(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 359 799**

(51) Int. Cl.:
H03G 3/30 (2006.01)
H03G 7/00 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Número de solicitud europea: **07754779 .2**
(96) Fecha de presentación : **30.03.2007**
(97) Número de publicación de la solicitud: **2011234**
(97) Fecha de publicación de la solicitud: **07.01.2009**

(54) Título: **Control de ganancia de audio usando detección de eventos auditivos basada en la sonoridad específica.**

(30) Prioridad: **27.04.2006 US 795808 P**

(45) Fecha de publicación de la mención BOPI:
27.05.2011

(45) Fecha de la publicación del folleto de la patente:
27.05.2011

(73) Titular/es: **DOLBY LABORATORIES LICENSING CORPORATION**
100 Potrero Avenue
San Francisco, California 94103-4813, US

(72) Inventor/es: **Crockett, Brett, Graham y**
Seefeldt, Alan, Jeffrey

(74) Agente: **Torner Lasalle, Elisabet**

DESCRIPCIÓN

Control de ganancia de audio usando detección de eventos auditivos basada en la sonoridad específica.

Campo Técnico

La presente invención se refiere a métodos y aparatos para controlar el rango dinámico de audio en los que un dispositivo de procesamiento de audio analiza una señal de audio y cambia el nivel, ganancia y rango dinámico del audio, y todos o algunos de los parámetros del procesamiento dinámico y de ganancia de audio se generan en función de eventos auditivos. La invención también se refiere a programas informáticos para poner en práctica tales métodos o controlar tales aparatos.

La presente invención también se refiere a métodos y aparatos que usan una detección de eventos auditivos basada en la sonoridad específica. La invención también se refiere a programas informáticos para poner en práctica tales métodos o controlar tales aparatos.

Antecedentes de la Técnica

Procesamiento dinámico de audio

Las técnicas de control automático de ganancia (AGC) y control de rango dinámico (DRC) son muy conocidas y son un elemento común de muchas trayectorias de señal de audio. En un sentido abstracto, ambas técnicas de alguna manera miden el nivel de una señal de audio y entonces modifican en ganancia la señal en una cantidad en función del nivel medido. En un sistema de procesamiento dinámico 1:1, lineal, la entrada de audio no se procesa y la señal de audio de salida idealmente coincide con la señal de audio de entrada. Adicionalmente, si se tiene un sistema de procesamiento dinámico de audio que automáticamente mide las características de la señal de entrada y usa esa medición para controlar la señal de salida, si la señal de entrada aumenta de nivel en 6 dB y la señal de salida se procesa de tal manera que solamente aumenta de nivel en 3 dB, entonces la señal de salida se ha comprimido en una razón de 2:1 con respecto a la señal de entrada. La publicación internacional número WO 2006/047600 A1 (“Calculating and Adjusting the Perceived Loudness and/or the Perceived Spectral Balance of an Audio Signal” de Alan Jeffrey Seefeldt) proporciona una detallada visión general de los cinco tipos básicos de procesamiento dinámico de audio: compresión, limitación, control automático de ganancia (AGC), expansión y conmutación de compuertas (*gating*).

Eventos auditivos y detección de eventos auditivos

La división de sonidos en unidades o segmentos percibidos como separados y distintos se denomina a veces “análisis de eventos auditivos” o “análisis de la escena auditiva” (“ASA”) y los segmentos se denominan a veces “eventos auditivos” o “eventos de audio”. Albert S. Bregman expone en su libro Auditory Scene Analyis -- The Perceptual Organization of Sound, Instituto de Tecnología de Massachussets, 1991, Cuarta edición, 2001, (Segunda edición en tapa blanda MIT Press) una amplia explicación sobre el análisis de la escena auditiva. Además, la patente estadounidense n.º 6.002.776 de Bhadkamkar, et al, 14 de diciembre de 1990, cita publicaciones con fecha desde 1976 como “trabajos de la técnica anterior relativos a la separación de sonido mediante análisis de la escena auditiva”. Sin embargo, la patente de Bhadkamkar, et al, desalienta el uso práctico del análisis de la escena auditiva, concluyendo que “las técnicas implicadas en el análisis de la escena auditiva, aunque interesantes desde un punto de vista científico como modelos de procesamiento auditivo humano, son actualmente demasiado exigentes computacionalmente y especializadas para ser consideradas técnicas prácticas de separación de sonido hasta que se produzcan avances fundamentales”.

Crockett and Crockett et al exponen en varias solicitudes de patente y documentos mencionados a continuación bajo la cabecera de “Incorporación como referencia”, una manera útil de identificar eventos auditivos. Según esos documentos, una señal de audio se divide en eventos auditivos, cada uno de los cuales tiende a percibirse como separado y distinto, mediante la detección de cambios en la composición espectral (amplitud en función de la frecuencia) con respecto al tiempo. Esto puede hacerse, por ejemplo, calculando el contenido espectral de bloques de tiempo sucesivos de la señal de audio, calculando la diferencia en el contenido espectral entre bloques de tiempo sucesivos de la señal de audio, e identificando un límite de evento auditivo como el límite entre bloques de tiempo sucesivos cuando la diferencia en el contenido espectral entre tales bloques de tiempo sucesivos excede un umbral. Alternativamente, pueden calcularse cambios en la amplitud con respecto al tiempo en lugar o además de los cambios en la composición espectral con respecto al tiempo.

En su implementación con menor exigencia computacional, el proceso divide el audio en segmentos de tiempo analizando toda la banda de frecuencia (audio de ancho de banda completo) o sustancialmente toda la banda de frecuencia (en implementaciones prácticas, a menudo se emplea filtrado de limitación de banda en los extremos del espectro) y otorgando el mayor peso a las componentes de señal de audio más intensas. Este enfoque aprovecha un fenómeno psicoacústico en el que a escalas de tiempo más pequeñas (20 milisegundos (ms) y menos) el oído puede tender a enfocar un solo evento auditivo en un momento dado. Esto implica que, aunque puedan ocurrir múltiples

eventos a la vez, una componente tiende a ser la más prominente desde el punto de vista de la percepción y puede procesarse individualmente como si fuese el único evento que está teniendo lugar. Aprovechando este efecto, también permite la detección de eventos auditivos a escala con la complejidad del audio que está procesándose. Por ejemplo, si la señal de audio de entrada que está procesándose es un instrumento solista, los eventos de audio que se identifiquen probablemente serán las notas individuales que están tocándose. De manera similar, para una señal de voz de entrada, las componentes individuales de la locución, las vocales y consonantes por ejemplo, probablemente se identificarán como elementos de audio individuales. A medida que aumenta la complejidad del audio, tal como música con toques de tambor o múltiples instrumentos y voz, la detección de eventos auditivos identifica el elemento de audio "más prominente" (es decir, el más intenso) en un momento dado.

A expensas de una complejidad computacional mayor, el proceso puede también tomar en consideración cambios en la composición espectral con respecto al tiempo en subbandas de frecuencia discretas (subbandas fijas o determinadas de manera dinámica o tanto fijas como determinadas de manera dinámica) en lugar del ancho de banda completo. Este enfoque alternativo tiene en cuenta más de un flujo de audio en subbandas de frecuencia diferentes en lugar de asumir que solamente puede percibirse un único flujo en un momento particular.

La detección de eventos auditivos puede implementarse dividiendo una forma de onda de audio en el dominio del tiempo en intervalos o bloques de tiempo y entonces convirtiendo los datos en cada bloque al dominio de la frecuencia, usando o bien un banco de filtros o bien una transformación de tiempo-frecuencia, tal como la FFT. La amplitud del contenido espectral de cada bloque puede normalizarse con el fin de eliminar o reducir el efecto de cambios de amplitud. Cada representación en el dominio de la frecuencia resultante proporciona una indicación del contenido espectral del audio en el bloque particular. El contenido espectral de bloques sucesivos se compara y pueden tomarse cambios mayores que un umbral para indicar el inicio temporal o el fin temporal de un evento auditivo.

Preferiblemente, los datos en el dominio de la frecuencia se normalizan tal como se describe a continuación. El grado al que los datos en el dominio de la frecuencia tienen que normalizarse da una indicación de amplitud. Por consiguiente, si un cambio en este grado excede un predeterminado umbral, ello también puede tomarse para indicar un límite de evento. A los puntos de inicio y fin de evento que resultan de cambios espectrales y de cambios de amplitud se les puede aplicar conjuntamente una operación O, de modo que se identifiquen límites de evento resultantes de cualquiera de los tipos de cambio.

Aunque las técnicas descritas en dichas solicitudes y documentos de Crockett and Crockett et al son particularmente útiles en conexión con aspectos de la presente invención, otras técnicas para identificar eventos auditivos y límites de evento pueden emplearse en aspectos de la presente invención.

Descripción de la invención

El procesamiento dinámico de audio de la técnica anterior convencional implica multiplicar el audio por una señal de control que varía en el tiempo que ajusta la ganancia del audio produciendo un resultado deseado. "Ganancia" es un factor de ajuste a escala que ajusta a escala la amplitud de audio. Esta señal de control puede generarse de manera continua o a partir de bloques de datos de audio, pero generalmente se deriva mediante alguna forma de medición del audio que está procesándose, y su tasa de cambio se determina mediante filtros de alisado, a veces con características fijas y a veces con características que varían con la dinámica del audio. Por ejemplo, los tiempos de respuesta pueden ser ajustables según cambios en la magnitud o la potencia del audio. Los métodos de la técnica anterior, tales como el control automático de ganancia (AGC) y la compresión de rango dinámico (DRC), no evalúan de ninguna forma basada en la psicoacústica los intervalos de tiempo durante los cuales los cambios de ganancia pueden percibirse como deficiencias y cuándo pueden aplicarse sin conllevar artefactos audibles. Por lo tanto, los procesos dinámicos de audio convencionales pueden a menudo introducir artefactos audibles, por ejemplo, los efectos del procesamiento dinámico pueden introducir cambios perceptibles no deseados en el audio.

El análisis de la escena auditiva identifica eventos auditivos discretos desde el punto de vista de la percepción, ocurriendo cada evento entre dos límites de evento auditivo consecutivos. Las deficiencias audibles causadas por un cambio de ganancia pueden reducirse en gran medida garantizando que dentro de un evento auditivo la ganancia es casi constante y restringiendo la mayor parte del cambio a la proximidad de un límite de evento. En el contexto de compresores o expansores, la respuesta a un aumento en el nivel de audio (a menudo llamado ataque) puede ser rápida, comparable con o más breve que la duración mínima de eventos auditivos, pero la respuesta a una disminución (la liberación o recuperación) puede ser más lenta de modo que los sonidos que deberían aparecer constantes o decrecer gradualmente pueden perturbarse de manera audible. Bajo tales circunstancias, es muy beneficioso retardar la recuperación de ganancia hasta el siguiente límite o ralentizar la tasa de cambio de ganancia durante un evento. Para aplicaciones de control automático de ganancia, en las que el nivel o la sonoridad a medio y largo plazo del audio se normaliza y tanto el tiempo de ataque como el de liberación pueden por lo tanto ser largos en comparación con la duración mínima de un evento auditivo, es beneficioso durante los eventos retardar los cambios o ralentizar las tasas de cambio en ganancia hasta el siguiente límite de evento tanto para ganancias en aumento como en disminución.

Según un aspecto de la presente invención, un sistema de procesamiento de audio recibe una señal de audio y analiza y altera las características de ganancia y/o de rango dinámico del audio. La modificación del rango dinámico del audio se controla a menudo mediante parámetros de un sistema de procesamiento dinámico (tiempo de ataque y liberación, razón de compresión, etc.) que tienen efectos significativos en los artefactos perceptivos introducidos por el procesamiento dinámico. Cambios en las características de señal con respecto al tiempo en la señal de audio se detectan e identifican como límites de evento auditivo, de manera que un segmento de audio entre límites consecutivos constituye un evento auditivo en la señal de audio. Las características de los eventos auditivos de interés pueden incluir características de los eventos tales como intensidad perceptiva o duración. Algunos de dichos uno o más parámetros de procesamiento dinámico se generan al menos en parte en respuesta a eventos auditivos y/o al grado de cambio en características de señal asociadas con dichos límites de evento auditivo.

Normalmente, un evento auditivo es un segmento de audio que tiende a percibirse como separado y distinto. Una medida que puede usarse de características de señal incluye una medida del contenido espectral del audio, por ejemplo, tal como se describe en los citados documentos de Crockett and Crockett et al. Todos o algunos de los uno o más parámetros de procesamiento dinámico de audio pueden generarse al menos en parte en respuesta a la presencia o ausencia de características de uno o más eventos auditivos. Un límite de evento auditivo puede identificarse como un cambio en las características de señal con respecto al tiempo que excede un umbral. Alternativamente, todos o algunos de los uno o más parámetros pueden generarse al menos en parte en respuesta a una medida continua del grado de cambio en las características de señal asociadas con dichos límites de evento auditivo. Aunque, en principio, los aspectos de la invención pueden implementarse en los dominios analógico y/o digital, las implementaciones prácticas es probable que se implementen en el dominio digital en el que cada una de las señales de audio se representa por muestras individuales o muestras dentro de bloques de datos. En este caso, las características de señal pueden ser el contenido espectral de audio dentro de un bloque, la detección de cambios en las características de señal con respecto al tiempo pueden ser la detección de cambios en el contenido espectral de audio de un bloque a otro, y los límites de inicio y detención temporal de evento auditivo coinciden cada uno con un límite de un bloque de datos. Debe observarse que, para el caso más tradicional de realización de cambios de ganancia dinámicos muestra a muestra, el análisis de la escena auditiva descrito puede realizarse por bloques y la información de evento auditivo resultante usarse para realizar cambios de ganancia dinámicos que se aplican muestra a muestra.

Al controlar los parámetros clave del procesamiento dinámico de audio usando los resultados del análisis de la escena auditiva, puede conseguirse una reducción drástica de artefactos audibles introducidos mediante procesamiento dinámico.

La presente invención presenta dos maneras de realizar análisis de eventos auditivos. La primera realiza análisis espectral e identifica la ubicación de eventos de audio perceptibles que se usan para controlar los parámetros de ganancia dinámicos mediante la identificación de cambios en el contenido espectral. La segunda manera transforma el audio en un dominio de sonoridad perceptiva (que puede proporcionar información más relevante desde el punto de vista psicoacústico que la primera manera) e identifica la ubicación de eventos auditivos que posteriormente se usan para controlar los parámetros de ganancia dinámicos. Debe observarse que la segunda manera requiere que el procesamiento de audio sea conciente de los niveles de reproducción acústica absolutos, lo que puede no ser posible en algunas implementaciones. El hecho de presentar ambos métodos de análisis de la escena auditiva permite implementaciones de modificación de ganancia dinámica controlada por ASA usando procesos o dispositivos que pueden o no estar calibrados para tener en cuenta niveles de reproducción absolutos.

En el presente documento se describen aspectos de la presente invención en un entorno de procesamiento dinámico de audio que incluye aspectos de otras invenciones. Tales otras invenciones se describen en varias solicitudes de patente internacionales y estadounidenses en tramitación de Dolby Laboratories Licensing Corporation, el titular de la presente solicitud, solicitudes que se identifican en el presente documento.

Descripción de los dibujos

La figura 1 es un diagrama de flujo que muestra un ejemplo de las etapas de procesamiento para realizar el análisis de la escena auditiva.

La figura 2 muestra un ejemplo de procesamiento de bloques, división en ventanas y realización de DFT en audio mientras se realiza el análisis de la escena auditiva.

La figura 3 es en forma de un diagrama de flujo o diagrama de bloques funcional, que muestra el procesamiento paralelo en el que el audio se usa para identificar eventos auditivos y para identificar las características de los eventos auditivos de manera que el evento y sus características se usan para modificar parámetros de procesamiento dinámico.

La figura 4 es en forma de un diagrama de flujo o diagrama de bloques funcional, que muestra el procesamiento en el que el audio se usa solamente para identificar eventos auditivos y las características de los eventos se determinan a partir de la detección de eventos de audio de manera que los eventos y sus características se usan para modificar los parámetros de procesamiento dinámico.

La figura 5 es en forma de un diagrama de flujo o diagrama de bloques funcional, que muestra el procesamiento en el que el audio se usa solamente para identificar eventos auditivos y las características de los eventos se determinan a partir de la detección de eventos de audio y de manera que solamente se usan las características de los eventos auditivos para modificar los parámetros de procesamiento dinámico.

La figura 6 muestra un conjunto de respuestas características de filtro auditivo idealizadas que aproximan la banda crítica en la escala ERB. La escala horizontal es la frecuencia en hercios y la escala vertical es el nivel en decibelios.

La figura 7 muestra los contornos de igual sonoridad de la norma ISO 226. La escala horizontal es la frecuencia en hercios (escala logarítmica de base 10) y la escala vertical es el nivel de presión sonora en decibelios.

Las figuras 8a-c muestran características de entrada/salida idealizadas y características de ganancia de entrada de un compresor de rango dinámico de audio.

Las figuras 9a-f muestran un ejemplo del uso de eventos auditivos para controlar el tiempo de liberación en una implementación digital de un controlador de rango dinámico (DRC) tradicional en el que el control de ganancia se deriva de la potencia de raíz cuadrática media (RMS) de la señal.

Las figuras 10a-f muestran un ejemplo del uso de eventos auditivos para controlar el tiempo de liberación en una implementación digital de un controlador de rango dinámico (DRC) tradicional en el que el control de ganancia se deriva de la potencia de raíz cuadrática media (RMS) de la señal para una señal alternativa a la usada en la figura 9.

La figura 11 representa un conjunto apropiado de curvas AGC y DRC idealizadas para la aplicación de AGC seguido por DRC en un sistema de procesamiento dinámico en el dominio de la sonoridad. La finalidad de la combinación es hacer que todo el audio procesado tenga aproximadamente la misma sonoridad percibida al tiempo que aún mantiene al menos parte de la dinámica del audio original.

Mejor modo de llevar a cabo la invención

Análisis de la escena auditiva (método no en el dominio de la sonoridad, original)

Según una realización de un aspecto de la presente invención, el análisis de la escena auditiva puede componerse de cuatro etapas de procesamiento general, tal como se muestra en una parte de la figura 1. La primera etapa 1-1 (“Realizar análisis espectral”) toma una señal de audio en el dominio del tiempo, la divide en bloques y calcula un perfil espectral o contenido espectral para cada uno de los bloques. El análisis espectral transforma la señal de audio en el domino de la frecuencia a corto plazo. Esto puede realizarse usando cualquier banco de filtros, ya sea basándose en transformadas o bancos de filtros pasa banda, y en cualquier espacio de frecuencia lineal o distorsionada (tal como la banda crítica o escala de Bark, que aproxima mejor las características del oído humano). Con cualquier banco de filtros existe un equilibrio entre el tiempo y la frecuencia. Una mayor resolución de tiempo, y por lo tanto intervalos de tiempo más cortos, lleva a una resolución de frecuencia más baja. Una mayor resolución de frecuencia, y por lo tanto subbandas más estrechas, lleva a intervalos de tiempo más largos.

La primera etapa, ilustrada conceptualmente en la figura 1, calcula el contenido espectral de segmentos de tiempo sucesivos de la señal de audio. En una realización práctica, el tamaño de bloque de ASA puede ser de cualquier número de muestras de la señal de audio de entrada, aunque 512 muestras proporcionan un buen equilibrio de resolución de tiempo y frecuencia. En la segunda etapa 1-2, se determinan las diferencias en el contenido espectral de un bloque a otro (“Realizar mediciones de diferencia de perfil espectral”). De esta manera, la segunda etapa calcula la diferencia en contenido espectral entre segmentos de tiempo sucesivos de la señal de audio. Como se mencionó anteriormente, se cree que un potente indicador del inicio o fin de un evento auditivo percibido es un cambio en el contenido espectral. En la tercera etapa 1-3 (“Identificar ubicación de límites de evento auditivo”), cuando la diferencia espectral entre un bloque de perfil espectral y el siguiente es mayor que un umbral, el límite de bloque se toma como un límite de evento auditivo. El segmento de audio entre límites consecutivos constituye un evento auditivo. De esta manera, la tercera etapa establece un límite de evento auditivo entre segmentos de tiempo sucesivos cuando la diferencia en el contenido de perfil espectral entre tales segmentos de tiempo sucesivos excede un umbral, definiendo de esta manera eventos auditivos. En esta realización, los límites de evento auditivo definen eventos auditivos que tienen una longitud que es un múltiplo entero de bloques de perfil espectral con una longitud mínima de un bloque de perfil espectral (512 muestras en este ejemplo). En principio, los límites de evento no tienen que estar limitados de este modo. Como una alternativa a las realizaciones prácticas mencionadas en el presente documento, el tamaño de bloque de entrada puede variar, por ejemplo, para tener básicamente del tamaño de un evento auditivo.

Tras la identificación de los límites de evento, se identifican características clave de los eventos auditivos, tal como se muestra en la etapa 1-4.

Segmentos solapados o no solapados del audio pueden dividirse en ventanas y usarse para calcular perfiles espectrales del audio de entrada. El solapamiento da como resultado una resolución más fina en cuanto a la ubicación de eventos auditivos y, también, hace que sea menos probable perder un evento, tal como un transitorio breve. Sin embargo, el solapamiento también aumenta la complejidad computacional. De esta manera, el solapamiento puede omitirse. La figura 2 muestra una representación conceptual de N bloques de muestra no solapados divididos en ventanas y transformados al dominio de la frecuencia mediante la transformada discreta de Fourier (DFT). Cada bloque puede dividirse en ventanas y transformarse al dominio de la frecuencia, por ejemplo usando la DFT, preferiblemente implementada como una transformada rápida de Fourier (FFT) para mayor rapidez.

Las siguientes variables pueden usarse para calcular el perfil espectral del bloque de entrada:

M=número de muestras divididas en ventanas en un bloque usado para calcular el perfil espectral.

P=número de muestras de solapamiento de cálculo espectral

En general, cualquier número entero puede usarse para las variables anteriores. Sin embargo, la implementación será más eficaz si M se establece igual a una potencia de 2 de modo que puedan usarse FFT estándar para los cálculos de perfil espectral. En una realización práctica del proceso de análisis de la escena auditiva, los parámetros enumerados pueden establecerse en:

M=512 muestras (o 11,6 ms a 44,1 kHz)

P=0 muestras (sin solapamiento)

Los valores anteriormente enumerados fueron determinados experimentalmente y se encontró que generalmente identificaban con suficiente precisión la ubicación y duración de eventos auditivos. Sin embargo, establecer el valor P a 256 muestras (solapamiento del 50%) en lugar que a cero muestras (sin solapamiento) ha resultado ser útil en la identificación algunos eventos difíciles de encontrar. Aunque pueden usarse muchos tipos diferentes de ventanas para minimizar los artefactos espectrales debidos a la división en ventanas, la ventana usada en los cálculos de perfil espectral es una ventana de Kaiser-Bessel, de Hanning de M puntos, u otra apropiada, preferiblemente no rectangular. Los valores indicados anteriormente y un tipo de ventana de Hanning se seleccionaron después de un análisis experimental extensivo ya que mostraron que proporcionaban excelentes resultados en una amplia gama de material de audio. La división en ventanas no rectangulares se prefiere para el procesamiento de señales de audio con contenido predominantemente de baja frecuencia. La división en ventanas rectangulares produce artefactos espectrales que pueden causar una detección incorrecta de eventos. A diferencia de ciertas aplicaciones de codificador/decodificador (códec) en las que un proceso de solapamiento/suma global debe proporcionar un nivel constante, tal restricción no se aplica aquí y la ventana puede escogerse por características tales como su resolución de tiempo/frecuencia y rechazo de supresión de banda.

En la etapa 1-1 (figura 1), el espectro de cada bloque de M muestras puede calcularse mediante la división en ventanas de los datos con una ventana de Kaiser-Bessel, de Hanning de M puntos, u otra apropiada, la conversión al dominio de la frecuencia usando una transformada rápida de Fourier de M puntos, y calculando la magnitud de los coeficientes FFT complejos. Los datos resultantes se normalizan de manera que la mayor magnitud se establece en la unidad, y la matriz normalizada de M números se convierte al dominio logarítmico. Los datos pueden también normalizarse mediante alguna otra métrica tal como el valor de magnitud media o el valor de potencia media de los datos. La matriz no tiene que convertirse al dominio logarítmico, pero la conversión simplifica el cálculo de la medida de diferencia en la etapa 1-2. Además, el dominio logarítmico se adapta de manera más próxima a la naturaleza del sistema auditivo humano. Los valores en el dominio logarítmico resultantes tienen un rango de menos infinito a cero. En una realización práctica, puede imponerse un límite más bajo al rango de valores; el limite puede ser fijo, por ejemplo -60 dB, o ser dependiente en frecuencia para reflejar la audibilidad más baja de sonidos suaves a frecuencias bajas y muy altas. (Obsérvese que sería posible reducir el tamaño de la matriz a M/2 en la que la FFT representa frecuencias negativas así como positivas).

La etapa 1-2 calcula una medida de la diferencia entre los espectros de bloques adyacentes. Para cada bloque, cada uno de los M coeficientes espectrales (logarítmicos) de la etapa 1-1 se resta del coeficiente correspondiente para el bloque anterior, y la magnitud de la diferencia calculada (el signo se ignora). Esas M diferencias se suman entonces dando un número. Esta medida de diferencia puede también expresarse como una diferencia promedio por coeficiente espectral dividiendo la medida de diferencia por el número de coeficientes espectrales usados en la suma (en este caso M coeficientes).

La etapa 1-3 identifica las ubicaciones de límites de evento auditivo aplicando un umbral a la matriz de medidas de diferencia de la etapa 1-2 con un valor de umbral. Cuando una medida de diferencia excede un umbral, el cambio en el espectro se considera suficiente para señalar un nuevo evento y el número de bloque del cambio se registra como un límite de evento. Para los valores de M y P dados anteriormente y para valores del dominio logarítmico (en la etapa 1-1) expresados en unidades de dB, el umbral puede establecerse igual a 2500 si se compara la magnitud completa FFT

(incluyendo la parte reflejada) o 1250 si se compara la mitad de la FFT (como se observó anteriormente, la FFT representa frecuencias negativas así como positivas – para la magnitud de la FFT, una es la imagen reflejada de la otra). Este valor se escogió experimentalmente y proporciona una buena detección de límites de evento auditivo. Este valor de parámetro puede cambiarse para reducir (aumenta el umbral) o aumentar (disminuir el umbral) la detección de eventos.

El proceso de la figura 1 puede representarse de manera más general mediante las disposiciones equivalentes de las figuras 3, 4 y 5. En la figura 3, una señal de audio se aplica en paralelo a una función de "Identificar eventos auditivos" o etapa 3-1 que divide la señal de audio en eventos auditivos, cada uno de los cuales tiende a percibirse como separado y distinto, y a una función de "Identificar características de eventos auditivos" o etapa 3-2 opcional. El proceso de la figura 1 puede emplearse para dividir la señal de audio en eventos auditivos y sus características identificadas o algún otro proceso apropiado pueden emplearse. La información de eventos auditivos, que puede ser una identificación de límites de evento auditivo, determinada por la función o etapa 3-1 se usa entonces para modificar los parámetros de procesamiento dinámico de audio (tal como ataque, liberación, razón, etc.), según se desee, por una función "Modificar parámetros dinámicos" o etapa 3-3. La función "Identificar características" o etapa 3-3 opcional también recibe la información de eventos auditivos. La función "Identificar características" o etapa 3-3 puede caracterizar algunos o todos los eventos auditivos mediante una o más características. Tales características pueden incluir una identificación de la subbanda dominante del evento auditivo, tal como se describe en conexión con el proceso de la figura 1. Las características pueden también incluir una o más características de audio, incluyendo, por ejemplo, una medida de potencia del evento auditivo, una medida de amplitud del evento auditivo, una medida de planeidad espectral del evento auditivo, y si el evento auditivo es sustancialmente silencioso, u otras características que ayudan a modificar parámetros dinámicos de manera que se reducen o eliminan los artefactos audibles negativos del procesamiento. Las características también pueden incluir otras características tales como si el evento auditivo incluye un transitorio.

En las figuras 4 y 5 se muestran alternativas a la disposición de la figura 3. En la figura 4, la señal de entrada de audio no se aplica directamente a la función "Identificar características" o etapa 4-3, sino que recibe información de la función "Identificar eventos auditivos" o etapa 4-1. La disposición de la figura 1 es un ejemplo específico de tal disposición. En la figura 5, las funciones o etapas 5-1, 5-2 y 5-3 se disponen en serie.

Los detalles de esta realización práctica no son críticos. Pueden emplearse otras maneras de calcular el contenido espectral de segmentos de tiempo sucesivos de la señal de audio, calcular las diferencias entre segmentos de tiempo sucesivos, y establecer límites de evento auditivo en los respectivos límites entre segmentos de tiempo sucesivos cuando la diferencia en el contenido de perfil espectral entre tales segmentos de tiempo sucesivos excede un umbral.

Análisis de la escena auditiva (método en el dominio de la sonoridad, nuevo)

La solicitud internacional según el Tratado de Cooperación en materia de Patentes n.º de serie PCT/US2005/038579, presentada el 25 de octubre de 2005, publicada como publicación internacional número WO 2006/047600, titulada "Calculting and Adjusting the Perceived Loudness and/or the Perceived Spectral Balance of an Audio Signal" por Alan Jeffrey Seefeldt da a conocer, entre otras cosas, un medida objetiva de la sonoridad percibida en un modelo psicoacústico. Dicha solicitud se incorpora por la presente como referencia en su totalidad. Como se describe en dicha solicitud, a partir de una señal de audio, *x[n]*, se calcula una señal de excitación *E[b,t]* que aproxima la distribución de energía a lo largo de la membrana basilar del oído interno en la banda crítica *b* durante un bloque de tiempo *t*. Esta excitación puede calcularse a partir de la transformada discreta de Fourier de tiempo corto (STDFT) de la señal de audio como sigue:

$$E[b,t] = \lambda_b E[b,t-1] + (1-\lambda_b)\sum_k |T[k]|^2 |C_b[k]|^2 |X[k,t]|^2 \qquad (1)$$

donde *X[k,t]* representa la STDFT de *x[n]* en el bloque de tiempo *t* y el intervalo *k*. Obsérvese que en la ecuación 1, *t* representa tiempo en unidades discretas de bloques de transformada en oposición a una medición continua, tal como segundos. *T[k]* representa la respuesta en frecuencia de un filtro que simula la transmisión de audio a través del oído externo y medio, y $C_b[k]$ representa la respuesta en frecuencia de la membrana basilar en la ubicación correspondiente a la banda crítica *b*. La figura 6 representa un conjunto apropiado de respuestas de filtro de banda crítica en el que 40 bandas están espaciadas uniformemente a lo largo de la escala de ancho de banda rectangular equivalente (ERB), definida por Moore y Glasberg. Cada forma de filtro se describe por una función exponencial redondeada y las bandas están distribuidas usando un espaciamiento de 1 ERB. Por último, la constante de tiempo de alisamiento λ_b en la ecuación 1 puede ventajosamente escogerse proporcional al tiempo de integración de la percepción de la sonoridad humana en la banda *b*.

Al usar contornos de igual sonoridad, tales como los representados en la figura 7, la excitación en cada banda se transforma en un nivel de excitación que generará la misma sonoridad percibida a 1 kHz. La sonoridad específica, una medida de sonoridad perceptiva distribuida por tiempo y frecuencia, se calcula a partir de la excitación transformada,

$E_{1kHz}[b,t]$, a través de una no linealidad compresiva. Una función apropiada de este tipo para calcular la sonoridad específica $N[b,t]$ viene dada por:

$$N[b,t] = \beta\left(\left(\frac{E_{1kHz}[b,t]}{TQ_{1kHz}}\right)^{\alpha} - 1\right)$$

(2)

donde TQ_{1kHz} es el umbral en silencio a 1 kHz y las constantes β y α se escogen para adecuarse al crecimiento de datos de sonoridad a medida que se recopilan a partir de experimentos de audición. De manera abstracta, esta transformación de excitación a sonoridad específica puede presentarse por la función $\psi\{\}$ de tal manera que:

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

Finalmente, la sonoridad total, $L[t]$, representada en unidades de sonios, se calcula sumando la sonoridad específica a través de las bandas:

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

(3)

La sonoridad específica $N[b,t]$ es una representación espectral con la intención de simular la manera en que un humano percibe el audio como una función de tiempo y frecuencia. Captura variaciones en la sensibilidad a diferentes frecuencias, variaciones en la sensibilidad al nivel, y variaciones en la resolución de frecuencia. Como tal, es una representación espectral bien adaptada a la detección de eventos auditivos. Aunque es más complejo computacionalmente, comparar la diferencia de $N[b,t]$ a través de las bandas entre bloques de tiempo sucesivos puede en muchos casos dar como resultado una detección más precisa desde el punto de vista de la percepción de eventos auditivos en comparación con el uso directo de espectros FFT sucesivos descritos anteriormente.

En dicha solicitud de patente, se dan a conocer varias aplicaciones para modificar el audio basándose en este modelo de sonoridad psicoacústica. Entre estos se encuentran varios algoritmos de procesamiento dinámico, tales como AGC y el DRC. Estos algoritmos dados a conocer pueden beneficiarse del uso de eventos auditivos para controlar varios parámetros asociados. Puesto que la sonoridad específica ya se ha calculado, está fácilmente disponible para los propósitos de detectar dichos eventos. Detalles de una realización preferida se comentan a continuación.

Control de parámetros de procesamiento dinámico de audio con eventos auditivos

A continuación se presentan dos ejemplos de realizaciones de la invención. El primero describe el uso de eventos auditivos para controlar el tiempo de liberación en una implementación digital de un controlador de rango dinámico (DRC) en el que el control de ganancia se deriva a partir de la potencia de raíz cuadrática media (RMS) de la señal. La segunda realización describe el uso de eventos auditivos para controlar ciertos aspectos de una combinación más sofisticada de AGC y DRC implementada en el contexto del modelo de sonoridad psicoacústica descrito anteriormente. Estas dos realizaciones pretenden servir como ejemplos de la invención solamente, y debería entenderse que el uso de eventos auditivos para controlar parámetros de un algoritmo de procesamiento dinámico no se restringe a las especificaciones descritas a continuación.

Control de rango dinámico

La implementación digital descrita de un DRC segmenta una señal de audio $x[n]$ en bloques solapados hasta la mitad, divididos en ventanas, y para cada bloque se calcula una ganancia de modificación basándose en una medida de la potencia local de la señal y una curva de compresión seleccionada. La ganancia se alisa por los bloques y entonces se multiplica por cada bloque. Los bloques modificados finalmente se solapan y suman para generar la señal de audio modificada $y[n]$.

Debe observarse que, mientras que el análisis de la escena auditiva y la implementación digital de DRC según se ha descrito aquí dividen la señal de audio en el dominio del tiempo en bloques para realizar el análisis y el procesamiento, el procesamiento DRC no necesita realizarse usando segmentación en bloques. Por ejemplo el análisis de la escena auditiva puede realizarse usando segmentación en bloques y análisis espectral como se describió anteriormente y las características y ubicaciones de evento auditivo resultantes pueden usarse para proporcionar información de control a

una implementación digital de una implementación de DRC tradicional que normalmente opera muestra a muestra. Aquí, sin embargo, la misma estructura en bloques usada para el análisis de la escena auditiva se emplea para el DRC para simplificar la descripción de su combinación.

Procediendo con la descripción de una implementación de DRC basada en bloques, los bloques solapados de la señal de audio pueden representarse como:

$$x[n,t] = w[n]x[n + tM/2] \qquad \text{para} \qquad 0 < n < M - 1$$

(4)

donde M es la longitud de bloque y el tamaño de salto es M/2, *w[n]* es la ventana, *n* es el índice de muestra dentro del bloque, y *t* es el índice de bloque (obsérvese que aquí *t* se usa de la misma manera que con la STDFT en la ecuación 1; representa el tiempo en unidades discretas de bloques en lugar de en segundos, por ejemplo). De manera ideal, la ventana *w[n]* se estrecha hasta cero en ambos extremos y suma la unidad cuando está solapada hasta la mitad consigo misma; la ventana sinusoidal comúnmente usada cumple estos criterios, por ejemplo.

Para cada bloque, puede entonces calcularse la potencia RMS para generar una medida de potencia *P[t]* en dB por bloque:

$$P[t] = 10 * \log 10\left(\frac{1}{M}\sum_{n=1}^{M} x^2[n,t]\right)$$

(5)

Como se mencionó anteriormente, puede alisarse esta medida de potencia con un ataque rápido y una liberación lenta antes del procesamiento con una curva de compresión, aunque, como alternativa, la potencia instantánea *P[t]* se procesa y la ganancia resultante se alisa. Este enfoque alternativo tiene la ventaja de que puede usarse una curva de compresión simple con puntos de inflexión bruscos, pero las ganancias resultantes siguen siendo lisas ya que la potencia pasa a través del punto de inflexión. Al representar una curva de compresión, como se muestra en la figura 8c, como una función *F* del nivel de señal que genera una ganancia, la ganancia de bloque *G[t]* viene dada por:

$$G[t] = F\{P[t]\}$$

(6)

Suponiendo que la curva de compresión aplica una mayor atenuación a medida que aumenta el nivel de la señal, la ganancia disminuirá cuando la señal está en “modo de ataque” y aumentará cuando está en “modo de liberación”. Por lo tanto, una ganancia alisada $\overline{G}[t]$ puede calcularse según:

$$\overline{G}[t] = \alpha[t] \cdot \overline{G}[t-1] + (1 - \alpha[t])G[t]$$

(7a)

donde

$$\alpha[t] = \begin{cases} \alpha_{\text{ataque}} & G[t] < \overline{G}[t-1] \\ \alpha_{\text{liberación}} & G[t] \geq \overline{G}[t-1] \end{cases}$$

(7b)

y

$$\alpha_{\text{liberación}} >> \alpha_{\text{ataque}}$$

(7c)

Finalmente, la ganancia alisada $\overline{G}[t]$, que está en dB, se aplica a cada bloque de la señal, y los bloques modificados se solapan y suman para producir el audio modificado:

$$y[n+tM/2] = \left(10^{\overline{G}[t]/20}\right)x[n,t] + \left(10^{\overline{G}[t-1]/20}\right)x[n+M/2,t-1] \quad \text{for } 0 < n < M/2$$

(8)

Obsérvese que puesto que los bloques se han multiplicado por una ventana de forma ahusada, tal como se muestra en la ecuación 4, la síntesis de solapamiento-suma mostrada anteriormente alisa efectivamente la ganancia por las muestras de la señal procesada *y[t]*. De esta manera, la señal de control de ganancia recibe alisamiento además de lo mostrado en la ecuación 7a. En una implementación más tradicional de DRC que opera muestra a muestra en lugar que bloque a bloque, podría ser necesario un alisamiento de ganancia más sofisticado que el sencillo filtro de un polo mostrado en la ecuación 7a con el fin de evitar una distorsión audible en la señal procesada. También, el uso de un procesamiento basado en bloques introduce un retardo inherente de *M/2* muestras en el sistema, y siempre que el tiempo de decrecimiento asociado con α_{ataque} esté próximo a este retardo, la señal *x[n]* no tiene que retardarse más antes de la aplicación de las ganancias con el fin de evitar la sobremodulación.

Las figuras 9a a 9c representan el resultado de aplicar el procesamiento DRC descrito a una señal de audio. Para esta implementación particular, se usa una longitud de bloque *M*=512 a una tasa de muestreo de 44,1 kHz. Se usa una curva de compresión similar a la mostrada en la figura 8b: por encima de -20 dB, en relación a la escala digital completa, la señal se atenúa con una razón de 5:1, y por debajo de -30 dB, la señal se amplifica con una razón de 5:1. La ganancia se alisa con un coeficiente de ataque α_{ataque} correspondiente a una mitad de tiempo de decrecimiento de 10 ms y un coeficiente de liberación $\alpha_{liberación}$ correspondiente a una mitad de tiempo de decrecimiento de 500 ms. La señal de audio original representada en la figura 9a consiste en seis acordes de piano, con el acorde final, ubicado alrededor de la muestra $1{,}75 \times 10^5$, decreciendo hasta llegar al silencio. Al examinar un trazado de la ganancia $\overline{G}[t]$ en la figura 9b, debe observarse que la ganancia se mantiene próxima a 0 dB mientras se tocan los seis acordes.

Esto se debe a que la señal de energía se mantiene, en su mayor parte, entre -30 dB y -20 dB, región dentro de la cual la curva de DRC no requiere modificación alguna. Sin embargo, tras el último acorde, la energía de la señal cae por debajo de -30 dB y la ganancia comienza a elevarse, eventualmente más allá de los 15 dB, a medida que el acorde decrece. La figura 9c representa la señal de audio modificada resultante, y puede verse que la cola del acorde final se intensifica significativamente.

De manera audible, esta intensificación del sonido natural del acorde, con decrecimiento de nivel bajo, crea un resultado extremadamente antinatural. La finalidad de la presente invención es evitar problemas de este tipo que están asociados con un procesador dinámico tradicional.

Las figuras 10a a 10c representan los resultados de aplicar exactamente el mismo sistema de DRC a una señal de audio distinta. En este caso, la primera mitad de la señal consiste en una pieza musical de ritmo rápido a un nivel alto, y entonces aproximadamente en la muestra 10×10^4 la señal cambia a una segunda pieza musical de ritmo rápido, pero a un nivel significativamente más bajo. Examinando la ganancia en la figura 6b, se ve que la señal está atenuada en aproximadamente 10 dB durante la primera mitad, y entonces la ganancia vuelve a elevarse hasta 0 dB durante la segunda mitad cuando se toca la pieza más suave. En este caso, la ganancia se comporta según se desea. Se desea intensificar la segunda pieza con respecto a la primera, y la ganancia debería incrementar rápidamente después de la transición a la segunda pieza para que sea discreta desde el punto de vista audible. Se observa un comportamiento de la ganancia que es similar al de la primera señal mencionada, pero en este caso el comportamiento es deseable. Por tanto, se desearía solucionar el primer caso sin afectar al segundo. El uso de eventos auditivos para controlar el tiempo de liberación de este sistema DRC proporciona tal solución.

En la primera señal que se examinó en la figura 9, la intensificación del decrecimiento del último acorde parece antinatural porque el acorde y su decrecimiento se perciben como un evento auditivo independiente cuya integridad se espera que se mantenga. En el segundo caso, sin embargo, muchos eventos auditivos ocurren mientras aumenta la ganancia, lo que significa que para cualquier evento individual, tienen lugar pequeños cambios. Por lo tanto, el cambio de ganancia global no es tan cuestionable. Se puede por lo tanto argumentar que un cambio de ganancia sólo debería permitirse en la proximidad temporal de un límite de evento auditivo. Este principio puede aplicarse a la ganancia mientras se encuentra o bien en modo de ataque o bien de liberación, pero para las implementaciones más prácticas de DRC, la ganancia se mueve tan rápido en el modo de ataque en comparación con la resolución temporal humana de la percepción de eventos que no es necesario ningún control. Pueden usarse por lo tanto eventos para controlar el alisamiento de la ganancia de DRC solamente cuando esté en modo de liberación.

A continuación se describe un comportamiento apropiado del control de liberación. En términos cualitativos, si se detecta un evento, la ganancia se alisa con la constante de tiempo de liberación como se especificó anteriormente en la ecuación 7a. A medida que el tiempo evoluciona más allá del evento detectado, y si no se detectan eventos posteriores, la constante de tiempo de liberación aumenta de manera continua de modo que eventualmente la ganancia alisada se “congela” en su lugar. Si se detecta otro evento, entonces la constante de tiempo de alisamiento se reinicia al valor

original y el proceso se repite. Con el fin de modular el tiempo de liberación, puede generarse en primer lugar una señal de control basándose en los límites de evento detectados.

Como se mencionó anteriormente, los límites de evento pueden detectarse buscando cambios en los espectros sucesivos de la señal de audio. En esta implementación particular, la DFT de cada bloque solapado *x[n,t]* puede calcularse para generar la STDFT de la señal de audio *x[n]*:

$$X[k,t] = \sum_{n=0}^{M-1} x[n,t] e^{-j\frac{2\pi kn}{M}}$$

(9)

Luego, la diferencia entre los espectros de magnitud logarítmica normalizada de bloques sucesivos puede calcularse según:

$$D[t] = \sum_{k} \left| X_{NORM}[k,t] - X_{NORM}[k,t-1] \right|$$

(10a)

donde

$$X_{NORM}[k,t] = \log\left(\frac{|X[k,t]|}{\max_{k}\{|X[k,t]|\}} \right)$$

(10b)

Aquí, el máximo de $|X[k,t]|$ a través del intervalo *k* se usa para la normalización, aunque se pueden emplear otros factores de normalización; por ejemplo, el promedio de $|X[k,t]|$ a través de los intervalos. Si la diferencia *D[t]* excede un umbral D_{min}, entonces se considera que ha ocurrido un evento. Adicionalmente, puede asignarse una intensidad a este evento, situada entre cero y uno, basándose en el tamaño de *D[t]* en comparación con un umbral máximo D_{max}. La señal de intensidad de evento auditivo resultante *A[t]* puede calcularse como:

$$A[t] = \begin{cases} 0 & D[t] \le D_{\min} \\ \dfrac{D[t] - D_{\min}}{D_{\max} - D_{\min}} & D_{\min} < D[t] < D_{\max} \\ 1 & D[t] \ge D_{\max} \end{cases}$$

(11)

Al asignar una intensidad al evento auditivo proporcional a la cantidad de cambio espectral asociado con ese evento, se consigue un mayor control sobre el procesamiento dinámico en comparación con una decisión de evento binaria. Los inventores han encontrado que cambios de ganancia mayores son aceptables durante eventos más fuertes, y la señal en la ecuación 11 permite un control variable de este tipo.

La señal *A[t]* es una señal impulsiva, ocurriendo un impulso en la ubicación de un límite de evento. Para los propósitos de controlar el tiempo de liberación, puede además alisarse la señal *A[t]* de modo que decrezca suavemente hasta cero tras la detección de un límite de evento. La señal de control de evento alisada $\overline{A}[t]$ puede calcularse a partir de *A[t]* según:

$$\overline{A}[t] = \begin{cases} A[t] & A[t] > \alpha_{event}\overline{A}[t-1] \\ \alpha_{event}\overline{A}[t-1] & \text{si no} \end{cases}$$

(12)

Aquí, α_{event} controla el tiempo de decrecimiento de la señal de control de evento. Las figuras 9d y 10d representan la señal de control de evento $\overline{A}[t]$ para las dos señales de audio correspondientes, con la mitad del tiempo de decrecimiento de la más lisa ajustado a 250 ms. En el primer caso, se observa que se detecta un límite de evento para cada uno de los seis acordes de piano, y que la señal de control de evento decrece suavemente hacia cero después de cada evento. Para la segunda señal, se detectan muchos eventos muy próximos unos a otros en el tiempo, y por lo tanto la señal de control de evento nunca decrece completamente a cero.

Puede usarse ahora la señal de control de evento $\overline{A}[t]$ para variar la constante de tiempo de liberación usada para alisar la ganancia. Cuando la señal de control es igual a uno, el coeficiente de alisamiento *α[t]* de la ecuación 7a es igual a $\alpha_{liberación}$, como antes, y cuando la señal de control es igual a cero, el coeficiente es igual a uno de modo que se evita que cambie la ganancia alisada. El coeficiente de alisamiento se interpola entre estos dos extremos usando la señal de control según:

$$\alpha[t] = \begin{cases} \alpha_{\text{ataque}} & G[t] < \overline{G}[t-1] \\ \overline{A}[t]\alpha_{\text{liberación}} + \left(1 - \overline{A}[t]\right) & G[t] \geq \overline{G}[t-1] \end{cases}$$

(13)

Al interpolar el coeficiente de alisamiento de manera continua en función de la señal de control de evento, el tiempo de liberación se reinicia a un valor proporcional a la intensidad del evento al comienzo de un evento y entonces aumenta suavemente al infinito tras producirse un evento. La tasa de este aumento viene impuesta por el coeficiente α_{evento} usado para generar la señal de control de evento alisada.

Las figuras 9e y 10e muestran el efecto de alisar la ganancia con el coeficiente controlado por evento de la ecuación 13 en oposición al coeficiente no controlado por evento de la ecuación 7b. En el primer caso, la señal de control de evento cae a cero después del último acorde de piano, evitando así que la ganancia se mueva hacia arriba. Como resultado, el correspondiente audio modificado en la figura 9f no sufre una intensificación antinatural del decrecimiento del acorde. En el segundo caso, la señal de control de evento nunca se aproxima a cero, y por lo tanto la señal de ganancia alisada se inhibe muy poco mediante la aplicación del control de evento. La trayectoria de la ganancia alisada es casi idéntica a la ganancia no controlada por evento en la figura 10b. Éste es exactamente el efecto deseado.

AGC y DRC basados en la sonoridad

Como una alternativa a las técnicas de procesamiento dinámico tradicionales en las que las modificaciones de la señal son una función directa de mediciones de señal simples tales como potencia pico o RMS, la solicitud de patente internacional n.º de serie PCT/US2005/038579 da a conocer el uso del modelo de sonoridad basado en psicoacústica descrito anteriormente como marco dentro del cual se realiza el procesamiento dinámico. se citan algunas ventajas. En primer lugar, las mediciones y modificaciones se especifican en unidades de sonios, que es una medida más precisa de la percepción de la sonoridad que medidas más básicas tales como potencia pico o RMS. En segundo lugar, el audio puede modificarse de tal manera que el equilibrio espectral percibido del audio original se mantenga a medida que cambia la sonoridad global. De esta manera, cambios en la sonoridad global resultan menos aparentes desde el punto de vista de la percepción, en comparación con un procesador dinámico que utiliza una ganancia de banda ancha, por ejemplo, para modificar el audio. Por último, el modelo psicoacústico es inherentemente multibanda, y por lo tanto el sistema se configura fácilmente para realizar procesamiento dinámico multibanda con el fin de paliar los problemas de bombeo espectral cruzado ampliamente conocidos, asociados con un procesador dinámico de banda ancha.

Aunque realizar procesamiento dinámico en este dominio de sonoridad ya tiene varias ventajas frente al procesamiento dinámico tradicional, la técnica puede mejorarse adicionalmente a través del uso de eventos auditivos para controlar varios parámetros. Considérese el segmento de audio que contiene acordes de piano tal como se representa en 27a y el DRC asociado mostrado en las figuras 10b y c. Es posible realizar un DRC similar en el dominio de la sonoridad, y en

este caso, cuando la sonoridad del decrecimiento del acorde de piano final se intensifica, la intensificación será menos aparente porque el equilibrio espectral de la nota en decrecimiento se mantendrá mientras se aplica la intensificación. Sin embargo, una mejor solución es no intensificar el decrecimiento en absoluto, y por lo tanto ventajosamente es posible aplicar el mismo principio de controlar los tiempos de ataque y de liberación con eventos auditivos en el dominio de la sonoridad que se describió previamente para el DRC tradicional.

El sistema de procesamiento dinámico en el dominio de la sonoridad que se describe ahora consiste en AGC seguido de DRC. El objetivo de esta combinación es hacer que todo audio procesado tenga aproximadamente la misma sonoridad percibida al tiempo que sigue manteniendo al menos parte de la dinámica del audio original. La figura 11 representa un conjunto apropiado de curvas de AGC y DRC para esta aplicación. Obsérvese que la entrada y la salida de ambas curvas se representan en unidades de sonios ya que el procesamiento se realiza en el dominio de la sonoridad. La curva de AGC intenta llevar el audio de salida más cerca de algún nivel objetivo, y, como se mencionó anteriormente, lo hace con constantes de tiempo relativamente lentas. Puede pensarse que el AGC hace que la sonoridad a largo plazo del audio sea igual al objetivo, pero a corto plazo, la sonoridad puede fluctuar significantemente alrededor de este objetivo. Por lo tanto, puede emplearse DRC, que actúa más rápido, para limitar estas fluctuaciones hasta un punto que se considere aceptable para la aplicación particular. La figura 11 muestra una curva de DRC de este tipo en la que el objetivo del AGC entra dentro de la “banda nula” de DRC, la parte de la curva que no requiere modificación alguna. Con esta combinación de curvas, el AGC sitúa la sonoridad a largo plazo del audio dentro de la banda nula de la curva de DRC de modo que tienen que aplicarse mínimas modificaciones del DRC que actúa rápido. Si la sonoridad a corto plazo aún fluctúa fuera de la banda nula, el DRC entonces actúa para mover la sonoridad del audio hacia esta banda nula. Como observación general final, puede aplicarse el AGC que actúa lentamente de modo que todas las bandas del modelo de sonoridad reciban la misma cantidad de modificación de la sonoridad, manteniendo así el equilibrio espectral percibido, y puede aplicarse el DRC que actúa rápido de una manera que permita que la modificación de la sonoridad varíe a través de las bandas con el fin de paliar el bombeo espectral cruzado que, de lo contrario, podría obtenerse como resultado de una modificación de sonoridad independiente de la banda, que actúa rápido.

Pueden utilizarse eventos auditivos para controlar el ataque y la liberación tanto de AGC como de DRC. En el caso del AGC, tanto el tiempo de ataque como el de liberación son grandes en comparación con la resolución temporal de la percepción de eventos, y por lo tanto el control de eventos puede emplearse ventajosamente en ambos casos. Con el DRC, el ataque es relativamente corto, y por lo tanto el control de eventos puede necesitarse solamente para la liberación como con el DRC tradicional descrito anteriormente.

Como se mencionó anteriormente, puede usarse un espectro de sonoridad específica asociado con el modelo de sonoridad empleado con la finalidad de detección de eventos. Una señal de diferencia D[t], similar a la de las ecuaciones 10a y b, puede calcularse a partir de la sonoridad específica N[b,t], definida en la ecuación 2, como sigue:

$$D[t] = \sum_b \left| N_{NORM}[b,t] - N_{NORM}[b,t-1] \right|$$

(14a)

donde

$$N_{NORM}[b,t] = \frac{N[b,t]}{\max_b \{N[b,t]\}}$$

(14b)

Aquí el máximo de $\left|N[b,t]\right|$ a través de las bandas de frecuencia b se usa para la normalización, aunque podrían emplearse otros factores de normalización; por ejemplo, el promedio de $\left|N[b,t]\right|$ a través de las bandas de frecuencia. Si la diferencia *D[t]* excede un umbral D_{min}, entonces se considera que ha ocurrido un evento. La señal de diferencia puede procesarse entonces de la misma manera mostrada en las ecuaciones 11 y 12 para generar una señal de control de evento alisada $\overline{A}[t]$ usada para controlar los tiempos de ataque y de liberación.

La curva de AGC representada en la figura 11 puede representarse como una función que toma como su entrada una medida de sonoridad y genera una sonoridad de salida deseada:

$$L_o = F_{AGC}\{L_i\}$$

(15a)

La curva de DRC puede representarse de manera similar:

$$L_o = F_{DRC}\{L_i\}$$

(15b)

Para el AGC, la sonoridad de entrada es una medida de la sonoridad a largo plazo del audio. Puede calcularse tal medida alisando la sonoridad instantánea *L[t]*, definida en la ecuación 3, usando constantes de tiempo relativamente largo (en el orden de varios segundos). Se ha mostrado que al valorar la sonoridad a largo plazo de un segmento de audio, los seres humanos ponderan las partes más fuertes en mayor medida que las más suaves, y puede usarse un ataque más rápido que la liberación en el alisamiento para simular este efecto. Con la incorporación de control de evento tanto para el ataque como para la liberación, la sonoridad a largo plazo usada para determinar la modificación de AGC puede por lo tanto calcularse según:

$$L_{AGC}[t] = \alpha_{AGC}[t]L_{AGC}[t-1] + (1-\alpha_{AGC}[t])L[t]$$

(16a)

donde

$$\alpha_{AGC}[t] = \begin{cases} \overline{A}[t]\alpha_{AGC_{\text{ataque}}} + (1-\overline{A}[t]) & L[t] > L_{AGC}[t-1] \\ \overline{A}[t]\alpha_{AGC_{\text{liberación}}} + (1-\overline{A}[t]) & L[t] \le L_{AGC}[t-1] \end{cases}$$

(16b)

Además, puede calcularse un espectro de sonoridad específica a largo plazo asociado que después se usará para el DRC multibanda:

$$N_{AGC}[b,t] = \alpha_{AGC}[t]N_{AGC}[b,t-1] + (1-\alpha_{AGC}[t])N[b,t]$$

(16c)

En la práctica pueden escogerse los coeficientes de alisamiento de modo que el tiempo de ataque sea aproximadamente la mitad que el de liberación. Dada la medida de sonoridad a largo plazo, puede calcularse el ajuste a escala de la modificación de sonoridad asociada con el AGC como la razón de sonoridad de salida respecto a sonoridad de entrada.

$$S_{AGC}[t] = \frac{F_{AGC}\{L_{AGC}[t]\}}{L_{AGC}[t]}$$

(17)

La modificación de DRC puede ahora calcularse a partir de la sonoridad tras la aplicación del ajuste a escala de AGC. En lugar de alisar una medida de la sonoridad antes de la aplicación de la curva de DRC, puede alternativamente aplicarse la curva de DRC a la sonoridad instantánea y entonces posteriormente alisar la modificación resultante. Esto es similar a la técnica descrita anteriormente para el alisamiento de la ganancia del DRC tradicional. Además, el DRC puede aplicarse en un modo multibanda, lo que significa que la modificación de DRC es una función de la sonoridad específica *N[b,t]* en cada banda *b*, en lugar de la sonoridad global *L[t]*. Sin embargo, con el fin de mantener el equilibrio espectral promedio del audio original, puede aplicarse DRC a cada banda de manera que las modificaciones resultantes tengan el mismo efecto promedio que se obtendría como resultado al aplicar DRC a la sonoridad global. Esto puede lograrse ajustando a escala cada banda por la razón de sonoridad global a largo plazo (después de la aplicación del ajuste a escala de AGC) respecto a sonoridad específica a largo plazo, y usando este valor como el argumento para la función DRC. El resultado se ajusta entonces a escala de nuevo por la inversa de dicha razón para producir la sonoridad específica de salida. De esta manera, el ajuste a escala de DRC en cada banda puede calcularse según:

$$S_{DRC}[b,t] = \frac{\frac{N_{AGC}[b,t]}{S_{AGC}[t]L_{AGC}[t]} F_{DRC}\left\{\frac{S_{AGC}[t]L_{AGC}[t]}{N_{AGC}[t]} N[b,t]\right\}}{N[b,t]} \quad (18)$$

Las modificaciones de AGC y DRC pueden entonces combinarse para formar un ajuste a escala de sonoridad total por banda:

$$S_{TOT}[b,t] = S_{AGC}[t]S_{DRC}[b,t] \quad (19)$$

Este ajuste a escala total puede entonces alisarse a través del tiempo independientemente para cada banda con un ataque rápido y una liberación lenta y control de evento aplicado a la liberación solamente. Idealmente, el alisamiento se realiza en el logaritmo del ajuste a escala de manera análoga a las ganancias del DRC tradicional que se alisan en su representación en decibelios, aunque esto no es esencial. Para garantizar que el ajuste a escala total alisado se mueva en sincronización con la sonoridad específica en cada banda, los modos de ataque y de liberación pueden determinarse a través del alisamiento simultáneo de la propia sonoridad específica:

$$\overline{S}_{TOT}[b,t] = \exp\left(\alpha_{TOT}[b,t]\log\left(\overline{S}_{TOT}[b,t-1]\right) + (1-\alpha_{TOT}[b,t])\log\left(S_{TOT}[b,t]\right)\right) \quad (20a)$$

$$\overline{N}[b,t] = \alpha_{TOT}[b,t]\overline{N}[b,t-1] + (1-\alpha_{TOT}[b,t])N[b,t] \quad (20b)$$

donde

$$\alpha_{TOT}[b,t] = \begin{cases} \alpha_{TOT_{\text{ataque}}} & N[b,t] > \overline{N}[b,t-1] \\ \overline{A}[t]\alpha_{TOT_{\text{liberación}}} + \left(1-\overline{A}[t]\right) & N[b,t] \le \overline{N}[b,t-1] \end{cases} \quad (20c)$$

Finalmente puede calcularse una sonoridad específica del objetivo basándose en el ajuste a escala aplicado a la sonoridad específica original

$$\hat{N}[b,t] = \overline{S}_{TOT}[b,t]N[b,t] \quad (21)$$

y entonces resolver las ganancias *G[b,t]* que, cuando se aplican a la excitación original, dan como resultado una sonoridad específica igual al objetivo:

$$\hat{N}[b,t] = \Psi\left\{G^2[b,t]E[b,t]\right\} \quad (22)$$

Las ganancias pueden aplicarse a cada banda del banco de filtros usado para calcular la excitación, y el audio modificado puede generarse entonces invirtiendo el banco de filtros para producir una señal de audio en el dominio del tiempo modificada.

Control de parámetros adicionales

Aunque la explicación anterior se ha enfocado en el control de parámetros de ataque y liberación de AGC y DRC mediante análisis de la escena auditiva del audio que está procesándose, otros parámetros importantes puede también beneficiarse de ser controlados mediante los resultados del ASA. Por ejemplo, la señal de evento de control $\overline{A}[t]$ de la ecuación 12 puede usarse para variar el valor del parámetro de razón de DRC que se usa para ajustar dinámicamente

la ganancia del audio. El parámetro de razón, de manera similar a los parámetros de tiempo de ataque y de liberación, puede contribuir significativamente a los artefactos perceptivos introducidos por el ajuste de ganancia dinámico.

Implementación

La invención puede implementarse en hardware o software, o una combinación de ambos (por ejemplo, disposiciones lógicas programables). A menos que se especifique lo contrario, los algoritmos incluidos como parte de la invención no están relacionados de manera inherente a ningún ordenador u otro aparato particular. En particular, pueden usarse diversas máquinas de propósito general con programas escritos según las enseñanzas en el presente documento, o puede ser más conveniente construir aparatos más especializados (por ejemplo, circuitos integrados) para realizar las etapas de método requeridas. De esta manera, la invención puede implementarse en uno o más programas informáticos que se ejecuten en uno o más sistemas informáticos programables comprendiendo cada uno al menos un procesador, al menos un sistema de almacenamiento de datos (incluyendo memoria volátil y no volátil y/o elementos de almacenamiento), al menos un puerto o dispositivo de entrada, y al menos un puerto o dispositivo de salida. Se aplica código de programa a 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 este tipo puede implementarse en cualquier lenguaje informático deseado (incluyendo lenguajes máquina, ensambladores, o para procedimientos de alto nivel, lógicos, o de programación orientados a objetos) para la comunicación con un sistema informático. En cualquier caso, el lenguaje puede ser un lenguaje compilado o interpretado.

Cada programa informático de este tipo se almacena o descarga preferiblemente en un dispositivo o medio de almacenamiento (por ejemplo medios o memoria de estado sólido, o medios magnéticos u ópticos) legibles por un ordenador programable de propósito especial o general, para configurar y operar el ordenador cuando el sistema informático lee el dispositivo o medios de almacenamiento para realizar los procedimientos descritos en el presente documento. También puede considerarse que el sistema de la invención puede implementarse como un medio de almacenamiento legible por ordenador, configurado con un programa informático, en el que el medio de almacenamiento así configurado hace que un sistema informático funcione de una manera específica y predefinida para realizar las funciones descritas en el presente documento.

Se han descrito varias realizaciones de la invención. Sin embargo, se entenderá que pueden realizarse diversas modificaciones sin apartarse del alcance de la invención. Por ejemplo, algunas de las etapas descritas en el presente documento pueden ser independientes del orden, y por tanto pueden realizarse en un orden diferente del descrito.

Debe entenderse que la implementación de otras variaciones y modificaciones de la invención y sus diversos aspectos resultará evidente para los expertos en la técnica, y que la invención no se limita a estas realizaciones específicas descritas. Por lo tanto se contempla que la presente invención cubra todas y cada una de las modificaciones, variaciones, o equivalentes que entren dentro del alcance de los principios subyacentes básicos dados a conocer y reivindicadas en el presente documento.

Las siguientes patentes, solicitudes de patente y publicaciones dan a conocer documentos de la técnica anterior adicionales:

Procesamiento dinámico de audio

Audio Engineer's Reference Book, editado por Michael Talbot-Smith, 2ª edición. Limiters and Compressors, Alan Tutton, 2-1493-165. Focal Press, Reed Educational and Professional Publishing, Ltd., 1999.

Detectar y usar eventos auditivos

Solicitud de patente estadounidense n.º de serie 10/474.387, "High Quality Time-Scaling and Pitch-Scaling of Audio Signals" de Brett Graham, Crockett, publicada el 24 de junio de 2004 como US 2004/0122662 A1.

Solicitud de patente estadounidense n.º de serie 10/478.398, "Method for Time Aligning Audio Signals Using Characterizations Based on Auditory Events" de Brett G. Crockett et al, publicada el 29 de julio de 2004 como US 2004/0148159 A1.

Solicitud de patente estadounidense n.º de serie 10/478.538, "Segmenting Audio Signals Into Auditory Events" de Brett G. Crockett, publicada el 26 de agosto de 2004 como US 2004/0165730 A1. Aspectos de la presente invención proporcionan una manera para detectar eventos auditivos además de los dados a conocer en dicha solicitud de Crockett.

Solicitud de patente estadounidense n.º de serie 10/478.397, "Comparing Audio Using Characterizations Based on Auditory Events" de Brett G. Crockett et al, publicada el 2 de septiembre de 2004 como US 2004/0172240 A1.

Solicitud internacional según el Tratado de Cooperación en materia de Patentes n.º de serie PCT/US 05/24630 presentada el 13 de julio de 2005, titulada “Method for Combining Audio Signals Using Auditory Scene Analysis”, de Michael John Smithers, publicada el 9 de marzo de 2006 como documento WO 2006/026161.

Solicitud internacional según el Tratado de Cooperación en materia de Patentes n.º de serie PCT/US 2004/016964 presentada el 27 de mayo de 2004, titulada “Method, Apparatus and Computer Program for Calculating and Adjusting the Perceived Loudness of an Audio Signal”, de Alan Jeffrey Seefeldt et al, publicada el 23 de diciembre de 2004 como WO 2004/1994 A2.

Solicitud internacional según el Tratado de Cooperación en materia de Patentes n.º de serie PCT/US 2005/038579 presentada el 25 de octubre de 2005, titulada “Calculating and Adjusting the Perceived Loudness and/or the Perceived Spectral Balance of an Audio Signal”, de Alan Jeffrey Seefeldt, y publicada como publicación internacional número WO 2006/047600.

“A Method for Characterizing and Identifying Audio Based on Auditory Scene Analysis”, por Brett Crockett y Michael Smithers, artículo 6416 de la convención de la Audio Engineering Society, 118ª Convención, Barcelona, 28-31 de mayo de 2005.

“High Quality Multichannel Time Scaling and Pitch-Shifting using Auditory Scene Analysis”, por Brett Crockett, artículo 5948 de la convención de la Audio Engineering Society, Nueva York, octubre de 2003.

“A New Objective Measure of Perceived Loudness”, por Alan Seefeldt et al, artículo 6236 de la convención de la Audio Engineering Society, San Francisco, 28 de octubre de 2004.

Handbook for Sound Engineers, The New Audio Cyclopedia, editado por Glen M- Ballou, 2ª edición. Dynamics, 850-851. Focal Press en impresión de Butterworth-Heinemann, 1998.

Audio Engineer’s Reference Book, editado por Michael Talbot-Smith, 2ª edición, sección 2.9 (“Limiters and Compressors” por Alan Tutton), páginas 2.149-2.165, Focal Press, Reed Educational and Professional Publishing, Ltd., 1999.

REIVINDICACIONES

1. Método para modificar un parámetro de procesamiento dinámico de audio, que comprende:

detectar cambios en las características espectrales con respecto al tiempo en una señal de audio,

identificar, como límites de evento auditivo, cambios mayores que un umbral en características espectrales con respecto al tiempo en dicha señal de audio, en el que un segmento de audio entre límites consecutivos constituye un evento auditivo,

generar una señal de control de modificación de parámetros basándose en dichos límites de evento identificados, y

modificar los parámetros de procesamiento dinámico de audio en función de la señal de control.

2. Método según la reivindicación 1, en el que el parámetro es uno de tiempo de ataque, tiempo de liberación, y razón.

3. Método según la reivindicación 1, en el que el parámetro modificado es una constante de tiempo con alisamiento de ganancia.

4. Método según la reivindicación 3, en el que la constante de tiempo con alisamiento de ganancia es una constante de tiempo de ataque con alisamiento de ganancia.

5. Método según la reivindicación 3, en el que la constante de tiempo con alisamiento de ganancia es una constante de tiempo de liberación con alisamiento de ganancia.

6. Método según una cualquiera de las reivindicaciones 1 a 5, en el que dicha señal de control de modificación de parámetros se basa en la ubicación de dichos límites de evento auditivo identificados y el grado de cambio en características espectrales asociadas con cada uno de dichos límites de evento auditivo.

7. Método según la reivindicación 6, en el que generar una señal de control de modificación de parámetros comprende:

proporcionar un impulso en cada uno de los límites de evento auditivo, teniendo cada impulso de este tipo una amplitud proporcional al grado de dichos cambios en las características espectrales, y

alisar en el tiempo cada impulso de modo que su amplitud decrezca suavemente hasta cero, obteniendo de esta manera la señal de control de modificación de parámetros.

8. Método según una cualquiera de las reivindicaciones 1 a 7, en el que los cambios en las características espectrales con respecto al tiempo se detectan mediante la comparación de diferencias en la sonoridad específica.

9. Método según la reivindicación 8, en el que dicha señal de audio se representa por una secuencia de tiempo discreta x[n] que se ha muestreado a partir de una fuente de audio a una frecuencia de muestreo f_s y los cambios en las características espectrales con respecto al tiempo se calculan mediante la comparación de la diferencia en la sonoridad específica $N[b,t]$ a través de las bandas de frecuencia b entre bloques de tiempo t sucesivos.

10. Método según la reivindicación 9, en el que la diferencia en el contenido espectral entre bloques de tiempo sucesivos de la señal de audio se calcula según

$$D[t] = \sum_b \left| N_{NORM}[b,t] - N_{NORM}[b,t-1] \right|$$

donde

$$N_{NORM}[b,t] = \frac{N[b,t]}{\max_b \{N[b,t]\}}.$$

11. Método según la reivindicación 9, en el que la diferencia en el contenido espectral entre bloques de tiempo sucesivos de la señal de audio se calcula según

$$D[t] = \sum_b \left| N_{NORM}[b,t] - N_{NORM}[b,t-1] \right|$$

donde

$$N_{NORM}[b,t] = \frac{N[b,t]}{\underset{b}{avg}\{N[b,t]\}}.$$

12. Aparato que comprende medios adaptados para realizar el método según una cualquiera de las reivindicaciones 1 a 11.

13. Programa informático, almacenado en un medio legible por ordenador, para hacer que un ordenador realice el método según una cualquiera de las reivindicaciones 1 a 11.

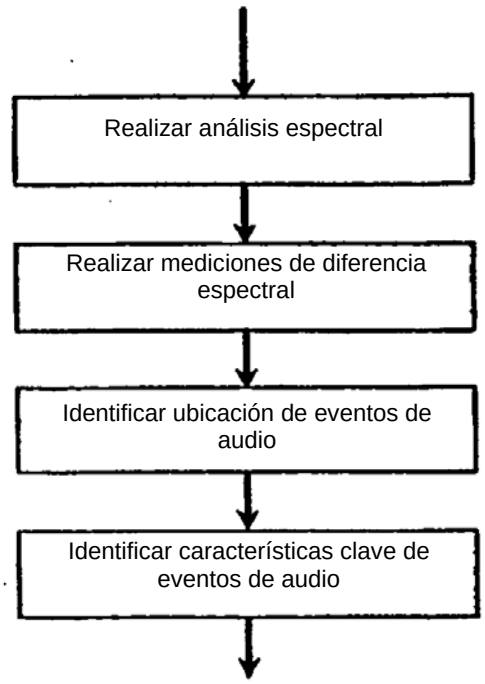

Realizar análisis espectral
Realizar mediciones de diferencia espectral
Identificar ubicación de eventos de audio
Identificar características clave de eventos de audio

FIG. 1

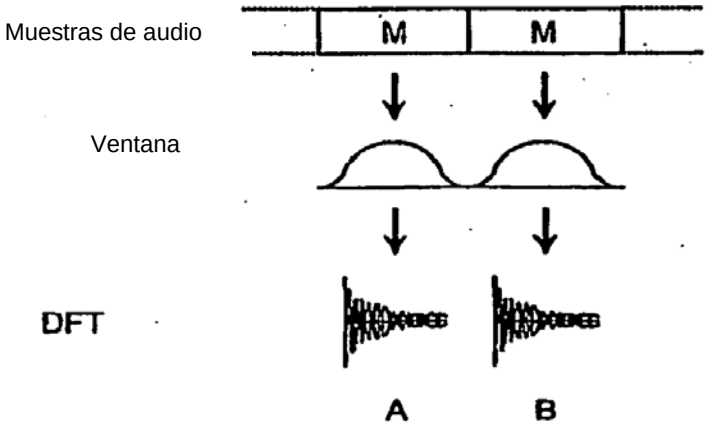

Muestras de audio
M
M
Ventana
DFT
A
B

FIG. 2

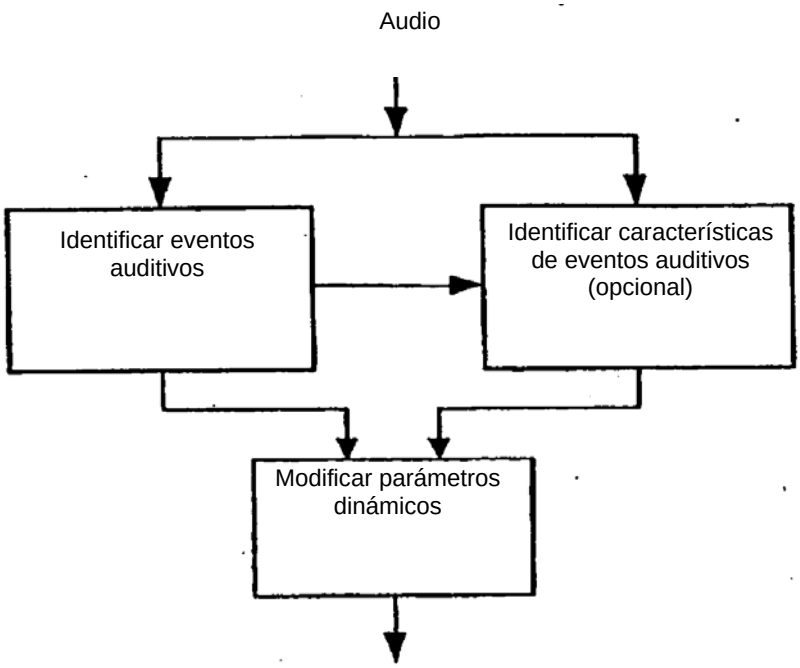

Audio
Identificar eventos auditivos
Identificar características de eventos auditivos (opcional)
Modificar parámetros dinámicos

FIG. 3

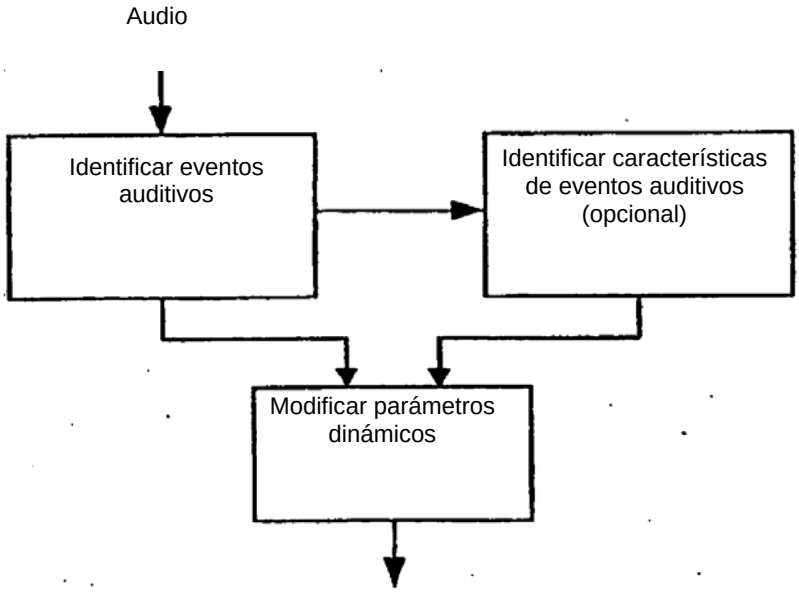

Audio
Identificar eventos auditivos
Identificar características de eventos auditivos (opcional)
Modificar parámetros dinámicos

FIG. 4

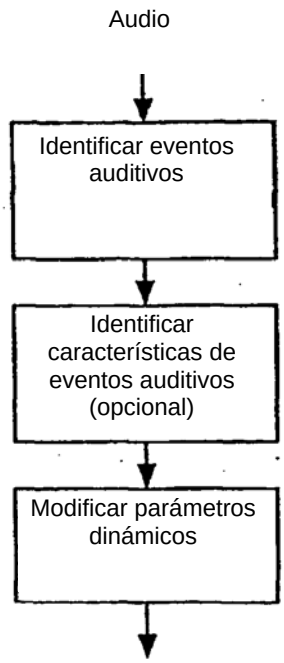

Audio
Identificar eventos auditivos
Identificar características de eventos auditivos (opcional)
Modificar parámetros dinámicos

FIG. 5

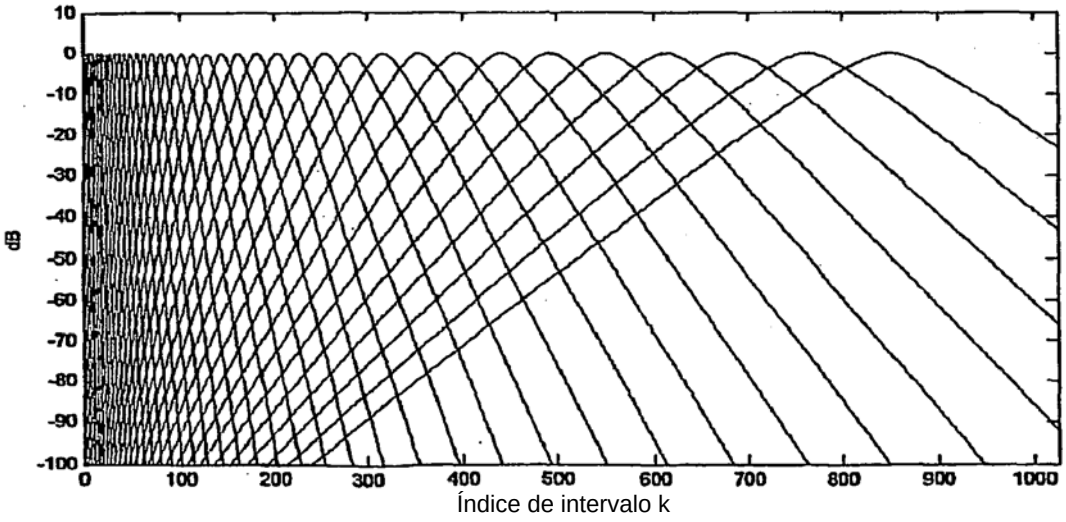

dB
10
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
-100
0
100
200
300
400
500
600
700
800
900
1000
Índice de intervalo k

FIG. 6

Anexo A de la norma ISO226 “Contornos de igual sonoridad”

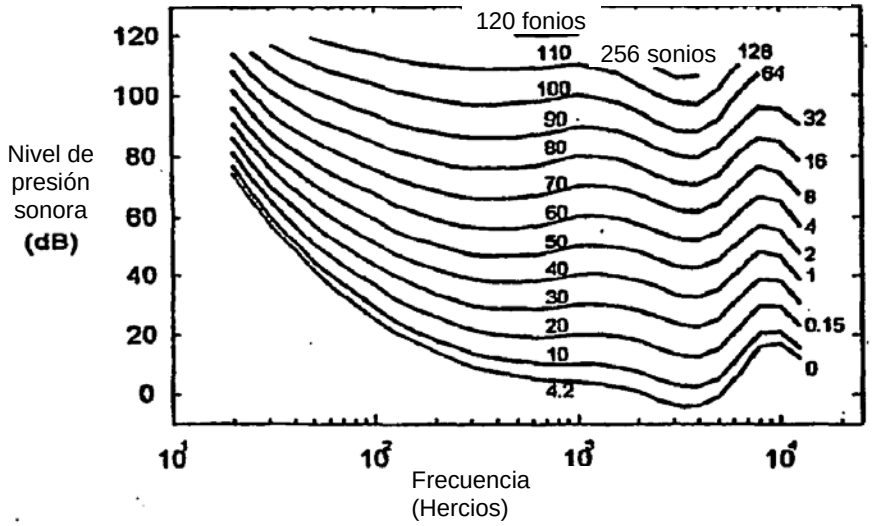


FIG. 7

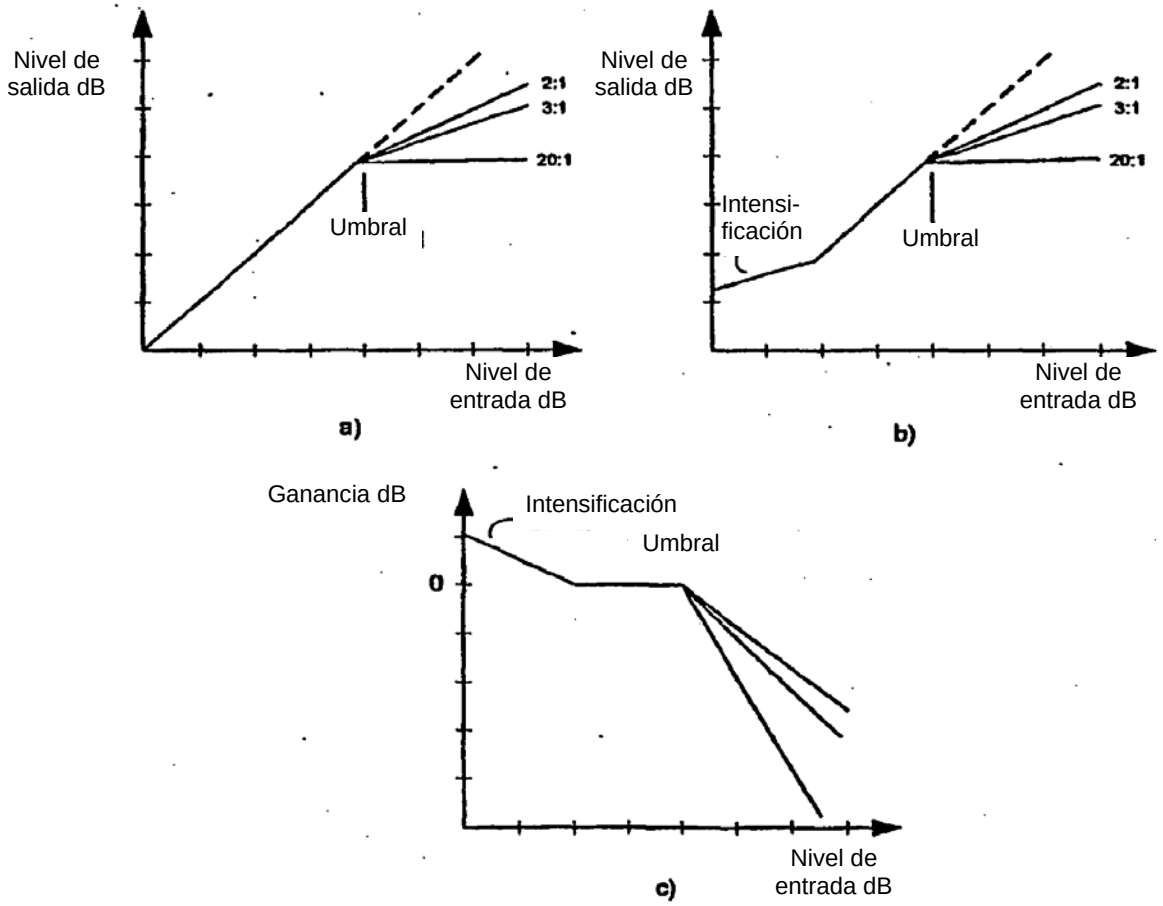
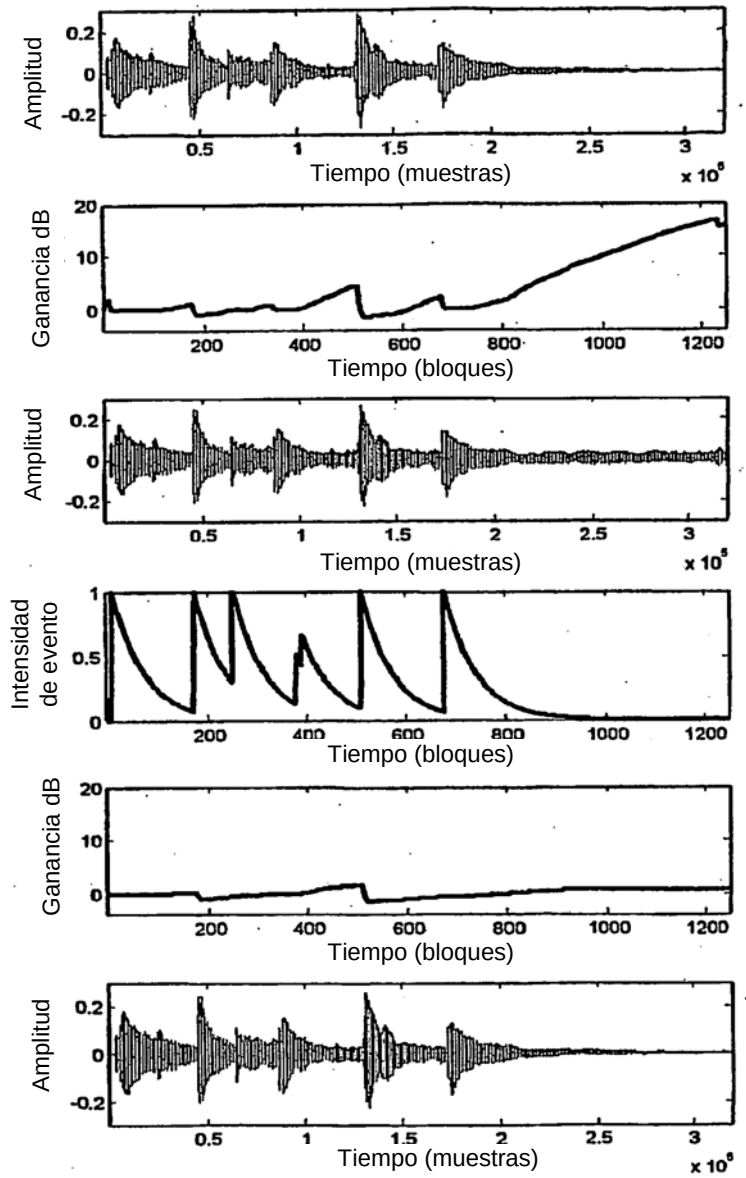


FIG. 8



a) Señal original

b) Ganancia de DRC

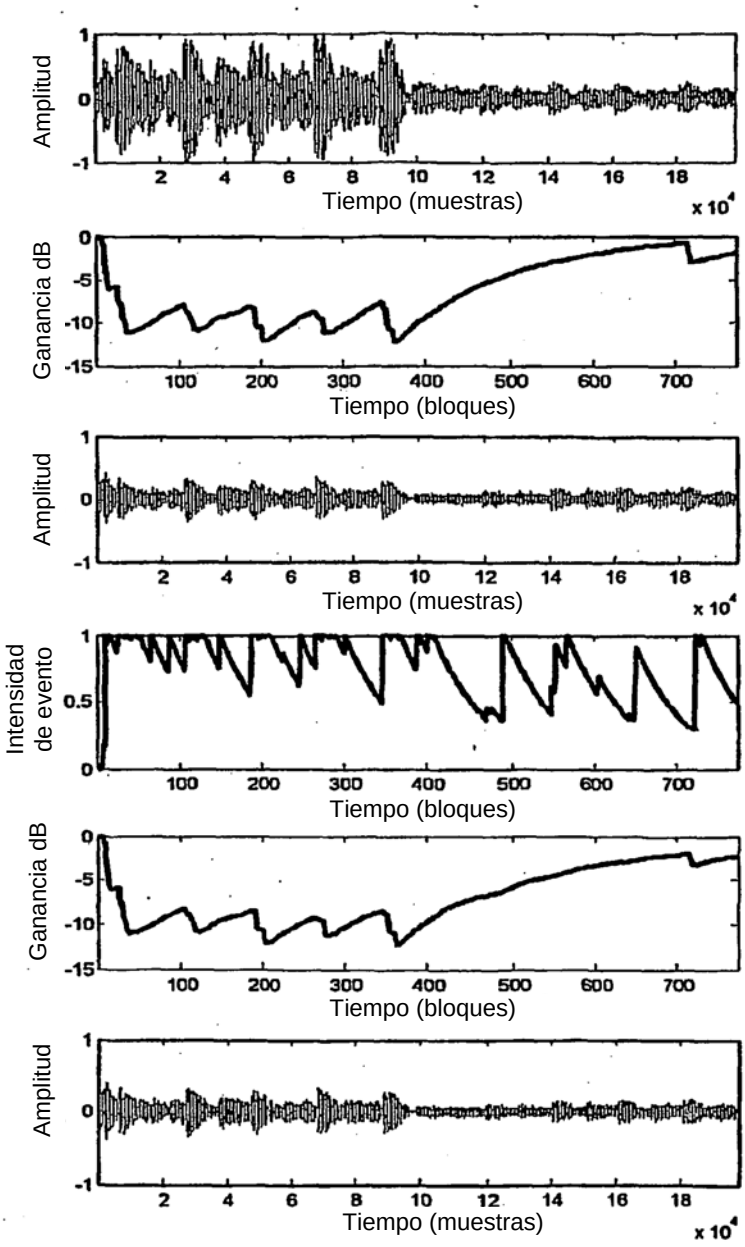
c) Señal modificada con la ganancia de DRC

d) Señal de control de evento

e) Ganancia de DRC controlada por evento

f) Señal modificada con ganancia de DRC controlada por evento

FIG. 9



a) Señal original

b) Ganancia de DRC

c) Señal modificada con la ganancia de DRC

d) Señal de control de evento

e) Ganancia de DRC controlada por evento

f) Señal modificada con ganancia de DRC controlada por evento

FIG. 10

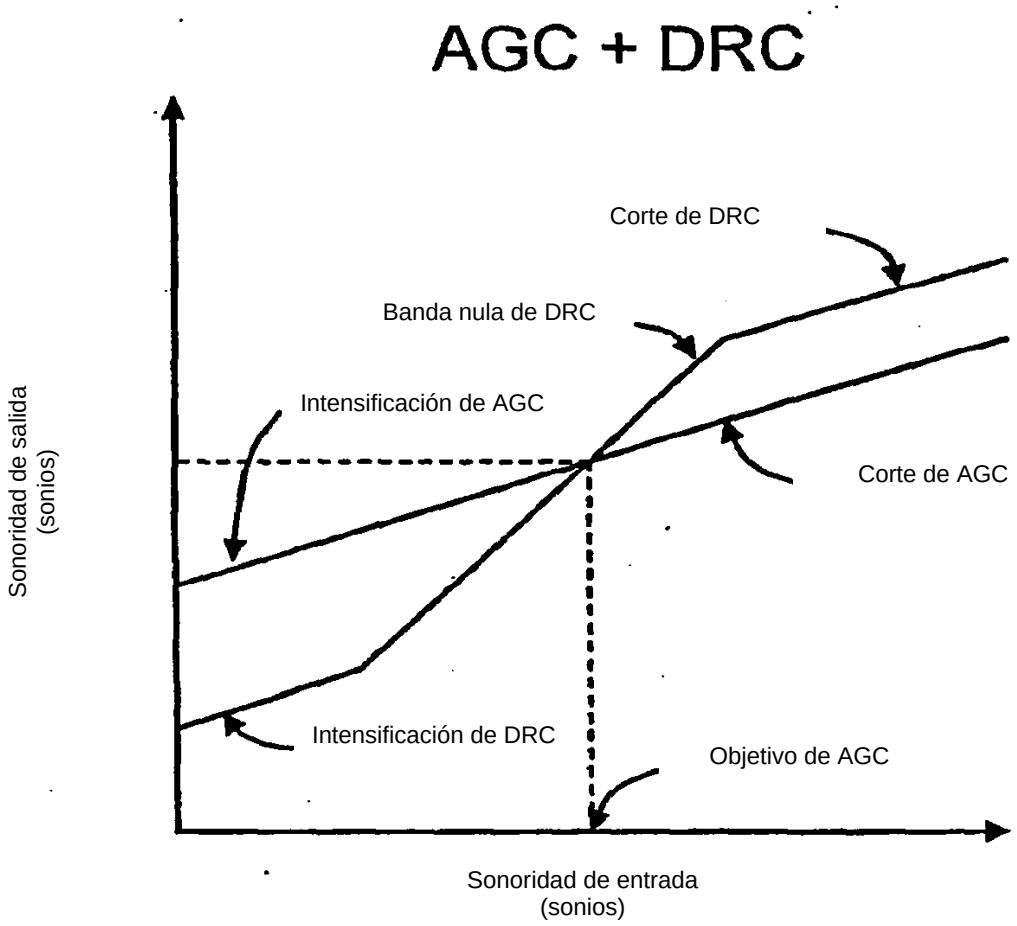

AGC + DRC
Corte de DRC
Banda nula de DRC
Intensificación de AGC
Corte de AGC
Sonoridad de salida (sonios)
Intensificación de DRC
Objetivo de AGC
Sonoridad de entrada (sonios)

FIG. 11