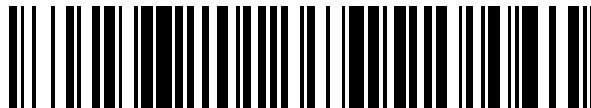


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 530 360**

51 Int. Cl.:

H03G 7/00 (2006.01)

H04S 7/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **17.10.2007** **E 13169247 (7)**

97 Fecha y número de publicación de la concesión europea: **17.12.2014** **EP 2632044**

54 Título: **Generación jerárquica de parámetros de control para el procesamiento de la dinámica del audio**

30 Prioridad:

01.11.2006 US 856586 P

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

02.03.2015

73 Titular/es:

**DOLBY LABORATORIES LICENSING CORPORATION (100.0%)
100 Potrero Avenue San Francisco
California 94103-4813 , US**

72 Inventor/es:

**SEEFELDT, ALAN y
GUNDRY, KENNETH**

74 Agente/Representante:

LINAGE GONZÁLEZ, Rafael

ES 2 530 360 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

Generación jerárquica de parámetros de control para el procesamiento de la dinámica del audio

5 **Campo técnico**

La invención se refiere a procesadores de señales de audio, al procesamiento de señales de audio y a software para controlar procesadores de señales de audio y el procesamiento de señales de audio. De acuerdo con algunos aspectos de la invención, se proporciona una ruta de control jerárquico para el procesamiento de la dinámica y se imponen restricciones, de manera jerárquica, sobre el procesamiento de la dinámica de señales de audio para mejorar el rendimiento.

Técnica antecedente

15 El objeto del procesamiento de la dinámica de señales de audio consiste en alterar el nivel relativo o dinámica de una señal de audio para que se mantenga dentro de los límites deseados. Por lo general, se logra creando algún tipo de medida, variable con el tiempo, del nivel de una señal de audio (por ejemplo, nivel r.m.s. o nivel pico) y, a continuación, computando y aplicando una modificación de la señal (por ejemplo, un cambio de ganancia) que sea una función de la medida del nivel. En la solicitud internacional de patente PCT/US2005/038579, a nombre de Alan Jeffrey Seefeldt y publicada con el número WO 2006/47600 el 4 de mayo de 2006, se dan a conocer procesadores de dinámica que comparten dicho modo de funcionamiento y que incluyen controles automáticos de ganancia (AGC), controles de gama dinámica (DRC), expansores, limitadores, puertas de ruido, etc. La solicitud de Seefeldt designa Estados Unidos entre otras entidades. Puede encontrarse técnica anterior relevante adicional en los documentos EP1251715A2, US 2006/0215852A1

25 La figura 1 ilustra un diagrama de bloques de alto nivel de un procesador genérico de dinámica de audio que procesa una señal de audio (un único canal de una señal de audio multicanal o una señal de audio de un solo canal). Se puede considerar que el procesador incluye dos rutas: una ruta superior de «señal» 2 y una ruta inferior de «control» 4. En la ruta inferior de control, el nivel de la señal de audio se mide mediante un dispositivo o proceso de medición («medida de nivel») 6 y esta medición, una medida del nivel de la señal, es usada más tarde por un dispositivo o proceso de control de dinámica («control de dinámica») 8 para computar uno o más parámetros de modificación de señales. Dichos parámetros funcionan como señales de control de modificación de señales y se usan para modificar la señal de audio de acuerdo con una función de procesamiento de dinámica, y dicha función puede ser un perfil deseado de procesamiento de dinámica como el que se muestra en la figura 3b, que se describe más adelante. Tal como se muestra, los parámetros de modificación se obtienen a partir de la señal de audio de entrada.

40 Alternativamente, los parámetros de modificación pueden obtenerse a partir del audio procesado (de salida) o de una combinación de las señales de audio de entrada y de salida. En la ruta de la señal de audio 2, los parámetros de modificación generados por el control de dinámica 8 se aplican al audio para controlar la modificación del audio, con lo cual se genera el audio procesado. La aplicación de los parámetros de modificación a una señal de audio se puede lograr de diversas maneras conocidas y se indica de modo genérico mediante el símbolo de multiplicación 12. En la ruta de la señal de audio 2, se puede retardar el audio mediante un dispositivo o proceso de retardo («retardo») 10 para compensar cualquier retardo relacionado con los procesos de estimación de niveles y control de dinámica.

50 Cuando se trabaja con material de audio multicanal complejo, es preciso actuar con precaución a la hora de computar y aplicar las modificaciones de las señales con el fin de evitar que se introduzcan artefactos perceptibles. Un procesador de dinámica básico que reciba la entrada de una señal de audio multicanal podría computar un nivel de señal que fuese representativo de todos los canales combinados en total y, a continuación, aplicar la misma modificación a todos los canales en función de dicha medida de nivel total. Al modificar todos los canales de la misma manera, dicho enfoque presenta la ventaja de mantener los niveles relativos entre todos los canales, con lo cual se conserva la imagen espacial (incluida, por ejemplo, la ubicación de imágenes virtuales distribuidas panorámicamente entre los canales, así como la dispersión percibida). Dicho enfoque puede funcionar bien si las modificaciones deseadas no son demasiado agresivas.

60 No obstante, pueden surgir problemas cuando las modificaciones deseadas son más severas. Consideremos una señal de audio multicanal (por ejemplo, 5.1 canales) a la que se le aplica un controlador de gama dinámica con una relación de compresión muy alta. Con este tipo de procesador, las señales que superan el umbral de compresión se atenúan de manera considerable para que el nivel de la señal se aproxime al umbral. Supongamos que la señal de audio contiene música de fondo con un nivel relativamente constante en todos los canales, para los cuales el nivel total tras combinar todos los canales se encuentre por debajo del umbral de compresión. Supongamos también que en el canal central se introduce un segmento de diálogo breve pero con un volumen alto. A causa del diálogo, el nivel total de todos los canales combinados supera ahora el umbral de compresión y, por lo tanto, se atenúa la totalidad de la señal. Cuando el diálogo termina, el nivel de la señal vuelve a caer por debajo del umbral de compresión y no se aplica la atenuación. A consecuencia de ello, en la música de fondo de los canales izquierdo,

derecho, envolvente izquierdo y envolvente derecho, se oyen fluctuaciones en el nivel o el denominado «bombeo» oscilante en función del diálogo del canal central. El efecto puede sonar muy poco natural y puede resultar desconcertante para el oyente. Esta clase de artefactos, un tipo de modulación cruzada o intermodulación, ha sido reconocida ampliamente dentro del campo del procesamiento de la dinámica del audio y una solución típica de la técnica anterior conlleva la aplicación del control de gama dinámica a cada canal de manera independiente. Aunque dicha solución puede llegar a corregir el problema mencionado, es posible que presente la desventaja de alterar la imagen espacial del audio. En particular, las fuentes virtuales distribuidas panorámicamente entre dos canales pueden dar la impresión de «fluctuar» debido a las diferentes cantidades de atenuación aplicadas a los dos canales. De este modo, existe la necesidad de hallar una solución que haga frente a los problemas de bombeo y de imagen inestable.

Existen problemas análogos cuando se considera el espectro de un único canal de audio. Consideremos un único canal que contiene una nota de cuerda sostenida, a frecuencias entre medias y altas, para la cual el nivel de señal se encuentra por debajo del umbral de compresión. Consideremos ahora un golpe de bombo con un volumen muy alto introducido en las frecuencias bajas y que hace que el nivel de la señal aumente momentáneamente y supere el umbral de compresión. Momentáneamente, se atenúa la totalidad de la señal, por lo cual se percibe en las cuerdas un efecto de bombeo en el nivel vinculado al bombo, que resulta muy poco natural. Una solución típica de la técnica anterior a este problema consiste en dividir la señal de audio en múltiples bandas de frecuencia y aplicar después el control de gama dinámica a cada banda de manera independiente. Esto reduce el problema del bombeo, pero puede alterar el balance espectral o el timbre percibidos. Por tanto, existe la necesidad de hallar una solución que reduzca el bombeo al tiempo que reduce los cambios en el balance espectral percibido.

Exposición de la invención

De acuerdo con la presente invención, se proporciona un método según la reivindicación 1. Por otra parte, según la invención, se proporciona un aparato tal como se define en la reivindicación 12 y se proporciona un programa informático tal como se define en la reivindicación 13. Los aspectos adicionales de la invención se definen en las reivindicaciones dependientes.

La figura 2a ilustra una vista general esquemática de un ejemplo de arquitectura jerárquica de ruta de control en la que se hace pasar una información desde arriba hacia abajo a través de una ruta de control jerárquico (un sistema jerárquico descendente). En este ejemplo de jerarquía de ruta de control descendente, la señal de entrada introducida en el procesador o el proceso de la dinámica consiste en una señal de audio multibanda y los niveles jerárquicos son: (1) todas las bandas de todos los canales (el nivel jerárquico «total»), (2) los canales contenidos en la señal de audio multicanal (el nivel jerárquico de «canales»), y (3) las bandas de frecuencia contenidas en cada canal (el nivel jerárquico de «bandas»). Las salidas de los respectivos dispositivos o procesos de control de dinámica en el nivel jerárquico de bandas, a la derecha de la figura, pueden actuar como señales de control de modificación de señales y se puede aplicar con el fin de modificar las señales de audio en los respectivos canales y bandas de una señal de audio que se esté procesando, tal como se muestra en la figura 2b, descrita más adelante. Otra posibilidad consiste en que dichas señales de control estén incluidas en la señal de audio a modo de metadatos, por ejemplo, de manera que la información se pueda aplicar a la señal de audio, del modo que se muestra en la figura 2b, por ejemplo, en un punto espacial y/o temporalmente remoto. Como se explica más adelante, la información que se hace descender a través de la jerarquía puede incluir otros parámetros además de los de modificación, que resulten adecuados para controlar la modificación de la señal de audio si se aplicasen a la misma- la información que desciende a través de la jerarquía puede incluir parámetros de control que se pueden utilizar dentro de los dispositivos o procesos de control de dinámica para influir sobre el comportamiento de dichos dispositivos o procesos (dichos parámetros de control se pueden denominar «parámetros de control de dinámica» y dichos parámetros pueden incluir «parámetros de modificación» adecuados para controlar la señal de audio de manera directa o indirecta).

La señal de entrada aplicada a la ruta de control jerárquico de la figura 2a se puede obtener a partir de una señal de audio aplicada a un procesador de dinámica de audio del que forme parte la ruta de control de la figura 2a, o, como otra posibilidad, a partir del audio procesado (de salida) de dicho procesador de dinámica de audio o a partir de una combinación de dichas señales de audio de entrada y salida.

En la figura 2a, las líneas continuas representan el flujo de audio de una señal de audio multicanal que presenta los canales $x_1 \dots x_c$, y cada canal presenta una o más bandas de frecuencia o «subbandas» $b_1 \dots b_n$. Aunque es posible un procesamiento analógico, digital o híbrido analógico/digital, en las formas de realización prácticas el procesamiento del audio se puede llevar a cabo en el dominio digital. La línea discontinua de salida de cada dispositivo o proceso de control de dinámica representa uno o más parámetros de control de dinámica generados por un respectivo dispositivo o proceso de control de dinámica. Los parámetros de control de dinámica se pueden asociar a dispositivos o procesos de control de dinámica en niveles jerárquicos inferiores de la jerarquía de la ruta de control con el fin de imponer restricciones. Por ejemplo, tal como se explica en relación con un ejemplo descrito más adelante, se pueden llevar en sentido descendente valores de nivel de señal a largo plazo, ganancia a largo plazo y ganancia a corto plazo desde el nivel de jerarquía total al nivel de jerarquía de canales y desde el nivel de jerarquía de canales al nivel de jerarquía de bandas para establecer restricciones. Entre estos valores, solo se aplica la

ganancia a corto plazo en el nivel de jerarquía de bandas para controlar la dinámica de la señal de audio en los ejemplos descritos en la presente memoria descriptiva (y dicha ganancia a corto plazo puede estar suavizada en el tiempo, al igual que en un ejemplo descrito).

5 En principio, la información que se hace pasar en sentido descendente a través de la jerarquía puede no aplicarse a uno o más niveles jerárquicos. Además, en principio, la información procedente de niveles jerárquicos distintos del nivel jerárquico de bandas se puede aplicar (además o en lugar de la información procedente del nivel jerárquico de bandas) para controlar la dinámica de la señal de audio. Dicha información puede estar suavizada en el tiempo.

10 De este modo, en el ejemplo de ruta de control de la figura 2a, se aplica una señal de audio multicanal que presenta los canales $x_1 \dots x_c$ a un primer nivel de ruta de control jerárquico, el nivel jerárquico total, en el que un dispositivo o proceso («medida de nivel total») 14 computa una medida del nivel de la señal media para la totalidad de canales de audio combinados. El funcionamiento del nivel total 14 es similar al del nivel 6 de la figura 1, a excepción de que el nivel total 14 actúa en todos los canales de una señal de audio multicanal en lugar de en el único canal de una señal de audio multicanal o una señal de audio que sólo presente un canal. La medición obtenida por el nivel total 14, una medida del nivel de señal total, puede ser empleada después por un dispositivo o proceso de control de dinámica («control de dinámica») 16 para computar uno o más parámetros de control de dinámica del nivel jerárquico total que pueden incluir uno o más parámetros de modificación de señal. Algunos o todos los parámetros de control de dinámica computados por el control de dinámica 16 se pueden llevar en sentido descendente hasta los controles de dinámica 20 en el nivel jerárquico de canales para ser utilizados por dichos controles de dinámica en sus operaciones. Se pueden aplicar a la señal de audio los parámetros de modificación computados por el control de dinámica 16 del nivel jerárquico total, o funciones del mismo, pero en este ejemplo no se aplican.

25 A continuación, en un segundo nivel de ruta de control jerárquico, el nivel jerárquico de canales, el nivel jerárquico inmediatamente inferior en este ejemplo, cada canal desde x_1 hasta x_c se aplica a su propio par de dispositivos o procesos de control de dinámica y medida de nivel 18-1 hasta 18-c y 20-1 hasta 20-c, respectivamente. Cada medida de nivel 18-1 hasta 18-c puede funcionar de manera similar a la medida de nivel 6 de la figura 1 para computar una medida de nivel de señal promedio para el canal de audio concreto que se le aplica. A partir de esa medida de nivel, cada uno de los controles de dinámica 20-1 hasta 20-c del par nivel/control de dinámica computa uno o más parámetros de control de dinámica que pueden incluir uno o más parámetros de modificación de señal. No obstante, además de diferenciarse del control de dinámica 8 de la figura 1 por producir parámetros de control de dinámica en lugar de parámetros de modificación, los controles de dinámica 20-1 hasta 20-c en el esquema de la figura 2a también se diferencian por recibir uno o más parámetros de control de dinámica calculados por el control de dinámica 16 en la etapa jerárquica inmediatamente superior. Dichos parámetros de control de dinámica de un nivel jerárquico superior se pueden emplear para restringir el funcionamiento de los respectivos controles de dinámica 20-1 hasta 20-c, o para influir sobre ellos. Algunos o todos los parámetros de control de dinámica computados por los controles de dinámica 20 se pueden llevar en sentido descendente hasta los controles de dinámica 26 en el nivel jerárquico de bandas para ser utilizados por dichos controles de dinámica en sus operaciones. Se pueden aplicar a la señal de audio parámetros de modificación, o funciones de los mismos, computados por uno o más de los controles de dinámica 20, pero en este ejemplo no se aplican.

45 A continuación, en un tercer nivel de ruta de control jerárquico, el nivel jerárquico de bandas, cada canal desde x_1 hasta x_c de la señal de audio se aplica a un dispositivo o función de división de banda («división de banda») 22-1 hasta 22-c, que divide cada canal de audio en una pluralidad de bandas de frecuencia b_1 hasta b_n . En la técnica se conocen muchos dispositivos y procesos para dividir una señal de audio en bandas de frecuencia (bancos de filtros, filtros de paso de banda, filtros espejo en cuadratura, transformadas de tiempo a frecuencia, etc.). Cada banda de frecuencia de cada canal se aplica a su propio par de dispositivos o procesos de control de dinámica y medida de nivel 24-1-1 hasta 24-c-n y 26-1-1 hasta 26-c-n, respectivamente. Cada medida de nivel 24-1-1 hasta 24-c-n puede actuar de manera similar a la medida de nivel 6 de la figura 1 para computar una medida de nivel de señal medio para la banda de frecuencia concreta del canal de audio que se le aplica - por tanto, funciona a un nivel jerárquico de banda de frecuencia en lugar de un nivel jerárquico de canal. A partir de la medida del nivel de la señal proporcionada por la medida de nivel con la que está emparejada, cada uno de los controles de dinámica 26-1-1 hasta 26-c-n computa parámetros de modificación de señal, de manera similar al control de dinámica 8 de la figura 1. No obstante, a diferencia del control de dinámica 8 de la figura 1, los controles de dinámica 26-1-1 hasta 26-c-n del ejemplo de la figura 2a también pueden recibir uno o más parámetros de control de dinámica calculados por el control de dinámica 20 asociado con el mismo canal en la etapa jerárquica inmediatamente superior. Dichos parámetros de modificación de un nivel jerárquico superior se pueden emplear para restringir el funcionamiento de los respectivos controles de dinámica 26-1-1 hasta 26-c-n que actúan sobre bandas comprendidas en un canal, o para influir en dichos controles. Uno o más de los parámetros de modificación calculados por cada uno de los controles de dinámica 26-1-1 hasta 26-c-n se pueden aplicar a bandas de cada canal del audio, tal como se muestra en la figura 2b. En el nivel jerárquico de bandas, debido a que constituye el nivel más bajo de la jerarquía en este ejemplo, los controles de dinámica 26 no están obligados a proporcionar una salida de parámetros de control de dinámica (no son necesarios), sino que, en cambio, proporcionan una salida de uno o más parámetros de modificación que pueden ser utilizados para controlar la modificación de la señal de audio, tal como se muestra en la figura 2b.

En referencia a la figura 2b, de manera similar a lo que ocurre con la ruta de señal superior de la figura 1, se aplica una banda concreta de cada canal de audio (canal de audio 1 banda 1 hasta canal de audio c banda n) a una ruta de señal correspondiente que presenta un retardo correspondiente 10-1-1 hasta 10-c-n y un multiplicador correspondiente 12-1-1 hasta 12-c-n. Al igual que en la figura 1, los multiplicadores se representan genéricamente mediante un símbolo multiplicador para indicar cualquier manera adecuada de modificar el audio en función de uno o más parámetros de modificación. Cada uno de los multiplicadores recibe uno o más parámetros de modificación procedentes de un respectivo control de dinámica 26-1-1 hasta 26-c-n (es decir, el audio del canal 1 banda 1 es modificado por los parámetros de modificación procedentes del control de dinámica para el canal 1 banda 1, etc.). Otra posibilidad consistiría en aplicar parámetros de modificación, o funciones de estos parámetros, producidos por otros niveles de la jerarquía a la señal de audio, uno o más de sus canales y/o una o más bandas de uno o más de sus canales.

Aunque el ejemplo de las figuras 2a y 2b muestra una jerarquía de tres niveles constituida por el nivel jerárquico total, el nivel jerárquico de los canales y el nivel jerárquico de las bandas, la jerarquía puede ser de un orden diferente. Por ejemplo, es posible que no se desee llevar a cabo modificaciones multibanda, en cuyo caso, podría resultar apropiada una jerarquía de dos niveles que incluya únicamente un nivel jerárquico total y un nivel jerárquico de canal. Asimismo, podría darse el caso de que se no deseara ningún grado de independencia entre los canales y, por lo tanto, se podría utilizar una jerarquía de dos niveles que incluya únicamente un nivel jerárquico total y un nivel jerárquico de banda. Es decir, puede que no se deseen aplicar diferentes parámetros de modificación a diferentes canales - las bandas pueden ser tratadas de diferente modo, pero cada banda se puede modificar de la misma manera en todos los canales. Como se mencionó anteriormente, cada nivel jerárquico puede comprender uno o más dispositivos o procesos. El ejemplo de jerarquía de tres niveles de las figuras 2a y 6 se puede ampliar para que incluya más niveles jerárquicos. Por ejemplo, en lugar de dividir la señal en una pluralidad de bandas en un nivel, se podría dividir la señal en dos bandas en un primer nivel de banda. Después, cada una de estas bandas se puede subdividir en dos bandas más en un segundo nivel de banda, etc. De igual modo, la división en canales se puede producir en más de un nivel jerárquico. Por ejemplo, después del nivel total, los canales se podrían dividir en tres grupos en un primer nivel jerárquico de canal: (1) izquierdo y derecho, (2) central, y (3) envolvente izquierdo y envolvente derecho. A continuación, en el siguiente nivel jerárquico de canal, se puede dividir explícitamente el grupo «izquierdo y derecho» en «izquierdo» y «derecho». Asimismo, el grupo «envolvente izquierdo y envolvente derecho» se puede dividir de manera explícita en «envolvente izquierdo» y «envolvente derecho».

Una jerarquía de ruta de control que presente restricciones al modo del ejemplo de las figuras 2a y 2b y variaciones de las mismas se puede aplicar a una gran variedad de aplicaciones de procesamiento de dinámica, tales como compresores, limitadores, expansores, etc. No obstante, para que la explicación resulte más sencilla, se describen ejemplos en el contexto de los controles de gama dinámica (DRC) en el que se reduce la gama dinámica original de una señal de audio. Es decir, las partes del audio con sonidos fuertes se suavizan y/o se aumenta el volumen de las partes con sonidos más bajos. Dicho procesamiento resulta útil, por ejemplo, cuando se desea escuchar una selección de audio, bien entrada la noche y a un volumen promedio reducido.

La figura 3a ilustra un típico perfil de entrada/salida o función de transferencia para un DRC en el que el eje horizontal es una representación logarítmica del nivel de señal de entrada, y el eje vertical es una representación logarítmica del nivel de señal de salida deseado. El nivel de señal representa genéricamente una cualquiera de diversas medidas posibles; por ejemplo, el nivel r.m.s. o el nivel de sonoridad psicoacústica. En la zona de «banda nula» del perfil o función de transferencia, la pendiente de la curva es uno, lo que indica que no se realiza ninguna modificación de la señal. Por encima y por debajo de la banda nula, la pendiente es inferior a uno, lo que indica que se aplica la compresión - por encima de la banda nula, el audio se atenúa y, por debajo de la banda nula, el audio se potencia. La figura 3b ilustra una representación equivalente del perfil o función de transferencia de entrada/salida de la figura 3a que sustituye al nivel de señal de salida en el eje vertical por el logaritmo de la ganancia, que, al ser aplicado al nivel de señal de entrada, da lugar al nivel de señal de salida deseado. En este contexto, se entiende que el término «ganancia» representa genéricamente una modificación que potencia o atenúa la señal. Dicha «ganancia» puede ser un valor que se multiplique directamente con la señal o podría ser algo más sofisticado, como por ejemplo un escalado de la sonoridad perceptual. Debido a la aplicación de restricciones a dicha ganancia, es esta interpretación alternativa del perfil de dinámica o función de transferencia (concretamente, la representación de la figura 3b) la que se utiliza para el resto de este ejemplo.

Restricciones para promediar

Por lo general, el nivel de señal de entrada L en un procesador de dinámica de audio es una medida variable en el tiempo del nivel de señal que utiliza un nivel de señal r.m.s. constante en el tiempo integrado a lo largo de un periodo de unos milisegundos, por ejemplo. Con el objeto de imponer restricciones, también se puede computar un nivel de señal correspondiente \bar{L} a largo plazo, o promedio. \bar{L} puede ser la misma medida básica (por ejemplo: r.m.s., pico, etc.) que L , pero con una constante de tiempo considerablemente más larga, del orden de un segundo o más, por ejemplo.

Aunque las constantes de tiempo para L y \bar{L} no son de una importancia fundamental, se puede observar que la

5 constante de tiempo para \bar{L} es mayor que la de L. En el caso del control de gama dinámica (DRC), la constante a corto plazo puede guardar proporción con la resolución temporal de la percepción humana de la sonoridad (el intervalo de tiempo más corto a lo largo del cual una persona puede percibir variaciones en la sonoridad), y a largo plazo puede guardar proporción con un periodo de tiempo suficiente para que una persona integre la sonoridad «a corto plazo» en una impresión psicoacústica de la sonoridad de una fuente de sonido.

10 En una jerarquía de ruta de control de tres niveles, como la del ejemplo de la figura 2a, se puede computar una medida del nivel de señal tanto a corto plazo como promedio (a largo plazo) en los respectivos controles de dinámica 16, 20 y 26 para la totalidad de los canales audio combinados, para cada canal individual, y para cada una de una pluralidad de bandas de frecuencia en cada canal. Dichas medidas del nivel de señal se pueden denominar L_T y \bar{L}_T , L_C y \bar{L}_C , y L_B y \bar{L}_B , respectivamente, y pueden representar valores logarítmicos, como, por ejemplo, los que se muestran en las figuras 3a y 3b.

15 Con respecto a un esquema de ruta de control de tres capas jerárquicas, como la del ejemplo de la figura 2a, se puede comenzar la aplicación de restricciones especificando el perfil de procesamiento de dinámica que se desee, como el que se muestra en la figura 3b, para el nivel de jerarquía total. Dicho perfil se puede representar genéricamente como una función $F_T\{L\}$ que relacione la ganancia G con el nivel de señal L en el nivel jerárquico total. De este modo:

20
$$G = F_T\{L\} \quad (1a)$$

Se puede considerar que cada medida de nivel de señal L, independientemente de cómo se mida (por ejemplo, nns, pico, etc.), tiene asociada una ganancia teórica G computada a partir de la función de procesamiento de dinámica F_T asociada. De este modo, se pueden computar las ganancias G_T y \bar{G}_T a partir de L_T y \bar{L}_T de acuerdo con:

25
$$G_T = F_T\{L_T\} \quad (1b)$$

$$\bar{G}_T = F_T\{\bar{L}_T\} \quad (1c)$$

A continuación, para cada canal del nivel jerárquico de canales de la ruta de control, se puede obtener una función de procesamiento de dinámica $F_C\{L\}$ que relacione la ganancia G con el nivel de señal L en el nivel jerárquico de los canales, a partir de la F_T restringida, de manera que $\bar{G}_C = \bar{G}_T$ (donde G_C y \bar{G}_C son las ganancias teóricas asociadas con L_C y \bar{L}_C , respectivamente). Se puede obtener la función de dinámica $F_C\{L\}$ con las restricciones deseadas mediante el cambio del argumento de entrada de F_T por la diferencia, en caso de que existiera, y dependiendo de las condiciones de la señal, entre los niveles de señal promedio \bar{L}_T y \bar{L}_C (véase la figura 4b):

35
$$F_C\{L\} = F_T\{(\bar{L}_T - \bar{L}_C) + L\} \quad (2a)$$

Dependiendo de las condiciones de la señal, el efecto puede consistir en un desplazamiento de la función F_T en los ejes ganancia/nivel, conservando la forma del perfil de procesamiento de dinámica deseado, pero moviendo todos los umbrales horizontalmente la cantidad $\bar{L}_C - \bar{L}_T$, de tal modo que la relación entre el nivel de señal de canal promedio \bar{L}_C y los diversos umbrales relacionados con la función de canal F_C sea la misma que la relación entre el nivel de señal total \bar{L}_T y los umbrales relacionados con la función total F_T . Dicho desplazamiento se muestra en las figuras 4a-c, descritas más adelante. La ecuación 2a indica que la información que define F_T y \bar{L}_T se lleva desde el nivel jerárquico total hasta el nivel jerárquico de canales.

45 A partir de la función de procesamiento de dinámica de canal F_C obtenida, se pueden computar las ganancias G_C y \bar{G}_C :

$$G_C = F_C\{L_C\} \quad (2b)$$

$$\bar{G}_C = F_C\{\bar{L}_C\} \quad (2c)$$

A partir de las anteriores ecuaciones, se puede observar que:

$$\bar{G}_C = F_C \{ \bar{L}_C \} = F_T \{ (\bar{L}_T - \bar{L}_C) + \bar{L}_C \} = F_T \{ \bar{L}_T \} = \bar{G}_T \quad (2d)$$

A la hora de construir la función de dinámica de canal F_C a partir de la función de dinámica total F_T de manera que $\bar{G}_C = \bar{G}_T$, el valor promedio de G_C a lo largo del tiempo es igual al valor medio de G_T a lo largo del tiempo. Es decir, la modificación efectuada en un canal es, en promedio, igual a la modificación deseada en la señal en total. No obstante, a corto plazo, \bar{L}_T y \bar{L}_C permanecen relativamente constantes, y G_C actúa como una función del nivel de señal de canal L_C para reducir o prevenir el bombeo entre canales.

A modo de ejemplo, consideremos de nuevo la señal de audio multicanal con música en todos los canales y, después, un breve segmento de diálogo a un volumen alto introducido en el canal central. Antes de introducir el diálogo, el nivel de señal total promedio \bar{L}_T queda en la banda nula de la función de dinámica F_T junto con el nivel de señal total a corto plazo L_T . En consecuencia, la función no requiere ninguna modificación. Lo mismo se puede decir para los niveles de señal \bar{L}_C y L_C con respecto a F_C en todos los canales. Cuando se introduce el diálogo, el nivel de señal a corto plazo L_C situado en el canal central aumenta considerablemente y supera el umbral de compresión de F_C , lo que da lugar a una atenuación. Sin embargo, el comportamiento de los niveles de señal en el resto de los canales permanece inalterado y no se aplica ninguna atenuación. De este modo, el bombeo entre canales se reduce o se evita, a costa de una alteración momentánea en la imagen espacial. Sin embargo, si después el diálogo a alto volumen continúa, el nivel de señal total promedio \bar{L}_T aumenta lentamente y supera el umbral de compresión. A su vez, esto da lugar a un lento aumento de la atenuación aplicada a todos los canales debido a las restricciones empleadas en la construcción de las funciones de canal F_C . Esto continúa hasta que la atenuación promedio aplicada a todos los canales alcanza el nivel de atenuación que se aplica al diálogo continuado en el canal central. Por lo tanto, un esquema de restricciones jerárquicas de puede proporcionar un equilibrio entre la reducción o la prevención del bombeo entre canales y la conservación de la imagen espacial.

Los mismos principios se aplican a las bandas individuales comprendidas en un canal. La función de procesamiento de dinámica F_B para una banda concreta se puede construir a partir de la función del canal del que forma parte, al igual que la función de canal se puede construir a partir de la función total. La función F_B se puede obtener desplazando el argumento de entrada de F_C una distancia correspondiente a la diferencia, en caso de haberla, y dependiendo de las condiciones de la señal, entre los niveles de señal promedio \bar{L}_C y \bar{L}_B (véase la figura 4c):

$$F_B \{ L \} = F_C \{ (\bar{L}_C - \bar{L}_B) + L \} \quad (3a)$$

La ecuación 3a indica que la información que define F_C y \bar{L}_C se hace pasar desde el nivel jerárquico de canales hasta el nivel jerárquico de canales.

Las ganancias de banda se computan de acuerdo con las siguientes fórmulas:

$$G_B = F_B \{ L_B \} \quad (3b)$$

$$\bar{G}_B = F_B \{ \bar{L}_B \} \quad (3c)$$

Esta construcción restringe $\bar{G}_B = \bar{G}_C$, lo cual se muestra a continuación:

$$\bar{G}_B = F_B \{ \bar{L}_B \} = F_C \{ (\bar{L}_C - \bar{L}_B) + \bar{L}_B \} = F_C \{ \bar{L}_C \} = \bar{G}_C \quad (3d)$$

Debido a que $\bar{G}_B = \bar{G}_C$, el valor promedio de la ganancia de banda G_B a lo largo del tiempo es aproximadamente igual al de G_C , pero, a corto plazo, la ganancia G_B varía en función del nivel de señal de banda L_B , para reducir o evitar el bombeo a través del espectro. Las figuras 4a-c ilustran el proceso descrito de construcción de F_C como una versión desplazada de F_T y F_B como una versión desplazada de F_C , de tal modo que $\bar{G}_B = \bar{G}_C = \bar{G}_B$. El cambio de nivel desde el nivel jerárquico total al nivel jerárquico de canales se muestra en la figura 4b. El cambio desde el nivel jerárquico de canales al nivel jerárquico de bandas se muestra en la figura 4c. Las figuras 4a-c se pueden contemplar en conjunto para observar el modo en que el perfil deseado de procesamiento de dinámica se hace

pasar por la jerarquía en sentido descendente desde el nivel total hasta el nivel de bandas.

Restricciones de gama

5 Idealmente, las restricciones empleadas en la construcción de F_B a partir de F_C y de F_C a partir de F_T garantizan que la modificación promedio aplicada a cada banda de un canal sea sustancialmente la misma que para todo el canal, y que la modificación promedio aplicada a cada canal sea sustancialmente la misma que para todos los canales combinados. Debido a que dichas restricciones no imponen límites en lo que respecta a la gama de modificaciones en torno a su valor promedio, ciertos casos pueden resultar problemáticos. Consideremos de nuevo la señal de audio multicanal con música en todos los canales y, después, un breve segmento de diálogo a volumen alto introducido en el canal central. Cuando se introduce el diálogo, el nivel de señal total L_T aumenta una cierta cantidad y supera el nivel de señal total promedio \bar{L}_T , y la función de dinámica F_T requiere una atenuación general de la señal. En el canal central, el nivel de señal L_C también aumenta por encima del nivel de señal promedio \bar{L}_C , pero el aumento de L_C con respecto a \bar{L}_C es mucho mayor que el aumento de L_T con respecto a \bar{L}_T , ya que al valor de \bar{L}_T contribuyen todos los demás canales que contienen música. En consecuencia, la atenuación computada por F_C es considerablemente mayor que la atenuación general deseada, computada por F_T . Por lo tanto, para reducir dichos problemas, resulta deseable establecer restricciones en la gama de modificaciones además de en su comportamiento promedio.

15 Dichas restricciones de gama se podrían implementar de diversas maneras, pero una técnica sencilla y eficaz conlleva establecer restricciones en la ganancia de canal G_C para que quede entre las ganancias a corto plazo y promedio G_T y \bar{G}_T y seguir restringiendo la ganancia de banda G_B para que quede entre las ganancias a corto plazo y promedio G_C y \bar{G}_C . De este modo, cuando también se emplean restricciones de gama, la información que se hace pasar desde el nivel jerárquico total hasta el nivel jerárquico de canales incluye no solo información que define F_T y \bar{L}_T , sino también información que define G_T y \bar{G}_T , y la información que se hace pasar desde el nivel jerárquico de canales hasta el nivel jerárquico de bandas incluye no solo información que define F_C y \bar{L}_C , sino también información que define G_C y \bar{G}_C . Si \bar{G}_C y \bar{G}_B representan estas ganancias de canal y banda con restricción de gama, las restricciones se pueden representar formalmente por escrito según la siguiente lógica:

- 30 I) Restricción de total a canal:
- A) If $G_T \leq \bar{G}_T$ then
 - i) If $G_C < G_T$ then $G'_C = G_T$
 - ii) Else if $G_C > \bar{G}_T$ then $G'_C = \bar{G}_T$
 - iii) Else $G'_C = G_C$
 - 35 B) Else if $G_T > \bar{G}_T$ then
 - i) If $G_C > G_T$ then $G'_C = G_T$
 - ii) Else if $G_C < \bar{G}_T$ then $G'_C = \bar{G}_T$
 - iii) Else $G'_C = G_C$
- II) Restricción de canal a banda:
- 40 A) If $G'_C \leq \bar{G}_C$ then
 - i) If $G_B < G'_C$ then $G'_B = G'_C$
 - ii) Else if $G_B > \bar{G}_C$ then $G'_B = \bar{G}_C$
 - iii) Else $G'_B = G_B$
 - 45 B) Else if $G'_C > \bar{G}_C$ then
 - i) If $G_B > G'_C$ then $G'_B = G'_C$
 - ii) Else if $G_B < \bar{G}_C$ then $G'_B = \bar{G}_C$
 - iii) Else $G'_B = G_B$

Las figuras 5a y 5b ilustran dos ejemplos de las restricciones de gama desde el nivel total al de canal y al de banda. En cada caso, la zona de restricción desde un nivel actual de la jerarquía se lleva en sentido descendente hasta el siguiente nivel. En la fila superior, se observa que en el nivel jerárquico de canales, se impone la restricción I.A.i de la lógica anterior. En el nivel jerárquico de bandas, se impone entonces la restricción II.A.ii. En la fila inferior, se impone la restricción I.A.iii en el nivel jerárquico de canales, y se impone la restricción II.A.i en el nivel jerárquico de bandas.

Restricciones «ascendentes»

En los anteriores ejemplos, se han impuesto restricciones «descendentes», pasando desde los niveles jerárquicos superiores a los inferiores (desde «total» a «canales» y de ahí a «bandas» en el ejemplo de las figuras 2a y 2b). También puede resultar beneficiosa la implementación de otro aspecto de la invención en el que se implementen

unas restricciones secundarias «ascendentes» junto con las restricciones principales descendentes.

De acuerdo con la invención, la figura 6 ilustra una vista general, en el contexto del ejemplo de la figura 2a de la aplicación de restricciones jerárquicas «ascendentes» además de las restricciones jerárquicas descendentes de la figura 2a. Durante la aplicación de restricciones descendentes en el ejemplo de la figura 2a, cada uno de los controles de dinámica produce uno o más parámetros restringidos, representados por las líneas discontinuas sencillas de la figura 6.

Por lo que respecta al ejemplo de la figura 2a, la señal de entrada aplicada a la ruta de control jerárquico de la figura 6 se puede obtener a partir de una señal de audio aplicada a un procesador de dinámica de audio del cual forme parte la ruta de control de la figura 6 o, como otra posibilidad, a partir del audio procesado (salida) de dicho procesamiento de la dinámica del audio o a partir de una combinación de dichas señales de audio de entrada y salida.

Para comenzar el proceso de aplicación de restricciones ascendentes, se pueden analizar en primer lugar los parámetros descendentes en un dispositivo o proceso analizador («analizar») 27 en todas las bandas de todos los canales para generar una o más restricciones jerárquicas de banda secundarias que se pueden volver a llevar a todos los controles de dinámica en el nivel de bandas y, opcionalmente, en sentido ascendente en la jerarquía hasta todos los canales. Como se explica más adelante, «analizar» 27 puede computar, por ejemplo, un promedio o un mínimo de todas sus entradas. En la figura 6, se muestran dichas una o más restricciones de banda secundarias indicadas mediante las líneas discontinuas dobles que parten desde «analizar» 27 en el nivel jerárquico de bandas. Dichas una o más restricciones de banda secundarias se aplican a cada control de dinámica 26'-1-1 hasta 26'-c-n del nivel jerárquico de bandas y son utilizadas por estos para aumentar las restricciones en los parámetros de control de dinámica descendentes de la figura 2a. Dichos uno o más parámetros de control de dinámica doblemente restringidos (es decir, restringidos en sentido descendente y restringidos en sentido ascendente) se representan mediante las líneas discontinuas dobles que parten de la salida de cada control de dinámica de banda 26'-1-1 hasta 26'-c-n. Es decir, los parámetros de control de dinámica doblemente restringidos producidos por cada uno de los controles de dinámica 26'-1-1 hasta 26'-c-n, que se muestran con líneas discontinuas dobles, son el resultado de los parámetros de modificación restringidos en sentido descendente producidos por los controles de dinámica 26-1-1 hasta 26 de la figura 2a, restringidos también por las una o más restricciones secundarias ascendentes procedentes de la salida de «analizar» 27.

Opcionalmente, las restricciones secundarias procedentes de la salida de «analizar» 27 también se hacen pasar en sentido ascendente por la jerarquía hasta el nivel jerárquico de canales. En ese caso, en el nivel jerárquico de canales, los parámetros de control de dinámica restringidos en sentido descendente de todos los canales, representados por las líneas discontinuas simples, son analizados por un dispositivo o función de análisis («analizar») 21 junto con las una o más restricciones de banda secundarias recibidas desde «analizar» 27 para generar parámetros de modificación doblemente restringidos, representados por las líneas discontinuas dobles que salen desde «analizar» 21 en el nivel jerárquico de canales. Dicha restricción secundaria del nivel jerárquico de canales es utilizada por cada uno de los controles de dinámica 20-1 hasta 20-c en el nivel jerárquico de canal para producir uno o más parámetros de control de dinámica doblemente restringidos, representados mediante las líneas discontinuas dobles que parten desde la salida de cada uno de los controles de dinámica 20' en el nivel jerárquico de canales.

Opcionalmente, las restricciones secundarias procedentes de la salida de «analizar» 21 también se hacen pasar en sentido ascendente por la jerarquía hasta el nivel jerárquico total. En ese caso, los parámetros de dinámica restringidos en sentido descendente para el nivel jerárquico total son analizados por un dispositivo o proceso de análisis («analizar») 17 junto con las restricciones de canal secundarias para generar una restricción total secundaria, que a su vez se usa para crear uno o más parámetros de control de dinámica doblemente restringidos en el nivel jerárquico total.

Tras la aplicación de dichas restricciones secundarias ascendentes, se pueden usar los parámetros de modificación doblemente restringidos en la salida de algunos de los controles de dinámica para modificar el audio. Por ejemplo, los parámetros obtenidos en las salidas de los controles de dinámica 26'-1-1 hasta 26'-c-n en el nivel jerárquico de bandas se pueden usar para modificar bandas individuales dentro de cada canal de la señal de audio. Otra posibilidad consiste en usar los parámetros en el nivel jerárquico de canales para modificar cada canal de manera uniforme a lo largo de bandas, en usar los parámetros en el nivel jerárquico total para modificar todos los canales de manera uniforme, o en usar alguna combinación de parámetros en varios niveles jerárquicos, o funciones de los mismos, para modificar la señal de audio.

Dichas restricciones ascendentes secundarias resultan particularmente útiles cuando el sistema de procesamiento de dinámica se encuentra en modo de «liberación» (release), lo cual significa que el nivel de señal total a corto plazo L_T es menor que el nivel de señal total promedio \bar{L}_T . Para entender por qué son necesarias dichas restricciones, consideremos de nuevo la señal de audio multicanal con música en todos los canales y diálogo en el canal central. Supongamos que el diálogo a alto volumen ha permanecido constante en el canal central durante un periodo de

tiempo considerable. Por lo tanto, las restricciones descendentes han convergido en un comportamiento de estado estacionario, por lo que todos los canales reciben aproximadamente la misma cantidad de atenuación. Supongamos ahora que el diálogo se interrumpe repentinamente, dejando únicamente la señal musical, que tiene un nivel más bajo, en todos los canales. El nivel de señal total L_T cae inmediatamente por debajo del nivel de señal total promedio

5 \bar{L}_T y entra en la banda nula de la función de dinámica F_T , lo cual quiere decir que G_T no requiere ninguna modificación, en comparación con la atenuación previa. Lo mismo se puede decir de G'_C en el canal central en el que estaba presente el diálogo. Sin embargo, en el resto de canales, el comportamiento de L_C con respecto a \bar{L}_C permanece inalterado, lo cual quiere decir que la atenuación previa continúa. Por lo tanto, el canal central se ve ahora potenciado con respecto al resto de los canales, lo que distorsiona la imagen espacial de la música. Esta

10 distorsión continúa hasta que el nivel de señal total promedio \bar{L}_T disminuye hasta el nivel promedio de la música, quedando en la banda nula; en ese momento, las restricciones descendentes dan lugar a que no se aplique ninguna modificación al resto de los canales. Por lo tanto, se distorsiona la imagen espacial de la música durante un periodo de tiempo considerable a medida que \bar{L}_T disminuye lentamente.

15 Para hacer frente a este problema, se podría aplicar una restricción ascendente secundaria cuando $L_T < \bar{L}_T$. Para el problema concreto que se acaba de describir, se podrían analizar los valores de G'_C en todos los canales y restringir estos valores junto con G_T para que todos sean iguales al valor mínimo de G'_C en todos los canales. Así se evita que algún canal se potencie más que cualquier otro. El mismo principio se puede extender a las bandas. Si $L_T < \bar{L}_T$, los valores de G'_B de todas las bandas y canales se pueden restringir para que sean iguales al mínimo de este conjunto

20 de valores. A continuación, este mínimo se hace pasar de nuevo hacia arriba hasta los canales, en los que se establece un valor de G'_C para todos los canales igual al mínimo del conjunto que incluye G'_C en todos los canales junto con el valor mínimo de la restricción de banda. Este mínimo de canal puede volver a llevarse finalmente hasta el nivel total, en el que se establece un valor de G_T igual al valor mínimo de G'_C y a dicho valor mínimo de canal.

25 Suavizado en el tiempo

La aplicación de las restricciones jerárquicas descritas, tanto descendentes como ascendentes, pueden introducir a lo largo del tiempo discontinuidades abruptas en las trayectorias de las ganancias. Si se dejan sin modificar, dichas discontinuidades pueden provocar la aparición de artefactos perceptibles en el audio procesado. Por lo tanto, en una

30 forma de realización práctica de la invención, es deseable suavizar las ganancias restringidas a lo largo del tiempo antes de su usarlas para modificar el audio. Dicho suavizado se puede llevar a cabo de diversas maneras, y más adelante se explicará uno de estos métodos.

35 **Breve descripción de los dibujos**

La figura 1 es un diagrama de bloques de alto nivel de un procesador de dinámica de audio genérico de la técnica anterior, que procesa un único canal de una señal de audio multicanal o una señal de audio que solo presenta un canal.

40 La figura 2a es una vista general esquemática de un ejemplo de arquitectura de ruta de control jerárquico en el que la información se hace pasar en sentido descendente por una jerarquía de ruta de control desde arriba hacia abajo (en un esquema jerárquico «descendente»).

45 La figura 2b es una vista general esquemática de un ejemplo de la ruta de audio en el que una señal de audio se puede modificar en función de la información obtenida en una ruta de control como la de la figura 2a y/o la de la figura 2a y la figura 6.

50 La figura 3a es un ejemplo de perfil o función de transferencia de entrada/salida para un control de gama dinámica (DRC) en el que el eje horizontal es una representación logarítmica del nivel de señal de entrada, y el eje vertical es una representación logarítmica del nivel de señal de salida deseado.

55 La figura 3b es una representación equivalente del perfil o función de transferencia de entrada/salida de la figura 3a en la que se sustituye el nivel de señal de salida en el eje vertical por el logaritmo de la ganancia, que al aplicarlo al nivel de señal de entrada, da lugar al nivel de señal de salida deseado.

60 Las figuras 4a-c son ejemplos que resultan útiles para comprender algunos aspectos de la invención, en los que se muestran perfiles o funciones de transferencia de entrada/salida al modo de la figura 3b, y en los que se reconstruyen versiones de un perfil inicial de un nivel jerárquico superior inicial en niveles jerárquicos más bajos, teniendo en cuenta ciertas restricciones de promedio.

Las figuras 5a y 5b son ejemplos que resultan útiles para comprender algunos aspectos de la invención, en los que se muestran perfiles o funciones de transferencia de entrada/salida al modo de la figura 3b, y en los que se reconstruyen versiones de un perfil inicial de un nivel jerárquico superior inicial en niveles jerárquicos más bajos, teniendo en cuenta ciertas restricciones de gama.

La figura 6 es una vista general esquemática de un ejemplo de arquitectura de ruta de control jerárquico de acuerdo con otros aspectos de la invención en el que la información se hace pasar en sentido ascendente por una jerarquía de ruta de control desde abajo hacia arriba (en un esquema jerárquico «ascendente»), además de la información jerárquica descendente del ejemplo de la figura 2a.

La figura 7 ilustra un conjunto adecuado de respuestas del filtro de banda crítica que contiene 40 bandas espaciadas uniformemente a lo largo de la escala de ancho de banda rectangular equivalente (ERB), según la definición de Moore y Glasberg.

La figura 8 es un conjunto de curvas isofónicas.

Mejor modo para llevar a cabo la invención

La citada solicitud internacional de patente PCT/US2005/038579 a nombre de Alan Jeffrey Seefeldt y publicada con el número WO 2006/047600 expone, entre otros aspectos, esquemas tanto para medir como para modificar la sonoridad percibida y/o el balance espectral percibido de una señal de audio mediante el uso de un modelo psicoacústico de percepción de la sonoridad. La exposición discute el uso de tales esquemas para el procesamiento de la dinámica del audio y cita varias ventajas para llevar a cabo dicho procesamiento en el dominio de la sonoridad perceptual. No obstante, los esquemas de Seefeldt se pueden mejorar mediante diversos aspectos de la presente invención. Por lo tanto, a continuación se describe una forma de realización preferida de la presente invención en el contexto de un modelo psicoacústico de sonoridad.

Como primera etapa, se computa la sonoridad específica, una medida de la sonoridad perceptual distribuida a lo largo de la frecuencia y el tiempo, de una señal de audio digital multicanal $x_c[n]$, $c = 1 \dots C$, a través de las siguientes etapas. En primer lugar, se computa una señal de excitación $E_c[b,t]$ para cada canal, aproximando la distribución de energía a lo largo de la membrana basilar del oído interno en una banda crítica b durante el bloque de tiempo t . esta excitación se puede computar a partir de la transformada discreta de Fourier de tiempo corto (STDFT) de la señal de audio, mediante la siguiente fórmula:

$$E_c[b,t] = \lambda_b E_c[b,t-1] + (1 - \lambda_b) \sum_k |T[k]|^2 |C_b[k]|^2 |X_c[k,t]|^2 \tag{4}$$

en la que $X_c[k,t]$ representa la STDFT de $X_c[n]$ en el bloque de tiempo t y el bin k . $T[k]$ representa la respuesta en frecuencia de un filtro que simula la transmisión de audio a través del oído externo y medio, y $C_b[k]$ representa la respuesta en frecuencia de la membrana basilar en un punto correspondiente a la banda crítica b . La figura 7 ilustra un conjunto adecuado de respuestas de filtro de banda crítica que contiene 40 bandas espaciadas uniformemente a lo largo de la escala de ancho de banda rectangular equivalente (ERB), según la definición de Moore y Glasberg (B.C.J. Moore, M. Glasberg, T. Baer, "A Model for the Prediction of Thresholds, Loudness, and Partial Loudness", Journal of the Audio Engineering Society, vol. 45, núm. 4, abril de 1997, pp. 224-240). Cada forma de filtro se describe por medio de una función exponencial redondeada y las bandas se distribuyen usando un espaciado de 1 ERB. La constante de tiempo λ_b del suavizado de la ecuación 4 se puede escoger de manera que, ventajosamente, guarde proporción con el tiempo de integración de la percepción humana de la sonoridad dentro de la banda b .

Además de la excitación por canal, se computa una excitación representativa de la señal de audio total mediante la suma de las excitaciones de canal de los canales:

$$E_T[b,t] = \sum_{c=1}^C E_c[b,t] \tag{5}$$

A continuación, usando curvas isofónicas, como las que se ilustran en la figura 8, las excitaciones total y de canal en cada banda se transforman en niveles de excitación que generarían la misma sonoridad a 1 kHz. Por último, se computan los espectros de la sonoridad específicos total y de canal a partir de estas excitaciones transformadas, $E'_c[b,t]$ y $E'_T[b,t]$, a través de una no linealidad compresiva. Si nos referimos de forma genérica a la función para computar la sonoridad específica N a partir de la excitación transformada E como $N = \Phi\{E\}$, una de dichas funciones adecuadas viene dada por la siguiente ecuación:

$$\Phi\{E\} = \beta \left(\left(\frac{E}{TQ_{1kHz}} \right)^\alpha - 1 \right) \tag{6}$$

en la que $T_{Q_{1\text{kHz}}}$ es el umbral en silencio a 1 kHz y las constantes β y α se escogen de manera que coincidan con el crecimiento publicado de datos de sonoridad. Mediante el uso de esta función, se computan los espectros de sonoridad específicos de acuerdo con las siguientes ecuaciones:

$$N_T[b, t] = \Psi\{E'_T[b, t]\} \quad (7a)$$

$$5 \quad N_c[b, t] = \Psi\{E'_c[b, t]\} \quad (7b)$$

A continuación se computan los valores de sonoridad total y sonoridad de canal mediante la suma de los espectros específicos de la sonoridad total y de canal a lo largo de las bandas b:

$$N_T[t] = \sum_b N_T[b, t] \quad (8a)$$

$$10 \quad N_c[t] = \sum_b N_c[b, t] \quad (8b)$$

Con el objeto de imponer restricciones en el procesamiento de la dinámica, se computan espectros específicos de la sonoridad promedio y valores de sonoridad promedio mediante el suavizado de los anteriores espectros específicos de la sonoridad y valores de sonoridad usando una constante de tiempo de suavizado $\bar{\lambda}$ grande:

$$15 \quad \bar{N}_T[t] = \bar{\lambda}\bar{N}_T[t-1] + (1-\bar{\lambda})N_T[t] \quad (9a)$$

$$\bar{N}_c[t] = \bar{\lambda}\bar{N}_c[t-1] + (1-\bar{\lambda})N_c[t] \quad (9b)$$

$$\bar{N}_c[b, t] = \bar{\lambda}\bar{N}_c[b, t-1] + (1-\bar{\lambda})N_c[b, t] \quad (9c)$$

20 Por último, los valores correspondientes a los términos genéricos L_T , \bar{L}_T , L_c , \bar{L}_c , L_B y \bar{L}_B usados en la descripción general de la invención se computan a partir de los valores de sonoridad y los espectros específicos de la sonoridad del siguiente modo:

$$L_T \rightarrow L_T[t] = \log(N_T[t]) \quad (10a)$$

$$\bar{L}_T \rightarrow \bar{L}_T[t] = \log(\bar{N}_T[t]) \quad (10b)$$

$$L_c \rightarrow L_c[t] = \log(N_c[t]) \quad (10c)$$

$$\bar{L}_c \rightarrow \bar{L}_c[t] = \log(\bar{N}_c[t]) \quad (10d)$$

$$L_B \rightarrow L_c[b, t] = \log(N_c[b, t]) \quad (10e)$$

$$\bar{L}_B \rightarrow \bar{L}_c[b, t] = \log(\bar{N}_c[b, t]) \quad (10f)$$

Comenzando con una función de dinámica total deseada F_T que toma como entrada el logaritmo de la sonoridad

perceptual y genera como salida el logaritmo del escalado de sonoridad perceptual deseado, se aplican las restricciones descendentes y ascendentes que se detallan en la descripción general de la invención para generar valores de escalado del logaritmo de la sonoridad total, de canal y de banda restringidos, representados por $G'_T[t]$, $G'_c[t]$, y $G'_c[b,t]$, respectivamente.

5 El escalado del logaritmo de la sonoridad de banda restringido $G'_c[b,t]$ se usa para modificar la señal de audio. No obstante, esta cantidad debe someterse primero a un suavizado en el tiempo para eliminar cualquier discontinuidad temporal introducida por la aplicación de las restricciones. El suavizado puede emplear, ventajosamente, una constante de tiempo rápida y lenta, y la decisión de qué constante de tiempo se debe usar se basa en el suavizado
10 simultáneo de $L_c[b,t]$ y $G'_c[b,t]$. Si $L_c[b,t]$ aumenta con respecto a su versión suavizada mientras que $G'_c[b,t]$ disminuye con respecto a su versión suavizada, entonces se usa la constante de tiempo rápida. En caso contrario, se usa la constante de tiempo lenta. Más concretamente, se computan versiones suavizadas temporalmente de $L_c[b,t]$ y $G'_c[b,t]$ mediante las siguientes ecuaciones:

$$\tilde{G}'_c[b,t] = \tilde{\lambda}_c[b,t]\tilde{G}'_c[b,t] + (1 - \tilde{\lambda}_c[b,t])G'_c[b,t] \quad (11a)$$

$$\tilde{L}_c[b,t] = \tilde{\lambda}_c[b,t]\tilde{L}_c[b,t] + (1 - \tilde{\lambda}_c[b,t])L_c[b,t] \quad (11b)$$

en las que

$$\tilde{\lambda}_c[b,t] = \begin{cases} \lambda_{rápida} & G'_c[b,t] < \tilde{G}'_c[b,t-1] \\ \lambda_{lenta} & \text{en caso contrario} \end{cases} \quad \text{y} \quad \begin{cases} \lambda_{rápida} & L_c[b,t] > \tilde{L}_c[b,t-1] \\ \lambda_{lenta} & \text{en caso contrario} \end{cases} \quad (11c)$$

20 En la siguiente etapa de la computación de la señal modificada, se computa un espectro deseado de sonoridad específica modificada en cada canal multiplicando la sonoridad específica original con el escalado de banda suavizado $G'_c[b,t]$:

$$\hat{N}_c[b,t] = \exp(\tilde{G}'_c[b,t])N[b,t] \quad (12)$$

Si la función $\Psi\{\cdot\}$ representa la transformación de excitación en sonoridad específica, se calcula una ganancia variable de banda $g_c[b,m]$ para cada canal, de manera que:

$$\hat{N}_c[b,t] = \Psi\{g_c^2[b,m]E_c[b,m]\} \quad (13a)$$

Reordenando (13a), se llega a la solución:

$$g_c[b,m] = \sqrt{\frac{\Psi^{-1}\{\hat{N}_c[b,m]\}}{E_c[b,m]}} \quad (13b)$$

35 En la citada solicitud internacional de patente PCT/US2005/038579, a nombre de Alan Jeffrey Seefeldt y publicada con el número WO 2006/47600 el 4 de mayo de 2006, se describen varias técnicas para computar la función inversa Ψ^{-1} en (13b), entre ellas: una expresión en forma cerrada, una tabla de consulta y una búsqueda iterativa.

40 Por último, la ganancia variable de banda $g_c[b,m]$ se aplica en cada banda crítica a los correspondientes bins de la STDFT original para generar una STDFT modificada:

$$\hat{X}_c[k,m] = \sum_b g_c[b,m]C_b[k]X_c[k,m] \quad (14)$$

45 La STDFT modificada se somete a una transformación inversa y a una operación de solapamiento y suma para producir la señal de audio digital multicanal modificada final $\hat{X}_C[n]$. En la patente de EE.UU. 5.899.969 a nombre de Fielder y col. ("Frame-based audio with gain-control words"), se exponen más detalles de un esquema de procesamiento de audio digital que emplea bloques de muestras solapados.

Aunque en la presente memoria descriptiva se expresan diversas restricciones jerárquicas (por ejemplo, $\bar{G}_B = \bar{G}_C = \bar{G}_D$) como igualdades, dichas igualdades son ejemplos idealizados. La invención contempla la imposición de restricciones que tienden a hacer que dichos valores estén más cerca de ser iguales de lo que lo estarían en el caso de que no existiera dicha imposición de restricciones. Asimismo, aunque el paso de los perfiles de procesamiento de dinámica desde un nivel jerárquico a otro se ilustran (por ejemplo, en la figura 3b) como el paso de la misma forma de perfil, dicha uniformidad de las formas no supone más que un ejemplo idealizado. La invención contempla el paso de perfiles de procesamiento de dinámica desde un nivel jerárquico a otro en el que el perfil que se hace pasar es una aproximación de uno o más perfiles de nivel más alto, y el grado de proximidad responde a una opción de implementación del creador del dispositivo o proceso de procesamiento de dinámica.

Implementación

La invención se puede implementar en hardware o software, o en una combinación de los dos (por ejemplo, matrices lógicas programables). A menos que se indique lo contrario, los algoritmos incluidos como parte de la invención no están relacionados inherentemente con ningún ordenador en concreto u otro aparato. En particular, se pueden emplear diversas máquinas de uso general con programas escritos de acuerdo con las indicaciones ofrecidas en la presente descripción, o es posible que resulte más conveniente construir más aparatos especializados (por ejemplo, circuitos integrados) para llevar a cabo las etapas necesarias del método. Por tanto, la invención se puede implementar en uno o más programas informáticos que se ejecuten en uno o más sistemas informáticos programables, y cada uno de ellos comprende al menos un procesador, al menos un sistema de almacenamiento de datos (incluida una memoria volátil o no volátil y/o elementos de almacenamiento), al menos un dispositivo o puerto de entrada, y al menos un dispositivo o puerto de salida. Se aplica un código de programa a los datos de entrada para llevar a cabo las funciones descritas en la presente memoria descriptiva y para generar información de salida. La información de salida se aplica a uno o más dispositivos de salida, de manera ya conocida.

Cada uno de dichos programas se puede implementar en cualquier lenguaje informático que se desee (incluidos: código máquina, ensamblador, o lenguajes de programación de alto nivel lógicos, orientados al método, u orientados a objetos) para comunicarse con un sistema informático. En cualquier caso, el lenguaje puede ser un lenguaje compilado o interpretado.

Cada uno de dichos programas informáticos se almacena o se descarga preferentemente en un medio o dispositivo de almacenamiento (por ejemplo, memoria o medios de estado sólido, o medios magnéticos u ópticos) que puede ser leído por un ordenador programable de uso general o especial, para configurar y manejar el ordenador cuando el medio o dispositivo de almacenamiento es leído por el sistema informático para llevar a cabo los procedimientos descritos en la presente memoria. También se puede considerar la implementación del sistema de la invención como un medio de almacenamiento legible por ordenador, configurado con un programa informático, en el que el medio de almacenamiento así configurado hace que un sistema informático se comporte de un modo concreto y predefinido para llevar a cabo las funciones descritas en la presente memoria.

Se han descrito varias formas de realización de la invención. No obstante, se entenderá que se pueden realizar diversas modificaciones. Por ejemplo, algunas de las etapas descritas en la presente memoria pueden ser independientes del orden, y, por tanto, se pueden llevar a cabo en un orden diferente al descrito.

REIVINDICACIONES

1. Un método para el procesamiento de la dinámica del audio, que comprende la modificación de la dinámica de una señal de audio en respuesta a la información obtenida de uno o más dispositivos o procesos que operan en un nodo de cada uno de dos niveles de jerarquía, presentando cada nivel uno o más nodos, en el que el uno o más dispositivos o procesos que actúan en cada nivel jerárquico toma una medida de una o más características de la señal de audio de manera que el uno o más dispositivos o procesos que actúan en un nivel jerárquico más bajo toma una medida de una o más características de una subdivisión más pequeña de la señal de audio, en el que dichos uno o más dispositivos o uno o más procesos en cada nodo de cada nivel toman una medida del audio en su nivel jerárquico y obtienen información a partir de dicha medida de la señal de audio y a partir de la información que reciben procedente de uno o más dispositivos o procesos en el otro nivel jerárquico, y en el que dicha subdivisión más pequeña del audio incluye una de (1) subdivisiones de canal y (2) subdivisiones de banda de frecuencia.
2. Un método según la reivindicación 1, en el que el otro nivel jerárquico comprende un nodo raíz que toma una medida de una totalidad de la señal de audio.
3. Un método según una cualquiera de las reivindicaciones 1-2, en el que dicha señal de audio es una señal de audio multicanal y el otro nivel jerárquico comprende una pluralidad de nodos que toman cada uno una medida de una subdivisión de canal de la señal de audio, y el nivel jerárquico más bajo comprende una pluralidad de nodos que toman cada uno una medida de una subdivisión de frecuencia de los canales de la señal de audio.
4. Un método según una cualquiera de las reivindicaciones 1-2, en el que dicha señal de audio es una señal de audio de un único canal y el nivel jerárquico más bajo comprende una pluralidad de nodos que toman cada uno una medida de una subdivisión de frecuencia de la señal de audio.
5. Un método según una cualquiera de las reivindicaciones 1-4, en el que la información obtenida a partir de una medida de la señal de audio en el otro nivel jerárquico se hace pasar en sentido descendente en la jerarquía hasta el nivel jerárquico más bajo con el fin de influir en el funcionamiento de los dispositivos o procesos que actúan en un nivel jerárquico más bajo.
6. Un método según la reivindicación 5, en el que la información obtenida a partir de una medida de la información de audio en el otro nivel jerárquico comprende no sólo información útil para influir en el funcionamiento de los dispositivos y procesos que actúan en el nivel jerárquico más bajo, sino también información útil para modificar la dinámica de la señal de audio.
7. Un método según la reivindicación 5 o la reivindicación 6, en el que la información obtenida a partir de una medida de la señal de audio en el nivel jerárquico más bajo y a partir de la información que pasa al nivel jerárquico más bajo se usa para influir en el funcionamiento de los dispositivos o procesos que actúan en el nivel jerárquico más bajo.
8. Un método según la reivindicación 7, en el que la influencia en el funcionamiento de los dispositivos o procesos que actúan en el nivel jerárquico más bajo da lugar a que el nivel jerárquico más bajo genere una modificación de la información que generaría en caso de no ejercerse dicha influencia.
9. Un método según la reivindicación 5 o la reivindicación 6, en el que la información obtenida a partir de una medida de la señal de audio en el nivel jerárquico más bajo y a partir de la información que pasa al nivel jerárquico más bajo se hace pasar en sentido ascendente en la jerarquía hasta el otro nivel jerárquico con el fin de influir en el funcionamiento de los dispositivos o procesos que actúan en el otro nivel jerárquico.
10. Un método según la reivindicación 9, en el que la influencia en el funcionamiento de los dispositivos o procesos que actúan en el otro nivel jerárquico da lugar a que el otro nivel jerárquico genere una modificación de la información que, en caso contrario, haría pasar al nivel jerárquico más bajo o que se usaría para modificar la señal de audio.
11. Un método según la reivindicación anterior, en el que la información para modificar la dinámica de una señal de audio es suavizada en el tiempo.
12. Aparato adaptado para llevar a cabo los métodos de una cualquiera de las reivindicaciones 1 a 11.
13. Un programa informático, almacenado en un medio legible por ordenador para dar lugar a que un ordenador lleve a cabo los métodos de una cualquiera de las reivindicaciones 1 a 11.

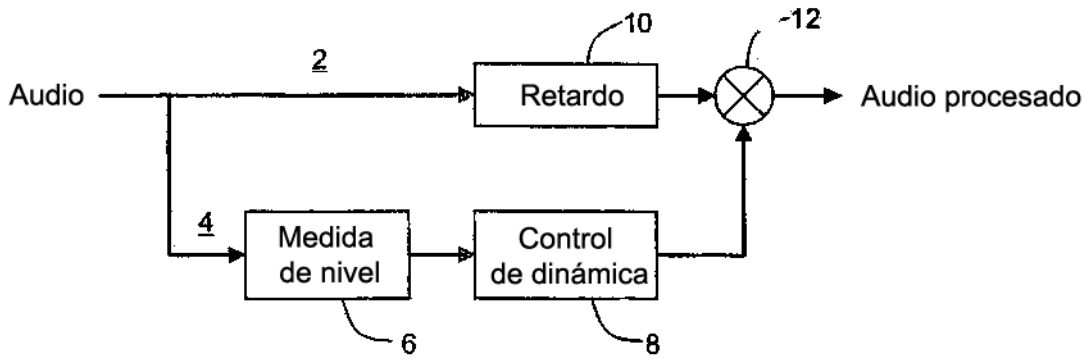


FIG. 1

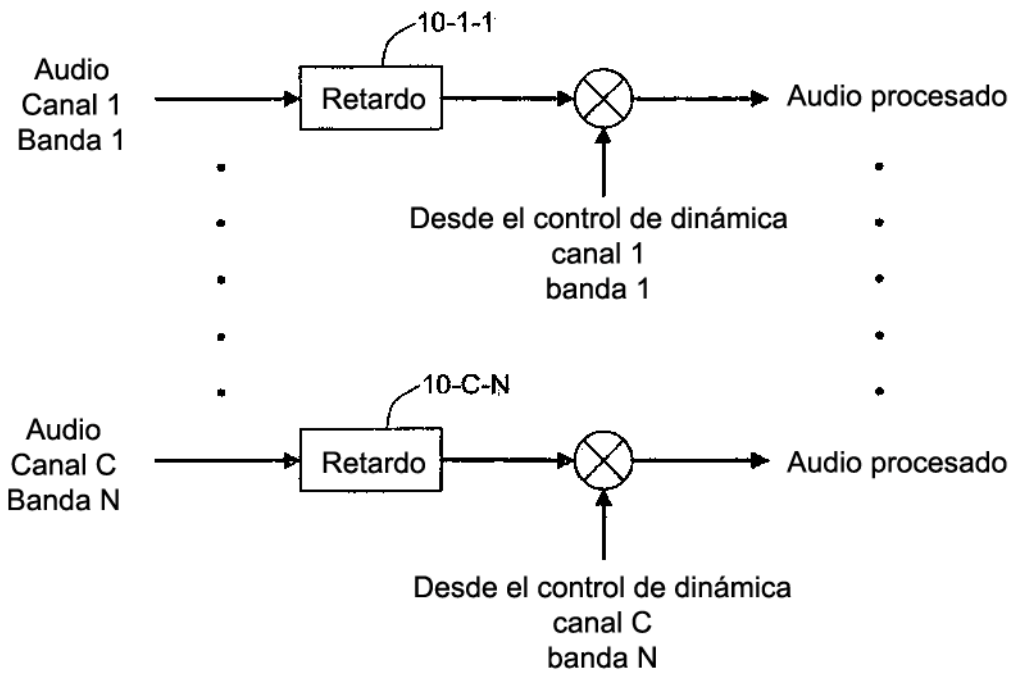


FIG. 2b

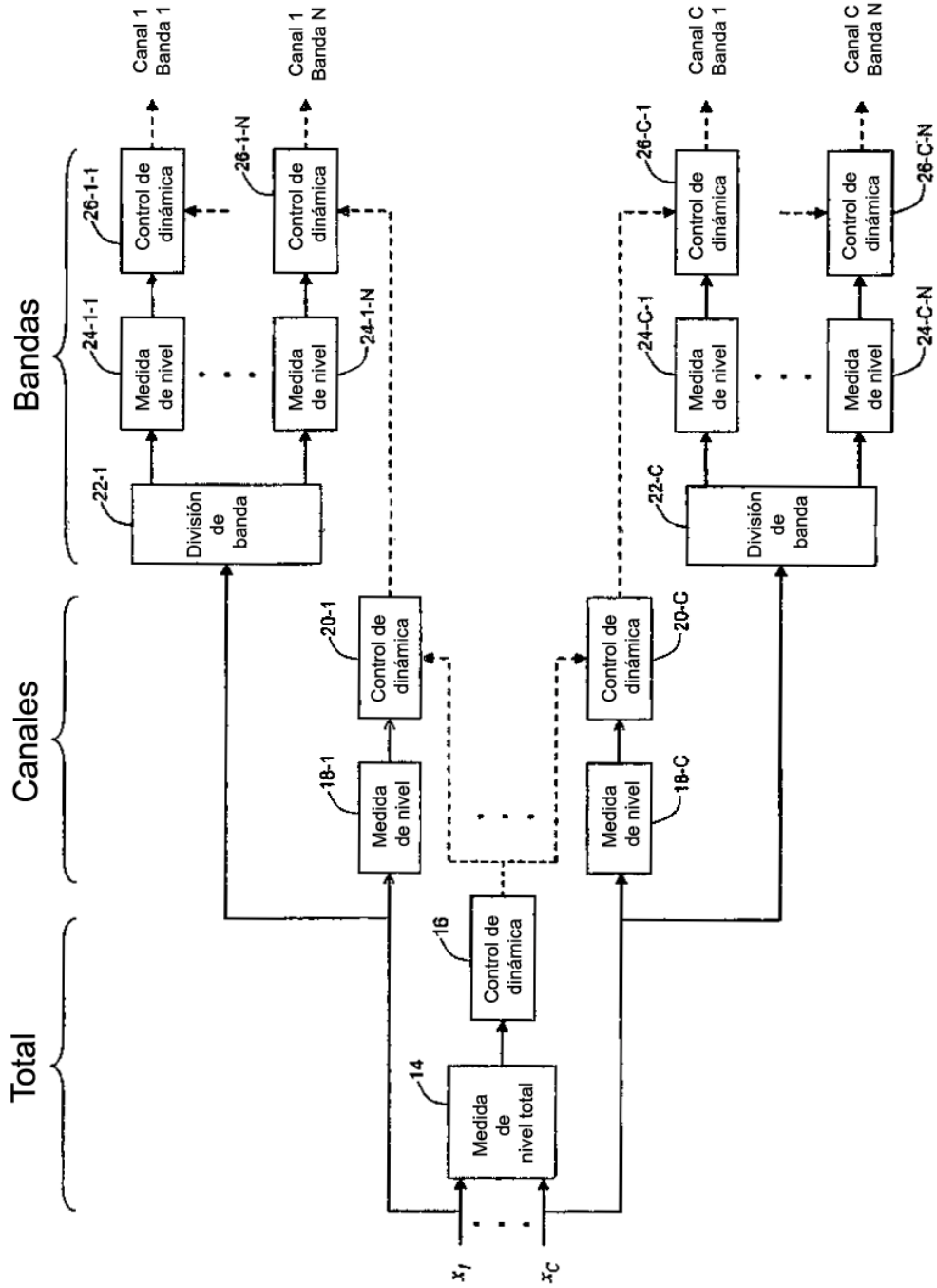


FIG. 2a

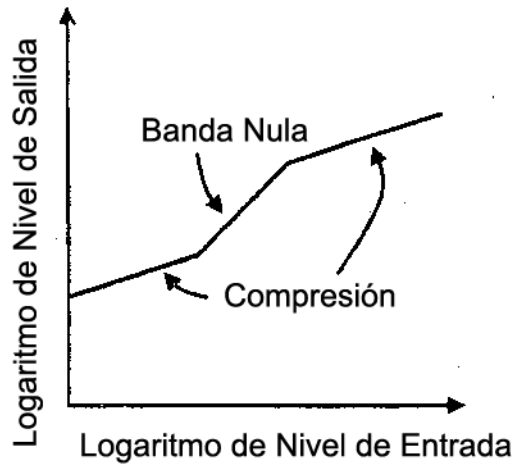


FIG. 3a

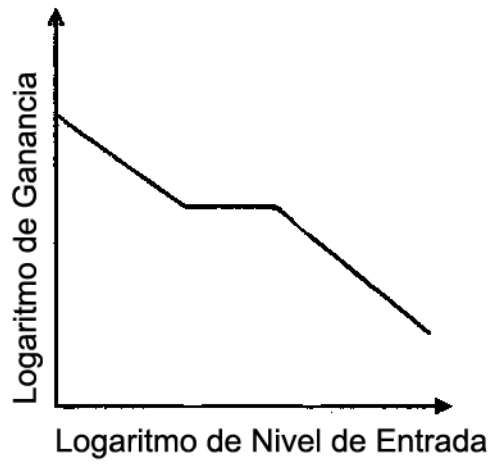


FIG. 3b

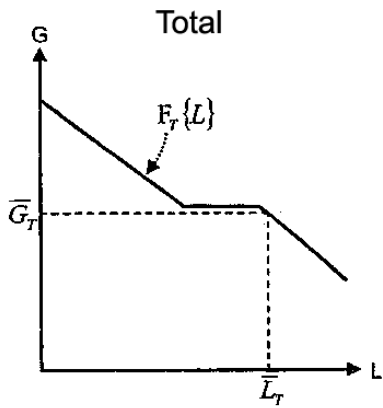


FIG. 4a

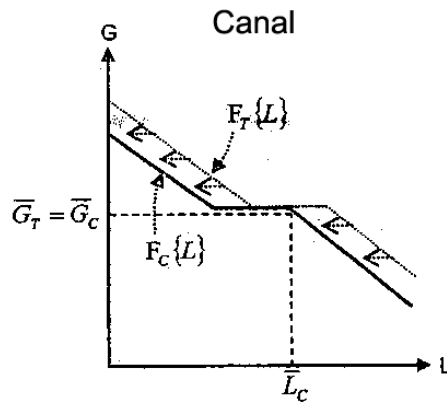


FIG. 4b

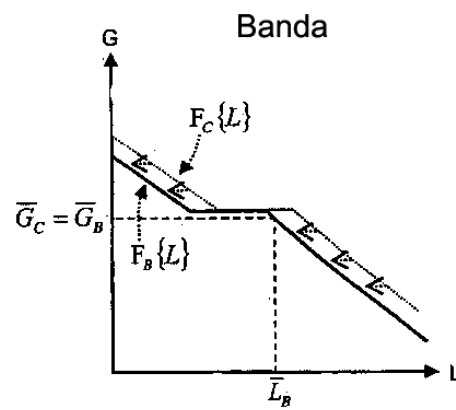


FIG. 4c

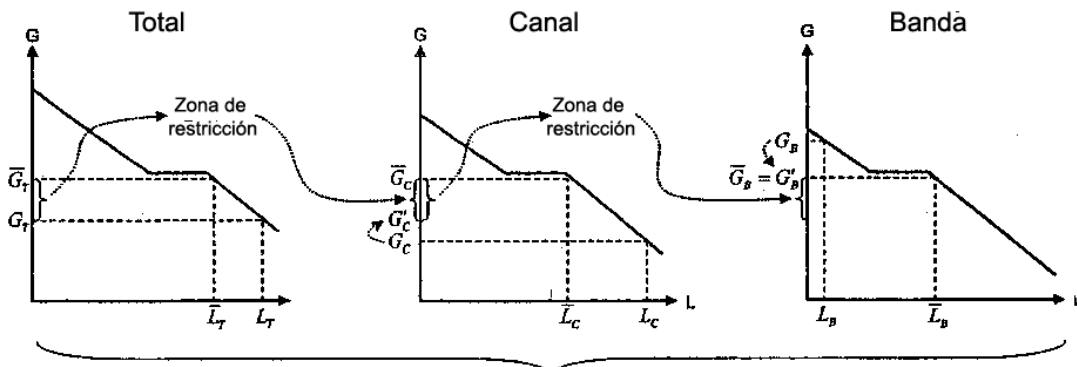


FIG. 5a

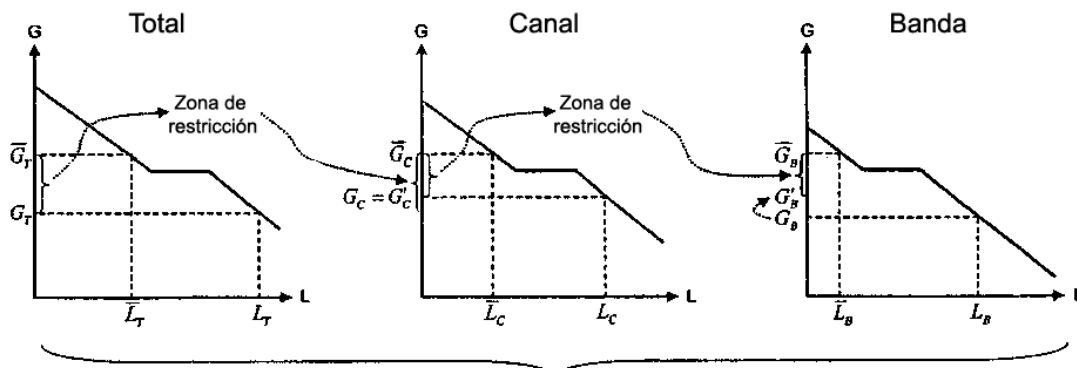


FIG. 5b

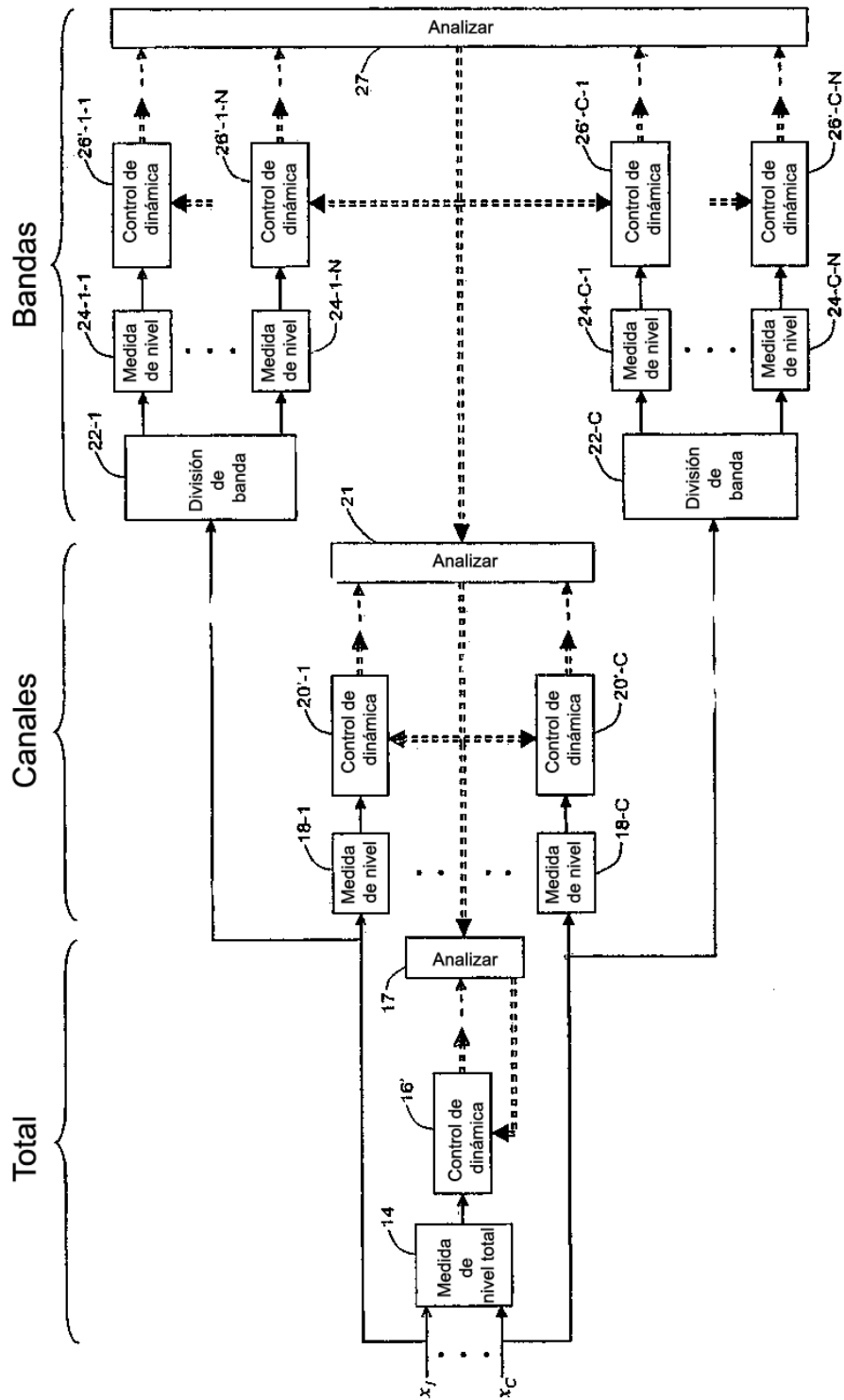


FIG. 6

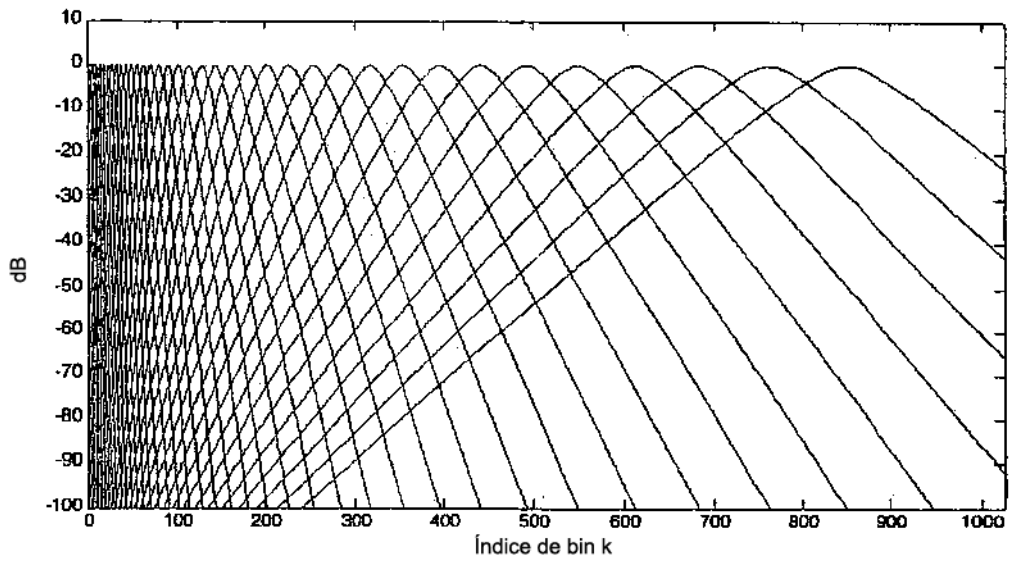


FIG. 7

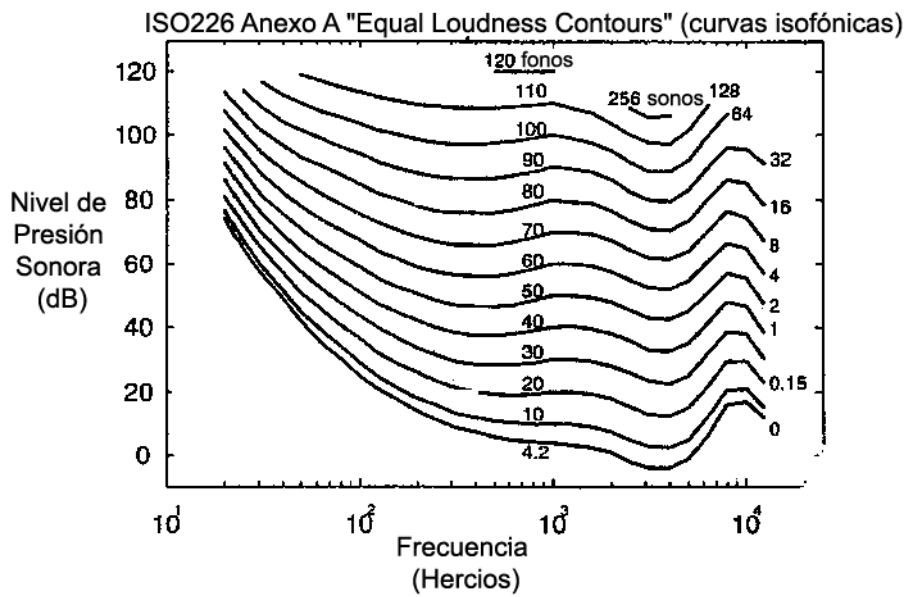


FIG. 8