



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 377 719**

51 Int. Cl.:
G10L 21/02 (2006.01)
H03G 3/20 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **08780174 .2**
96 Fecha de presentación : **11.07.2008**
97 Número de publicación de la solicitud: **2168122**
97 Fecha de publicación de la solicitud: **31.03.2010**

54 Título: **Procesamiento de audio utilizando un análisis de escenas auditivas y oblicuidad espectral.**

30 Prioridad: **13.07.2007 US 959463 P**

45 Fecha de publicación de la mención BOPI:
30.03.2012

45 Fecha de la publicación del folleto de la patente:
30.03.2012

73 Titular/es:
Dolby Laboratories Licensing Corporation
100 Potrero Avenue
San Francisco, California 94103-4813, US

72 Inventor/es: **Seefeldt, Alan, Jeffrey y**
Smithers, Michael, John

74 Agente/Representante:
Torner Lasalle, Elisabet

ES 2 377 719 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

ES 2 377 719 T3

DESCRIPCIÓN

Procesamiento de audio utilizando un análisis de escenas auditivas y oblicuidad espectral.

5 **Campo técnico**

La presente invención versa, en general, acerca del procesamiento de audio y, en particular, acerca del análisis de escenas auditivas y de la oblicuidad espectral.

10 **Referencias e Incorporación por referencia**

Los siguientes documentos son incorporados al presente documento por referencia en su integridad:

15 Crockett y Seefeldt, Solicitud Internacional bajo el Tratado de Cooperación de Patentes, con número de serie PCT/US2007/008313, titulada "Controlling Dynamic Gain Parameters of Audio using Auditory Scene Analysis and Specific-Loudness-Based Detection of Auditory Events", que nombra a Brett Graham Crockett y Alan Jeffrey Seefeldt como inventores, presentada el 30 de marzo de 2007, con Expediente de Agente DOL186 PCT, y publicada el 8 de noviembre de 2007 como WO 2007/127023;

20 Seefeldt y otros, Solicitud Internacional bajo el Tratado de Cooperación de Patentes, con número de serie PCT/US 2004/016964, titulada "Method, Apparatus and Computer Program for Calculating and Adjusting the Perceived Loudness of an Audio Signal", que nombra a Alan Jeffrey Seefeldt y otros como inventores, presentada el 27 de mayo de 2004, con Expediente de Agente n° DOL119 PCT, y publicada el 23 de diciembre de 2004 como WO 2004/111994 A2;

25 Seefeldt, Solicitud Internacional bajo el Tratado de Cooperación de Patentes, con número de serie PCT/US2005/038579, titulada "Calculating and Adjusting the Perceived Loudness and/or the Perceived Spectral Balance of an Audio Signal", que nombra a Alan Jeffrey Seefeldt como inventor, presentada el 25 de octubre de 2005, con Expediente de Agente n° DOL15202 PCT, y publicada el 4 de mayo de 2006 como WO 2006/047600;

30 Crockett, Solicitud de Patente Estadounidense con número de serie 10/474,387, titulada "High Quality Time-Scaling and Pitch-Scaling of Audio Signals", que nombra a Brett Graham Crockett como inventor, presentada el 10 de octubre de 2003, con Expediente de Agente n° DOL07503, y publicada el 24 de junio de 2004 como US 2004/0122662 A1;

35 Crockett y otros, Solicitud de Patente Estadounidense con número de serie 10/478,398, titulada "Method for Time Aligning Audio Signals Using Characterizations Based on Auditory Events", que nombra a Brett G. Crockett y otros como inventores, presentada el 20 de noviembre de 2003, con Expediente de Agente n° DOL09201, y publicada el 29 de julio de 2004 como US 2004/0148159 A1;

40 Crockett, Solicitud de Patente Estadounidense con número de serie 10/478,538, titulada "Segmenting Audio Signals Into Auditory Events", que nombra a Brett G. Crockett como inventor, presentada el 20 de noviembre de 2003, con Expediente de Agente n° DOL098, y publicada el 26 de agosto de 2004 como US 2004/0165730 A1;

45 Crockett y otros, Solicitud de Patente Estadounidense con número de serie 10/478,397, titulada "Comparing Audio Using Characterizations Based on Auditory Events", que nombra a Brett G. Crockett y otros como inventores, presentada el 20 de noviembre de 2003, con Expediente de Agente n° DOL092, y publicada el 2 de septiembre de 2004 como US 2004/0172240 A1;

50 Smithers, Solicitud Internacional bajo el Tratado de Cooperación de Patentes, con número de serie PCT/US 05/24630, titulada "Method for Combining Audio Signals Using Auditory Scene Analysis", que nombra a Michael John Smithers como inventor, presentada el 13 de julio de 2005, con Expediente de Agente n° DOL148 PCT, y publicada el 9 de marzo de 2006 como WO 2006/026161;

55 Crockett, B. y Smithers, M., "A Method for Characterizing and Identifying Audio Based on Auditory Scene Analysis", Audio Engineering Society Convention Paper 6416, 118th Convention, Barcelona, 28-31 de mayo de 2005;

60 Crockett, B., "High Quality Multichannel Time Scaling and Pitch-Shifting using Auditory Scene Analysis", Audio Engineering Society Convention Paper 5948, Nueva York, octubre de 2003; y

Seefeldt y otros, "A New Objective Measure of Perceived Loudness", Audio Engineering Society Convention Paper 6236, San Francisco, 28 de octubre de 2004.

65

Técnica antecedente*Eventos auditivos y detección de eventos auditivos*

5 La división de sonidos en unidades o segmentos percibidos de forma separada y diferenciada es denominada a veces como “análisis de eventos auditivos” o “análisis de escenas auditivas” (“ASA”). Los segmentos son denominados a veces “eventos auditivos” o “eventos de audio”. Albert S. Bregman, “Auditory Scene Analysis-The Perceptual Organizaron of Sound” (Massachusetts Institute of Technology, 1991, cuarta impresión, 2001, segunda edición en rústica de la MIT Press) presenta exhaustivamente el análisis de escenas auditivas. Además, Bhadkamkar y otros, patente estadounidense n° 6.002.776 (14 de diciembre de 1999) cita publicaciones que se remontan a 1976 como “investigaciones de la técnica anterior relativas a la separación de sonidos por medio del análisis de escenas auditivas”. Sin embargo, Bhadkamkar y otros desalientan el uso práctico del análisis de escenas auditivas, llegando a la conclusión de que las “[t]écnicas que implican el análisis de escenas auditivas, aunque interesantes desde el punto de vista científico como modelos del procesamiento auditivo humano, son en la actualidad demasiado exigentes y especializadas desde el punto de vista del cálculo como para ser consideradas técnicas prácticas para la separación de sonidos hasta que se logre un progreso fundamental”.

Crockett y Crocket y otros, en las diversas solicitudes de patente y en las monografías enumeradas más arriba, identifican los eventos auditivos. Esos documentos enseñan la división de una señal de audio en eventos auditivos (cada uno de los cuales tiende a ser percibido como separado y diferenciado) detectando cambios en la composición espectral (amplitud como función de la frecuencia) con respecto al tiempo. Esto puede realizarse, por ejemplo, calculando el contenido espectral de sucesivos bloques temporales de la señal de audio, comparando el contenido espectral entre bloques temporales sucesivos e identificando un límite de los eventos auditivos como el límite entre bloques en los que la diferencia en el contenido espectral supera un umbral. Alternativamente, pueden calcularse los cambios en la amplitud con respecto al tiempo en lugar de o en adición a los cambios en la composición espectral con respecto al tiempo.

Los marcadores de los límites de eventos auditivos están dispuestos a menudo en una señal de control temporal por lo que el intervalo, típicamente de cero a uno, indica la intensidad del límite del evento. Además, esta señal de control es a menudo filtrada, de forma que permanece la intensidad del límite del evento, y los intervalos temporales entre los límites de los eventos se calculan como valores en decaimiento del límite de evento precedente. Esta intensidad filtrada del evento auditivo es usada entonces por otros procedimientos de procesamiento de audio, incluyendo el control automático de ganancia y el control dinámico de gama.

35 *Procesamiento de audio por dinámica*

Las técnicas de control automático de ganancia (AGC) y de control dinámico de gama (DRC) son bien conocidas y comunes en muchos recorridos de señales de audio. En un sentido abstracto, ambas técnicas miden el nivel de una señal de audio y luego modifican la ganancia de la señal en una cantidad que es función del nivel medido. En un sistema lineal 1:1 de procesamiento de dinámica, el audio de entrada no es procesado e, idealmente, la señal de audio de salida coincide con la señal de audio de entrada. Además, imaginemos un sistema de procesamiento de audio por dinámica que mida automáticamente la señal de entrada y controle con esa medida la señal de salida. Si la señal de entrada aumenta su nivel en 6 dB y la señal procesada de salida aumenta su nivel en solo 3 dB, entonces la señal de salida ha sido comprimida en una proporción de 2:1 con respecto a la señal de entrada.

En Crockett y Seefeldt, el análisis de escenas auditivas mejora el rendimiento de los procedimientos de AGC y DRC minimizando el cambio en ganancia entre los límites de los eventos auditivos y confinando gran parte del cambio en ganancia a las inmediaciones de un límite de eventos. Esto lo realiza modificando el comportamiento de liberación del procesamiento de dinámica. Así, los eventos auditivos suenan coherentes y naturales.

Las notas tocadas en un piano son un ejemplo. Con los procedimientos convencionales de AGC o DRC, la ganancia aplicada a la señal de audio aumenta durante la cola de cada nota, haciendo que cada nota sea amplificada de forma poco natural. Con el análisis de escenas auditivas, la ganancia del AGC o el DRC, la ganancia aplicada a la señal de audio se mantiene constante dentro de cada nota y cambia únicamente cerca del inicio de cada nota cuando se detecta un límite de eventos auditivos. La señal resultante de audio ajustada en ganancia suena natural, al irse apagando la cola de cada nota.

Las implementaciones típicas del análisis de escenas auditivas (como en las referencias en lo que antecede) son deliberadamente invariables en nivel. Es decir, detectan límites de eventos auditivos con independencia del nivel absoluto de la señal. Aunque la invariabilidad del nivel es útil en muchas aplicaciones, algunos análisis de escenas auditivas se benefician de alguna dependencia del nivel.

Un caso tal es el procedimiento descrito en Crockett y Seefeldt. En él, el control de ASA del AGC y el DRC impide cambios de ganancia grandes entre los límites de eventos auditivos. Sin embargo, los cambios de ganancia más a la larga pueden seguir siendo poco deseables en algunos tipos de señales de audio. Cuando una señal de audio pasa de una sección de más volumen a una de menos volumen, la ganancia del AGC o el DRC, limitada a cambiar solo cerca de los límites de eventos, puede permitir que el nivel de la señal de audio de procesamiento aumente de forma poco

deseable y poco natural durante la sección de poco volumen. Esta situación ocurre frecuentemente en películas, en las que el diálogo esporádico alterna con sonidos de fondo de poco volumen. Dado que la señal de audio de poco volumen de fondo también contiene eventos auditivos, la ganancia del AGC o el DRC cambia cerca de esos límites de eventos y aumenta el nivel global de la señal de audio.

Es poco deseable realizar una ponderación simple de la importancia de los eventos auditivos mediante una medida del nivel, de la potencia o de la sonoridad de la señal de audio. En muchas situaciones, no se conoce la relación entre la medida de la señal y el nivel absoluto de reproducción. Idealmente, sería útil una medida que discriminase o detectase las señales de audio perceptivamente de menos volumen con independencia del nivel absoluto de la señal de audio.

Aquí, "perceptivamente de menos volumen" se refiere no a menos volumen en una medida objetiva de sonoridad (como en Seefeldt y otros y Seefeldt), sino más bien a menos volumen en base a la sonoridad esperada del contenido. Por ejemplo, la experiencia humana indica que un susurro es un sonido de poco volumen. Si un sistema de procesamiento de dinámica mide que este es de poco volumen y, en consecuencia, aumenta la ganancia del AGC para lograr alguna sonoridad o nivel de salida nominal, el susurro resultante ajustado en ganancia tendría más volumen de lo que la experiencia dice que debería tener.

Revelación de la invención

En el presente documento se enseñan procedimientos y un aparato, según se reivindica en las reivindicaciones independientes, para controlar la sonoridad de eventos auditivos en una señal de audio. en una realización, el procedimiento incluye ponderar los eventos auditivos (teniendo un evento auditivo un espectro y una sonoridad), usar la oblicuidad en los espectros y controlar la sonoridad de los eventos auditivos usando los coeficientes de ponderación. Diversas realizaciones de la invención son como sigue: la ponderación es proporcional a la medida de oblicuidad en los espectros; la medida de oblicuidad es una medida de oblicuidad estabilizada; la ponderación es insensible a la amplitud de la señal de audio; la ponderación es insensible a la potencia; la ponderación es insensible a la sonoridad; en el momento de la ponderación no se conoce ninguna relación entre la medida de la señal y el nivel absoluto de reproducción; la ponderación incluye la ponderación de la importancia de los límites de eventos auditivos usando la oblicuidad en los espectros y reduciendo la amplificación del nivel de procesamiento del AGC o el DRC durante los segmentos perceptivamente de menos volumen de la señal de audio en comparación con procedimientos que no llevan a cabo la ponderación reivindicada.

En otras realizaciones, la invención es una memoria legible por ordenador que contiene un programa de ordenador para llevar a cabo uno cualquiera de los procedimientos anteriores.

En otras realizaciones adicionales, la invención es un sistema de ordenador que incluye una CPU, una de las memorias mencionadas anteriormente y un bus que acopla de forma comunicativa la CPU y la memoria.

En otra realización adicional, la invención es un procesador de señales de audio que incluye un calculador de oblicuidad espectral para calcular la oblicuidad espectral en una señal de audio, un identificador de eventos auditivos para identificar y ponderar eventos auditivos en la señal de audio usando la oblicuidad espectral calculada, un modificador de parámetros para modificar parámetros para controlar la sonoridad de eventos auditivos en la señal de audio y un controlador para controlar la sonoridad de eventos auditivos en la señal de audio.

En otra realización adicional, la invención es un procedimiento para controlar la sonoridad de eventos auditivos en una señal de audio, incluyendo el cálculo de medidas de oblicuidad de espectros de eventos auditivos sucesivos de una señal de audio, la generación de coeficientes de ponderación para los eventos auditivos en base a las medidas de oblicuidad, la derivación de una señal de control a partir de los coeficientes de ponderación y el control de la sonoridad de los eventos auditivos usando la señal de control.

Las diversas características de la presente invención y sus realizaciones preferentes pueden entenderse mejor por referencia a la siguiente exposición y a los dibujos adjuntos, en los que números de referencia similares se refieren a elementos similares.

Descripción de los dibujos

La Fig. 1 ilustra un dispositivo para llevar a cabo dos procedimientos de Crockett y Seefeldt de análisis de escenas auditivas y de control de parámetros de ganancia de dinámica.

La Fig. 2 ilustra un procesador de audio para identificar eventos auditivos y calcular la oblicuidad para modificar los eventos auditivos, los cuales, a su vez, modifican los parámetros de procesamiento de dinámica según una realización de la invención.

La Fig. 3 es una serie de gráficos que ilustra el uso de eventos auditivos para controlar el tiempo de liberación en una implementación digital de un controlador dinámico de gama (DRC) según una realización de la invención.

La Fig. 4 es una respuesta característica idealizada de un filtro lineal adecuado como filtro de transmisión según una realización de la invención.

La Fig. 5 muestra un conjunto de respuestas características idealizadas de un filtro auditivo que se aproximan a un bandaje crítico en la escala de ERB.

Mejor modo de llevar a cabo la invención

La Fig. 1 ilustra un dispositivo 1 para analizar escenas auditivas y controlar los parámetros de ganancia de dinámica según Crockett y Seefeldt. El dispositivo incluye un identificador 10 de eventos auditivos, un identificador opcional 11 de características de eventos auditivos y un modificador 12 de parámetros de dinámica. El identificador 10 de eventos auditivos recibe audio como entrada y produce una entrada para el modificador 12 de parámetros de dinámica (y una entrada para el identificador 11 de características de eventos auditivos, si está presente). El modificador 12 de parámetros de dinámica recibe la salida del identificador 10 de eventos auditivos (y del identificador 11 de características de eventos auditivos, si está presente) y produce una salida.

El identificador 10 de eventos auditivos analiza el espectro y, a partir de los resultados, identifica la ubicación de eventos de audio perceptibles que son para controlar los parámetros de ganancia de dinámica. Alternativamente, el identificador 10 de eventos auditivos transforma el audio en un dominio de sonoridad perceptiva (que puede proporcionar más información psicoacústicamente relevante que el primer procedimiento) y el dominio de sonoridad perceptiva identifica la ubicación de eventos auditivos que son para controlar los parámetros de ganancia de dinámica. (En esta alternativa, el procesamiento de audio está al tanto de los niveles absolutos de reproducción acústica).

El modificador 12 de parámetros de dinámica modifica los parámetros de dinámica en base a la salida del identificador 10 de eventos auditivos (y del identificador 11 de características de eventos auditivos, si está presente). En ambas alternativas, se segmenta en bloques una señal $x[n]$ de audio digital y, para cada bloque t , $D[t]$ representa la diferencia espectral entre el bloque actual y el bloque anterior.

Para la primera alternativa, $D[t]$ es la suma, en todos los coeficientes espectrales, de la magnitud de la diferencia entre los coeficientes espectrales logarítmicos normalizados (en dB) para el bloque actual t y el bloque anterior $t-1$. En esta alternativa, $D[t]$ es proporcional a las diferencias absolutas en los espectros (en dB ellas mismas). Para la segunda alternativa, $D[t]$ es, en todos los coeficientes de sonoridad específica, de la magnitud de la diferencia entre los coeficientes sonoridad específica normalizados para el bloque actual t y el bloque anterior $t-1$. En esta alternativa, $D[t]$ es proporcional a las diferencias absolutas en sonoridad específica (en sonios).

En ambas alternativas, si $D[t]$ supera un umbral D_{\min} , se considera que ha ocurrido un evento. El evento puede tener una intensidad entre cero y uno, en base a la proporción de $D[t]$ menos D_{\min} con respecto a la diferencia entre D_{\max} y D_{\min} . La intensidad $A[t]$ puede calcularse como:

$$A[t] = \begin{cases} 0 & D[t] \leq D_{\min} \\ \frac{D[t] - D_{\min}}{D_{\max} - D_{\min}} & D_{\min} < D[t] < D_{\max} \\ 1 & D[t] \geq D_{\max} \end{cases} \quad (1)$$

Los límites máximo y mínimo son diferentes para cada alternativa debido a sus diferentes unidades. Sin embargo, el resultado de ambos es una intensidad de evento en el intervalo de 0 a 1. Otras alternativas pueden calcular una intensidad de evento, pero la alternativa expresada en la Ecuación (1) se ha demostrado en varias áreas, incluyendo el control del procesamiento de dinámica. La asignación de una intensidad (proporcional a la cantidad de cambio espectral asociado con ese evento) al evento auditivo permite un mayor control sobre el procesamiento de dinámica en comparación con una decisión binaria de eventos. Son aceptables cambios mayores de ganancia durante eventos más intensos, y la señal en la Ecuación (1) permite un control variable tal.

La señal $A[t]$ es una señal de impulsos con un impulso que ocurre en la ubicación de un límite de evento. Para los fines de controlar el tiempo de liberación, se puede estabilizar adicionalmente la señal $A[t]$ para que decaiga de forma estable hasta cero tras la detección de un límite de evento. La señal $A[t]$ de control estabilizada del evento puede calcularse a partir de $A[t]$ según:

$$\bar{A}[t] = \begin{cases} A[t] & A[t] > \alpha_{\text{evento}} \bar{A}[t-1] \\ \alpha_{\text{evento}} \bar{A}[t-1] & \text{en los demás casos} \end{cases} \quad (2)$$

ES 2 377 719 T3

Aquí, α_{evento} a evento controla el tiempo de decaimiento de la señal de control del evento.

La Fig. 3 es una secuencia de gráficos que ilustra la operación y el efecto de la invención, según una realización, “b)” en la Fig. 3 representa la señal $\bar{A}[t]$ de control del evento para la correspondiente señal de audio de “a)” en la Fig. 3, con el tiempo de decaimiento medio del conjunto más estabilizado a 250 ms. La señal de audio contiene tres ráfagas de diálogo, intercaladas con sonidos de fondo de poco volumen del chisporroteo de una hoguera. La señal de control del evento muestra muchos eventos auditivos en los sonidos tanto del diálogo como del segundo plano.

En la Fig. 3, “c)” muestra la señal de ganancia del DRC en la que se usa la señal $\bar{A}[t]$ de control del evento para variar la constante del tiempo de liberación para la estabilización de la ganancia del DRC. Según describen Crocket y Seefeldt, cuando la señal de control es igual a uno, el coeficiente de estabilización de liberación no se ve afectado y la ganancia estabilizada cambia según el valor de la constante de tiempo. Cuando la señal de control es igual a cero, se impide que cambie la ganancia estabilizada. Cuando la señal de control está entre cero y uno, se permite que cambie la ganancia estabilizada, pero con una tasa reducida, en proporción a la señal de control.

En “c)” de la Fig. 3, la ganancia del DRC aumenta durante los sonidos de fondo de poco volumen debido al número de eventos detectados en segundo plano. La señal resultante modificada por el DRC en “d)” de la Fig. 3 tiene una amplificación audible y poco deseable del ruido de fondo entre las ráfagas de diálogo.

Para reducir el cambio de ganancia durante los sonidos de fondo de poco volumen, una realización de la invención modifica o pondera la intensidad auditiva $A[t]$ usando una medida de la asimetría del espectro de la señal de audio. Una realización de la invención calcula la oblicuidad espectral de la excitación de la señal de audio.

La oblicuidad es una medida estadística de la asimetría de una distribución de probabilidad. Una distribución simétrica en torno a la media tiene una oblicuidad de cero. Una distribución con su volumen o masa concentrada por encima de la media y con una larga cola que se extiende por debajo de la media tiene una oblicuidad negativa. Una distribución concentrada por debajo de la media y con una larga cola que se extiende por encima de la media tiene una oblicuidad positiva. La magnitud o el espectro de potencia de una señal de audio típica tiene oblicuidad positiva. Es decir, el grueso de la energía en el espectro se concentra más abajo en el espectro, y el espectro tiene una larga cola hacia la parte superior del espectro.

La Fig. 2 ilustra un procesador 2 de audio según una realización de la invención. El procesador 2 de audio incluye un modificador 12 de parámetros de dinámica y el identificador opcional 11 de características de eventos auditivos de la Fig. 1, así como un identificador 20 de eventos auditivos y un calculador 21 de oblicuidad. El calculador 21 de oblicuidad y el identificador 20 de eventos auditivos reciben ambos la señal 13 de audio, y el calculador 21 de oblicuidad produce una entrada para el identificador 20 de eventos auditivos. El identificador 20 de eventos auditivos, el identificador 11 de características de eventos auditivos y el modificador 12 de parámetros de dinámica están conectados, por lo demás, como sus homólogos de la Fig. 1.

En la Fig. 2, el calculador 21 de oblicuidad calcula la oblicuidad a partir de una representación espectral de la señal 13 de audio, y el identificador 20 de eventos auditivos calcula el análisis de escenas auditivas a partir de la misma representación espectral. La señal 13 de audio puede ser agrupada en bloques de M muestras que se solapan en un 50 por ciento, y la transformada discreta de Fourier puede calcularse como sigue:

$$X[k, t] = \sum_{n=0}^{M-1} x[n, t] e^{-j \frac{2\pi kn}{M}} \quad (3)$$

siendo $M = 2*N$ muestras y denotando $x[n, t]$ un bloque de muestras.

Se supone que el tamaño de bloque para la transformada es el mismo que para el cálculo de la señal del evento auditivo. Sin embargo, no es preciso que sea así. Cuando existen tasas de bloques diferentes, pueden interpolarse señales en una tasa de bloques o ser convertidas en tasa en la misma escala temporal que las señales en la otra tasa de bloques.

Se calcula la señal $E[b, t]$ de excitación que se aproxima a la distribución de energía a lo largo de la membrana basilar del oído interno en la banda crítica b durante el bloque temporal t :

$$E[b, t] = \sum_k |T[k]|^2 |C_b[k]|^2 |X[k, t]|^2 \quad (4)$$

representando $T[k]$ la respuesta frecuencial de un filtro que simula la transmisión de audio a través del oído externo y medio y representando $C_b[k]$ la respuesta frecuencial de la membrana basilar en una ubicación correspondiente a la banda crítica b .

La Fig. 4 representa la respuesta frecuencial de un filtro adecuado $T[k]$ de transmisión. La Fig. 5 representa un conjunto adecuado de respuestas de filtro de banda crítica, correspondientes a $C_b[k]$, en el cual se espacian uniformemente 40 bandas por toda la escala del ancho de banda rectangular equivalente (ERB) de Moore y Glasberg, para una tasa de muestreo de 48 kHz y un tamaño de transformada de $M = 2048$. Una función exponencial redondeada describe cada forma de filtro, y 1 ERB separa las bandas.

Si los límites de eventos auditivos se calculan a partir del espectro de sonoridad específica, según Crocket y Seefeldt, la señal $E[b,t]$ de excitación ya existe como parte del cálculo de la sonoridad específica.

Por último, la oblicuidad espectral se calcula a partir de la señal $E[b,t]$ de excitación como:

$$SK[t] = \frac{1}{B-1} \frac{\sum_{b=0}^{B-1} (E[b,t] - \mu)^3}{\sigma^3} \quad (5)$$

siendo μ la media aritmética de la excitación:

$$\mu = \frac{1}{B} \sum_{b=0}^{B-1} E[b,t] \quad (6)$$

y siendo σ la varianza de la señal de excitación:

$$\sigma = \sqrt{\frac{1}{B} \sum_{b=0}^{B-1} (E[b,t] - \mu)^2} \quad (7)$$

La señal $SK[t]$ de oblicuidad de la Ecuación (5) fluctúa considerablemente y requiere estabilización para evitar aberraciones cuando se modifica la señal de control del evento y los parámetros subsiguientes de procesamiento de dinámica. Una realización usa un estabilizador con una constante α_{SK} de decaimiento que tiene un tiempo de decaimiento medio de aproximadamente 6,5 ms:

$$SK'[t] = \alpha_{SK} SK'[t-1] + (1 - \alpha_{SK}) SK[t] \quad (8)$$

Limitar la oblicuidad a un máximo y un mínimo SK_{max} y SK_{min} , respectivamente, puede resultar útil. Puede calcularse una oblicuidad limitada $SK''[t]$ como:

$$SK''[t] = \begin{cases} 0 & SK'[t] \leq SK_{min} \\ \frac{SK'[t] - SK_{min}}{SK_{max} - SK_{min}} & SK_{min} < SK'[t] < SK_{max} \\ 1 & SK[t] \geq SK_{max} \end{cases} \quad (7)$$

Típicamente, los valores bajos (valores cercanos a 0,0) de la señal $SK''[t]$ de oblicuidad corresponden a señales característicamente de menos volumen, mientras que los valores de oblicuidad altos (valores cercanos a 1,0) corresponden típicamente a señales característicamente de más volumen. En la Fig. 3, el gráfico "e)" muestra la señal de oblicuidad que corresponde a la señal de audio en "a)" de la Fig. 3. La oblicuidad es elevada para las ráfagas de diálogo de más volumen y baja para los sonidos de fondo.

La señal $SK''[t]$ de oblicuidad pasa al identificador 20 de eventos auditivos de la Fig. 2, que pondera la medida $D[t]$ de diferencia espectral como:

$$D_{SK}[t] = SK''[t] D[t] \quad (8)$$

La señal $A_{SK}[t]$ de intensidad auditiva modificada en oblicuidad se calcula de la misma manera que $A[t]$ en la Ecuación (1):

$$A_{SK}[t] = \begin{cases} 0 & D_{SK}[t] \leq D_{\min} \\ \frac{D_{SK}[t] - D_{\min}}{D_{\max} - D_{\min}} & D_{\min} < D_{SK}[t] < D_{\max} \\ 1 & D_{SK}[t] \geq D_{\max} \end{cases} \quad (9)$$

La señal $A_{SK}[t]$ de intensidad auditiva modificada en oblicuidad es estabilizada de la misma manera que $A[t]$ en la Ecuación (2):

$$\bar{A}_{SK}[t] = \begin{cases} A[t] & A_{SK}[t] > \alpha_{evento} \bar{A}_{SK}[t-1] \\ \alpha_{evento} \bar{A}_{SK}[t-1] & \text{en los demás casos} \end{cases} \quad (10)$$

En la Fig. 3, “f)” representa la señal $\bar{A}_{SK}[t]$ de control de eventos modificada en oblicuidad para la correspondiente señal de audio en “a)” de la Fig. 3. Aparecen menos eventos auditivos durante los sonidos de fondo, mientras que permanecen los eventos correspondientes al diálogo de mayor volumen.

En la Fig. 3, “g)” muestra la señal del DRC controlada por el evento modificada en oblicuidad. Con menos eventos auditivos en los sonidos de fondo, la ganancia del DRC permanece relativamente constante y se mueve únicamente para las secciones de diálogo de mayor volumen, “h)” en la Fig. 3 muestra la señal resultante de audio modificada por el DRC.

La señal de audio modificada por el DRC no tiene nada de la amplificación poco deseable de nivel en los sonidos de fondo.

La señal $SK''[t]$ de oblicuidad baja a veces para señales perceptualmente de más volumen. Para estas señales de volumen alto, el valor de la medida $D[t]$ de diferencia espectral es suficientemente grande incluso tras la ponderación por medio de la señal $SK''[t]$ de oblicuidad de la Ecuación 8, la medida $D_{SK}[t]$ de diferencia espectral ponderada sigue siendo típicamente lo bastante grande como para indicar un límite del evento auditivo. La señal de $A_{SK}[t]$ control del evento no se ve adversamente afectada.

REIVINDICACIONES

- 5 1. Un procedimiento para controlar el procesamiento de señal de una señal de audio, que comprende dividir dicha señal de audio en eventos auditivos detectando cambios en la composición espectral con respecto al tiempo, incluyendo dicha división:
- 10 la detección de la ubicación y de la intensidad de los límites de los eventos auditivos en dicha señal de audio,
- la obtención de una medida de la oblicuidad espectral de dicha señal de audio, en el que la oblicuidad espectral es una medida estadística de la asimetría de la distribución de probabilidad del espectro de la señal de audio,
- 15 la modificación de la intensidad de un evento auditivo en respuesta a dicha medida, de modo que cuanto menos oblicuo sea el espectro en la ubicación de un evento auditivo, más se reduce la intensidad del evento auditivo, y
- el control del procesamiento de señal en respuesta a la intensidad modificada en ubicación e intensidad de dicho evento auditivo.
- 20 2. Un procedimiento según la reivindicación 1 en el que la división de dicha señal de audio en eventos auditivos incluye el análisis del espectro de la señal de audio.
3. Un procedimiento según la reivindicación 2 en el que la obtención de una medida de la oblicuidad espectral de la señal de audio incluye el cálculo de la oblicuidad a partir de una representación espectral de la señal de audio.
- 25 4. Un procedimiento según la reivindicación 3 en el que el cálculo de la oblicuidad incluye el cálculo de una señal de excitación que se aproxima a la distribución de energía a lo largo de la membrana basilar del oído interno.
5. Un procedimiento según la reivindicación 1 en el que la división de dicha señal de audio en eventos auditivos incluye la transformación de la señal de audio en un dominio de sonoridad perceptiva, incluyendo dicha transformación el cálculo de una señal de excitación que se aproxima a la distribución de energía a lo largo de la membrana basilar del oído interno.
- 30 6. Un procedimiento según la reivindicación 5 en el que la obtención de una medida de la oblicuidad espectral de la señal de audio incluye el cálculo de la oblicuidad a partir de dicha señal de excitación.
- 35 7. Un procedimiento según una cualquiera de las reivindicaciones 1-6 en el que dicha medida de la oblicuidad espectral es una medida estabilizada.
- 40 8. Un aparato que comprende medios adaptados para llevar a cabo el procedimiento de una cualquiera de las reivindicaciones 1 a 7.
9. Un programa de ordenador, almacenado en un medio legible por ordenador, que, cuando es ejecutado en un ordenador, lleva a cabo el procedimiento de una cualquiera de las reivindicaciones 1 a 7.
- 45 10. Una memoria legible por ordenador que contiene el programa de ordenador de la reivindicación 9.
11. Un sistema de ordenador que comprende:
- 50 una CPU:
- la memoria de la reivindicación 10; y
- un bus que acopla de forma comunicativa la CPU y la memoria.
- 55
- 60
- 65

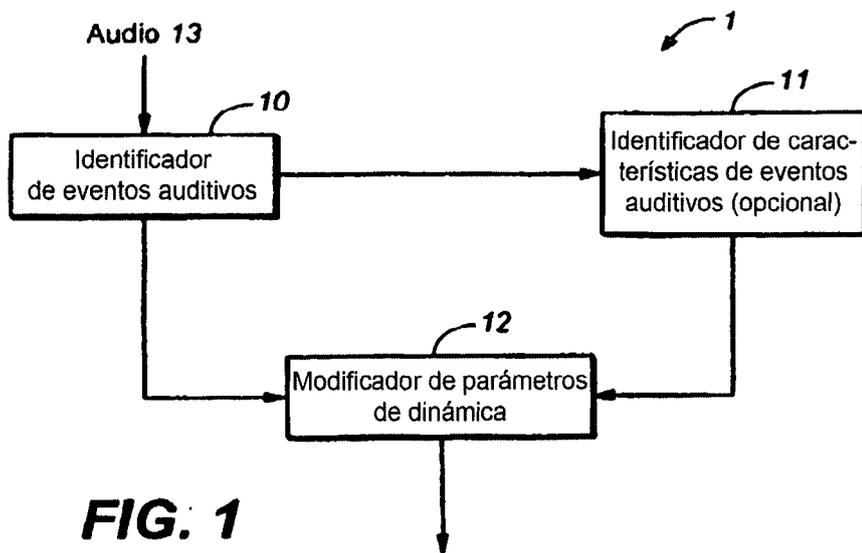


FIG. 1

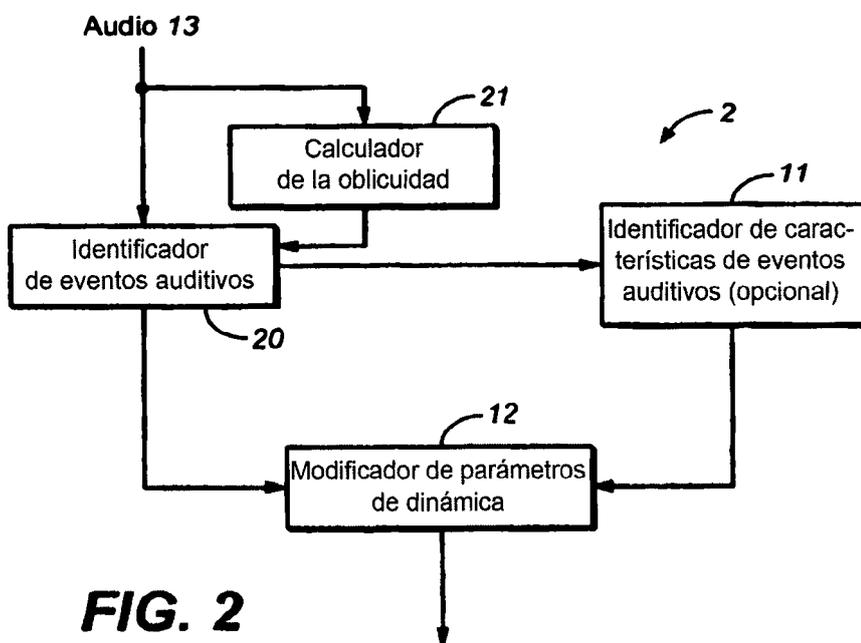


FIG. 2

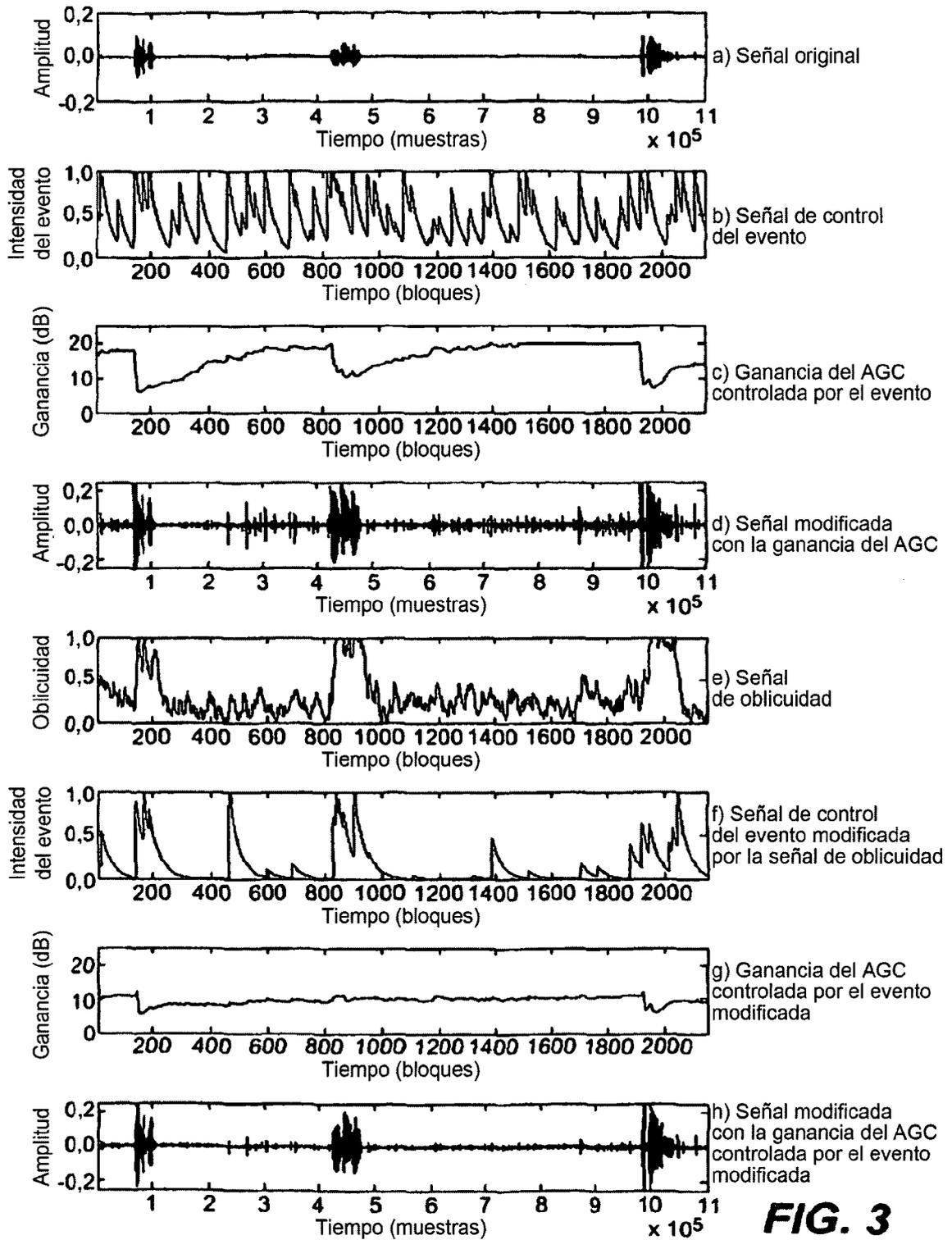


FIG. 3

