

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 738 534**

51 Int. Cl.:

G10L 21/04 (2013.01)

G10L 19/025 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **17.02.2009 PCT/EP2009/001108**

87 Fecha y número de publicación internacional: **17.09.2009 WO09112141**

96 Fecha de presentación y número de la solicitud europea: **17.02.2009 E 09719651 (3)**

97 Fecha y número de publicación de la concesión europea: **01.05.2019 EP 2250643**

54 Título: **Dispositivo y método para manipular una señal de audio que tiene un evento transitorio**

30 Prioridad:

10.03.2008 US 35317 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
23.01.2020

73 Titular/es:

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

72 Inventor/es:

**DISCH, SASCHA;
NAGEL, FREDERIK;
RETTTELACH, NIKOLAUS;
MULTRUS, MARKUS y
FUCHS, GUILLAUME**

74 Agente/Representante:

ARIZTI ACHA, Monica

ES 2 738 534 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

Dispositivo y método para manipular una señal de audio que tiene un evento transitorio

5 La presente invención es concerniente con el procesamiento de señales de audio y particularmente con la manipulación de señales de audio en el contexto de aplicación de efectos de audio a una señal que contiene eventos transitorios.

10 Es conocido manipular las señales de audio de tal manera que la velocidad de reproducción es cambiada, en tanto que se mantiene el tono. Métodos conocidos para tal procedimiento son implementados por vocodificadores de fase o métodos como superposición-adición (cabecero sincrónico) (P)SOLA, como se describe por ejemplo en J.L. Flanagan and R. M. Golden, The Bell System Technical Journal, noviembre 1966, pág. 1394 a 1509; patente estadounidense 6549884 expedida a Laroche, J. & Dolson, M.: Phase-vocoder pitch-shifting; Jean Laroche and Mark Dolson, New Phase-Vocoder Techniques for Pitch-Shifting, Harmonizing And Other Exotic Effects”, Proc. 1999 IEEE
15 Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, Nueva York, Oct. 17-20, 1999; and Zölzer, U: DAFX: Digital Audio Effects; Wiley & Sons; Edición: 1 (26 de febrero de 2002); pág. 201-298.

20 Adicionalmente, las señales de audio pueden ser sometidas a una transposición utilizando tales métodos, esto es, vocodificadores de fase o (P)SOLA en donde la cuestión especial de esta clase de transposición es que la señal de audio transpuesta tiene la misma duración de reproducción/repetición como la señal de audio original antes de la transposición, en tanto que el tono es cambiado. Esto es obtenido mediante una reproducción acelerada de las señales estiradas, en donde el factor de aceleración para efectuar la reproducción acelerada depende del factor de estiramiento para estirar la señal de audio original en el tiempo. Cuando se tiene una representación de señal discreta en el tiempo, este procedimiento corresponde a una toma de muestras de descendente de la señal estirada
25 o decimación de la señal estirada por un factor igual al factor de estiramiento, en donde se mantiene la frecuencia de toma de muestras.

30 Un desafío específico en tales manipulaciones de señal de audio son los eventos transitorios. Los eventos transitorios son eventos en una señal en los cuales la energía de la señal en toda la banda o en un cierto intervalo de frecuencia está cambiando rápidamente, esto es, incrementándose rápidamente o disminuyendo rápidamente. Elementos característicos de los transitorios específicos (eventos transitorios) son la distribución de energía de señal en el espectro. Comúnmente, la energía de la señal de audio durante un evento transitorio es distribuida sobre toda la frecuencia, mientras que en las porciones de señal no transitorias, la energía es normalmente concentrada en la porción de baja frecuencia de la señal de audio o en bandas específicas. Esto significa que una porción de señal no
35 transitoria, que es también llamada una porción de señal estacionaria o porción de señal tonal tiene un espectro que no es plano. En otras palabras, la energía de la señal está incluida en un número comparativamente pequeño de líneas espectrales/bandas espectrales, que son fuertemente elevadas sobre un piso de ruido de una señal de audio. En una porción transitoria, sin embargo, la energía de la señal de audio será distribuida sobre muchas bandas de frecuencias diferentes y específicamente, será distribuida en la porción de alta frecuencia, de tal manera que un
40 espectro para una porción transitoria de la señal de audio será comparativamente plana y en cualquier evento serán más planas que un espectro de una porción tonal de la señal de audio. Comúnmente, un evento transitorio es un cambio fuerte en el tiempo, lo que significa que la señal incluirá muchas armónicas más altas cuando se efectúa una descomposición de Fourier. Una característica importante de estas muchas armónicas superiores es que las fases de estas armónicas superiores están en una relación mutua muy específica, de tal manera que una superposición de
45 todas estas ondas seno dará como resultado un cambio rápido en la energía de señal. En otras palabras, existe una fuerte correlación a través del espectro.

50 La situación de fase específica entre todas las armónicas, puede también ser denominada como “coherencia vertical”. Esta “coherencia vertical” está relacionada con una representación de espectrograma de tiempo/frecuencia de la señal, en donde una dirección horizontal corresponde al desarrollo de la señal a lo largo del tiempo y en donde la dimensión vertical describe la interdependencia con respecto a la frecuencia de los componentes espectrales (binarios de frecuencia de transformada) en un espectro de corto tiempo sobre la frecuencia.

55 Debido a las etapas de procesamiento típicas que son efectuadas con el fin de estirar o acortar el tiempo una señal de audio, esta coherencia vertical es destruida, lo que significa que un transitorio es “dañado” a lo largo del tiempo cuando un transitorio es sometido a una operación de estiramiento en el tiempo o acortamiento en el tiempo, como por ejemplo tal como es efectuada por un vocodificador de fase o cualquier otro método, que efectúa un procesamiento dependiente de la frecuencia que introduce desplazamiento de fase a la señal de audio, que son diferentes para diferentes coeficientes de frecuencia.

60 Cuando la coherencia vertical de los transitorios es destruida por un método de procesamiento de señal de audio, la señal manipulada será muy similar a la señal original en las porciones estacionaria o no transitoria, pero las porciones transitorias tendrán una calidad reducida en la señal manipulada. La manipulación sin control de la coherencia vertical de un transitorio da como resultado dispersión temporal del mismo, puesto que muchos

componentes armónicos contribuyen a un evento transitorio y el cambio de las fases de todos estos componentes de manera incontrolada inevitablemente da como resultado tales artefactos.

5 Sin embargo, las porciones transitorias son extremadamente importantes para la dinámica de una señal de audio, tal como una señal de música o una señal de habla en donde cambios repentinos de energía en un tiempo específico representan mucho de la impresión subjetiva del usuario de la calidad de la señal manipulada. En otras palabras, los eventos transitorios en una señal de audio son comúnmente "hitos" bastantes notables de una señal de audio, que tienen una influencia sobreproporcionada de la impresión de la calidad subjetiva. Los transitorios manipulados en los cuales la coherencia vertical ha sido destruida por una operación de procesamiento de señal o ha sido degradada con respecto a la porción transitoria de la señal original serán sonido distorsionado, reverberante y no natural al usuario que escucha.

15 Algunos métodos actuales estiran el tiempo alrededor de los transitorios a una extensión más alta para tener que efectuar subsecuentemente, durante la duración del transitorio, ninguno o solamente un estiramiento en el tiempo menor. Tales referencias en patentes y técnicas anteriores describen métodos para la manipulación del tiempo y/o tono. Las referencias de la técnica anterior son: Laroche L., Dolson M.: Improved phase vocoder timescale modification of audio", IEEE Trans. Speech and Audio Processing, vol. 7, n.º 3, pág. 323 – 332; Emmanuel Ravelli, Mark Sandler and Juan P. Bello: Fast implementation for non-linear time-scaling of stereo audio; Proc. of the 8ª Int. Conference on Digital Audio Effects (DAFx'05), Madrid, España, 20-22 de septiembre de 2005; Duxbury, C. M. Davies, and M. Sandler (diciembre de 2001). Separation of transient information in musical audio using multiresolution analysis techniques. In Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-01), Limerick, Ireland; and Röbel, A.: A NEW APPROACH TO TRANSIENT PROCESSING IN THE PHASE VOCODER; Proc. of the 6ª Int. Conference on Digital Audio Effects (DAFx-03), Londres, RU, 8-11 de septiembre de 2003.

25 Durante el estiramiento en el tiempo de las señales de audio por vocodificadores de fase, las porciones de señal transitorias son "borrosas" mediante dispersión, puesto que la llamada coherencia vertical de la señal es deteriorada. Métodos que usan los llamados métodos de sobreposición-adición, como (P)SOLA pueden generar pre- y post-ecos alterantes de eventos de sonido transitorios. Estos problemas pueden realmente ser tratados por un estiramiento en el tiempo incrementado en el medio ambiente de transitorios; sin embargo, si se va a presentar una trasposición, el factor de trasposición ya no será constante en el medio ambiente de los transitorios, esto es, el tono de los componentes de señal superpuestos (posiblemente tonales) cambiarán y serán percibidos como alteración.

35 La publicación "A Sines+Transient+Noise Audio Representation for Data Compression and Time/Pitch Scale Modifications", Proceedings of the 105ª AES Convention, 1º de enero de 1998, páginas 1-21, da a conocer un algoritmo de codificación de audio de baja velocidad binaria que permite modificaciones en el dominio comprimido. El audio de entrada está segregado en tres representaciones diferentes: sinusoides, transitorios y ruido. Cada representación puede cuantificarse individualmente y, entonces fácilmente escalarse en el tiempo y/o modificarse el tono. Debido a que se han separado los transitorios del resto de la señal, pueden tratarse de manera diferente de los senos o el ruido. Con el fin de escalar en el tiempo el audio, los componentes de senos and ruido se estirarán en el tiempo mientras que los transitorios se traducirán a tiempo.

Es un objeto de la presente invención proveer un concepto de calidad superior para la manipulación de la señal de audio.

45 Este objeto es obtenido por un aparato para manipular una señal de audio según la reivindicación 1, un método para manipular una señal de audio según la reivindicación 7 o un programa informático según la reivindicación 7 o un programa de ordenador según la reivindicación 8.

50 Para tratar los problemas de calidad que se presentan en el procesamiento sin control de las porciones transitorias, la presente invención se asegura que las porciones transitorias no sean procesadas de manera perjudicial, esto es, se procesan los eventos transitorios, pero se retiran de la señal procesada y se reemplazan por eventos transitorios no procesados.

55 Preferiblemente, las porciones transitorias insertadas a la señal procesada son copias de las porciones transitorias correspondientes en la señal de audio original, de tal manera que la señal manipulada consiste de una porción procesada que no incluye un transitorio y una porción no procesada que incluye el transitorio. Por razones de eficiencia, sin embargo, es preferido copiar una porción de la señal de audio original antes de la manipulación e insertar esta copia a la señal de audio procesada, puesto que este procedimiento garantiza que la porción transitoria en la señal procesada es idéntica al transitorio de la señal original. Este procedimiento asegurará que la alta influencia específica de transitorios en una percepción de señal de sonido sea mantenida en la señal procesada en comparación con la señal original antes del procesamiento. Así, una calidad subjetiva u objetiva con respecto a los transitorios no es degradada por cualquier clase de procesamiento de señal de audio para manipular una señal de audio.

En realizaciones preferidas, la presente solicitud proporciona un método novedoso para un tratamiento favorable perceptual de eventos de sonido transitorios dentro de la estructura de tal procesamiento, que de otra manera generaría una "borrosidad" temporal mediante dispersión de una señal.

5 Todas las apariciones de las palabras "realización/realizaciones", excepto las relacionadas a las reivindicaciones, se refieren a ejemplos útiles para entender la invención que se presentaron originalmente pero que no representan realizaciones de la presente invención reivindicada. Estos ejemplos se muestran sólo para fines ilustrativos.

10 Realizaciones preferidas de la presente invención se explican subsecuentemente con referencia a los dibujos adjuntos, en los que:

La figura 1 ilustra una realización preferida de un método o aparato de la invención para manipular una señal de audio que tiene un transitorio;

15 La figura 2 ilustra una implementación preferida de un eliminador de señal transitoria de la figura 1;

La figura 3a ilustra una implementación preferida de un procesador de señales de la figura 1;

20 La figura 3b ilustra una realización preferida adicional para implementar el procesador de señales de la figura 1;

La figura 4 ilustra una implementación preferida del insertador de señal de la figura 1;

25 La figura 5a ilustra una vista general de la implementación de un vocodificador que va a usarse en el procesador de señal de la figura 1;

La figura 5b muestra una implementación de partes (análisis) de un procesador de señales de la figura 1;

La figura 5c ilustra otras partes (estiramiento) de un procesador de señales de la figura 1;

30 La figura 5d ilustra otras partes (síntesis) de un procesador de señales de la figura 1;

La figura 6 ilustra una implementación de transformada de un vocodificador de fase que va a usarse en el procesador de señal de la figura 1;

35 La figura 7a ilustra un lado del codificador de un esquema de procesamiento de extensión de ancho de banda;

La figura 7b ilustra el lado del decodificador de un esquema de extensión de ancho de banda;

40 La figura 8a ilustra una representación de energía de una señal de entrada de audio con un evento transitorio;

La figura 8b ilustra la señal de la figura 8a, pero con un transitorio en ventana;

La figura 8c ilustra una señal sin la porción transitoria antes de ser estirada;

45 La figura 8d ilustra la señal de la figura 8c subsecuentemente a ser estirada; y

La figura 8e ilustra la señal manipulada después que la porción correspondiente de la señal original ha sido insertada.

50 La figura 9 ilustra un aparato para generar información lateral para una señal de audio.

La figura 1 ilustra un aparato preferido para manipular una señal de audio que tiene un evento transitorio. Preferiblemente, el aparato comprende un eliminador 100 de señal transitoria que tiene una entrada 101 para una señal de audio con un evento transitorio. La salida 102 del eliminador de señal transitoria se conecta a un procesador 110 de señales. La salida 111 del procesador de señales se conecta a un insertador 120 de señal. La salida 121 del insertador de señal en la cual una señal de audio manipulada con un transitorio "natural" sin procesar o sintetizado está disponible puede ser conectada a un dispositivo adicional tal como un acondicionador 130 de señal, que puede efectuar cualquier procesamiento adicional de la señal manipulada tal como toma de muestras descendente/decimación a ser requerida por propósitos de extensión de ancho de banda como se discute en relación con las figuras 7A y 7B.

Sin embargo, el acondicionador 130 de señal no puede ser usado en absoluto si la señal de audio manipulada obtenida en la salida del insertador 120 de señal es usada tal como está, esto es, es almacenada para procesamiento adicional, es transmitida a un receptor o es transmitida a un convertidor digital/análogo que, en el

extremo, es conectado a un equipo de altavoz para generar finalmente una señal de sonido que representa la señal de audio manipulada.

5 En el caso de extensión de ancho de banda, la señal 121 en la línea puede ya ser la señal de banda alta. Luego, el procesador de señales ha generado la señal de banda alta a partir de la señal de banda baja de entrada y la porción transitoria de banda baja extraída de la señal de audio 101 tendría que ser puesta en el intervalo de frecuencias de la banda ancha, que se hace preferiblemente por un procesamiento de señal que no altera la coherencia vertical, tal como decimación. Esta decimación sería efectuada antes del insertador de señal, de tal manera que la porción transitoria decimada es insertada en la señal de banda alta en la salida del bloque 110. En esta realización, el
10 acondicionador de señal efectuaría cualquier procesamiento adicional de la señal de banda alta tal como formación de envuelta, adición de ruido, filtración inversa o adición de armónicas, etc., como se hace por ejemplo en la replicación de banda espectral de MPEG 4.

15 El insertador 120 de señal recibe preferiblemente información lateral del eliminador 100 a través de la línea 123 con el fin de escoger la porción correcta de la señal sin procesar a ser insertada en 111.

20 Cuando se implementa la realizaciones que tiene los dispositivos 100, 110, 120, 130 puede obtenerse una secuencia de señales como se discute en relación con las figuras 8a a 8e. Sin embargo, no es necesariamente requerido eliminar la porción transitoria antes de efectuar la operación de procesamiento de señal en el procesador 110 de señal. En esta realización, el eliminador 100 de señal transitorio no es requerido y el insertador 120 de señal determina una porción de señal a ser cortada de la señal procesada en la salida 111 y para reemplazar esta señal cortada por una porción de la señal original como se ilustra esquemáticamente por la línea 121 o por una señal sintetizada como se ilustra por la línea 141, en donde esta señal sintetizada puede ser generada en un generador 140 de señales transitorias. Con el fin de poder generar un transitorio apropiado, el insertador 120 de señal está
25 configurado para comunicar parámetros de descripción de transitorio al generador de señales transitorias. Por consiguiente, la conexión entre los bloques 140 y 120 como se indica por el ítem 141 es ilustrada como una conexión bidireccional. Cuando se provee un detector de transitorios específico en el aparato para manipulación, entonces la información en cuanto al transitorio puede ser provista de este detector transitorio (no mostrado en la figura 1) al generador 140 de señales transitorias. El generador de señales transitorias puede ser implementado para tener muestras transitorias, que pueden directamente ser usadas o para tener muestras transitorias pre-almacenadas, que pueden ser ponderadas utilizando parámetros transitorios con el fin de generar/sintetizar realmente un transitorio que va a usarse por el insertador 120 de señal.
30

35 En una realización, el eliminador 100 de señal transitoria está configurado para eliminar una primera porción de tiempo de la señal de audio para obtener una señal de audio transitorio-reducida, en donde la primera porción de tiempo comprende el evento transitorio.

40 Además, el procesador de señales está configurado preferiblemente para procesar la señal de audio transitorio-reducida en la cual una primera porción de tiempo que comprende el evento transitorio es eliminado o para procesamiento de la señal de audio que incluye el evento transitorio para obtener la señal de audio procesada en la línea 111.

45 Preferiblemente, el insertador 120 de señal está configurado para insertar una segunda porción de tiempo a la señal de audio procesada en una ubicación de señal en donde la primera porción de tiempo ha sido eliminada o donde el evento transitorio está ubicado en la señal de audio, en donde la segunda porción de tiempo comprende un evento transitorio no influenciado por el procesamiento efectuado por el procesador 110 de señal de tal manera que se obtiene la señal de audio manipulada en la salida 121.

50 La figura 2 ilustra una realización preferida del eliminador 100 de señal transitoria. En una realización en la cual la señal de audio no incluye ninguna información lateral/meta información en cuanto a transitorios, el eliminador 100 de señal transitoria comprende un detector 103 de transitorios, un calculador 104 de desvanecimiento hacia fuera/desvanecimiento hacia adentro y una primera porción 105 de eliminador. En una realización alternativa en la cual la información en cuanto a transitorios en la señal de audio han sido recogidos como anexados a la señal de audio por un dispositivo de codificación como se discute posteriormente con respecto a la figura 9, el eliminador 100
55 de señal transitoria comprende un extractor 106 de información lateral, que extrae la información lateral anexada a la señal de audio como se indica por la línea 107. La información en cuanto al tiempo transitorio puede ser provista al calculador 104 de desvanecimiento hacia fuera/desvanecimiento hacia adentro como se ilustra por la línea 107. Sin embargo, cuando la señal de audio incluye meta-información, no (solamente) el tiempo transitorio, esto es, el tiempo exacto en el cual el evento transitorio está ocurriendo, pero el tiempo de inicio/parada de la porción que va a excluirse de la señal de audio, esto es, el tiempo de inicio y el tiempo de parada de la "primera porción" de la señal de audio, entonces el calculador 104 de desvanecimiento hacia fuera/desvanecimiento hacia adentro no es requerido también y la información de tiempo de inicio/parada puede ser enviada directamente al eliminador 105 de la primera porción como se ilustra por la línea 108. La línea 108 ilustra una opción y todas las otras líneas que son indicadas por líneas discontinuas, son opcionales también.
60

En la figura 2, el calculador 104 de desvanecimiento hacia dentro/desvanecimiento hacia fuera emite preferiblemente información 109 lateral. Esta información 109 lateral es diferente de los tiempos de inicio/parada de la primera porción, puesto que se toma en cuenta la naturaleza del procesamiento en el procesador 110 de la figura 1. Además, la señal de audio de entrada es alimentada preferiblemente al eliminador 105.

Preferiblemente, el calculador 104 de desvanecimiento hacia fuera/desvanecimiento hacia dentro proporciona los tiempos de inicio/parada de la primera porción. Estos tiempos son calculados basados en el tiempo transitorio, de tal manera que no solamente el evento transitorio, si no también algunas muestras que rodean el evento transitorio son eliminadas por el eliminador 105 de la primera porción. Además, es preferido no solo cortar la porción transitoria por una ventana rectangular de dominio de tiempo, sino efectuar la extracción mediante una porción de desvanecimiento hacia fuera y una porción de desvanecimiento hacia dentro. Para efectuar una porción de desvanecimiento hacia fuera y/o desvanecimiento hacia dentro, se puede aplicar cualquier clase de ventana que tiene una transición más suave en comparación con un filtro rectangular tal como una ventana de coseno elevada de tal manera que la respuesta de frecuencia de esta extracción no es problemática como sería cuando una ventana rectangular sería aplicada, aunque esto es también una opción. Esta operación de formación de ventana de dominio de tiempo emite el resto de la operación de ventana, esto es, la señal de audio sin la porción de ventana.

Se puede aplicar cualquier método de supresión de transitorios en este contexto incluyendo tales métodos de supresión de transitorios que conducen a una señal residual plenamente de preferencia sin transitorios o de transitorios reducidos después de la eliminación de transitorios. En comparación con la eliminación completa de la porción transitoria, en la cual la señal de audio es ajustada a cero en una cierta posición de tiempo, la supresión transitoria es ventajosa en situaciones en las cuales un procesamiento adicional de la señal de audio sufriría de porciones ajustadas a cero, puesto que tales porciones ajustadas a cero no son muy naturales para una señal de audio.

Naturalmente, todos los cálculos efectuados por el detector 103 de transitorios y el calculador 104 de desvanecimiento hacia fuera/desvanecimiento hacia dentro pueden ser aplicados también en el lado de codificación como se discute en relación con la figura 9 siempre que los resultados de estos cálculos, tales como en tiempo transitorio y/o en tiempos de inicio/parada de la primera porción sean transmitidos a un manipulador de señal, ya sea como información lateral o meta información junto con la señal de audio o separadamente de la señal de audio, tal como dentro de una señal de metadatos de audio separada que va a transmitirse a través de un canal de transmisión separado.

La figura 3a ilustra una implementación preferida del procesador 110 de señales de la figura 1. Esta implementación comprende un analizador 112 selectivo de frecuencias y un dispositivo 113 de procesamiento frecuencia-selectivo conectado subsecuentemente. El dispositivo 113 de procesamiento de frecuencia-selectivo es implementado de tal manera que aplica una influencia negativa sobre la influencia vertical de la señal de audio original. Ejemplos para este procesamiento es el estiramiento de una señal en el tiempo o el acortamiento de una señal en el tiempo en donde este estiramiento o acortamiento es aplicado de manera frecuencia-selectiva, de tal manera que, por ejemplo, el procesamiento introduce desplazamientos de fase a la señal de audio procesada, que son diferentes para las diferentes bandas de frecuencia.

Una manera de procesamiento preferida es ilustrada en la figura 3b en el contexto de un procesamiento vocodificador de fase. En general, un vocodificador de fase comprende un analizador 114 de sub-banda/transformada, un procesador 115 conectado subsecuentemente para efectuar un procesamiento frecuencia-selectivo de una pluralidad de señales de salida provistas por el ítem 114 y subsecuentemente, un combinador 116 de sub-banda/transformada, que combina las señales procesadas con el ítem 115 con el fin de obtener finalmente una señal procesada en el dominio de tiempo en la salida 117, en donde esta señal procesada en el dominio de tiempo, otra vez, una señal de pleno ancho de banda o una señal filtrada de pase bajo siempre que el ancho de banda de la señal procesada 117 sea mayor que el ancho de banda representado por una sola rama entre el ítem 115 y 116, puesto que el combinador 116 de sub-banda/transformada efectúa una combinación de señales frecuencia-selectivas.

Detalles adicionales en cuanto al vocodificador de fase son discutidos subsecuentemente en relación con las figuras 5A, 5B, 5C y 6.

Subsecuentemente, se discute y se ilustra en la figura 4 una implementación preferida del insertador 120 de señal de la figura 1. El insertador de señal comprende preferiblemente un calculador 122 para calcular la duración de la segunda porción de tiempo. Con el fin de ser aptos de calcular la duración para la segunda porción de tiempo en la realización en la cual la porción transitoria ha sido removida antes del procesamiento de señal en el procesador 110 de señal en la figura 1, la duración de la primera porción eliminada y el factor de estiramiento de tiempo (o el factor de acortamiento de tiempo) son requeridos de tal manera que se calcula la duración de la segunda porción de tiempo en el ítem 122. Estos ítems de datos pueden ser introducidos desde el exterior como se discute en relación

con las figuras 1 y 2. A modo de ejemplo, la duración de la segunda porción de tiempo es calculada al multiplicar la duración de la primera porción por el factor de estiramiento.

5 La duración de la segunda porción de tiempo es enviada al calculador 123 para calcular la primera frontera y la segunda frontera de la segunda porción de tiempo en la señal de audio. En particular, el calculador 133 puede ser implementado para efectuar un procesamiento de correlación cruzada entre la señal de audio procesada sin el evento transitorio suministrado en la entrada 124 y la señal de audio con el evento transitorio, que provee la segunda porción tal como es suministrada en la entrada 125. Preferiblemente, el calculador 123 es controlado por una entrada de control adicional 126 de tal manera que un desplazamiento positivo del evento transitorio dentro de la
10 segunda porción de tiempo es preferido contra un desplazamiento negativo del evento transitorio como se discute posteriormente.

La primera frontera y la segunda frontera de la segunda porción en el tiempo son provistas a un extractor 127. Preferiblemente, el extractor 127 corta la porción, esto es, la segunda porción de tiempo de la señal de audio original provista en la entrada 125. Puesto que se usa un desvanecedor 128 cruzado subsecuente, el corte toma lugar utilizando un filtro rectangular. En el desvanecedor 128 cruzado, la porción de inicio de la segunda porción de tiempo y la segunda porción de parada de la segunda porción de tiempo son ponderadas por un peso incrementado de 0 a 1 para la porción de inicio y/o disminución de peso de 1 a 0 en la porción del extremo, de tal manera que en esa región de desvanecimiento cruzado, la porción del extremo de la señal procesada junto con la porción de inicio de la
15 señal extraída, cuando son tomados conjuntamente, dan como resultado una señal útil. Un procesamiento similar es efectuado en el desvanecedor 128 cruzado para el extremo de la segunda porción de tiempo y el comienzo de la señal de audio procesada antes de la extracción. El desvanecimiento cruzado asegura que no se presente ningún artefacto de dominio de tiempo que de otra manera sería perceptible como artefacto de chasquido cuando las fronteras de la señal de audio procesadas sin la porción transitoria y las fronteras de la segunda porción de tiempo
20 no coinciden perfectamente de manera conjunta.

Subsecuentemente, se hace referencia a las figuras 5a, 5b, 5c y 6 con el fin de ilustrar una implementación preferida del procesador 110 de señal en el contexto de un vocodificador de fase.

30 En lo siguiente, con referencia a las figuras 5 y 6, se ilustran implementaciones preferidas para un vocodificador según la invención. La figura 5a muestra una implementación de bancos de filtros de un vocodificador de fases, en donde una señal de audio es alimentada en una entrada 500 y obtenida en una salida 510. En particular, cada canal del banco de filtros esquemáticos ilustrado en la figura 5a incluye un filtro de paso 501 de banda y un oscilador 502 aguas abajo. Las señales de salida de todos los osciladores de cada canal son combinadas por un combinador, que es implementado por ejemplo, como un adicionador e indicado en 503, con el fin de obtener la señal de salida. Cada filtro 501 es implementado de tal manera que provee una señal de amplitud por una parte y una señal de frecuencia por otra parte. La señal de amplitud y la señal de frecuencia son señales de tiempo que ilustran un desarrollo de la amplitud en un filtro 501 a lo largo del tiempo, en tanto que la señal de frecuencia representa un desarrollo de la frecuencia de la señal filtrada por un filtro 501.
35

40 Un montaje esquemático de filtro 501 es ilustrado en la figura 5b. Cada filtro 501 de la figura 5a puede ser establecido como la figura 5b, en donde, sin embargo, solamente las frecuencias f_i suministradas a los dos mezcladores 551 de entrada y el adicionador 552 son diferentes de un canal a otro. Las señales de salida del mezclador son ambas filtradas por pase bajo por los filtros 553 de pase bajo, en donde las señales de pase bajo son diferentes ya que fueron generadas por frecuencias de oscilador locales (frecuencias LO), que están fuera de fase por 90° . El filtro 553 de pase bajo superior provee una señal 554 de cuadratura, mientras que el filtro 553 inferior proporciona una señal 555 en fase. Estas dos señales, esto es, I y Q son suministradas a un transformador 556 de coordenadas que genera una representación de fase de magnitud a partir de la representación rectangular. La señal de magnitud o señal de amplitud, respectivamente, de la figura 5a con respecto al tiempo es emitida en una salida 557. La señal de fase es suministrada a un desenvolvedor 558 de fase. En la salida del elemento 558, no hay ningún valor de fase presente que está siempre entre 0 y 380° , sino un valor de fase que se incrementa linealmente. Este valor de fase "desenvuelto" es suministrado a un convertidor 559 de fase/frecuencia que puede ser implementado por ejemplo, como un formador de diferencia de fase simple que resta una fase de un punto en el tiempo previo de una fase en un punto en el tiempo actual para obtener un valor de frecuencia para el punto en el tiempo actual. Este valor de frecuencia es agregado al valor de frecuencia constante f_i del canal de filtros i para obtener un valor de frecuencia variable temporalmente en la salida 560. El valor de frecuencia en la salida 160 tiene un componente directo = f_i y un componente alternante = desviación de frecuencia por la cual una frecuencia actual de la señal en el canal del filtro se desvía de la frecuencia promedio f_i .
45
50
55

60 Así, como se ilustra en las figuras 5a y 5b, el vocodificador de fase obtiene una separación de la información espectral e información de tiempo. La información espectral está en el canal especial o en la frecuencia f_i que proporciona la porción directa de la frecuencia para cada canal, en tanto que la información de tiempo está contenida en la desviación de secuencia o la magnitud a lo largo del tiempo, respectivamente.

La figura 5c muestra una manipulación tal como es ejecutada por el incremento de ancho de banda según la inversión, en particular, en el vocodificador y en particular, en la ubicación del circuito ilustrado trazado en líneas discontinuas en la figura 5a.

5 Para el escalamiento en el tiempo, por ejemplo, las señales de amplitud $A(t)$ en cada señal o la frecuencia de las señales $f(t)$ en cada señal puede ser decimada o interpolada, respectivamente. Para fines de transposición, como es útil para la presente invención, se efectúa una interpolación esto es, una extensión o esparcimiento temporal de las señales $A(t)$ y $f(t)$ para obtener señales dispersas $A(t)$ y $f(t)$, en donde la interpolación es controlada por un factor de dispersión en un escenario de extensión de ancho de banda. Mediante la interpolación de la variación de fases, esto es, el valor antes de la adición de la frecuencia constante por el adionador 552, la frecuencia de cada oscilador 502 individual, la frecuencia de cada oscilador 502 individual en la figura 5a no es cambiada. El cambio temporal de la señal de audio global es frenado, sin embargo, esto es por el factor 2. El resultado es un tono esparcido temporalmente que tiene la tonalidad original, esto es, la onda fundamental original con sus armónicas.

15 Al efectuar el procesamiento de señales ilustrado en la figura 5c, en donde tal procesamiento es ejecutado en cada canal de banda de filtro en la figura 5a y por la señal que es luego decimada en un decimador, la señal de audio es encogida de regreso a su duración original en tanto que todas las frecuencias son duplicadas simultáneamente. Esto conduce a una transposición de tonalidad por el factor 2, en donde sin embargo, se obtiene una señal de audio que tiene la misma tonalidad como la señal de audio original, esto es, el mismo número de muestra.

20 Como una alternativa a la implementación de bancos de filtros ilustrada en la figura 5a, también se puede usar una implementación de transformada de un vocodificador de fase como se ilustra en la figura 6. Aquí, la señal 100 de audio es alimentada a un procesador de PPT o más en general, a un procesador 600 de transformación de Fourier de tiempo corto como una secuencia de muestras de tiempo. El procesador 600 de FFT es implementado esquemáticamente en la figura 6 para efectuar una formación de ventanas en el tiempo de una señal de audio con el fin de luego, por medio un FFT, calcular la magnitud y fase del espectro, en donde este cálculo es efectuado para espectros respectivos que están relacionados con bloques de la señal de audio, que están fuertemente superpuestos.

30 En un caso extremo, para cada nueva muestra de señal de audio se puede calcular un nuevo espectro, en donde un nuevo espectro puede ser calculado también, por ejemplo, solamente para cada vigésima y nueva muestra. Esta distancia a en la muestra entre dos espectros es dada preferiblemente por un controlador 602. El controlador 602 es implementado adicionalmente para alimentar un procesador 604 de IFFT que es implementado para operar en una operación de sobreposición. En particular, el procesador 604 de IFFT es alimentado de tal manera que efectúa una transformación de Fourier de corto tiempo inversa al efectuar una IFFT por espectro en base a la magnitud de fase de un espectro modificado, con el fin de efectuar luego una operación de traslape – adición de la cual se obtiene la señal de tiempo resultante. La operación de superposición-adición elimina los efectos de la ventana de análisis.

40 Se logra una dispersión de la señal de tiempo por la distancia b entre dos espectros, como son procesadas por el procesador 604 de IFFT, que es mayor que la distancia a entre los espectros en la generación de los espectros FFT. La idea básica es esparcir la señal de audio por la FFT inversa simplemente que están separadas adicionalmente, que la FFT de análisis como resultado, los cambios temporales en la señal de audio sintetizada ocurre más lentamente que en la señal de audio original.

45 Sin un rescalamiento de fase en bloque 606, sin embargo, esto conduciría a artefactos. Cuando por ejemplo, se considera un solo binario de frecuencia para el cual valores de fase sucesivos por 45° son implementados, esto implica que la señal dentro de este banco de filtros se incrementa en la fase con una proporción de $1/8$ de un ciclo, esto es, por 45° por intervalo de tiempo, en donde el intervalo de tiempo en el presente documento es el intervalo de tiempo entre FFT sucesivas. Si ahora, las FFT inversas están espaciadas entre sí, esto significa que el incremento de fase de 45° ocurre a través de un intervalo de tiempo más largo. Esto significa que debido al desplazamiento de fase, se presenta un desajuste en el proceso de traslape-adición subsecuente que conduce a una cancelación de señal indeseable. Para eliminar este artefacto, la fase es reescalada por exactamente el mismo factor por el cual la señal de audio fue esparcida en el tiempo. La fase de cada valor espectral de FFT es así incrementada por el factor b/a de tal manera que se elimina este desajuste.

55 Aunque en la realización ilustrada en la figura 5c, se obtuvo el esparcimiento mediante interpolación de las señales de control de amplitud/frecuencia para un oscilador de señal en la implementación de banco de filtros de la figura 5a, el esparcimiento en la figura 6 se obtiene por la distancia entre dos espectros de IFFT que es mayor que la distancia entre dos espectros de FFT, esto es, b es mayor que a , sin embargo, en donde para una prevención del artefacto, un reescalamiento de fase es ejecutado según b/a .

Con respecto a una descripción detallada de vocodificador de fase, se hace referencia a los siguientes documentos:

“The phase Vocoder: A tutorial”, Mark Dolson, Computer Music Journal, vol. 10, n.º 4, pág. 14 -- 27, 1986 o “New

phase Vocodificador techniques for pitch-shifting, harmonizing and other exotic effects”, L. Laroche und M. Dolson, Proceedings 1999 IEEE Workshop on applications of signal processing to audio and acoustics, New Paltz, Nueva York, 17 – 20 de octubre, 1999, páginas 91 a 94; “New approached to transient processing interphase vocodificador”, A. Röbel, Proceeding of the 6ª international conference on digital audio effects (DAFx-03), Londres, RU, 8-11 de septiembre, 2003, páginas DAFx-1 a DAFx-6; “Phase-locked Vocoder”, Meller Puckette, Proceedings 1995, IEEE ASSP, Conference on applications of signal processing to audio and acoustics o solicitud de patente estadounidense n.º 6.549.884.

Alternativamente, otros métodos para el esparcimiento de señal están disponibles, tal como por ejemplo, el método de “Pitch Synchronous Overlap Add”. Sobreposición-adición sincrónico de altura en PSOLA, es un método de síntesis en el cual las grabaciones de señales de audio están ubicadas en la base de datos. Ya que estas son señales periódicas, las mismas están dotadas de información en la frecuencia (tonalidad) fundamental y el comienzo de cada período es marcado. En la síntesis, estos períodos son cortados con un cierto medio ambiente por medio de una función de ventana y agregados a la señal que va a sintetizarse en un sitio apropiado: Dependiendo de si la frecuencia fundamental deseada es más alta o más baja que de la entrada de la base de datos, son combinadas de acuerdo con densas o menos densas que en el original. Para el ajuste de la duración del audible, los períodos pueden ser omitidos o emitidos el doble. Este método es también llamado TD-PSOLA, en donde TD significa dominio de tiempo y enfatiza que los métodos operan en el dominio de tiempo. Un desarrollo adicional es el método de sobreposición-adición de la re-síntesis de multibanda, en breve MBROLA. Aquí, los segmentos en la base de datos son traídos a una frecuencia fundamental uniforme mediante un pre-procesamiento y la posición de fases en la armónica es normalizada. Mediante esto, en la síntesis de una transmisión de un segmento al siguiente, se tienen como resultado menos interferencias perceptibles y la calidad de habla obtenida es más alta.

En una alternativa adicional, la señal de audio ya está filtrada por paso de banda antes de la dispersión, de tal manera que la señal tras la dispersión y decimación ya contiene las porciones deseadas y la filtración de paso de banda subsecuente puede ser omitida. En este caso, el filtro de paso de banda es ajustado de tal manera que la porción de la señal de audio que habría sido filtrada después de la extensión de ancho de banda está todavía contenida en la señal de salida del filtro de paso de banda. El filtro de paso de banda así contiene un intervalo de frecuencia que no está contenido en la señal de audio tras la dispersión y decimación. La señal con este intervalo de frecuencia es la señal deseada que forma la señal de alta frecuencia sintetizada.

El manipulador de señal como se ilustra en la figura 1 puede comprender adicionalmente el acondicionador 130 de señal para procesar adicionalmente la señal de audio con el transitorio “natural” sin procesar o transitorio sintetizado en la línea 121. Este acondicionador de señal puede ser un decimador de señal dentro de una aplicación de extensión de ancho de banda, que en su salida, genera una señal de alta banda que puede luego ser adaptada adicionalmente para asemejarse estrechamente a las características de la señal de alta banda original al usar los parámetros de alta frecuencia (HF) que van a transmitirse junto con una corriente de datos de HFR (reconstrucción de alta frecuencia).

Las figuras 7a y 7b ilustran un escenario de extensión de ancho de banda que puede usar ventajosamente la señal de salida del acondicionador de señal dentro del codificador de extensión de ancho de banda 720 de la figura 7b. Una señal de audio es alimentada a una combinación de pase bajo/pase alto en una entrada 700. La combinación del pase bajo/pase alto por una parte incluye un pase bajo (LP), para generar una versión filtrada por pase bajo de la señal de audio 700, ilustrada en 703 en la figura 7a. Esta señal de audio filtrada por pase bajo es codificada con un codificador de audio 704. El codificador de audio es, por ejemplo, un codificador MP3 (capa 3 de MPEG1) o un codificador AAC, también conocido como codificador MP4 y descrito en el estándar de MPEG4. Codificadores de audio alternativos que proveen una representación transparente o ventajosamente transparentes perceptualmente de la señal de audio de banda limitada 703 pueden ser usados en el codificador 704 para generar una señal de audio completamente codificada o perceptualmente codificada y de preferencia codificada de manera transparente perceptualmente 705, respectivamente.

La banda superior de la señal de audio es emitida en una salida 706 por la porción de pase alto del filtro 702, designada por “HP”. La porción de pase alto de la señal de audio, esto es, la banda superior o banda HF, también designada como porción de HF, es suministrada a un calculador 707 de parámetros que es implementado para calcular los diferentes parámetros. Estos parámetros son, por ejemplo, la envuelta espectral de la banda 706 superior en una resolución relativamente gruesa, por ejemplo, mediante representación de un factor de escala para cada grupo de frecuencias psicoacústicas o para cada banda Bark en la escala de Bark, respectivamente. Un parámetro adicional que puede ser calculado por el calculador 707 de parámetros es el piso de ruido en la banda superior, cuya energía por banda puede preferiblemente estar relacionada con la energía de la envuelta en esta banda. Parámetros adicionales que pueden ser calculados por el calculador 707 de parámetros incluyen una medida de tonalidad para cada banda parcial de la banda superior que indica cómo la energía espectral está distribuida en una banda, esto es, si la energía espectral en la banda está distribuida relativamente de manera uniforme, en donde luego existe una señal sin tono en esta banda o si la energía en esta banda está relativamente fuerte concentrada en un cierto sitio en la banda, en donde entonces más bien existe una señal tonal para esta banda.

5 Parámetros adicionales consisten en codificar explícitamente picos relativamente fuertes que sobresalen en la banda superior con respecto a su altura y su frecuencia, como el concepto de extensión de ancho de banda, en la reconstrucción sin tal codificación explícita de porciones sinusoidales prominentes en la banda superior, solamente recuperarán la misma rudimentariamente o no.

10 En cualquier caso, el calculador 707 de parámetros es implementado para generar solamente parámetros 708 para la banda superior que puede ser sometida a etapas de reducción de entropía similares ya que pueden también ser efectuados en el codificador de audio 704 para valores espectrales cuantificados, tales como por ejemplo codificación diferencial, predicción o codificación de Huffman, etc. La representación de parámetro 708 y la señal 705 de audio son luego suministrados a un formateador 709 de corriente de datos que es implementado para proveer una corriente 710 de datos lateral de salida que comúnmente será una corriente de bits según un cierto formato como es por ejemplo estandarizado en el estándar de MPEG4.

15 El lado del descodificador, ya que es especialmente apropiado para la presente invención, es en lo siguiente ilustrado con respecto a la figura 7b. La corriente 710 de datos entra a un interpretador 711 de corriente de datos que es implementado para separar la porción de parámetros relacionada con la extensión 708 de ancho de banda para la porción 705 de señal de audio. La porción 708 de parámetro es descodificada por un descodificador de parámetro 712 para obtener parámetros 713 descodificados. En paralelo a esto, la porción 705 de señal de audio es
20 decodificada por un descodificador 714 de audio para obtener una señal de audio.

25 Dependiendo de la implementación, la señal 100 de audio puede ser emitida a través de una primera salida 715. En la salida 715, una señal de audio con un ancho de banda pequeño y así también una baja calidad puede luego ser obtenida. Para una mejora de la calidad, sin embargo, la extensión 720 de ancho de banda de la invención es efectuada para obtener la señal 712 de audio en el lado de salida con un ancho de banda extendido o alto, respectivamente, y así una alta calidad.

30 Es conocido a partir del documento WO 98/57436 someter a la señal de audio a una limitación de banda en tal situación en el lado del codificador y codificar solamente una banda inferior de la señal de audio por medio de un codificador de audio de alta calidad. La banda superior, sin embargo, es solamente caracterizada muy burdamente, esto es, por un conjunto de parámetros que reproducen la envuelta espectral de la banda superior. En el lado del decodificador, la banda superior es luego sintetizada. Para este propósito, se propone una transposición armónica, en donde la banda inferior de la señal de audio descodificada es suministrada a un banco de filtros. Canales de banco de filtros de la banda inferior son conectados a canales de banco de filtros de la banda superior, o son
35 "parchados", y cada señal de paso de banda parchada es sometida a un ajuste de envuelta. El banco de filtros de síntesis perteneciente a un banco de filtros de análisis especial en el presente documento recibe así señales de paso de banda de la señal de audio en la banda inferior y señales de paso de banda envuelta-ajustada de la banda inferior que fueron parchadas armónicamente en la banda superior. La señal de salida del banco de filtro de síntesis es una señal de audio extendida con respecto a su ancho de banda, que fue transmitida del lado del codificador al
40 lado del descodificador con una velocidad de datos muy baja. En particular, los cálculos de banco de filtros y parche en el dominio de banco de filtros se pueden convertir en un alto esfuerzo computacional.

45 El método presentado en el presente documento resuelve los problemas mencionados. La novedad inventiva del método consiste en que a diferencia de los métodos existentes, una porción de ventana, que contiene el transitorio, es eliminado de la señal a ser manipulada, y que de la señal original una segunda porción de ventana (en general diferente de la primera porción) es seleccionada adicionalmente que puede ser reinsertada a la señal manipulada, de tal manera que la envuelta temporal es conservada tanto como sea posible en el medioambiente del transitorio. Esta segunda porción es seleccionada de tal manera que encajará exactamente al rebajo cambiado por la operación de estiramiento en el tiempo. El encaje o ajuste exacto es efectuado mediante el cálculo del máximo de la
50 correlación cruzada de los bordes del rebajo resultante con los bordes de la porción transitoria original.

Entonces, la calidad de audio subjetiva del transitorio ya no es deteriorada por la dispersión y efectos de eco.

55 La determinación precisa de la posición del transitorio para el fin de seleccionar una porción apropiada puede ser efectuada por ejemplo utilizando un cálculo de centroide móvil de la energía en un período de tiempo apropiado.

60 Junto con el factor de estiramiento de tiempo, el tamaño de la primera porción determina el tamaño requerido de la segunda porción. Preferiblemente, este tamaño será seleccionado de tal manera que más de un transitorio es acomodado por la segunda porción usada para reinserción solamente si el intervalo de tiempo entre los transitorios estrechamente adyacentes está por debajo del umbral para la perceptibilidad humana de eventos temporales individuales.

El ajuste óptimo del transitorio según la correlación cruzada máxima puede requerir un ligero desplazamiento en tiempo en relación con la posición original del mismo. Sin embargo, debido a la existencia de efectos de pre- y

particularmente post-enmascaramiento temporales, la posición del transitorio reinsertado no necesita coincidir de manera precisa con la posición original. Debido al período de acción prolongado del post-enmascaramiento, un desplazamiento del transitorio en la dirección de tiempo positiva será preferido.

5 Al insertar la porción de señal original, el timbre o tonalidad de la misma será cambiado cuando la velocidad de toma de muestras sea cambiada por una etapa de decimación subsecuente. En general, sin embargo, esto es enmascarado por el transitorio mismo por medio de mecanismos de enmascaramiento temporal psicoacústicos. En particular, si se presenta el estiramiento por un factor entero, el timbre solamente será cambiado ligeramente, puesto que fuera del medio ambiente del transitorio, solamente cada n-ésimo onda armónica ($n = \text{factor de estiramiento}$) será ocupada.

15 Utilizando el nuevo método, se impiden efectivamente artefactos (dispersión, pre- y post-ecos) que resultan durante el procesamiento de transitorios por medio de métodos de transposición y estiramiento de tiempo. Se evita el deterioro potencial de la calidad de porciones de señal superpuestas (posibles tonales).

El método es apropiado para cualquier aplicación de audio en donde las velocidades de reproducción de señales de audio o sus tonalidades van a ser cambiados.

20 Subsecuentemente, se discute una realización preferida en el contexto de las figuras 8a a 8e. La figura 8a ilustra una representación de la señal de audio, pero a diferencia de una secuencia de muestras de audio de dominio de tiempo directa, la figura 8a ilustra una representación de envuelta de energía, que puede por ejemplo, ser obtenida cuando cada muestra de audio en una ilustración de muestra de dominio de tiempo es elevada al cuadrado. Específicamente, la figura 8a ilustra una señal de audio 800 que tiene un evento 801 transitorio, en donde el evento transitorio está caracterizado por un incremento y disminución aguda de energía a lo largo del tiempo. Naturalmente, un transitorio también sería un incremento agudo de energía cuando esta energía permanece en un cierto alto nivel o una disminución aguda de energía cuando la energía ha estado en un alto nivel por un cierto tiempo antes de la disminución. Un patrón específico para un transitorio es, por ejemplo un aplauso de manos o cualquier otro tono generado por un instrumento de percusión. Adicionalmente, los transitorios son ataques rápidos de un instrumento, que empieza a tocar un tono fuertemente, esto es, proporciona energía de sonido a una cierta banda o una pluralidad de bandas por encima de un cierto nivel de umbral debajo de un cierto tiempo de umbral. Naturalmente, otra fluctuación de energía, tal como la fluctuación 802 de energía de la señal 800 de audio en la figura 8a no son detectados como transitorios. Detectores de transitorios son conocidos en el arte y son descritos extensamente en la literatura y dependen de muchos diferentes algoritmos que pueden comprender procesamiento frecuencia-selectivo y una comparación de un resultado de un procesamiento frecuencia-selectivo con un umbral y una decisión subsecuente si hubo o no un transitorio.

40 La figura 8b ilustra un transitorio en ventana. El área delimitada por la línea continua es restada de la señal ponderada por la forma de ventana ilustrada. El área marcada por la línea discontinua es agregada después del procesamiento. Específicamente, el transitorio que se presenta a un cierto tiempo 803 transitorio tiene que ser cortado de la señal 800 de audio. Para estar en el lado seguro, no solamente el transitorio, sino también algunas muestras adyacentes/vecinas van a ser cortadas de la señal original. Por consiguiente, la primera porción 804 de tiempo es determinada, en donde la primera porción de tiempo se extiende desde un instante de tiempo 805 de partida a un instante 806 de tiempo de parada. En general, la primera porción 804 de tiempo es seleccionado de tal manera que el tiempo 803 transitorio está incluido dentro de la primera porción 804 de tiempo. La figura 8c ilustra una señal sin un transitorio antes de ser estirada. Como puede observarse de los bordes 807 y 808 que decaen lentamente, la primera porción de tiempo no es cortada por un ajustador rectangular/formador de ventanas, sino que se efectúa una probación de ventanas para tener bordes que decaen lentamente o flancos de la señal de audio.

50 De manera importante, la figura 8c ilustra ahora la señal 102 de audio en la línea de la figura 1, esto es, subsecuente a la eliminación de la señal transitoria. Los flancos 807, 808 de decaimiento/incremento lento proporcionan la región de desvanecimiento hacia adentro o desvanecimiento hacia fuera a ser usada por el desvanecedor 120 cruzado de la figura 4. La figura 8d ilustra la señal de la figura 8c, pero en un estado estirado, esto es, subsecuente al procesamiento aplicado por el procesador 110 de señales. Así, la señal en la figura 8d es la señal en la línea 111 de la figura 1. Debido a la operación de estiramiento, la primera porción 804 se ha vuelto mucho más larga. Así, la primera porción 804 de la figura 8d ha sido estirada a la segunda porción 809 de tiempo, que tiene el instante 810 de inicio de la segunda porción de tiempo y el instante 811 de parada de la segunda porción de tiempo. Al estirar la señal, los flancos 807, 808 tienen que ser estirados también, de tal manera que la tonalidad de tiempo de los flancos 807', 808' ha sido estirada también. Este estiramiento ha sido tomado en cuenta cuando se calcula la duración de la segunda porción de tiempo tal como se efectúa por el calculador 122 de la figura 4.

60 Tan pronto como se determina la duración de la segunda porción de tiempo, una porción correspondiente a la duración de la segunda porción de tiempo es cortada de la señal de audio original ilustrada en la figura 8a como se indica por las líneas discontinuas en la figura 8b. Para este fin, la segunda porción 809 de tiempo ha entrado a la figura 8e. Como se discute, el instante 812 de tiempo de inicio, esto es, la primera frontera de la segunda porción

809 de tiempo en la señal de audio original y el instante 813 de tiempo de parada de la segunda porción de tiempo, esto es, la segunda frontera de la segunda porción de tiempo en la señal de audio original no tienen que ser necesariamente simétricas con respecto al tiempo 803, 803' de evento transitorio de tal manera que el transitorio 801 está localizado en exactamente el mismo instante de tiempo como estuvo en la señal original. En lugar de esto, los instantes 812, 813 de tiempo de la figura 8b se pueden hacer variar ligeramente, de tal manera que la correlación cruzada da como resultado una forma de señal sobre estas fronteras en la señal original es tanto como sea posible, similar a porciones correspondientes en la señal estirada. Así, la posición real del transitorio 803 puede ser movida fuera de centro de la segunda porción de tiempo hasta un cierto grado, que es indicado en la figura 8e por el número de referencia 803' que indica un cierto tiempo con respecto a la segunda porción de tiempo, que se desvía del tiempo correspondiente 803 con respecto a la segunda porción de tiempo en la figura 8b. Como se discute en relación con la figura 4, el ítem 126, un desplazamiento positivo del transitorio a un tiempo 803' con respecto a un tiempo 803 es preferido debido al efecto de post-enmascaramiento, que es más pronunciado que el efecto de pre-enmascaramiento. La figura 8e ilustra adicionalmente las regiones 813a, 813b de cruce/transición en las cuales el desvanecedor 128 cruzado proporciona un desvanecimiento cruzado entre la señal estirada sin el transitorio y la copia de la señal original que incluye el transitorio.

Como se ilustra en la figura 4, el calculador para calcular la duración de la segunda porción 122 de tiempo está configurado para recibir la duración de la primera porción de tiempo y el factor de estiramiento. Alternativamente, el calculador 122 puede también recibir información en cuanto a la permisibilidad de transitorios vecinos a ser incluidos dentro de una y la misma primera porción de tiempo. Por consiguiente, basándose en esta permisibilidad, el calculador puede determinar la duración de la primera porción 804 de tiempo por sí mismo y, dependiendo del factor de estiramiento/acortamiento, calcula luego la duración de la segunda porción 809 de tiempo.

Como se discutió anteriormente, la funcionalidad del insertador de señal es que el insertador de señal remueve un área apropiada para el espacio en la figura 8e, que es ampliado dentro de la señal estirada de la señal original y encaja a esta área apropiada, esto es, la segunda porción de tiempo a la señal procesada utilizando un cálculo de correlación cruzada para determinar el instante 812 y 813 de tiempo y de preferencia, efectuando una operación de desvanecimiento cruzado en las regiones 813a y 813b de desvanecimiento cruzado también.

La figura 9 ilustra un aparato para generar información lateral para una señal de audio, que puede ser usada en el contexto de la presente invención cuando la detección de transitorios es efectuada en el lado del codificador y la información lateral concerniente con esta detección de transitorios es calculada y transmitida a un manipulador de señal, que representaría entonces el lado del decodificador. Para este fin, un detector de transitorios similar al detector de transitorio 103 en la figura 2 es aplicado para analizar la señal de audio que incluye un evento transitorio. El detector de transitorios calcula un tiempo transitorio, esto es, en el tiempo 803 en la figura 1 y envía este tiempo transitorio a un calculador 104' de metadatos, que puede ser estructurado similarmente al calculador 104' de desvanecimiento hacia adentro/desvanecimiento hacia fuera en la figura 2. En general, el calculador 104' de metadatos puede calcular metadatos que van a enviarse a una interfaz 900 de salida de señal en donde estos metadatos comprenden fronteras para la eliminación de transitorios, esto es, fronteras para la primera porción de tiempo, esto es fronteras 805 y 806 de la figura 8b o fronteras para la inserción del transitorio (segunda porción de tiempo) como se ilustra en 812, 813 en la figura 8b o el instante de tiempo del evento transitorio 803 o aún 803'. Aún en este último caso, el manipulador de señales estaría en posición para determinar todos los datos requeridos, esto es, los datos de la primera porción de tiempo, los datos de la segunda porción de tiempo, esto es, en base a un instante de tiempo 803 de evento transitorio.

Los metadatos tal como son generados por el ítem 104' son enviados a la interfaz de salida de señal de tal manera que la interfaz de salida de señal genera una señal, esto es, una señal de salida para transmisión o almacenamiento. La señal de salida puede incluir solamente los metadatos o puede incluir los metadatos y la señal de audio en donde, en el último caso, los metadatos representarían información lateral para la señal de audio. Para este fin, la señal de audio puede ser enviada a la interfaz 900 de salida de señal a través de la línea 901. La señal de salida generada por la interfaz 900 de salida de señal puede ser almacenada en cualquier clase de medio de almacenamiento o puede ser transmitida a través de cualquier clase de canal de transmisión a un manipulador de señal o cualquier otro dispositivo que requiere información transitoria.

Debe observarse que aunque la presente invención ha sido descrita en el contexto de diagramas de bloques, en donde los bloques representan componentes de hardware reales o lógicos, la presente invención puede también ser implementada mediante un método implementado por ordenador. En este último caso, los bloques representan etapas de métodos correspondientes, en donde estas etapas significan las funcionalidades efectuadas por bloques de hardware lógicos o físicos correspondientes.

Las realizaciones descritas son solamente ilustrativas para los principios de la presente invención. Se comprende que modificaciones y variaciones de los fragmentos y los detalles descritos en el presente documento resultarán evidentes para otros expertos en la técnica. Es la intención, por consiguiente, estar limitados solamente por el alcance de las reivindicaciones de patente pendientes y no por los detalles específicos presentados a manera de

descripción y explicación de las realizaciones en el presente documento.

5 Dependiendo de ciertos requerimientos de implementación de los métodos de la invención, los métodos de la invención pueden ser implementados en hardware o en software. La implementación puede ser efectuada utilizando un medio de almacenamiento digital, en particular, un disco, un DVD o un CD que tiene señales de control que se pueden leer electrónicamente almacenadas en el mismo, que cooperan con sistemas informáticos programables, de tal manera que los métodos de la invención son efectuados. En general, la presente invención puede por consiguiente ser implementada como un producto de programas informáticos con códigos de programa almacenados en un portador que se puede leer con la máquina, los códigos de programa son puestos en operación para efectuar los métodos de la invención cuando el producto de programas informáticos se ejecuta en un ordenador. En otras palabras, los métodos de la invención son por consiguiente un programa informático que tiene un código de programa para efectuar al menos uno de los métodos de la invención cuando el programa informático se ejecuta en un ordenador. La señal de metadatos de la invención puede ser almacenada en cualquier medio de almacenamiento que se puede leer por la máquina tal como un medio de almacenamiento digital.

10

15

REIVINDICACIONES

1. Aparato para manipular una señal de audio que comprende un evento (801) transitorio en una primera porción (804) de tiempo de la señal de audio, comprendiendo el aparato:
 - 5 un procesador (110) de señal para procesar la señal de audio que comprende el evento (801) transitorio en la primera porción (804) de tiempo para obtener una señal de audio procesada que comprende una porción transitoria perceptualmente degradada;
 - 10 un insertador (120) de señales para insertar una segunda porción (809) de tiempo a la señal de audio procesada en un sitio de señal, en donde la porción transitoria perceptualmente degradada se sitúa en la señal de audio procesada, de modo que se obtiene una señal de audio manipulada,
 - 15 en el que la segunda porción (809) de tiempo comprende el evento (801) transitorio no influenciado por el procesamiento efectuado por el procesador (110) de señal,
 - 20 en el que el procesador (110) de señal se configura para generar la porción transitoria perceptualmente degradada en la señal de audio procesada mediante estiramiento o acortamiento de la señal de audio que comprende el evento (801) transitorio de modo que la señal de audio procesada tiene una duración mayor que o menor que la señal de audio, y
 - 25 en el que la segunda porción (809) de tiempo tiene una duración diferente de una duración de la primera porción (804) de tiempo, en la que en el caso de estiramiento, la segunda porción (809) de tiempo es mayor que la primera porción (804) de tiempo o en el caso de acortamiento, la segunda porción (809) de tiempo es menor que la primera porción (804) de tiempo.
2. Aparato según la reivindicación 1, en el que el procesador (110) de señales comprende un vocodificador, un vocodificador de fase o un procesador (P)SOLA.
- 30 3. Aparato según una cualquiera de las reivindicaciones anteriores, que comprende además un acondicionador (130) de señal para acondicionar la señal de audio manipulada mediante decimación o interpolación de una versión de tiempo-discreta de la señal de audio manipulada.
- 35 4. Aparato según una cualquiera de las reivindicaciones anteriores, en el que el insertador (120) de señales está configurado:
 - 40 para determinar (122) la duración de tiempo de la segunda porción (809) de tiempo a ser copiada de la señal de audio que tiene el evento (801) transitorio en la primera porción (804) de tiempo;
 - 45 para determinar (123) un instante de tiempo de inicio de la segunda porción (809) de tiempo o un instante de tiempo de parada de la segunda porción (809) de tiempo al encontrar un máximo de un cálculo de correlación cruzada, de tal manera que la frontera de la segunda porción (809) de tiempo coincide con una frontera correspondiente de la señal de audio procesada, tanto como sea posible,
 - 50 en el que una porción en el tiempo (803') del evento (801) transitorio en la señal de audio manipulada coincide con la posición (803) en el tiempo del evento (801) transitorio en la señal de audio o se desvía de la posición en el tiempo (803) del evento (801) transitorio en la señal de audio por una diferencia de tiempo menor que un grado psicoacústicamente tolerable determinado por un pre-enmascaramiento o post-enmascaramiento del evento (801) transitorio.
5. Aparato según una cualquiera de las reivindicaciones anteriores, que comprende además:
 - 55 un detector (103) de transitorios para detectar el evento (801) transitorio en la señal de audio o
 - un extractor (106) de información lateral para extraer o interpretar una información lateral asociada con la señal de audio, la información lateral indica una posición (803) de tiempo del evento (801) transitorio o que indica un instante de tiempo de inicio o un instante de tiempo de parada de la primera porción (804) de tiempo o la segunda porción (809) de tiempo.
- 60 6. Método de manipulación de una señal de audio que comprende un evento (801) transitorio en una primera porción (804) de tiempo de la señal de audio, comprendiendo el método:
 - procesar (110) la señal de audio que comprende el evento (801) transitorio en la primera porción (804) de tiempo para obtener una señal de audio procesada que comprende una porción transitoria perceptualmente

degradada;

5 insertar (120) una segunda porción (809) de tiempo a la señal de audio procesada en un sitio de señal, en donde la porción transitoria perceptualmente degradada está ubicada en la señal de audio procesada, de tal manera que se obtiene una señal de audio manipulada.

10 en el que la segunda porción (809) de tiempo comprende el evento (801) transitorio no influenciado por la etapa de procesamiento (110), en el que la etapa de procesamiento (110) genera la porción transitoria perceptualmente degradada en la señal de audio procesada mediante estiramiento o acortamiento de la señal de audio que comprende el evento (801) transitorio, de modo que la señal de audio procesada tiene una duración mayor que o menor que la señal de audio, y

15 en el que la segunda porción (809) de tiempo tiene una duración diferente de una duración de la primera porción (804) de tiempo, en el que en el caso de estiramiento, la segunda porción (809) de tiempo es mayor que la primera porción (804) de tiempo o en el caso de acortamiento, la segunda porción (809) de tiempo es menor que la primera porción (804) de tiempo.

20 7. Programa de ordenador que tiene un código de programa para efectuar, cuando es ejecutado en un ordenador, el método según la reivindicación 6.

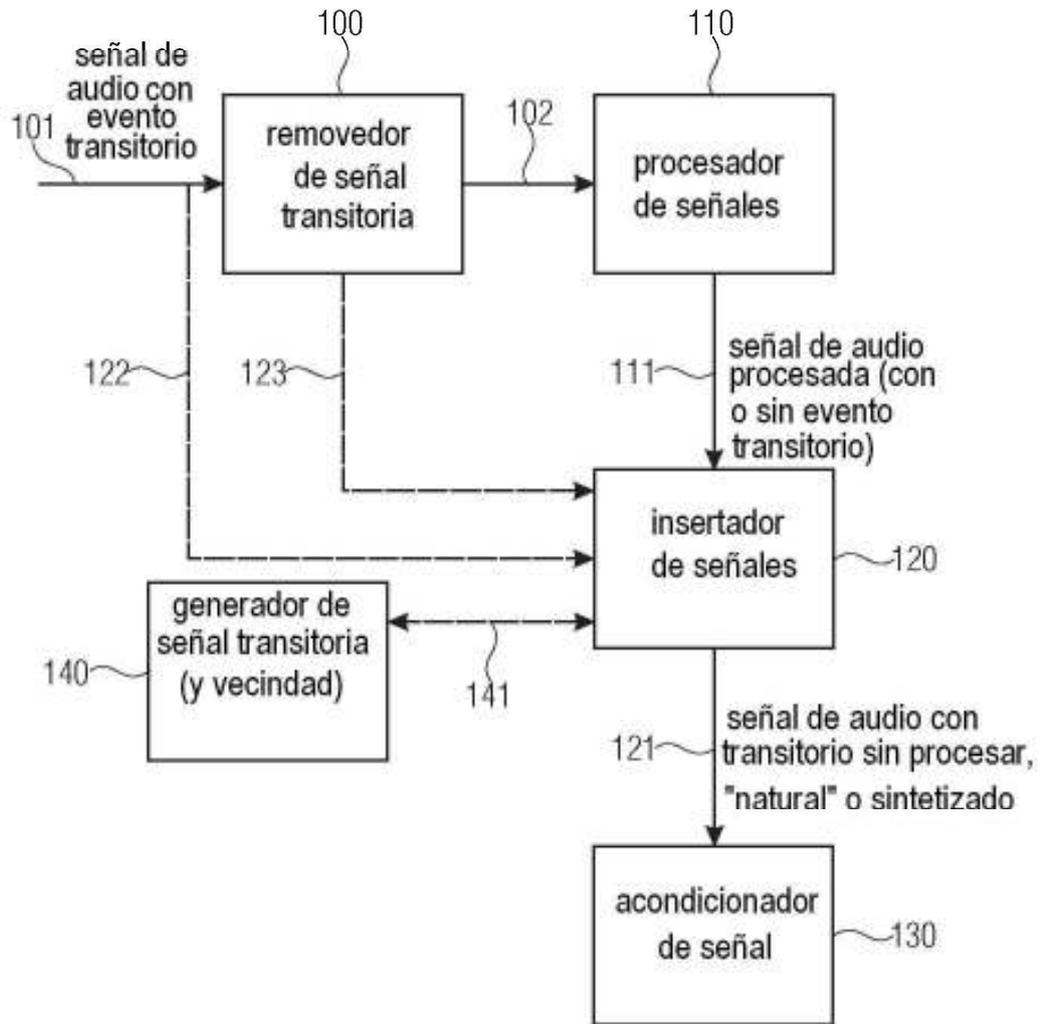


FIGURA 1

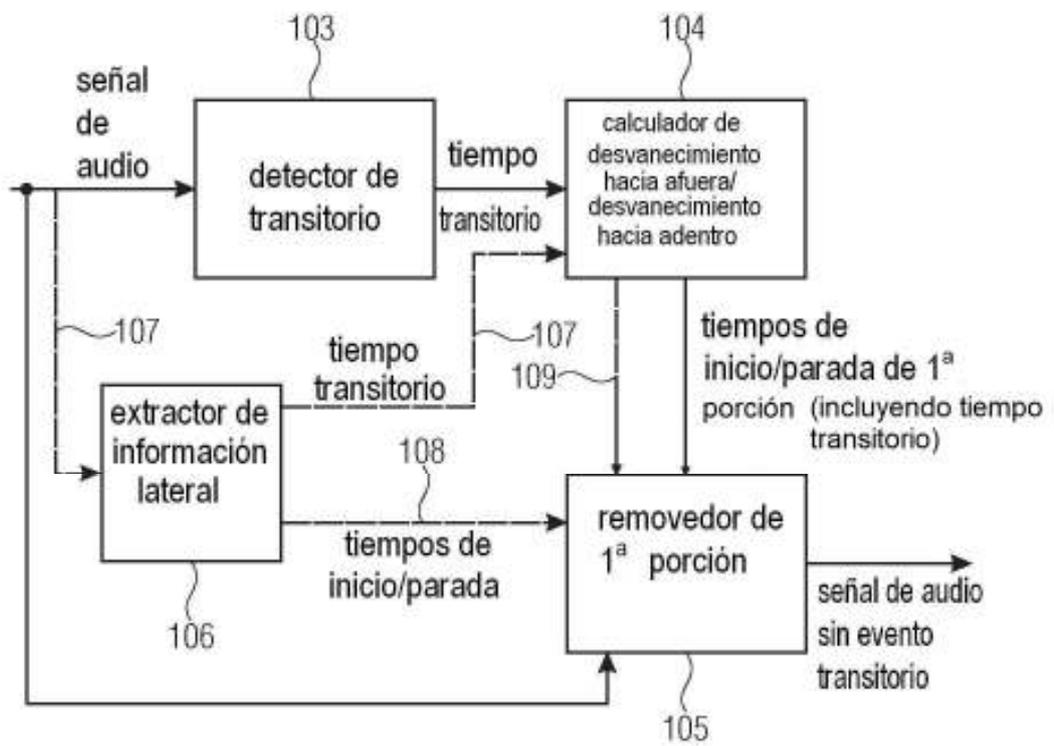


FIGURA 2

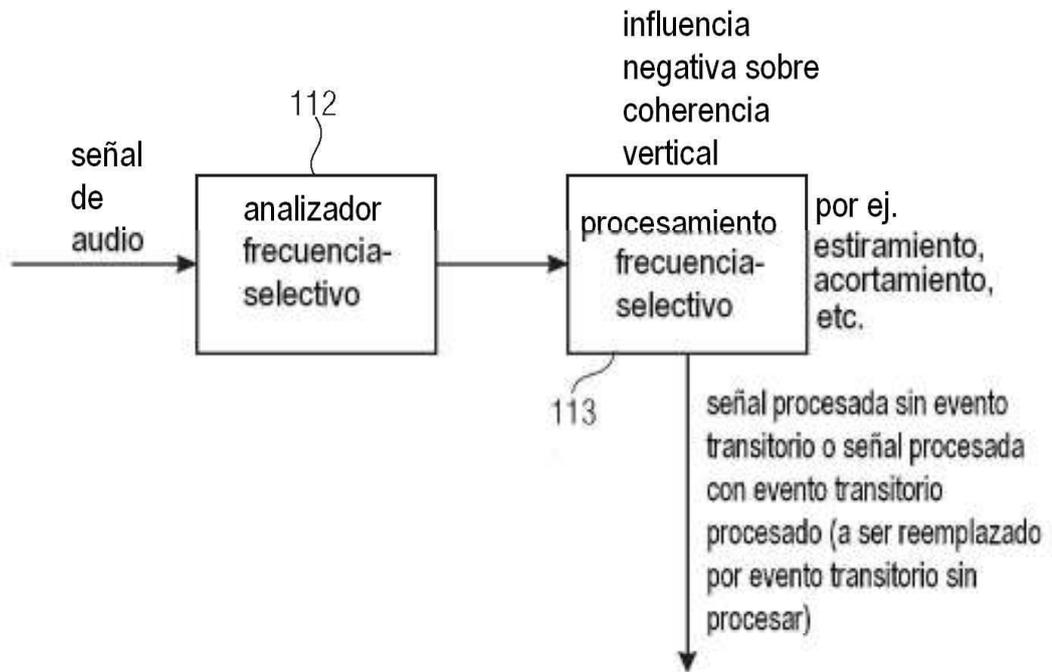


FIGURA 3A

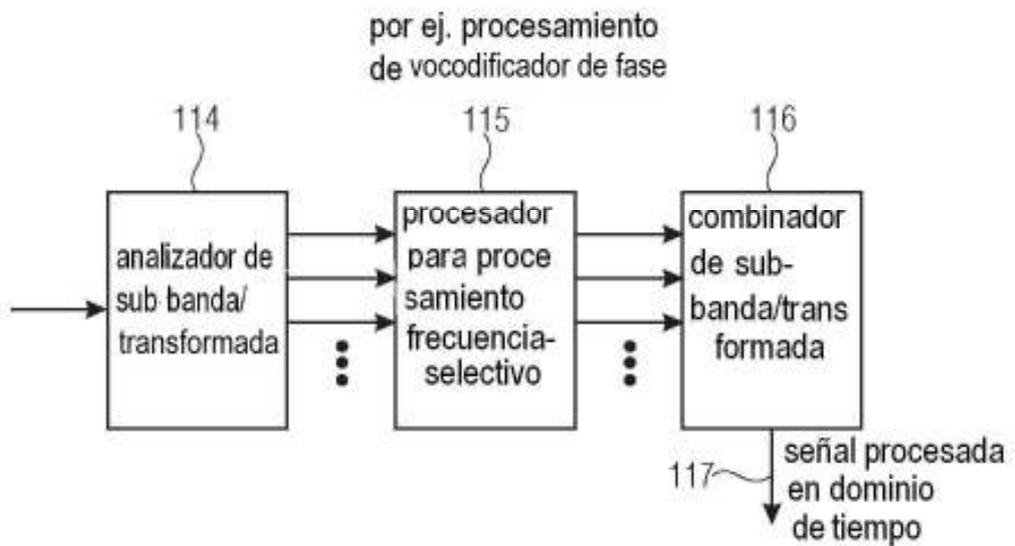


FIGURA 3B

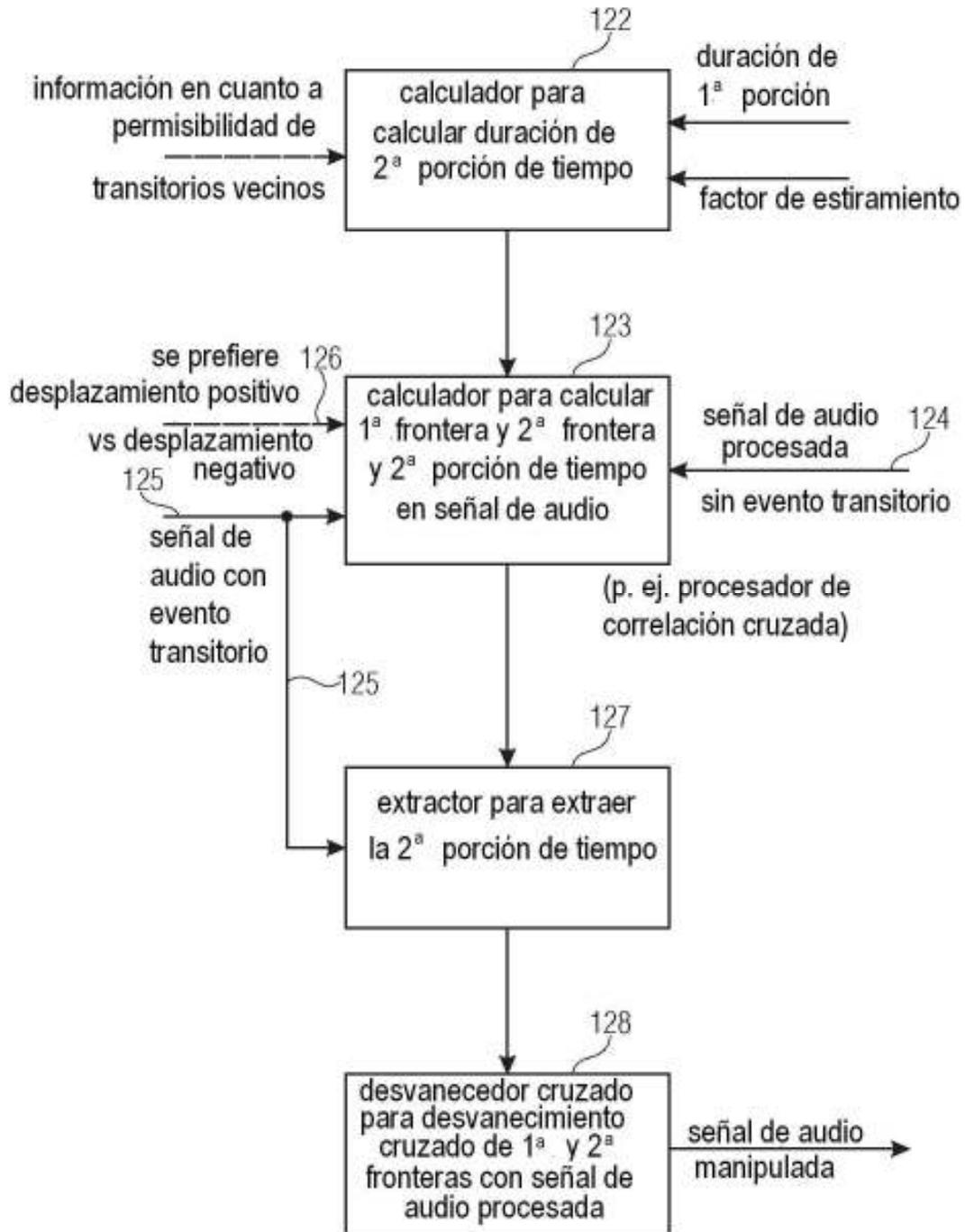


FIGURA 4

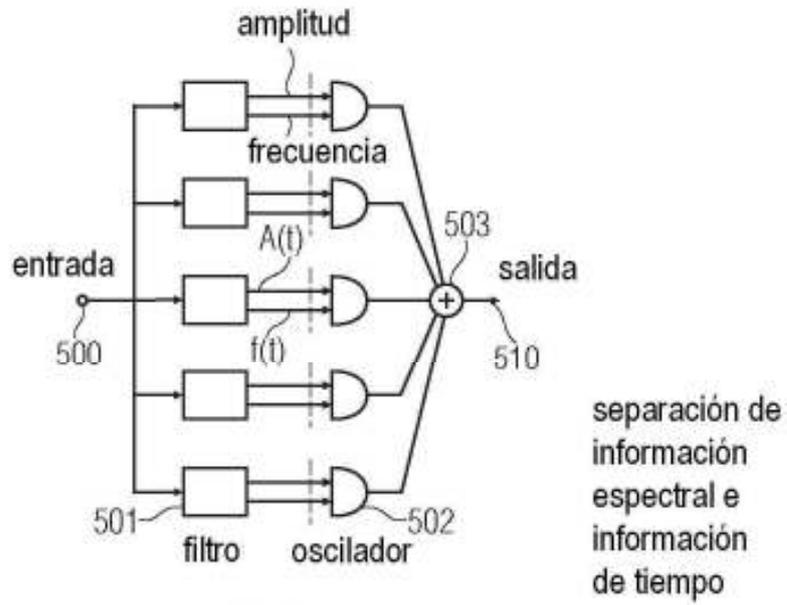


FIGURA 5A

(implementación de banco de filtros)

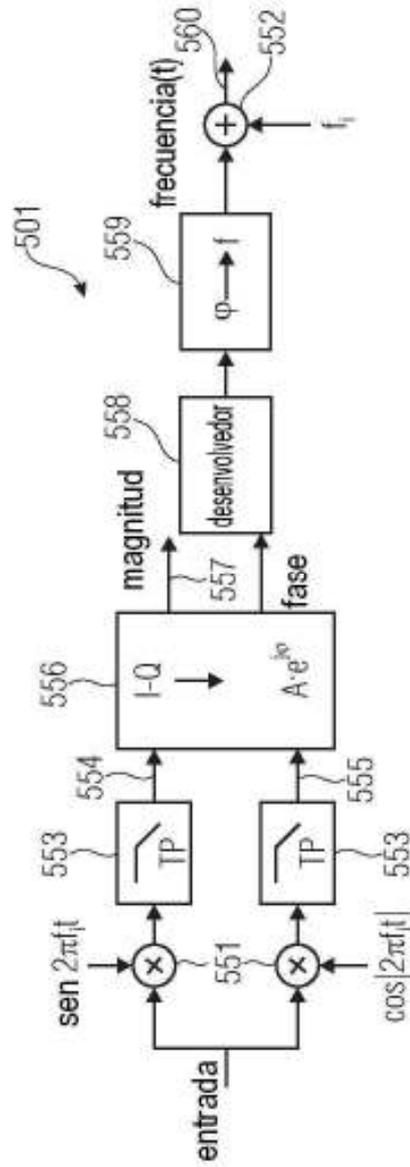


FIGURA 5B

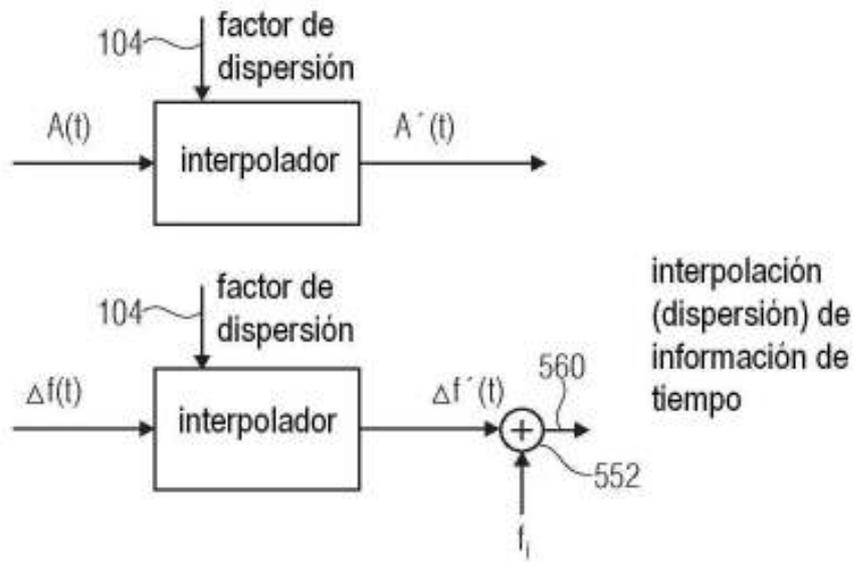


FIGURA 5C

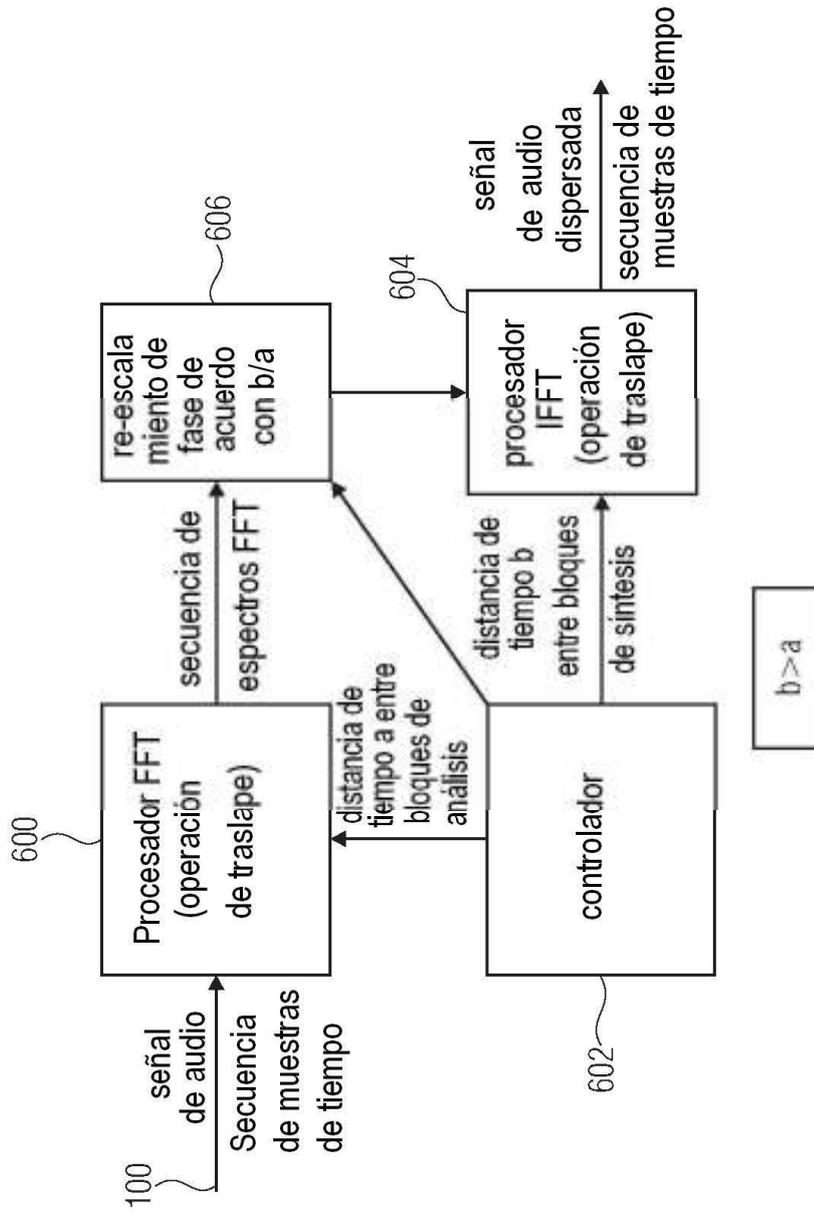


FIGURA 6
(implementación de transformada)

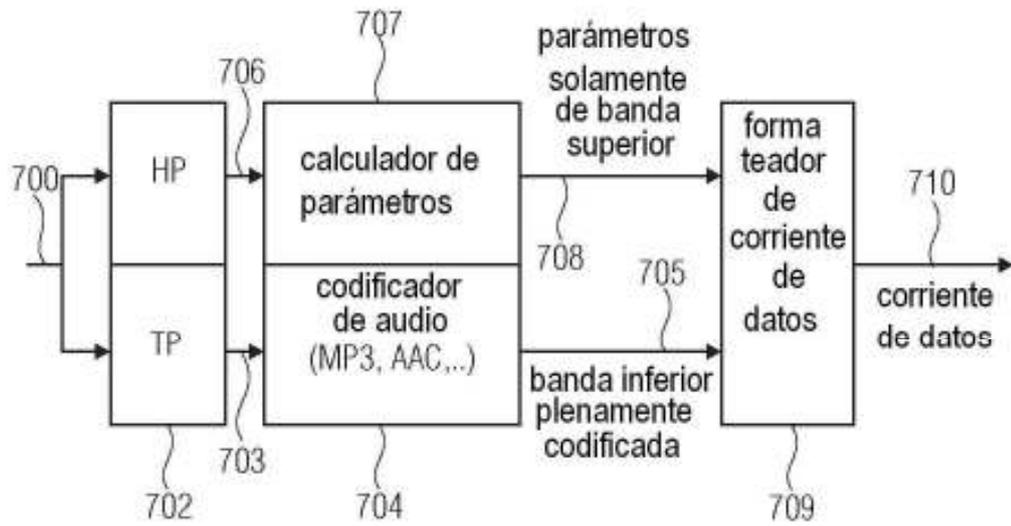


FIGURA 7A
(Lado del codificador)

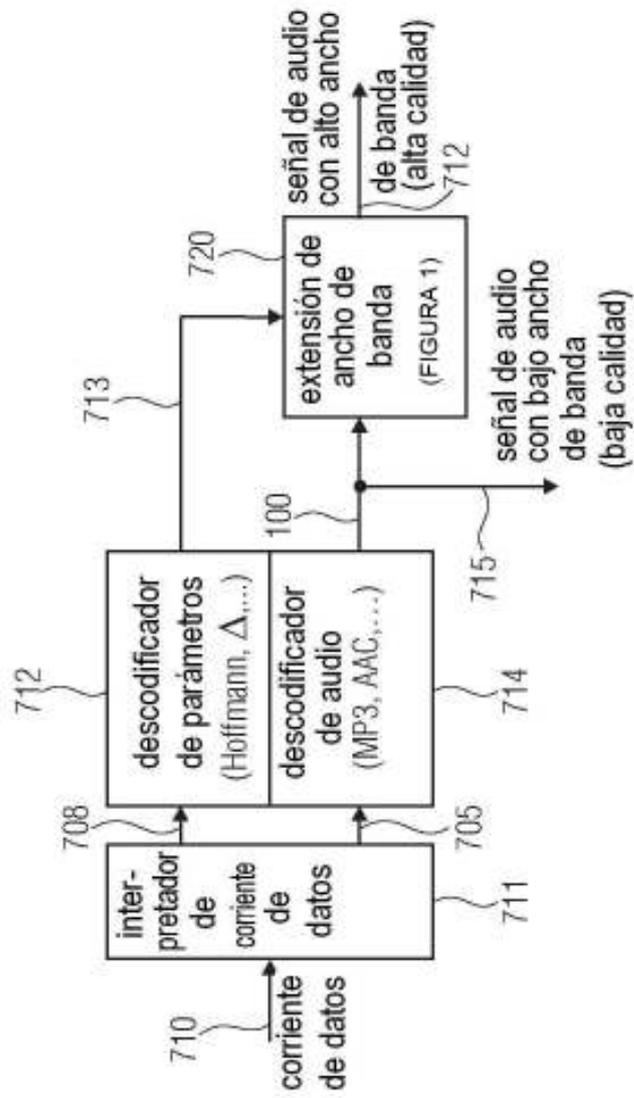


FIGURA 7B
(Lado del descodificador)

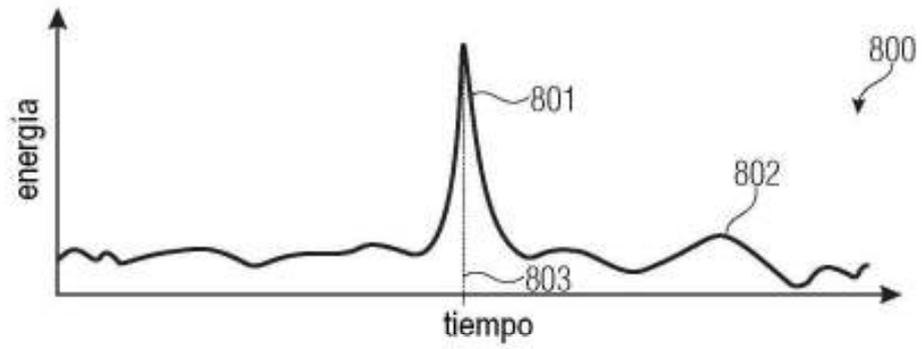


FIGURA 8A

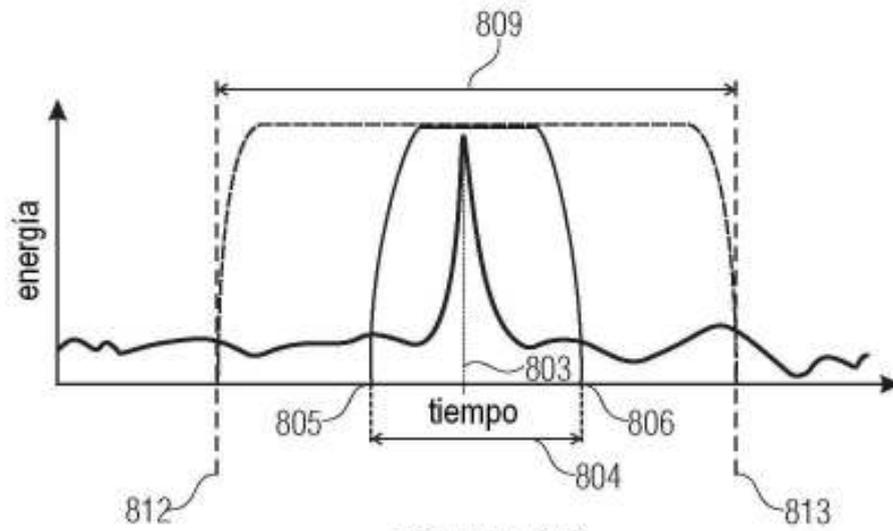


FIGURA 8B

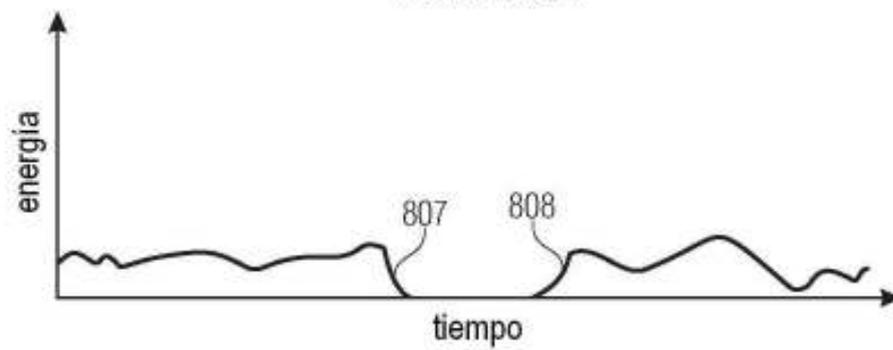


FIGURA 8C

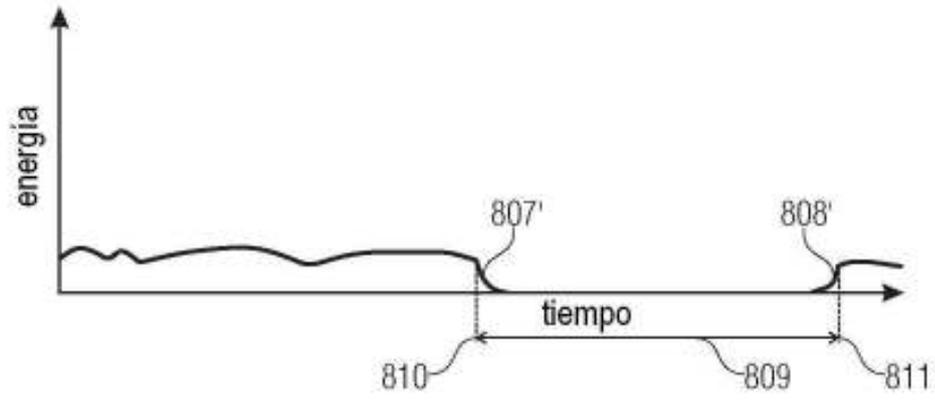


FIGURA 8D

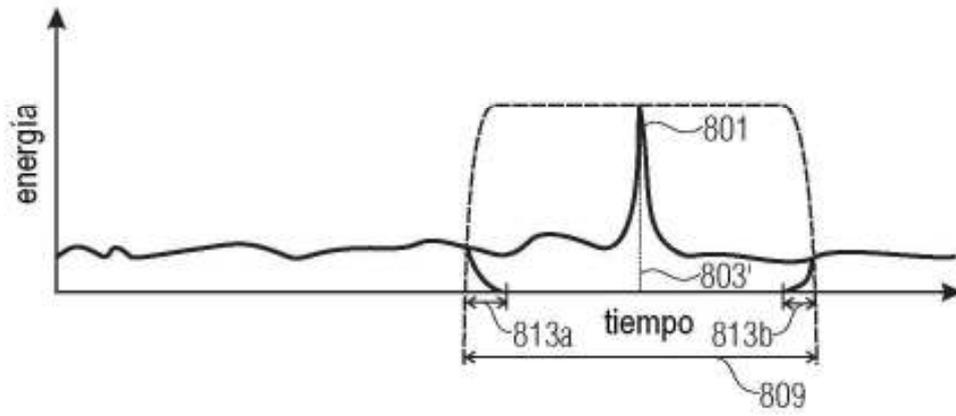


FIGURA 8E

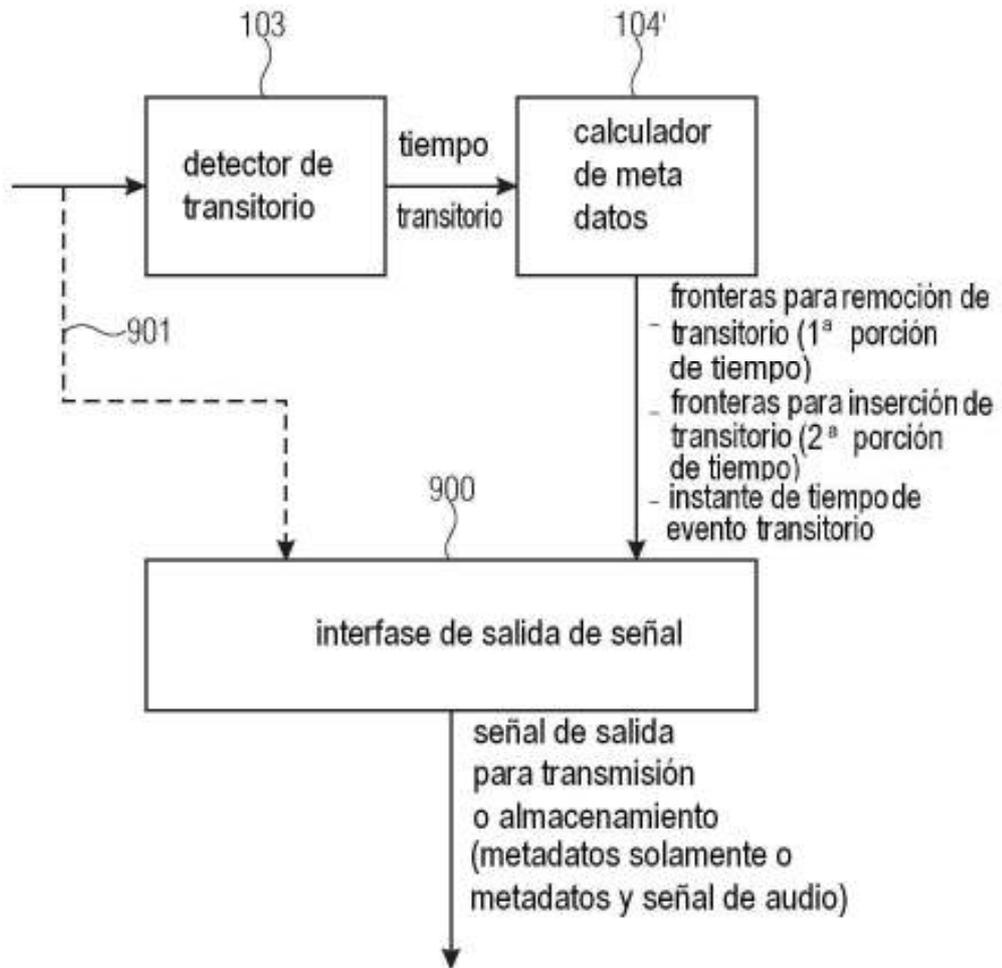


FIGURA 9