



OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA



11) Número de publicación: 2 400 160

51 Int. Cl.:

G10L 19/00 (2006.01)
G10L 21/02 (2006.01)
H03G 3/32 (2006.01)
H04R 25/00 (2006.01)
H03G 5/16 (2006.01)
H03G 9/02 (2006.01)
H03G 9/00 (2006.01)

12 TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Fecha de presentación y número de la solicitud europea: 30.03.2007 E 07754463 (3)
 97 Fecha y número de publicación de la concesión europea: 21.11.2012 EP 2002429
- (54) Título: Control de una característica percibida del volumen sonoro de una señal de audio
- (30) Prioridad:

04.04.2006 US 789539 P

Fecha de publicación y mención en BOPI de la traducción de la patente: **08.04.2013**

73) Titular/es:

DOLBY LABORATORIES LICENSING CORPORATION (100.0%) 100 PORTRERO AVENUE SAN FRANCISCO, CALIFORNIA 94103-4813, US

(72) Inventor/es:

SEEFELDT, ALAN JEFFREY

74 Agente/Representante:

LINAGE GONZÁLEZ, Rafael

DESCRIPCIÓN

Control de una característica percibida del volumen sonoro de una señal de audio

5 Campo Técnico

La invención se refiere al procesamiento de señales de audio. Más específicamente, la invención se refiere a la medición y control del volumen sonoro percibido y/o al equilibrio espectral percibido de una señal de audio. La invención es útil, por ejemplo, en uno o más de: control de volumen compensador del volumen sonoro, control de ganancia automático, control de gama dinámico (que incluye, por ejemplo, limitadores, compresores, expansores, etc.), ecualización dinámica y compensación de la interferencia del ruido de fondo en un entorno de reproducción de audio. La invención incluye no solamente métodos, sino también los correspondientes programas de ordenador y aparatos.

Técnica anterior

15

20

25

10

Ha habido muchos intentos de desarrollar un método objetivo satisfactorio para medir el volumen sonoro. Fletcher y Munson determinaron en 1933 que el oído humano es menos sensible a frecuencias altas y bajas que a frecuencias medias (o de voz). También hallaron que el cambio relativo en la sensibilidad disminuía según aumentaba el nivel del sonido. Un primitivo medidor del volumen sonoro consistía en un micrófono, un amplificador, un medidor y una combinación de filtros diseñados para imitar groseramente la respuesta de frecuencia del oído a niveles de sonido bajos, medios y altos.

Incluso aunque tales dispositivos proporcionaban una medición del volumen sonoro de un tono aislado, individual, de nivel constante, las mediciones de sonidos más complejos no coincidían muy bien con las impresiones subjetivas del volumen sonoro. Los medidores del nivel sonoro de este tipo han sido estandarizados, pero se usan solamente para tareas específicas, tales como la monitorización y control del ruido industrial.

En los primeros años de la década de 1950, Zwicker y Stevens, entre otros, extendieron el trabajo de Fletcher y Munson al desarrollar un modelo más realista del proceso de percepción del volumen sonoro. Stevens publicó un método para el "Cálculo del volumen sonoro del ruido complejo" en la Revista de la Sociedad Acústica de América en 1956, y Zwicker publicó su artículo "Base psicológica y metodológica del volumen sonoro" en Acoustica en 1958. En 1959 Zwicker publicó un método gráfico para el cálculo del volumen sonoro, así como varios artículos similares poco después. Los métodos de Stevens y Zwicker fueron estandarizados como el estándar ISO 532, partes A y B (respectivamente). Ambos métodos implican etapas similares.

35

40

45

50

30

En primer lugar, la distribución variable en el tiempo de la energía a lo largo de la membrana basilar del oído interno, denominada la excitación, es simulada haciendo pasar el audio a través de un banco de filtros auditivos de paso de banda, con frecuencias centrales separadas uniformemente en una escala crítica de velocidades de banda. Cada filtro auditivo está diseñado para simular la respuesta de frecuencia en una ubicación específica a lo largo de la membrana basilar del oído interno, correspondiendo la frecuencia central del filtro a esta ubicación. Un ancho de banda crítica se define como el ancho de banda de un filtro de ese tipo. Medida en unidades de Hercios, el ancho de banda crítica de estos filtros auditivos aumenta con la frecuencia central creciente. Por lo tanto, es útil definir una escala de frecuencia distorsionada de modo que el ancho de banda crítica para todos los filtros auditivos medidos en esta escala distorsionada sea constante. Una escala distorsionada de tal tipo se denomina la escala crítica de velocidades de banda, y es muy útil para comprender y simular una amplia gama de fenómenos psicoacústicos. Véase, por ejemplo, Psicoacústica - Hechos y Modelos de E. Zwicker y H. Fastl, Springer-Verlag, Berlín, 1990. Los métodos de Stevens y Zwicker utilizan una escala crítica de velocidades de banda denominada la escala Bark, en la cual el ancho de banda crítica es constante por debajo de los 500 Hz y aumenta por encima de los 500 Hz. Más recientemente, Moore y Glasberg definieron una escala crítica de velocidades de banda, que llamaron la escala del Ancho de Banda Rectangular Equivalente (ERB) (B. C. J. Moore, B. Glasberg, T. Baer, "Un modelo para la predicción de umbrales, volumen sonoro y volumen sonoro parcial", Revista de la Sociedad de Ingeniería de Audio, Vol. 45, Nº 4, Abril de 1997, págs. 224 a 240). Mediante experimentos psicoacústicos, usando enmascaradores de ruido estriado, Moore y Glasberg demostraron que el ancho de banda crítica continúa disminuyendo por debajo de los 500 Hz, a diferencia de la escala Bark, donde el ancho de banda crítica permanece constante.

55

60

A continuación del cálculo de la excitación hay una función compresiva no lineal que genera una cantidad denominada "volumen sonoro específico". El volumen sonoro específico es una medida del volumen sonoro perceptivo como una función de la frecuencia y del tiempo, y puede ser medido en unidades de volumen sonoro perceptivo por frecuencia unitaria, a lo largo de una escala crítica de velocidades de banda, tal como la escala Bark o ERB expuesta anteriormente. Idealmente, el volumen sonoro específico representa una distribución continua del volumen sonoro como una función de la frecuencia y del tiempo, y el "volumen sonoro total" variable en el tiempo se calcula integrando esta distribución con respecto a la frecuencia. En la práctica, una representación exacta del volumen sonoro específico se obtiene muestreando esta distribución uniformemente a lo largo de una escala crítica de velocidades de banda, mediante el uso, por ejemplo, de los filtros auditivos mencionados anteriormente. En este caso, el volumen sonoro total puede calcularse sumando sencillamente el volumen sonoro específico proveniente de cada filtro. Para reducir la complejidad, algunas aplicaciones pueden calcular una aproximación más grosera al volumen sonoro específico, al precio de leves

inexactitudes en la estimación y modificación del volumen sonoro percibido. Tales aproximaciones se expondrán en más detalle más adelante.

El volumen sonoro puede ser medido en unidad de phon. El volumen sonoro de un sonido dado en phon es el nivel de presión sonora (SPL) de un tono de 1 kHz con un volumen sonoro subjetivo igual al del sonido. Convencionalmente, los 0 dB de referencia para el SPL es una presión de raíces cuadradas mínimas de 2 x 10⁻⁵ Pascal, y esto también es, por lo tanto, el 0 phon de referencia. Usando esta definición al comparar el volumen sonoro de los tonos a frecuencias distintas a 1 kHz con el volumen sonoro a 1 kHz, puede determinarse un perfil de igual volumen sonoro para un nivel dado de phon. La Figura 11 muestra perfiles de igual volumen sonoro para frecuencias entre 20 Hz y 12,5 kHz, y para niveles de phon entre 4,2 phon (considerado el umbral de la audición) y 120 phon (ISO226: 1087 (E), "Acústica - Perfiles normales de igual nivel de volumen sonoro"). La medición en phon tiene en cuenta la sensibilidad variable del oído humano con la frecuencia, pero los resultados no permiten la evaluación de los relativos volúmenes sonoros subjetivos de los sonidos a niveles variables, porque no hay ningún intento de corregir la no linealidad del crecimiento del volumen sonoro con el SPL, es decir, el hecho de que la separación de los perfiles varía.

15

20

10

El volumen sonoro también puede medirse en unidades de "sone". Hay una correlación de uno a uno entre las unidades de phon y las unidades de sone, según lo indicado en la figura 11. Un sone se define como el volumen sonoro de una onda sinusoidal pura de 40 dB (SPL) y 1kHz, y es equivalente a 40 phon. Las unidades de sone son tales que un aumento doble en los sone corresponde a una duplicación del volumen sonoro percibido. Por ejemplo, 4 sone son percibidos como el doble de sonoros que 2 sone. Así, expresar niveles de volumen sonoro en sone es más informativo. Dada la definición del volumen sonoro específico como una medida del volumen sonoro perceptivo como una función de la frecuencia y del tiempo, el volumen sonoro específico puede ser medido en unidades de sone por frecuencia unitaria. De ese modo, al usar la escala Bark, el volumen sonoro específico tiene unidades de sone por Bark y, análogamente, al usar la escala ERB, las unidades son sone por ERB.

25

30

Como se ha mencionado anteriormente, la sensibilidad del oído humano varía tanto con la frecuencia como con el nivel, un hecho bien documentado en la bibliografía de la psicoacústica. Uno de los resultados es que el espectro o timbre percibido de un sonido dado varía con el nivel acústico en el cual se oye el sonido. Por ejemplo, para un sonido que contiene frecuencias bajas, medias y altas, las proporciones relativas percibidas de tales componentes de frecuencia cambian con el volumen sonoro global del sonido; cuando es bajo, los componentes de frecuencias bajas y altas suenan más bajos con respecto a las frecuencias medias en que suenan cuando es alto. Este fenómeno es bien conocido y ha sido mitigado en equipos reproductores de sonido por los así denominados controles de volumen sonoro. Un control de volumen sonoro es un control de volumen que aplica estímulos de frecuencia baja y, a veces, también alta, según se reduce el volumen. Así, la sensibilidad inferior del oído en los extremos de la frecuencia es compensada por un estímulo artificial de esas frecuencias. Tales controles son completamente pasivos; el grado de compensación aplicado es una función de la configuración del control de volumen o de algún otro control operado por el usuario, no una función del contenido de las señales de audio.

35

40

En la práctica, los cambios en el equilibrio espectral relativo percibido entre las frecuencias bajas, medias y altas dependen de la señal, en particular, de su espectro efectivo, y de si se pretende que sea alta o suave. Considérese la grabación de una orquesta sinfónica. Reproducida al mismo nivel que un miembro de la audiencia oiría en la sala de conciertos, el equilibrio por todo el espectro puede ser correcto ya sea que la orquesta esté tocando alto o suavemente. Si la música es reproducida 10 dB más baja, por ejemplo, el equilibrio percibido por todo el espectro cambia de una manera para los pasajes altos y cambia de otra manera para los pasajes suaves. Un control convencional pasivo del volumen sonoro no aplica distintas compensaciones en función de la música.

45

50

En la solicitud internacional de patente nº PCT/US2004/016964, registrada el 27 de mayo de 2004, y publicada el 23 de diciembre de 2004 como el documento WO 2004/111994 A2, Seefeldt et al divulgan, entre otras cosas, un sistema para medir y ajustar el volumen sonoro percibido de una señal de audio. En dicha solicitud, un modelo psicoacústico calcula el volumen sonoro de una señal de audio en unidades perceptivas. Además, la solicitud introduce técnicas para calcular una ganancia multiplicativa de banda ancha, que, cuando se aplica al audio, da como resultado que el volumen sonoro del audio modificado en la ganancia sea esencialmente el mismo que un volumen sonoro de referencia. La aplicación de tal ganancia de banda ancha, sin embargo, cambia el equilibrio espectral percibido del audio.

55

65

El documento US 2004/190740 A1 (Chalupper et al.) divulga un método para calcular parámetros de modificación utilizables para modificar la señal de audio, a fin de reducir la diferencia entre la característica global del volumen sonoro de la señal de audio y un volumen sonoro global de destino.

60

El documento US 2002/076072 A1 (Cornelisse) divulga un enfoque similar, en el cual se usa un número reducido de canales de frecuencia.

Divulgación de la invención

La invención está definida por las reivindicaciones independientes. Las reivindicaciones dependientes se refieren a características optativas de algunas realizaciones de la invención.

ES 2 400 160 T3

En un aspecto, la invención está orientada a los métodos según lo definido en las reivindicaciones 1 y 3, para controlar una característica específica del volumen sonoro de una señal de audio, en donde la característica específica del volumen sonoro es o bien una aproximación del volumen sonoro específico o bien una aproximación del volumen sonoro específico parcial.

De acuerdo a aspectos adicionales de la invención, se proporcionan un aparato según la reivindicación 6 y un programa de ordenador según la reivindicación 7.

5

10

25

30

35

40

60

El volumen sonoro específico es una medida del volumen sonoro perceptivo como función de la frecuencia y del tiempo. En implementaciones prácticas, puede hacerse que el volumen sonoro específico de la señal de audio modificada se aproxime al volumen sonoro específico de destino. La aproximación puede estar afectada no solamente por consideraciones del procesamiento de señales ordinarias, sino también por el allanamiento con respecto al tiempo y/o a la frecuencia, que puede ser empleado en la modificación, según se describe más adelante.

Debido a que el volumen sonoro específico es una medida del volumen sonoro perceptivo de una señal de audio como función de la frecuencia y del tiempo, a fin de reducir la diferencia entre el volumen sonoro específico de la señal de audio y el volumen sonoro específico de destino, la modificación puede modificar la señal de audio en función de la frecuencia. Aunque en algunos casos el volumen sonoro específico de destino también puede ser invariante con respecto al tiempo y la señal de audio en sí misma puede ser una señal de estado estable invariante con respecto al tiempo, habitualmente, la modificación también puede modificar la señal de audio como función del tiempo.

Aspectos de la presente invención también pueden ser empleados para compensar el ruido de fondo que interfiere en un entorno de reproducción de audio. Cuando el audio es oído en presencia de ruido de fondo, el ruido puede enmascarar, parcialmente o completamente, el audio, de una manera que depende tanto del nivel y del espectro del audio como del nivel y espectro del ruido. El resultado es una alteración en el espectro percibido del audio. De acuerdo a estudios psicoacústicos (véase, por ejemplo, de Moore, Glasberg y Baer, "Un modelo para la predicción de umbrales, volumen sonoro y volumen sonoro parcial", Rev. de la Soc. de Ing. de Audio, Vol. 45, Nº 4, Abril de 1997), puede definirse el "volumen sonoro específico parcial" del audio como el volumen sonoro perceptivo del audio en presencia de una señal de sonido secundaria de interferencia, tal como el ruido.

Un aspecto de la invención incluye obtener parámetros de modificación utilizables para modificar la señal de audio, a fin de reducir la diferencia entre su característica específica de volumen sonoro y la aproximación a un volumen sonoro específico de destino. Hacerlo así mitiga los efectos del ruido de una manera perceptivamente precisa. En este y otros aspectos de la invención que tienen en cuenta una señal de ruido de interferencia, se supone que hay un acceso a la señal de audio por sí misma y a la señal secundaria de interferencia por sí misma.

Otro aspecto de la invención incluye controlar la aproximación del volumen sonoro específico de una señal de audio, modificando la señal de audio a fin de reducir la diferencia entre su aproximación del volumen sonoro específico y la aproximación a un volumen sonoro específico de destino.

Otro aspecto de la invención incluye controlar la aproximación del volumen sonoro específico parcial de una señal de audio, modificando la señal de audio a fin de reducir la diferencia entre su aproximación del volumen sonoro específico parcial y la aproximación a un volumen sonoro específico de destino.

Cuando el volumen sonoro específico de destino no es una función de la señal de audio, puede ser un volumen sonoro específico de destino almacenado o recibido. Cuando el volumen sonoro específico de destino no es una función de la señal de audio, la modificación o la obtención puede calcular, explícitamente o implícitamente, el volumen sonoro específico o el volumen sonoro específico parcial. Los ejemplos de cálculo implícito incluyen una tabla de búsqueda o una expresión matemática de "forma cerrada", en la cual el volumen sonoro específico y/o el volumen sonoro específico parcial está inherentemente determinado (se pretende que el término "forma cerrada" describa una expresión matemática que pueda ser representada exactamente usando un número finito de operaciones y funciones matemáticas estándar, tales como la exponenciación y el coseno). Además, cuando el volumen sonoro específico de destino no es una función de la señal de audio, el volumen específico de destino puede ser invariante con respecto tanto al tiempo como a la frecuencia, o bien puede ser solamente invariante con respecto al tiempo.

Las divulgaciones que puedan ser útiles con respecto a la presente invención admiten el procesamiento de la señal de audio o una medida de la señal de audio, de acuerdo a uno o más procesos y a uno o más parámetros de control de procesos, para producir un volumen sonoro específico de destino. Aunque el volumen sonoro específico de destino puede ser invariante con respecto al tiempo ("fijo"), el volumen sonoro específico de destino puede ser, ventajosamente, una función del volumen sonoro específico de la señal de audio. Aunque puede ser una señal estática, invariante con respecto a la frecuencia y al tiempo, habitualmente, la señal de audio en sí misma es variable con respecto a la frecuencia y al tiempo, causando de ese modo que el volumen sonoro específico de destino sea variable con respecto a la frecuencia y al tiempo cuando es una función de la señal de audio.

El audio y un volumen sonoro específico de destino, o una representación de un volumen sonoro específico de destino, pueden ser recibidos desde una transmisión o reproducidos a partir de un medio de almacenamiento.

La representación de un volumen sonoro específico de destino puede ser uno o más factores de escala que ajusten a escala la señal de audio, o una medida de la señal de audio.

- El volumen sonoro específico de destino puede ser una función de la señal de audio, o una medida de la señal de audio. Una medida adecuada de la señal de audio es el volumen sonoro específico de la señal de audio. La función de la señal de audio, o la medida de la señal de audio, puede ser un ajuste a escala de la señal de audio, o de la medida de la señal de audio. Por ejemplo, el ajuste a escala puede ser uno, o una combinación de ajustes a escala:
- 10 (a) un ajuste a escala, por un factor ∑ [b,t] de escala variable con respecto al tiempo y a la frecuencia, del volumen sonoro específico, según la relación

$$\hat{N}[b,t] = \Xi[b,t]N[b,t];$$

15 (b) un ajuste a escala, por un factor Φ[t] de escala invariante con respecto a la frecuencia, y variable con respecto al tiempo, del volumen sonoro específico, según la relación

$$\hat{N}[b,t] = \Phi[t]N[b,t];$$

20 (c) un ajuste a escala, por un factor Θ[b] de escala variable con respecto a la frecuencia, e invariante con respecto al tiempo, del volumen sonoro específico, según la relación

$$\hat{N}[b,t] = \Theta[b]N[b,t];$$

25 y

(d) un ajuste a escala, por un factor α de escala invariante con respecto a la frecuencia e invariante con respecto al tiempo, del volumen sonoro específico de la señal de audio según la relación

$$\hat{N}[b,t] = \alpha N[b,t],$$

en donde $\hat{N}[b,t]$ es el volumen sonoro específico de destino, N[b,t] es el volumen sonoro específico de la señal de audio, b es una medida de la frecuencia y t es una medida del tiempo.

En el caso (a) de un factor de escala variable en el tiempo y en la frecuencia, el ajuste a escala puede ser determinado, al menos en parte, por una razón entre el volumen sonoro multibanda deseado y el volumen sonoro multibanda de la señal de audio. Un ajuste a escala de ese tipo puede ser utilizable como un control de gama dinámico. Detalles adicionales del empleo de aspectos de la invención como un control de gama dinámico se estipulan más adelante.

También en el caso (a) de un factor de escala variable en el tiempo y en la frecuencia, el volumen sonoro específico puede ser ajustado a escala por la razón entre una medida de una forma espectral deseada y la medida de una forma espectral de la señal de audio. Un ajuste a escala de ese tipo puede ser empleado para transformar el espectro percibido de la señal de audio, de un espectro percibido variable en el tiempo a un espectro percibido esencialmente invariable en el tiempo. Cuando el volumen sonoro específico es ajustado a escala por la razón entre una medida de una forma espectral deseada y la medida de una forma espectral de la señal de audio, un ajuste a escala de ese tipo puede ser utilizable como un ecualizador dinámico.

45

40

35

En el caso (b) de un factor de escala variable en el tiempo e invariable en cuanto a la frecuencia, el ajuste a escala puede ser determinado, al menos en parte, por una razón entre el volumen sonoro de banda ancha deseado y el volumen sonoro de banda ancha de la señal de audio. Un ajuste a escala de ese tipo puede ser utilizable como un control de ganancia automático o un control de gama dinámico.

50

En el caso (a) (un factor de escala variable en el tiempo y en la frecuencia) o en el caso (b) (un factor de escala variable en el tiempo e invariante en cuanto a la frecuencia), el factor de escala puede ser una función de la señal de audio o una medida de la señal de audio.

Tanto en el caso (c) de un factor de escala invariante en el tiempo y variable en cuanto a la frecuencia, como en el caso (d) de un factor de escala invariante en el tiempo e invariante en cuanto a la frecuencia, la modificación o la obtención puede incluir almacenar el factor de escala, o bien el factor de escala puede ser recibido desde un origen externo.

ES 2 400 160 T3

En cualquiera de los casos (c) y (d), el factor de escala puede no ser una función de la señal de audio o una medida de la señal de audio.

La modificación, obtención o producción puede calcular, variadamente, explícita o implícitamente, (1) el volumen sonoro específico y/o (2) el volumen sonoro específico parcial y/o (3) el volumen sonoro específico de destino. Los cálculos implícitos pueden implicar, por ejemplo, una tabla de búsqueda o una expresión matemática de forma cerrada.

Los parámetros de modificación pueden ser temporalmente allanados. Los parámetros de modificación pueden ser, por ejemplo, (1) una pluralidad de factores de escala de amplitud relacionados con las bandas de frecuencia de la señal de audio o (2) una pluralidad de coeficientes de filtro para controlar uno o más filtros, tales como un filtro FIR con tomas múltiples o un filtro IIR multipolar. Los factores de escala o los coeficientes de filtro (y los filtros a los cuales se aplican) pueden ser variables en el tiempo.

Al calcular la función del volumen sonoro específico de la señal de audio que define el volumen sonoro específico de destino, o la inversa de esa función, el proceso, o procesos, que realiza(n) tales cálculos opera(n) en lo que puede ser caracterizado como el dominio perceptivo (psicoacústico) del volumen sonoro – la entrada y la salida del cálculo son volúmenes sonoros específicos. Por el contrario, al aplicar factores de escala de amplitud a bandas de frecuencia de la señal de audio o al aplicar coeficientes de filtro a un filtrado controlable de la señal de audio, los parámetros de modificación operan para modificar la señal de audio fuera del dominio perceptivo (psicoacústico) del volumen sonoro, en lo que puede ser caracterizado como el dominio de señales eléctricas. Aunque las modificaciones a la señal de audio pueden ser hechas a la señal de audio en el dominio de las señales eléctricas, tales cambios en el dominio de las señales eléctricas son obtenidos de cálculos en el dominio perceptivo (psicoacústico) del volumen sonoro, de modo que la señal de audio modificada tenga un volumen sonoro específico que se aproxime al volumen sonoro específico de destino deseado.

Al obtener parámetros de modificación de cálculos en el dominio del volumen sonoro, puede lograrse un control sobre el volumen sonoro perceptivo y el equilibrio espectral percibido, mayor que si tales parámetros de modificación fueran obtenidos en el dominio de las señales eléctricas. Además, el uso de una membrana basilar que simula un banco de filtros psicoacústico, o su equivalente, al realizar los cálculos del dominio del volumen sonoro, puede proporcionar un control más detallado del espectro percibido que en las disposiciones que obtienen parámetros de modificación en el dominio de las señales eléctricas.

Cualquiera entre la modificación, la obtención y la producción puede ser dependiente de uno o más entre una medida de una señal de audio de interferencia, un volumen sonoro específico de destino, una estimación del volumen sonoro específico de la señal de audio no modificada obtenida del volumen sonoro específico o un volumen sonoro específico parcial de la señal de audio modificada, el volumen sonoro específico de la señal de audio no modificada, y una aproximación al volumen sonoro específico de destino obtenido del volumen sonoro específico, o del volumen sonoro específico parcial, de la señal de audio modificada.

La modificación u obtención puede obtener parámetros de modificación, al menos en parte, de uno o más entre una medida de una señal de audio de interferencia, un volumen sonoro específico de destino, una estimación del volumen sonoro específico de la señal de audio no modificada obtenida del volumen sonoro específico, o del volumen sonoro específico parcial, de la señal de audio modificada, el volumen sonoro específico de la señal de audio no modificada, y una aproximación al volumen sonoro específico de destino obtenido del volumen sonoro específico, o del volumen sonoro específico parcial, de la señal de audio modificada.

Más específicamente, la modificación u obtención puede obtener parámetros de modificación, al menos en parte, de

50 (1) uno entre

30

40

45

un volumen sonoro específico de destino, y

una estimación del volumen sonoro específico de la señal de audio no modificada, recibida desde el volumen sonoro específico de la señal de audio modificada, y

(2) uno entre

55 el volumen sonoro específico de la señal de audio no modificada, y

una aproximación al volumen sonoro específico de destino obtenido del volumen sonoro específico de la señal de audio modificada.

o bien, cuando ha de tenerse en cuenta una señal de audio de interferencia, la modificación u obtención puede obtener 60 parámetros de modificación, al menos en parte, de

- (1) una medida de una señal de audio de interferencia,
- (2) uno entre

un volumen sonoro específico de destino, y

una estimación del volumen sonoro específico de la señal de audio no modificada, obtenido desde el volumen sonoro específico parcial de la señal de audio modificada, y

(3) uno entre

el volumen sonoro específico de la señal de audio no modificada, y

una aproximación al volumen sonoro específico de destino obtenido desde el volumen sonoro específico parcial de la señal de audio modificada.

5

10

15

20

Puede ser empleada una disposición de alimentación hacia adelante, en la cual el volumen sonoro específico se obtiene de la señal de audio, y en la cual el volumen sonoro específico de destino es recibido desde un origen externo al método, o desde un almacenamiento, cuando la modificación u obtención incluye almacenar un volumen sonoro específico de destino. Alternativamente, puede ser empleada una disposición híbrida de alimentación hacia adelante o de retroalimentación, en la cual una aproximación al volumen sonoro específico de destino es obtenida de la señal de audio modificada, y en la cual el volumen sonoro específico de destino es recibido desde un origen externo al método, o desde un almacenamiento, cuando la modificación u obtención incluye almacenar un volumen sonoro específico de destino.

La modificación u obtención puede incluir uno o más procesos para obtener, explícitamente o implícitamente, el volumen sonoro específico de destino, uno o más de los cuales calcula(n), explícitamente o implícitamente, una función de la señal de audio, o una medida de la señal de audio. En una alternativa, puede ser empleada una disposición de alimentación hacia adelante, en la cual el volumen sonoro específico y el volumen sonoro específico de destino son obtenidos de la señal de audio, empleando la obtención del volumen sonoro específico de destino una función de la señal de audio, o una medida de la señal de audio. En otra alternativa, puede ser empleada una disposición híbrida de alimentación hacia adelante o retroalimentación, en la cual una aproximación al volumen sonoro específico de destino es obtenida a partir de la señal de audio modificada y el volumen sonoro específico de destino es obtenido de la señal de audio, empleando la obtención del volumen sonoro específico de destino una función de la señal de audio o una medida de la señal de audio.

La modificación u obtención puede incluir uno o más procesos para obtener, explícitamente o implícitamente, una estimación del volumen sonoro específico de la señal de audio no modificada en respuesta a la señal de audio modificada, uno o más de los cuales calcula(n), explícitamente o implícitamente, la inversa de una función de la señal de audio, o una medida de la señal de audio. En una alternativa, es empleada una disposición de retroalimentación en la cual una estimación del volumen sonoro específico de la señal de audio no modificado, y una aproximación al volumen sonoro específico de destino, son obtenidas de la señal de audio modificada, siendo calculada la estimación del volumen sonoro específico usando la inversa de una función de la señal de audio, o una medida de la señal de audio. En otra alternativa, es empleada una disposición híbrida de alimentación hacia adelante o retroalimentación, en la cual el volumen sonoro específico es obtenido de la señal de audio y la estimación del volumen sonoro específico de la señal de audio no modificada es obtenida a partir de la señal de audio modificada, siendo calculada la obtención de la estimación usando la inversa de dicha función de la señal de audio, o la medida de la señal de audio.

Los parámetros de modificación pueden ser aplicados a la señal de audio para producir una señal de audio modificada.

Puede haber una separación temporal y/o espacial de los procesos o dispositivos, de modo que haya, de hecho, un codificador o una codificación, y también un descodificador o una descodificación. Por ejemplo, puede haber un sistema de codificación/descodificación en el cual la modificación u obtención puede bien transmitir y recibir, o bien almacenar y también reproducir la señal de audio, y bien (1) los parámetros de modificación o bien (2) una aproximación a un volumen sonoro específico de destino, o una representación de la aproximación a ese volumen sonoro específico de destino. Alternativamente, puede haber, de hecho, solamente un codificador, o una codificación, en donde hay bien una transmisión o almacenamiento de la señal de audio y (1) parámetros de modificación o (2) una aproximación a un volumen sonoro específico de destino. Alternativamente, como se ha mencionado anteriormente, puede haber, de hecho, solamente un descodificador, o una descodificación, en donde haya también una recepción y reproducción de la señal de audio y (1) parámetros de modificación o (2) una aproximación a un volumen sonoro específico de destino o una representación de la aproximación a ese volumen sonoro específico de destino.

Descripción de los dibujos

La figura1 es un diagrama de bloques funcionales que ilustra un ejemplo de una implementación de alimentación hacia adelante según aspectos de la invención.

La figura 2 es un diagrama de bloques funcionales que ilustra un ejemplo de una implementación de retroalimentación según aspectos de la invención.

La figura 3 es un diagrama en bloques funcionales que ilustra un ejemplo de una implementación híbrida de alimentación hacia adelante o retroalimentación, según aspectos de la invención.

La figura 4 es un diagrama en bloques funcionales que ilustra un ejemplo de otra implementación híbrida de alimentación hacia adelante o retroalimentación, según aspectos de la invención.

65

La figura 5 es un diagrama en bloques funcionales que ilustra la manera en que la señal de audio no modificada y los

ES 2 400 160 T3

parámetros de modificación, según lo determinado por cualquiera de las disposiciones de alimentación hacia adelante, de retroalimentación y la híbrida de alimentación hacia adelante o retroalimentación, pueden ser almacenados o transmitidos para su uso, por ejemplo, en un dispositivo o proceso temporal o espacialmente separado.

- La figura 6 es un diagrama de bloques funcionales que ilustra la manera en la cual la señal de audio no modificada y un volumen sonoro específico de destino, o una representación del mismo, según lo determinado por cualquiera entre las disposiciones de alimentación hacia adelante, retroalimentación, y la híbrida de alimentación hacia adelante o retroalimentación, pueden ser almacenados o transmitidos para su uso, por ejemplo, en un dispositivo o proceso temporal o espacialmente separado.
 - La figura 7 es un diagrama esquemático de bloques funcionales, o diagrama de flujo esquemático, que muestra un panorama de un aspecto de la presente invención.
- La figura 8 es una respuesta característica idealizada de un filtro lineal P(z) adecuado como filtro de transmisión, en una realización de la presente invención en la cual el eje vertical es la atenuación en decibelios (dB) y el eje horizontal es una frecuencia en base logarítmica 10, en Hercios (Hz).
 - La figura 9 muestra la relación entre la escala de frecuencias ERB (eje vertical) y la frecuencia en Hercios (eje horizontal).
- 20 La figura 10 muestra un conjunto de respuestas idealizadas de características de filtros auditivos, que se aproximan al efecto crítico de banda en la escala ERB. La escala horizontal es la frecuencia en Hercios y la escala vertical es el nivel en decibelios.
- La figura 11 muestra los perfiles de igual volumen sonoro del estándar ISO 226. La escala horizontal es la frecuencia en Hercios (escala en base logarítmica 10) y la escala vertical es el nivel de presión sonora en decibelios.
 - La figura 12 muestra los perfiles de igual volumen sonoro del estándar ISO 226, normalizados por el filtro P(z) de transmisión. La escala horizontal es la frecuencia en Hercios (escala de base logarítmica 10) y la escala vertical es el nivel de presión sonora en decibelios.
 - La figura 13a es un diagrama idealizado que muestra las ganancias de banda ancha y de multibanda para un ajuste a escala del volumen sonoro de 0,25 sobre un segmento de habla femenina. La escala horizontal es de las bandas ERB y la escala vertical es la ganancia relativa en decibelios (dB).
- La figura 13b es un diagrama idealizado que muestra el volumen sonoro específico, respectivamente, de una señal original, una señal modificada por ganancia de banda ancha y una señal modificada por ganancia de multibanda. La escala horizontal es de las bandas ERB y la escala vertical es el volumen sonoro específico (sone/ERB).
- La figura 14a es un diagrama idealizado que muestra: $L_0[t]$ como una función de $L_1[t]$ para el AGC típico. La escala 40 horizontal es $\log(L_1[t])$ y la escala vertical es $\log(L_0[t])$.
 - La figura 14b es un diagrama idealizado que muestra: $L_0[t]$ como una función de $L_j[t]$ para el DRC típico. La escala horizontal es $\log (L_j[t])$ y la escala vertical es $log(L_0[t])$.
- La figura 15 es un diagrama idealizado que muestra una típica función de allanamiento de banda para el DRC de multibanda. La escala horizontal es el número de banda y la escala vertical es la ganancia emitida para la banda b.
 - La figura 16 es un diagrama esquemático de bloques funcionales, o diagrama esquemático de flujo, que muestra un panorama de un aspecto de la presente invención.
 - La figura 17 es un diagrama esquemático de bloques funcionales, o diagrama esquemático de flujo, similar a la figura 1, que también incluye la compensación para el ruido en un entorno de reproducción.

Mejor modo de llevar a cabo la invención

30

50

55

- Las figuras 1 a 4 muestran diagramas de bloques funcionales que ilustran posibles ejemplos de implementación, de alimentación hacia adelante, de retroalimentación y dos versiones de alimentación híbrida hacia adelante o de retroalimentación, según aspectos de la invención.
- Con referencia al ejemplo de una topología de alimentación hacia adelante en la figura 1, una señal de audio es aplicada a dos trayectos: (1) un trayecto de señal con un proceso o dispositivo 2 ("Modificar señal de audio") capaz de modificar el audio en respuesta a parámetros de modificación y (2) un trayecto de control con un proceso o dispositivo 4 ("Generar parámetros de modificación") capaz de generar tales parámetros de modificación. El bloque Modificación de señal de audio 2 en el ejemplo de topología de alimentación hacia adelante de la figura 1 y en cada uno de los ejemplos de las figuras 2 a 4 puede ser un dispositivo o proceso que modifica la señal de audio, por ejemplo, su amplitud, de una manera variable en la frecuencia y/o en el tiempo, de acuerdo a los parámetros M de modificación recibidos desde el bloque

Generar parámetros de modificación 4 (o desde los procesos o dispositivos 4', 4" y 4" de contrapartida en cada uno de los ejemplos de las figuras 2 a 4, respectivamente). El bloque Generar parámetros de modificación 4 y cada una de sus contrapartes en las figuras 2 a 4 operan al menos parcialmente en el dominio del volumen sonoro perceptivo. El bloque Modificar señal de audio 2 opera en el dominio de señales eléctricas y produce una señal de audio modificada en cada uno de los ejemplos de las figuras 1 a 4, el bloque Modificar señal de audio 2 y el bloque Generar parámetros de modificación 4 (o sus contrapartes) modifican la señal de audio para reducir la diferencia entre su volumen sonoro específico y un volumen sonoro específico de destino.

En el ejemplo de alimentación hacia adelante de la figura 1, el proceso o dispositivo 4 puede incluir varios procesos y/o dispositivos: un proceso o dispositivo 6 para "Calcular volumen sonoro específico de destino", que calcula un volumen sonoro específico de destino en respuesta a la señal de audio o a una medida de la señal de audio, tal como el volumen sonoro específico de la señal de audio, un proceso o dispositivo 8 para "Calcular volumen sonoro específico" que calcula el volumen sonoro específico de la señal de audio en respuesta a la señal de audio o a una medida de las señales de audio, tal como su excitación, y un proceso o dispositivo 10 para "Calcular parámetros de modificación" que calcula los parámetros de modificación en respuesta al volumen sonoro específico y al volumen sonoro específico de destino. El bloque Calcular volumen sonoro específico de destino 6 puede realizar una o más funciones "F", cada una de las cuales puede tener parámetros de función. Por ejemplo, puede calcular el volumen sonoro específico de la señal de audio y luego aplicar una o más funciones F al mismo para proporcionar un volumen sonoro específico de destino. Esto es indicado esquemáticamente en la figura 1 como un bloque para "Seleccionar la(s) función(es) F y parámetro(s) de función(es)", ingresado al proceso o dispositivo 6. En lugar de ser calculado por el dispositivo o proceso 6, el volumen sonoro específico de destino puede ser proporcionado por un proceso o dispositivo de almacenamiento (mostrado esquemáticamente como un bloque "Almacenado" ingresado al proceso o dispositivo 10) incluido en, o asociado a, el bloque Generar parámetros de modificación 4, o por una fuente externa al proceso o dispositivo global (mostrado esquemáticamente como el bloque "Externo" ingresado al proceso o dispositivo 10). De este modo, los parámetros de modificación están basados, al menos en parte, en cálculos en el dominio del volumen sonoro perceptual (psicoacústico) (es decir, al menos el volumen sonoro específico y, en algunos casos, los cálculos del volumen sonoro específico de destino).

Los cálculos realizados por los procesos o dispositivos 6, 8 y 10 (y por los procesos o dispositivos 12, 14, 10' en el ejemplo de la figura 2, 6, 14, 10" en el ejemplo de la figura 3, y 8, 12, 10" en el ejemplo de la figura 4) pueden ser realizados explícitamente y/o implícitamente. Los ejemplos de realización implícita incluyen (1) una tabla de búsqueda cuyas entradas están basadas, totalmente o en parte, en el volumen sonoro específico y/o el volumen sonoro específico de destino y/o cálculos de parámetros de modificación, y (2) una expresión matemática de forma cerrada que está inherentemente basada, totalmente o en parte, en el volumen sonoro específico y/o el volumen sonoro específico de destino y/o los parámetros de modificación.

Aunque los procesos o dispositivos 6, 8 y 10 de cálculo del ejemplo de la figura 1 (y los procesos o dispositivos 12, 14, 10' en el ejemplo de la figura 2, 6, 14, 10'' en el ejemplo de la figura 3, y 8, 12, 10''' en el ejemplo de la figura 4) se muestran esquemáticamente y se describen como separados, esto es solamente con fines explicativos. Se entenderá que uno de, o todos, estos procesos o dispositivos puede(n) combinarse en un único proceso o dispositivo, o combinarse de forma variada en múltiples procesos o dispositivos. Por ejemplo, en la disposición de la figura 9 más adelante, una topología de alimentación hacia adelante como en el ejemplo de la figura 1, el proceso o dispositivo que calcula parámetros de modificación lo hace en respuesta a la excitación allanada obtenida de una señal de audio y de un volumen sonoro específico de destino. En el ejemplo de la figura 9, el dispositivo o proceso que calcula parámetros de modificación calcula implícitamente el volumen sonoro específico de la señal de audio.

Como un aspecto de la presente invención, en el ejemplo de la figura 1 y en otros ejemplos de realizaciones de la invención en el presente documento, el volumen sonoro específico de destino $(\hat{N}[b,t])$ puede ser calculado ajustando a escala el volumen sonoro específico (N[b,t]) con uno o más factores de ajuste a escala. El ajuste a escala puede ser un ajuste a escala por el factor $\Xi[b,t]$ de escala, variable en el tiempo y en la frecuencia, del volumen sonoro específico, según la relación

$$\hat{N}[b,t] = \Xi[b,t]N[b,t],$$

10

15

20

25

40

45

50

un ajuste a escala por el factor Φ[t] de escala, variable en el tiempo e invariante en cuanto a la frecuencia, del volumen sonoro específico, según la relación

$$\hat{N}[b,t] = \Phi[t]N[b,t],$$

60 un ajuste a escala, por el factor Θ[b] de escala, invariante en el tiempo y variable en cuanto a la frecuencia, del volumen sonoro específico, según la relación

$$\hat{N}[b,t] = \Theta[b]N[b,t],$$

o un ajuste a escala, por el factor α de escala, del volumen sonoro específico de la señal de audio, según la relación

$$\hat{N}[b,t] = \alpha N[b,t],$$

5

15

20

25

35

40

45

50

donde b es una medida de frecuencia (p. ej., el número de banda) y t es una medida de tiempo (p. ej., el número de bloque). También pueden emplearse múltiples ajustes a escala, usando múltiples casos de un ajuste a escala específico y/o combinaciones de ajustes a escala específicos. Los ejemplos de tales múltiples ajustes a escala se dan más adelante. En algunos casos, según lo explicado más adelante, el ajuste a escala puede ser una función de la señal de audio o de la medida de la señal de audio. En otros casos no referidos a la invención, también según se explica más adelante, cuando el ajuste a escala no es una función de una medida de la señal de audio, el ajuste a escala puede ser determinado o suministrado de otro modo. Por ejemplo, un usuario podría seleccionar o aplicar un ajuste a escala por un factor α de escala, invariante en cuanto al tiempo y a la frecuencia, o por un factor $\Theta[b]$ de escala, invariante en el tiempo y variable en cuanto a la frecuencia.

Así, el volumen sonoro específico de destino puede ser expresado como una o más funciones *F* de la señal de audio o de la medida de la señal de audio (siendo el volumen sonoro específico una posible medida de la señal de audio):

$$\hat{N}[b,t] = F(N[b,t]).$$

Siempre que la función, o funciones F, sea(n) invertible(s), el volumen sonoro específico (N[b,t]) de la señal de audio no modificada puede ser calculado como la función, o funciones, inversa(s) F^1 del volumen sonoro específico de destino ($\hat{N}[b,t]$):

$$N[b,t] = F^{-1}(\hat{N}[b,t])$$
.

Como se verá más adelante, la función, o las funciones, inversa(s) F^1 se calcula(n) en los ejemplos de retroalimentación y de alimentación híbrida hacia adelante y retroalimentación de las figuras 2 y 4.

Se muestra un bloque "Seleccionar función(es) y parámetro(s) de función", ingresado para el bloque Calcular volumen sonoro específico de destino 6, para indicar que el dispositivo o proceso 6 puede calcular el volumen sonoro específico de destino aplicando una o más funciones de acuerdo a uno o más parámetros de función. Por ejemplo, el bloque Calcular volumen sonoro específico de destino 8 puede calcular la función o funciones "F" del volumen sonoro específico de la señal de audio, a fin de definir el volumen sonoro específico de destino. Por ejemplo, el bloque de entrada "Seleccionar función(es) y parámetro(s) de función" puede seleccionar una o más funciones específicas, que caen en uno o más de los anteriores tipos de ajuste a escala, junto con uno o más parámetros de función, tales como constantes (p. ej., factores de escala) pertenecientes a las funciones.

Los factores de escala asociados a un ajuste a escala pueden servir como una representación del volumen sonoro específico de destino, en la medida en que el volumen sonoro específico de destino pueda ser calculado como un ajuste a escala del volumen sonoro específico, según lo indicado anteriormente. Así, en el ejemplo de la figura 9, descrito más adelante y mencionado anteriormente, la tabla de búsqueda puede estar indizada por factores de escala y excitaciones, de modo que el cálculo del volumen sonoro específico y el volumen sonoro específico de destino sean inherentes en la tabla.

Ya sea empleando una tabla de búsqueda, una expresión matemática de forma cerrada, o alguna otra técnica, el funcionamiento del bloque Generar parámetros de modificación 4 (y sus procesos o dispositivos 4', 4" y 4" de contrapartida, en cada uno de los ejemplos de las figuras 2 a 4) es tal que los cálculos se basan en el dominio del volumen sonoro perceptivo (psicoacústico), incluso aunque el volumen sonoro específico y el volumen sonoro específico de destino no puedan ser calculados explícitamente. O bien hay un volumen sonoro específico explícito o bien hay un volumen sonoro específico explícito de destino o bien hay un volumen sonoro específico de destino ficticio e implícito. En cualquier caso, el cálculo de los parámetros de modificación busca generar parámetros de modificación que modifiquen la señal de audio para reducir la diferencia entre el volumen sonoro específico y un volumen sonoro específico de destino.

En un entorno de reproducción con una señal secundaria de audio de interferencia, tal como el ruido, el bloque Calcular parámetros de modificación 10 (y sus procesos o dispositivos 10', 10" y 10" de contrapartida en cada uno de los

ejemplos de las figuras 2 a 4, respectivamente) también puede recibir como una entrada optativa una medida de una tal señal secundaria de audio de interferencia, o la misma señal secundaria de interferencia, como una de sus entradas. Una entrada optativa de ese tipo se muestra en la figura 1 (y en las figuras 2 a 4) con una línea conductora discontinua. La medida de una señal secundaria de interferencia puede ser su excitación, tal como en el ejemplo de la figura 17, descrito más adelante. La aplicación de una medida de la señal de interferencia, o la misma señal (suponiéndose que la señal de interferencia está disponible por separado para su procesamiento) a los procesos o dispositivos 10 para Calcular parámetros de modificación en la figura 1(y sus procesos o dispositivos 10', 10" y 10" de contrapartida en cada uno de los ejemplos de las figuras 2 a 4, respectivamente) permite a un tal proceso o dispositivo, debidamente configurado, calcular parámetros de modificación que tienen en cuenta la señal de interferencia, según se explica más adelante bajo el título "Compensación de ruido". En los ejemplos de las figuras 2 a 4, el cálculo del volumen sonoro específico parcial supone que una medida adecuada de una señal de interferencia es aplicada no solamente a los respectivos bloques Calcular parámetros de modificación 10', 10" o 10", sino también a un proceso o dispositivo 12 para "Calcular aproximación de volumen sonoro específico de audio no modificado", y/o a un proceso o dispositivo 14 para "Calcular aproximación de volumen sonoro específico de destino", a fin de facilitar el cálculo del volumen sonoro específico parcial por esa función o dispositivo. En el ejemplo de alimentación hacia adelante de la figura 1, el volumen sonoro específico parcial no es calculado explícitamente: el bloque Calcular parámetros de modificación 10 de la figura 1 calcula los parámetros de modificación adecuados para hacer que el volumen sonoro específico parcial del audio modificado se aproxime al volumen sonoro específico de destino. Esto se explica más adelante bajo el título "Compensación de ruido" mencionado anteriormente.

20

25

30

35

40

45

50

55

60

65

10

15

Como se ha mencionado anteriormente, en cada uno de los ejemplos de las figuras 1 a 4, los parámetros M de modificación, cuando son aplicados a la señal de audio por el Modificador de señal de audio 2, reducen la diferencia entre el volumen sonoro específico, el volumen sonoro específico parcial del audio modificado resultante, y el volumen sonoro específico de la señal de audio modificada se aproxima estrechamente a, o es el mismo que, el volumen sonoro específico de destino. Los parámetros M de modificación, por ejemplo, pueden tomar la forma de factores de ganancia variables en el tiempo, aplicados a las bandas de frecuencia obtenidas de un banco de filtros, o a los coeficientes de un filtro variable en el tiempo. En consecuencia, en todos los ejemplos de las figuras 1 a 4, el bloque Modificar señal de audio 2 puede ser implementado, por ejemplo, como una pluralidad de ajustadores a escala de la amplitud, funcionando cada uno en una banda de frecuencia, o un filtro variable en el tiempo (p. ej., un filtro FIR con múltiples tomas o un filtro IIR multipolar).

Aquí y en el resto de este documento, el uso del mismo número de referencia indica que el dispositivo o proceso puede ser esencialmente idéntico a otro, u otros, que lleve(n) el mismo número de referencia. Los números de referencia que lleven números con comillas (p. ej., "10") indican que el dispositivo o proceso es similar en estructura o función a, pero puede ser una modificación de, otro u otros que lleven el mismo número básico de referencia o versiones con comillas del mismo.

Bajo ciertas restricciones, puede ser realizada una disposición de retroalimentación casi equivalente al ejemplo de alimentación hacia adelante de la figura 1. La figura 2 ilustra un tal ejemplo, en el cual la señal de audio también es aplicada a un proceso o dispositivo 2 para Modificar señal de audio, en un trayecto de señal. El proceso o dispositivo 2 también recibe los parámetros M de modificación desde un trayecto de control, en el cual un proceso o dispositivo 4' para Generar parámetros de modificación, en una disposición de retroalimentación, recibe como su entrada la señal de audio modificada desde la salida del bloque Modificar señal de audio 2. Así, en el ejemplo de la figura 2, el audio modificado, en lugar del audio no modificado, es aplicado a un trayecto de control. El proceso o dispositivo 2 para Modificar señal de audio y el proceso o dispositivo 4' para Generar parámetros de modificación modifican la señal de audio para reducir la diferencia entre su volumen sonoro específico y un volumen sonoro específico de destino. El proceso o dispositivo 4' puede incluir varias funciones y/o dispositivos: un proceso o dispositivo 12 para "Calcular aproximación de volumen sonoro específico de audio no modificado", un proceso o dispositivo 14 para "Calcular aproximación de volumen sonoro específico de destino" y un proceso o dispositivo 10' para "Calcular parámetros de modificación", que calcula los parámetros de modificación.

Con la restricción de que la función, o funciones, F, sea(n) invertible(s), el proceso o dispositivo 12 estima el volumen sonoro específico o la señal de audio no modificada aplicando la función inversa F^1 al volumen sonoro específico o al volumen sonoro específico parcial de la señal de audio modificada. El dispositivo o proceso 12 puede calcular una función inversa F^1 , según lo descrito anteriormente. Esto está indicado esquemáticamente en la figura 2 como la entrada "Seleccionar función(es) inversa(s) F^1 y parámetro(s) de función(es)" al proceso o dispositivo 12. El bloque "Calcular aproximación de volumen sonoro específico de destino" 14 funciona calculando el volumen sonoro específico, o el volumen sonoro específico parcial, de la señal de audio modificada. Tal volumen sonoro específico, o volumen sonoro específico de la señal de audio no modificada y la aproximación del volumen sonoro específico de destino son usadas por el bloque Calcular parámetros de modificación 10' para obtener los parámetros M de modificación, que, si son aplicados a la señal de audio por el bloque Modificar señal de audio por el bloque funcionado anteriormente, estos parámetros M de modificación, por ejemplo, pueden tomar la forma de ganancias variables en el tiempo aplicadas a las bandas de frecuencia de un banco de filtros, o a los coeficientes de un filtro variable en el tiempo. En el bloque Calcular parámetros de modificación 10", realizaciones

ES 2 400 160 T3

prácticas del bucle de retroalimentación pueden introducir un retardo entre el cálculo y la aplicación de los parámetros M de modificación.

Como se ha mencionado anteriormente, en un entorno de reproducción con una señal secundaria de audio de interferencia, tal como el ruido, cada uno de los bloques Calcular parámetros de modificación 10', Calcular aproximación de volumen sonoro específico de audio no modificado o12, y Calcular aproximación de volumen sonoro específico de destino 14 puede recibir también, como una entrada optativa, una medida de tal señal secundaria de audio de interferencia, o la misma señal de interferencia secundaria, como una de sus entradas, y tanto el proceso o dispositivo 12 como el proceso o dispositivo 14 puede calcular el volumen sonoro específico parcial de la señal de audio modificada. Tales entradas optativas se muestran en la figura 2 usando líneas conductoras discontinuas.

10

15

20

25

30

35

45

60

65

Como se ha mencionado anteriormente, son posibles ejemplos de implementación híbrida de alimentación hacia adelante o retroalimentación de aspectos de la invención. Las figuras 3 y 4 muestran dos ejemplos de tales implementaciones. En los ejemplos de las figuras 3 y 4, como en los ejemplos de las figuras 1 y 2, la señal de audio también es aplicada a un proceso o dispositivo 2 para Modificar señal de audio en un trayecto de señal, pero los bloques para Generar parámetros de modificación (4" en la figura 3 y 4" en la figura 4), en los respectivos trayectos de control, reciben cada uno tanto la señal de audio no modificada como la señal de audio modificada. En ambos ejemplos de las figura 3 y 4, los bloques para Modificar señal de audio 2 y para Generar parámetros de modificación (4" y 4", respectivamente) modifican la señal de audio para reducir la diferencia entre su volumen sonoro específico, que puede estar implícito, y un volumen sonoro específico de destino, que también puede estar implícito.

En el ejemplo de la figura 3, el proceso o dispositivo 4' para Generar parámetros de modificación puede incluir varias funciones y/o dispositivos: un bloque para Calcular volumen sonoro específico de destino 6, como en el ejemplo de la figura 1, un bloque para Calcular aproximación del volumen sonoro específico de destino 14, como en el ejemplo de retroalimentación de la figura 2, y un proceso o dispositivo 10" para "Calcular parámetros de modificación". Como en el ejemplo de la figura 1, en la parte de alimentación hacia adelante de este ejemplo híbrido de alimentación hacia adelante o retroalimentación, el bloque Calcular volumen sonoro específico de destino 6 puede realizar una o más funciones "F", cada una de las cuales puede tener parámetros de función. Esto está indicado esquemáticamente en la figura 3 como una entrada para "Seleccionar función(es) F y parámetro(s) de función(es)" al proceso o dispositivo 6. En la parte de retroalimentación de este ejemplo híbrido de alimentación hacia adelante o retroalimentación, la señal de audio modificada es aplicada a un bloque para Calcular aproximación de volumen sonoro específico de destino 14, como en el ejemplo de retroalimentación de la figura 2. El proceso o dispositivo 14 funciona en el ejemplo de la figura 3 como lo hace en el ejemplo de la figura 2, calculando el volumen sonoro específico o el volumen sonoro específico parcial de la señal de audio modificada. Tal volumen sonoro específico, o volumen sonoro específico parcial, es una aproximación del volumen sonoro específico de destino. El volumen sonoro específico de destino (proveniente del proceso o dispositivo 6) y la aproximación del volumen sonoro específico de destino (proveniente del proceso o dispositivo 14) son aplicados al bloque Cálculo de parámetros de modificación 10" para obtener parámetros M de modificación, que, si son aplicados a la señal de audio por el bloque Modificar señal de audio 2, reducen la diferencia entre el volumen sonoro específico de la señal de audio no modificada y el volumen sonoro específico de destino. Como se ha mencionado anteriormente, estos parámetros M de modificación, por ejemplo, pueden tomar la forma de ganancias variables en el tiempo aplicadas a las bandas de frecuencia de un banco de filtros, o a los coeficientes de un filtro variable en el tiempo. En realizaciones prácticas, el bucle de retroalimentación puede introducir un retardo entre el cálculo y la aplicación de los parámetros M de modificación. Como se ha mencionado anteriormente, en un entorno de reproducción con una señal secundaria de audio de interferencia, tal como el ruido, cada uno de los bloques Calcular parámetros de modificación 10" y Calcular aproximación de volumen sonoro específico de destino 14 también puede recibir, como una entrada optativa, una medida de una tal señal secundaria de ruido de interferencia, o la misma señal secundaria de interferencia, como una de sus entradas, y el proceso o dispositivo 14 puede calcular el volumen sonoro específico parcial de la señal de audio modificada. Las entradas optativas se muestran en la figura 3 usando líneas conductoras discontinuas.

El bloque Calcular parámetros de modificación 10" puede emplear un dispositivo o función de detección de errores, de modo que las diferencias entre su volumen sonoro específico de destino y las entradas de aproximación del volumen sonoro de la especificación de destino ajusten los Parámetros de Modificación a fin de reducir las diferencias entre la aproximación del volumen sonoro específico de destino y el volumen sonoro específico "real" de destino. Tales ajustes reducen las diferencias entre el volumen sonoro específico de la señal de audio no modificada y el volumen sonoro específico de destino, que puede estar implícito. Así, los parámetros M de modificación pueden ser actualizados en base a un error entre el volumen sonoro específico de destino, calculado en el trayecto de alimentación hacia adelante, a partir del volumen sonoro específico del audio original, usando la función F, y la aproximación del volumen sonoro específico de destino, calculada en el trayecto de retroalimentación a partir del volumen sonoro específico, o volumen sonoro específico parcial, del audio modificado.

En el ejemplo de la figura 4, se muestra un ejemplo alternativo de alimentación hacia adelante o retroalimentación. Esta alternativa difiere del ejemplo de la figura 3 en cuanto a que la función o funciones inversa(s) F^1 se calcula(n) en el trayecto de retroalimentación, en lugar de que la función, o funciones, F sea(n) calculada(s) en el trayecto de alimentación hacia adelante. En el ejemplo de la figura 4, el proceso o dispositivo 4' para Generar parámetros de modificación puede incluir varias funciones y/o dispositivos: un bloque para Calcular volumen sonoro específico 8, como en el ejemplo de alimentación hacia adelante de la figura 1, el bloque Calcular aproximación de volumen sonoro

específico de audio no modificado 12, como en el ejemplo de retroalimentación de la figura 2, y un bloque para Calcular parámetros de modificación 10"". El bloque para Calcular el volumen sonoro específico 8, como en el ejemplo de alimentación hacia adelante de la figura 1, proporciona, como entrada al bloque para Calcular parámetros de modificación 10"", el volumen sonoro específico de la señal de audio no modificada. Como en el ejemplo de retroalimentación de la figura 2, con la restricción de que la función, o funciones, F sea(n) invertible(s), el proceso o dispositivo 12 estima el volumen sonoro específico de la señal de audio no modificada aplicando la función inversa F^1 al volumen sonoro específico, o al volumen sonoro específico parcial, de la señal de audio modificada. Una entrada para "Seleccionar función(es) inversa(s) y parámetro(s) de función(es) inversa(s))", para el bloque Calcular aproximación de volumen sonoro específico de audio no modificado 12, se muestra para indicar que el dispositivo o proceso 12 puede calcular una función inversa F^1 , según lo descrito anteriormente. Esto está indicado esquemáticamente en la figura 4 como una entrada "Seleccionar función(es) inversa(s) F^1 y parámetro(s) de función(es)" al proceso o dispositivo 12. De esa manera, el proceso o dispositivo 2 proporciona, como otra entrada al bloque para Calcular parámetros de modificación 10", una aproximación al volumen sonoro específico de la señal de audio no modificada.

10

30

35

40

45

50

55

60

65

Como en los ejemplos de las figuras 1 a 3, el bloque Calcular parámetros de modificación 10" obtiene parámetros M de 15 modificación, que, si son aplicados a la señal de audio por el bloque Modificar señal de audio 2, reducen la diferencia entre el volumen sonoro específico de la señal de audio no modificada y el volumen sonoro específico de destino, que está implícito en este ejemplo. Como se ha mencionado anteriormente, los parámetros M de modificación, por ejemplo, pueden tomar la forma de ganancias variables en el tiempo aplicadas a las bandas de frecuencia de un banco de filtros o 20 a los coeficientes de un filtro variable en el tiempo. En realizaciones prácticas, el bucle de retroalimentación puede introducir un retardo entre el cálculo y la aplicación de los parámetros M de modificación. Como se ha mencionado anteriormente, en un entorno de reproducción con una señal secundaria de audio de interferencia, tal como el ruido, tanto el bloque para Calcular parámetros de modificación 10" como el bloque para Calcular aproximación del volumen sonoro específico del audio no modificado 12 puede recibir también, como una entrada optativa, una medida de tal señal 25 secundaria de audio de interferencia, o la misma señal secundaria de interferencia, como una de sus entradas, y el proceso o dispositivo 12 puede calcular el volumen sonoro específico parcial de la señal de audio modificada. Las entradas optativas se muestran en la figura 4 usando líneas conductoras discontinuas.

El bloque Calcular parámetros de modificación 10" puede emplear un dispositivo o función de detección de errores, de modo que las diferencias entre su volumen sonoro específico y las entradas de aproximación del volumen sonoro específico produzcan salidas que ajusten los Parámetros de Modificación a fin de reducir las diferencias entre la aproximación del volumen sonoro específico y el volumen sonoro específico "real". Debido a que la aproximación del volumen sonoro específico parcial, del audio modificado, que puede verse como una aproximación del volumen sonoro específico de destino, tales ajustes reducen las diferencias entre el volumen sonoro específico de la señal de audio modificada y el volumen sonoro específico de destino, que es inherente en la función o funciones F^1 . De ese modo, los parámetros M de modificación pueden ser actualizados en base a un error entre el volumen sonoro específico, calculado en el trayecto de alimentación hacia adelante desde el audio original, y la aproximación del volumen sonoro específico calculado, usando la función o funciones inversa(s) F^1 , en el trayecto de retroalimentación a partir del volumen sonoro específico, o del volumen sonoro específico parcial, del audio modificado. Debido al trayecto de retroalimentación, las implementaciones prácticas pueden introducir un retardo entre la actualización y la aplicación de los parámetros de modificación.

Aunque los parámetros M de modificación en los ejemplos de las figuras 1 a 4, cuando se aplican a un proceso o dispositivo 2 para Modificar señal de audio, reducen la diferencia entre el volumen sonoro específico de la señal de audio y el volumen sonoro específico de destino, en las realizaciones prácticas los correspondientes parámetros de modificación producidos en respuesta a la misma señal de audio pueden no ser idénticos entre sí.

Aunque no es crítico ni esencial para los aspectos de la presente invención, el cálculo del volumen sonoro específico de la señal de audio, o de la señal de audio modificada, puede emplear ventajosamente técnicas estipuladas en dicha solicitud internacional de patente nº PCT/US2004/016964, publicada como el documento WO 2004/111964 A2, en la cual el cálculo selecciona, entre un grupo de dos o más funciones del modelo de volumen sonoro específico, una, o una combinación de dos o más de las funciones del modelo de volumen sonoro específico, cuya selección está controlada por la medida de las características de la señal de audio de entrada. La descripción del Volumen sonoro específico 104 de la figura 1, más adelante, describe una disposición de ese tipo.

De acuerdo a aspectos adicionales de la invención, la señal de audio no modificada y (1) los parámetros de modificación,

o bien (2) el volumen sonoro específico de destino, o una representación del volumen sonoro específico de destino (p. ej., los factores de escala utilizables para calcular, explícitamente o implícitamente, el volumen sonoro específico de destino), pueden ser almacenados o transmitidos para su uso, por ejemplo, en un dispositivo o proceso temporalmente y/o espacialmente separado. Los parámetros de modificación, el volumen sonoro específico de destino, o la representación del volumen sonoro específico de destino, pueden ser determinados de cualquier manera adecuada, como, por ejemplo, en uno de los ejemplos de disposiciones de alimentación hacia adelante, de retroalimentación y de alimentación híbrida hacia adelante o retroalimentación de las figuras 1 a 4, según lo descrito anteriormente. En la práctica, una disposición de alimentación hacia adelante, tal como en el ejemplo de la figura 1, es el menos complejo y el

más rápido, en cuanto evita cálculos basados en la señal de audio modificada. Un ejemplo de transmisión o almacenamiento del audio no modificado y de los parámetros de modificación se muestra en la figura 5, mientras que un

ES 2 400 160 T3

ejemplo de la transmisión o almacenamiento del audio no modificado y del volumen sonoro específico de destino, o de una representación del volumen sonoro específico de destino, se muestra en la figura 6.

Una disposición tal como la del ejemplo de la figura 5 puede ser usada para separar temporalmente y/o espacialmente la aplicación de los parámetros de modificación a la señal de audio de la generación de tales parámetros de modificación. Una disposición tal como la del ejemplo de la figura 6 puede ser usada para separar temporalmente y/o espacialmente, tanto la generación como la aplicación de los parámetros de modificación, de la generación del volumen sonoro específico de destino, o de una representación del mismo. Ambos tipos de disposiciones posibilitan una disposición sencilla, de bajo coste, de reproducción o de recepción, que evite la complejidad de generar los parámetros de modificación, o la de generar el volumen sonoro específico de destino. Aunque una disposición del tipo de la figura 5 es más sencillo que una disposición del tipo de la figura 6, la disposición de la figura 6 tiene la ventaja de que la información requerida para ser almacenada o transmitida puede ser mucho menos, especialmente cuando una representación del volumen sonoro específico de destino, tal como uno o más factores de escala, se almacenan o transmiten. Tal reducción en el almacenamiento o transmisión de información puede ser especialmente útil en entornos de audio de baja velocidad de bits.

10

15

20

25

30

35

40

45

50

55

60

65

En consecuencia, aspectos adicionales de la presente invención son la provisión de un dispositivo o proceso (1) que recibe o reproduce, desde un dispositivo o proceso de almacenamiento o transmisión, parámetros M de modificación, y los aplica a una señal de audio que también es recibida, o bien (2) que recibe o reproduce, a partir de un dispositivo o proceso de almacenamiento o transmisión, un volumen sonoro específico de destino, o una representación de un volumen sonoro específico de destino, genera parámetros M de modificación aplicando el volumen sonoro específico de destino, o una representación del mismo, a la señal de audio que también es recibida (o a una medida de la señal de audio, tal como su volumen sonoro específico, que puede ser obtenida de la señal de audio), y aplica los parámetros M de modificación a la señal de audio recibida. Tales dispositivos o procesos pueden ser caracterizados como procesos de descodificación o descodificadores; mientras que los dispositivos o procesos requeridos para producir la información almacenada o transmitida pueden ser caracterizados como procesos de codificación o codificadores. Tales procesos de codificación, o codificadores, son aquellas partes de los ejemplos de disposición de las figuras 1 a 4 que son utilizables para producir la información requerida por los respectivos procesos de descodificación, o descodificadores. Tales procesadores de descodificación, o descodificadores, pueden estar asociados a, o ser operativos con, prácticamente cualquier tipo de proceso o dispositivo que procese y/o reproduzca el sonido.

En un aspecto de la invención, como en el ejemplo de la figura 5, la señal de audio no modificada y los parámetros M de modificación producidos, por ejemplo, por un proceso generador, o un generador, de parámetros de modificación, tal como el bloque Generar parámetros de modificación 4 de la figura 1, 4' de la figura 2, 4" de la figura 3 o 4" de la figura 4, pueden ser aplicados a cualquier dispositivo o función adecuado de almacenamiento o transmisión ("Almacenar o transmitir") 16. En el caso de usar el ejemplo de alimentación hacia adelante de la figura 1 como un proceso de codificación, o un codificador, el bloque para Modificar la señal de audio 2 no sería requerido para generar el audio modificado, y podría ser omitido si no hay ningún requisito para proporcionar el audio modificado en la ubicación temporal o espacial del codificador, o proceso de codificación. El bloque Almacenar o transmitir 16 puede incluir, por ejemplo, dispositivos adecuados cualesquiera, magnéticos, ópticos o de estado sólido, de almacenamiento y reproducción, o dispositivos adecuados cualesquiera, cableados o inalámbricos, de transmisión y recepción, no siendo crítica la elección de los mismos para la invención. Los parámetros de modificación reproducidos o recibidos pueden ser aplicados luego a un bloque para Modificar señal de audio 2, del tipo empleado en los ejemplos de las figuras 1 a 4, a fin de modificar la señal de audio reproducida o recibida, de modo que su volumen sonoro específico se aproxime al volumen sonoro específico de destino de, o inherente a, la disposición en la cual se obtuvieron los parámetros de modificación. Los parámetros de modificación pueden ser almacenados o transmitidos de una cualquiera entre diversas maneras. Por ejemplo, pueden ser almacenados o transmitidos como metadatos que acompañan a la señal de audio, pueden ser enviados en trayectos o canales separados, pueden ser codificados esteganográficamente en el audio, pueden ser multiplexados, etc. El uso de los parámetros de modificación para modificar la señal de audio puede ser optativo y, si es optativo, su uso puede ser seleccionable, por ejemplo, por un usuario. Por ejemplo, los parámetros de modificación, si son aplicados a la señal de audio, podrían reducir la gama dinámica de la señal de audio. El emplear o no tal reducción dinámica de gama podría ser seleccionable por un usuario.

En otro aspecto de la invención, como en el ejemplo de la figura 6, la señal de audio no modificada y el volumen sonoro específico de destino, o la representación del volumen sonoro específico de destino, pueden ser aplicados a cualquier dispositivo o función adecuado de almacenamiento o transmisión ("Almacenar o transmitir") 16. En el caso de usar una configuración de alimentación hacia adelante, tal como el ejemplo de la figura 1, como un proceso de codificación o un codificador, no se requeriría ni un proceso o dispositivo del tipo del bloque Calcular parámetros de modificación 10, ni un proceso o dispositivo del tipo del bloque Modificar señal de audio 2, y podría omitirse si no hay ningún requisito de proporcionar ya sea los parámetros de modificación o bien el audio modificado en la ubicación temporal o espacial del codificador o proceso de codificación. Como en el caso del ejemplo de la figura 5, el bloque Almacenar o transmitir 16 puede incluir, por ejemplo, dispositivos adecuados cualesquiera, magnéticos, ópticos o de estado sólido, de almacenamiento y reproducción, o dispositivos adecuados cualesquiera, cableados o inalámbricos, de transmisión y recepción, no siendo crítica la elección de los mismos para la invención. El volumen sonoro específico de destino reproducido o recibido, o la representación del volumen sonoro específico de destino, puede ser aplicado luego, junto con el audio no modificado, a un bloque para Calcular parámetros de modificación 10, del tipo empleado en el ejemplo de la

figura 1, o a un bloque para Calcular parámetros de modificación 10", del tipo empleado en el ejemplo de la figura 3, a fin de proporcionar parámetros M de modificación que puedan ser luego aplicados al bloque para Modificar señal de audio 2, del tipo empleado en los ejemplos de las figuras 1 a 4, a fin de modificar la señal de audio reproducida o recibida, de modo que su volumen sonoro específico se aproxime al volumen sonoro específico de destino de, o inherente a, la disposición en la cual los parámetros de modificación fueron obtenidos. Aunque el volumen sonoro específico de destino, o la representación del mismo, puede ser obtenida de manera inmediata, en un proceso de codificación, o codificador, del tipo del ejemplo de la figura 1, el volumen sonoro específico de destino, o la representación del mismo, o una aproximación al volumen sonoro específico de destino, o una representación de la misma, pueden ser obtenidos en un proceso de codificación, o codificador, de los tipos de los ejemplos desde la figura 2 hasta la figura 4 (las aproximaciones son calculadas en los procesos o dispositivos 14 de las figuras 2 y 3, y en el proceso o dispositivo 12 de la figura 4). El volumen sonoro específico de destino, o la representación del mismo, puede ser almacenado o transmitido de cualquiera entre diversas maneras. Por ejemplo, puede ser almacenado o transmitido como metadatos que acompañan a la señal de audio, puede ser enviado en trayectos o canales separados, puede ser codificado esteganográficamente en el audio, puede ser multiplexado, etc. El uso de los parámetros de modificación obtenidos del volumen sonoro específico de destino, almacenado o transmitido, o de una representación para modificar la señal de audio, puede ser optativo y, si es optativo, su uso puede ser seleccionable, por ejemplo, por un usuario. Por ejemplo, los parámetros de modificación, si son aplicados a la señal de audio, podrían reducir la gama dinámica de la señal de audio. El emplear o no tal reducción de gama dinámica podría ser seleccionable por un usuario.

10

15

25

Al implementar la invención divulgada como un sistema digital, una configuración de alimentación hacia adelante es la más práctica, y los ejemplos de tales configuraciones, por lo tanto, se describen más adelante en detalle, entendiéndose que el alcance de la invención no está limitado por ello.

A lo largo de este documento, los términos tales como "filtro" o "banco de filtros" se usan en el presente documento para incluir esencialmente cualquier forma de filtrado recursivo y no recursivo, tal como los filtros IIR o transformadores, y la información "filtrada" es el resultado de aplicar tales filtros. Las realizaciones descritas más adelante emplean bancos de filtros implementados por transformaciones.

La figura 7 ilustra mayores detalles de una realización ejemplar de un aspecto de la invención, realizada en una 30 disposición de alimentación hacia adelante. El audio atraviesa primero una función o dispositivo de banco de filtros de análisis ("Banco de filtros de análisis") 100, que divide la señal de audio en una pluralidad de bandas de frecuencia (por tanto, la figura 5 muestra múltiples salidas del Banco 100 de filtros de análisis, representando cada salida una banda de frecuencia, cuya salida lleva, a través de las diversas funciones o dispositivos, hasta un banco de filtros de síntesis, que suma las bandas en una señal combinada de banda ancha, según se describe más adelante). La respuesta del filtro asociado a cada banda de frecuencia en el Banco 100 de filtros de análisis está diseñada para simular la respuesta en 35 una ubicación específica de la membrana basilar en el oído interno. La salida de cada filtro en el Banco 100 de filtros de análisis pasa a continuación a un filtro de transmisión, o función de filtro de transmisión ("Filtro de transmisión") 101 que simula el efecto de filtrado de la transmisión del audio a través del oído externo y medio. Si solamente el volumen sonoro del audio hubiera de medirse, el filtro de transmisión podría ser aplicado antes del banco de filtros de análisis, pero, 40 debido a que las salidas del banco de filtros de análisis son usadas para sintetizar el audio modificado, es ventajoso aplicar el filtro de transmisión después del banco de filtros. Las salidas del Filtro 101 de transmisión pasan a continuación a una función o dispositivo de excitación ("Excitación") 102, cuyas salidas simulan la distribución de energía a lo largo de la membrana basilar. Los valores de energía de excitación pueden ser allanados a lo largo del tiempo por una función o dispositivo de allanamiento ("Allanamiento") 103. Las constantes temporales de la función de allanamiento se fijan de 45 acuerdo a los requisitos de una aplicación deseada. Las señales de excitaciones allanadas son subsiguientemente convertidas en volumen sonoro específico en una función o dispositivo de volumen sonoro específico ("Volumen sonoro específico (SL)") 104. El volumen sonoro específico está representado en unidades de sone por frecuencia unitaria. El componente de volumen sonoro específico asociado a cada banda pasa a una función o dispositivo de modificación de volumen sonoro específico ("Modificación de SL") 105. El bloque Modificación de SL 105 toma como su entrada el 50 volumen sonoro específico original y luego emite un volumen sonoro específico deseado o "de destino", que, según un aspecto de la presente invención, es preferiblemente una función del volumen sonoro específico original (véase el próximo encabezamiento más adelante, titulado "Volumen sonoro específico de destino"). El bloque Modificación de SL 105 puede operar independientemente en cada banda, o puede existir una interdependencia entre dos o más bandas (un allanamiento de frecuencia según lo sugerido por las líneas interconectoras en la figura 7), según el efecto deseado. 55 Tomando como sus entradas los componentes de bandas de frecuencia de excitación allanada, provenientes del bloque Excitación 102, y el volumen sonoro específico de destino proveniente del bloque Modificación de SL 105, una función o dispositivo discriminador de ganancia ("Discriminador de ganancia") 106 determina la ganancia que debe ser aplicada a cada banda de la salida del Banco 100 de filtros de análisis a fin de transformar el volumen sonoro específico medido en el volumen sonoro específico de destino. El Discriminador de ganancia puede ser implementado de diversas maneras. Por ejemplo, el Discriminador de ganancia puede incluir un proceso iterativo, tal como al estilo del divulgado en dicha 60 solicitud internacional de patente nº PCT/US2004/016964, publicada como el documento WO 2004/111964 A2 o, alternativamente, una tabla de búsqueda. Aunque las ganancias por banda generadas por el Discriminador 106 de ganancia pueden ser allanadas adicionalmente a lo largo del tiempo por una función o dispositivo optativo de allanamiento ("Allanamiento") 107, a fin de minimizar las distorsiones perceptivas, se prefiere que el allanamiento temporal sea aplicado en otras partes en el proceso o dispositivo global, según lo descrito en otra parte. Finalmente, las 65 ganancias son aplicadas a las respectivas bandas del Banco 100 de filtros de análisis, a través de una respectiva función

combinadora o combinador 108 multiplicativo, y el audio procesado o "modificado" es sintetizado a partir de las bandas modificadas por ganancia en una función o dispositivo de banco de filtros de síntesis ("Banco de filtros de síntesis") 110. Además, las salidas del banco de filtros de análisis pueden ser retardadas por una función o dispositivo de retardo ("Retardo") 109 antes de la aplicación de las ganancias, a fin de compensar cualquier latencia asociada al cálculo de ganancias. Alternativamente, en lugar de calcular ganancias para su uso en la aplicación de modificaciones de ganancias en bandas de frecuencia, los Discriminadores 106 de ganancias pueden calcular coeficientes de filtro que controlan un filtro variable en el tiempo, tal como un filtro FIR de múltiples tomas o un filtro IIR multipolar. Para simplificar la exposición, los aspectos de la invención son descritos principalmente como empleadores de factores de ganancia aplicados a bandas de frecuencia, entendiéndose que los coeficientes de filtro y los filtros variables en el tiempo también pueden ser empleados en realizaciones prácticas.

En las realizaciones prácticas, el procesamiento del audio puede ser efectuado en el dominio digital. En consecuencia, la señal de entrada de audio está indicada por la secuencia temporal discreta x[n], que ha sido muestreada a partir del origen de audio en alguna frecuencia f_s de muestreo. Se supone que la secuencia x[n] ha sido adecuadamente ajustada a escala de modo que la potencia de RMS (Raíces Cuadradas Mínimas) de x[n] en decibelios, dada por

$$RMS_{dB} = 10\log_{10}\left(\frac{1}{L}\sum_{n=0}^{L}x^{2}[n]\right)$$

10

15

25

30

40

sea igual al nivel de presión sonora en dB, en el cual el audio está siendo oído por un oyente humano. Además, se supone que la señal de audio es monofónica, para simplificar la exposición.

El Banco 100 de filtros de análisis, el Filtro 101 de transmisión, la Excitación 12, el Volumen sonoro específico 104, la Modificación 105 del volumen sonoro específico, el Discriminador 106 de ganancia y el Banco 110 de filtros de síntesis pueden ser descritos en mayor detalle de la siguiente manera.

Banco 100 de filtros de análisis

La señal de entrada de audio es aplicada a un banco de filtros de análisis, o función de banco de filtros ("Banco de filtros de análisis") 100. Cada filtro en el Banco 100 de filtros de análisis está diseñado para simular la respuesta de frecuencia en una ubicación específica a lo largo de la membrana basilar en el oído interno. El Banco 100 de filtros puede incluir un conjunto de filtros lineales cuyo ancho de banda y espaciado sean constantes en la escala de frecuencia del Ancho de Banda Rectangular Equivalente (ERB), según lo definido por Moore, Glasberg y Baer (B. C. J. Moore, B. Glasberg, T. Baer, "Un modelo para la predicción de umbrales, volumen sonoro y volumen sonoro parcial", *supra*).

Aunque la escala de frecuencia ERB coincide más estrechamente con la percepción humana y muestra prestaciones mejoradas al producir mediciones objetivas de volumen sonoro que coinciden con resultados subjetivos de volumen sonoro, la escala de frecuencias de Bark puede ser empleada con prestaciones reducidas.

Para una frecuencia central f en Hercios, el ancho de una banda ERB en hercios puede ser aproximado como:

$$ERB(f) = 24.7(4.37 f/1000 + 1)$$
 (1)

A partir de esta relación, se define una escala de frecuencias distorsionada, de modo que en cualquier punto a lo largo de la escala distorsionada, el ERB correspondiente, en unidades de la escala distorsionada, sea igual a uno. La función para convertir la frecuencia lineal en hercios a esta escala de frecuencias de ERB se obtiene integrando la recíproca de la Ecuación 1:

$$HzAERB(f) = \int \frac{1}{24.7(4.37f/1000+1)} df = 21.4\log_{10}(4.37f/1000+1) \quad (2a)$$

También es útil expresar la transformación desde la escala ERB de vuelta a la escala de frecuencia lineal, despejando *f* en la Ecuación 2a.

ERBAHz(e) =
$$f = \frac{1000}{4.37} 10^{(e/21.4-1)}$$
, (2b)

donde e está en unidades de la escala de ERB. La figura 9 muestra la relación entre la escala de ERB y la frecuencia en hercios.

El Banco 100 de filtros de análisis puede incluir *B* filtros auditivos, denominados bandas, en las frecuencias centrales $f_c[1]...f_c[B]$, uniformemente espaciados a lo largo de la escala de ERB. Más específicamente,

$$f_{c}[1] = f_{min} \tag{3a}$$

$$f_c[b] = f_c[b-1] + ERBAHz(HzAERB(f_c[b-1]) + \Delta)$$
 $b = 2...B$ (3b)

$$f_c[B] < f_{max}$$
, (3c)

- donde Δ es el espaciado deseado de ERB del Banco 100 de filtros de análisis, y donde f_{min} y f_{max} son, respectivamente, las frecuencias centrales mínimas y máximas. Se puede escoger Δ = 1, y teniendo en cuenta la gama de frecuencias en las cuales el oído humano es sensible, se puede fijar f_{min} = 50 Hz y f_{max} = 20.000 Hz. Con tales parámetros, por ejemplo, la aplicación de las Ecuaciones 3a a 3c produce B = 40 filtros auditivos.
- 10 La respuesta de frecuencia de magnitud de cada filtro auditivo puede ser caracterizada por una función exponencial redondeada, según lo sugerido por Moore y Glasberg. Específicamente, la respuesta de magnitud de un filtro con frecuencia central f_c[b] puede calcularse como:

$$H_b(f) = (1 + pg)e^{-pg}$$
 (4a)

15 donde

20

35

40

45

$$g = \left| \frac{f - f_c[b]}{f_c[b]} \right|, \tag{4b}$$

$$p = \frac{4f_c[b]}{ERB(f_c[b])} \tag{4c}$$

Las respuestas de magnitud de tales filtros auditivos *B*, que aproximan la formación de bandas críticas en la escala de ERB, se muestran en la figura 10.

Las operaciones de filtrado del Banco 100 de filtros de análisis pueden ser adecuadamente aproximadas usando una Transformación Discreta de Fourier de longitud finita, usualmente denominada la Transformación Discreta de Fourier de Corto Plazo (STDFT), porque se cree que una implementación que hace funcionar los filtros a la velocidad de muestreo de la señal de audio, denominada una implementación a velocidad completa, proporciona más resolución temporal de lo necesario para mediciones precisas del volumen sonoro. Al usar la STDFT en lugar de una implementación a velocidad completa, pueden lograrse una mejora en la eficacia y una reducción en la complejidad de cálculo.

La STDFT de la señal x[n] de audio de entrada está definida como:

$$X[k,t] = \sum_{n=0}^{N-1} w[n] x[n+tT] e^{-j\frac{2\pi k}{N}},$$
 (5a)

donde k es el índice de frecuencia, t es el índice de bloque temporal, N es el tamaño de la DFT, T es el tamaño del salto y w[n] es una ventana de longitud N, normalizada de modo que

$$\sum_{n=0}^{N-1} w^2[n] = 1 \tag{5b}$$

Obsérvese que la variable t en la Ecuación 5a es un índice discreto que representa el bloque temporal de la STDFT, en contraste con una medida del tiempo en segundos. Cada incremento en t representa un salto de T muestras a lo largo de la señal x[n]. Las referencias subsiguientes al índice t suponen esta definición. Si bien pueden usarse distintas configuraciones de parámetros y formas de ventana, según los detalles de la implementación, para $f_s = 44.100$ Hz, escogiendo N = 2.048, T = 1.024, y haciendo que w[n] sea una ventana de Hanning, se proporciona un equilibrio

adecuado del tiempo y de la resolución de la frecuencia. La STDFT descrita anteriormente puede ser más eficaz usando la Transformación Rápida de Fourier (FFT).

En lugar de la STDFT, la Transformación de Coseno Discreta Modificada (MDCT) puede ser utilizada para implementar el banco de filtros de análisis. La MDCT es una transformación usualmente usada en codificadores de audio perceptivos, tales como el Dolby AC-3. Si el sistema divulgado es implementado con tal audio perceptivamente codificado, la medición y modificación divulgadas del volumen sonoro pueden ser implementadas más eficazmente procesando los coeficientes de MDCT existentes del audio codificado, eliminando por ello la necesidad de realizar la transformación del banco de filtros de análisis. La MDCT de la señal x[n] de audio de entrada está dada por:

$$X[k,t] = \sum_{n=0}^{N-1} w[n]x[n+tT]\cos((2\pi/N)(k+1/2)(n+n_0)),$$

donde

10

25

30

35

50

55

$$n_0 = \frac{(N/2) + 1}{2} \tag{6}$$

Generalmente, el tamaño de salto T se escoge para que sea exactamente la mitad de la longitud N de la transformación, de modo que la reconstrucción perfecta de la señal x[n] sea posible.

20 Filtro 101 de Transmisión

Las salidas del Banco 100 de filtros de análisis se aplican a un filtro de transmisión, o función de filtro de transmisión, ("Filtro de transmisión") 101 que filtra cada banda del banco de filtros de acuerdo a la transmisión de audio a través del oído externo y medio. La figura 8 ilustra una respuesta de frecuencia de magnitud adecuada del filtro de transmisión, P(f), sobre la gama de frecuencia audible. La respuesta es de una unidad por debajo de 1 kHz y, por encima de 1 kHz, sigue la inversa del umbral del oído, según lo especificado en el estándar ISO226, con el umbral normalizado para que sea igual a la unidad a 1 kHz.

Excitación 102

A fin de calcular el volumen sonoro de la señal de audio de entrada, se necesita una medida de la energía a corto plazo de las señales de audio en cada filtro del Banco 100 de filtros de análisis después de la aplicación del Filtro 101 de transmisión. Esta medida, variable en cuanto al tiempo y a la frecuencia, se denomina la excitación. La salida de energía a corto plazo de cada filtro en el Banco 100 de filtros de análisis puede ser aproximada en la Función de excitación 102, mediante la multiplicación de respuestas de filtro en el dominio de frecuencia, con el espectro de potencia de la señal de entrada:

$$E[b,t] = \frac{1}{N} \sum_{k=0}^{N-1} |H_b[k]|^2 |P[k]|^2 |X[k,t]|^2, \tag{7}$$

donde *b* es el número de banda, *t* es el número de bloque y H_b[k] y P[k] son las respuestas de frecuencia del filtro auditivo y del filtro de transmisión, respectivamente, muestreados a una frecuencia correspondiente al índice k de compartimiento de la STDFT o la MDCT. Debería observarse que pueden usarse en la Ecuación 7 formas para la respuesta de magnitud de los filtros auditivos distintas a la especificada en las Ecuaciones 4a a 4c, para lograr resultados similares. Por ejemplo, dicha Solicitud Internacional Nº PCT/US2004/016964, publicada como el documento WO 2004/111964 A2, describe dos alternativas: un filtro auditivo caracterizado por una función de transferencia IIR de 12º orden, y una aproximación de paso de banda de "pared de ladrillos" de bajo coste.

En resumen, la salida de la Función 102 de excitación es una representación del dominio de frecuencia de la energía *E* en las respectivas bandas *b* de ERB por periodo *t* de tiempo.

Promedio temporal ("Allanamiento") 103

Para ciertas aplicaciones de la invención divulgada, según se describe más adelante, puede ser deseable allanar la excitación *E[b,t]* antes de su transformación en un volumen sonoro específico. Por ejemplo, el allanamiento puede ser realizado recursivamente en la función 103 de Allanamiento según la ecuación:

$$\overline{E}[b,t] = \lambda_b \overline{E}[b,t] + (1 - \lambda_b) E[b,t], \tag{8}$$

donde las constantes temporales λ_b en cada banda b están seleccionadas de acuerdo a la aplicación deseada. En la mayoría de los casos, las constantes temporales pueden ser escogidas ventajosamente para que sean proporcionales al tiempo de integración de la percepción humana del volumen sonoro dentro de la banda b. Watson y Gengel realizaron experimentos que demuestran que este tiempo de integración está dentro de la gama de entre 150 y 175 ms a bajas frecuencias (125 a 200 Hz) y entre 40 y 60 ms a altas frecuencias (Charles S. Watson y Roy W. Gengel, "Duración de señal y frecuencia de señal con relación a la sensibilidad auditiva", Revista de la Sociedad Acústica de América, Vol. 46, N° 4 (Parte 2), 1969, págs. 989 a 997).

10 Volumen sonoro específico 104

5

15

20

25

30

35

En el convertidor, o función de conversión, del volumen sonoro específico ("Volumen sonoro específico") 104, cada banda de frecuencia de la excitación es convertida en un valor componente del volumen sonoro específico, que es medido en sone por ERB.

Inicialmente, al calcular el volumen sonoro específico, el nivel de excitación en cada banda de E[b,t] puede ser transformado en un nivel equivalente de excitación a 1 kHz, según lo especificado por los perfiles de igual volumen sonoro del estándar ISO 226 (figura 11), normalizados por el filtro P(z) de transmisión (figura 12):

$$\overline{E}_{1kll_k}[b,t] = T_{1kll_k}(\overline{E}[b,t], f_c[b]), \qquad (9)$$

donde $T_{1kHz}(E,f)$ es una función que genera el nivel a 1 kHz, que es igualmente alto al nivel E en la frecuencia f. En la práctica, $T_{1kHz}(E,f)$ se implementa como una interpolación de una tabla de búsqueda de los perfiles de igual volumen sonoro, normalizados por el filtro de transmisión. La transformación a niveles equivalentes a 1 kHz simplifica el siguiente cálculo de volumen sonoro específico.

Luego, el volumen sonoro específico en cada banda puede ser calculado como:

$$N[b,t] = \alpha[b,t]N_{NB}[b,t] + (1-\alpha[b,t])N_{NB}[b,t], \qquad (10)$$

donde $N_{NB}[b,t]$ y $N_{WB}[b,t]$ son valores de volumen sonoro específico basados en un modelo de señales de banda estrecha y de banda ancha, respectivamente. El valor $\alpha[b,t]$ es un factor de interpolación situado entre 0 y 1, que se calcula a partir de la señal de audio. Dicha Solicitud Internacional N° PCT/US2004/016964, publicada como el documento WO 2004/111964 A2, describe una técnica para calcular $\alpha[b,t]$ a partir de la uniformidad espectral de la excitación. También describe los modelos de señal de "banda estrecha" y de "banda ancha" en mayor detalle.

Los valores $N_{NB}[b,t]$ y $N_{WB}[b,t]$ de volumen sonoro específico de banda estrecha y de banda ancha pueden ser estimados a partir de la excitación transformada, usando las funciones exponenciales:

$$N_{NB}[b,t] = \begin{cases} G_{NB}\left(\left(\frac{\overline{E}_{1kHz}[b,t]}{TQ_{1kHz}}\right)^{a_{NB}} - 1\right), & \overline{E}_{1kHz}[b,t] > TQ_{11kHz} \\ 0, & \text{en otros casos} \end{cases}$$
 (11a)

$$N_{WB}[m,t] = \begin{cases} G_{WB} \left(\left(\frac{\overline{E}_{1kHz}[b,t]}{TQ_{1kHz}} \right)^{b_{WB}} - 1 \right), & \overline{E}_{1kHz}[b,t] > TQ_{1kHz}, \\ 0, & \text{en otros casos} \end{cases}$$
(11b)

donde *TQ_{1kHz}* es el nivel de excitación en silencio para un tono de 1 kHz. A partir de los perfiles de igual volumen sonoro (figuras 11 y 12), *TQ_{1kHz}* es igual a 4,2 dB. Se observa que ambas funciones de volumen sonoro específico son iguales a cero cuando la excitación es igual al umbral en silencio. Para excitaciones mayores que el umbral en silencio, ambas funciones crecen monótonamente con una ley de potencia de acuerdo a la ley de Stevens de la sensación de intensidad.
 El exponente para la función de banda estrecha se escoge mayor que el de la función de banda ancha, haciendo que la función de banda estrecha aumente más rápidamente que la función de banda ancha. La selección específica de los exponentes β y las ganancias G para los casos de banda estrecha y de banda ancha se escogen para que coincidan con los datos experimentales sobre el crecimiento del volumen sonoro para los tonos y el ruido.

Moore y Glasberg sugieren que el volumen sonoro específico debería ser igual a algún valor pequeño, en lugar de cero, cuando la excitación está en el umbral del oído. El volumen sonoro específico debería disminuir luego monótonamente hasta el cero, según la excitación disminuye hasta el cero. La justificación es que el umbral del oído es un umbral probabilístico (el punto en el cual un tono es detectado el 50% del tiempo), y que un cierto número de tonos, cada uno en el umbral, presentados juntos, pueden sumarse en un sonido que es más audible que cualquiera de los tonos individuales. En la solicitud divulgada, el aumento de las funciones de volumen sonoro específico con esta propiedad tiene el beneficio añadido de hacer que el discriminador de ganancia, expuesto más adelante, se comporte más adecuadamente cuando la excitación está cerca del umbral. Si el volumen sonoro específico está definido para que sea cero cuando la excitación está en, o por debajo de, el umbral, entonces no existe una solución única para el discriminador de ganancia, para excitaciones en, o por debajo de, el umbral. Si, por otra parte, el volumen sonoro específico está definido para que sea monótonamente creciente para todos los valores de excitación mayores que, o iguales a, cero, según lo sugerido por Moore y Glasberg, entonces existe una única solución. El ajuste a escala mayor que la unidad del volumen sonoro específico en las Ecuaciones 11a y 11b pueden ser alteradas para que tengan la propiedad deseada, según:

$$N_{NB}[b,t] = \begin{cases} G_{NB} \left(\frac{\overline{E}_{1kHz}[b,t]}{TQ_{1kHz}} \right)^{\rho_{NB}} - 1 \\ \exp \left\{ K_{NB} \left(-\log \left(\frac{\overline{E}_{1kHz}[b,t]}{TQ_{1kHz}} \right) + C_{NB} \right)^{\eta_{NB}} \right\}, & \overline{E}_{1kHz}[b,t] > \lambda TQ_{1kHz} \end{cases}$$

$$N_{NB}[m,t] = \begin{cases} G_{NB} \left(\frac{\overline{E}_{1kHz}[b,t]}{TQ_{1kHz}} \right)^{\rho_{NB}} - 1 \\ \exp \left\{ K_{NB} \left(-\log \left(\frac{\overline{E}_{1kHz}[b,t]}{TQ_{1kHz}} \right) + C_{NB} \right)^{\eta_{NB}} \right\}, & \overline{E}_{1kHz}[b,t] > \lambda TQ_{1kHz} \end{cases}$$

$$(11c)$$

$$\exp \left\{ K_{NB} \left(-\log \left(\frac{\overline{E}_{1kHz}[b,t]}{TQ_{1kHz}} \right) + C_{NB} \right)^{\eta_{NB}} \right\}, & \text{en otros casos} \end{cases}$$

donde la constante λ es mayor que uno, el exponente η es menos que uno, y las constantes K y C están escogidas de modo que la función de volumen sonoro específico y su primera derivada sean continuas en el punto $\overline{E}_{1kHz}[b,t] = \lambda TQ_{1kHz}$.

A partir del volumen sonoro específico, el volumen sonoro global o "total" *L[t]* está dado por la suma del volumen sonoro específico para todas las bandas b:

$$L[t] = \sum_{b} N[b,t] \tag{12}$$

Modificación del volumen sonoro específico 105

So En la función de modificación del volumen sonoro específico ("Modificación del volumen sonoro específico") 105, el volumen sonoro específico de destino, denominado $\hat{N}[b,t]$, puede ser calculado a partir del volumen sonoro específico de SL 104 (figura 7) de diversas maneras, según la aplicación deseada del dispositivo o proceso global. Según se describe en mayor detalle más adelante, un volumen sonoro específico de destino puede ser calculado usando un factor α de escala, por ejemplo, en el caso de un control de volumen. Véase la Ecuación 16 más adelante y su descripción asociada. En el caso del control automático de ganancia (AGC) y del control de gama dinámico (DRC), un volumen sonoro específico de destino puede ser calculado usando una razón entre el volumen sonoro de salida deseado y el volumen sonoro de entrada. Véanse las Ecuaciones 17 y 18 más adelante y sus descripciones asociadas. En el caso de la ecualización dinámica, un volumen sonoro específico de destino puede ser calculado usando una relación estipulada en la Ecuación 23 y su descripción asociada.

Discriminador 106 de ganancia

10

15

20

25

40

En este ejemplo, para cada banda b y cada intervalo temporal t, el Discriminador 106 de ganancia toma como sus entradas la excitación allanada E[b,t] y el volumen sonoro específico de destino $\hat{N}[b,t]$, y genera las ganancias G[b,t], usadas subsiguientemente para modificar el audio. Dejando que la función $\psi\{^*\}$ represente a la transformación no lineal, desde la excitación al volumen sonoro específico, de modo que

$$N[b,t] = \Psi\{\overline{E}[b,t]\},\tag{13}$$

el Discriminador de ganancia halla G[b,t] de modo que

5

10

15

20

35

$$\widehat{N}[b,t] = \Psi \left\{ G^2[b,t] \overline{E}[b,t] \right\}. \tag{14a}$$

Los Discriminadores 106 de ganancia determinan ganancias variables en cuanto a la frecuencia y al tiempo, que, cuando se aplican a la excitación original, dan como resultado un volumen sonoro específico que, idealmente, es igual al volumen sonoro específico de destino deseado. En la práctica, los Discriminadores de ganancia determinan ganancias variables en cuanto a la frecuencia y al tiempo, que cuando son aplicadas a la versión del dominio de frecuencia de la señal de audio, dan como resultado la modificación de la señal de audio a fin de reducir la diferencia entre su volumen sonoro específico y el volumen sonoro específico de destino. Idealmente, la modificación es tal que la señal de audio modificada tiene un volumen sonoro específico que es una aproximación cercana al volumen sonoro específico de destino. La solución a la Ecuación 14a puede ser implementada de una cierta variedad de maneras. Por ejemplo, si existe una expresión matemática de forma cerrada para la inversa del volumen sonoro específico, representada por ψ^{-1} (**), entonces las ganancias pueden ser calculadas directamente reordenando la ecuación 14a:

$$G[b,t] = \sqrt{\frac{\Psi^{-1}(\hat{N}[b,t])}{\bar{E}[b,t]}}$$
 (14b)

Alternativamente, si no existe una solución de forma cerrada para ψ⁻¹{*}, puede ser empleado un enfoque iterativo en el cual, para cada iteración, se evalúa la ecuación 14a usando una estimación actual de las ganancias. El volumen sonoro específico resultante es comparado con el objetivo deseado y las ganancias son actualizadas en base al error. Si las ganancias son actualizadas debidamente, convergerán a la solución deseada. Otro método implica pre-calcular la función ψ{*} para una gama de valores de excitación en cada banda, para crear una tabla de búsqueda. A partir de esta tabla de búsqueda, se obtiene una aproximación de la función inversa ψ⁻¹{*} y las ganancias pueden ser luego calculadas a partir de la ecuación 14b. Como se ha mencionado anteriormente, el volumen sonoro específico de destino puede ser representado por un ajuste a escala del volumen sonoro específico:

$$\hat{N}[b,t] = \Xi[b,t]N[b,t] \tag{14c}$$

Sustituyendo la ecuación 13 en 14c y luego 14c en 14b, se obtiene una expresión alternativa para las ganancias:

$$G[b,t] = \sqrt{\frac{\Psi^{-1}(\Xi[b,t]\Psi(\overline{E}[b,t]))}{\overline{E}[b,t]}}$$
(14d)

Vemos que las ganancias pueden ser expresadas exclusivamente como una función de la excitación E[b,t] y el ajuste a escala $\Xi[b,t]$ del volumen sonoro específico. Por lo tanto, las ganancias pueden ser calculadas mediante la evaluación de 14d o una tabla de búsqueda equivalente, sin calcular jamás explícitamente el volumen sonoro específico o el volumen sonoro específico de destino como valores intermedios. Sin embargo, estos valores son calculados implícitamente mediante el uso de la ecuación 14d. Pueden idearse otros métodos equivalentes para calcular los parámetros de modificación, mediante el cálculo, bien explícito o bien implícito, del volumen sonoro específico y del volumen sonoro específico de destino, y esta invención está concebida para cubrir todos los métodos de ese tipo.

Banco 110 de filtros de síntesis

Como se ha descrito anteriormente, el Banco 100 de filtros de análisis puede ser implementado eficazmente mediante el uso de la Transformación Discreta de Fourier a Corto Plazo (STDFT) o la Transformación de Coseno Discreta Modificada, y la STDFT o la MDCT pueden ser usadas de manera similar para implementar el Banco 110 de filtros de síntesis. Específicamente, dejando que X[k,t] represente a la STDFT o la MDCT del audio de entrada, según lo definido

anteriormente, la STDFT o la MDCT del audio procesado (modificado) en el Banco 110 de filtros de síntesis puede ser calculada como

$$\hat{X}[k,t] = \sum_{b} G[b,t] S_{b}[k] X[k,t-d], \qquad (15)$$

donde $S_b[k]$ es la respuesta del filtro de síntesis asociado a la banda b, y d es el retardo asociado al bloque 109 de retardo en la figura 7. La forma de los filtros $S_b[k]$ de síntesis puede ser escogida como la misma que la de los filtros utilizados en el banco de filtros de análisis, $H_b[k]$, o bien pueden ser modificados para proporcionar una reconstrucción perfecta en ausencia de cualquier modificación de ganancia (es decir, cuando G[b,t]=1). El audio final procesado puede

ser luego generado mediante la transformación inversa de Fourier, o la transformación de coseno modificada de X[k, t], y la síntesis de superposición y agregado, como resulta familiar para alguien experto en la técnica.

Volumen sonoro específico de destino

5

10

25

30

35

55

60

El comportamiento de las disposiciones que realizan aspectos de la invención, tales como los ejemplos de las figuras 1 a 7, está dictado principalmente por la manera en la cual se calcula el volumen sonoro específico de destino $\hat{N}[b,t]$. Aunque la invención no está limitada por ninguna función, o función inversa, específica para calcular el volumen sonoro específico de destino, se describirán ahora varias de tales funciones y aplicaciones adecuadas para ellas.

20 Función invariable en cuanto al tiempo e invariable en cuanto a la frecuencia, adecuada para el control de volumen

Un control de volumen estándar ajusta el volumen sonoro de una señal de audio aplicando una ganancia de banda ancha al audio. En general, la ganancia está acoplada con una perilla o deslizador que es ajustado por un usuario hasta que el volumen sonoro del audio esté en el nivel deseado. Un aspecto de la presente invención admite una manera psicoacústicamente más coherente de implementar un control de ese tipo. Según este aspecto de la invención, en lugar de tener una ganancia de banda ancha acoplada al control de volumen que da como resultado un cambio de ganancia en la misma magnitud para todas las bandas de frecuencia, lo que puede causar un cambio en el espectro percibido, un factor de escala del volumen sonoro específico se asocia, en cambio, al ajuste del control de volumen, de modo que la ganancia en cada una de las múltiples bandas de frecuencia sea cambiada en una magnitud que tenga en cuenta el modelo del oído humano, de modo que, idealmente, no haya ningún cambio en el espectro percibido. En el contexto de este aspecto de la invención y de una aplicación ejemplar del mismo, se concibe que "constante" o "invariante en cuanto al tiempo" admitan cambios en la configuración de un factor de escala del control de volumen de tanto en tanto, por ejemplo, por parte de un usuario. Tal "invariancia en cuanto al tiempo" se denomina a veces "cuasi-invariante en cuanto al tiempo", "cuasi-estática", "invariante a trozos en cuanto al tiempo", "estática a trozos", "gradualmente invariante en cuanto al tiempo" y "gradualmente estática". Dado un tal factor de escala, α , el volumen sonoro específico de destino puede ser calculado como el volumen sonoro específico medido, multiplicado por α :

$$\hat{N}[b,t] = \alpha N[b,t]. \tag{16}$$

Debido a que el volumen sonoro total *L[t]* es la suma del volumen sonoro específico *N[b,t]* para todas las bandas b, la modificación anterior también ajusta a escala el volumen sonoro total en un factor de α, pero lo hace de una manera que preserva el mismo espectro percibido en un momento específico, por cambios en el ajuste del control de volumen. En otras palabras, en cualquier momento específico, un cambio en el ajuste del control de volumen da como resultado un cambio en el volumen sonoro percibido, pero ningún cambio en el espectro percibido del audio modificado con respecto al espectro percibido del audio no modificado. La figura 13a ilustra las ganancias multibanda resultantes *G[b,t]* para todas las bandas "b" en un momento "t" específico cuando α = 0,25 para una señal de audio que consiste en habla femenina. Para comparar, la ganancia de banda ancha requerida para ajustar a escala el volumen sonoro total original en 0,25 (la línea horizontal), como en un control de volumen estándar, también se grafica. La ganancia multibanda *G[b,t]* aumenta en las bandas de frecuencia alta y baja, en comparación con las bandas de frecuencia media. Esto es congruente con los perfiles de igual volumen sonoro, lo que indica que el oído humano es menos sensible a frecuencias bajas y altas.

La figura 13b ilustra el volumen sonoro específico para la señal de audio original, la señal modificada en ganancia de banda ancha, según lo modificado de acuerdo a un control de volumen de la técnica anterior, y la señal modificada en ganancia de multibanda, según lo modificado de acuerdo a este aspecto de la invención. El volumen sonoro específico de la señal modificada en ganancia de multibanda es el del original, ajustado a escala en 0,25. El volumen sonoro específico de la señal modificada en ganancia de banda ancha ha cambiado su forma espectral con respecto al de la señal original no modificada. En este caso, el volumen sonoro específico, en un sentido relativo, ha perdido volumen sonoro en ambas frecuencias bajas y altas. Esto es percibido como un oscurecimiento del audio según su volumen se baja, un problema que no ocurre con la señal modificada de multibanda cuyo volumen sonoro está controlado por ganancias obtenidas en el dominio del volumen sonoro perceptivo.

Junto con la distorsión del equilibrio espectral percibido, asociado a un control de volumen tradicional, existe un segundo problema. Una propiedad de la percepción del volumen sonoro, que se refleja en el modelo de volumen sonoro reflejado en las Ecuaciones 11a a 11d, es que el volumen sonoro de una señal, en cualquier frecuencia, disminuye más rápidamente según el nivel de la señal se aproxima al umbral del oído. Como resultado, la atenuación eléctrica requerida para impartir la misma atenuación del volumen sonoro a una señal más suave es menor que la requerida para una señal más alta. Un control de volumen tradicional imparte una atenuación constante, independientemente del nivel de señal y, por lo tanto, las señales suaves se tornan "demasiado suaves" con respecto a las señales más altas, según el volumen se baja. En muchos casos, esto da como resultado la pérdida de detalles en el audio. Consideremos la grabación de una castañuela en una habitación reverberante. En tal grabación, el "golpe" principal de la castañuela es bastante alto en comparación a los ecos reverberantes, pero son los ecos reverberantes los que transmiten el tamaño de la habitación. Según se baja el volumen con un control de volumen tradicional, los ecos reverberantes se suavizan con respecto al golpe principal y eventualmente desaparecen por debajo del umbral del oído, dejando una castañuela con sonido "seco". El control de volumen basado en el volumen sonoro impide la desaparición de las partes más suaves de las grabaciones, estimulando la parte reverberante más suave de la grabación con respecto al golpe principal más alto, de modo que el volumen sonoro relativo entre estas secciones permanezca constante. A fin de lograr este efecto, las ganancias de multibanda G/b,t] deben variar a lo largo del tiempo a una velocidad que sea conmensurable con la resolución temporal humana de la percepción del volumen sonoro. Debido a que las ganancias de multibanda G[b,t] son calculadas como una función de la excitación allanada $\overline{E}[b, t]$, la selección de las constantes temporales λ_b en la Ecuación 8 dicta cuán rápidamente pueden variar las ganancias a lo largo del tiempo en cada banda b. Como se ha mencionado anteriormente, estas constantes temporales pueden ser seleccionadas para que sean proporcionales al tiempo de integración de la percepción humana del volumen sonoro dentro de la banda b, y así producir la variación adecuada de G[b,t] a lo largo del tiempo. Debería observarse que si las constantes temporales son escogidas inadecuadamente (ya sea demasiado rápidas o demasiado lentas), entonces pueden introducirse distorsiones perceptivamente objetables en el audio procesado.

Función invariante en cuanto al tiempo y variable en cuanto a la frecuencia, adecuada para la ecualización fija

En algunas aplicaciones, puede desearse aplicar una ecualización perceptiva fija al audio, en cuyo caso el volumen sonoro específico de destino puede ser calculado aplicando un factor $\Theta[b]$ de escala invariante en cuanto al tiempo pero variable en cuanto a la frecuencia, según la relación

$\hat{N}[b,t] = \Theta[b]N[b,t],$

10

15

20

25

30

40

45

50

en la que $\hat{N}[b, t]$ es el volumen sonoro específico de destino, N[b, t] es el volumen sonoro específico de la señal de audio, b es una medida de frecuencia y t es una medida de tiempo. En este caso, el ajuste a escala puede variar entre una banda y otra. Una aplicación de ese tipo puede ser útil para realzar, por ejemplo, la parte del espectro dominada por frecuencias del habla, a fin de potenciar la inteligibilidad.

Función invariante en cuanto a la frecuencia y variable en cuanto al tiempo, adecuada para el control de ganancia automática y de gama dinámica

Las técnicas de control de ganancia automática y de gama dinámica (AGC y DRC) son bien conocidas en el campo del procesamiento del audio. En un sentido abstracto, ambas técnicas miden el nivel de una señal de audio de alguna manera, y luego modifican en ganancia la señal, en una magnitud que es una función del nivel medido. Para el caso del AGC, la señal está modificada en ganancia de modo que su nivel medido esté más cerca de un nivel de referencia seleccionado por el usuario. Con el DRC, la señal es modificada en ganancia de modo que la gama del nivel medido de la señal se transforme en alguna gama deseada. Por ejemplo, puede desearse hacer que las partes tranquilas del audio sean más altas y las partes altas más tranquilas. Un tal sistema es descrito por Robinson y Gundry (Charles Robinson y Kenneth Gundry, "Control de gama dinámica mediante metadatos", 107ª Convención de la AES, Preimpresión 5028, septiembre 24 a 27, 1999, Nueva York). Las implementaciones tradicionales de AGC y DRC utilizan generalmente una medición simple del nivel de la señal de audio, tal como un pico allanado o una amplitud de la raíz cuadrada mínima (rms), para controlar la modificación de ganancia. Tales mediciones sencillas se correlacionan en cierto grado con el volumen sonoro percibido del audio, pero los aspectos de la presente invención admiten un AGC y DRC más relevantes perceptivamente, controlando las modificaciones de ganancia con una medida del volumen sonoro basada en un modelo psicoacústico. Además, muchos sistemas tradicionales de AGC y DRC aplican la modificación de ganancia con una ganancia de banda ancha, incurriendo por ello en las precitadas distorsiones tímbricas (espectrales) en el audio procesado. Los aspectos de la presente invención, por otra parte, utilizan una ganancia de multibanda para modelar el volumen sonoro específico, de manera tal que reduzca o minimice tales distorsiones.

Ambas aplicaciones de AGC y DRC, que emplean aspectos de la presente invención, están caracterizadas por una función que transforma, o correlaciona, un volumen sonoro $L_i[t]$ de banda ancha de entrada en un volumen sonoro becomo el sono. El volumen sonoro $L_i[t]$ de banda ancha de entrada es una función del volumen sonoro N[b,t] específico de

la señal de audio de entrada. Aunque puede ser el mismo que el volumen sonoro total de la señal de audio de entrada, puede ser una versión temporalmente allanada del volumen sonoro total de la señal de audio.

Las figuras 14a y 14b ilustran ejemplos de funciones de correlación típicas para un AGC y un DRC, respectivamente. Dada una correlación de ese tipo, en la cual $L_o[t]$ es una función de $L_i[t]$, el volumen sonoro específico de destino puede ser calculado como

$$\hat{N}[b,t] = \frac{L_o[t]}{L_i[t]} N[b,t]. \tag{17}$$

5

20

30

35

45

50

55

El volumen sonoro N[b,t] específico original de la señal de audio es sencillamente ajustado a escala por la razón entre el volumen sonoro de banda ancha de salida deseado y el volumen sonoro de banda ancha de entrada, para producir un volumen sonoro $\hat{N}[b,t]$ específico de salida. Para un sistema de AGC, el volumen sonoro $L_i[t]$ de banda ancha de entrada debería ser generalmente una medida del volumen sonoro total a largo plazo del audio. Esto puede lograrse allanando el volumen sonoro total L[t] a lo largo del tiempo para generar $L_i[t]$.

En comparación con un AGC, un sistema de DRC reacciona ante cambios a corto plazo en el volumen sonoro de una señal y, por lo tanto, L[t], sencillamente, puede hacerse igual a L[t]. Como resultado, el ajuste a escala del volumen sonoro específico, dado por $L_0[t]/L[t]$, puede fluctuar rápidamente, llevando a distorsiones indeseadas en el audio procesado. Una distorsión típica es la modulación audible de una parte del espectro de frecuencias por alguna otra parte relativamente ajena del espectro. Por ejemplo, una selección de música clásica podría contener altas frecuencias dominadas por una nota sostenida de cuerdas, mientras que las frecuencias bajas contienen un timbal alto y retumbante. Toda vez que el timbal ataca, aumenta el volumen sonoro global L[t], y el sistema de DRC aplica la atenuación al volumen sonoro específico total. Se oye luego que las cuerdas "bombean" hacia arriba y hacia abajo en volumen sonoro con los timbales. Tal bombeo cruzado en el espectro es asimismo un problema con los sistemas tradicionales de DRC de banda ancha, y una solución típica implica aplicar el DRC independientemente a distintas bandas de frecuencia. El sistema divulgado aquí es inherentemente multibanda, debido al banco de filtros y al cálculo del volumen sonoro específico que emplea un modelo de volumen sonoro perceptivo y, por lo tanto, la modificación de un sistema de DRC para que funcione de manera multibanda, de acuerdo a los aspectos de la presente invención, es relativamente inmediata, y se describe a continuación.

Función variable en cuanto a la frecuencia y en cuanto al tiempo, adecuada para el control de gama dinámica

El sistema de DRC puede ser expandido para funcionar en estilo multibanda, o variable en cuanto a la frecuencia, permitiendo que el volumen sonoro de entrada y de salida varíen independientemente con la banda b. Estos valores de volumen sonoro de multibanda son denominados L[b,t] y $L_o[b,t]$, y el volumen sonoro específico de destino puede entonces estar dado por

$$\hat{N}[b,t] = \frac{L_a[b,t]}{L_t[b,t]} N[b,t],$$
 (18)

donde $L_0[b,t]$ ha sido calculado, o correlacionado, a partir de $L_1[b,t]$, según lo ilustrado en la figura 14b, pero independientemente para cada banda b. El volumen sonoro $L_1[b,t]$ de multibanda de entrada es una función del volumen sonoro N[b,t] específico de la señal de audio de entrada. Aunque puede ser el mismo que el volumen sonoro específico de la señal de audio de entrada, puede ser una versión temporalmente allanada y/o allanada en cuanto a la frecuencia del volumen sonoro específico de la señal de audio.

La manera más inmediata de calcular $L_i[b,t]$ es fijarlo igual al volumen sonoro específico N[b,t]. En este caso, el DRC es realizado independientemente en cada banda en el banco de filtros auditivos del modelo de volumen sonoro perceptivo, en lugar de hacerlo según la misma razón de volumen sonoro entre la entrada y la salida para todas las bandas, según lo recién descrito anteriormente bajo el encabezamiento "Función invariante en cuanto a la frecuencia y variable en cuanto al tiempo, adecuada para el control de ganancia automática y de gama dinámica". En una realización práctica que emplea 40 bandas, el espaciado de estas bandas a lo largo del eje de la frecuencia es relativamente fino, a fin de proporcionar una medida precisa del volumen sonoro. Sin embargo, la aplicación de un factor de escala del DRC independientemente a cada banda puede causar que el audio procesado suene "desgarrado". Para evitar este problema, se puede escoger calcular $L_i[b,t]$ allanando el volumen sonoro específico N[b,t] sobre las bandas, de modo que la magnitud del DRC aplicada desde una banda a la siguiente no varíe tan drásticamente. Esto puede lograrse definiendo un filtro Q(b) de allanamiento de bandas y allanando luego el volumen sonoro específico sobre todas las bandas c, según la suma estándar de convolución:

$$L_{t}[b,t] = \sum_{c} Q(b-c)N[c,t].$$
 (19)

5

10

15

20

25

30

50

en la que N[c,t] es el volumen sonoro específico de la señal de audio y Q(b-c) es la respuesta desfasada en banda del filtro de allanamiento. La figura 15 ilustra un ejemplo de un filtro de allanamiento de banda de ese tipo.

Si la función de DRC que calcula $L_i[b,t]$ como una función de $L_o[b,t]$ está fijada para cada banda b, entonces el tipo de cambio incurrido para cada banda del volumen sonoro específico N[b,t] variará según el espectro del audio que está siendo procesado, incluso si el volumen sonoro global de la señal permanece igual. Por ejemplo, una señal de audio con bajos altos y agudos silenciosos puede tener los bajos recortados y el agudo realzado. A una señal con bajos silenciosos y agudos altos puede ocurrirle lo contrario. El efecto neto es un cambio en el timbre o espectro percibido del audio, y esto puede ser deseable en ciertas aplicaciones.

Sin embargo, se puede desear realizar el DRC de multibanda sin modificar el espectro percibido medio del audio. Se podría querer que la modificación media en cada banda sea aproximadamente la misma, admitiendo sin embargo a la vez que las variaciones a corto plazo de las modificaciones operen independientemente entre dos o más bandas. El efecto deseado puede ser logrado forzando que el comportamiento promedio del DRC en cada banda sea el mismo que el de algún comportamiento de referencia. Se puede escoger este comportamiento de referencia como el DRC deseado para el volumen sonoro $L_i[t]$ de entrada de banda ancha. Sea la función $L_i[t]$ = DRC $\{L_i[t]\}$ la que represente la deseada correlación del DRC para el volumen sonoro de banda ancha. Entonces, sea $\overline{L}_i[t]$ la que represente una versión promediada en el tiempo del volumen sonoro de entrada de banda ancha, y $\overline{L}_i[b,t]$ la que represente una versión promediada en el tiempo del volumen sonoro $L_i[b,t]$ de entrada de multibanda. El volumen sonoro de salida de multibanda puede ser calculado entonces como

$$L_o[b,t] = \frac{\overline{L}_i[b,t]}{\overline{L}_i[t]} DRC \left\{ \frac{\overline{L}_i[t]}{\overline{L}_i[b,t]} L_i[b,t] \right\}.$$
 (20)

Obsérvese que el volumen sonoro de entrada de multibanda se ajusta primero a escala para que esté en la misma gama media que el volumen sonoro de entrada de banda ancha. La función de DRC diseñada para el volumen sonoro de banda ancha se aplica luego. Por último, el resultado es reajustado a escala descendiendo hasta la gama media del volumen sonoro de multibanda. Con esta formulación del DRC de multibanda, se retienen los beneficios del bombeo espectral reducido, preservando a la vez el espectro medio percibido del audio.

Función variable en cuanto a frecuencia y variable en el tiempo para la ecualización dinámica

Otra aplicación de aspectos de la presente invención es la transformación intencionada del espectro del audio, variable 35 en el tiempo, en un espectro percibido invariante en el tiempo, preservando aún la gama dinámica original del audio. Se puede denominar a este procesamiento Ecualización Dinámica (DEQ). Con la tradicional ecualización estática, se aplica un filtrado fijo sencillo al audio a fin de cambiar su espectro. Por ejemplo, se podría aplicar un estímulo fijo de bajos o de agudos. Tal procesamiento no tiene en cuenta el espectro actual del audio y, por lo tanto, puede ser inadecuado para algunas señales, es decir, señales que ya contienen una magnitud relativamente grande de bajos o agudos. Con la DEQ, 40 el espectro de la señal es medido y la señal luego es dinámicamente modificada a fin de transformar el espectro medido en una forma deseada, esencialmente estática. Para los aspectos de la presente invención, una forma deseada de ese tipo se especifica sobre las bandas en el banco de filtros y se denomina EQ[b]. En una realización práctica, el espectro medido debería representar la forma espectral media del audio que puede ser generado allanando el volumen sonoro específico N[b,t] a lo largo del tiempo. Es posible referirse al volumen sonoro específico allanado como N[b,t]. Igual que con el DRC de multibanda, se puede no querer que la modificación de la DEQ varíe drásticamente desde una banda a la 45 siguiente y, por lo tanto, pueda ser aplicada una función de allanamiento de banda para generar un espectro $\overline{L}[b,\,t]$ allanado por bandas:

$$\overline{L}[b,t] = \sum_{c} Q(b-c)\overline{N}[c,t]. \tag{21}$$

A fin de preservar la gama dinámica original del audio, el espectro deseado EQ[b] debería ser normalizado para que tenga el mismo volumen sonoro global que la forma espectral medida dada por $\overline{L}[b,t]$. Se puede hacer referencia a esta forma espectral normalizada como $\overline{L}_{EQ}[b,t]$:

$$\overline{L}_{EQ}[b,t] = \left(\frac{\sum_{c} \overline{L}[c,t]}{\sum_{c} EQ[c]}\right) EQ[b].$$
(22)

Finalmente, el volumen sonoro específico de destino se calcula como

$$\hat{N}[b,t] = \left(\frac{\overline{L}[b,t]}{\overline{L}_{EQ}[b,t]}\right)^{1-\beta} \frac{\overline{L}_{EQ}[b,t]}{\overline{L}[b,t]} N[b,t], \qquad (23)$$

donde β es un parámetro especificado por el usuario que oscila entre cero y uno, indicando el grado de la DEQ que ha de aplicarse. Mirando a la Ecuación 23, se observa que cuando β = 0, el volumen sonoro específico original está sin modificar y que, cuando β = 1, el volumen sonoro específico está ajustado a escala por la razón entre la forma espectral deseada y la forma espectral medida.

Una manera conveniente de generar la forma espectral deseada EQ[b] es que un usuario la fije igual a $\overline{L}[b, f]$, según lo medido para algún trozo de audio cuyo equilibrio espectral el usuario halle placentero. En una realización práctica, por ejemplo, según se muestra en la figura 16, el usuario puede estar dotado de un botón u otro activador adecuado 507 que, cuando se activa, causa una captura de la medida corriente de la forma espectral $\overline{L}[b, f]$ del audio, y luego almacena esta medida como una preselección (en el bloque Captura y almacenamiento de preselección de volumen sonoro específico de destino 506) que pueda ser cargada más tarde en EQ[b] cuando se habilite la DEQ (según la selección preseleccionada 508). La figura 16 es una versión simplificada de la figura 7, en la cual solamente se muestra una única línea para representar múltiples bandas desde el Banco 100 de filtros de análisis hasta el Banco 110 de filtros de síntesis. El ejemplo de la figura 17 también proporciona un bloque de Modificación de volumen sonoro específico (SL) de EQ dinámico 505, que proporciona una modificación para el volumen sonoro específico medido por la función o dispositivo 104, de acuerdo a la ecualización dinámica, según lo explicado anteriormente.

Procesamiento combinado

5

10

15

20

25

30

Se puede desear combinar todo el procesamiento anteriormente descrito, incluso el Control de volumen (VC), el AGC, el DRC y la DEQ, en un único sistema. Debido a que cada uno de estos procesos puede estar representado como un ajuste a escala del volumen sonoro específico, todos ellos se combinan fácilmente de la siguiente manera:

$$\hat{\mathcal{N}}[b,t] = \left(\Xi_{VC}[b,t]\Xi_{AGC}[b,t]\Xi_{DEC}[b,t]\Xi_{DEQ}[b,t]\right)N[b,t],\tag{24}$$

donde Ξ , [b, t] representa a los factores de escala asociados al proceso "*". Un único conjunto de ganancias G[b,t] puede ser calculado luego para el volumen sonoro específico de destino que representa al procesamiento combinado.

En algunos casos, los factores de escala de uno, o una combinación, de los procesos de modificación del volumen sonoro, pueden fluctuar demasiado rápidamente a lo largo del tiempo y producir distorsiones en el audio procesado resultante. Por lo tanto, puede ser deseable allanar algún subconjunto de estos factores de escala. En general, los factores de escala de VC y DEQ varían suavemente a lo largo del tiempo, pero puede requerirse el allanamiento de la combinación de los factores de escala del AGC y del DRC. Sea la combinación de estos factores de escala representada por

$$\Xi_{C}[b,t] = \Xi_{AGC}[b,t]\Xi_{DRC}[b,t] \tag{25}$$

La noción básica tras el allanamiento es que los factores de escala combinados deberían reaccionar rápidamente cuando está aumentando el volumen sonoro específico, y que los factores de escala deberían ser allanados más intensamente cuando el volumen sonoro específico está disminuyendo. Esta noción corresponde a la práctica bien conocida de utilizar un ataque rápido y una liberación lenta en el diseño de los compresores de audio. Las constantes temporales adecuadas para allanar los factores de escala pueden ser calculadas allanando a lo largo del tiempo una versión allanada por bandas del volumen sonoro específico. Primero se calcula una versión allanada por bandas del volumen sonoro específico:

$$L[b,t] = \sum_{c} Q(b-c)N[c,t]. \tag{26}$$

en la que N[c,t] es el volumen sonoro específico de la señal de audio y Q(b-c) es la respuesta desfasada en banda del filtro de allanamiento, según la Ecuación 19 anterior.

La versión allanada a lo largo del tiempo de este volumen sonoro específico allanado por banda se calcula luego como

$$\overline{L}[b,t] = \lambda[b,t]L[b,t] + (1-\lambda[b,t])\overline{L}[b,t-1]$$
(27)

10 donde el coeficiente λ[b,t] de allanamiento dependiente de la banda está dado por

$$\lambda[b,t] = \begin{cases} \lambda_{fast}, & L[b,t] > \overline{L}[b,t] \\ \lambda_{slow} & L[b,t] \le \overline{L}[b,t] \end{cases}$$
 (28)

Los factores de escala combinados y allanados se calculan luego como

$$\overline{\Xi}_{C}[b,t] = \lambda_{M}[b,t]\Xi_{C}[b,t] + (1 - \lambda_{M}[b,t])\overline{\Xi}_{C}[b,t-1], \tag{29}$$

donde $\lambda_M[b,t]$ es una versión allanada por banda de $\lambda[b,t]$:

$$\lambda_{M}[b,t] = \left(\frac{1}{\sum_{c} Q(c)}\right) \sum_{c} Q(b-c)\lambda[b,t]. \tag{30}$$

El allanamiento de bandas del coeficiente de allanamiento impide que los factores de escala allanados en cuanto al tiempo cambien drásticamente entre las bandas. El allanamiento descrito, en cuanto al tiempo y a la banda, del factor de escala da como resultado un audio procesado que contiene menos distorsiones perceptivas objetables.

Compensación de ruido

5

15

20

25

30

35

40

45

En muchos entornos de reproducción de audio existe un ruido de fondo que interfiere con el audio que un oyente desea oír. Por ejemplo, un oyente en un automóvil en movimiento puede estar reproduciendo música por el sistema estéreo instalado, y el ruido del motor y de la carretera puede alterar significativamente la percepción de la música. En particular, para partes del espectro en las cuales la energía del ruido es significativa con respecto a la energía de la música, el volumen sonoro percibido de la música se reduce. Si el nivel del ruido es lo bastante grande, la música es enmascarada por completo. Con respecto a un aspecto de la actual invención, se desearía escoger ganancias G[b,t] de modo que el volumen sonoro específico del audio procesado en presencia del ruido de interferencia sea igual al volumen sonoro $\hat{N}[b,t]$ específico de destino. Para lograr este efecto, se puede utilizar el concepto de volumen sonoro parcial, según lo definido por Moore y Glasberg, *supra*. Supongamos que es posible obtener una medición del ruido en sí y una medición del audio en sí. Sea $E_N[b,t]$ lo que represente la excitación proveniente del ruido y $E_A[b,t]$ lo que represente la excitación proveniente del ruido está entonces dado por

$$N_{TOT}[b,t] = \Psi \{ E_A[b,t] + E_N[b,t] \}, \qquad (31)$$

donde, nuevamente, ψ {*} representa la transformación no lineal de la excitación en volumen sonoro específico. Se puede suponer que el oído de un oyente divide el volumen sonoro específico combinado entre el volumen sonoro específico parcial del audio y el volumen sonoro específico parcial del ruido, de una manera que preserve el volumen sonoro específico combinado:

$$N_{TOT}[b,t] = N_A[b,t] + N_N[b,t]. \tag{32}$$

El volumen sonoro específico parcial del audio, $N_A[b,t]$, es el valor que se desea controlar y, por lo tanto, hay que despejar este valor. El volumen sonoro específico parcial del ruido puede ser aproximado como

$$N_{N}[b,t] = \left(\frac{E_{TN}[b,t]}{E_{A}[b,t]}\right)^{k} \left(\Psi\{E_{N}[b,t] + E_{TN}[b,t]\} - \Psi\{E_{TQ}[b]\}\right)$$
(33)

5

20

25

30

45

50

55

donde $E_{TN}[b,t]$ es el umbral enmascarado en presencia de ruido, $E_{TO}[b]$ es el umbral del oído en silencio en la banda b y k es un exponente entre cero y uno. Combinando las Ecuaciones 31 a 33, se llega a una expresión para el volumen sonoro específico parcial del audio:

$$N_{A}[b,t] = \Psi\{E_{A}[b,t] + E_{N}[b,t]\} - \left(\frac{E_{TN}[b,t]}{E_{A}[b,t]}\right)^{\kappa} \left(\Psi\{E_{N}[b,t] + E_{TN}[b,t]\} - \Psi\{E_{TQ}[b]\}\right)$$
(34)

Se observa que cuando la excitación del audio es igual al umbral enmascarado del ruido ($E_A[b,t]=E_{TN}[b,t]$), el volumen sonoro específico parcial del audio es igual al volumen sonoro de una señal en el umbral en silencio, que es el resultado deseado. Cuando la excitación del audio es mucho mayor que la del ruido, el segundo término en la Ecuación 34 desaparece, y el volumen sonoro específico del audio es aproximadamente igual al que tendría si el ruido no estuviera presente.

En otras palabras, según el audio se hace mucho más alto que el ruido, el ruido es enmascarado por el audio. El exponente k se escoge empíricamente para dar un buen ajuste a los datos sobre el volumen sonoro de un tono en el ruido, como una función de la razón entre señal y ruido. Moore y Glasberg han hallado que un valor de k = 0,3 es adecuado. El umbral enmascarado del ruido puede ser aproximado como una función de la misma excitación del ruido:

$$E_{TN}[b,t] = K[b]E_{N}[b,t] + E_{TD}[b]$$
(35)

donde *K[b]* es una constante que aumenta en las bandas de menor frecuencia. Así, el volumen sonoro específico parcial del audio, dado por la Ecuación 34, puede ser representado abstractamente como una función de la excitación del audio y de la excitación del ruido:

$$N_{A}[b,t] = \Phi\{E_{A}[b,t], E_{N}[b,t]\}. \tag{36}$$

Un discriminador de ganancia modificado puede ser utilizado luego para calcular las ganancias *G[b,t]*, de modo que el volumen sonoro específico parcial del audio procesado en presencia del ruido sea igual al volumen sonoro específico de destino:

$$\hat{N}[b,t] = \Phi \left\{ G^{2}[b,t] E_{A}[b,t], E_{N}[b,t] \right\}$$
(37)

La figura 17 ilustra el sistema de la figura 7 con el Discriminador 106 de ganancia reemplazado por el Discriminador 206 de ganancia compensador de ruido descrito (obsérvese que las múltiples líneas verticales entre los bloques, que representan a las múltiples bandas del banco de filtros, han sido reemplazadas por una única línea para simplificar el diagrama). Además, la figura ilustra la medición de la excitación del ruido (por el Banco 200 de filtros de análisis, el Filtro 201 de transmisión, la Excitación 202 y el Allanamiento 203, de manera correspondiente al funcionamiento de los bloques 100, 101, 102 y 103) que se suministra al nuevo discriminador 206 de ganancia junto con la excitación del audio (proveniente del Allanamiento 103) y el volumen sonoro específico de destino (proveniente de la Modificación 105 de SL).

En su modalidad más básica de funcionamiento, la Modificación 105 de SL en la figura 17 puede simplemente fijar el volumen sonoro específico de destino $\hat{N}[b,t]$ igual al volumen sonoro específico original del audio N[b,t]. En otras palabras, la Modificación de SL proporciona un ajuste a escala por el factor α , invariante en cuanto a la frecuencia, del volumen sonoro específico de la señal de audio, en el que $\alpha=1$. Con una disposición tal como en la figura 17, las ganancias son calculadas de modo que el espectro del volumen sonoro percibido del audio procesado en presencia del ruido sea igual al espectro del volumen sonoro del audio en ausencia del ruido. Adicionalmente, una cualquiera, o una combinación, de las técnicas previamente descritas para calcular el volumen sonoro específico de destino como una función del original, incluso VC, AGC, DRC y DEQ, puede ser utilizada conjuntamente con el sistema de modificación de volumen sonoro compensador de ruido.

En una realización práctica, la medición del ruido puede ser obtenida de un micrófono colocado en, o cerca de, el entorno en el cual se reproducirá el audio. Alternativamente, un conjunto predeterminado de excitaciones de ruido de plantilla puede ser utilizado para aproximarse al espectro de ruido anticipado, en diversas condiciones. Por ejemplo, el ruido en

una cabina de automóvil puede ser pre-analizado a diversas velocidades de conducción, y luego almacenado como una tabla de búsqueda de la excitación del ruido con respecto a la velocidad. La excitación de ruido suministrada al Discriminador 206 de ganancia en la figura 17 puede ser aproximada luego a partir de esta tabla de búsqueda, según varíe la velocidad del automóvil.

Aproximaciones al volumen sonoro específico

5

10

15

20

25

30

35

40

45

Si bien las mejores prestaciones se obtienen cuando se usa una medida precisa del volumen sonoro específico, algunas aplicaciones pueden requerir el uso de una aproximación más grosera, a fin de reducir la complejidad de cálculo. Con una aproximación adecuada, de acuerdo a la invención, aún pueden lograrse una estimación y una modificación aceptables del volumen sonoro percibido. Una tal aproximación debería intentar preservar, al menos en parte, varios aspectos clave de la percepción del volumen sonoro. En primer lugar, la aproximación debería, al menos, capturar aproximadamente la sensibilidad variable en la percepción del volumen sonoro con la frecuencia. En términos generales, la aproximación debería exhibir menos sensibilidad en las frecuencias más bajas y más altas, en comparación con las frecuencias medias. En segundo lugar, la aproximación debería exhibir un crecimiento no lineal del volumen sonoro con el nivel de la señal. Específicamente, el crecimiento del volumen sonoro específico debería ser sumamente rápido para señales de bajo nivel cerca del umbral del oído, y luego disminuir hasta una velocidad constante de crecimiento según aumenta el nivel de la señal. Por último, la aproximación debería exhibir la propiedad de adición del volumen sonoro, lo que significa que, para un nivel de señal constante, el volumen sonoro total (la integración del volumen sonoro específico entre las frecuencias) aumenta según aumenta el ancho de banda de una señal.

Una forma de reducir la complejidad del cálculo del volumen sonoro específico, reteniendo a la vez las propiedades deseadas de la percepción del volumen sonoro, es usar un banco de filtros con menos bandas, y en el cual las bandas pueden no estar uniformemente espaciadas en una escala crítica de velocidades de banda. Por ejemplo, se podría usar un banco de filtros de 5 bandas, con las bandas uniformemente distribuidas en una escala lineal de frecuencia, al contrario que las 40 bandas descritas anteriormente. Existen muchas técnicas para la implementación eficaz de tales bancos de filtros, por ejemplo, los bancos de filtros modulados por cosenos (P. P. Vaidyanthan, Sistemas de MultiVelocidad y Bancos de Filtros, 1993 Prentice Hall). Como un ejemplo general, consideremos un banco de filtros con B bandas, donde cada banda está descrita por una respuesta $h_b[n]$ de impulso del dominio temporal. Además, supongamos que el banco de filtros es una reconstrucción casi perfecta, lo que significa que:

$$x[n] \cong \sum_{b=1}^{B} h_b[n] * x[n]$$
(38)

La respuesta de frecuencia de cada banda b puede ser caracterizada por una frecuencia central f_b y un ancho de banda Δf_b en Hercios. La frecuencia central equivalente y el ancho de banda en unidades de ERB están entonces dados por:

$$e_b = HzAERB(f_b) (39a)$$

$$Ae_b = HzAERB \left(f_b + \frac{\Delta f_b}{2} \right) - HzAERB \left(f_b - \frac{\Delta f_b}{2} \right)$$
(39b)

Si el número de bandas B es relativamente pequeño, entonces el ancho de banda e_b de cada banda, probablemente, será mayor que 1 ERB.

Haciendo que $x_b[n] = h_b[n] * x[n]$ represente a la señal de audio asociada a cada banda, una señal $\overline{E}[b, n]$ de excitación allanada puede ser calculada de manera similar a las Ecuaciones 7 y 8, calculando la potencia de RMS allanada de $x_b[n]$, ponderada por la respuesta de frecuencia del filtro de transmisión muestreado a la frecuencia f_b y la inversa del ancho de banda, en unidades de ERB:

$$\overline{E}[b,n] = \lambda_b \overline{E}[b,n] + (1-\lambda_b) \frac{|P(f_b)|^2}{e_b} x_b^2[n]$$
(40)

La ponderación de la excitación de la banda b por $1/e_b$ distribuye efectivamente la energía dentro de esa banda uniformemente por todas las bandas críticas subsumidas dentro de ella. Se podría asignar alternativamente toda la energía a la banda crítica cuya frecuencia central esté más cerca de la frecuencia central f_b de la banda, pero distribuir la energía uniformemente es una mejor aproximación para la mayoría de las señales de audio del mundo real.

Con la excitación $\overline{E}[b,n]$ se puede avanzar para calcular el volumen sonoro específico N[b,n], un volumen sonoro $\hat{N}[b,n]$ específico de destino y ganancias G[b,n], según lo expuesto anteriormente, excepto porque aquí el número total de

bandas puede ser mucho más pequeño, reduciendo por ello la complejidad. Además, se debe hacer una modificación al calcular el volumen sonoro global *L[n]*: el volumen sonoro específico debe ser ponderado por el número de los ERB dentro de la banda b, según se suma sobre las bandas:

$$L[n] = \sum_{b} e_b N[b, n] \tag{41}$$

Finalmente, puede producirse una señal modificada y[n] sumando cada una de las señales de banda, ponderadas por la ganancia proveniente de la banda respectiva:

$$y[n] = \sum_{b} G[b, n] x_b[n] \tag{42}$$

Para simplificar la exposición, los cálculos anteriores se muestran como realizados para cada periodo n de muestreo de la señal *x[n]*. En la práctica, sin embargo, la excitación puede ser submuestreada a una velocidad mucho menor a lo largo del tiempo, y luego todo el procesamiento subsiguiente del volumen sonoro puede ser llevado a cabo a esta velocidad reducida. Cuando las ganancias son aplicadas al final, pueden luego ser muestreadas por exceso, a través de la interpolación, antes de ser aplicadas a las señales de banda.

Lo anterior es solamente un ejemplo de una aproximación más grosera al volumen sonoro específico, que es adecuada para la invención divulgada.

Implementación

5

10

15

20

30

35

La invención puede ser implementada en hardware o software, o una combinación de ambos (p. ej., formaciones de lógica programable). A menos que se especifique lo contrario, los algoritmos incluidos como parte de la invención no están inherentemente relacionados con ningún ordenador específico, u otro aparato. En particular, pueden ser usadas diversas máquinas de propósito general con programas escritos de acuerdo a las divulgaciones en el presente documento, o puede ser más conveniente construir aparatos más especializados (p. ej., circuitos integrados) para llevar a cabo las etapas de método requeridas. De ese modo, la invención puede ser implementada en uno o más programas de ordenador ejecutándose en uno o más sistemas de ordenadores programables, comprendiendo cada uno al menos un procesador, al menos un sistema de almacenamiento de datos (que incluye elementos de memoria y/o almacenamiento volátiles y no volátiles), al menos un dispositivo o puerto de entrada y al menos un dispositivo o puerto de salida. El código de programa se aplica a los datos de entrada para realizar las funciones descritas en el presente documento y generar información de salida. La información de salida se aplica a uno o más dispositivos de salida, de una manera conocida.

Cada programa de ese tipo puede ser implementado en cualquier lenguaje de ordenador deseado (incluso lenguajes de máquina, de ensamblaje, o de programación de alto nivel, procedural, lógica u orientada a objetos) para comunicarse con un sistema de ordenador. En cualquier caso el lenguaje puede ser un lenguaje compilado o interpretado.

Cada programa de ordenador de ese tipo es preferiblemente almacenado en, o descargado a, un medio o dispositivo de almacenamiento (p. ej., memoria o medios de estado sólido, o medios magnéticos u ópticos) legible por un ordenador programable de propósito general o especial, para configurar y operar el ordenador cuando el medio o dispositivo de almacenamiento es leído por el sistema de ordenador para realizar los métodos descritos en el presente documento. El sistema de la invención puede también ser considerado para ser implementado como un medio de almacenamiento
 legible por ordenador, configurado con un programa de ordenador, donde el medio de almacenamiento así configurado causa que un sistema de ordenador funcione de una manera específica y predefinida para llevar a cabo las funciones descritas en el presente documento.

Se han descrito un cierto número de realizaciones de la invención. No obstante, se entenderá que pueden ser hechas diversas modificaciones sin apartarse del alcance de la invención, según lo definido por las reivindicaciones adjuntas. Por ejemplo, algunas de las etapas descritas en el presente documento pueden ser independientes del orden y, por tanto, pueden ser llevadas a cabo en un orden distinto al descrito.

REIVINDICACIONES

- 1. Un método para controlar una característica específica del volumen sonoro de una señal de audio, en el cual la característica específica del volumen sonoro es bien una aproximación del volumen sonoro específico, que es una medida del volumen sonoro perceptivo como una función de la frecuencia y del tiempo, o bien una aproximación del volumen sonoro específico parcial, que es una medida del volumen sonoro perceptivo de la señal en presencia de una señal secundaria de interferencia, como una función de la frecuencia y del tiempo, que comprende:
- calcular, como una función de la señal de audio, una aproximación a un volumen sonoro específico de destino, en el que dicha aproximación a un volumen sonoro específico de destino se obtiene usando un banco de filtros en el cual hay menos bandas de las que se obtendrían usando un banco de filtros en el cual las bandas estuvieran uniformemente espaciadas en la escala de ERB con un espaciado deseado de ERB de 1,
- obtener parámetros de modificación utilizables para modificar la señal de audio de manera variable en cuanto a la
 frecuencia y/o al tiempo, a fin de reducir la diferencia entre su característica específica de volumen sonoro y la aproximación a un volumen sonoro específico de destino, y
 - a) aplicar los parámetros de modificación a la señal de audio para reducir la diferencia entre su característica específica de volumen sonoro y la aproximación a un volumen sonoro específico de destino, o
 - b) transmitir o almacenar los parámetros de modificación y la señal de audio para la aplicación, temporalmente y/o espacialmente separada, de los parámetros de modificación a la señal de audio, para reducir la diferencia entre su característica específica de volumen sonoro y la aproximación a un volumen sonoro específico de destino.
- 25 2. Un método según la reivindicación 1, en el cual cada banda en dicho banco de filtros es mayor que 1 ERB.

20

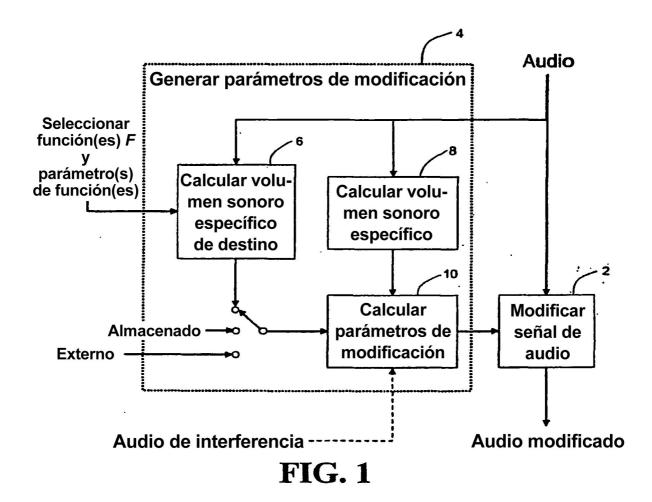
30

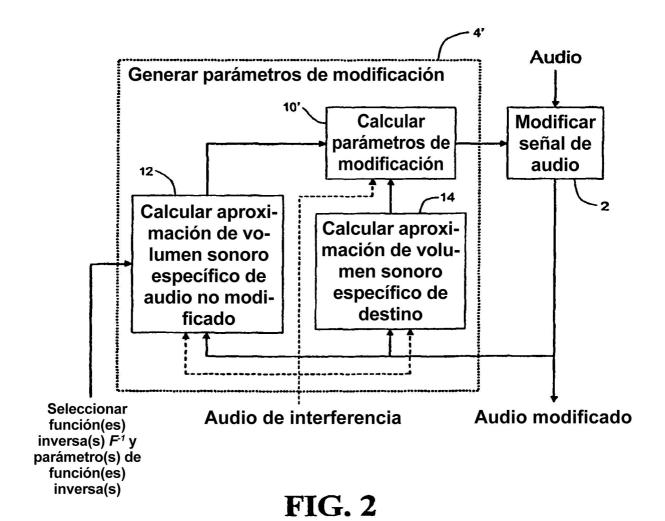
40

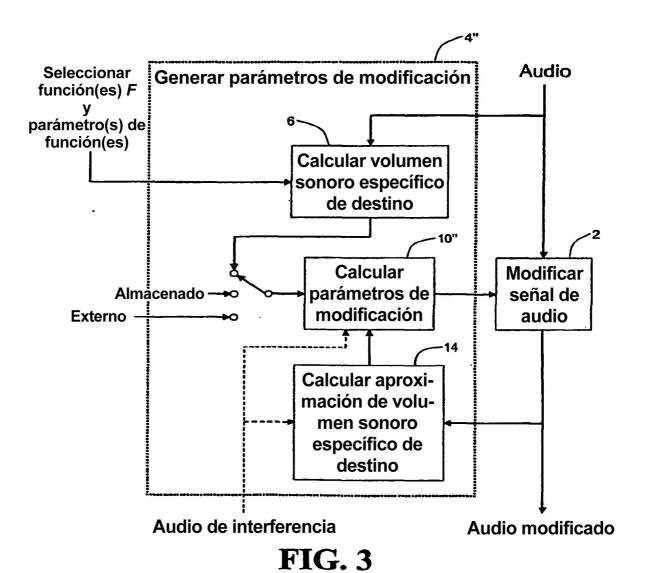
45

50

- 3. Un método para controlar una característica específica del volumen sonoro de una señal de audio, en el cual la característica específica del volumen sonoro es bien una aproximación del volumen sonoro específico, que es una medida del volumen sonoro perceptivo como una función de la frecuencia y del tiempo, o bien una aproximación del volumen sonoro específico parcial, que es una medida del volumen sonoro perceptivo de la señal de audio en presencia de una señal secundaria de interferencia como una función de la frecuencia y del tiempo, que comprende:
- recibir de una transmisión, o reproducir desde un medio de almacenamiento, la señal de audio y
- a) parámetros de modificación para modificar la señal de audio, habiendo sido obtenidos los parámetros de modificación a partir de una aproximación a un volumen sonoro específico de destino que emplea un banco de filtros en el cual hay menos bandas de las que se obtendrían usando un banco de filtros en el cual las bandas estuvieran uniformemente espaciadas en la escala de ERB, con un espaciado deseado de ERB de 1, habiendo sido calculada dicha aproximación como una función de la señal de audio, o
 - b) una aproximación a un volumen sonoro específico de destino que emplea un banco de filtros en el cual hay menos bandas de las que se obtendrían usando un banco de filtros en el cual las bandas estuvieran uniformemente espaciadas en la escala de ERB, con un espaciado deseado de ERB de 1, o una representación de la aproximación a ese volumen sonoro específico de destino, habiendo sido calculada dicha aproximación como una función de la señal de audio, y
 - modificar, de manera variable en cuanto a la frecuencia y/o al tiempo, la señal de audio en respuesta a a) los parámetros de modificación recibidos o b) los parámetros de modificación obtenidos de dicha aproximación a un volumen sonoro específico de destino, o su representación, a fin de reducir la diferencia entre la característica específica del volumen sonoro de la señal de audio y la aproximación a un volumen sonoro específico de destino.
 - 4. Un método según la reivindicación 3, en el cual cada banda en dicho banco de filtros que es empleada en la aproximación al volumen sonoro específico de destino es mayor que 1 ERB.
- 5. Un método según cualquiera de las reivindicaciones 1 a 4, en el cual las bandas en el banco de filtros están espaciadas no uniformemente en una escala crítica de velocidades de banda.
 - 6. Aparato adaptado para realizar todas las etapas del método de cualquiera de las reivindicaciones 1 a 5.
- 7. Un programa de ordenador, almacenado en un medio legible por ordenador, adaptado para hacer que un ordenador realice todas las etapas del método de cualquiera de las reivindicaciones 1 a 5.







34

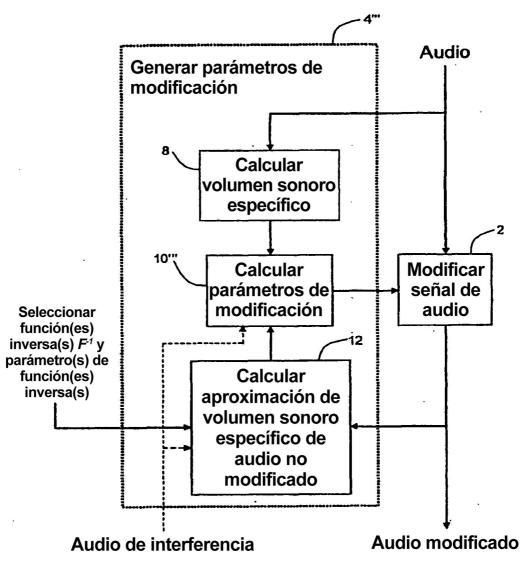


FIG. 4

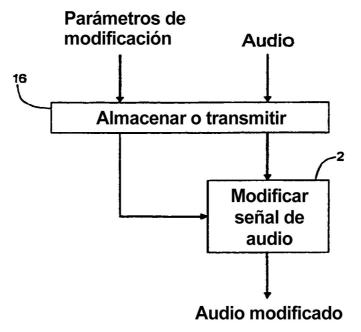
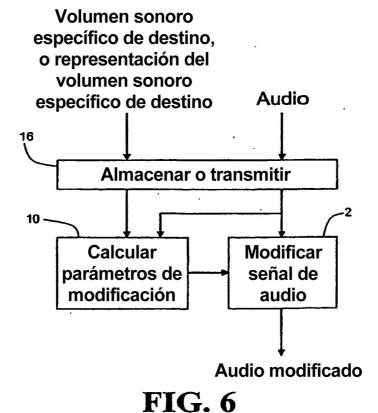


FIG. 5



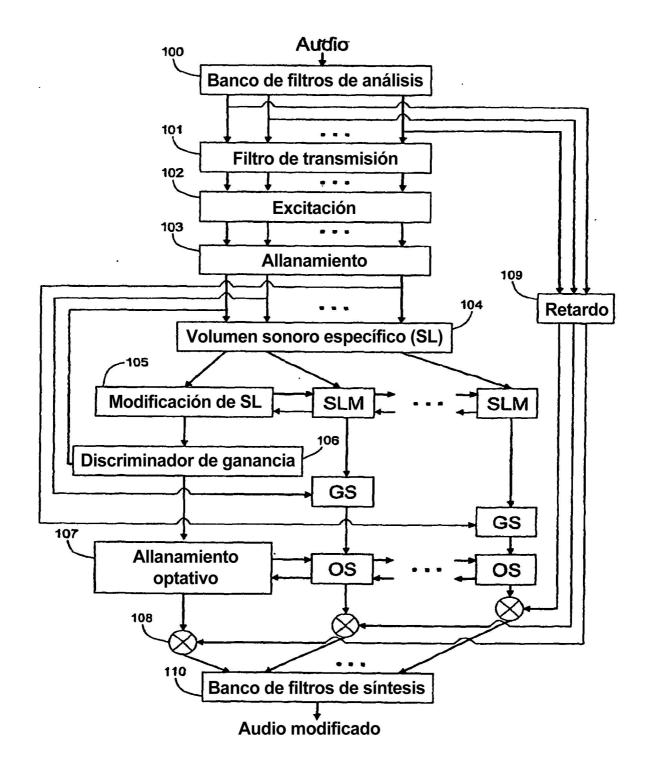


FIG. 7

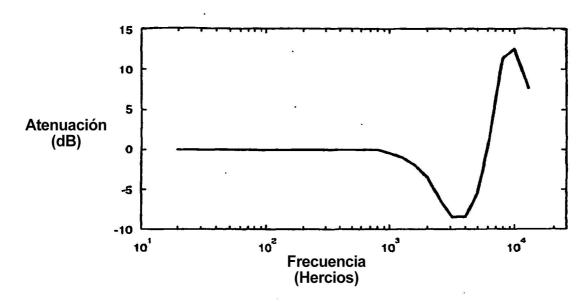


FIG. 8

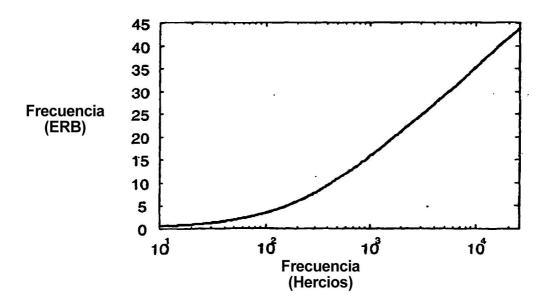


FIG. 9

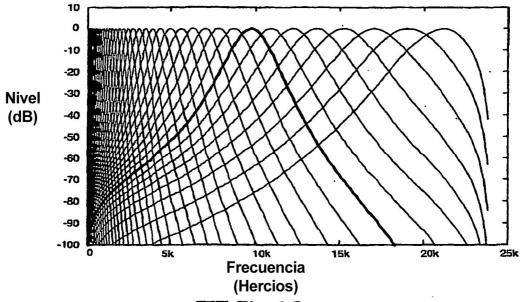


FIG. 10

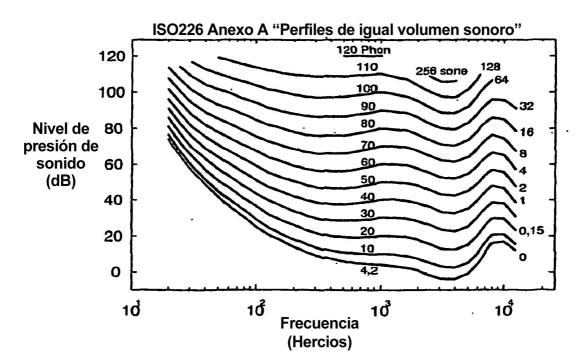


FIG. 11

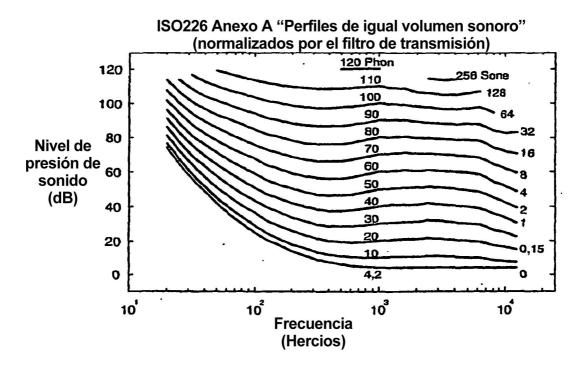


FIG. 12

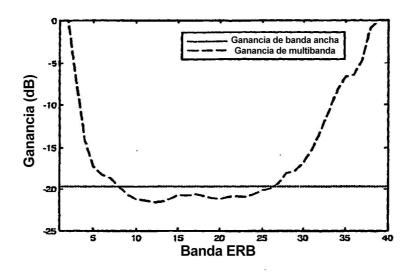


FIG. 13a

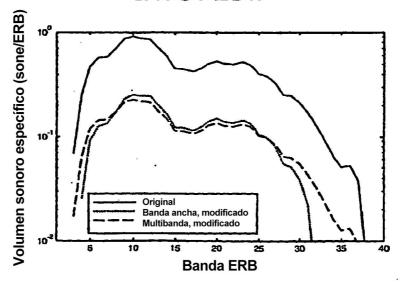
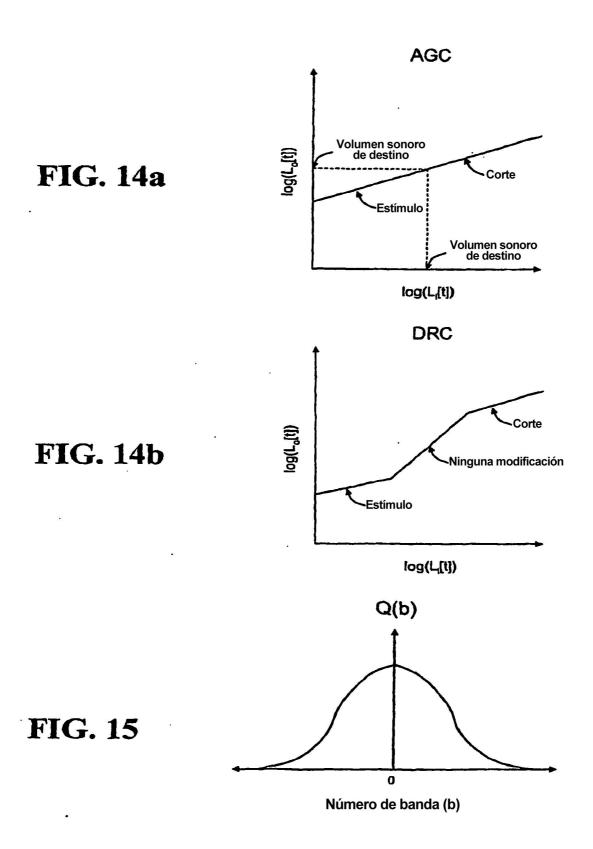


FIG. 13b



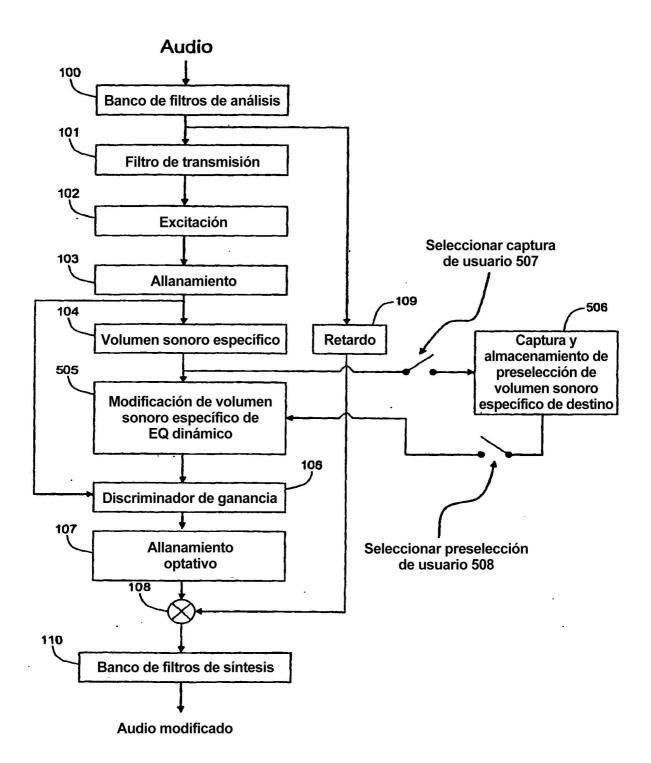


FIG. 16

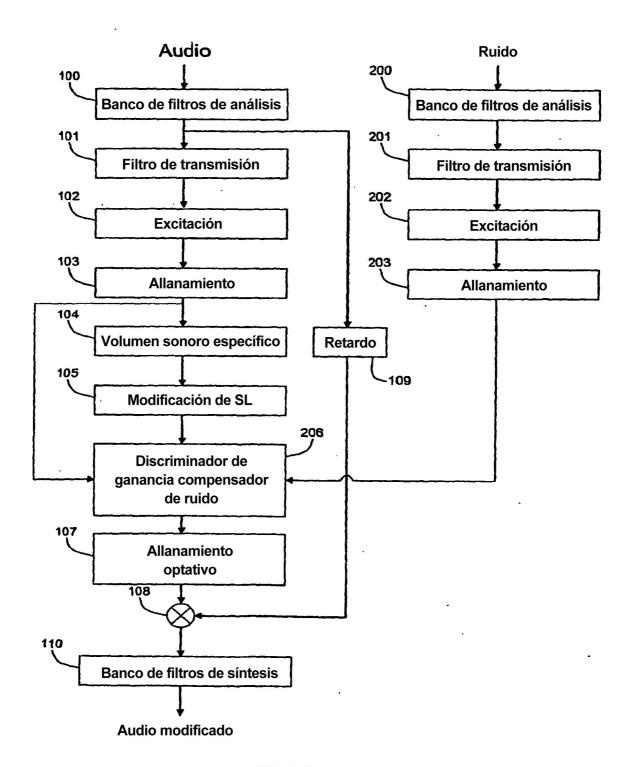


FIG. 17