

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 763 838**

51 Int. Cl.:

H03G 9/00 (2006.01)

H03G 9/14 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **07.03.2011 E 17170132 (9)**

97 Fecha y número de publicación de la concesión europea: **27.11.2019 EP 3236586**

54 Título: **Sistema para combinar mediciones de sonoridad en un único modo de reproducción**

30 Prioridad:

10.03.2010 US 312561 P

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

01.06.2020

73 Titular/es:

**DOLBY INTERNATIONAL AB (50.0%)
Apollo Building, 3E, Herikerbergweg 1-35
1101 CN Amsterdam Zuidoost, NL y
DOLBY LABORATORIES LICENSING CORP.
(50.0%)**

72 Inventor/es:

**WOLTERS, MARTIN;
SCHUG, MICHAEL;
MUNDT, HARALD y
RIEDMILLER, JEFFREY**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 763 838 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

Sistema para combinar mediciones de sonoridad en un único modo de reproducción

Campo técnico

5 El presente documento está relacionada con procesamiento de datos multimedia, en particular la codificación, la transmisión, la descodificación y la presentación de datos multimedia, p. ej., archivos de audio o flujos de bits. En particular, el presente documento está relacionado con la implementación de control de sonoridad en reproductores multimedia.

Antecedentes de la invención

10 El problema de niveles variables de mezcla y reproducción de contenido de audio se aborda en la industria de películas al especificar las recomendaciones de la SMPTE (*Society of Motion Picture and Television Engineers*, Sociedad de Ingenieros de Imágenes en Movimiento y Televisión) que garantiza un nivel de reproducción consistente en cines y para diferente contenido. Las recomendaciones de la SMPTE aseguran que el contenido de audio es reproducido en un nivel consistente agradable para los consumidores.

15 La situación en la difusión es más desafiante, dado que los sistemas de reproducción individuales de los usuarios no son controlados por técnicos y debido a canales de distribución y redes más complejos para la difusión. Con la introducción de la difusión digital, la industria estableció el concepto de metadatos variables en el tiempo que permite controlar valores de ganancia en el extremo de recepción para confeccionar contenido a un ambiente de escucha específico. Un ejemplo son los metadatos incluidos en Dolby Digital que incluyen normalización general de información de sonoridad ("dialnorm") para diálogos, así como palabras de ganancia ("dynrng" y "comp") para reducir el rango dinámico de un programa. Cabe señalar que por toda esta memoria descriptiva y en las reivindicaciones, referencias a Dolby Digital se entenderán como que engloban ambos sistemas de codificación Dolby Digital y Dolby Digital Plus. Tales sistemas son específicamente potentes para situaciones donde se especifican los modos de funcionamiento en el receptor relativos al ambiente de escucha y las preferencias de escucha. A modo de ejemplo, el estándar dialnorm permite la especificación de los llamados "modo de línea" y "modo de RF" para Dolby Digital. El "modo RF" se diseña para situaciones de limitación de pico donde el programa descodificado está pensado para entrega a través de una entrada de RF en una televisión, tal como a través de la salida de antena de un descodificador. El "Modo línea" proporciona menos compresión del rango dinámico que el "modo RF" y también permite ajuste de usuario de los parámetros de realce de nivel bajo y de corte de alto nivel dentro de un descodificador doméstico. El ajuste o "escalado" de las áreas de realce y corte permite al usuario personalizar la reproducción de audio para su ambiente de escucha específico. Estas tecnologías también son parte de discos de audio/vídeo actuales como DVD y Blu-ray.

20 Un canal de distribución importante para contenido de audio es todavía el CD que contiene datos PCM de 16 bits sin metadatos. Se dice que la normalización de pico usada típicamente para los CD es la razón principal para la llamada "guerra de sonoridad" que ha llevado a menor rango dinámico del contenido de audio con alto promedio de niveles de audio. Sin embargo, el comportamiento del consumidor ha cambiado en años recientes con contenido codificado (p. ej. contenido en formatos de datos reducido tales como mp3) convirtiéndose más popular e importante para la distribución y el almacenamiento de contenido. Tales formatos permiten rango dinámico virtualmente ilimitado de cuyo contenido se aprovechan los propietarios y los entusiastas de audio. Adicionalmente, la creciente popularidad de los teléfonos móviles, teléfonos inteligentes y otros dispositivos electrónicos portátiles como reproductores multimedia personales ha creado nuevos desafíos para diseñar dispositivos de reproducción de alta calidad que cumplan las expectativas de cliente de nivelación consistente de audio y mejor calidad de audio en diversas condiciones de escucha. El gran número de contenido en colecciones de música personal (a menudo exceden miles de archivos) así como la gran gama de formatos de audio tales como mp3, HE-AAC, OGG, WMA, y Dolby Digital complican aún más el problema de proporcionar dispositivos de reproducción de audio con nivelación de audio consistente.

El informe europeo de búsqueda cita los siguientes documentos:

- 45 - ROBINSON D J M: "Perceptual model for assessment of coed audio", TESIS PRESENTADA PARA EL TÍTULO DE DOCTOR DE FILOSOFÍA, UNIVERSIDAD DE ESSEX, páginas 278-307 (en adelante en esta memoria: "D1");
- "Recommendation ITU-R BS.1770-1 Algorithm a measure audio programme loudness and true-peak audio level", ITU, páginas 1-19 (en adelante en esta memoria: "D2"); y
- 50 - US 2010/027812 A1 (en adelante en esta memoria "D3").

55 D1 describe un modelo auditivo binaural en dominio de tiempo para la evaluación de señales de audio codificadas de alta calidad. El modelo se diseña para que coincida con el procesamiento dentro del sistema auditivo humano tan cercanamente como sea posible. Las transformaciones de oído exterior y medio se simulan mediante filtros lineales; la respuesta de membrana basilar se simula mediante un banco de filtros de parasitación gamma dependiente de amplitud, y la acción compresiva de las celdas de pelo interior se simula mediante un circuito de adaptación. Una aplicación del modelo es como medidor perceptual de sonoridad.

D2 es una recomendación para algoritmos de medición de audio con el propósito de determinar sonoridad subjetiva de programa, y nivel de señal de pico verdadero. La primera fase del algoritmo aplica un prefiltrado que tiene en cuenta los efectos acústicos de la cabeza. La segunda fase aplica la curva de ponderación RLB, que consiste en un simple filtro paso alto. Con el prefiltro y el filtrado RLB aplicados, se mide entonces la media cuadrática de energía en el intervalo de medición T. Una vez se ha computado el nivel ponderado de media cuadrática para cada canal, la etapa final es sumar los canales.

D3 describe un aparato y un método para controlar un volumen de una señal de audio, por el que un volumen de una señal de entrada puede ser controlado usando una curva de ganancia no lineal y un objetivo de volumen de la señal de entrada. Se almacena una curva de ganancia no lineal que indica relaciones entre volúmenes de una señal de entrada y una señal de salida. Se recibe una señal de entrada e información de objetivo de volumen. Se determina una primera ganancia usando el volumen de la señal de entrada y el objetivo de información de volumen. Se determina una segunda ganancia a partir de uno de a) el volumen de la señal de entrada y el objetivo de información de volumen o b) la primera ganancia, usando la curva de ganancia no lineal. El volumen de señal de entrada se ajusta aplicando la segunda ganancia a la señal de entrada.

15 **Compendio de la invención**

El presente documento aborda el problema de proporcionar nivelación de audio consistente usando valores normalizados de sonoridad en dispositivos electrónicos, p. ej., reproductores multimedia. Se pone énfasis particular en los desafíos de tales dispositivos, es decir, en particular los diversos ambientes de escucha que influyen en el nivel de salida deseado así como la máxima tolerancia de rango dinámico y el amplio rango de diferentes fuentes y formatos del contenido de audio que hace difícil reproducir todo el contenido en el mismo nivel de salida deseado. Es más, se pueden tener en cuenta aspectos tales como la complejidad computacional y la aplicabilidad independiente de códec de valores normalizados de sonoridad en reproductores multimedia portátiles. Además, la solución no debe ser destructiva con relación al contenido presentado, es decir, la carga útil o señal PCM reales antes de la fase de descodificación deben permanecer sin cambiar. Un posible planteamiento para abordar el último asunto es el uso de metadatos que se almacenan y/o son enviados junto con el archivo multimedia o flujo de bits.

La invención es definida por la reivindicación independiente. Realizaciones adicionales son definidas por las reivindicaciones dependientes.

Según un aspecto de la invención, se describe un método para usar metadatos generados por codificador en operaciones de decodificación de audio. El método comprende recibir, por un descodificador de audio, una señal de audio de entrada codificada para una configuración de canal de audio. La señal de audio de entrada incluye muestras de audio. El método comprende además recibir los metadatos generados por codificador. Los metadatos generados por codificador comprenden: un primer valor relacionado con sonoridad; una pluralidad de valores de ganancia según un perfil de control de rango dinámico que define las características operacionales de un compresor de rango dinámico; parámetros que identifican varios canales de audio y la configuración de canal de audio. El método comprende además convertir las muestras de audio en una señal de audio de salida con un objetivo de nivel de salida. La conversión comprende: realizar control de rango dinámico usando dicha pluralidad de valores de ganancia, resultando de ese modo muestras de audio controladas por rango dinámico; convertir dicho primer valor relacionado con sonoridad a un segundo valor relacionado con sonoridad basado en una relación reversible configurada para convertir primeros valores relacionados con sonoridad determinados según un primer procedimiento en segundos valores relacionados con sonoridad determinados según un segundo procedimiento; derivar un valor de ganancia a partir del objetivo de nivel de salida y el segundo valor relacionado con sonoridad; y aplicar dicho valor de ganancia a las muestras de audio controladas por rango dinámico.

Además, se describe un ejemplo de método para proporcionar datos relacionados con sonoridad a un reproductor multimedia. El reproductor multimedia puede ser, p. ej., un teléfono móvil, un teléfono inteligente, un reproductor mp3, un ordenador portátil o un reproductor multimedia personal. Datos relacionados con sonoridad pueden ser un valor de sonoridad o un valor relacionado con sonoridad que se asocian con una señal de audio. La señal de audio puede ser almacenada como muestras en un archivo multimedia, p. ej., un archivo mp3, un archivo WMA o un archivo multimedia de iTunes. Como alternativa o adicionalmente, la señal de audio se puede proporcionar como flujo de bits, p. ej., un flujo de bits AAC, HE-AAC, Dolby Pulse o Dolby Digital.

El método puede comprender la etapa de proporcionar un primer valor relacionado con sonoridad asociado con la señal de audio, en donde el primer valor relacionado con sonoridad ha sido determinado según un primer procedimiento o algoritmo. A un valor relacionado con sonoridad también se le puede hacer referencia como información de nivelación. Un procedimiento o un algoritmo para determinar un valor relacionado con sonoridad pueden ser un conjunto de manipulaciones de la señal de audio a fin de determinar un valor relacionado con sonoridad que representa la sonoridad perceptual, es decir, la energía percibida, de una señal de audio. Tal procedimiento o algoritmo puede ser el algoritmo ITU-R BS.1770 para medir el esquema de cálculo de sonoridad de programa de audio y/o de sonoridad de Replay Gain. En un ejemplo, el primer procedimiento puede ser el algoritmo ITU-R BS.1770 y el primer valor relacionado con sonoridad puede ser el valor de sonoridad BS.1770. Cabe señalar que también se pueden usar variantes del algoritmo ITU-R BS.1770, p. ej., una variante que no considera periodos de silencio de la señal de audio.

El método puede comprender la etapa de convertir el primer valor relacionado con sonoridad en un segundo valor relacionado con sonoridad usando un modelo que comprende una relación reversible, en donde el segundo valor relacionado con sonoridad se asocia con un segundo procedimiento para determinar valores relacionados con sonoridad. En un ejemplo, el segundo procedimiento puede ser el esquema de cálculo de sonoridad de Replay Gain. En tal caso, el segundo valor de sonoridad puede ser el valor de ajuste de Replay Gain, es decir, un valor de ganancia que amplifica o atenúa la señal de audio a un nivel de salida predeterminado. Típicamente, el segundo valor relacionado con sonoridad cuando es derivado por conversión del primer valor relacionado con sonoridad es una estimación de un valor relacionado con sonoridad que se determina para la señal de audio según el segundo procedimiento o algoritmo. Esto se debe al hecho de que el modelo que comprende una relación reversible puede ser una aproximación de la relación real entre los valores relacionados con sonoridad determinados por el primer y el segundo procedimiento.

La relación reversible puede ser una relación lineal entre el primer valor relacionado con sonoridad y el segundo valor relacionado con sonoridad medidos en el espacio logarítmico, p. ej., medidos en dB. Este tipo de relación reversible puede ser obtenida al determinar una pluralidad de primeros valores relacionados con sonoridad asociados con una correspondiente pluralidad de señales de audio, en donde la pluralidad de primeros valores relacionados con sonoridad se determina según el primer procedimiento. Es más, se puede determinar una pluralidad de segundos valores relacionados con sonoridad asociados con la correspondiente pluralidad de señales de audio, en donde la pluralidad de segundos valores relacionados con sonoridad se determina según el segundo procedimiento. Un modelo que comprende uno o más parámetros de modelo se puede encajar en la pluralidad de parejas de valores primeros y segundos relacionados con sonoridad. El modelo puede ser ajustado y los parámetros de modelo pueden ser determinados usando un esquema de minimización de error, p. ej., usando un criterio de error cuadrático medio mínimo, un criterio de error total de mínimos cuadrados o un criterio de error mínimo de desviación absoluta, resultando de ese modo la relación reversible. En un ejemplo se puede usar un criterio de error de mediana. A modo de ejemplo, los parámetros de modelo pueden ser determinados de manera que la mediana del error de estimación del segundo valor relacionado con sonoridad puede ser cero. La relación reversible también puede ser determinada usando regresión lineal.

El método puede comprender la etapa de almacenar el segundo valor relacionado con sonoridad en metadatos asociados con la señal de audio. Los metadatos pueden tener una sintaxis o un formato predeterminados. En un ejemplo, el formato predeterminado usa la sintaxis de Replay Gain. Como alternativa o adicionalmente, el formato predeterminado puede cumplir con metadatos estilo iTunes o etiquetas ID3v2. En otro ejemplo, el primer valor relacionado con sonoridad puede ser transmitido en un flujo de bits Dolby Pulse o HE-AAC como Elemento de Relleno, p. ej., como parámetro de "nivel de referencia de programa", según la MPEG norma ISO 14496-3.

El método puede comprender la etapa de proporcionar los metadatos al reproductor multimedia. Los metadatos se pueden proporcionar junto con la señal de audio. En un ejemplo, la señal de audio y los metadatos pueden ser almacenados en uno o más archivos. Los archivos pueden ser almacenados en un soporte de almacenamiento, p. ej., memoria de acceso aleatorio (RAM) o disco compacto. En un ejemplo, la señal de audio y los metadatos pueden ser transmitidos al reproductor multimedia, p. ej., dentro de un flujo de bits multimedia tal como HE-AAC.

Como se ha resumido anteriormente, el primer y/o el segundo procedimiento para determinar valores relacionados con sonoridad pueden ser el algoritmo ITU-R BS.1770 para medir el esquema de cálculo de sonoridad de programa de audio y/o de sonoridad Replay Gain. Típicamente, el primer y el segundo procedimiento son diferentes. El primer y/o el segundo procedimiento pueden comprender la etapa de procesar la señal de audio de acuerdo con percepción humana de sonoridad, p. ej., usando ponderación-A y/o ponderación-K. Es más, el procedimiento puede comprender la etapa de determinar la energía de una o más partes de la señal de audio procesada. Adicionalmente, el procedimiento puede comprender la etapa de desplazar la energía para determinar un valor relacionado con sonoridad. Tal desplazamiento puede ser realizado para normalizar el valor relacionado con sonoridad con la Escala Completa (FS, *Full Scale*). Como alternativa o adicionalmente, un valor de ganancia puede ser derivado al desplazar la energía con un objetivo de energía. En otras palabras, un valor de ganancia puede ser determinado como el desplazamiento entre la energía y un objetivo de energía. A este valor de ganancia también se le puede hacer referencia como valor relacionado con sonoridad.

En un ejemplo, la relación reversible puede ser un polinomio de primer o segundo orden. En particular, la relación reversible puede ser dada por $L_2 = A + BL_1$, en donde L_2 es el segundo valor relacionado con sonoridad en el espacio logarítmico (p. ej. dB), L_1 es el primer valor relacionado con sonoridad en el espacio logarítmico (p. ej. dB) y A y B son números reales con $-17 \leq A \leq -15$ y/o $-0,7 \leq B \leq -0,9$. A y B pueden tomar cualquiera de los números reales en los intervalos anteriores. En particular, $-16,1 \leq A \leq 15,9$ o $-16,01 \leq A \leq 15,99$ o $A = -16,00$. En particular:

$-0,80 \leq B \leq -0,82$ o $-0,811 \leq B \leq -0,813$ o $B = -0,812$.

Como alternativa, B puede estar restringido a $B = -1,0$. En este caso, la relación entre las mediciones de potencia para el primer y el segundo valor relacionado con sonoridad es modelada por un factor constante. Es más, A se puede establecer como

$-19 \leq A \leq -18$, o $-18,4 \leq A \leq -18,3$, o $-18,31 \leq A \leq -18,29$ o $A = -18,30$ o $A = -18,10$ o $A = -18,00$.

La relación reversible puede ser dada por $L_2 = A + BL_1 + CL_1^2$ en donde L_2 es el segundo valor relacionado con sonoridad medido en el espacio logarítmico (p. ej. en dB), L_1 es el primer valor relacionado con sonoridad medido en el espacio logarítmico (p. ej. en dB) y A, B y C son números reales.

5 Como alternativa o adicionalmente, la relación reversible puede ser segmentada en intervalos del primer y/o el segundo valor relacionado con sonoridad y se puede determinar una relación reversible separada para la pluralidad de intervalos. Los parámetros de la relación reversible pueden ser determinados usando Regresión Lineal Segmentada. A modo de ejemplo, la relación reversible global puede ser segmentada en una pluralidad de intervalos $[L_{1,t-1}, L_{1,t})$ con $t=1, \dots, N_T$, en donde N_T es el número total de intervalos o segmentos ($N_T > 0$). Para cada uno de los intervalos $[L_{1,t-1}, L_{1,t})$ se puede determinar una relación reversible diferente, p. ej., un polinomio de primer o segundo orden. Es más, puede ser beneficioso imponer que la relación reversible global sea continua en las fronteras de la pluralidad de intervalos.

15 La etapa de convertir el primer valor relacionado con sonoridad en un segundo valor relacionado con sonoridad puede comprender además la etapa de desplazar el primer valor relacionado con sonoridad en un valor predeterminado si la señal de audio es una señal mono. En un ejemplo, la relación reversible está relacionada con una señal de audio estéreo. En tales casos, la correspondiente relación reversible para una señal de audio mono puede ser obtenida al desplazar el primer valor relacionado con sonoridad por el valor predeterminado, antes de aplicar la relación reversible. El valor predeterminado puede ser 3 dBFS.

20 El método para proporcionar datos de sonoridad a un reproductor multimedia puede comprender además la etapa de extraer el segundo valor relacionado con sonoridad de los metadatos y/o la etapa de producir la señal de audio usando el segundo valor relacionado con sonoridad. La señal de audio puede ser producida en un objetivo de nivel de salida. En tales casos, la etapa de producir puede comprender la amplificación o la atenuación de la señal de audio de acuerdo con el segundo valor relacionado con sonoridad. En otras palabras, el segundo valor relacionado con sonoridad o un valor derivado del objetivo de nivel de salida y el segundo valor relacionado con sonoridad, p. ej., un valor correspondiente a la diferencia entre el objetivo de nivel de salida y el segundo valor relacionado con sonoridad, pueden ser aplicados a las muestras de la señal de audio.

25 El método puede comprender además la etapa de convertir el segundo valor relacionado con sonoridad en el primer valor relacionado con sonoridad usando la relación reversible. En particular, se puede usar la inversa de la relación reversible. En tales casos, el reproductor multimedia puede producir el archivo de audio usando el primer valor relacionado con sonoridad. Como se ha resumido antes, la etapa de producción puede comprender la etapa de aplicar el primer valor relacionado con sonoridad o un valor derivado del primer valor relacionado con sonoridad a muestras de la señal de audio. Adicionalmente, la etapa de producción puede comprender la etapa de controlar el rango dinámico de la señal de audio procesada de sonoridad. Es más, la etapa de producción puede comprender limitar la señal de audio procesada de sonoridad para evitar recorte.

30 Es más, se describe un ejemplo de sistema configurado para proporcionar datos relacionados con sonoridad a un reproductor multimedia. El sistema puede comprender un codificador de sonoridad configurado para proporcionar un primer valor relacionado con sonoridad asociado con una señal de audio, en donde el primer valor relacionado con sonoridad ha sido determinado según un primer procedimiento. El codificador de sonoridad se puede configurar además para convertir el primer valor relacionado con sonoridad en un segundo valor relacionado con sonoridad usando un modelo que comprende una relación reversible, en donde el segundo valor relacionado con sonoridad se asocia con un segundo procedimiento para determinar valores relacionados con sonoridad. Adicionalmente, el codificador de sonoridad se puede configurar para almacenar el segundo valor relacionado con sonoridad en metadatos asociados con la señal de audio.

35 El sistema puede comprender un medio de transmisión configurado para proporcionar los metadatos al reproductor multimedia. Tal medio de transmisión puede ser un medio de transmisión inalámbrico o cableado. Es más, el medio de transmisión puede ser un soporte de almacenamiento tal como RAM o un disco compacto.

40 El sistema puede comprender un reproductor multimedia configurado para extraer el segundo valor relacionado con sonoridad de los metadatos; y para producir la señal de audio basándose en el segundo valor relacionado con sonoridad. Adicionalmente, el reproductor multimedia se puede configurar para convertir el segundo valor relacionado con sonoridad en el primer valor relacionado con sonoridad usando la relación reversible; y para producir la señal de audio usando el primer valor relacionado con sonoridad.

45 Es más, se describe un ejemplo de codificador de sonoridad. El codificador de sonoridad se puede configurar para proporcionar un primer valor relacionado con sonoridad asociado con una señal de audio; en donde el primer valor relacionado con sonoridad ha sido determinado según un primer procedimiento; para convertir el primer valor relacionado con sonoridad en un segundo valor relacionado con sonoridad usando un modelo que comprende una relación reversible; en donde el segundo valor relacionado con sonoridad se asocia con un segundo procedimiento para determinar valores relacionados con sonoridad; y/o para almacenar el segundo valor relacionado con sonoridad en metadatos asociados con la señal de audio.

Es más, se describe un ejemplo de reproductor multimedia. El reproductor multimedia se puede configurar para extraer

un segundo valor relacionado con sonoridad de metadatos asociados con una señal de audio; en donde el segundo valor relacionado con sonoridad se asocia con un segundo procedimiento para determinar valores relacionados con sonoridad. Es más, el reproductor multimedia se puede configurar para convertir el segundo valor relacionado con sonoridad en un primer valor relacionado con sonoridad usando un modelo que comprende una relación reversible; en donde el primer valor relacionado con sonoridad ha sido determinado según un primer procedimiento. Adicionalmente, el reproductor multimedia se puede configurar para producir la señal de audio usando el primer valor relacionado con sonoridad.

Es más, se describe un ejemplo de método para producir una señal de audio en un reproductor multimedia usando un primer valor relacionado con sonoridad. El método puede comprender la etapa de recibir metadatos asociados con la señal de audio en el reproductor multimedia. El método puede proceder a extraer un segundo valor relacionado con sonoridad de los metadatos asociados con la señal de audio; en donde el segundo valor relacionado con sonoridad se asocia con un segundo procedimiento para determinar valores relacionados con sonoridad. El segundo valor relacionado con sonoridad puede ser convertido en un primer valor relacionado con sonoridad usando un modelo que comprende una relación reversible; en donde el primer valor relacionado con sonoridad ha sido determinado según un primer procedimiento. Finalmente, el método puede comprender la etapa de producción la señal de audio usando el primer valor relacionado con sonoridad. En un ejemplo, el segundo valor relacionado con sonoridad se deriva en un correspondiente codificador de sonoridad por conversión del primer valor relacionado con sonoridad usando la relación reversible. En particular, el segundo valor relacionado con sonoridad puede ser determinado a partir de un valor de sonoridad ITU-R BS.1770 usando la relación reversible. En otro ejemplo, el segundo valor relacionado con sonoridad puede corresponder al valor relacionado con sonoridad determinado directamente usando el segundo procedimiento. En particular, el segundo valor relacionado con sonoridad puede corresponder a un valor de ajuste de Replay Gain computado por un dispositivo legado.

Es más, se describe un ejemplo de método para determinar una relación reversible para convertir un primer valor relacionado con sonoridad en un segundo valor relacionado con sonoridad. El método puede comprender la etapa de determinar una pluralidad de primeros valores relacionados con sonoridad asociados con una correspondiente pluralidad de señales de audio, en donde la pluralidad de primeros valores relacionados con sonoridad se determina según un primer procedimiento. El método puede comprender además la etapa de determinar una pluralidad de segundos valores relacionados con sonoridad asociados con la correspondiente pluralidad de señales de audio, en donde la pluralidad de segundos valores relacionados con sonoridad se determina según un segundo procedimiento. Adicionalmente, el método puede comprender la etapa de proporcionar un modelo para la relación reversible que comprende un parámetro de modelo. Adicionalmente, el método puede comprender la etapa de determinar el parámetro de modelo al ajustar el modelo a la pluralidad de valores primeros y segundos relacionados con sonoridad usando un criterio apropiado de minimización de error tal como mínimo error de cuadrados medios, resultando de ese modo la relación reversible.

Es más, se describe un ejemplo de programa de software adaptado para ejecución en un procesador. El programa de software se puede adaptar además para realizar un método según cualquiera de los aspectos resumidos en el presente documento cuando se lleva a cabo en un dispositivo informático.

Es más, se describe un ejemplo de soporte de almacenamiento que comprende un programa de software adaptado para ejecución en un procesador. El programa de software se puede adaptar además para realizar un método según cualquiera de los aspectos resumidos en el presente documento cuando se lleva a cabo en un dispositivo informático.

Según un aspecto adicional de la invención, se describe un producto de programa informático. El producto de programa informático comprende instrucciones ejecutables para realizar un método según la invención como se resume en el presente documento cuando se lleva a cabo en un dispositivo informático.

Breve descripción de los dibujos

La invención y ejemplos adicionales se explican a continuación de manera ejemplar con referencia a los dibujos adjuntos, en donde

la figura 1a ilustra un ejemplo configuración de medición para sonoridad BS.1770;

la figura 1b ilustra ejemplos de curvas de ponderación de frecuencia aplicados para Replay Gain y ITU-R BS.1770;

la figura 2a ilustra la distribución de géneros de un ejemplo de base de datos de archivos de música;

la figura 2b ilustra la sonoridad BS.1770 media y la desviación típica para los diferentes géneros del ejemplo de base de datos de la figura 2a;

la figura 2c ilustra la distribución de información anual proporcionada para el ejemplo de base de datos de la figura 2a;

la figura 2d ilustra el promedio de sonoridad BS.1770 y la desviación típica frente a información anual para el ejemplo de base de datos de la figura 2a;

la figura 2e ilustra los valores de ajuste de Replay Gain vs. valores de sonoridad BS.1770 para el ejemplo de base de datos de la figura 2a;

la figura 2f muestra un ejemplo de histograma para el error de transcodificación entre sonoridad BS.1770 y sonoridad Replay Gain para el ejemplo de base de datos de la figura 2a;

5 la figura 2g muestra valores modificados de ajuste de Replay Gain frente a valores de sonoridad BS.1770 para el ejemplo de base de datos de la figura 2a;

la figura 3 es un diagrama de bloques esquemático de un ejemplo de dispositivo de descodificación dentro de un reproductor multimedia;

10 la figura 4 es un diagrama de bloques esquemático de un ejemplo de dispositivo de codificación en un lugar de producción o transmisión de audio; y

la figura 5 ilustra un diagrama de bloques de un ejemplo de sistema de codificación y descodificación.

Descripción detallada

15 Un planteamiento para proporcionar salida de audio en un nivel percibido constante es definir un objetivo de nivel de salida en el que se va a producir el contenido de audio. Tal objetivo de nivel de salida puede ser, p. ej., -11 dBFS (decibelios relativos a Escala Completa). En particular, el objetivo de nivel de salida puede depender del ambiente de escucha actual. Es más, se puede determinar el nivel de sonoridad real del contenido de audio, también se le hace referencia como nivel de referencia. El nivel de sonoridad se proporciona preferiblemente junto con el contenido multimedia, p. ej., como metadatos proporcionados conjuntamente con el contenido multimedia. A fin de producir el contenido de audio en el objetivo de nivel de salida durante la reproducción se puede aplicar un valor de ganancia coincidente. El valor de ganancia coincidente se puede determinar como la diferencia entre el objetivo de nivel de salida y el nivel de sonoridad real.

20 Es más, se debe tener en cuenta posible recorte en casos en los que el contenido se tiene que realzar para que coincida con el objetivo de nivel de salida, es decir, en caso de que se tenga que aplicar una ganancia coincidente positiva. Es más, si el nivel de sonoridad real o el nivel de referencia se proporcionan en metadatos que se asocian con el contenido multimedia, se tiene que abordar el manejo de contenido multimedia legado, es decir, el manejo de contenido multimedia que no comprende los metadatos requeridos.

25 Como ya se ha indicado anteriormente, sistemas para retrasmisión y difusión, como, p. ej. Dolby Digital, típicamente dependen de transmitir metadatos que comprenden valor "dialnorm" que indica el nivel de sonoridad del programa actual al dispositivo de descodificación. El valor "dialnorm" puede ser diferente para programas diferentes. En vista del hecho de que el valor o valores "dialnorm" se determinan en el codificador, se permite al propietario de contenido controlar la cadena completa de señal hasta el descodificador real. Es más, se puede reducir la complejidad computacional en el dispositivo de descodificación, ya que no se requiere determinar valores de sonoridad para el programa actual en el descodificador. En cambio los valores de sonoridad se proporcionan en los metadatos asociados con el programa actual.

30 Para sistemas basados en archivos, es decir, para sistemas que están relacionados con archivos multimedia en lugar de flujos de bits multimedia, un valor de sonoridad o un nivel de referencia típicamente no cambian para un archivo dado. En otras palabras, se determinan valores de sonoridad por archivo. En un ejemplo, los niveles de sonoridad se codifican en la carga útil, es decir, se modifica el contenido multimedia real, a fin de encontrar ciertos objetivos de niveles de salida. Por otro lado, se han diseñado sistemas de archivos multimedia que dependen de un único valor de sonoridad por archivo, en donde el valor de sonoridad puede ser almacenado como metadatos adicionales asociados con el contenido multimedia. Un ejemplo para este tipo de sistema de archivos multimedia se basa en la llamada iniciativa "Replay Gain" y el correspondiente formato "Replay Gain" que se especifica en <http://www.replaygain.org>. El formato Replay Gain se ha implementado en diversos reproductores multimedia, que de ese modo se les permite realizar ajustes de sonoridad a archivos multimedia que comprenden metadatos correspondientes al formato Replay Gain.

35 El formato Replay Gain define una sintaxis Replay Gain predeterminada para especificar una pluralidad de parámetros asociados con un archivo multimedia particular. Posibles parámetros pueden ser un valor de ajuste de Replay Gain por pista multimedia o un valor de ajuste de Replay Gain por álbum de pistas multimedia. El primer parámetro especifica un valor de sonoridad para un archivo multimedia, mientras que el segundo parámetro especifica un valor de sonoridad común para un álbum, es decir, para un conjunto de pistas multimedia. Parámetros adicionales pueden ser la amplitud de señal de Pico Replay Gain por pista y/o la amplitud de señal de Pico Replay Gain por álbum. Estos parámetros especifican la amplitud absoluta máxima de una pista o un álbum, respectivamente.

40 La iniciativa Replay Gain también especifica un procedimiento para determinar los parámetros anteriores para una pista o álbum particulares, es decir, se especifican las semánticas para los parámetros. Los valores de ajuste de Replay Gain, es decir, los valores relacionados con sonoridad, se determinan usando un procedimiento resumido en <http://www.replaygain.org>. Los valores de ajuste de Replay Gain se supone que ajustan la sonoridad musical a la

sonoridad de ruido rosa a 20 dBFS reproducido en altavoces estéreo. El nivel de presión sonora asociada, es decir, el objetivo predeterminado de nivel de salida, es 83 dB SPL. La amplitud de señal de Pico Replay Gain se determina determinando la amplitud absoluta máxima del archivo multimedia. Cabe señalar, sin embargo, que la motivación perceptual de estos procedimientos, en particular el procedimiento relacionado con la determinación de un valor de sonoridad, puede ser cuestionado y es preferible aplicar procedimientos que han sido aceptados generalmente por ser perceptivamente significativos.

Como consecuencia, existe el problema de definir parámetros apropiados perceptivamente relevantes para contenido de audio. En particular, se tiene que abordar el problema de determinar un nivel de sonoridad de un archivo multimedia o un grupo de archivos multimedia que satisfaga igualmente la percepción de diferentes usuarios. Como materia de hecho, la sonoridad es una cantidad sumamente subjetiva y a menudo no hay un único nivel de sonoridad que satisfaga a todos a los oyentes (o incluso un único oyente) todo el tiempo. Un estudio realizado por Dolby Laboratories concluyó que incluso cuando una programación de audio ha sido "normalizada" por un grupo de personas "a oído", los programas "normalizados" no satisfacen completamente a un grupo diferente de oyentes un 100 por ciento del tiempo. No obstante, un método para medir la sonoridad de programa de audio, así como el nivel de audio pico verdadero, ha sido especificado como recomendación ITU-R BS.1770. El método especificado para medir la sonoridad de audio ha sido ampliamente aceptado y se puede usar como medición de sonoridad común para diferentes fuentes y formatos de contenido multimedia. Como tal, se puede asegurar que diferentes herramientas de codificación se comportan coherentemente. Esto aumentará la satisfacción del consumidor puesto que el contenido de fuentes diferentes se comportará de manera similar. Cabe señalar que ITU-R BS.1770 define un objetivo de sonoridad u objetivo de nivel de salida predeterminados y está abierto a que el usuario derive una ganancia apropiada basándose en la medida de sonoridad. El algoritmo BS.1770 ha sido verificado para predecir la sonoridad correctamente para diferente contenido (discurso, música y película) y para mono, estéreo y multicanal 5.0/5.1 igualmente. Adicionalmente, la complejidad computacional es baja comparada con modelos sicoacústicos más sofisticados (p. ej. el modelo Zwicker) que emplean análisis de banco de filtro y enmascaramiento de modelos.

En compendio, la llamada iniciativa Replay Gain especifica un formato Replay Gain para asociar datos de sonoridad con un archivo multimedia particular. Este formato ha sido ampliamente aceptado por diversos reproductores multimedia. La iniciativa Replay Gain también especifica un algoritmo o procedimiento para computar un valor de ajuste de Replay Gain para normalizar sonoridad por pistas y/o álbumes. En vista del extenso saber hacer sicoacústico y psicológico que se ha tenido en cuenta en la recomendación BS.1770, se espera, sin embargo, que se pueda lograr un mayor grado de satisfacción perceptual cuando se usa la medición de sonoridad especificada en la recomendación BS.1770.

Dadas las consideraciones anteriores, se ha propuesto implementar normalización de sonoridad en reproductores multimedia basándose en la medición de sonoridad como se define en la recomendación ITU-R BS.1770, y al trasladar tales valores de sonoridad como valores de ajuste de Replay Gain equivalentes o correspondientes como metadatos con archivos multimedia. En otras palabras, se ha propuesto medir la sonoridad basándose en ITU-R BS.1770. Adicionalmente, a fin de soportar metadatos Replay Gain existentes y mantener nivelación consistente con ambos tipos de información de nivelación, se describe una conversión entre sonoridad Replay Gain y ITU-R BS.1770. Para lograr esta meta se investigan ambos planteamientos de nivelación, es decir, Replay Gain y ITU-R BS. 1770, por medio de una base de datos de música estadísticamente relevante.

A continuación se resumen ciertos aspectos en relación con la medición de sonoridad según la recomendación ITU-R BS.1770. La figura 1a muestra un diagrama de bloques de los diversos componentes del algoritmo de medición de sonoridad 100. El diagrama de bloques muestra entradas para cinco canales principales (izquierdo, centro, derecho, envolvente izquierdo y envolvente derecho); esto permite la monitorización de archivos multimedia que comprende de uno a cinco canales. Para un archivo multimedia que tiene menos de cinco canales no se usarían algunas entradas. El canal de efectos de baja frecuencia (LFE, *low frequency effects*) típicamente no se incluye en la medición de sonoridad.

La primera fase del algoritmo aplica un prefiltrado 101 a la señal como se muestra en la figura 2 de ITU-R BS.1770-1. El prefiltrado considera los efectos acústicos de la cabeza del oyente, donde la cabeza se modela como esfera rígida. El prefiltro 101 es definido por el filtro mostrado en la figura 3 de ITU-R BS.1770-1 con los coeficientes especificados en la Tabla 1 de ITU-R BS.1770-1. La segunda fase 102 del algoritmo aplica la curva de ponderación RLB, que consiste en un filtro paso alto como se muestra en la figura 4 de ITU-R BS.1770-1, en donde la curva de ponderación RLB se especifica como filtro de segundo orden como se muestra en la figura 3 de ITU-R BS.1770-1, con los coeficientes especificados en la Tabla 2 de ITU-R BS.1770-1.

Posteriormente, la energía media cuadrática de cada canal en un intervalo de medición T se mide usando la Ecuación (1) de ITU-R BS.1770-1 en la unidad 103. Finalmente, se ponderan los valores de energía (signo de referencia 104) y los valores de energía ponderados para los diferentes canales se suman (signo de referencia 105) para dar como resultado el valor de sonoridad del archivo multimedia en el respectivo intervalo de medición T. A la ponderación 104 se le hace referencia como ponderación K. Es más, la determinación de sonoridad puede comprender un desplazamiento de la suma de los valores de energía ponderados. Por lo tanto, al valor de sonoridad determinado según ITU-R BS.1770-1 se le puede hacer referencia como Sonoridad, ponderada K, relativa a la Escala Completa normal (LKFS).

En compendio, ITU-R BS.1770 recomienda las siguientes etapas para determinar la sonoridad BS.1770: (1) Filtrar todos los canales con la curva B de Baja Frecuencia Revisada (RLB) y en serie con un filtro de aproximación de cabeza esférica (unidades 101 y 102); (2) Computar potencia a largo plazo para cada canal relativo a escala completa (unidad 103); (3) Si está presente, aumentar las potencias de canales envolventes por un factor 1,41 (1,5 dB) (unidad 104); (4) Añadir todas las potencias de canal, convertir en dB y añadir un nivel de calibración de 0,691 dB (unidad 105). El resultado es la sonoridad dada en LKFS. El nivel de calibración asegura que a una onda sinusoidal mono de escala completa se le asigna una sonoridad de -3,01 LKFS.

El algoritmo para la determinación de valores de ganancia usado en la iniciativa Replay Gain difiere de la medición de sonoridad según ITU-R BS.1770-1. Los valores de ganancia se determinan usando un preprocesamiento con un filtro de sonoridad igual promedio (o con más precisión una aproximación de filtro IIR (filtro de Respuesta de Impulso Infinito) a la curva de sonoridad promedio inverso como se resume en <http://www.replaygain.Org/>) Posteriormente se calculan valores de energía RMS (raíz cuadrática media) durante intervalos de tiempo de 50 ms de la respectiva señal de audio. Las diferentes energías RMS se clasifican en orden numérico y únicamente el 5 % de los intervalos de tiempo que tienen las energías RMS más altas se consideran para estimar la sonoridad percibida global de la señal de audio. Finalmente, el valor de ajuste de Replay Gain del archivo de audio puede ser determinado al calibrar el valor de ajuste de Replay Gain de manera que la señal de audio que es producida con el valor de ajuste de Replay Gain se produce en un promedio de nivel de repetición de 83 dB SPL (nivel de presión sonora) según una señal de referencia de ruido rosa en -20 dBFS reproducida en dos altavoces. Esta calibración se realiza de acuerdo con la calibración definida en la recomendación SMPTE.

En compendio, para la determinación de los valores de ajuste de Replay Gain se proponen las siguientes etapas: (1) Filtrar todos los canales con un promedio de filtro de sonoridad igual; (2) Computar potencias relativas a escala completa para bloques no solapados de longitudes 50 ms y promedio en canales; (3) Computar la potencia de bloque que es superada en un 5 % de todos los bloques por pista y derivar la sonoridad al convertir en dB. Finalmente, los valores de ajuste de Replay Gain se computan como la diferencia entre el objetivo de sonoridad y la sonoridad medida donde el objetivo de sonoridad se basa en la señal de referencia de ruido rosa [-25,5 dB - sonoridad]. Como tal, un valor de ganancia se obtiene como el valor de ajuste de Replay Gain.

Ambos planteamientos de nivelación, es decir, ITU-R BS.1770 y Replay Gain, miden una potencia ponderada en frecuencia. Las principales diferencias entre ambos planteamientos de nivelación son las características de filtro y el análisis de potencia estadística del que se deriva la sonoridad. Si bien ITU-R BS.1770 aplica una curva de ponderación en frecuencia 110 que tiene una característica de filtro paso alto, la curva de ponderación en frecuencia de Replay Gain 120 tiene una característica de filtro pasobanda como se muestra en la figura 1b.

En ITU-R BS.1770 se promedia la energía en la pista de música completa que incluye potencialmente silencio que típicamente no contribuye a la sonoridad subjetiva. En las investigaciones realizadas por los inventores, los periodos de silencio se excluyen preferiblemente de la medición. Como tal, se puede usar un algoritmo modificado ITU-R BS.1770. Para esta finalidad, se pueden detectar periodos de silencio usando uno o más criterios. Según un primer criterio, el nivel pico (absoluto) de una señal de audio no supera un umbral pico (absoluto) durante un periodo de silencio de la señal de audio. En un ejemplo, este tipo de umbral pico puede ser de -60 dBFS o -70 dBFS. Según un criterio adicional, la energía de la señal de audio puede no superar un umbral de energía durante un periodo de silencio de la señal de audio. En un ejemplo, este tipo de umbral de energía puede ser -70 LKFS. Un criterio adicional puede ser una duración mínima y/o máxima de un periodo de silencio. Como tal, se puede detectar un periodo de silencio si se cumple el primer y/o el segundo criterio para una duración mínima y/o máxima de la señal de audio. Duraciones mínimas típicas pueden ser de 200 ms o 1 s, en donde duraciones máximas típicas pueden ser de 400 ms o 10 s.

Como tal, el primer o el segundo procedimiento para determinar un valor relacionado con sonoridad puede ser un algoritmo modificado ITU-R BS.1770 que excluye periodos de silencio de la señal de audio para la determinación del valor relacionado con sonoridad. Los periodos de silencio de la señal de audio pueden ser determinados como se ha resumido antes. A la exclusión de periodos de silencio también se le puede hacer referencia como *gating* de silencios, en donde al intervalo de tiempo dado por la duración mínima y la duración máxima de los periodos de silencio se le puede hacer referencia como *gating* de silencios. Como tal, al algoritmo modificado ITU-R BS.1770 también se le puede hacer referencia como el algoritmo ITU-R BS.1770 usando *gating* de silencios. En términos más generales, se puede indicar que un procedimiento para determinar un valor relacionado con sonoridad puede o no considerar *gating* de silencios.

Sin embargo puesto que los datos de música usualmente no exhiben una cantidad significativa de silencio el impacto del silencio es limitado. Replay Gain, por otro lado, mide la potencia de trama que es superada únicamente por un 5 % de todas las potencias de trama que están cerca de la potencia de trama máxima absoluta, es decir, Replay Gain determina una potencia casi máxima para un archivo de música particular.

En vista del hecho de que los valores de ajuste de Replay Gain y los valores de ITU-R BS.1770 de sonoridad se determinan usando diferentes procedimientos o algoritmos, la relación entre ambos valores para un archivo multimedia particular es sumamente compleja. No obstante, basándose en el análisis de los inventores resumido en este documento, se puede determinar una relación significativa entre ambos procedimientos y sus valores resultantes relacionados con sonoridad. Como se mostrará, la medición de potencia casi máxima aplicada en Replay Gain y la

potencia a largo plazo aplicada en ITU-R BS.1770 tienen un impacto significativo en la relación entre los valores de ajuste de Replay Gain y los valores de sonoridad BS.1770. Es más, se mostrará que la relación entre los resultados de los diferentes esquemas de medición de potencia depende del rango dinámico del archivo de música particular.

5 Como se indica anteriormente, se sugiere trasladar valores de sonoridad según ITU-R BS.1770-1 como valores equivalentes o correspondientes de ajuste de Replay Gain. En este contexto, se sugiere hacer coincidir las semánticas
 10 Replay Gain con resultados de sonoridad BS.1770 basándose en una relación derivada estadísticamente, p. ej., una ecuación lineal. Esto se puede lograr determinando valores de sonoridad BS.1770 y valores de ajuste de Replay Gain para un número estadísticamente relevante de archivos de sonido o audio (p. ej. en formato mp3 y m4a).
 15 Posteriormente, se puede determinar una relación determinística y preferiblemente invertible entre valores de sonoridad BS.1770 y valores de ajuste de Replay Gain. Este tipo de relación puede ser un polinomio, p. ej., un polinomio de primer o segundo orden, que se hace coincidir con la pluralidad de parejas de valores de sonoridad BS.1770 y valores de ajuste de Replay Gain usando un criterio apropiado de minimización de error. En un ejemplo, la relación puede ser una ecuación lineal en el espacio logarítmico que se determina usando un criterio de mínimo error cuadrático medio u otros criterios de error tales como un criterio de error total de mínimos cuadrados o un criterio de error de desviación absoluta mínima. Para determinar la relación se pueden usar técnicas de regresión lineal.

20 En un ejemplo, una base de datos de música estadísticamente relevante consiste en 21 220 archivos estéreo originados en diferentes colecciones privadas de música. Formatos de compresión son mp3 y AAC en diversas tasas de bits y tasas de muestreo entre 32 y 48 kHz. Los valores de ajuste de Replay Gain se calculan para todos los archivos. La sonoridad según ITU-R BS.1770 se computa sin considerar periodos de silencio, en donde se identifica
 25 silencio cuando el nivel pico máximo relativo a escala completa permanece por debajo de -60 dBFS durante más de un segundo. En promedio únicamente el 0,6 % de la duración de pista de audio se identificó como silencio, lo que indica que el efecto del silencio es relativamente pequeño. Archivos que tienen bajos valores de sonoridad parecen tener un grado más alto de periodos de silencio que archivos más sonoros (0,3 % de periodos de silencio a -5 LKFS a 1,2 % de periodos de silencio a -30 LKFS). La figura 2a muestra la distribución de géneros musicales en el ejemplo de base de datos según metadatos ID3.

30 Basándose en el ejemplo anterior de base de datos de música, se ha encontrado que en hay moderada variación general de sonoridad en la amplitud de 5-10 LKFS entre géneros. Sin embargo la música Clásica y contenido hablado tienen sonoridad especialmente baja comparada con otros géneros, como se puede ver en la figura 2b. Según metadatos ID3 aproximadamente la mitad de todos los archivos de música no son anteriores a 2001 como se puede ver en la figura 2c. Como ya se ha indicado anteriormente, existe la tendencia de creciente sonoridad desde principios de los 90 hasta hoy. Esto se puede ver en la figura 2d. Si restringir el análisis a archivos de música muy recientes de 2009, la sonoridad media es de -8,5 LKFS.

35 La figura 2e muestra una gráfica donde se han trazado las parejas de valor de ajuste de Replay Gain y valor de sonoridad BS.1770 para la base de datos de música mencionada anteriormente. La línea 210 se ha obtenido usando una línea recta de mínimos cuadrados (primer orden polinomio) ajustada a la pluralidad de parejas de datos. La línea tiene una pendiente de -0,81 en el espacio logarítmico (dB). Una pendiente que es diferente de -1,0 en el espacio logarítmico indica una relación no lineal entre ambas medidas subyacentes de potencia. La línea discontinua 220 representa la solución de mínimos cuadrados cuando la pendiente se restringe a -1,0 en el espacio logarítmico de modo que la relación entre las diferentes medidas de potencia se modela linealmente por un factor constante.

40 Se puede aplicar regresión lineal para determinar una relación que se puede usar para convertir un valor de ajuste de Replay Gain a un valor de sonoridad ITU-R BS.1770 y viceversa. La línea discontinua 220 representa el ajuste a línea recta donde la pendiente se restringe para ser -1,0 en el espacio de dB con un desplazamiento entre sí de -18,3 dB (numeral de referencia 221). Sin esta restricción la pendiente óptima en el espacio dB es -0,81 y el desplazamiento óptimo entre sí -16,0 dB como se representa con la línea continua 210 (numeral de referencia 211).

45 En otras palabras, se sugiere determinar una estimación del valor de ajuste de Replay Gain usando la fórmula:

$$\text{Replay Gain}' = -16,00 - 0.812 \cdot \text{BS1770}, \quad (1)$$

50 en donde $\text{Replay Gain}'$ es la estimación del valor de ajuste de Replay Gain medido en dB, y BS1770 es el valor de sonoridad BS.1770 medido en dB (o LKFS), para el mismo archivo multimedia. La fórmula (1) ha sido derivada aplicando un criterio de error cuadrático medio y una relación lineal en el espacio dB en una pluralidad estadísticamente relevante de parejas de valores de sonoridad BS.1770 y valores de ajuste de Replay Gain, en donde cada pareja de valores se determina del mismo archivo multimedia.

Como alternativa, la pendiente puede ser restringida a -1,0 en el espacio dB. En este caso, se puede determinar una estimación para el valor de ajuste de Replay Gain usando la fórmula:

$$\text{Replay Gain}' = -18,3 - 1,0 \cdot \text{BS1770}, \quad (2).$$

55 Cabe señalar que en un ejemplo alternativo, la pendiente puede ser restringida a -1,0 en el espacio dB y el desplazamiento entre sí del polinomio de primer orden puede ser encajado en el espacio dB con la pendiente predefinida de -1,0 de modo que la mediana del error de estimación de Replay Gain es 0,0 dB para la base de datos

subyacente. Al usar estos criterios de error, se puede determinar una relación reversible de manera que el número de puntos de datos medidos con error de estimación positivo es igual al número de puntos de datos con error de estimación negativo. Al usar este criterio de error basado en mediana, se puede obtener un desplazamiento entre sí de -18,1 para la base de datos investigada, es decir,

5
$$\text{Replay Gain}' = -18,1 - 1,0 \cdot \text{BS1770}.$$

Cabe señalar que en algunos casos, puede ser ventajoso modificar la fórmula anterior para usar un desplazamiento entre sí de -18,0 en lugar de -18,1, es decir,

$$\text{Replay Gain}' = -18,0 - 1,0 \cdot \text{BS1770}.$$

10 El error de estimación de la relación entre valores de sonoridad BS.1770 y valores de ajuste de Replay Gain se define como Error = Replay Gain' - Replay Gain (dB). Características estadísticas del valor absoluto de este Error para las mediciones mencionadas anteriormente se proporcionan en la Tabla 1.

	Fórmula (1)	Fórmula (2)
RMS	1,04	1,32
Media	0,78	1,04
Máx. 95 %	2,11	2,66
Máx. 75 %	1,08	1,45

Tabla 1

15 La figura 2f muestra un histograma de error correspondiente para el error correspondiente a la diferencia entre el valor de ajuste de Replay Gain estimado y el real. El histograma 230 corresponde a las estimaciones obtenidas de la fórmula (1) y el histograma 240 corresponde a las estimaciones obtenidas de la fórmula (2).

20 Cabe señalar que las fórmulas anteriores (1) y (2) se han derivado de contenido de audio estéreo. Sin embargo, se ha observado que para contenido mono valores de sonoridad de ITU-R BS.1770 y de ajuste de Replay Gain se comportan de manera diferente. Si bien la determinación de valores de ajuste de Replay Gain da como resultado el mismo resultado para una señal mono y una correspondiente señal mono dual, la determinación de sonoridad ITU-R BS.1770 de la señal mono dual es 3 dB más alta que la de la correspondiente señal mono.

Por lo tanto, las reglas de conversión anteriores se podrían adaptar para contenido general mono o estéreo a:

$$\text{Replay Gain}' = -16,00 - 0,812 \cdot (\text{BS1770} + x), \quad (1')$$

$$\text{Replay Gain}' = -18,3 - 1,0 \cdot (\text{BS1770} + x), \quad (2')$$

en donde $x = 3$ dBFS para contenido mono y $x = 0$ dBFS de otro modo.

25 La observación de que la pendiente óptima no es igual a -1,0 indica una dependencia de la relación entre el valor de ajuste de Replay Gain y el valor de sonoridad BS.1770 en el nivel de sonoridad real. Esto se puede explicar por los diferentes análisis de potencia estadística realizados por ambos procedimientos. La potencia de trama casi máxima determinada por Replay Gain no cambia de la misma manera que lo hace la potencia a largo plazo determinada por ITU-R BS.1770. Esto es particularmente verdadero para valores de sonoridad más altos cuando se está reduciendo el rango dinámico. Por lo tanto se puede lograr un ajuste mejorado por un polinomio de 2º orden. El polinomio de 2º orden puede tener una pendiente de -1,0 para sonoridad baja y una pendiente reducida (mayor de -1,0) para valores de sonoridad moderados y más altos (p. ej. >-25 LKFS). Sin embargo usar una pendiente de -1,0 con un desplazamiento de nivel óptimo (de -18,3 dB) ya puede ser suficiente para muchos archivos de música.

35 El hecho de que la relación no lineal entre las medidas de potencia obtenidas cuando se usa el procedimiento Replay Gain y cuando se usa el procedimiento ITU-R BS.1770 se debe principalmente a diferentes esquemas de análisis estadísticos de potencia usados en ambos procedimientos se puede ver en la figura 2g. Para esta finalidad, la medición de potencia de trama casi máxima en Replay Gain es sustituida por la medición de potencia a largo plazo usado en ITU-R BS.1770. Los valores de ajuste de Replay Gain modificados se trazan contra los correspondientes valores de sonoridad BS.1770 y en la figura 2g se puede ver que una línea 250 que tiene una pendiente de -1,0 encaja bien en la pluralidad de puntos de datos/parejas.

45 Usando la fórmula de conversión anterior (1) o (2), se pueden realizar mediciones de sonoridad según BS.1770 y posteriormente convertirse a valores equivalentes de ajuste de Replay Gain que pueden ser almacenados y/o transmitidos de acuerdo con el formato Replay Gain. Dependiendo de las capacidades del descodificador de audio en el dispositivo electrónico, la sonoridad puede ser normalizada de acuerdo con la especificación Replay Gain. En otras palabras, si el reproductor multimedia es compatible con Replay Gain, el archivo multimedia puede ser producido en el objetivo de nivel de salida de la iniciativa Replay Gain usando el valor de ajuste de Replay Gain.

Por otro lado, si el descodificador de audio se configura para realizar normalización de sonoridad de acuerdo con

BS.1770, los valores de sonoridad BS.1770 pueden ser recalculados a partir de los valores equivalentes de ajuste de Replay Gain usando la inversa de la fórmula (1) o (2). En otras palabras, el reproductor multimedia puede determinar los valores de sonoridad BS.1770 a partir de los valores de ajuste de Replay Gain y producir los archivos multimedia en el objetivo de nivel de salida al determinar la ganancia apropiada de acuerdo con el estándar BS.1770. En vista del hecho de que los valores de sonoridad BS.1770 se normalizan a Escala Completa (FS), la ganancia a aplicar para lograr un objetivo de nivel de salida medido en FS se puede determinar como la diferencia entre el objetivo de nivel de salida y el valor de sonoridad BS.1770.

Cabe señalar que debido a reversibilidad de la fórmula de conversión (1) o (2), un valor de sonoridad BS.1770 original se puede recalcular a partir de un valor equivalente de ajuste de Replay Gain sin pérdida de información de sonoridad. Este es el caso si el valor de ajuste de Replay Gain se ha derivado de la sonoridad ITU por medio de la fórmula de conversión (1) o (2).

A continuación, se considera el aspecto de almacenar los parámetros Replay Gain de acuerdo con la sintaxis Replay Gain. Como se ha resumido antes, el formato Replay Gain típicamente comprende dos tipos de parámetros relacionados con el contenido multimedia, una "amplitud de señal Pico Replay Gain" y un "valor de ajuste de Replay Gain". Estos parámetros se pueden calcular pista por pista o álbum a álbum. Los valores basados en pista son más idóneos para usar casos y listas de reproducción donde se mezclan pistas de diferentes álbumes. Los valores basados en álbumes son más idóneos para usar casos donde todas las pistas de un álbum se reproducen consecutivamente. La "amplitud de señal pico Replay Gain" indica la amplitud absoluta máxima de la señal de audio y puede ser usada para impedir recorte en el reproductor multimedia. En un ejemplo, el valor de "nivel de audio pico verdadero" determinado según la recomendación ITU-R BS.1770-1 puede ser transmitido como parámetro de "amplitud de señal pico Replay Gain".

Además de los parámetros mencionados anteriormente, el formato Replay Gain permite la especificación del originador de los parámetros Replay Gain. Posibles valores de tal originador de parámetros Replay Gain pueden ser, p. ej., el ingeniero, el artista, el productor o el usuario.

Los parámetros Replay Gain anteriores se pueden almacenar usando la sintaxis descrita a continuación. En un ejemplo, contenido multimedia almacenado en archivos que cumplen el estándar de archivo MPEG-4 puede usar preferiblemente metadatos de estilo iTunes. Otros formatos pueden almacenar los parámetros Replay Gain en etiquetas ID3v2 especificadas en la especificación ID3v2. Una sintaxis para ambos casos se resume a continuación:

1) parámetros Replay Gain en metadatos estilo iTunes

Se pueden añadir parámetros Replay Gain como caja de extensión de tipo '----', conformada al estándar metadatos estilo iTunes. Puede haber presente una caja 'significada' dentro de la caja '----' y comprender el significado "org.hydrogenaudio.replaygain" especificado en <http://www.replaygain.org>. Puede haber presente una caja 'nombre' dentro de la caja '----' y comprender el nombre del valor: replaygain_track_gain; replaygain_track_peak; replaygain_album_gain; y/o replaygain_album_peak.

Puede haber presente una caja 'datos' dentro de la caja '----' y comprender el valor en los siguientes formatos:

- Los valores de ajuste de ganancia se escriben preferiblemente como valor de dB en coma flotante con 2 lugares decimales y un prefijo -/+. (p. ej. "-4,65 dB").
- La amplitud de señal pico se escribe preferiblemente como valor con coma flotante (p. ej. "0,860931396"). La amplitud de señal pico debe ser más alta que 1,0.

Para compatibilidad, preferiblemente los reproductores multimedia deben coincidir únicamente en el valor en la caja 'nombre' e ignorar el valor en la caja 'media'.

Se pueden definir metadatos estilo iTunes adicionales para el código de originador de Replay Gain: La Replay Gain puede incluir información de "código de originador". Para esta finalidad, puede haber presente una caja 'significado' dentro de la caja '----' y contener el significado "org.hydrogenaudio.replaygain" de acuerdo con <http://www.replaygain.org>. Puede haber presente una caja 'nombre' dentro de la caja '----' y contener la nombre "replaygain_originator_code". Se pueden usar los siguientes códigos de originador: 000 = Replay Gain no especificado; 001 = Replay Gain preestablecido por el artista / productor / ingeniero de generación; 010 = Replay Gain establecido por el usuario; y/o 011= Replay Gain determinado automáticamente.

Es más, la caja 'datos' puede comprender una cadena de texto que representa la concatenación de los códigos de originador de 3 bits para los valores ReplayGain en el siguiente orden: replaygain_track_gain; replaygain_track_peak; replaygain_album_gain; y/o replaygain_album_peak. Por ejemplo, "011011000000" puede asignarse a valores generados automáticamente para ganancia y pico de pista y valores no especificados para ganancia y pico de álbum.

En un mínimo, un archivo con metadatos Replay Gain debe incluir uno de un valor de ganancia de pista o un valor de ganancia de álbum.

2) Replay Gain en etiquetas ID3v2

Valores de Replay Gain se pueden almacenar en campos 'TXXX' que siguen la siguiente sintaxis:

<pre> <Header for 'User defined text information frame', ID: "TXXX"> Text encoding Sxx Description <string according to encoding> \$00 (00) Value <string according to encoding> </pre>

Tabla 2

5 Cada parámetro Replay Gain puede estar contenido en su propio elemento específico 'TXXX'. Para distinguir parámetros, la cadena "Descripción" puede tomar los mismos valores que los escritos en la caja 'nombre' de iTunes (ver arriba), es decir en particular `replaygain_track_gain`; `replaygain_track_peak`; `replaygain_album_gain`; `replaygain_album_peak`; y/o `replaygain_originator_code`. El valor de parámetro correspondiente a estas descripciones de parámetro puede ser almacenado en el campo "Valor". Puede usar el mismo formato que se describe en la sección iTunes anterior.

En un ejemplo que usa la fórmula (1), los ajustes Replay Gain deben estar entre -16 dB y +9 dB (correspondientes a una amplitud de valores de sonoridad de 0 a -31,25 dBFS). Valores fuera de esta amplitud deben ser limitados a -16 dB y +9 dB. En otro ejemplo que usa la fórmula (2), los valores de ajuste de Replay Gain correspondientes a los valores de sonoridad ITU-R BS.1770 de 0 dBFS y -31,25 dBFS se pueden tomar de la gráfica 220 en la figura 2e.

15 A continuación, se describen diferentes aspectos en relación con un ejemplo de sistema de reproducción que soporta el método propuesto para normalización de sonoridad. En particular, estos aspectos están relacionados con el objetivo deseado de nivel de salida, el control del rango dinámico, y el manejo de archivos multimedia que no comprenden metadatos de sonoridad.

20 Reproductores multimedia portátiles, p. ej., teléfonos móviles, reproductores dedicados de música personal, u ordenadores portátiles, a menudo necesitan soportar diferentes ambientes de escucha. Ejemplos de ambientes de escucha pueden ser un ambiente que usa altavoces integrados, un ambiente que usa salida de auriculares, y/o una salida de línea usada en combinación con una base analógica o digital que puede soportar salida multicanal.

25 Dependiendo del ambiente de escucha seleccionado, puede ser necesario seleccionar objetivos apropiados de niveles de salida. A modo de ejemplo, para el último caso de uso usando una base que puede conectar el dispositivo a equipo Hifi, un menor nivel de objetivo de -31d BFS como el especificado por ejemplo en "modo línea" para Dolby Digital es lo más apropiado, permitiendo así capacidades de rango dinámico completo.

30 Es más, el reproductor multimedia portátil debe poder controlar el rango dinámico según la normalización aplicada de sonoridad. Dado que el valor de sonoridad soportado más bajo es -31 dBFS, todos los objetivos de niveles de salida por encima de -31 dBFS deben soportar prevención de recorte a través de control de rango dinámico. Es decir, si se va a producir un archivo de audio que tiene un valor de sonoridad o nivel de referencia de -31 dBFS en un objetivo de nivel de salida de más de -31 dBFS, se tiene que aplicar una ganancia coincidente positiva que puede provocar recorte de la señal de audio amplificada. Por tanto, los reproductores multimedia portátiles deben al menos proporcionar un limitador a fin de impedir tal posible recorte. Formatos que soportan metadatos para control de rango dinámico tales como Dolby Digital también pueden aplicar tales metadatos antes que la señal sea alimentada al limitador. Por ejemplo, un decodificador Dolby Digital que funciona en modo RF, que tiene un objetivo de nivel de -20 dBFS, requerirá un realce adicional de 9 dB seguido por un limitador para lograr un objetivo de nivel de -11 dBFS sin recortar artefactos. A modo de ejemplo, se podría usar un limitador adelantado con ataque dependiente de señal y tiempos de liberación, que puede impedir recorte incluso para contenido crítico (p. ej. dinámico) sin artefactos audibles.

40 Adicionalmente, un reproductor multimedia portátil debe poder manejar archivos multimedia que no comprenden metadatos relacionados con sonoridad. Cuando se prepara para reproducción de un archivo, un reproductor multimedia puede comprobar primero si hay disponible un valor de replay gain. En casos donde se está reproduciendo un álbum completo, la ganancia de álbum puede preferirse sobre la ganancia de pista. De otro modo, el uso de la ganancia de pista se puede establecer como predeterminado. En casos donde no hay disponible valor de ajuste de Replay Gain, el sistema puede comprobar la presencia de un valor de sonoridad dependiente de formato tal como el parámetro "dialnorm" en Dolby Digital o el nivel de referencia de programa en MPEG AAC. En tales casos, se pueden usar estos valores de sonoridad dependientes de formato. Si no están disponibles valores de ajuste de Replay Gain ni valores de sonoridad dependientes de formato, el reproductor multimedia portátil puede usar un valor predeterminado de sonoridad. A modo de ejemplo, se puede usar un valor predeterminado de sonoridad, es decir, valor predeterminado de ajuste de Replay Gain, de -11 dBFS para contenido de música estéreo y -27 dBFS para contenido de audio/vídeo y multicanal. Estos valores de sonoridad predeterminados se han determinado en las

mediciones estadísticas anteriores en un número estadísticamente relevante de archivos de música.

- 5 En un ejemplo, se establece un objetivo de nivel de salida de -11 dBFS (opcionalmente -8 o -14 dBFS) y se determinan las ganancias para establecer este objetivo de nivel de salida sobre la base de los valores de sonoridad o niveles de referencia proporcionados junto con los archivos multimedia. El rango dinámico de los archivos multimedia amplificados es controlado ya sea con algoritmos de único lado o combinando un limitador de único lado con control de rango dinámico posiblemente derivado de otros metadatos disponibles.

A fin de ilustrar aún más la implementación de un reproductor multimedia que aplica valores de sonoridad de acuerdo con los métodos resumidos en el presente documento, un ejemplo de implementación pseudocódigo de los métodos se proporciona en la Tabla 3.

```
// Determinar objetivo de nivel de salida:
Switch ( currentListeningEnvironment() ) {
    Case BuiltInSpeaker:
        TargetLevel = -8; //dBFS
        break;
    Case LineOut:
        TargetLevel = -31; //dBFS
        break;
    Default:
        TargetLevel = -11; //dBFS;
        // -14 para teléfonos de alta gama con buenos amplificadores
        break;
}
// Determinar nivel de referencia:
If ( ReplayGainPresent() ) {
    Switch ( currentPlaybackMode() ) {
        Case albumPlayback:
            If ( albumGainPresent() ) {
                ReplayGain = getAlbumGain();
            }
    }
}
```

```

        } else {
            ReplayGain = getTrackGain();
        }
        Break;
default:
    If ( trackGainPresent() ) {
        ReplayGain = getTrackGain();
    } else {
        ReplayGain = getAlbumGain();
    }
    Break;
}
ReferenceLevel =
    (ReplayGain + 16) / -0.8;
// conversión a valor de sonoridad BS.1770
} else {
    If ( LevelInfoInPayload() )
        ReferenceLevel =
            getLevelInfoFromPayload();
    } else {
        // Niveles predeterminados
        Switch ( currentMusicFormat ) {
            Case StereoMusic:
                ReferenceLevel = -11; //dBFS
                Break;
            default:
                ReferenceLevel = -27; //dBFS
                Break;
        }
    }
}
// Aplicar compresión de rango dinámico, si está disponible

```

```

If ( metadataAvailable() &&
    TargetLevel >= -20) {
    applyMetadata(RF_MODE);
    //, p. ej., Dolby Digital

    ReferenceLevel = -20; //dBFS
}
// Aplicar ganancia/limitador final

Gain = TargetLevel - ReferenceLevel;
If ( Gain > 0 ) {
    applyLimiterAndGain();
} else {
    applyGain();
}

```

Tabla 3

El dispositivo receptor/descodificador 10 de un ejemplo de reproductor multimedia se ilustra en la figura 3. El dispositivo 10 recibe una señal de entrada codificada del medio de transmisión o camino de señal 11, aplica procesos adecuados en el desformateador 12 para extraer información codificada de audio y metadatos asociados de la señal de entrada, pasa la información de audio codificada al descodificador 14 y pasa los metadatos a lo largo del camino de señal 13. La información de audio codificada puede incluir señales subbanda codificadas que representan contenido espectral de la señal de audio y los metadatos pueden especificar valores de ganancia de sonoridad y otros parámetros que especifican compresión de rango dinámico según un perfil de compresión de rango dinámico. El término "perfil de compresión de rango dinámico" se refiere a rasgos tales como factores de ganancia (sonoridad), tiempos de ataque de compresión y tiempos de liberación de compresión que definen las características operacionales de un compresor de rango dinámico. El descodificador 14 aplica un proceso de descodificación a la información de audio codificada para obtener señales subbanda descodificadas, que se pasan al control de rango dinámico 16. El funcionamiento y las funciones del proceso de descodificación se puede adaptar en respuesta a parámetros de control de descodificación desde el camino de señal 13. Ejemplos de parámetros de control de descodificación que se pueden usar para adaptar el funcionamiento y las funciones del proceso de descodificación son parámetros que identifican el número y la configuración de los canales de audio representados por la información de audio codificada.

El control de rango dinámico 16 ajusta opcionalmente el rango dinámico de la información de audio descodificada. Este ajuste puede ser encendido o apagado y adaptado en respuesta a metadatos recibidos del camino de señal 13 y/o de señales de control que se pueden proporcionar en respuesta a entrada desde un oyente. Por ejemplo, se puede proporcionar una señal de control en respuesta a un oyente que maneja un interruptor o selecciona una opción de funcionamiento para el dispositivo 10. En implementaciones conformes al estándar ATSC, el estándar MPEG-2 AAC o el estándar de Audio MPEG-4, por ejemplo, la señal de entrada codificada incluye información de audio codificada dispuesta en una secuencia de segmentos o tramas. Cada trama contiene señales subbanda codificadas que representan componentes espectrales de una señal de audio con su rango dinámico completo. El control de rango dinámico 16 puede no realizar acción, lo que permite reproducir la señal de audio con una cantidad máxima de rango dinámico, o puede modificar las señales subbanda descodificadas para comprimir el rango dinámico en varios grados. El banco de filtros de síntesis 18 aplica un banco de filtros de síntesis a las señales subbanda descodificadas, que pueden haber sido ajustadas por el control de rango dinámico 16, y proporciona en su salida una señal de audio en dominio de tiempo que puede ser una señal digital o una analógica.

El limitador de ganancia 20 se usa en algunas implementaciones para ajustar la amplitud de la señal de audio en dominio de tiempo. La salida del limitador de ganancia 20 es pasada a lo largo del camino 21 para la subsiguiente presentación por un transductor acústico.

La figura 4 es un diagrama de bloques esquemático de un ejemplo de dispositivo codificador/trasmisor 30 en un emplazamiento de producción o transmisión de contenido de audio que puede incorporar diversos aspectos resumidos en el presente documento. El dispositivo 30 recibe una señal de entrada de audio desde el camino de señal 31. El dispositivo 30 aplica un banco de filtros de análisis a la señal de audio para obtener señales subbanda ya sea en una representación de dominio de frecuencia de la señal de audio de entrada o un conjunto de señales limitadas en ancho

de banda que representan la señal de audio de entrada. La calculadora de metadatos 34 analiza la señal de entrada de audio y/o una o más señales derivadas de la señal de entrada de audio tales como una versión modificada de la señal de entrada de audio o las señales subbanda del banco de filtros de análisis 32 para calcular metadatos que especifican valores para una variedad de parámetros que incluyen parámetros de control de codificación, uno o más parámetros de control de descodificación y uno o más parámetros que especifican compresión de rango dinámico según un perfil de compresión de rango dinámico. La calculadora de metadatos 34 puede analizar señales en dominio de tiempo, señales en dominio de frecuencia, o una combinación de señales en dominio de tiempo y dominio de frecuencia. Los cálculos realizados por la calculadora de metadatos 34 también se puede adaptar en respuesta a uno o más parámetros de metadatos recibidos del camino 33. El codificador 36 aplica un proceso de codificación a la salida del banco de filtros de análisis 32 para obtener información de audio codificada que incluye señales subbanda codificadas, que se pasa al formateador 38. El proceso de codificación se puede adaptar en respuesta a los parámetros de control de codificación recibidos del camino 33. El proceso de codificación también puede generar otros parámetros de control de descodificación a lo largo del camino 33 para uso por procesos realizados en el dispositivo 10 para descodificar la información de audio codificada. El formateador 38 ensambla la información de audio codificada y al menos algunos de los metadatos que incluye el uno o más parámetros de control de descodificación y el uno o más parámetros que especifican compresión de rango dinámico (en particular los valores de ganancia de sonoridad) en una señal de salida codificada que tiene un formato que es adecuado para transmisión o almacenamiento.

En implementaciones conformes al estándar ATSC, el estándar MPEG-2 AAC o el estándar de Audio MPEG-4, por ejemplo, la señal de salida codificada incluye información de audio codificada dispuesta en una secuencia de segmentos o tramas. Cada trama contiene señales subbanda codificadas que representan componentes espectrales de una señal de audio con su rango dinámico completo y que tiene amplitudes para reproducción en un nivel de reproducción de referencia.

La figura 5 ilustra un ejemplo de sistema global 500 que comprende un codificador 511 de audio y un descodificador de audio 521, así como Control de Rango Dinámico y Ajuste de Sonoridad opcionales 522 según Modo de línea/RF o de música. El codificador de audio 511 puede ser un codificador Dolby Digital, Dolby Digital Plus, Dolby Pulse, HE-AAC, mp3, o cualquier otro. Mediciones de sonoridad se realizan sobre la base de ITU-R BS.1770 como indica la unidad 512. La sonoridad BS.1770 medida o proporcionada es transmitida directamente en el flujo de bits como metadatos "dialnorm" en Dolby Digital/Dolby Digital Plus o como "prog_ref_level" en Dolby Pulse o MPEG HE-AAC. Esto es indicado por la flecha de puntos 513. Como alternativa o adicionalmente, la sonoridad puede ser transportada en una o más etiquetas ID3v2 o metadatos estilo iTunes por medio de una conversión a Replay Gain. Tal conversión puede ser realizada en la unidad de conversión de sonoridad 514. Las etiquetas ID3v2 o metadatos estilo iTunes 515 se proporcionan desde el lado de codificación 510 al lado de descodificación 520. Cabe señalar que un Codificador Dolby Pulse comprende los bloques de procesamiento como indica la unidad 510 y que un Descodificador Dolby Pulse comprende los bloques de procesamiento como indica la unidad 520. En el lado de descodificación 520 se convierten metadatos Replay Gain a sonoridad ITU-R BS.1770 usando la unidad de conversión inversa 523. La unidad de conversión 514 y la unidad de conversión inversa 523 se aplican típicamente a una relación reversible como se resume en el presente documento. En un ejemplo, los metadatos Replay Gain recibidos en el lado de descodificación 520 pueden ser metadatos Replay Gain originales 530 ya sea computados en el lado de codificador 510 como se muestra anteriormente o como posprocesamiento en el lado de descodificador o reproductor 520, el valor de sonoridad que es determinado en la unidad de conversión inversa 523 típicamente es una aproximación a la sonoridad ITU-BS.1770 verdadera. Cabe señalar que opcionalmente, se puede usar información de sonoridad en el lado de codificador 510 para computar metadatos de Control de Rango Dinámico (DRC). Estos metadatos DE DRC se pueden proporcionar al lado de descodificador 520 por medio del enlace 513 y pueden ser usados en la unidad de descodificación 521.

En el presente documento, se ha descrito un método y un sistema para normalización de sonoridad en reproductores multimedia portátiles. El método y el sistema dependen de la combinación de la medición estandarizada de sonoridad BS.1770 y los contenedores Replay Gain usados en diversos reproductores multimedia portátiles. Se ha determinado una fórmula a partir de mediciones estadísticas que se puede usar para convertir valores de sonoridad BS.1770 a valores Replay Gain y viceversa, en el lado de codificar, así como en el lado de descodificar. Como tal, se pueden lograr una normalización de sonoridad según BS. 1770 que es totalmente compatible con nivelación actual basada en dialnorm. El sistema resulta en normalización de sonoridad estadísticamente óptima incluso para contenido mezclado, p. ej., contenido que ha sido analizado parcialmente con BS.1770 y parcialmente analizado por el algoritmo Replay Gain. El método y el sistema se pueden implementar con baja complejidad computacional, en particular en el reproductor multimedia portátil.

Los métodos y los sistemas descritos en el presente documento pueden ser implementados como software, firmware y/o hardware. Ciertos componentes se pueden implementar, p. ej., como software ejecutado en un procesador de señales digitales o microprocesador. Otros componentes se pueden implementar, p. ej., como hardware y/o como circuitos integrados específicos de aplicación. Las señales encontradas en los métodos y sistemas descritos pueden ser almacenadas en medios tales como memoria de acceso aleatorio o soportes de almacenamientos ópticos. Puede ser transferidos por medio de redes, tales como redes de radio, redes de satélites, redes inalámbricas o redes cableadas, p. ej., internet. Dispositivos típicos que hacen uso de los métodos descritos en el presente documento son los reproductores multimedia que descodifican señales de audio. En el lado de codificación, los sistemas y métodos pueden ser usados en estaciones de difusión y en lugares de producción multimedia.

REIVINDICACIONES

- 1 Un método para usar metadatos generados por codificador en operaciones de decodificación de audio, el método comprende:
- 5 recibir, por un descodificador de audio (14, 521), una señal de audio de entrada codificada para una configuración de canal de audio, la señal de audio de entrada incluye muestras de audio;
- recibir los metadatos generados por codificador (13, 513, 515), los metadatos generados por codificador comprenden:
- un primer valor relacionado con sonoridad;
 - una pluralidad de valores de ganancia según un perfil de control de rango dinámico que define las características operacionales de un compresor de rango dinámico;
 - 10 parámetros que identifican varios canales de audio y la configuración de canal de audio;
- convertir las muestras de audio en una señal de audio de salida con un objetivo de nivel de salida, la conversión comprende:
- realizar control de rango dinámico (16) usando dicha pluralidad de valores de ganancia, resultando de ese modo muestras de audio controladas por rango dinámico;
 - 15 convertir dicho primer valor relacionado con sonoridad a un segundo valor relacionado con sonoridad basado en una relación reversible configurada para convertir primeros valores relacionados con sonoridad determinados según un primer procedimiento en segundos valores relacionados con sonoridad determinados según un segundo procedimiento;
 - 20 derivar un valor de ganancia a partir del objetivo de nivel de salida y el segundo valor relacionado con sonoridad; y
 - aplicar dicho valor de ganancia a las muestras de audio controladas por rango dinámico.
- 2 Un dispositivo de retrasmisión multimedia configurado para realizar el método de la reivindicación 1.
- 3 Un reproductor multimedia configurado para realizar el método de la reivindicación 1.
- 4 Un sistema (500) para usar datos de control generados por codificador en operaciones de decodificación de audio, el sistema comprende:
- 25 un reproductor multimedia configurado para realizar el método de la reivindicación 1;
- un medio de transmisión configurado para proporcionar la señal de audio de entrada y los metadatos generados por codificador (13, 513, 515) al reproductor multimedia;
- 30 un dispositivo de codificación de audio (36, 510) configurado para generar la señal de audio de entrada y los metadatos generados por codificador.
- 5 Un producto de programa informático que comprende instrucciones ejecutables para realizar el método de la reivindicación 1.

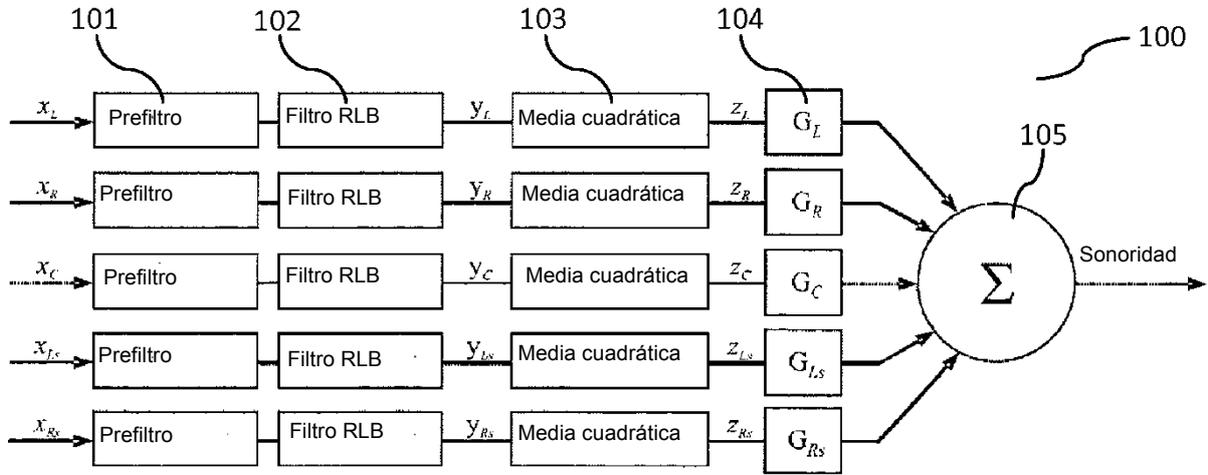


Fig. 1a

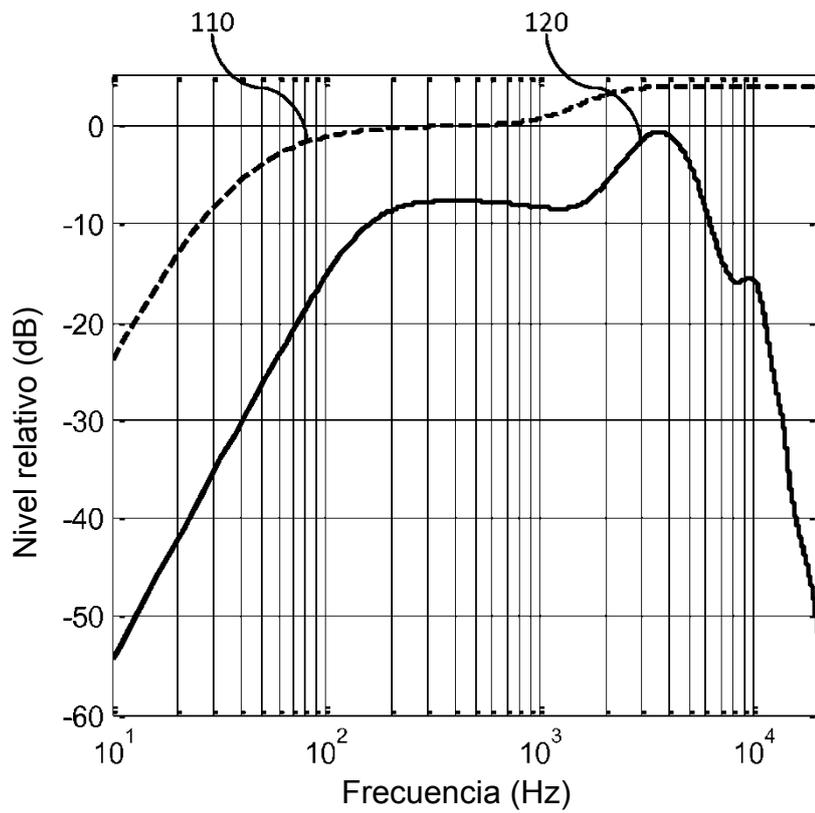


Fig. 1b

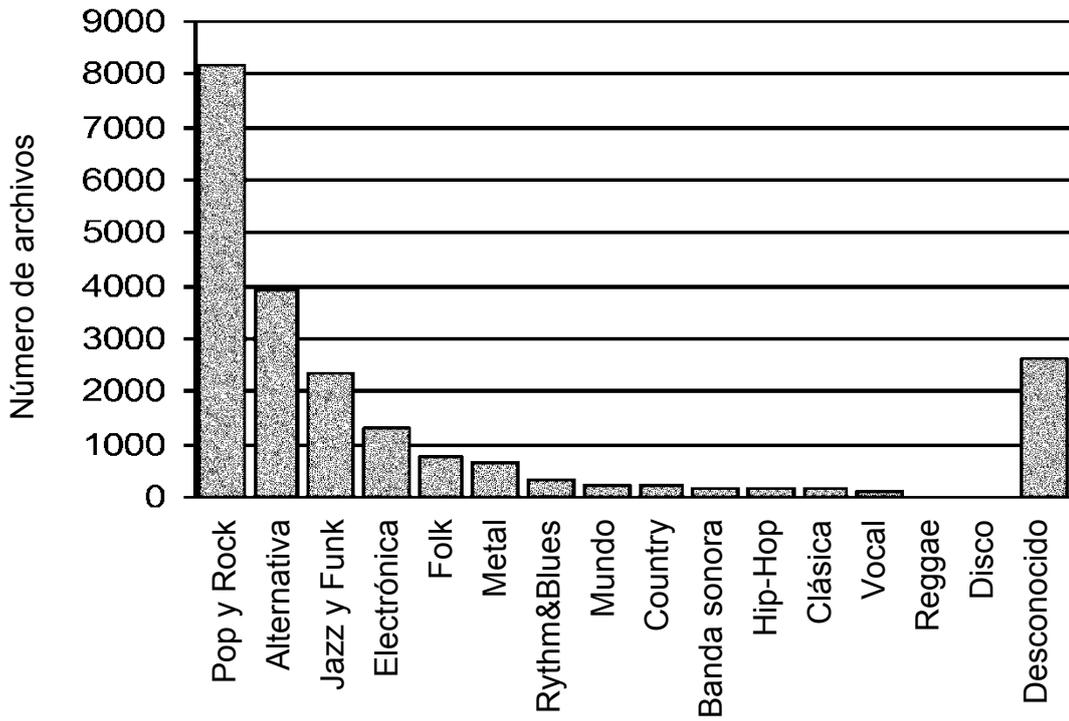


Fig. 2a

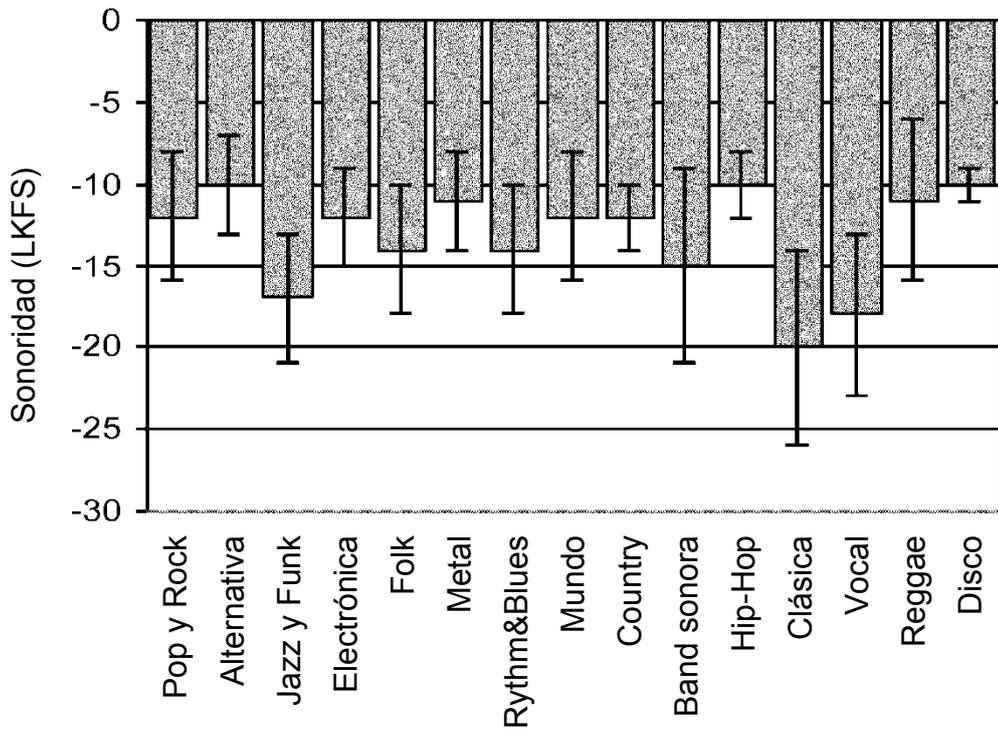


Fig. 2b

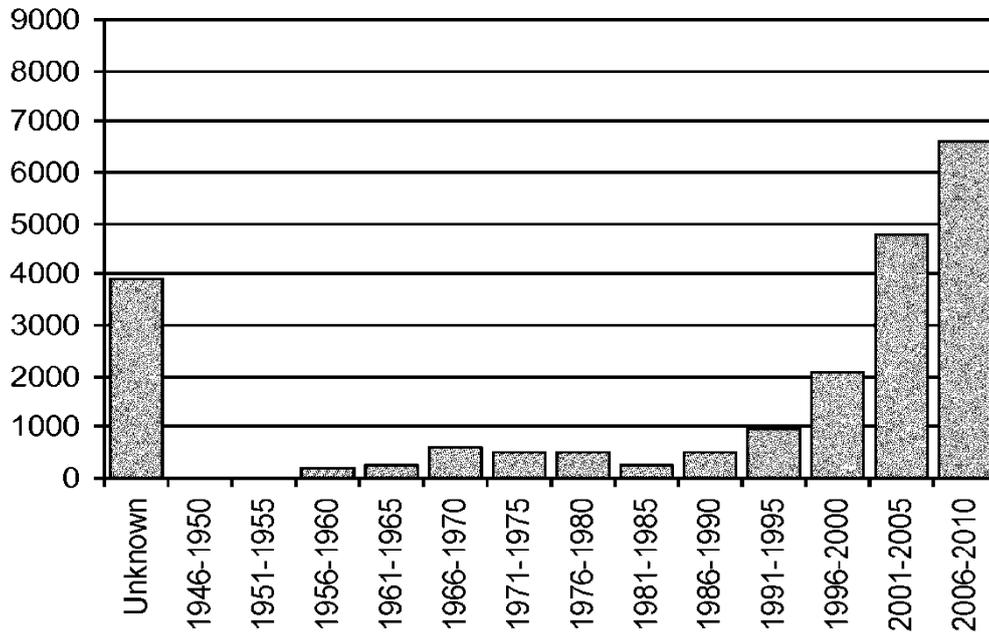


Fig. 2c

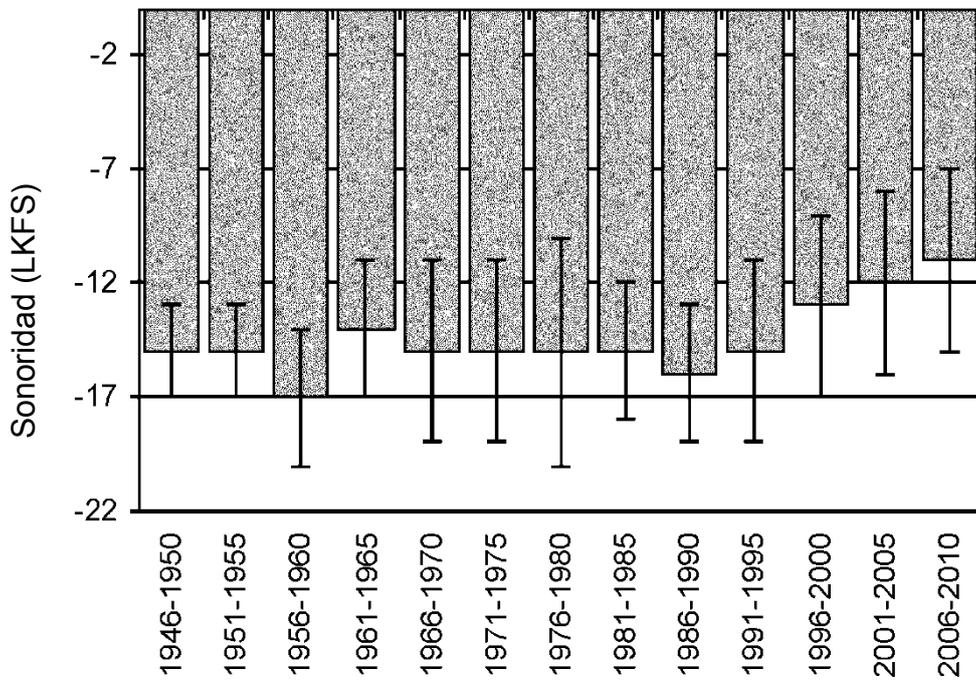


Fig. 2d

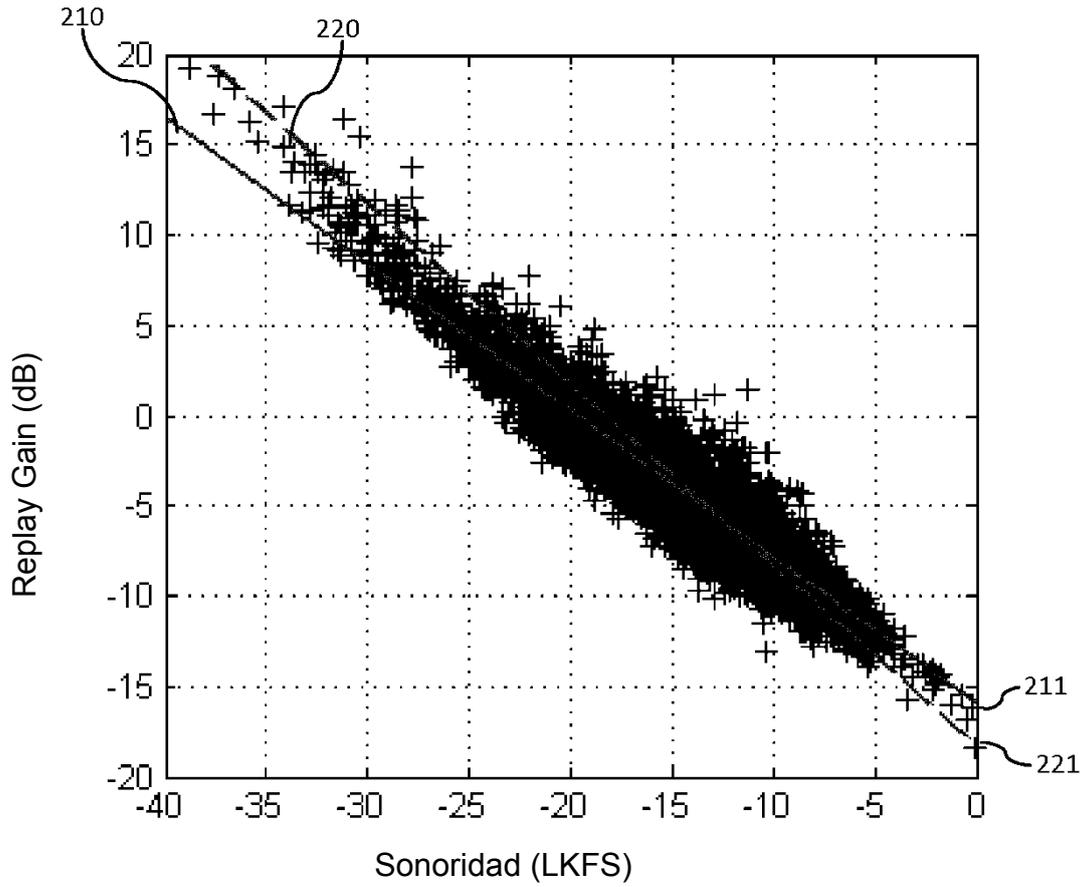


Fig. 2e

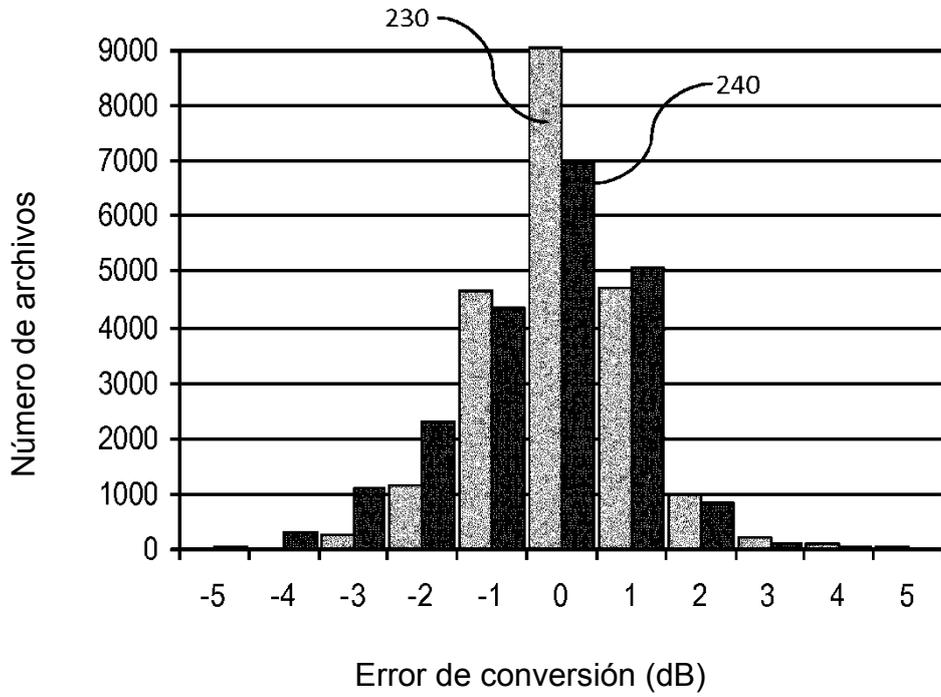


Fig. 2f

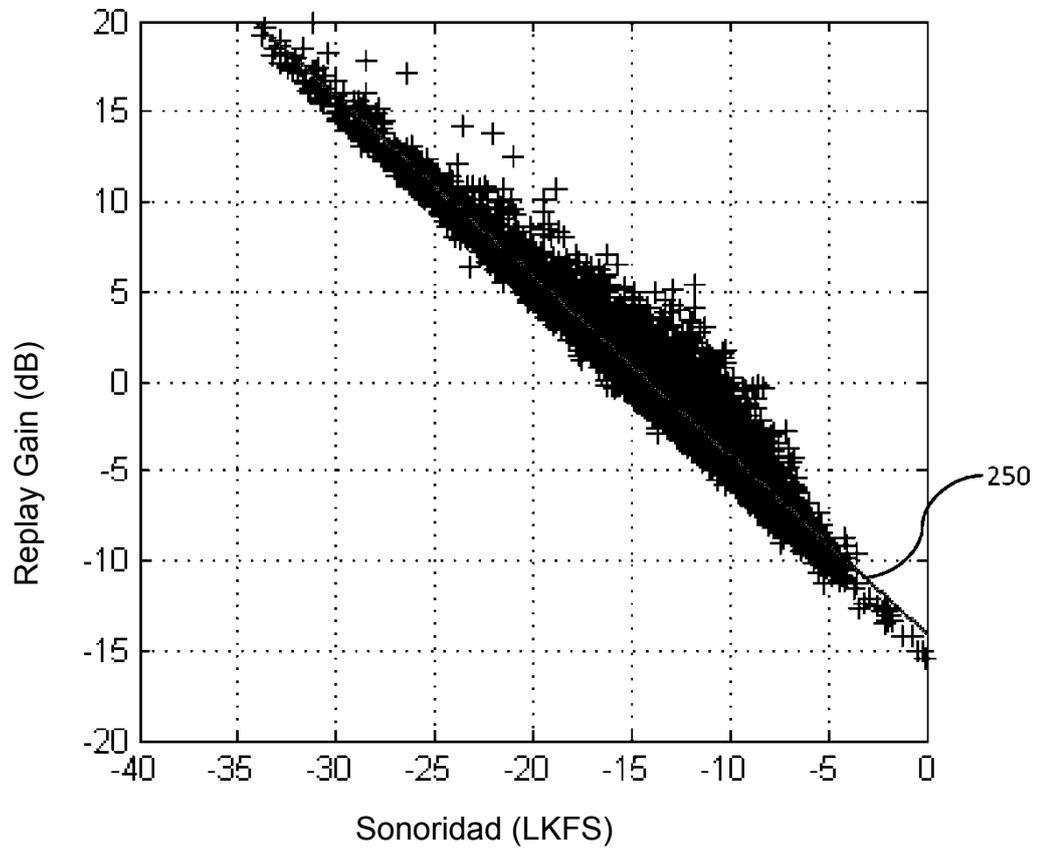


Fig. 2g

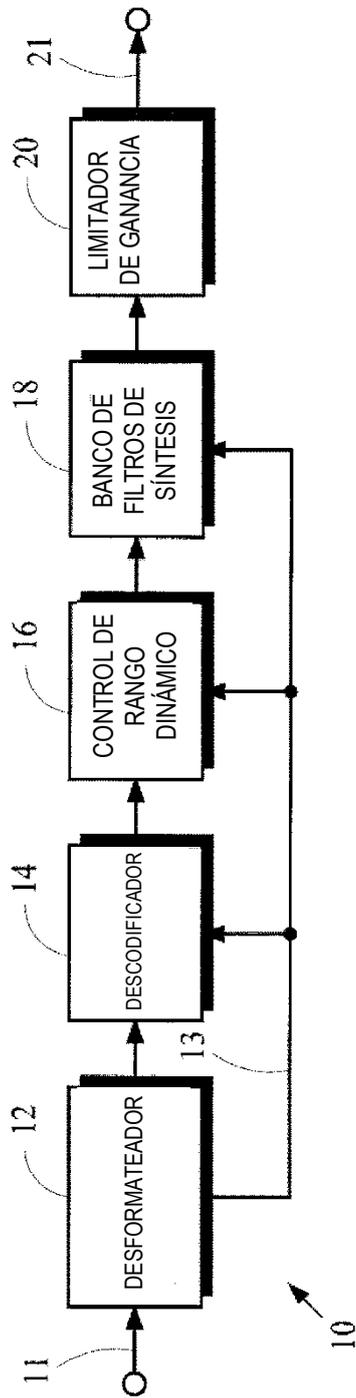


Fig. 3

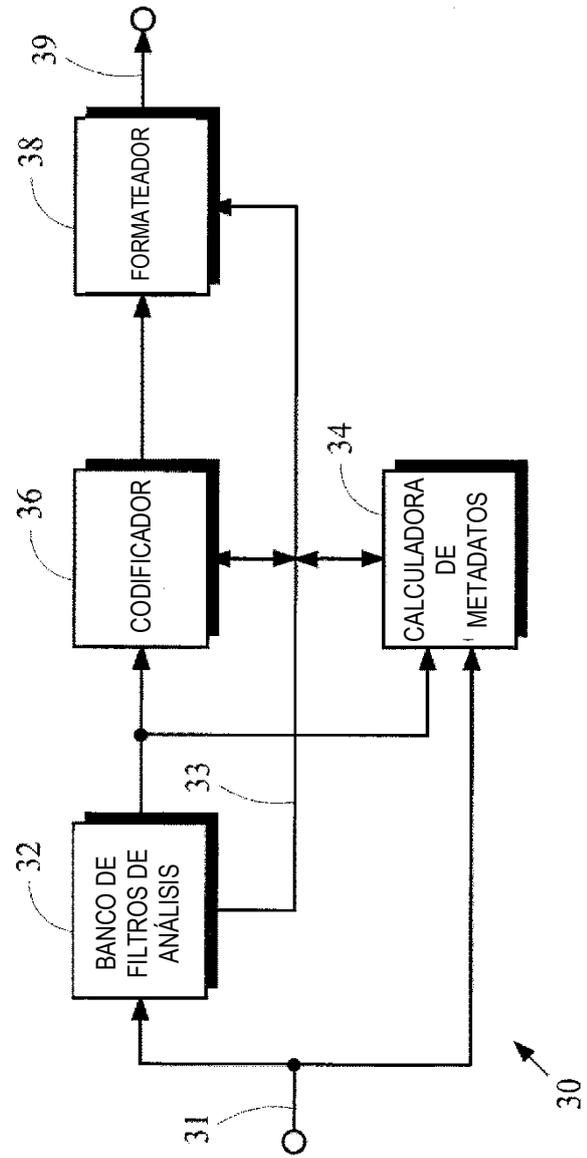


Fig. 4

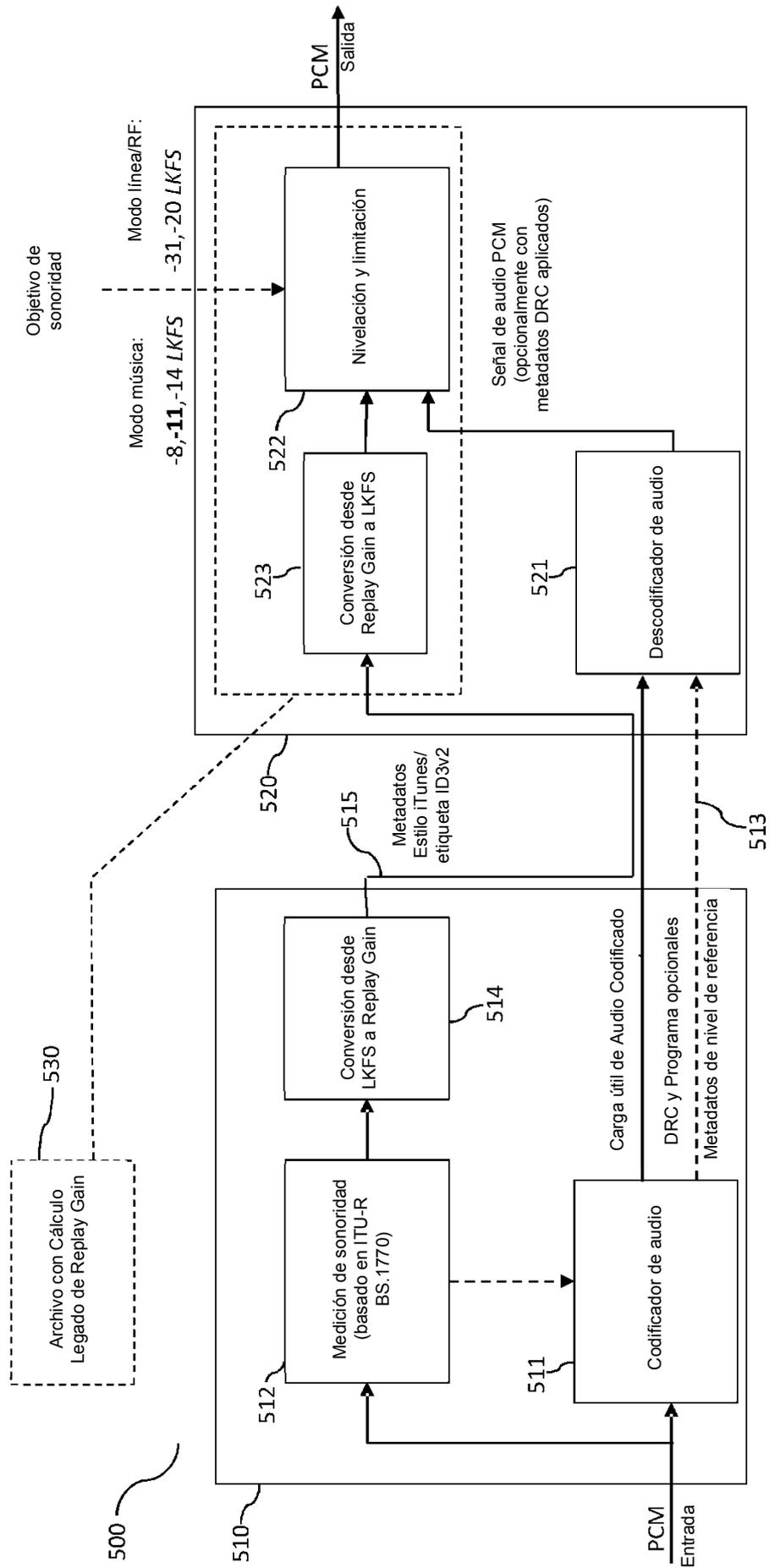


Fig. 5