar la transformación verdadera Hilbert o un algoritmo de transformación Hilbert aproximado. Esto se ilustra por el elemento 110b en la figura 1b. Luego, en la etapa 141, se realiza la implementación de la caja 110c, es decir, se determina la magnitud de la señal analítica con el fin de proporcionar la información AM. Básicamente, se obtiene la información AM en la misma resolución que la resolución de la señal de paso de banda en la salida del bloque 110a. Con el fin de comprimir esta gran cantidad de información AM, se pueden realizar cualquier técnica de parametrización o decimación que se discutirán después.

Con el fin de obtener información de frecuencia o de fase, la etapa 142 comprende una multiplicación de la señal analítica mediante una señal osciladora que tiene la frecuencia central del filtro de paso de banda. En el caso de una multiplicación, se prefiere una posterior operación de filtro de paso bajo para rechazar la porción de frecuencia alta generada por la multiplicación en la etapa 142. Cuando la señal osciladora es compleja, entonces, no se requiere filtración. La etapa 142 da como resultado una señal analítica mezclada baja, que se procesa en la etapa 143 para

extraer la información de fase instantánea como se indica por la caja 110f en la figura 1b. Esta información de fase puede generarse como información paramétrica además de la información AM, pero se prefiere diferenciar esta información fase en la caja 144 para obtener una información de modulación de frecuencia verdadera como se ilustra en la figura 1b en 114. De nuevo, puede utilizarse la información de fase para describir las fluctuaciones relacionadas con frecuencia/fase. Cuando la información de fase como información de parametrización es suficiente, entonces no es necesaria la diferenciación en el bloque 110g.

La figura 3a ilustra un aparato para modificar una representación parametrizada de una señal de audio que tiene, para una porción de tiempo, información de filtro de paso de banda desde una pluralidad de filtros de paso de banda, tal como el bloque 1 en el gráfico en el centro de la figura 4c. La información de filtro de paso de banda indica frecuencias centrales de filtro de paso de banda de tiempo variante (frecuencias portadoras) de filtros de paso de banda que tienen anchos de banda que dependen de los filtros de paso de banda y las frecuencias de los filtros de paso de banda, y que tienen modulación de amplitud o modulación de fase o información de modulación de frecuencia para cada filtro de paso de banda para la porción de tiempo respectiva. El aparato para modificar comprende un modificador 160 de información que es operativo para modificar las frecuencias centrales de tiempo variante y para modificar la información de modulación de amplitud o la información de modulación de frecuencia o la información de modulación de fase y que genera una representación parametrizada modificada que tiene frecuencias portadoras para una porción de señal de audio, información AM modificada, información PM modificada o información FM modificada.

La figura 3b ilustra un ejemplo del modificador 160 de información en la figura 3a. En el ejemplo, la información AM se introduce en una etapa de descomposición para descomponer la información AM en una estructura de escala áspera/fina. Esta descomposición es, a modo de ejemplo, una descomposición no lineal tal como la descomposición que se ilustra en la figura 3c. Con el fin de comprimir los datos transmitidos para la información AM, solo la estructura áspera, por ejemplo, se transmite a un sintetizador. Una porción de este sintetizador puede ser el sumador 160e y la fuente 160f de ruido de paso de banda. Sin embargo, estos elementos también pueden ser parte del modificador de información. En el ejemplo, sin embargo, una ruta de transmisión está entre el bloque 160a y 160e, y en este canal de transmisión, solo una representación parametrizada de la estructura áspera, y, por ejemplo, un valor de energía que representa o se deriva de la estructura fina se transmiten a través de la línea 161 desde un analizador hasta un sintetizador. A continuación, en el lado del sintetizador, se cambia la escala que la fuente 160f de ruido con el fin de proporcionar una señal de ruido de paso de banda para una señal de paso de banda específica, y la señal de ruido tiene una energía que se indica a través de un parámetro tal como el valor de energía en la línea 161. Luego, en el lado del decodificador /sintetizador, el ruido se forma temporalmente mediante la estructura áspera, se pondera por su energía objetivo y se agrega a la estructura áspera transmitida con el fin de sintetizar una señal que solo requiere una baja velocidad de bits para transmisión debido a la síntesis artificial de la estructura fina. En general, el sumador 160f de ruido es para agregar una señal de ruido (pseudoaleatoria) que tiene un cierto valor de energía global y una distribución de energía temporal predeterminada. Se controla a, través de información lateral transmitida o se establece de manera fija por ejemplo basado en una figura empírica tal como valores fijos determinados para cada banda. Alternativamente se controla mediante un análisis local en el modificador o el sintetizador, en el que se analiza la señal disponible y se derivan los valores de control del sumador de ruido. Estos valores de control son, a modo de ejemplo, valores relacionados con la energía.

El modificador 160 de información puede, adicionalmente, comprender una funcionalidad 160b de ajuste polinomial restrictiva y/o un reemisor 160d para las frecuencias portadoras, que también transponen la información FM a través del multiplicador 160c. Alternativamente, también puede ser útil modificar solo las frecuencias portadoras y no modificar la información FM o la información AM o solo modificar la información FM pero no modificar la información AM o la información de frecuencia portadora.

La figura 3c ilustra un ejemplo para extraer una estructura áspera de una señal de paso de banda. La figura 3c ilustra una estructura áspera típica para un tono producido mediante un cierto instrumento en el gráfico superior. Al inicio, el instrumento está en silencio, luego un momento de tiempo de ataque, puede observarse un fuerte aumento de la amplitud, que luego se mantiene constante en un denominado periodo de sostenimiento. Luego, el tono se libera. Esto se caracteriza por una clase de un decaimiento exponencial que empieza al final del periodo sostenido. Este es el principio del periodo de liberación, es decir, un momento de tiempo de liberación. El periodo de sostenimiento no está necesariamente en instrumentos. Cuando, por ejemplo, se considera una guitarra, llega a ser claro que el tono se genera al excitar una cuerda y después el ataque en el momento de tiempo de excitación, una porción de liberación, que es bastante larga, sigue inmediatamente la cual se caracteriza por el hecho de que la oscilación de la cuerda se amortigua hasta que la cuerda llega a un estado estacionario que, luego, es el fin del tiempo de liberación. Para instrumentos típicos, hay formas típicas o estructuras ásperas para tales tonos. Con el fin de extraer tales estructuras ásperas de una señal de paso de banda, se puede realizar un ajuste polinomial en la señal de paso de banda, en donde el ajuste polinomial tiene una forma general similar a la forma en el gráfico superior de la figura 3c, que se puede igualar al determinar los coeficientes polinomiales. Tan pronto como se obtiene un mejor ajuste polinomial, la señal se determina por la carga polinomial, que es la estructura áspera de la señal de paso de banda que se resta de la señal de paso de banda actual de tal manera que se obtiene la estructura fina que, cuando el ajuste polinomial es suficientemente bueno, es una señal bastante ruidosa que tiene una cierta energía que se puede transmitir desde el

lado del analizador hasta el lado del sintetizador además de la información de estructura áspera que serían los coeficientes polinomiales. La descomposición de la señal de paso de banda en su estructura áspera y su estructura fina es un ejemplo para una descomposición no lineal. Pueden realizarse otras composiciones lineales también con el fin de extraer otras características de la señal de paso de banda con el fin de reducir grandemente la velocidad de datos para transmitir información AM en una aplicación de baja velocidad de bits.

La figura 3d ilustra las etapas en tal procedimiento. En una etapa 165, la estructura áspera se extrae tal como mediante ajuste polinomial y mediante cálculo de los parámetros polinomiales que hay, entonces, la información de modulación de amplitud que va a transmitirse desde un analizador hasta un sintetizador. Con el fin de realizar más eficientemente esta transmisión, se realiza una operación 166 de codificación y cuantificación adicional de los parámetros para transmisión. La cuantificación puede ser uniforme o no uniforme, y la operación de codificación puede ser cualquiera de las operaciones de codificación de entropía bien conocidas, tal como codificación Huffman, con o sin tablas o codificación aritmética tal como codificación aritmética basada en contexto como la conocida de la compresión de videos.

Luego, una información AM de baja velocidad de bits o información FM/PM se forma la cual se puede transmitir sobre un canal de transmisión en una forma muy eficiente. Sobre un lado del sintetizador, se realiza una etapa 168 para decodificar y descuantificar los parámetros transmitidos. Luego, en una etapa 169, se reconstruye la estructura áspera, por ejemplo, al calcular realmente todos los valores definidos mediante un polinomio que tiene coeficientes polinomiales transmitidos. Adicionalmente, puede ser útil agregar ruido de adorno por banda a modo de ejemplo basado en los parámetros de energía transmitidos y con forma temporal mediante la información AM áspera, o, alternativamente, en una aplicación de ultra velocidad de bit, al agregar ruido (adorno) que tiene una energía empíricamente seleccionada.

Alternativamente, una modificación de señal puede incluir, como se discutió anteriormente, un mapeo de las frecuencias centrales para números MIDI, o, generalmente, a una escala musical y para transformar luego la escala con el fin de, por ejemplo, transformar una pieza de música que está a una escala mayor a una escala menor o viceversa. En este caso, de manera más importante, las frecuencias portadoras se modifican. A modo de ejemplo, la información AM o la información PM/FM no se modifica en este caso.

Alternativamente, otras clases de modificaciones de frecuencia portadoras pueden realizarse tal como la transposición de todas las frecuencias portadoras que utilizan el mismo factor de transposición que puede ser un número entero mayor que 1 o que puede ser un número fraccionario entre 1 y 0. En este último caso, la altura de los tonos será menor después de la modificación, y en el primer caso, la altura de los tonos será mayor después de la modificación que antes de la modificación.

La figura 4a ilustra un aparato para sintetizar una representación parametrizada de una señal de audio comprendiendo la representación parametrizada información de paso de banda tal como frecuencias portadoras o frecuencias centrales de paso de banda para los filtros de paso de banda. Los componentes adicionales de la representación parametrizada son la información sobre una modulación de amplitud, información sobre una modulación de frecuencia o información sobre una modulación de fase de una señal de paso de banda

Con el fin de sintetizar una señal, el aparato para sintetizar comprende una interfaz 200 de entrada que recibe una representación parametrizada modificada o no modificada que incluye información para todos los filtros de paso de banda. A modo de ejemplo, la figura 4a ilustra los módulos de síntesis para una única señal de filtro de paso de banda. Con el fin de sintetizar información AM, se proporciona un sintetizador AM 201 para sintetizar un componente AM basado en la modulación AM. Adicionalmente, también se proporciona un sintetizador FM/PM para sintetizar una frecuencia instantánea o información de fase basada en la información de las frecuencias portadoras y la información de modulación PM o FM. Ambos elementos 201, 202 se conectan a un módulo oscilador para generar una señal de salida, que es la señal 204 de oscilación modulada AM/FM/PM para cada canal de banco de filtro. Adicionalmente, se proporciona un combinador 205 para combinar señales de los canales de filtro de paso de banda, tal como las señales 204 de los osciladores para otros canales de filtro de paso de banda y para generar una señal de salida de audio que se basa en la señales de los canales de filtro de 'paso de banda. Tan pronto acaba de agregar señales de paso de banda en forma de muestra en un ejemplo, genera la señal 206 de audio sintetizada. Sin embargo, también se pueden utilizar otros métodos de combinación.

La figura 4b ilustra un ejemplo del sintetizador de la figura 4a. Una implementación ventajosa se basa en una operación de adición de solapamiento (OLA) en el dominio de modulación, es decir, en el dominio antes de generar la señal de paso de banda de dominio en el tiempo. Como se ilustra en el centro del gráfico de la figura 4c, la señal de entrada que puede ser un flujo de bits, pero que también puede ser una conexión directa a un analizador o modificador también, se separa en el componente 207a AM, el componente 207b FM, y el componente 207c de frecuencia portadora. El sintetizador 201 AM comprende a modo de ejemplo un sumador 201a de solapamiento y, adicionalmente, un controlador 201b de conexión de componente que, a modo de ejemplo no solo comprende el bloque 201a sino también el bloque 202a, que tiene un sumador de solapamiento dentro del sintetizador 202 FM. El sintetizador 202 FM

comprende adicionalmente un sumador 202a de solapamiento de frecuencia, un integrador 202b de fase, un combinador 202c de fase que, de nuevo, se puede implementar como un sumador regular y un cambiador 202d de fase que se puede controlar mediante el controlador 201b de conexión de componente con el fin de regenerar una fase constante de bloque a bloque de tal manera que la fase de una señal de un bloque anterior es continua con la fase de un bloque actual. Por lo tanto, puede decirse que la adición de fase en elementos 202d, 202c corresponde a una regeneración de una constante que se pierde durante la diferenciación en el bloque 110g en la figura 1b en el lado del analizador. Desde una perspectiva de pérdida de información en el dominio perceptivo, debe observarse que esta es la única pérdida de información, es decir, la pérdida de una porción constante mediante el dispositivo 110g de diferenciación en la figura 1b. Esta pérdida se recrea al agregar una fase constante determinada mediante el dispositivo 201b de conexión de componente en la figura 4b.

La señal se sintetiza sobre una base aditiva de todos los componentes. Para un componente la cadena de procesamiento se muestra en la figura 4b. Al igual que el análisis, la síntesis se realiza sobre una base bloque a bloque. Debido a que solo se utiliza la porción  $N/2$  centrada de cada bloque de análisis para síntesis, resulta un factor de solapamiento de  $\frac{1}{2}$ . Se utiliza un mecanismo de conexión de componente para mezclar AM y FM y alinear la fase absoluta para los componentes en la vecindad espectral de sus predecesores en un bloque previo. La vecindad espectral también se calcula sobre una base de escala de Bark para reflejar de nuevo la sensibilidad del oído humano con respecto a la percepción de la altura.

En primer lugar se agregan detalles a la señal FM a la frecuencia portadora y el resultado se pasa a la etapa de adición de solapamiento (OLA). Luego se integra para obtener la fase del componente que va a sintetizarse. Se carga un oscilador sinusoidal mediante la señal de fase resultante. La señal AM se procesa de la misma manera mediante otra etapa OLA. Finalmente se modula la salida del oscilador en su amplitud mediante la señal AM resultante para obtener la contribución aditiva del componente para la señal de salida.

La figura 4c, el bloque inferior muestra una implementación a modo de ejemplo de la operación de adición de solapamiento en el caso de solapamiento del 50%. En esta implementación, la primera parte de la información utilizada realmente a partir del bloque actual se agrega a la parte correspondiente que es la segunda, parte de un bloque anterior. Adicionalmente, la figura 4c, el bloque inferior, ilustra una operación de atenuación cruzada, en donde la porción del bloque que es atenuada recibe pesos reducidos desde 1 hasta 0 y, al mismo tiempo, el bloque que va a atenuarse recibe pesos crecientes desde 0 hasta 1. Estos pesos se pueden aplicar ya sobre el lado del analizador y, a continuación, solo es necesaria una operación sumadora sobre el lado del decodificador. Sin embargo, a modo de ejemplo, esos pesos no se aplican sobre el lado del codificador sino que se aplican sobre el lado del decodificador de manera predefinida. Como se discutió anteriormente, solo la porción  $N/2$  centrada de cada bloque de análisis se utiliza para la síntesis de tal manera que resulta un factor de solapamiento de  $1/2$  como se ilustra en la figura 4c. Sin embargo, también puede utilizarse la porción completa de cada bloque de análisis para solapamiento/suma de tal manera que se ilustra un solapamiento de cuatro veces como se ilustra en la porción superior de la figura 4c. La realización descrita, en la que se utiliza la parte central, se prefiere, ya que los cuartos externos incluyen balanceo espontáneo de la ventana de análisis y los cuartos centrales solo tienen una porción superior plana.

Todas las demás relaciones de solapamiento se pueden implementar como puede ser el caso.

La figura 4d ilustra una secuencia a modo de ejemplo de etapas que van a realizarse dentro de la realización preferida de las figuras 4a/4b. En una etapa 170, dos bloques adyacentes de información AM se mezclan y tienen atenuación cruzada. A modo de ejemplo, esta operación de atenuación cruzada se realiza en el dominio de parámetro de modulación a diferencia de en el dominio de señal de tiempo de paso de banda modulada, sintetizada fácilmente. Así, artefactos de pulsación entre dos señales que van a mezclarse se anulan cuando se comparan con el caso, en que se realizaría la atenuación cruzada se dominio del tiempo y no en el dominio de parámetro de modulación. En la etapa 171, se calcula una frecuencia absoluta para un cierto instante al combinar la frecuencia portadora en forma de bloques para una señal de paso de banda con la información FM de resolución fina utilizando el sumador 202c. Luego, en la etapa 171, dos bloques adyacentes de información de frecuencia absoluta se mezclan y tienen atenuación cruzada con el fin de obtener una frecuencia instantánea mezclada en la salida del bloque 202a. En la etapa 173, el resultado de la operación 202a OLA se integra como se ilustra en el bloque 202b en la figura 4b. Adicionalmente, la operación 201b de conexión de componente determina la fase absoluta de una frecuencia predecesora correspondiente en un bloque anterior como se ilustra en 174. Con base en la fase determinada, el conmutador 202d de fase se recrea al agregar una fase constante determinada por el dispositivo de conexión de componente 201b la figura 4b.

La señal se sintetiza sobre una base aditiva de todos los componentes. Para un componente la cadena de procesamiento se muestra en la figura 4b. Al igual que el análisis, la síntesis se realiza sobre una base bloque a bloque. Debido a que solo se utiliza la porción  $N/2$  centrada de cada bloque de análisis para síntesis, resulta un factor de solapamiento de  $\frac{1}{2}$ . Se utiliza un mecanismo de conexión de componente para mezclar AM y FM y alinear la fase absoluta para los componentes en la vecindad espectral de sus predecesores en un bloque previo. La vecindad espectral también se calcula sobre una base de escala de Bark para reflejar de nuevo la sensibilidad del oído humano con respecto a la percepción de la altura.

En primer lugar se agregan detalles a la señal FM a la frecuencia portadora y el resultado se pasa a la etapa de adición de solapamiento (OLA). Luego se integra para obtener la fase del componente que va a sintetizarse. Se carga un oscilador sinusoidal mediante la señal de fase resultante. La señal AM se procesa de la misma manera mediante otra etapa OLA. Finalmente se modula la salida del oscilador en su amplitud mediante la señal AM resultante para obtener la contribución aditiva del componente para la señal de salida.

La figura 4c, el bloque inferior muestra una implementación preferida de la operación de adición de solapamiento en el caso de solapamiento del 50%. En esta implementación, la primera parte de la información utilizada realmente a partir del bloque actual se agrega a la parte correspondiente que es la segunda, parte de un bloque anterior. Adicionalmente, la figura 4c, el bloque inferior, ilustra una operación de atenuación cruzada, en donde la porción del bloque que es atenuada recibe pesos reducidos desde 1 hasta 0 y, al mismo tiempo, el bloque que va a atenuarse recibe pesos crecientes desde 0 hasta 1. Estos pesos se pueden aplicar ya sobre el lado del analizador y, a continuación, solo es necesaria una operación sumadora sobre el lado del decodificador. Sin embargo, preferiblemente, esos pesos no se aplican sobre el lado del codificador sino que se aplican sobre el lado del decodificador de manera predefinida. Como se discutió anteriormente, solo la porción N/2 centrada de cada bloque de análisis se utiliza para la síntesis de tal manera que resulta un factor de solapamiento de 1/2 como se ilustra en la figura 4c. Sin embargo, también puede utilizarse la porción completa de cada bloque de análisis para solapamiento/suma de tal manera que se ilustra un solapamiento de cuatro veces como se ilustra en la porción superior de la figura 4c. La realización descrita, en la que se utiliza la parte central, se prefiere, ya que los cuartos externos incluyen balanceo espontáneo de la ventana de análisis y los cuartos centrales solo tienen una porción superior plana.

Todas las demás relaciones de solapamiento se pueden implementar como puede ser el caso.

La figura 4d ilustra una secuencia preferida de etapas que van a realizarse dentro de la realización preferida de las figuras 4a/4b. En una etapa 170, dos bloques adyacentes de información AM se mezclan y tienen atenuación cruzada. Preferiblemente, esta operación de atenuación cruzada se realiza en el dominio de parámetro de modulación a diferencia de en el dominio de señal de tiempo de paso de banda modulada, sintetizada fácilmente. Así, artefactos de pulsación entre dos señales que van a mezclarse se anulan cuando se comparan con el caso, en que se realizaría la atenuación cruzada se dominio del tiempo y no en el dominio de parámetro de modulación. En la etapa 171, se calcula una frecuencia absoluta para un cierto instante al combinar la frecuencia portadora en forma de bloques para una señal de paso de banda con la información FM de resolución fina utilizando el sumador 202c. Luego, en la etapa 171, dos bloques adyacentes de información de frecuencia absoluta se mezclan y tienen atenuación cruzada con el fin de obtener una frecuencia instantánea mezclada en la salida del bloque 202a. En la etapa 173, el resultado de la operación 202a OLA se integra como se ilustra en el bloque 202b en la figura 4b. Adicionalmente, la operación 201b de conexión de componente determina la fase absoluta de una frecuencia predecesora correspondiente en un bloque anterior como se ilustra en 174. Con base en la fase determinada, el conmutador 202d de fase de la figura 4b ajusta la fase absoluta de la señal mediante la adición de un  $\phi_0$  adecuado en el bloque 202c que también se ilustra por la etapa 175 en la figura 4d. Ahora, la fase ya está lista para el control de fase de un oscilador sinusoidal como se indica en la etapa 176. Finalmente, la señal de salida del oscilador tiene amplitud modulada en la etapa 177 utilizando la información de amplitud de atenuación cruzada del bloque 170. El modulador de amplitud tal como el multiplicador 203b genera finalmente una señal de paso de banda sintetizada para un cierto canal de paso de banda que, debido al procedimiento de la invención, tiene un ancho de banda de frecuencia que varía desde bajo hasta alto con aumento de la frecuencia central de paso de banda.

En lo siguiente, se presentan algunos espectrogramas que demuestran las propiedades de los esquemas de procesamiento de modulación propuestos. La figura 7a muestra el espectrograma log original de un extracto de un elemento de música clásica orquestal (Vivaldi).

De la figura 7b a la figura 7e muestran espectrogramas correspondientes después de diversos métodos de procesamiento de modulación con el fin de aumentar el detalle de la modulación restaurada. La figura 7b ilustra la reconstrucción de señal únicamente desde los portadores. Las regiones blancas corresponden a alta energía espectral y coinciden con la concentración de energía local en el espectrograma de la señal original en la figura 7a. La figura 7c representa los mismos portadores pero refinados mediante suavizado no lineal AM y FM. La adición de detalles es claramente visible. En la figura 7d adicionalmente la pérdida del detalle AM se compensa mediante la adición de ruido (de adorno) con forma de envoltura que puede de nuevo agregar más detalle a la señal. Finalmente el espectrograma de la señal sintetizada de los componentes de modulación no modificados se muestra en la figura 7e. Comparando el espectrograma en la figura 7e con el espectrograma de la señal original en la figura 7a ilustra la muy buena reproducción de los detalles completos.

Para evaluar el desempeño del método propuesto, se conduce una prueba de escucha subjetiva. Se condujo la prueba de escucha tipo MUSHRA [21] utilizando audífonos electrostáticos de alta calidad STAX. Un número total de 6 oyentes participaron en la prueba. Todos los sujetos se pueden considerar como oyentes experimentados.

El conjunto de prueba consistía en elementos enumerados en la figura 8 y las configuraciones bajo prueba se incluyeron en la figura 9.

5 El gráfico en la figura 8 visualiza el resultado. Se muestran los resultados promedio con intervalos de confianza del 95% para cada elemento. Los gráficos muestran los resultados después de análisis estadístico de los resultados de prueba para todos los oyentes. El eje X muestra el tipo de procesamiento y el eje Y representa la clasificación según la escala MUSHRA de 100 puntos que varía desde 0 (malo) hasta 100 (transparente).

10 A partir de los resultados puede observarse que las dos versiones que tienen detalles FM completos o ásperos y AM completos mejor en aproximadamente 80 puntos en media, pero se pueden distinguir aún del original. Debido a los intervalos de confianza de ambas versiones bastantes solapadas, puede concluirse que la pérdida de detalle fino FM de hecho es perceptivamente insignificante. La versión con ruido "de adorno" agregado y ruido AM y FM áspero se clasifica considerablemente más bajo pero en media aún en 60 puntos: esto refleja la propiedad de degradación adornada del método propuesto con la omisión aumentada de la información de detalle AM fino.

15 La mayor parte de la degradación se percibe de elementos que tienen contenido transitorio fuerte como carrillón y clavecín. Esto se debe a la pérdida de relaciones de fase original entre los diferentes componentes a través del espectro. Sin embargo, este problema se puede superar en versiones futuras del método de síntesis propuesto al ajustar la fase de portador en centros de gravedad temporales en la envoltura AM conjuntamente para todos los componentes.

Para los elementos de música clásica en el conjunto de prueba la degradación observada es estadísticamente insignificante.

25 El método de análisis/síntesis presentado puede ser de uso en diferentes escenarios de aplicación: El método de análisis/síntesis presentado puede ser de uso en diferentes escenarios de aplicación: para codificación de audio serviría como un elemento fundamental de un codificador de audio de escala de grano fino perceptivamente correcto mejorado cuyo principio básico se ha publicado en [1]. Con la reducción de la velocidad de bits se pueden transportar menos detalles al lado del receptor al reemplazar por ejemplo la envoltura AM completa mediante un ruido "de adorno" agregado y uno áspero.

30 Adicionalmente se pueden concebir nuevos conceptos de extensión de ancho de banda de audio [20] que por ejemplo utilizan componentes de base de banda alterados y cambiados para formar bandas grandes. Llegan a ser factibles experimentos mejorados sobre las propiedades auditivas humanas por ejemplo la creación mejorada de sonidos quiméricos con el fin de evaluar adicionalmente la percepción humana de la estructura de la modulación [11].

35 No menos importante, los efectos de audio artísticos nuevos y excitantes para la producción de música están dentro del alcance: ya sea en modo de escala y clave de un elemento de música que se puede alterar mediante el procesamiento adecuado de las señales de portador o la propiedad sicoacústica de la sensación de rugosidad que puede ser accedida mediante la manipulación de los componentes AM.

40 Se ha presentado una propuesta de un sistema para descomponer una señal de audio arbitraria en componentes AM/FM y de portadores perceptivamente significativos, que permiten el cambio de escala de grano fino de la modulación de detalle de modulación. Se ha dado un método de resíntesis apropiado. Algunos ejemplos de principios de procesamiento de modulación se han destacado y se han presentado los espectrogramas resultantes de un archivo de audio de ejemplo. Se ha conducido una prueba de audición para verificar la calidad perceptiva de diferentes tipos de procesamiento de modulación y la posterior resíntesis. Se han identificado escenarios de aplicación futuros para este método de análisis/síntesis nuevo promisorio. Los resultados demuestran que los métodos propuestos proporcionan medios apropiados para obviar el espacio entre el procesamiento de audio en forma de ondas y el procesamiento de audio paramétrico y más aún produce nuevos efectos de audio fascinantes posibles.

45 Las realizaciones descritas son meramente ilustrativas para los principios de la presente invención. Se entiende que resultarán evidentes modificaciones y variaciones de las disposiciones. y los detalles descritos en el presente documento para los expertos en la técnica. Por lo tanto, se pretenden estar limitado solo por el alcance de las reivindicaciones de patente pendientes y no por los detalles específicos presentados por vía de descripción y explicación de las realizaciones en el presente documento.

50 Dependiendo de ciertos requisitos de implementación de los métodos de la invención, los métodos de la invención se pueden implementar en hardware o en software. La implementación se puede realizar utilizando un medio de almacenamiento digital, en particular, un disco, un DVD o un CD que tiene señales de control legibles electrónicamente almacenadas allí, que cooperan con sistemas de ordenador programables de tal manera que se pueden realizar los métodos de la invención. De manera general, la presente invención es por lo tanto un producto de programa de ordenador con un código de programa almacenado en un portador legible por máquina, el código de programa se opera para realizar los métodos de la invención cuando el producto de programa de ordenador se ejecuta sobre un

ordenador. En otras palabras, los métodos de la invención son, por lo tanto, un programa de ordenador que tiene un código de programa para realizar al menos uno de los métodos de la invención cuando el programa se ejecuta en un ordenador.

## 5 Bibliografía

- [1] M. Vinton and L. Atlas, "A Scalable And Progressive Audio Codec", in Proc. of ICASSP 2001, páginas 3277-3280, 2001
- 10 [2] H. Dudley, "The vocoder," in Bell Labs Record, vol. 17, páginas 122-126, 1939
- [3] J. L. Flanagan and R. M. Golden, "Phase Vocoder," in Bell System Technical Journal, vol. 45, páginas 1493-1509, 1966
- 15 [4] J. L. Flanagan, "Parametric coding of speech spectra," J. Acoust. Soc. Am., vol. 68 (2), páginas 412-419, 1980
- [5] U. Zoelzer, DAFX: Digital Audio Effects, Wiley & Sons, páginas 201-298, 2002
- [6] H. Kawahara, "Speech representation and transformation using adaptive interpolation of weighted spectrum: vocoder revisited," in Proc. of ICASSP 1997, vol. 2, páginas 1303-1306, 1997
- 20 [7] A. Rao and R. Kumaresan, "On decomposing speech into modulated components," in IEEE Trans. on Speech and Audio Processing, vol. 8, páginas 240-254, 2000
- 25 [8] M. Christensen et al., "Multiband amplitude modulated sinusoidal audio modelling," in IEEE Proc. of ICASSP 2004, vol. 4, páginas 169-172, 2004
- [9] K. Nie and F. Zeng, "A perception-based processing strategy for cochlear implants and speech coding," in Proc. of the 26th IEEE-EMBS, vol. 6, páginas 4205-4208, 2004
- 30 [10] J. Thiemann and P. Kabal, "Reconstructing Audio Signals from Modified Non-Coherent Hilbert Envelopes," in Proc. Interspeech (Antuerpia, Bélgica), páginas 534-537, 2007
- [11] Z. M. Smith and B. Delgutte and A. J. Oxenham, "Chimaeric sounds reveal dichotomies in auditory perception," in Nature, vol. 416, páginas 87-90, 2002
- 35 [12] J. N. Anantharaman and A.K. Krishnamurthy, L.L Feth, "Intensity weighted average of instantaneous frequency as a model for frequency discrimination," in J. Acoust. Soc. Am., vol. 94 (2), páginas 723-729, 1993
- 40 [13] O. Ghitza, "On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception," in J. Acoust. Soc. Amer., vol. 110(3), páginas 1628-1640, 2001
- [14] E. Zwicker and H. Fastl, Psychoacoustics - Facts and Models, Springer, 1999
- 45 [15] E. Terhardt, "On the perception of periodic sound fluctuations (roughness)," in Acustica, vol. 30, páginas 201-213, 1974
- [16] P. Daniel and R. Weber, "Psychoacoustical Roughness: Implementation of an Optimized Model," in Acustica, vol. 83, páginas 113-123, 1997
- 50 [17] P. Loughlin and B. Tacer, "Comments on the interpretation of instantaneous frequency," in IEEE Signal Processing Lett., vol. 4, páginas 123-125, 1997.
- [18] D. Wei and A. Bovik, "On the instantaneous frequencies of multicomponent AM-FM signals," in IEEE Signal Processing Lett., vol. 5, páginas 84-86, 1998.
- 55 [19] Q.Li and L. Atlas, "Over-modulated AM-FM decomposition," in Proceedings of the SPIE, vol. 5559, páginas. 172-183, 2004
- 60 [20] M. Dietz, L. Liljeryd, K. Kjörling and O. Kunz, "Spectral Band Replication, a novel approach in audio coding," in 112º AES Convention, Múnich, mayo de 2002.
- [21] ITU-R Recommendation BS.1534-1, "Method for the subjective assessment of intermediate sound quality (MUSHRA)," International Telecommunications Union, Geneva, Suiza, 2001.

[22] "Sinusoidal modeling parameter estimation via a dynamic channel vocoder model" A.S. Master, 2002 IEEE International Conference on Acoustics, Speech and Signal Processing .

5

**REIVINDICACIONES**

1. Aparato para convertir una señal (100) de audio en una representación (180) parametrizada, que comprende:
  - 5 un analizador (102) de señal para analizar una porción de la señal (100) de audio para obtener un resultado (104) del análisis;
  - 10 un estimador (106) de paso de banda para estimar la información (108) de la pluralidad de filtros de paso de banda basados en el resultado (104) del análisis, en el que la información (108) en la pluralidad de filtros de paso de banda comprende, para cada filtro de paso de banda, información en una frecuencia (108) portadora que indica una frecuencia central del filtro de paso de banda para la porción de la señal (100) de audio, en el que los anchos de banda de los filtros de paso de banda de la pluralidad de filtros de paso de banda son diferentes de un espectro de audio y dependen de la frecuencia central de un filtro de paso de banda de la pluralidad de filtros de paso de banda;
  - 15 un estimador (110) de modulación para estimar una modulación (112) de amplitud y una modulación de frecuencia o una modulación (114) de fase para cada filtro de paso de banda de la pluralidad de filtros de paso de banda para la porción de la señal (100) de audio que utiliza la frecuencia (108) portadora para cada filtro de paso de banda de la pluralidad de filtros de paso de banda,
  - 20 en el que se configura el estimador (110) de modulación,
  - 25 para extraer (139), para cada filtro de paso de banda de la pluralidad de filtros de paso de banda, una señal de paso de banda de la porción de la señal (100) de audio que utiliza el filtro de paso de banda como se determina por la información (108) en la frecuencia central y el ancho de banda para el filtro de paso de banda,
  - 30 para formar (140, 110b), para cada filtro de paso de banda de la pluralidad de filtros de paso de banda, una señal analítica de la señal de paso de banda,
  - 35 para mezclar a la baja (142, 110d), para cada filtro de paso de banda de la pluralidad de filtros de paso de banda, la señal analítica con una señal del oscilador (110e) que tiene la frecuencia (108) portadora del respectivo filtro de paso de banda para obtener una señal analítica mezclada baja, en el que la mezcla a la baja (142, 110d) comprende, para cada filtro de paso de banda de la pluralidad de filtros de paso de banda, una multiplicación (110d) de la señal analítica por la señal del oscilador que tiene la frecuencia central del filtro de paso de banda, y en el que se realiza una posterior operación de filtro de paso bajo para rechazar una porción de frecuencia alta generada por la multiplicación, o en el que, cuando la señal del oscilador es compleja, el filtro de paso bajo no se realiza, y
  - 40 para extraer (143), para cada filtro de paso de banda de la pluralidad de filtros de paso de banda, información de fase de la señal analítica mezclada a la baja para obtener información en la modulación de fase en la banda del filtro de paso de banda, o para diferenciar (144) la información de fase en la banda del filtro de paso de banda para obtener información en la modulación de frecuencia en la banda del filtro de paso de banda; y
  - 45 una interfaz (116) de salida para transmitir o almacenar, para cada filtro de paso de banda de la pluralidad de filtros de paso de banda, como la representación (180) parametrizada,
  - 50 información sobre la modulación (112) de amplitud para la porción de la señal (100) de audio,
  - la información sobre la modulación de frecuencia o la información sobre la modulación (114) de fase para la porción de la señal (100) de audio, y
  - 55 la información (108) sobre la frecuencia portadora para la porción de la señal (100) de audio.
2. Aparato según la reivindicación 1, en el que el analizador (102) de señal está configurado para analizar (132) la porción de la señal (100) de audio con respecto a una distribución de potencia o amplitud sobre la frecuencia de la porción de la señal (100) de audio.
- 60 3. Aparato según la reivindicación 1 o 2, en el que el analizador (102) de señal está configurado para analizar (132) la distribución de potencia de la señal (100) de audio en las bandas de frecuencia dependiendo de las frecuencias centrales de las bandas (122).
4. Aparato según una de las reivindicaciones anteriores, en el que el estimador (106) de paso de banda está

configurado para estimar la información sobre la pluralidad de filtros de paso de banda, en el que un ancho de banda de un filtro de paso de banda de la pluralidad de filtros de paso de banda que tiene una frecuencia central mayor es más grande que un ancho de banda de un filtro de paso de banda de una pluralidad de filtros de paso de banda que tiene una frecuencia central menor.

5  
5. Aparato según una de las reivindicaciones anteriores, en el que la dependencia entre las frecuencias centrales y los anchos de banda es de modo que cualquiera de las dos frecuencias centrales adyacentes a la frecuencia tiene una distancia similar en frecuencia entre sí sobre una escala logarítmica.

10  
6. Aparato según una de las reivindicaciones anteriores, en el que el estimador (110) de modulación está configurado para operar sobre una base bloque a bloque con un solapamiento entre bloques, teniendo un bloque porciones externas y una porción central, comprendiendo las porciones externas un filtro que timbra o un balanceo espontáneo característico de una ventana de transformación,

15  
en el que el aparato está configurado para generar la representación (180) parametrizada para un bloque de una parte seleccionada del bloque, comprendiendo la parte seleccionada solo la porción central del bloque.

20  
7. Aparato según la reivindicación 1, en el que el estimador (110) de modulación está configurado para calcular (141), para cada filtro de paso de banda de la pluralidad de filtros de paso de banda, una magnitud de la señal analítica para obtener la información sobre la modulación de amplitud de la porción de la señal (100) de audio en la banda del filtro de paso de banda.

25  
8. Método para convertir una señal (100) de audio en una representación (180) parametrizada, que comprende:  
analizar (102) una porción de la señal (100) de audio para obtener un resultado (104) de análisis;

30  
estimar (106) la información (108) de una pluralidad de filtros de paso de banda basados en el resultado (104) del análisis, en el que la información (108) en la pluralidad de filtros de paso de banda comprende, para cada filtro de paso de banda, información en una frecuencia (108) portadora que indica una frecuencia central del filtro de paso de banda para la porción de la señal (100) de audio, en el que los anchos de banda de los filtros de paso de banda de la pluralidad de filtros de paso de banda son diferentes de un espectro de audio y dependen de la frecuencia central de un filtro de paso de banda de la pluralidad de filtros de paso de banda;

35  
estimar (110) una modulación (112) de amplitud y una modulación de frecuencia o una modulación (114) de fase para cada filtro de paso de banda de la pluralidad de filtros de paso de banda para la porción de la señal (100) de audio que utiliza la frecuencia (108) portadora para cada filtro de paso de banda de la pluralidad de filtros de paso de banda, comprendiendo la estimación (110):

40  
extraer (139), para cada filtro de paso de banda de la pluralidad de filtros de paso de banda, una señal de paso de banda de la porción de la señal (100) de audio que utiliza el filtro de paso de banda como se determina por la información (108) en la frecuencia central y el ancho de banda para el filtro de paso de banda,

45  
formar (140, 110b), para cada filtro de paso de banda de la pluralidad de filtros de paso de banda, una señal analítica de la señal de paso de banda,

50  
mezclar a la baja (142, 110d), para cada filtro de paso de banda de la pluralidad de filtros de paso de banda, la señal analítica con una señal del oscilador (110e) que tiene la frecuencia (108) portadora del respectivo filtro de paso de banda para obtener una señal analítica mezclada baja, en el que la mezcla a la baja (142, 110d) comprende, para cada filtro de paso de banda de la pluralidad de filtros de paso de banda, una multiplicación (110d) de la señal analítica por la señal del oscilador que tiene la frecuencia central del filtro de paso de banda, y en el que se realiza una posterior operación de filtro de paso bajo para rechazar una porción de frecuencia alta generada por la multiplicación, o en el que, cuando la señal del oscilador es compleja, el filtro de paso bajo no se realiza, y

55  
extraer (143), para cada filtro de paso de banda de la pluralidad de filtros de paso de banda, información de fase de la señal analítica mezclada a la baja para obtener información en la modulación de fase en la banda del filtro de paso de banda, o para diferenciar (144) la información de fase en la banda del filtro de paso de banda para obtener información en la modulación de frecuencia en la banda del filtro de paso de banda; y

60  
transmitir o almacenar, para cada filtro de paso de banda de la pluralidad de filtros de paso de banda, como la representación (180) parametrizada,

información sobre la modulación (112) de amplitud para la porción de la señal (100) de audio,

la información sobre la modulación de frecuencia o la información sobre la modulación (114) de fase para la porción de la señal (100) de audio, y

la información (108) sobre la frecuencia portadora para la porción de la señal (100) de audio.

- 5
9. Programa de ordenador para realizar, cuando se ejecuta sobre un ordenador, un método según la reivindicación 8.

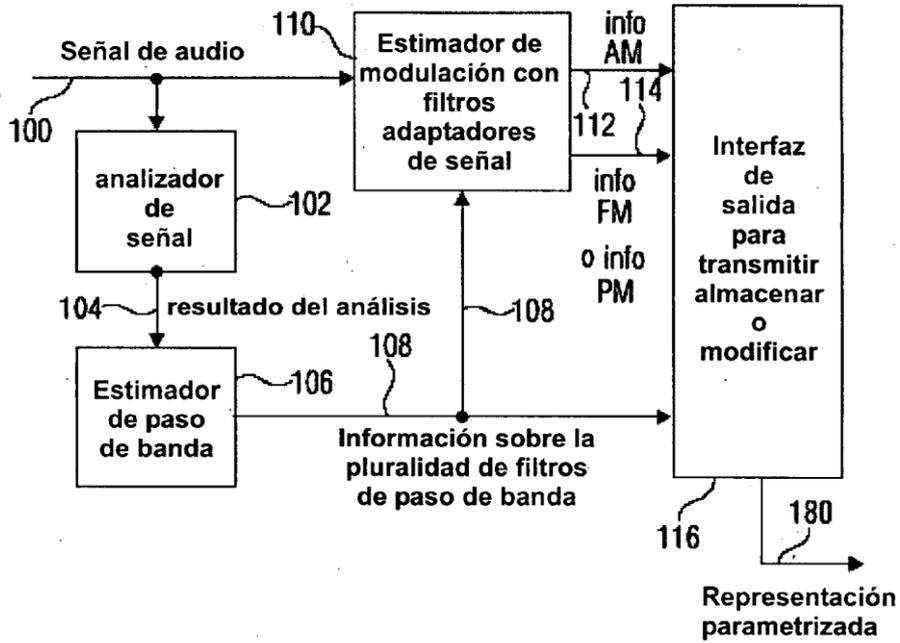


FIG 1A

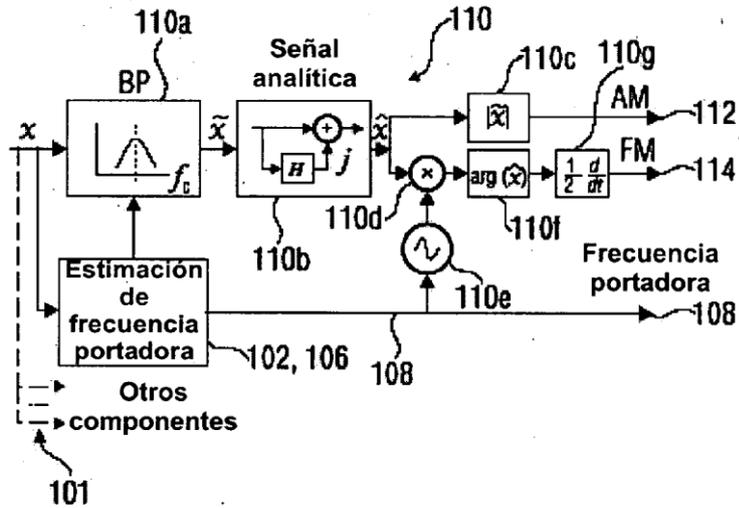


FIG 1B

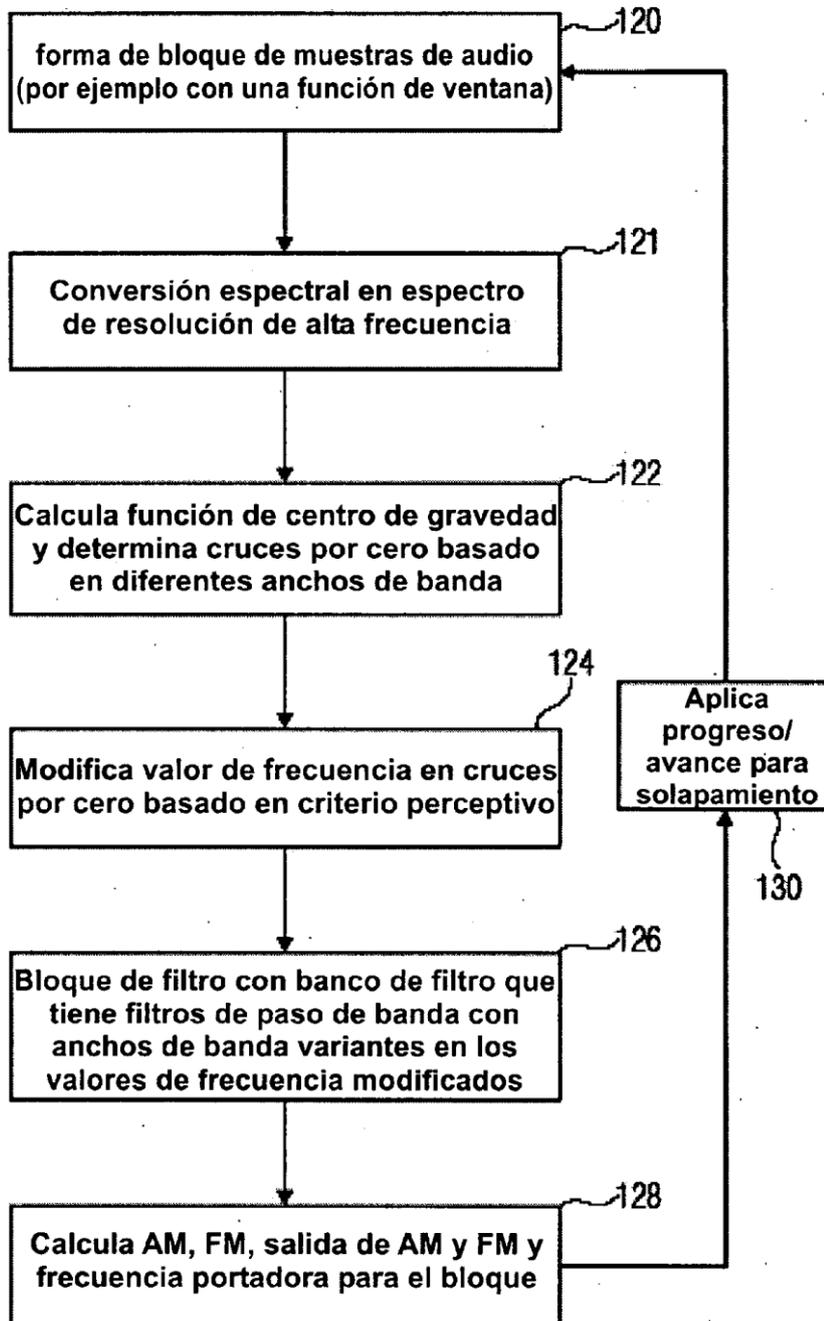


FIG 2A

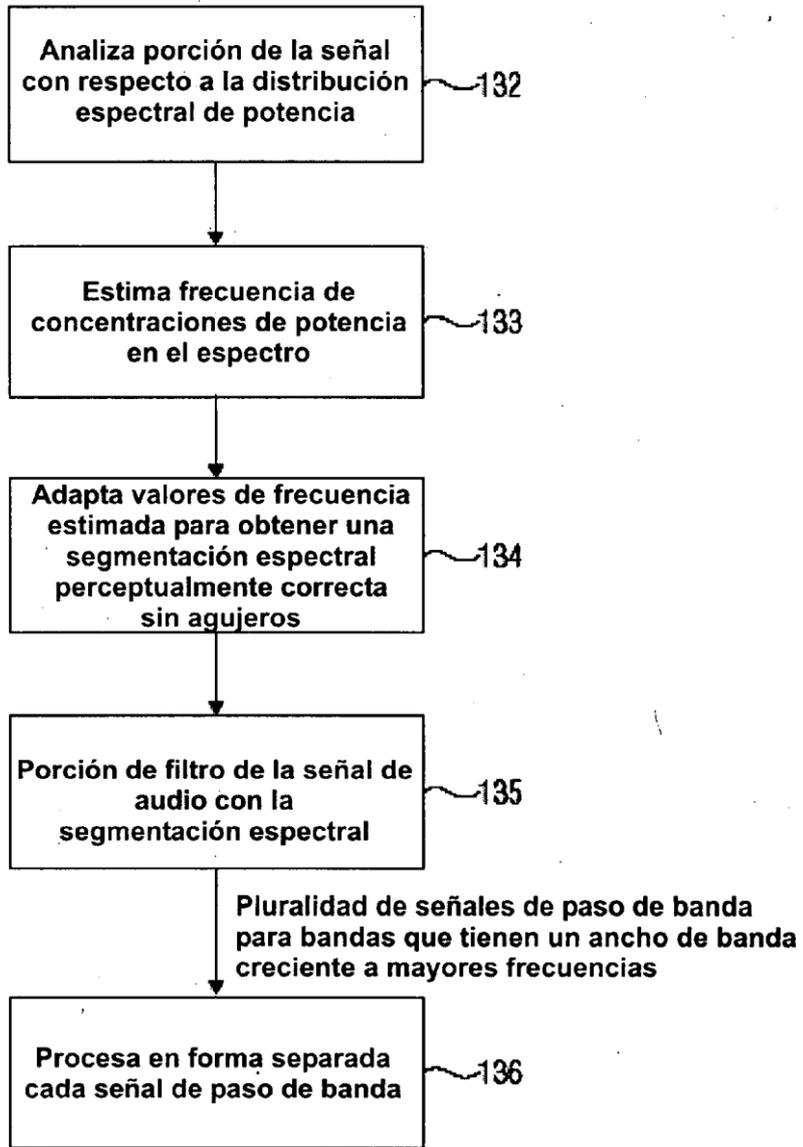


FIG 2B

Segmentación derivada de COG local de espectro de potencia (extracto)  
Elemento: Vivaldi, bloque de tiempo: No. 7

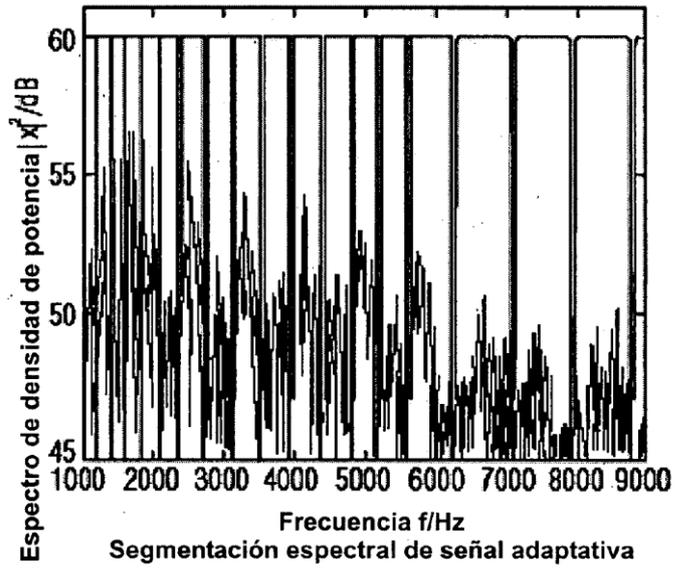


FIG 2C

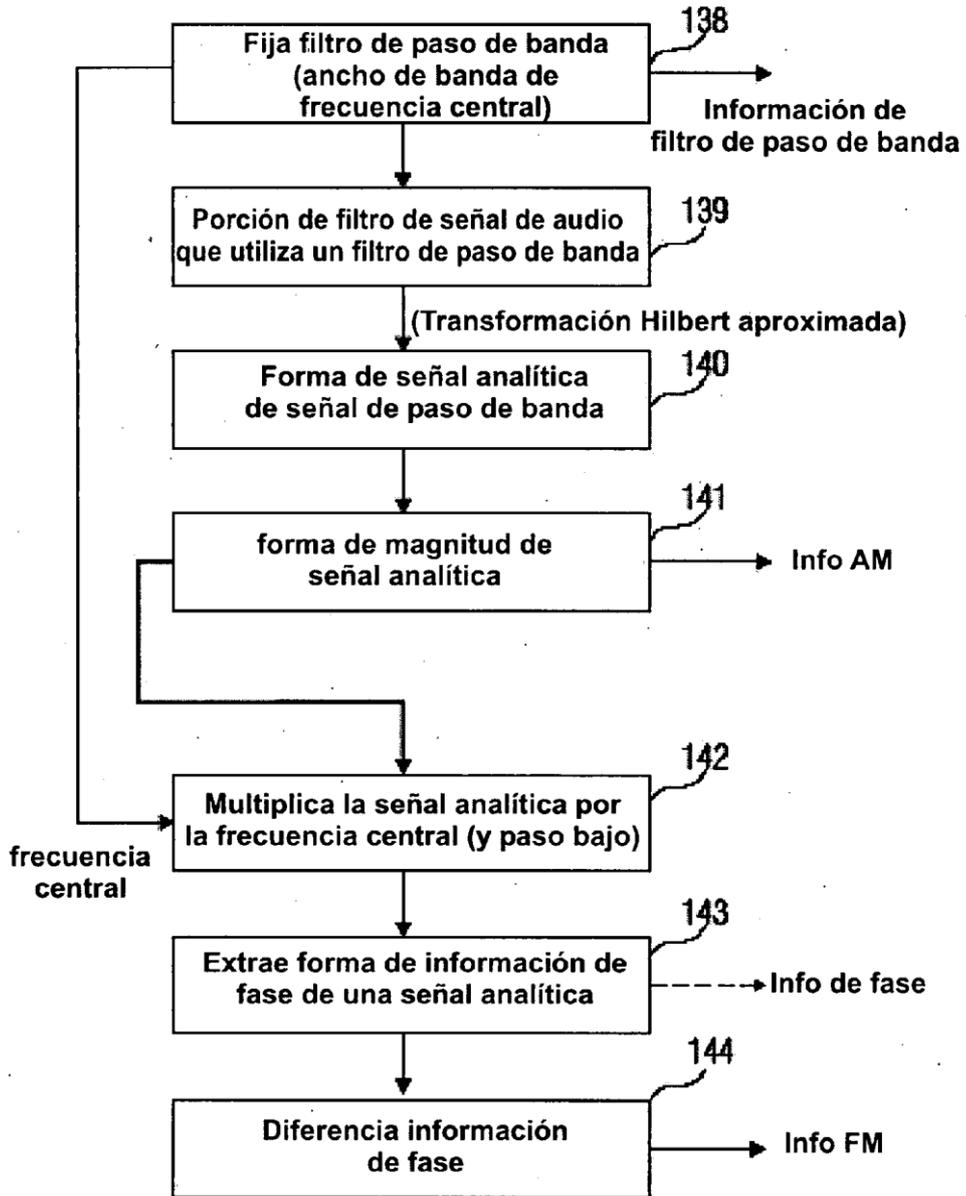


FIG 2D

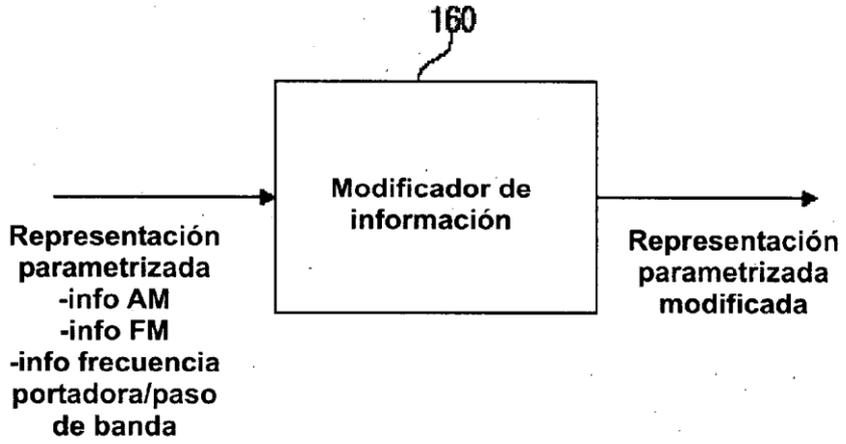


FIG 3A

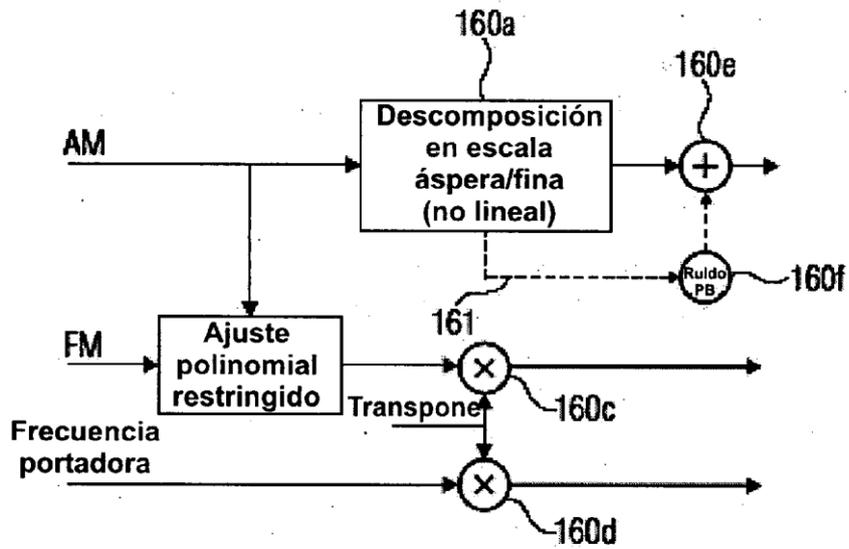


FIG 3B

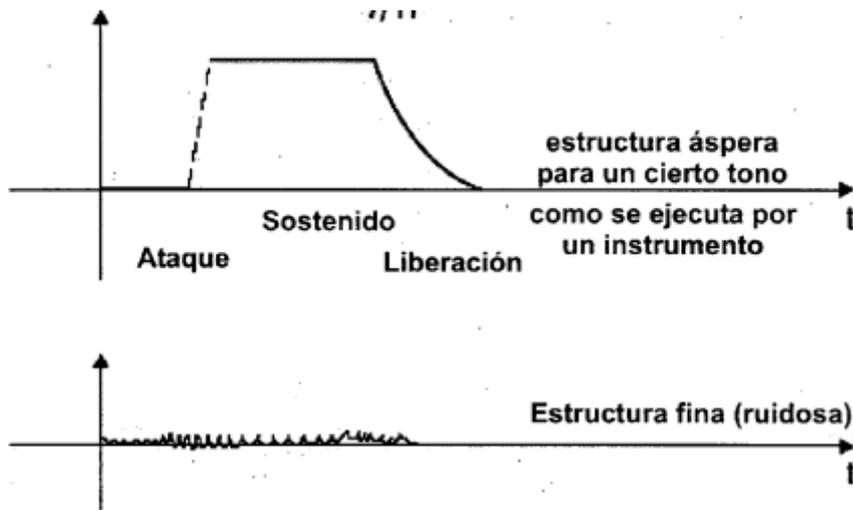


FIG 3C

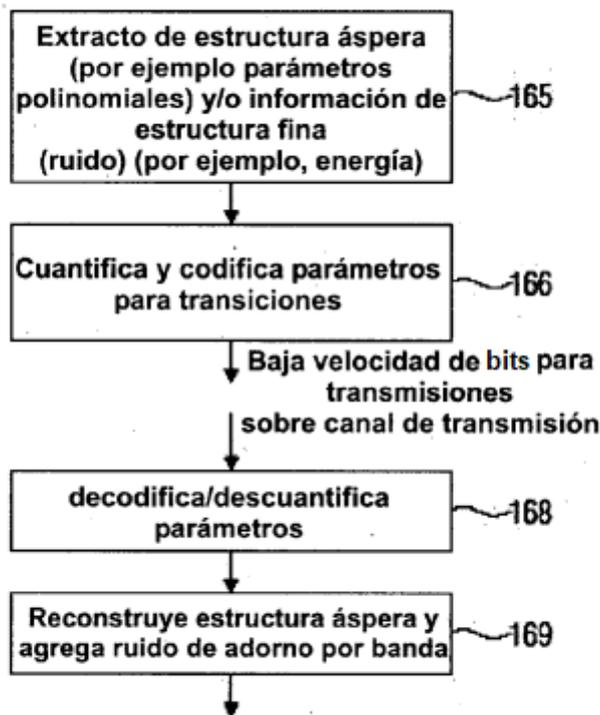


FIG 3D

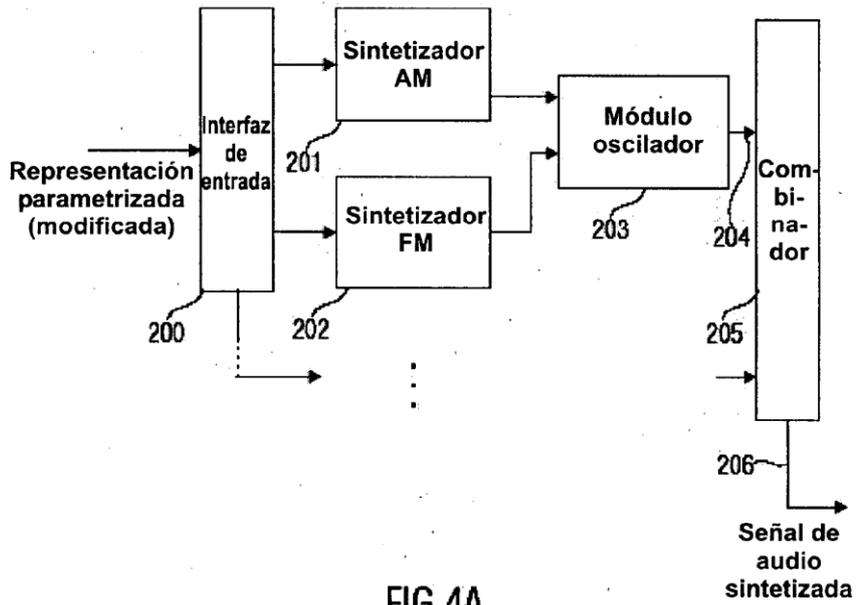


FIG 4A

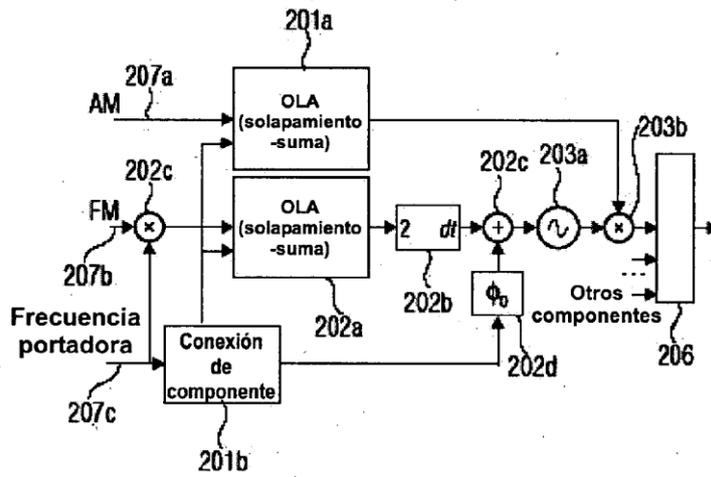


FIG 4B

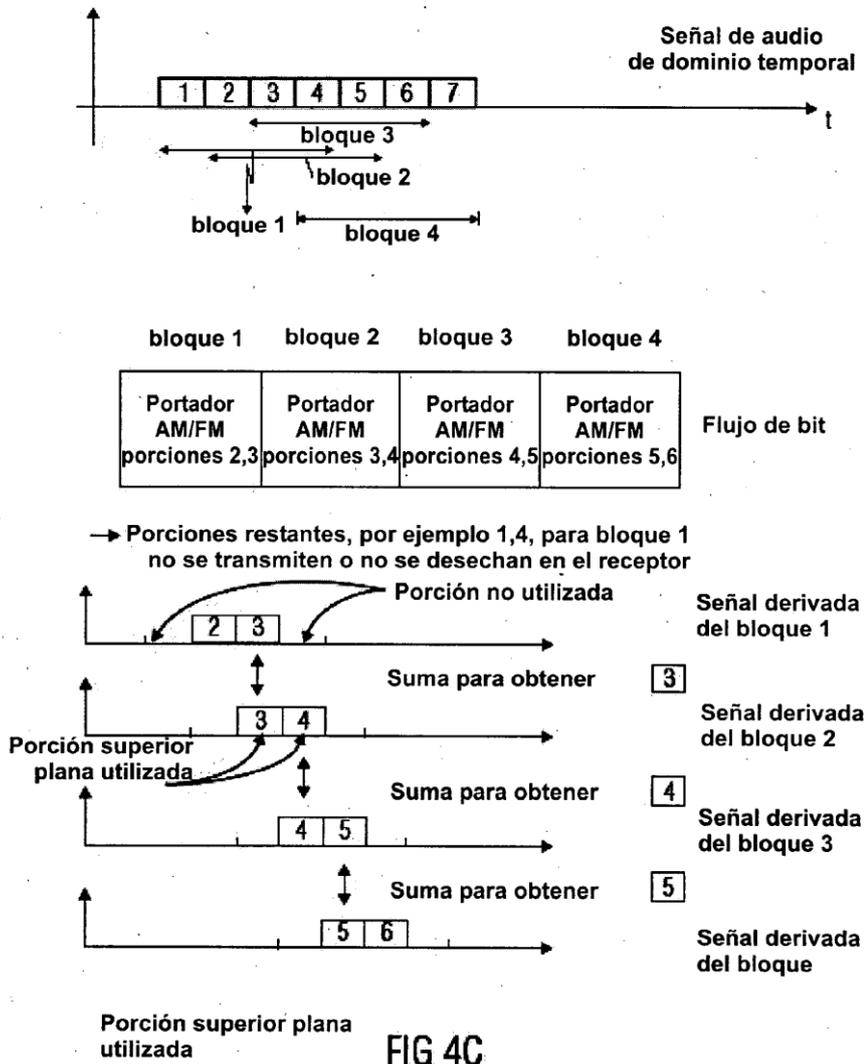


FIG 4C

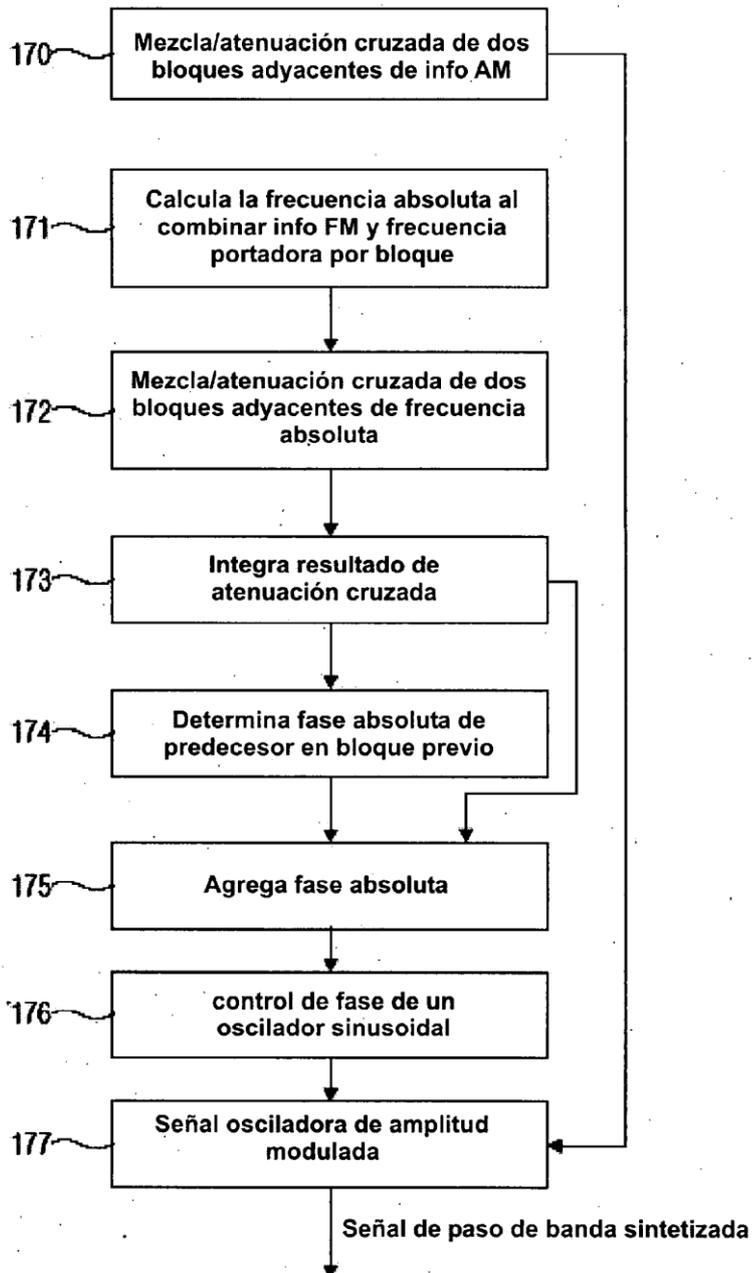


FIG 4D

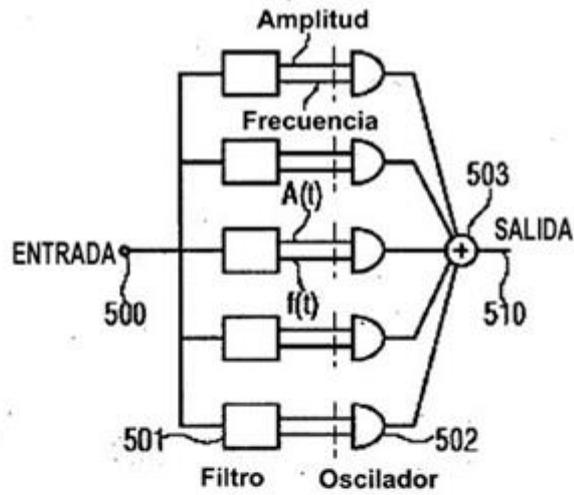


FIG 5 (Técnica Anterior)

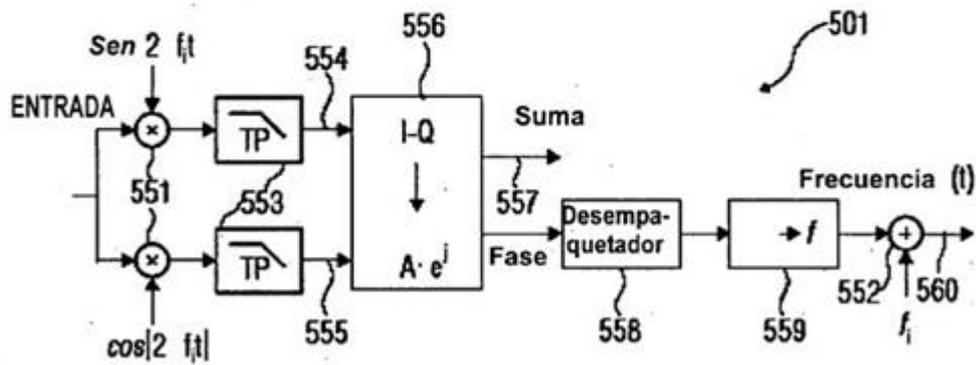
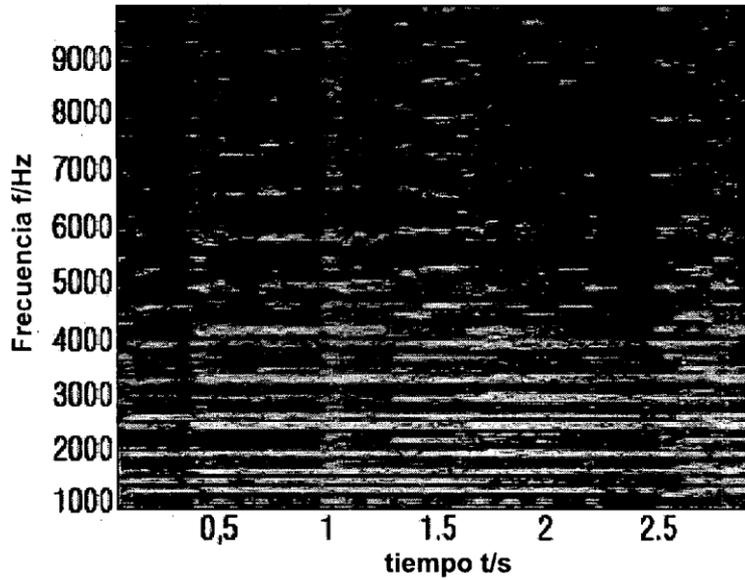


FIG 6 (Técnica anterior)

**FIG 7A**

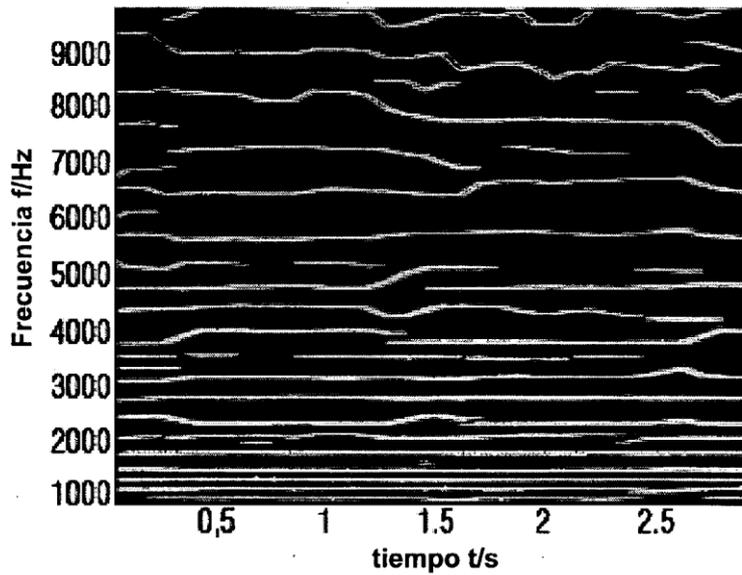
**Espectrograma Log (extracto) de elemento:  
Vivaldi\_48\_m**



**Espectrograma del elemento de música clásica original**

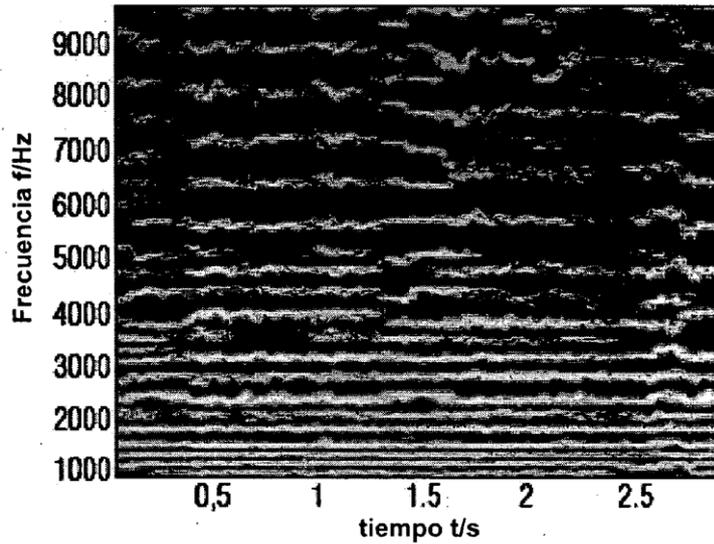
**FIG 7B**

**Espectrograma Log (extracto) de elemento:  
Vivaldi\_48\_m\_noAM\_noFM\_11\_processed**



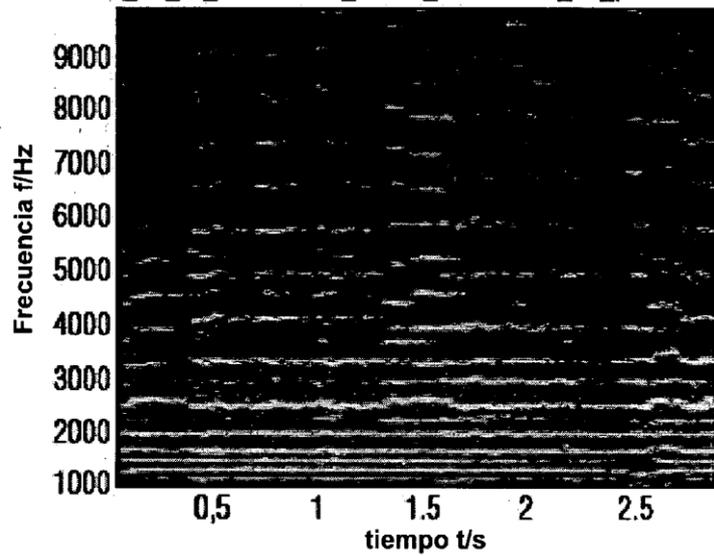
**Espectrograma del elemento portador sintetizado**

**FIG 7C** Espectrograma Log (extracto) de elemento:  
Vivaldi\_48\_m\_coarseAM\_coarseFM\_11\_processed



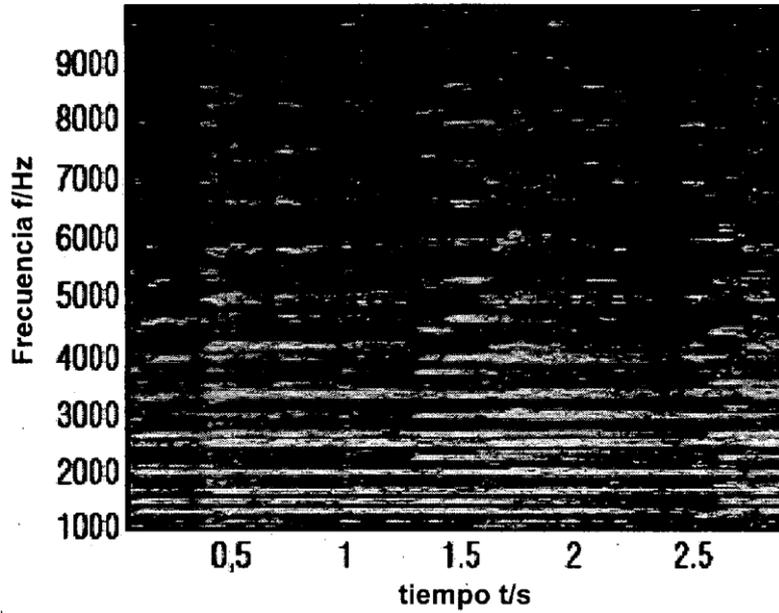
Espectrograma de portadores refinados por AM y FM áspero

**FIG 7D** Espectrograma Log (extracto) de elemento:  
Vivaldi\_48\_m\_coarseAM\_noise\_coarseFM\_11\_processed



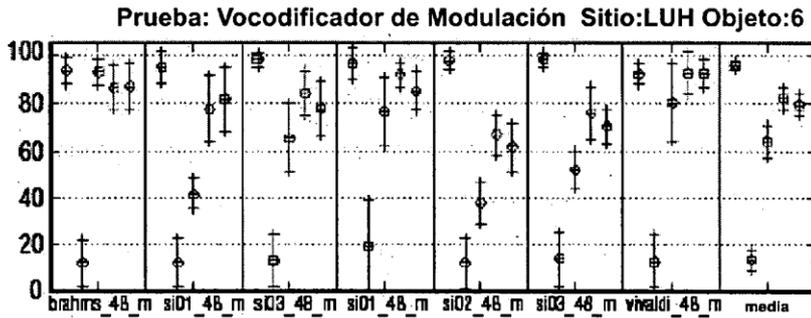
Espectrograma de portadores refinados por AM y FM áspero,  
y ruido de "adorno" agregado

**Espectrograma Log (extracto) de elemento:  
Vivaldi\_48\_m\_fullAM\_fullFM\_11\_processed**



**Espectrograma de portadores AM y FM no procesados**

**FIG 7E**



- hidden\_reference
- 3k5Hz\_reference
- coarseAM\_noise\_coarseFM
- full AM\_coarseFM
- fullAM\_fullFM

Resultado de prueba subjetiva de calidad de audio (MUSHRA)  
Elementos de Prueba

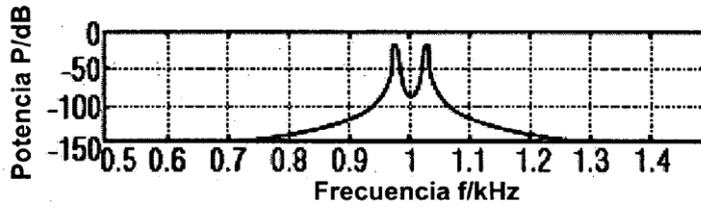
| Elemento     | Descripción                |
|--------------|----------------------------|
| vivaldi 48 m | Música clásica orquestal   |
| brahms 48 m  | Música clásica orquestal   |
| si01 48 m    | Clavecín/MPEG              |
| si03 48 m    | Diapazón/MPEG              |
| sm01 48 m    | Gaita/MPEG                 |
| sm02 48 m    | Carrillón/MPEG             |
| sm03 48 m    | Instrumento de cuerda/MPEG |

**Configuraciones bajo prueba**

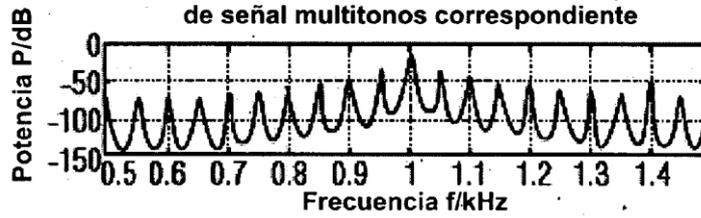
|                                |   |
|--------------------------------|---|
| Extensión de nombre de archivo | Método de procesamiento                                   |
| Referencia oculta              | Referencia Oculta   |
| 3k5Hz                          | Anclaje inferior  |
| fullAM_fullFM                  | Procesamiento sin modulación                              |
| fullAM_coarseFM                | Información FM áspera                                     |
| coarseAM_noise_coarseFM        | Información AM y FM áspera con ruido "de adorno" agregado |

**FIG 8**

Densidad espectral de potencia estimada (extracto) de señal de 2 tonos



Densidad espectral de potencia estimada (extracto) de señal multitonos correspondiente



Densidad espectral de potencia estimada (extracto) de señal multitonos limitada por banda perceptualmente equivalente

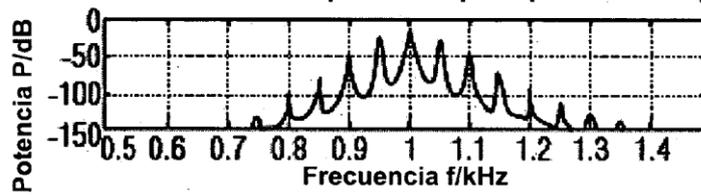
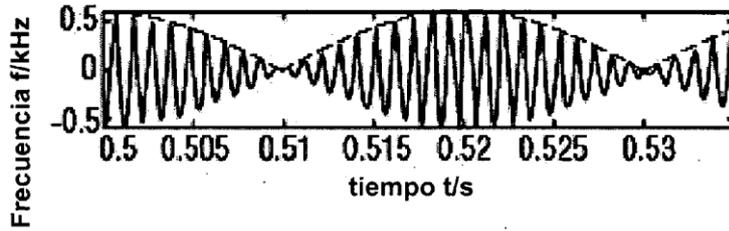
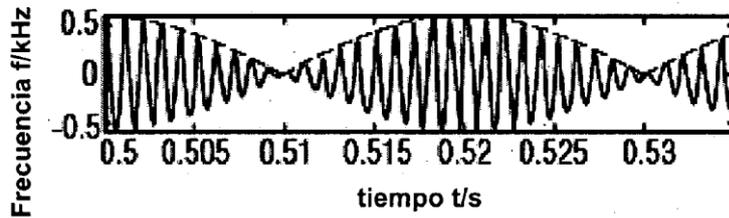


FIG 9A

forma de onda (línea), envoltura (discontinua) de señal de 2 tonos



forma de onda (línea), envoltura (discontinua) de señal multi tonos



forma de onda (línea), envoltura (discontinua) de señal multi tonos limitada por banda perceptivamente equivalente

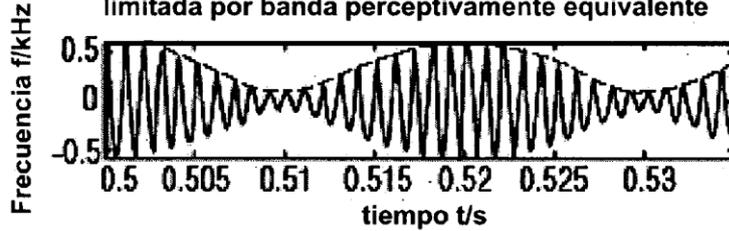


FIG 9B

$$s_t(t) = \text{sen}(2 f_1 t) + \text{sen}(2 f_2 t)$$

$$s_m(t) = 2 \text{sen}\left(2 \frac{f_1 + f_2}{2} t\right) \cdot \left| \cos\left(2 \frac{f_1 - f_2}{2} t\right) \right|$$

FIG 9C