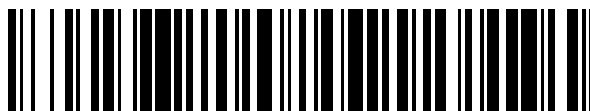


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 587 196**

51 Int. Cl.:

G10L 19/008 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **11.01.2011 PCT/EP2011/050265**

87 Fecha y número de publicación internacional: **21.07.2011 WO11086060**

96 Fecha de presentación y número de la solicitud europea: **11.01.2011 E 11700088 (5)**

97 Fecha y número de publicación de la concesión europea: **27.07.2016 EP 2524370**

54 Título: **Extracción de una señal directa/de ambiente de una señal de mezcla descendente e información paramétrica espacial**

30 Prioridad:

**26.08.2010 EP 10174230
15.01.2010 US 295278 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
21.10.2016

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
WISSENSCHAFT E.V. (100.0%)
Hansastraße 27c
80686 München, DE**

72 Inventor/es:

**VILKAMO, JUHA;
PLOGSTIES, JAN;
NEUGEBAUER, BERNHARD y
HERRE, JÜRGEN**

74 Agente/Representante:

ARIZTI ACHA, Monica

ES 2 587 196 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).

Extracción de una señal directa/de ambiente de una señal de mezcla descendente e información paramétrica espacial

DESCRIPCIÓN

5

La presente invención se refiere al procesamiento de señales de audio y, en particular, a un aparato y a un método para extraer una señal directa/del ambiente de una señal de mezcla descendente (downmix) e información paramétrica espacial. Otras formas de realización de la presente invención se refieren a la utilización de la separación de la señal directa/del ambiente para mejorar la reproducción binaural de señales de audio. Del mismo modo, otras formas de realización se refieren a la reproducción binaural de sonido de múltiples canales, donde el audio de múltiples canales significa audio que tiene dos o más canales. El contenido de audio típico que tiene sonido de múltiples canales son las bandas de sonido de las películas y las grabaciones musicales en múltiples canales.

10

15 El sistema auditivo espacial del ser humano tiende a procesar el sonido mayormente en dos partes. Por un lado, se trata de una parte localizable o directa y, por el otro, una parte no localizable o del ambiente. Hay muchas aplicaciones de procesamiento de audio, tal como la reproducción de sonido binaural y mezcla ascendente (upmixing) de múltiples canales, donde resulta conveniente tener acceso a estos dos componentes de audio.

20

25

30

En la técnica, se describen métodos de separación directa/del ambiente, según se describe en "Primary-ambient signal decomposition and vector-based localization for spatial audio coding and enhancement", Goodwin, Jot, IEEE Intl. Conf. On Acoustics, Speech and Signal proc, abril de 2007; "Correlation-based ambiente extraction from stereo recordings", Merimaa, Goodwin, Jot, AES 123rd Convention, Nueva York, 2007; "Multiple-loudspeaker playback of stereo signals", C. Faller, Journal of the AES, octubre de 2007; "Primary-ambient decomposition of stereo audio signals using a complex similarity index"; Goodwin et al., publicación número: US2009/0198356 A1, agosto de 2009; solicitud de patente que lleva el título "Method to Generate Multi-Channel Audio Signal from Stereo Signals", inventores: Christof Faller, agentes: FISH & RICHARDSON P.C., Cesionaria: LG ELECTRONICS, INC., origen: Minneapolis, MN, Estados Unidos, Clase IPC8: AH04R500FI, Clase USPC: 381 1; y "Ambience generation for stereo signals", Avendano et al., fecha de emisión: 28 de julio de 2009, solicitud 10/163.158, presentada: 4 de junio de 2002, los cuales pueden utilizarse para diversas aplicaciones. Los algoritmos de separación de la señal directa-del ambiente del estado de la técnica se basan en la comparación de señales inter-canal de sonido estéreo en bandas de frecuencia.

35

40

45

Además, en "Binaural 3-D Audio Rendering Based on Spatial Audio Scene Coding", Goodwin, Jot, AES 123rd Convention, Nueva York 2007, también se trata la reproducción binaural con extracción del ambiente. Del mismo modo, se hace referencia a la extracción del ambiente en relación con la reproducción binaural en J. Usher y J. Benesty, "Enhancement of spatial sound quality: a new reverberation-extraction audio upmixer", IEEE Trans. Audio, Speech, Language Processing, volumen 15, páginas 2141-2150, septiembre de 2007. Este último trabajo se centra en la extracción del ambiente en grabaciones con micrófono en estéreo, usando filtrado adaptativo en canales cruzados con criterio de mínimos cuadrados promediados del componente directo en cada canal. Típicamente, los codificadores-decodificadores de audio espacial, por ejemplo MPEG Surround, consisten en una corriente de audio de uno o dos canales en combinación con información lateral espacial, que extiende el audio en varios canales, tal como se describe en ISO/IEC 23003-1 - MPEG Surround; y en Breebaart, J., Herre, J., Villemoes, L., Jin, C., Kjörling, K., Plogsties, J., Koppens, J. (2006). "Multi-channel goes mobile: MPEG Surround binaural rendering". Proc. 29th AES conference, Seúl, Corea.

50

Sin embargo, las tecnologías de codificación de audio paramétricas modernas, tales como MPEG Surround (MPS, por sus siglas en inglés) y estéreo paramétrico (PS, por sus siglas en inglés) sólo proporcionan un número reducido de canales de mezcla descendente de audio - en algunos casos sólo uno - junto con información lateral espacial adicional. La comparación entre los canales de entrada "originales" sólo puede hacerse después de la primera decodificación del sonido en el formato de salida pretendido.

55

Por lo tanto, se requiere un concepto para extraer una porción de la señal directa o una porción de la señal del ambiente desde una señal de mezcla descendente e información paramétrica espacial. Sin embargo, no existen soluciones para la extracción de la señal directa/del ambiente usando la información lateral paramétrica.

60

La publicación de patente WO2005/101905 A1 desvela la extracción desde componentes directo y del ambiente de la señal de mezcla descendente basándose en información de nivel estimado (balance) obtenida de información paramétrica espacial.

La publicación de patente WO2007/110101 A1 desvela análogamente la generación de componentes directo y difuso desde una señal de mezcla descendente basándose en parámetros que tienen información acerca de la estructura temporal.

Por lo tanto, constituye un objeto de la presente invención proporcionar un concepto para extraer una porción de la señal directa o una porción de la señal del ambiente de una señal de mezcla descendente mediante el uso de información paramétrica espacial.

5 Este objeto se logra con un aparato según la reivindicación 1, un método según la reivindicación 15 o un programa informático según la reivindicación 16.

Una idea básica es que la extracción directa/del ambiente anteriormente mencionada pueda lograrse cuando se estima un nivel de información de una porción directa o una porción del ambiente de una señal de audio de múltiples canales según la información paramétrica espacial y una porción de la señal directa o una porción de la señal de ambiente se extrae de una señal de mezcla descendente basada en la información de nivel estimada. En este punto, la señal de mezcla descendente y la información paramétrica espacial representan la señal de audio de canales múltiples que tiene más canales que la señal de mezcla descendente. Esta medida permite una extracción directa y/o del ambiente de una señal de mezcla descendente que tiene uno o más canales de entrada usando la información paramétrica lateral espacial.

Por consiguiente un aparato para extraer una señal directa/del ambiente de una señal de mezcla descendente e información paramétrica espacial comprende un estimador directo/del ambiente y un extractor directo/del ambiente. La señal de mezcla descendente y la información paramétrica espacial representan una señal de audio de canales múltiples que tiene más canales que la señal de mezcla descendente. Además, la información paramétrica espacial comprende relaciones inter-canal de la señal de audio de múltiples canales. El estimador directo/del ambiente se configura para estimar una información de nivel de una porción directa o una porción del ambiente de la señal de audio de múltiples canales basada en la información paramétrica espacial. El extractor directo/del ambiente está configurado para extraer una porción de la señal directa o una porción de la señal del ambiente desde la señal de mezcla descendente basada en la información de nivel estimada de la porción directa o de la porción del ambiente.

Según otro ejemplo, el aparato para extraer una señal directa/del ambiente de una señal de mezcla descendente e información paramétrica espacial también comprende un dispositivo que genera sonido directo binaural, un dispositivo que genera sonido del ambiente binaural y un combinador. El dispositivo que genera sonido directo binaural está configurado para procesar la porción de la señal directa a fin de obtener una primera señal de salida binaural. El dispositivo que genera sonido del ambiente binaural está configurado para procesar la porción de la señal del ambiente a fin de obtener una segunda señal de salida binaural. El combinador está configurado para combinar las señales de salida binaural primera y segunda a fin de obtener una señal de salida binaural combinada. Por lo tanto, se puede proporcionar una reproducción binaural de una señal de audio, donde la porción de la señal directa y la porción de la señal del ambiente de la señal de audio se procesan por separado.

A continuación, se explican las formas de realización de la presente invención con referencia a los dibujos adjuntos, en los cuales:

40 La Figura 1 muestra un diagrama de bloques de una forma de realización de un aparato para extraer una señal directa/del ambiente de una señal de mezcla descendente e información paramétrica espacial que representa una señal de audio de múltiples canales;

45 La Figura 2 muestra un diagrama de bloques de una forma de realización de un aparato para extraer una señal directa/del ambiente de una señal de mezcla descendente mono e información paramétrica espacial que representa una señal de audio estéreo paramétrica;

50 La Figura 3a muestra una ilustración esquemática de la descomposición espectral de una señal de audio de múltiples canales según una forma de realización de la presente invención;

La Figura 3b muestra una ilustración esquemática para calcular las relaciones inter-canal de una señal de audio de múltiples canales basada en la descomposición espectral de la Figura 3a;

55 La Figura 4 muestra un diagrama de bloques de una forma de realización de un extractor directo/del ambiente mezclado descendente (downmixing) de la información de nivel estimada.

La Figura 5 muestra un diagrama de bloques de otra forma de realización de un extractor directo/del ambiente aplicando parámetros de ganancia a una señal de mezcla descendente;

60 La Figura 6 muestra un diagrama de bloques de otra forma de realización de un extractor directo/del ambiente basado en una solución LMS con mezcla cruzada de canales;

La Figura 7a muestra un diagrama de bloques de una forma de realización de un estimador directo/del ambiente

usando una fórmula de estimación del ambiente en estéreo;

- 5 La Figura 7b muestra un gráfico de un ejemplo de proporción de energía directa a total comparada con coherencia inter-canal;
- La Figura 8 muestra un diagrama de bloques de un sistema de codificador/decodificador según una forma de realización de la presente invención;
- 10 La Figura 9a muestra un diagrama de bloques de una vista general de la producción de un sonido directo binaural según una forma de realización de la presente invención;
- La Figura 9b muestra un diagrama de bloques de los detalles de la producción del sonido directo binaural de la Figura 9a;
- 15 La Figura 10a muestra un diagrama de bloques de una vista general de la producción de un sonido del ambiente binaural según una forma de realización de la presente invención;
- La Figura 10b muestra un diagrama de bloques de los detalles de la producción de sonido del ambiente binaural de la producción de sonido del ambiente binaural de la Figura 10a;
- 20 La Figura 11 muestra un diagrama de bloques conceptual de una forma de realización de reproducción binaural de una señal de audio de múltiples canales;
- La Figura 12 muestra un diagrama de bloques general de una forma de realización de una extracción directa/del ambiente que incluye la reproducción binaural;
- 25 La Figura 13a muestra un diagrama de bloques de una forma de realización de un aparato para extraer una señal directa/del ambiente de una señal de mezcla descendente mono en un dominio de banco de filtros;
- La Figura 13b muestra un diagrama de bloques de una forma de realización de un bloque de extracción de la señal directa/del ambiente de la Figura 13a; y
- 30 La Figura 14 muestra una ilustración esquemática de un ejemplo de un esquema de decodificación MPEG Surround según otra forma de realización de la presente invención.

35 La Figura 1 muestra un diagrama de bloques de una forma de realización de un aparato 100 para extraer una señal directa/de ambiente 125-1, 125-2 de una señal de mezcla descendente 115 e información paramétrica espacial 105. Tal como se muestra en la Figura 1, la señal de mezcla descendente 115 y la información paramétrica espacial 105 representan una señal de audio de múltiples canales 101 que tiene más canales $Ch_1 \dots Ch_N$ que la señal de mezcla descendente 115. La información paramétrica espacial 105 puede comprender relaciones inter-canales de la señal de audio de múltiples canales 101. En particular, el aparato 100 comprende un estimador directo/del ambiente 110 y un extractor directo/del ambiente 120. El estimador directo/del ambiente 110 puede configurarse para estimar una información de nivel 113 de una porción directa o una porción del ambiente de la señal de audio de múltiples canales 101 basada en la información paramétrica espacial 105. El extractor directo/del ambiente 120 puede estar configurado para extraer una porción de la señal directa 125-1 o una porción de la señal del ambiente 125-2 desde la señal de mezcla descendente 115 basada en la información de nivel estimada 113 de la porción directa o de la porción del ambiente.

50 La Figura 2 muestra un diagrama de bloques de una forma de realización de un aparato 200 para extraer una señal directa/del ambiente 125-1, 125-2 de una señal de mezcla descendente mono 215 e información paramétrica espacial 105 que representa una señal de audio estéreo paramétrica 201. El aparato 200 de la Figura 2 esencialmente comprende los mismos bloques que el aparato 100 de la Figura 1. Por lo tanto, los bloques idénticos que tienen implementaciones y/o funciones similares se identifican con las mismas referencias numéricas. Además, la señal de audio estéreo paramétrica 201 de la Figura 2 puede corresponder a la señal de audio de múltiples canales 101 de la Figura 1, y la señal de mezcla descendente mono 215 de la Figura 2 puede corresponder a la señal de mezcla descendente 115 de la Figura 1. En la forma de realización de la Figura 2, la señal de mezcla descendente mono 215 y la información paramétrica espacial 105 representan la señal de audio estéreo paramétrica 201. La señal de audio estéreo paramétrica puede comprender un canal izquierdo que lleva la indicación "L" y un canal derecho que lleva la indicación "R". En este punto, el extractor directo/de ambiente 120 está configurado para extraer la porción de la señal directa 125-1 o la porción de la señal del ambiente 125-2 de la señal de mezcla descendente mono 215 según la información de nivel estimada 113, que se puede derivar de la información paramétrica espacial 105 a través del uso del estimador directo/del ambiente 110.

60

En la práctica, los parámetros espaciales (información paramétrica espacial 105) de la forma de realización de la

Figura 1 o de la Figura 2, respectivamente, se refieren especialmente a MPEG Surround (MPS) o información lateral paramétrica estéreo (PS). Estas dos tecnologías son métodos de codificación de audio envolvente o estéreo de baja tasa de bits del estado de la técnica. Con referencia a la Figura 2, PS proporciona un canal de audio de mezcla descendente con parámetros espaciales y, con referencia a la Figura 1, MPS proporciona uno, dos o más canales de audio de mezcla descendente con parámetros espaciales.

Específicamente, las formas de realización de la Figura 1 y de la Figura 2 muestran claramente que la información lateral paramétrica espacial 105 puede usarse fácilmente en el campo de la extracción directa y/o del ambiente desde una señal (es decir, señal de mezcla descendente 115, 215) que tiene uno o más canales de entrada.

La estimación de niveles directos y/o del ambiente (información de nivel 113) se basa en la información sobre las relaciones inter-canal o diferencias inter-canal, tales como las diferencias y/o la correlación de nivel. Estos valores pueden calcularse a partir de una señal estéreo o de múltiples canales. La Figura 3a muestra una ilustración esquemática de la descomposición espectral 300 de una señal de audio de múltiples canales ($Ch_1 \dots Ch_N$) que se ha de usar para calcular las relaciones inter-canal de las respectivas $Ch_1 \dots Ch_N$. Tal como se puede apreciar a partir de la Figura 3a, una descomposición espectral de un canal inspeccionado Ch_i de la señal de audio de múltiples canales ($Ch_1 \dots Ch_N$) o una combinación lineal R del resto de los canales, respectivamente, comprende una pluralidad 301 de subbandas, donde cada subbanda 303 de la pluralidad 301 de subbandas se extiende a lo largo del eje horizontal (eje del tiempo 310) que tiene valores de subbanda 305, tal como lo indican los recuadros pequeños de la rejilla de tiempo/frecuencia. Además, las subbandas 303 están ubicadas consecutivamente a lo largo de un eje vertical (eje de frecuencia 320) correspondiente a diferentes regiones de frecuencia de un banco de filtros. En la Figura 3a, un

respectivo mosaico de tiempo/frecuencia $X_i^{n,k}$ o $X_R^{n,k}$ se indica con una línea discontinua. En este punto, el índice i denota el canal Ch_i y R y la combinación lineal del resto de los canales, mientras que los índices n y k corresponden a ciertos intervalos de tiempo del banco de filtros 307 y las subbandas del banco de filtros 303. Según

estos mosaicos de tiempo/frecuencia $X_i^{n,k}$ y $X_R^{n,k}$, por ejemplo ubicados en el mismo punto de tiempo/frecuencia (t_0, f_0) con respecto a los ejes de tiempo/frecuencia 310, 320, las relaciones inter-canales 335, tales como las coherencias inter-canal (ICC_i) o las diferencias del nivel de canales (CLD_i) del canal inspeccionado Ch_i , pueden calcularse en una etapa 330, tal como se muestra en la Figura 3b. En este punto, el cálculo de las relaciones inter-canal ICC_i y CLD_i puede realizarse usando las siguientes relaciones:

30

$$ICC_i = \frac{\langle Ch_i R^* \rangle}{\sqrt{\langle Ch_i Ch_i^* \rangle \langle RR^* \rangle}}$$

$$\sigma_i = \frac{\langle Ch_i Ch_i^* \rangle}{\langle RR^* \rangle}$$

35 donde Ch_i es el canal inspeccionado y R la combinación lineal de los canales restantes, mientras que $\langle \dots \rangle$ denota un promedio de tiempo. Un ejemplo de la combinación lineal R de los canales restantes es su suma normalizada por energía. Adicionalmente, la diferencia de nivel de canal (CLD_i) es, típicamente, un valor en decibeles del parámetro σ_i .

40 Con referencia a las ecuaciones citadas previamente, la diferencia del nivel de canal (CLD_i) o parámetro σ_i puede corresponder a un nivel P_i del canal Ch_i normalizado en un nivel P_R de la combinación lineal R del resto de los canales. En este punto, los niveles de P_i o P_R pueden derivarse del parámetro de diferencia de nivel inter-canal $ICLD_i$ del canal Ch_i y una combinación lineal $ICLD_R$ de los parámetros de diferencia de nivel inter-canal $ICLD_j$ ($j \neq i$) del resto de los canales.

45 En este punto, $ICLD_i$ e $ICLD_j$ pueden relacionarse con un canal de referencia Ch_{ref} , respectivamente. En otras formas de realización, los parámetros de diferencia de nivel inter-canal $ICLD_i$ e $ICLD_j$ también pueden relacionarse con cualquier otro canal de la señal de audio de múltiples canales ($Ch_1 \dots Ch_N$) que es el canal de referencia Ch_{ref} . En definitiva, esto lleva al mismo resultado para la diferencia de nivel de canal (CLD_i) o el parámetro σ_i .

50 Según otras formas de realización, las relaciones inter-canal 335 de la Figura 3b también pueden derivarse operando en diferentes o en todos los pares Ch_i, Ch_j de canales de entrada de la señal de audio de múltiples canales

(Ch₁ ... Ch_N). En este caso, los parámetros de coherencia inter-canal calculados en pares ICC_{i,j} o la diferencia del nivel del canal (CLD_{i,j}) o parámetros $\sigma_{i,j}$ (o ICLD_{i,j}) pueden obtenerse, los índices (i, j) denotan un cierto par de canales Ch_i y Ch_j, respectivamente.

5 La Figura 4 muestra un diagrama de bloques de una forma de realización 400 de un extractor directo/del ambiente 420, que incluye mezclado descendente de la información de nivel estimada 113. La forma de realización de la Figura 4, esencialmente, comprende los mismos bloques que la forma de realización de la Figura 1. Por lo tanto, los bloques idénticos que tienen implementaciones y/o funciones similares se identifican con las mismas referencias numéricas. Sin embargo, el extractor directo/del ambiente 420 de la Figura 4, que puede corresponder al extractor directo/del ambiente 120 de la Figura 1, está configurado para la mezcla descendente de la información de nivel estimada 113 de la porción directa o de la porción del ambiente de la señal de audio de múltiples canales para obtener información de nivel mezclada de manera descendente de la porción directa o de la porción del ambiente y para extraer la porción de la señal directa 125-1 o la porción de la señal del ambiente 125-2 desde la señal de mezcla descendente 115 basada en la información de nivel mezclada de manera descendente. Tal como se muestra en la Figura 4, la información paramétrica espacial 105 puede derivarse, por ejemplo, de la señal de audio de múltiples canales 101 (Ch₁ ... Ch_N) de la Figura 1 y puede comprender las relaciones inter-canales 335 de Ch₁ ... Ch_N que se introducen en la Figura 3b. La información paramétrica espacial 105 de la Figura 4 también puede comprender la mezcla descendente de la información 410 que se debe alimentar en el extractor directo/del ambiente 420. En ciertas formas de realización, la información de mezclado descendente 410 puede caracterizar una mezcla descendente de una señal de audio de múltiples canales original (por ejemplo, la señal de audio de múltiples canales 101 de la Figura 1) en la señal de mezcla descendente 115. Por ejemplo, la mezcla descendente puede realizarse usando un mezclador descendente (downmixer) (que no se muestra) que funciona en cualquier dominio de codificación, tal como en un dominio de tiempo o en un dominio espectral.

25 Según otras formas de realización, el extractor directo/del ambiente 420 también puede configurarse para realizar una mezcla descendente de la información de nivel estimada 113 de la porción directa o de la porción del ambiente de la señal de audio de múltiples canales 101, para lo cual se combina la información de nivel estimada de la porción directa con la suma coherente y la información de nivel estimada de la porción del ambiente con suma incoherente.

30 Cabe destacar que la información de nivel estimada puede representar niveles de energía o niveles de potencia de la porción directa o de la porción del ambiente, respectivamente.

En particular, el mezclado descendente de las energías (es decir, la información de nivel 113) de la parte directa/del ambiente puede realizarse asumiendo la incoherencia total o la coherencia total entre los canales. Las dos fórmulas que se pueden aplicar en el caso de la mezcla descendente basada en sumas incoherentes o coherentes, respectivamente, son como sigue.

Para señales incoherentes, la energía de mezcla descendente o la información de nivel con mezcla descendente puede calcularse con

$$E_{DMX} = \sum_{i=1}^N g_i^2 E_{Ch_i}$$

Para señales coherentes, la energía con mezcla descendente o la información de nivel con mezcla descendente puede calcularse con

$$E_{DMX} = \left(\sum_{i=1}^N g_i \sqrt{E_{Ch_i}} \right)^2$$

En este punto *g* es la ganancia de mezcla descendente, que puede obtenerse a partir de la información de mezclado descendente, mientras que E(Ch_i) denota la energía de la porción directa/del ambiente de un canal Ch_i de la señal de audio de múltiples canales. Como ejemplo típico de mezclado descendente incoherente, en caso de mezclar canales 5.1 en dos, la energía de la mezcla descendente izquierdo puede ser:

$$E_{L_DMX} = E_{izquierdo} + E_{envolvente_izquierdo} + 0,5 * E_{Centro}$$

La Figura 5 muestra otra forma de realización de un extractor directo/del ambiente 520 aplicando parámetros de ganancia *g_D*, *g_A* a una señal de mezcla descendente 115. El extractor directo/del ambiente 520 de la Figura 5 puede corresponder al extractor directo/del ambiente 420 de la Figura 4. En primer lugar, la información de nivel estimada

de una porción directa 545-1 o una porción del ambiente 545-2 puede recibirse desde un estimador directo/del ambiente, según ya describió. La información de nivel recibida 545-1, 542-1 puede combinarse/mezclarse de manera descendente en la etapa 550 a fin de obtener información de nivel mezclada de manera descendente de la porción directa 555-1 o la porción del ambiente 555-2, respectivamente. A continuación, en una etapa 560, se pueden derivar los parámetros de ganancia g_D 565-1 o g_A 565-2 desde la información de nivel mezclada de manera descendente 555-1, 555-2 para la porción directa o la porción del ambiente, respectivamente. Por último, el extractor directo/del ambiente 520 puede usarse para aplicar los parámetros de ganancia derivados 565-1, 565-2 a la señal de mezcla descendente 115 (etapa 570), de modo que se obtenga la porción de señal directa 125-1 o la señal del ambiente 125-2.

En este punto, cabe destacarse que en las formas de realización de las Figuras 1; 4; 5, la señal de mezcla descendente 115 puede consistir de una pluralidad de canales de mezcla descendente ($Ch_1 \dots Ch_M$) que se encuentran presentes en las entradas de los extractores directos/del ambiente 120; 420; 520, respectivamente.

En otras formas de realización, el extractor directo/del ambiente 520 está configurado para determinar una proporción de energía directa a total (DTT) o ambiente a total (ATT) de la información del nivel mezclada de manera descendente 555-1, 555-2 de la porción directa o de la porción del ambiente y el uso como los parámetros de extracción de los parámetros de ganancia 565-1, 565-2 según la proporción de energía DTT o de ATT.

También en otras formas de realización, el extractor directo/del ambiente 520 está configurado para multiplicar la señal de mezcla descendente 115 con un primer parámetro de extracción \sqrt{DTT} para obtener la porción de señal directa 125-1 y con un segundo parámetro de extracción \sqrt{ATT} para obtener la porción de la señal del ambiente 125-2. En este punto, la señal de mezcla descendente 115 puede corresponder a la señal de mezcla descendente mono 215, tal como se muestra en la forma de realización de la Figura 2 ("caso de mezcla descendente mono").

En el caso de mezcla descendente mono, la extracción del ambiente puede hacerse aplicando \sqrt{ATT} y \sqrt{DTT} . Sin embargo, el mismo enfoque es válido también para las señales mezcla descendente de múltiples canales, en particular, aplicando $\sqrt{ATT_i}$ y $\sqrt{DTT_i}$ para cada canal Ch_i .

Según otras formas de realización, en caso de que la señal de mezcla descendente 115 comprenda una pluralidad de canales ("caso de mezcla descendente de múltiples canales"), el extractor directo/del ambiente 520 puede configurarse para aplicar una primera pluralidad de parámetros de extracción, por ejemplo $\sqrt{DTT_i}$, a la señal de mezcla descendente 115 a fin de obtener la porción de señal directa 125-1 y una segunda pluralidad de los parámetros de extracción, por ejemplo $\sqrt{ATT_i}$, a la señal de mezcla descendente 115 para obtener la porción de señal del ambiente 125-2. En este punto, las pluralidades primera y segunda de parámetros de extracción pueden constituir una matriz diagonal.

En general, el extractor directo/del ambiente 120; 420; 520 también puede configurarse para extraer la porción de señal directa 125-1 o la porción de la señal del ambiente 125-2 aplicando una matriz de extracción cuadrática M por M a la señal de mezcla descendente 115, donde un tamaño (M) de la matriz de extracción cuadrática M por M corresponde a un número (M) de los canales de mezcla descendente ($Ch_1 \dots Ch_M$).

Por lo tanto, la aplicación de la extracción del ambiente puede describirse aplicando una matriz de extracción cuadrática M por M , donde M es el número de canales de mezcla descendente ($Ch_1 \dots Ch_M$). Esto puede incluir todas las formas posibles de manipular la señal de entrada para obtener la salida directa/del ambiente, incluyendo el enfoque relativamente simple basado en los parámetros $\sqrt{ATT_i}$ y $\sqrt{DTT_i}$ que representan los principales elementos de una matriz de extracción cuadrática M por M que está configurada como una matriz diagonal, o un enfoque de mezclado cruzado (crossmixing) de LMS como una matriz integral. A continuación se describe ésta última. En este punto, cabe destacar que el enfoque anterior de aplicar la matriz de extracción M por M cubre cualquier número de canales, incluyendo uno.

Según otras formas de realización, la matriz de extracción puede no ser necesariamente una matriz cuadrática del tamaño de matriz M por M , porque podemos tener un menor número de canales de salida. Por lo tanto, la matriz de extracción puede tener un número reducido de líneas. Un ejemplo de ello sería extraer una señal directa simple en lugar de M .

No es necesario tomar siempre todos los canales mezcla descendente M como la entrada correspondiente para tener M columnas de la matriz de extracción. Esto, en particular, puede ser pertinente para aplicaciones donde no se requiere tener todos los canales como entradas.

La Figura 6 muestra el diagrama de bloques de otra forma de realización 600 de un extractor directo/del ambiente 620 basándose en una solución LMS (criterio de mínimos cuadrados promediados) con mezclado cruzado de

canales. El extractor directo/del ambiente 620 de la Figura 6 puede corresponder al extractor directo/del ambiente 120 de la Figura 1. Por lo tanto, en la forma de realización de la Figura 6, los bloques idénticos que tienen implementaciones y/o funciones similares a la forma de realización de la Figura 1 se denotan con las mismas referencias numéricas. Sin embargo, la señal de mezcla descendente 615 de la Figura 6, que puede corresponder a la señal de mezcla descendente 115 de la Figura 1, puede comprender una pluralidad 617 de canales de mezcla descendente $Ch_1 \dots Ch_M$, donde el número de canales de mezcla descendente (M) es menor que el de los canales $Ch_1 \dots Ch_N$ (N) de la señal de audio de múltiples canales 101, es decir, $M < N$. Específicamente, el extractor directo/del ambiente 620 está configurado para extraer la porción de la señal directa 125-1 o la porción de la señal del ambiente 125-2 a través de una solución de mínimos cuadrados promediados (LMS, *por sus siglas en inglés*) que no requiere niveles de ambiente iguales. Dicha solución de LMS que no requiere niveles iguales de ambiente y que también se extiende a cualquier número de canales se proporciona a continuación. La solución de LMS que se acaba de mencionar no es obligatoria, sino que representa una alternativa más precisa a la anterior.

Los símbolos usados en la solución LMS para las ponderaciones de mezclado cruzado para la extracción directa/del ambiente son:

Ch_i canal i

a_i ganancia del sonido directo en el canal i

D y \hat{D} parte directa del sonido y su estimado

A_i y \hat{A}_i parte del ambiente del canal i y su estimado

$P_x = E[XX^*]$ energía estimada de X

$E[\cdot]$ expectativa

$E_{\hat{x}}$ error de estimación de X

$w_{\hat{D}i}$ ponderaciones de LMS de mezclado cruzado para el canal i en la parte directa

$w_{\hat{A}in}$ ponderaciones de LMS de mezclado cruzado para el canal n en el ambiente del canal i

En este contexto, también cabe destacar que la derivación de la solución de LMS puede basarse en una representación espectral de los respectivos canales de la señal de audio de múltiples canales, que significa que todo funciona en las bandas de frecuencia.

El modo de señal es se proporciona por

$$Ch_i = a_i D + A_i$$

En primer lugar, la derivación se ocupa de a) la parte directa y a continuación b) la parte del ambiente. Finalmente, se deriva la solución para las ponderaciones y se describe el método para una normalización de las ponderaciones.

a) Parte directa

La estimación de las ponderaciones de la parte directa es

$$\hat{D} = \sum_{i=1}^N w_{\hat{D}i} Ch_i = \sum_{i=1}^N w_{\hat{D}i} (a_i D + A_i)$$

El error de estimación se lee como

$$E_{\hat{D}} = D - \hat{D} = D - \sum_{i=1}^N w_{\hat{D}i} (a_i D + A_i)$$

Para tener la solución de LMS, necesitamos $E_{\hat{D}}$ ortogonal respecto de las señales de entrada.

$$E[E_{\hat{D}} Ch_k] = 0, \text{ para todas las } k$$

$$\begin{aligned}
 & E \left[\left(D - \sum_{i=1}^N w_{\hat{D}i} (a_i D + A_i) \right) (a_k D + A_k)^* \right] \\
 &= \left(a_k - \sum_{i=1}^N w_{\hat{D}i} a_i a_k \right) P_D - w_{\hat{D}k} P_{Ak} = 0 \\
 &\Leftrightarrow \sum_{i=1}^N w_{\hat{D}i} a_i a_k P_D + w_{\hat{D}k} P_{Ak} = a_k P_D
 \end{aligned}$$

En la forma de matriz, la relación anterior se lee

$$\mathbf{A}\bar{\mathbf{w}} \equiv \bar{\mathbf{P}}$$

$$\begin{bmatrix} (a_1 a_1 P_D + P_{A1}) & a_1 a_2 P_D & \cdots & a_1 a_N P_D \\ a_2 a_1 P_D & (a_2 a_2 P_D + P_{A2}) & & \vdots \\ \vdots & & \ddots & \\ a_N a_1 P_D & \cdots & & (a_N a_N P_D + P_{AN}) \end{bmatrix} \begin{bmatrix} w_{\hat{D}1} \\ w_{\hat{D}2} \\ \vdots \\ w_{\hat{D}N} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} P_D$$

5

b) Parte del ambiente

Partimos del mismo modelo de señal y estimamos las ponderaciones a partir de

10

$$\hat{A}_i = \sum_{n=1}^N w_{\hat{A}i,n} Ch_n = \sum_{n=1}^N w_{\hat{A}i,n} (a_n D + A_n)$$

El error de estimación es

$$E_{\hat{A}i} = A_i - \hat{A}_i = A_i - \sum_{n=1}^N w_{\hat{A}i,n} (a_n D + A_n)$$

y la ortogonalidad

15

$$E[E_{\hat{A}i} Ch_k] = 0, \text{ para todas las } k$$

$$\begin{aligned}
 & E \left[\left(A_i - \sum_{n=1}^N w_{\hat{A}i,n} (a_n D + A_n) \right) (a_k D + A_k)^* \right] \\
 &= \begin{cases} -\sum_{n=1}^N w_{\hat{A}i,n} a_n a_k P_D - w_{\hat{A}i,k} P_{Ak} = 0 & , \text{ si } i \neq k \\ -\sum_{n=1}^N w_{\hat{A}i,n} a_n a_k P_D - w_{\hat{A}i,k} P_{Ak} + P_{Ak} = 0 & , \text{ si } i = k \end{cases} \\
 &\Leftrightarrow \begin{cases} \sum_{n=1}^N w_{\hat{A}i,n} a_n a_k P_D + w_{\hat{A}i,k} P_{Ak} = 0 & , \text{ si } i \neq k \\ \sum_{n=1}^N w_{\hat{A}i,n} a_n a_k P_D + w_{\hat{A}i,k} P_{Ak} = P_{Ak} & , \text{ si } i = k \end{cases}
 \end{aligned}$$

20 En la forma de matriz, la relación precedente se lee

$$AW = P$$

$$\begin{bmatrix} (a_1 a_1 P_D + P_{A1}) & a_1 a_2 P_D & \dots & a_1 a_N P_D \\ a_1 a_2 P_D & (a_2 a_2 P_D + P_{A2}) & & \vdots \\ \vdots & & \ddots & \\ a_1 a_N P_D & \dots & & (a_N a_N P_D + P_{AN}) \end{bmatrix} \begin{bmatrix} w_{\hat{A}1,1} & w_{\hat{A}2,1} & \dots & w_{\hat{A}N,1} \\ w_{\hat{A}1,2} & w_{\hat{A}2,2} & & \vdots \\ \vdots & & \ddots & \\ w_{\hat{A}1,N} & \dots & & w_{\hat{A}N,N} \end{bmatrix} = \begin{bmatrix} P_{A1} & 0 & \dots & 0 \\ 0 & P_{A2} & & \vdots \\ \vdots & & \ddots & \\ 0 & \dots & & P_{AN} \end{bmatrix}$$

Soluciones para las ponderaciones

- 5 Las ponderaciones pueden resolverse invirtiendo la matriz A, que es idéntica en ambos cálculos tanto de la parte directa como de la parte del ambiente. En caso de señales estéreo, la solución es:

$$w_{D1} = \frac{a_1 P_D P_{A2}}{a_2 a_2 P_D P_{A1} + a_1 a_1 P_D P_{A2} + P_{A1} P_{A2}} = \frac{a_1 P_D P_{A2}}{div}$$

$$w_{D2} = \frac{a_2 P_D P_{A1}}{div}$$

$$w_{\hat{A}1,1} = \frac{a_2 a_2 P_D P_{A1} + P_{A1} P_{A2}}{div}$$

$$w_{\hat{A}1,2} = \frac{a_1 a_2 P_D P_{A1}}{div}$$

$$w_{\hat{A}2,1} = \frac{a_1 a_2 P_D P_{A2}}{div}$$

$$w_{\hat{A}2,2} = \frac{a_1 a_1 P_D P_{A2} + P_{A1} P_{A2}}{div}$$

- 10 donde *div* es el divisor $a_2 a_2 P_D P_{A1} + a_1 a_1 P_D P_{A2} + P_{A1} P_{A2}$

Normalización de las ponderaciones

- 15 Las ponderaciones son para la solución de LMS, pero debido a que se deben conservar los niveles de energía, las ponderaciones se normalizan. Esto también hace innecesaria la división por el término *div* en las fórmulas precedentes. La normalización se produce al garantizar que las energías de los canales directo y del ambiente sean P_D y P_{Ai} , donde *i* es el índice del canal.

- 20 Esto es directo suponiendo que conocemos las coherencias inter-canal, los factores de mezclado y las energías de los canales. Por cuestiones de simplicidad, nos centramos en el caso de dos canales y especialmente en el par de ponderaciones $w_{\hat{A}1,1}$ y $w_{\hat{A}1,2}$ que son las ganancias producen el primer canal del ambiente de los canales de entrada primero y segundo. Las etapas son las siguientes:

- 25 Etapa 1: calcular la energía de la señal de salida (donde la parte coherente se suma en toda la amplitud, y la parte incoherente en toda la energía)

$$P_{\hat{A}1} = \left(w_{\hat{A}1,1} \sqrt{|ICC| \cdot P_1} + \text{sign}(ICC) w_{\hat{A}1,2} \sqrt{|ICC| \cdot P_2} \right)^2 + (1 - |ICC|) P_1 w_{\hat{A}1,1}^2 + (1 - |ICC|) P_2 w_{\hat{A}1,2}^2$$

- 30 Etapa 2: calcular el factor de ganancia de la normalización

$$g = \sqrt{\frac{P_{A1}}{P_{\hat{A}1}}}$$

y aplicar el resultado a los factores de ponderación de mezclado cruzado $w_{\hat{A}1,1}$ y $w_{\hat{A}1,2}$. En la etapa 1, los valores absolutos y los operadores de signos para ICC se incluyen para tener en cuenta también el caso de que los canales

de entrada sean negativamente coherentes. El resto de los factores de ponderación también se normalizan de la misma manera.

5 En particular, y con referencia a lo anterior, el extractor directo/del ambiente 620 puede configurarse para derivar la solución de LMS suponiendo un modelo de señal de múltiples canales estable, de modo que la solución de LMS no se restrinja a una señal de mezcla descendente del canal estéreo.

10 La Figura 7a muestra un diagrama de bloques de una forma de realización 700 de un estimador directo/del ambiente 710 que se basa en una fórmula de estimación del ambiente estéreo. El estimador directo/del ambiente 710 de la Figura 7 puede corresponder al estimador directo/del ambiente 110 de la Figura 1. En particular, el estimador directo/del ambiente 710 de la Figura 7 está configurado para aplicar una fórmula de estimación del ambiente estéreo usando la información paramétrica espacial 105 para cada canal (Ch_i) de la señal de audio de múltiples canales 101, donde la fórmula de estimación del ambiente estéreo puede representarse como una dependencia funcional.

15

$$DTT_i = f_{DTT} [\sigma_i (Ch_i, R), ICC_i (Ch_i, R)],$$

$$ATT_i = 1 - DTT_i$$

20 que muestra explícitamente una dependencia de la diferencia del nivel de canal (CLD_i) o parámetro σ_i y un parámetro de coherencia inter-canal (ICC_i) del canal Ch_i . Tal como se indica en la Figura 7, la información paramétrica espacial 105 se alimenta en el estimador directo/del ambiente 710 y puede comprender los parámetros de relación inter-canal ICC_i y σ_i para cada canal Ch_i . Después de aplicar esta fórmula de estimación del ambiente estéreo con el uso del estimador directo/del ambiente 710, la proporción de energía directa a total (DTT_i) o ambiente a total (ATT_i), respectivamente, se obtendrá como su salida 715. Cabe destacar que la fórmula de estimación del ambiente estéreo indicada previamente que se usa para estimar las respectivas proporciones de energía DTT o ATT no se basa en una condición de ambiente igual.

25

En particular, la estimación de la proporción de señal directa/del ambiente puede realizarse porque la proporción (DTT) de la energía directa en un canal en comparación con la energía total de ese canal puede formularse de la siguiente manera:

30

$$Proporción = \frac{1}{2} \left[\left(1 - \frac{1}{\sigma} \right) + \sqrt{\left(\frac{1}{\sigma} - 1 \right)^2 + 4 \frac{ICC^2}{\sigma}} \right]$$

$$\sigma = \frac{\langle ChCh^* \rangle}{\langle RR^* \rangle} \quad ICC = \frac{\langle ChR^* \rangle}{\sqrt{\langle ChCh^* \rangle \langle RR^* \rangle}}$$

35 donde $\langle \rangle$ es el promedio de tiempo. Esta fórmula se aplica cuando se supone que el nivel del ambiente es igual en el canal y en la combinación lineal del resto de los canales, y su coherencia es cero.

40

La Figura 7b muestra un gráfico 750 de un ejemplo de una proporción de energía DTT (directa a total) 760 como una función del parámetro de coherencia inter-canal ICC 770. En la forma de realización de la Figura 7b, la diferencia de nivel del canal (CLD) o parámetro σ a modo de ejemplo se fija en 1 ($\sigma = 1$), de modo que el nivel P (Ch_i) del canal Ch_i y el nivel P(R) del resto de los canales será igual. En este caso, la proporción de energía DTT 760 será linealmente proporcional al parámetro ICC, según lo indica la línea recta 775 que se marca por $DTT \sim ICC$. En la Figura 7b puede apreciarse que en caso de que el $ICC = 0$, que puede corresponder a una relación inter-canal totalmente incoherente, la proporción de energía DTT 760 será 0, lo que puede corresponder a una situación totalmente de ambiente (caso 'R₁'). Sin embargo, un caso de $ICC = 1$, que puede corresponder a una relación inter-canal totalmente coherente, la proporción de energía DTT 760 puede ser 1, lo cual puede corresponder a una situación totalmente directa (caso 'R₂'). Por lo tanto, en el caso de R₁, no hay esencialmente energía directa, mientras que en el caso R₂, no hay esencialmente energía del ambiente en un canal con respecto a la energía total de ese canal.

45

50 La Figura 8 muestra un diagrama de bloques de un sistema de codificador/decodificador 800 según otras formas de realización de la presente invención. En el lado del sistema codificador/decodificador 800, se muestra una forma de realización del decodificador 820, que puede corresponder al aparato 100 de la Figura 1. Debido a la similitud de las formas de realización de las Figuras 1 y 8, los bloques idénticos que tienen implementaciones o funciones similares

en estas formas de realización se denotan con las mismas referencias numéricas. Tal como se muestra en las formas de realización de la Figura 8, el extractor directo/del ambiente 120 puede estar operativo en una señal de mezcla descendente 115 que tiene la pluralidad $Ch_1 \dots Ch_M$ de los canales de mezcla descendente. El estimador directo/del ambiente 110 de la Figura 8 también puede configurarse para recibir al menos dos canales de mezcla descendente 825 de la señal de mezcla descendente 815 (opcional), de modo que la información de nivel 113 de la porción directa o de la porción del ambiente de la señal de audio de múltiples canales 101 se estimará según la información paramétrica espacial 105 sobre, al menos, dos canales de mezcla descendente 825 recibidos. Finalmente, la porción de la señal directa 125-1 o la porción de la señal del ambiente 125-2 se obtendrán después de la extracción mediante el extractor directo/del ambiente 120.

En el lado del codificador del sistema codificador/decodificador 800, se muestra una forma de realización de un codificador 810, que puede comprender un mezclador descendente 815 para mezclar descendente de la señal de audio de múltiples canales ($Ch_1 \dots Ch_N$) en la señal de mezcla descendente 115 que tiene la pluralidad $Ch_1 \dots Ch_M$ de los canales de mezcla descendente, donde el número de canales se reduce de N a M. El mezclador descendente 825 también puede estar configurado para producir información paramétrica espacial 105 calculando las relaciones inter-canal a partir de la señal de audio de múltiples canales 101. En el sistema codificador/decodificador 800 de la Figura 8, la señal de mezcla descendente 115 y la información paramétrica espacial 105 pueden transmitirse desde el codificador 810 al decodificador 820. En este punto, el codificador 810 puede derivar una señal codificada basada en la señal de mezcla descendente 115 y la información paramétrica espacial 105 para transmisión desde el lado del codificador al lado del decodificador. Además, la información paramétrica espacial 105 se basa en información del canal de la señal de audio de múltiples canales 101.

Por otro lado, los parámetros de la relación inter-canal $\sigma_i(Ch_i, R)$ y $ICC_i(Ch_i, R)$ pueden calcularse entre el canal Ch_i y la combinación lineal R del resto de los canales en el codificador 810 y transmitirse dentro de la señal codificada. El decodificador 820 puede, a su vez, recibir la señal codificada y ser operativo en los parámetros de la relación inter-canal transmitidos $\sigma_i(Ch_i, R)$ e $ICC_i(Ch_i, R)$.

Por otro lado, el codificador 810 también puede estar configurado para calcular los parámetros de coherencia inter-canal $ICC_{i,j}$ entre pares de diferentes canales (Ch_i, Ch_j) a transmitir. En este caso, el decodificador 810 debe poder derivar los parámetros $ICC_i(Ch_i, R)$ entre el canal Ch_i y la combinación lineal R del resto de los canales de los parámetros calculados en pares transmitidos $ICC_{i,j}(Ch_i, Ch_j)$, de modo que puedan ponerse en práctica las formas de realización correspondientes que se describieron previamente. Cabe destacar en este contexto que el decodificador 820 no puede reconstruir los parámetros $ICC_i(Ch_i, R)$ a partir del conocimiento de la señal de mezcla descendente 115 solamente.

En ciertas formas de realización, los parámetros espaciales transmitidos no sólo son acerca de comparaciones de canales en pares.

Por ejemplo, el caso de MPS más típico es que haya dos canales de mezcla descendente. El primer conjunto de parámetros espaciales en la decodificación MPS forma los dos canales en tres: central, izquierdo y derecho. El conjunto de parámetros que guía este mapeo se llama coeficiente de predicción central (CPC, *por sus siglas en inglés*) y un parámetro ICC que es específico de esta configuración de dos a tres.

El segundo conjunto de parámetros espaciales divide a cada uno en dos: los canales laterales en los correspondientes canales delantero y trasero, y el canal central en el canal central e izquierdo. El mapeo se refiere a los parámetros ICC y CLD introducidos anteriormente.

No es práctico hacer reglas de cálculo para todas las clases de configuraciones de mezclado descendente y todas las clases de parámetros espaciales. Sin embargo, es práctico seguir virtualmente las etapas de mezclado descendente. Como sabemos la manera en que los dos canales se transforman en tres, y los tres en seis, en definitiva encontramos una relación entrada-salida de la manera que dos canales de entrada se encaminan a seis salidas. Las salidas son sólo combinaciones lineales de los canales mezcla descendente, más las combinaciones lineales de sus versiones no correlacionadas. No es necesario decodificar realmente la señal de salida y medirla, pero como conocemos esta "matriz de decodificación", podemos calcular con eficiencia los parámetros ICC y CLD entre cualquier canal o combinación de canales en el dominio paramétrico.

Independientemente de la configuración de la señal de mezcla descendente y de múltiples canales, cada salida de la señal decodificada es una combinación lineal de las señales mezcla descendente más una combinación lineal de una versión decorrelacionada de cada una de ellas.

$$Ch_out_i = \sum_{k=1}^{canales_dmx} (a_{k,i} Ch_dmx_k + b_{k,i} D[Ch_dmx_k])$$

5 donde el operador $D[]$ corresponde a un decorrelacionador, es decir, un proceso que hace un duplicado incoherente de la señal de entrada. Los factores a y b son conocidos, dado que derivan directamente de la información lateral paramétrica. Esto se debe a que, por definición, la información paramétrica es la guía para la manera en que el decodificador crea la salida de múltiples canales de las señales de mezcla descendente. La fórmula anterior puede simplificarse a

$$Ch_out_i = \sum_{k=1}^{canales_dmx} (a_{k,i} Ch_dmx_k) + D_i$$

10 dado que todas las partes decorrelacionadas pueden combinarse para hacer una comparación de energía/coherencia. Se conoce la energía de D , dado que los factores b también se conocían en la primera fórmula.

15 Desde este punto, cabe destacar que podemos hacer cualquier clase de comparación de coherencia y energía entre los canales de salida o entre las diferentes combinaciones lineales de los canales de salida. En caso de un ejemplo simple de dos canales de mezcla descendente, y un conjunto de canales de salida, entre los cuales, por ejemplo, los canales número 3 y 5 se comparan entre sí, la sigma se calcula de la siguiente manera:

$$\sigma_{3,5} = \frac{E[Ch_out_3^2]}{E[Ch_out_5^2]}$$

20 donde $E[]$ es la expectativa (en la práctica: promedio) del operador. Ambos términos pueden formularse de la siguiente manera:

$$\begin{aligned} E[Ch_out_i^2] &= E\left[\left(\sum_{k=1}^2 (a_{k,i} Ch_dmx_k) + D_i\right)^2\right] \\ &= E[D_i^2] + \sum_{k=1}^2 (a_{k,i}^2 E[Ch_dmx_k^2]) + 2a_{1,i}a_{2,i} (E[Ch_dmx_1 Ch_dmx_2]) \end{aligned}$$

25 Todos los parámetros indicados previamente son conocidos o mensurables a partir de las señales de mezcla descendente. Por definición, los términos cruzados $E[Ch_dmx*D]$ eran cero y, por lo tanto, no están en la última fila de la fórmula. Del mismo modo, la fórmula de coherencia es

$$ICC_{3,5} = \frac{E[Ch_out_3 Ch_out_5]}{\sqrt{E[Ch_out_3^2] E[Ch_out_5^2]}}$$

30 Nuevamente, dado que todas las partes de la fórmula anterior son una combinación lineal de las entradas más la señal decorrelacionada, la solución está fácilmente disponible.

35 En los ejemplos anteriores se compararon dos canales de salida, pero de la misma manera se puede hacer una comparación entre combinaciones lineales de los canales de salida, del mismo modo que con un proceso de ejemplo que se describirá más adelante.

40 En una síntesis de las formas de realización previas, la técnica/el concepto que se presentaron pueden comprender las siguientes etapas:

- 45 1. Recuperar las relaciones inter-canal (coherencia, nivel) de un conjunto "original" de canales que puede ser más alto que el número de canal o canales de mezcla descendente.
2. Estimar las energías directa y del ambiente en este conjunto de canales "original".
3. Mezclar de manera descendente las energías directa y del ambiente de este conjunto de canales "original" en un número menor de canales.

4. Usar las energías mezcladas de manera descendente para extraer las señales directa y del ambiente en los canales de mezcla descendente que se proporcionan aplicando factores de ganancia o una matriz de ganancia.

5 El uso de información lateral paramétrica espacial se explica y se sintetiza mejor en la forma de realización de la Figura 2. En la forma de realización de la Figura 2, tenemos una corriente estéreo paramétrica, que incluye un canal de audio simple e información lateral espacial acerca de las diferencias inter-canal (coherencia, nivel) del sonido estéreo que representa. Ahora, dado que conocemos las diferencias inter-canal, podemos aplicarles la fórmula de estimación del ambiente estéreo, y obtener las energías directa y del ambiente de los canales estéreo originales. A continuación, podemos “mezclar de manera descendente” (*downmix*) las energías de los canales agregando las energías directas juntas (con suma coherente) y energías del ambiente (con suma incoherente) y derivar las proporciones de energía directa a total y ambiente a total del canal mezcla descendente simple.

15 Con referencia a la forma de realización de la Figura 2, la información paramétrica espacial esencialmente comprende coherencia inter-canal (ICC_L , ICC_R) y parámetros de diferencia de nivel de canal (CLD_L , CLD_R) que corresponden al canal izquierdo (L) y derecho (R) de la señal de audio estéreo paramétrica, respectivamente. En este punto, cabe destacarse que los parámetros de coherencia inter-canal ICC_L e ICC_R son iguales ($ICC_L = ICC_R$), mientras que los parámetros de diferencia de nivel CLD_L y CLD_R se relacionan mediante $CLD_L = -CLD_R$. De manera correspondiente, dado que los parámetros de diferencia de nivel del canal CLD_L y CLD_R son, típicamente, valores en decibeles de los parámetros σ_L y σ_R , respectivamente, los parámetros σ_L y σ_R para los canales izquierdo (L) y derecho (R) se relacionan mediante $\sigma_L = 1/\sigma_R$. Estos parámetros de diferencias inter-canal puede utilizarse fácil y rápidamente para calcular las respectivas proporciones de energía directa a total (DTT_L , DTT_R) y ambiente a total (ATT_L , ATT_R) para ambos canales (L, R) basándose en la fórmula de estimación del ambiente estéreo. En la fórmula de estimación del ambiente estéreo, las proporciones de energía directa a total y ambiente a total (DTT_L , ATT_L) del canal izquierdo (L) dependen de los parámetros de diferencia inter-canal (CLD_L , ICC_L) para el canal L, mientras que las proporciones de energía directa a total y ambiente a total (DTT_R , ATT_R) del canal derecho (R) dependen de los parámetros de diferencia inter-canal (CLD_R , ICC_R) para el canal derecho R. Además, las energías (E_L , E_R) para ambos canales L, R, de la señal de audio estéreo paramétrica pueden derivarse según los parámetros de diferencia de nivel del canal (CLD_L , CLD_R) para el canal izquierdo (L) y para el canal derecho (R), respectivamente. En este punto, la energía (E_L) para el canal izquierdo L puede obtenerse aplicando el parámetro de diferencia de nivel de canal (CLD_L) para el canal izquierdo L a la señal de mezcla descendente mono, mientras que la energía (E_R) para el canal R puede obtenerse aplicando el parámetro de diferencia de nivel de canal (CLD_R) para el canal derecho R a la señal de mezcla descendente mono. A continuación, multiplicando las energías (E_L , E_R) para ambos canales (L, R) con los correspondientes parámetros basados en DTT_L , DTT_R y ATT_L , ATT_R - se pueden obtener las energías directa (E_{DL} , E_{DR}) y del ambiente (E_{AL} , E_{AR}) para ambos canales (L, R). A continuación, las energías directas (E_{DL} , E_{DR}) para ambos canales (L, R) pueden combinarse/sumarse usando una regla de mezclado descendente coherente para obtener una energía mezclada ($E_{D,mono}$) para la porción directa de la señal de mezcla descendente mono, mientras que las energías del ambiente (E_{AL} , E_{AR}) para ambos canales (L, R) pueden obtenerse/sumarse usando una regla de mezclado descendente incoherente para obtener una energía mezclada de manera descendente ($E_{A,mono}$) para la porción del ambiente de la señal de mezcla descendente. A continuación, relacionando las energías mezcladas ($E_{D,mono}$, $E_{A,mono}$) para la porción de la señal directa y la porción de la señal del ambiente respecto de la energía total (E_{mono}) de la señal de mezcla descendente mono, se obtiene la proporción de energía directa a total (DTT_{mono}) y ambiente a total (ATT_{mono}) de la señal de mezcla descendente mono. Finalmente, basándose en estas proporciones de energía DTT_{mono} y ATT_{mono} , la porción de la señal directa o la porción de la señal del ambiente pueden extraerse, esencialmente, de la señal de mezcla descendente mono.

45 En reproducción de audio, a menudo surge la necesidad de reproducir el sonido a través de auriculares. La escucha en auriculares tiene una característica específica que la hace drásticamente diferente de la escucha por altavoz y también es diferente del sonido natural del entorno. El audio se fija directamente en el oído izquierdo y derecho. Típicamente, el contenido de audio se produce para reproducción en altavoz. Por lo tanto, las señales de audio no contienen las propiedades y entradas que usa nuestro sistema auditivo en la percepción del sonido espacial. Este es el caso, salvo que se introduzca en el sistema el procesamiento binaural en el sistema.

55 Fundamentalmente, puede decirse que el procesamiento binaural es un proceso que toma el sonido de entrada y lo modifica para que contenga sólo las propiedades inter-aurales y monoaurales que son perceptualmente correctas (en cuanto a la manera en que el sistema auditivo procesa el sonido espacial). El procesamiento binaural no es una tarea directa y las soluciones existentes según el estado de la técnica tienen varias sub-optimizaciones.

60 Existe un gran número de aplicaciones en las cuales el procesamiento binaural para la reproducción de música y de películas ya está incluido, tal como los reproductores de medios y dispositivos de procesamiento que están diseñados para transformar las señales de audio de múltiples canales en la contraparte binaural para auriculares. Un enfoque típico es usar las funciones de transferencias relativas a la cabeza (HRTF) para hacer altavoces virtuales y agregar un efecto de habitación a la señal. Esto, en teoría, puede ser equivalente a escuchar con altavoces en una habitación específica.

5 Sin embargo, la práctica ha demostrado repetidamente que este enfoque no siempre ha dado satisfacción a los oyentes. Aparentemente, hay un compromiso de que la buena espacialización con este método directo provenga con el precio de perder calidad de audio, tal como tener cambios no preferidos en el color o timbre del sonido, lo cual perturba la percepción del efecto de la habitación y produce pérdida de dinámica. Entre otros problemas se incluye la localización imprecisa (por ejemplo, localización en cabeza, confusión frente-parte posterior), falta de distancia espacial de las fuentes del sonido y falta de coincidencia inter-aural, es decir sensación auditiva cerca de los oídos debido a entradas inter-aurales incorrectas.

10 Diferentes oyentes pueden juzgar los problemas de manera diferente. La sensibilidad también varía según el material de entrada, tal como música (estrictos criterios de calidad en cuanto a color del sonido), películas (menos estrictos) y juegos (aún menos estrictos, pero la localización es importante). También hay diferentes metas de diseño según el contenido.

15 Por lo tanto, la descripción a continuación se refiere a un enfoque para solucionar los problemas indicados previamente de la manera más satisfactoria posible para maximizar la calidad general percibida en promedio.

20 La Figura 9 muestra un diagrama de bloques de una vista general 900 de un dispositivo que produce sonido directo binaural 910 según otras formas de realización de la presente invención. Tal como se muestra en la Figura 9a, el dispositivo que produce sonido directo binaural 910 está configurado para procesar la porción de la señal directa 125-1, que puede estar presente en la salida del extractor directo/del ambiente 120 en la forma de realización de la Figura 1, a fin de obtener una primera señal de salida binaural 915. La señal de salida binaural 915 puede comprender un canal izquierdo que lleva la indicación "L" y un canal derecho que lleva la indicación "R".

25 En este punto, el dispositivo que produce sonido directo binaural 910 puede configurarse para alimentar la porción de señal directa 125-1 a través de funciones de transferencia relativas a la cabeza (HRTF) para obtener una porción de señal directa transformada. El dispositivo que produce sonido directo binaural 910 también puede configurarse para aplicar efecto de habitación a la porción de la señal directa transformada para obtener, finalmente, la primera señal de salida binaural 915.

30 La Figura 9b muestra un diagrama de bloques de detalles 905 del dispositivo que produce sonido directo binaural 910 de la Figura 9a. El dispositivo que produce sonido directo binaural 910 puede comprender un "transformador HRTF" indicado por un bloque 912 y un dispositivo para el procesamiento de efecto de habitación (reverberación paralela o simulación de reflexiones previas) que se indica con el bloque 914. Tal como se muestra en la Figura 9b, el transformador HRTF 912 y el dispositivo que procesa el efecto de habitación 914 pueden ser operativos en la porción de la señal directa 125-1 aplicando las funciones de transferencia relativas a la cabeza (HRTF) y efecto de habitación en paralelo, de modo que se obtenga la primera señal de salida binaural 915.

40 Específicamente, con referencia a la Figura 9b, este procesamiento del efecto de habitación también puede proporcionar una señal directa reverberada incoherente 919, que se puede procesar por medio de un filtro de mezclado cruzado posterior 920 para adaptar la señal a la coherencia interaural de los campos de sonido difuso. En este punto, la salida combinada del filtro 920 y el transformador HRTF 912 constituyen la primera señal de salida binaural 915. Según otras formas de realización, el procesamiento del efecto de habitación sobre el sonido directo también puede ser una representación paramétrica de reflexiones anteriores.

45 Por lo tanto, en formas de realización, el efecto de habitación puede aplicarse, preferentemente, en paralelo a los HRTF y no en serie (es decir, aplicando efecto de habitación después de alimentar la señal a través de los HRTF). Específicamente, sólo el sonido que se propaga directamente desde la fuente atraviesa, o es transformado por, los correspondientes HRTF. El sonido indirecto/reverberado puede aproximarse para entrar completamente al oído, es decir, de manera estadística (empleando control de coherencia en lugar de los HRTF). También puede haber implementaciones en serie, pero se prefiere el método paralelo.

50 La Figura 10a muestra un diagrama de bloques de una vista general 1000 de un dispositivo que produce sonido del ambiente binaural 1010 según otras formas de realización de la presente invención. Tal como se muestra en la Figura 10a, el dispositivo que produce sonido del ambiente binaural 1010 puede configurarse para procesar la salida de la porción de la señal del ambiente 125-2, por ejemplo, desde el extractor directo/del ambiente 120 de la Figura 1, para obtener la segunda señal de salida binaural 1015. La segunda señal de salida binaural 1015 también puede comprender un canal izquierdo (L) y un canal derecho (R).

60 La Figura 10b muestra un diagrama de bloques de detalles 1005 del dispositivo que produce sonido del ambiente binaural 1010 de la Figura 10a. En la Figura 10b puede observarse que el dispositivo que produce sonido del ambiente binaural 1010 puede configurarse de manera de aplicar el efecto de la habitación, según se indica con el bloque 1012 que se denota como "procesamiento del efecto de la habitación" a la porción de la señal del ambiente

125-2, de modo que se obtenga una señal del ambiente reverberada incoherente 1013. El dispositivo que produce sonido del ambiente binaural 1010 también puede configurarse para procesar la señal del ambiente reverberada incoherente 1013 aplicando un filtro tal como el filtro de mezclado cruzado que se indica con el bloque 1014, de modo que se proporcione la segunda señal de salida binaural 1015, la segunda señal de salida binaural 1015 se adaptará a la coherencia interaural de los campos de sonido difusos reales. El bloque 1012, denotado como “procesamiento del efecto de habitación” puede configurarse también de modo que produzca directamente la coherencia interaural de los campos de sonido difusos reales. En este caso no se usa el bloque 1014.

Según otra forma de realización, el dispositivo que produce sonido del ambiente binaural 1010 está configurado para aplicar el efecto de habitación y/o un filtro a la porción de la señal del ambiente 125-2 para proporcionar la segunda señal de salida binaural 1015, de modo que la segunda señal de salida binaural 1015 se adaptará a la coherencia interaural de los campos de sonidos difusos reales.

En las formas de realización indicadas previamente, se puede realizar la decorrelación y control de coherencia en dos etapas consecutivas, pero esto no es un requisito indispensable. También es posible lograr el mismo resultado con un proceso de una única etapa, sin la formulación intermedia de señales incoherentes. Ambos métodos son igualmente válidos.

La Figura 11 muestra un diagrama de bloques conceptual 1100 de una forma de realización de reproducción binaural de una señal de audio de entrada de múltiples canales 101. Específicamente, la forma de realización de la Figura 11 representa un aparato para la reproducción binaural de la señal de audio de entrada de múltiples canales 101, que comprende un primer convertidor 1110 (“transformación de frecuencia”), el separador 1120 (“separación directa-del ambiente”), el dispositivo que produce sonido directo binaural 910 (“producción de fuente directa”), el dispositivo que produce sonido del ambiente binaural 1010 (“producción de sonido del ambiente”), el combinador 1130 según se indica por el signo ‘más’ y un segundo convertidor 1140 (“transformación de frecuencia inversa”). En particular, el primer convertidor 1110 puede configurarse para convertir la señal de audio de entrada de múltiples canales 101 en una representación espectral 1115. El separador 1120 puede configurarse para extraer la porción de la señal directa 125-1 o la porción de la señal del ambiente 125-2 de la representación espectral 1115. En este punto, el separador 1120 puede corresponder al aparato 100 de la Figura 1, especialmente incluyendo el estimador directo/del ambiente 110 y el extractor directo/del ambiente 120 de la forma de realización de la Figura 1. Tal como ya se explicó, el dispositivo que produce el sonido directo binaural 910 puede estar operativo en la porción de la señal directa 125-1 a fin de obtener la primera señal de salida binaural 915. En correspondencia, el dispositivo que produce el sonido del ambiente binaural 1010 puede estar operativo en la porción de la señal del ambiente 125-2 a fin de obtener la segunda señal de salida binaural 1015. El combinador 1130 puede estar configurado para combinar la primera señal de salida binaural 915 y la segunda señal de salida binaural 1015 a fin de obtener una señal combinada 1135. Finalmente, el segundo convertidor 1140 puede estar configurado para convertir la señal combinada 1135 en un dominio de tiempo para obtener una señal de salida estéreo 1150 (“salida en estéreo para auriculares”).

La operación de transformación de la frecuencia de la forma de realización de la Figura 11 ilustra que el sistema funciona en un dominio de transformación de frecuencia, que es el dominio natural en el procesamiento perceptual de audio espacial. El sistema en sí mismo no necesariamente tiene transformación de frecuencia si se usa como expansión (*add-on*) en un sistema que ya funciona en un dominio de transformación en frecuencia.

El proceso de separación directa/del ambiente que se indicó puede subdividirse en dos partes diferentes. En la parte de estimación directa/del ambiente, los niveles y/o las proporciones de la parte del ambiente directa se estiman según la combinación de un modelo de señal y las propiedades de la señal de audio. En la parte de extracción directa/del ambiente, las proporciones conocidas y la señal de entrada pueden usarse para crear la directa de salida en las señales del ambiente.

Finalmente, la Figura 12 muestra un diagrama de bloques general de una forma de realización 1200 de la estimación/extracción directa/del ambiente incluida el caso de uso de la reproducción binaural. En particular, la forma de realización 1200 de la Figura 12 puede corresponder a la forma de realización 1100 de la Figura 11. Sin embargo, en la forma de realización 1200, se muestran los detalles del separador 1120 de la Figura 11 correspondientes a los bloques 110, 120 de la forma de realización de la Figura 1, que incluye el proceso de estimación/extracción basado en la información paramétrica espacial 105. Además, en oposición a la forma de realización 1100 de la Figura 11, no se muestra ningún proceso de conversión entre los diferentes dominios en la forma de realización 1200 de la Figura 12. Los bloques de la forma de realización 1200 también son explícitamente operativos en la señal de mezcla descendente 115, que se puede derivar de la señal de audio de múltiples canales 101.

La Figura 13a muestra un diagrama de bloques de una forma de realización de un aparato 1300 para extraer una señal directa/del ambiente de una señal de mezcla descendente mono en un dominio de banco de filtros. Tal como

se muestra en la Figura 13a, el aparato 1300 comprende un banco de filtros de análisis 1310, un banco de filtros de síntesis 1320 para la porción directa y un banco de filtros de síntesis 1322 para la porción del ambiente.

5 En particular, el banco de filtros de análisis 1310 del aparato 1300 puede implementarse para realizar una transformada de Fourier de tiempo corto (STFT, *por sus siglas en inglés*) o puede, por ejemplo, configurarse como un banco de filtros QMF, mientras que los bancos de filtros de síntesis 1320, 1322 del aparato 1300 pueden implementarse para realizar una transformada de Fourier de tiempo corto inversa (ISTFT, *por sus siglas en inglés*) o bien pueden configurarse como bancos de filtros QMF.

10 El banco de filtros de análisis 1310 está configurado para recibir una señal de mezcla descendente mono 1315, que puede corresponder a la señal de mezcla descendente mono 215 según se muestra en la forma de realización de la Figura 2, y convertir la señal de mezcla descendente mono 1315 en una pluralidad de subbandas del banco de filtros 1311. Tal como se puede apreciar en la Figura 13a, la pluralidad 1311 de subbandas del banco de filtros está conectada a una pluralidad 1350, 1352 de bloques de extracción directa/del ambiente, respectivamente, donde la pluralidad 1350, 1352 de los bloques de extracción directa/del ambiente está configurada para aplicar parámetros basados en DTT_{mono} - o ATT_{mono} - 1333, 1335 a las subbandas del banco de filtros, respectivamente.

20 Los parámetros basados en DTT_{mono} , ATT_{mono} 1333, 1335 pueden suministrarse a partir de un calculador DTT_{mono} , ATT_{mono} , tal como se muestra en la Figura 13b. En particular, el calculador DTT_{mono} , ATT_{mono} 1330 de la Figura 13b puede configurarse para calcular las proporciones de energía DTT_{mono} , ATT_{mono} o para derivar los parámetros basados en DTT_{mono} , ATT_{mono} de los parámetros de diferencia de nivel del canal y de coherencia inter-canal que se proporcionan (ICC_L , CLD_L , ICC_R , CLD_R) 105 correspondientes al canal izquierdo y derecho (L, R) de una señal de audio estéreo paramétrica (por ejemplo, la señal de audio estéreo paramétrica 201 de la Figura 2), que ya se ha descrito según corresponde. En este punto, para una subbanda del banco de filtros simple, se pueden usar los correspondientes parámetros 105 y los parámetros basados en DTT_{mono} , ATT_{mono} 1333, 1335. En este contexto, cabe destacar que aquellos parámetros no son constantes en toda la frecuencia.

30 Como resultado de la aplicación de los parámetros basados en DTT_{mono} o ATT_{mono} 1333, 1335, se obtendrá una pluralidad 1353, 1355 de las subbandas del banco de filtros modificadas, respectivamente. A continuación, se alimenta la pluralidad 1353, 1355 de las subbandas del banco de filtros modificadas en los bancos de filtros de síntesis 1320, 1322, respectivamente, que se configuran para sintetizar la pluralidad 1353, 1355 de subbandas del banco de filtros modificadas, para obtener la porción de señal directa 1325-1 o la porción de señal del ambiente 1325-2 de la señal de mezcla descendente mono 1315, respectivamente. En este punto, la porción de la señal directa 1325-1 de la Figura 13a puede corresponder a la porción de la señal directa 125-1 de la Figura 2, mientras que la porción de la señal del ambiente 1325-2 de la Figura 13a puede corresponder a la porción de la señal del ambiente 125-2 de la Figura 2.

40 Con referencia a la Figura 13b, un bloque de extracción directa/del ambiente 1380 de la pluralidad 1350, 1352 de bloques de extracción directa/del ambiente de la Figura 13a especialmente comprende el calculador de DTT_{mono} , ATT_{mono} 1330 y un multiplicador 1360. El multiplicador 1360 puede configurarse para multiplicar una subbanda del banco de filtros (FB, *por sus siglas en inglés*) simple 1301 de la pluralidad de subbandas del banco de filtros 1311 con el correspondiente parámetro basado en DTT_{mono}/ATT_{mono} 1333, 1335, de modo que se obtiene una subbanda del banco de filtros simple 1365 de la pluralidad de subbandas del banco de filtros 1353, 1355. En particular, el bloque de extracción directa/del ambiente 1380 está configurado para aplicar el parámetro basado en DTT_{mono} en caso de que el bloque 1380 pertenezca a la pluralidad 1350 de bloques, mientras que está configurado para aplicar el parámetro basado ATT_{mono} en caso de que el bloque 1380 pertenezca a la pluralidad 1352 de bloques. La subbanda del banco de filtros simple modificada 1365 también puede suministrarse al respectivo banco de filtros de síntesis 1320, 1322 para la porción directa o la porción del ambiente.

50 Según formas de realización, los parámetros espaciales y los parámetros derivados se dan en una resolución de frecuencia según las bandas críticas del sistema auditivo del ser humano, por ejemplo 28 bandas, que normalmente es menor que la resolución del banco de filtros.

55 Por lo tanto, la extracción directa/del ambiente según la forma de realización de la Figura 13a esencialmente opera en diferentes subbandas en un dominio del banco de filtros basado en la coherencia inter-canal calculada de toda la subbanda y los parámetros de diferencia de nivel de canal, que pueden corresponder a los parámetros de la relación inter-canal 335 de la Figura 3b.

60 La Figura 14 muestra una ilustración esquemática de un esquema de decodificación MPEG Surround 1400 ejemplar según otra forma de realización de la presente invención. En particular, la forma de realización de la Figura 14 describe una decodificación de una mezcla descendente estéreo 1410 en seis canales de salida 1420. En este punto, las señales que se denotan con "res" son señales residuales, que son reemplazos óptimos para las señales decorrelacionadas (de los bloques que se denotan con una "D"). Según la forma de realización de la Figura 14, la

información paramétrica espacial o los parámetros de relación inter-canal (ICC, CLD) transmitidos dentro de una corriente MPS de un codificador, tal como el codificador 810 de la Figura 8 a un decodificador, tal como el decodificador 820 de la Figura 8, puede usarse para generar matrices de decodificación 1430, 1440 que se notan con “matriz de pre-decorrelacionador M1” y “matriz mixta M2”, respectivamente. Es específico de la forma de realización de la Figura 14 que la generación de los canales de salida 1420 (es decir, los canales de mezcla ascendente L, LS, R, RS, C, LFE) de los canales laterales (L, R) y el canal central (C) (L, R, C 1435) usando la matriz mixta M2 1440, esté esencialmente determinada por información paramétrica espacial 1405, que puede corresponder a la información paramétrica espacial 105 de la Figura 1, que comprende los parámetros de relación inter-canal particulares (ICC, CLD) según el estándar de MPS Surround.

En este punto, una división del canal izquierdo (L) en los correspondientes canales de salida L, LS, el canal derecho (R) en los correspondientes canales de salida R, RS y el canal central (C) en los correspondientes canales de salida C, LFE, respectivamente, pueden estar representados por una configuración uno a dos (OTT) que tiene la respectiva entrada para los correspondientes parámetros ICC, CLD.

El esquema de decodificación MPEG Surround 1400 ejemplar que específicamente corresponde a la “configuración 5-2-5” puede, por ejemplo, comprender las siguientes etapas. En una primera etapa, los parámetros espaciales o la información lateral paramétrica pueden formularse en las matrices de decodificación 1430, 1440, que se muestran en la Figura 14, según el estándar MPS Surround existente. En una segunda etapa, las matrices de decodificación 1430, 1440 pueden usarse en el dominio de parámetro para proporcionar información inter-canal de los canales de mezcla ascendente 1420. En una tercera etapa, con la información inter-canal que se proporcionó de esta manera, se pueden calcular las energías directa/del ambiente de cada canal de mezcla ascendente. En una cuarta etapa, las energías directa/del ambiente obtenidas de esta manera pueden mezclarse de manera descendente al número de canales mezcla descendente 1410. En una quinta etapa, se pueden calcular las ponderaciones que se aplicarán a los canales de mezcla descendente 1410.

Antes de seguir avanzando, cabe destacar que el proceso de ejemplo que se acaba de mencionar requiere la medición de:

$$E \left[|L_{dmx}|^2 \right], E \left[|R_{dmx}|^2 \right]$$

que son las potencias medias de los canales de mezcla descendente, y

$$E \left[L_{dmx} R_{dmx}^* \right]$$

al que se puede hacer referencia como el espectro cruzado de los canales de mezcla descendente. En este punto, intencionalmente se hace referencia a las potencias medias de los canales de mezcla descendente como las energías, dado que el término “potencia media” no es un término de uso frecuente.

El operador de expectativa indicado por los corchetes puede reemplazarse en aplicaciones prácticas por un promedio de tiempo, sea recursivo o no recursivo. Las energías y el espectro cruzado se pueden medir fácilmente a partir de la señal de mezcla descendente.

También cabe destacar que la energía de una combinación lineal de dos canales puede formularse a partir de las energías de los canales, los factores de mezcla y el espectro cruzado (todo en el dominio paramétrico, donde no se requieren operaciones de las señales).

La combinación lineal

$$Ch = aL_{dmx} + bR_{dmx}$$

tiene la siguiente energía:

$$\begin{aligned} E \left[|Ch|^2 \right] &= E \left[|aL_{dmx} + bR_{dmx}|^2 \right] = a^2 E \left[|L_{dmx}|^2 \right] + b^2 E \left[|R_{dmx}|^2 \right] + ab \left(E \left[L_{dmx} R_{dmx}^* \right] + E \left[R_{dmx} L_{dmx}^* \right] \right) \\ &= a^2 E \left[|L_{dmx}|^2 \right] + b^2 E \left[|R_{dmx}|^2 \right] + 2ab \left(\operatorname{Re} \left\{ E \left[L_{dmx} R_{dmx}^* \right] \right\} \right) \end{aligned}$$

A continuación se describen las etapas individuales del proceso de ejemplo (es decir, el esquema de decodificación).

Primera etapa (parámetros espaciales a las matrices de mezclado)

Tal como se describió previamente, se crean las matrices M1 y M2 según el estándar de MPS Surround. El elemento de la fila a-ésima - columna b-ésima de M1 es M1(a,b).

5 Segunda etapa (mezcla de matrices con energías y espectros cruzados de la mezcla descendente a la información inter-canal de los canales mezclados).

10 Ahora tenemos las matrices de mezclado M1 y M2. Tenemos que formular la manera en que se crean los canales de salida desde el canal de mezcla descendente izquierdo (L_{dmx}) y del canal de mezcla descendente derecho (R_{dmx}). Suponemos que se usan los decorrelacionadores (área gris de la Figura 14). La decodificación/mezcla ascendente en el estándar MPS básicamente proporciona, al final, la siguiente fórmula para la relación general entrada-salida en todo el proceso:

15
$$L = a_L L_{dmx} + b_L R_{dmx} + c_L D_1 [S_1] + d_L D_2 [S_2] + e_L D_3 [S_3]$$

Lo que se presentó previamente es un ejemplo del canal izquierdo delantero mezclado. Los otros canales pueden formularse de la misma manera. Los elementos D son los decorrelacionadores, a-e son las ponderaciones que se pueden calcular a partir de las entradas de las matrices M1 y M2.

20 En particular, los factores a-e se pueden formular de manera simple a partir de las entradas de matrices:

$$a_L = \sum_{i=1}^3 M1_{i,1} M2_{1,i}$$

$$b_L = \sum_{i=1}^3 M1_{i,2} M2_{1,i}$$

$$c_L = M2_{1,4}$$

$$d_L = M2_{1,5}$$

$$e_L = M2_{1,6}$$

25 y para los otros canales de la manera correspondiente.

Las señales S son

30
$$S_n = M1_{n+3,1} L_{dmx} + M1_{n+3,2} R_{dmx}$$

Estas señales S son las entradas a los decorrelacionadores desde la matriz del lado izquierdo de la Figura 14. La energía

35
$$E \left[|D[S_n]|^2 \right] = E \left[|S_n|^2 \right]$$

puede calcularse de la manera que se explicó más arriba. El decorrelacionador no afecta a la energía.

40 Una manera perceptualmente motivada para hacer extracción del ambiente de múltiples canales es comparando un canal contra la suma de todos los otros canales. (Cabe destacar que ésta es sólo una opción entre muchas). Ahora, si consideramos a modo de ejemplo el caso del canal L, el resto de los canales se leen de la siguiente manera:

45
$$X_L = \sum_{Ch=(REST)} a_{Ch} L_{dmx} + \sum_{Ch=(REST)} b_{Ch} R_{dmx} + \sum_{Ch=(REST)} c_{Ch} D_1 [S_1] + \sum_{Ch=(REST)} d_{Ch} D_2 [S_2] + \sum_{Ch=(REST)} e_{Ch} D_3 [S_3]$$

Usamos el símbolo "X" aquí porque puede generarse una confusión al usar "R" para el resto de los canales.

Entonces, la energía del canal L es:

50
$$E \left[|L|^2 \right] = a_L^2 E \left[|L_{dmx}|^2 \right] + b_L^2 E \left[|R_{dmx}|^2 \right] + c_L^2 E \left[|S_1|^2 \right] + d_L^2 E \left[|S_2|^2 \right] + e_L^2 E \left[|S_3|^2 \right] + 2ab \operatorname{Re} \left\{ E \left[L_{dmx} R_{dmx}^* \right] \right\}$$

Entonces, la energía del canal X es:

$$E[|X_L|^2] = \left(\sum_{Ch=(REST)} a_{Ch} \right)^2 E[|L_{dmx}|^2] + \left(\sum_{Ch=(REST)} b_{Ch} \right)^2 E[|R_{dmx}|^2] + \left(\sum_{Ch=(REST)} c_{Ch} \right)^2 E[|S_1|^2] + \left(\sum_{Ch=(REST)} d_{Ch} \right)^2 E[|S_2|^2] + \left(\sum_{Ch=(REST)} e_{Ch} \right)^2 E[|S_3|^2] + 2 \left(\sum_{Ch=(REST)} a_{Ch} \sum_{Ch=(REST)} b_{Ch} \right) \text{Re}\{E[L_{dmx} R_{dmx}^*]\}$$

5

Y el espectro cruzado es:

$$E[LX_L^*] = \sum_{Ch=(REST)} a_{Ch} a_L E[|L_{dmx}|^2] + \sum_{Ch=(REST)} b_{Ch} b_L E[|R_{dmx}|^2] + \sum_{Ch=(REST)} c_{Ch} c_L E[|S_1|^2] + \sum_{Ch=(REST)} d_{Ch} d_L E[|S_2|^2] + \sum_{Ch=(REST)} e_{Ch} e_L E[|S_3|^2] + \sum_{Ch=(REST)} a_L b_{Ch} E[L_{dmx} R_{dmx}^*] + \sum_{Ch=(REST)} a_{Ch} b_L E[L_{dmx} R_{dmx}^*]^*$$

Ahora podemos formular el ICC

10

$$ICC_L = \frac{\text{Re}\{E[LX_L^*]\}}{\sqrt{E[|L|^2] E[|X_L|^2]}}$$

y la sigma

$$\sigma_L = \frac{E[|L|^2]}{E[|X_L|^2]}$$

15

Tercera etapa (información inter-canal en los canales de mezcla ascendente en parámetros DTT de los canales de mezcla ascendente).

Ahora podemos calcular el DTT del canal L según:

20

$$DTT_L = \frac{1}{2} \left[\left(1 - \frac{1}{\sigma_L} \right) + \sqrt{\left(\frac{1}{\sigma_L} - 1 \right)^2 + 4 \frac{ICC_L^2}{\sigma_L}} \right]$$

25 La energía de la señal directa de L es

$$E[|D_L|^2] = DTT \cdot E[|L|^2]$$

La energía de la señal del ambiente de L es

30

$$E[|A_L|^2] = (1 - DTT) \cdot E[|L|^2]$$

Cuarta etapa (mezcla descendente de las energías directa/del ambiente)

35 Si se usa una regla de mezclado descendente incoherente, la energía del ambiente del canal de mezcla descendente izquierdo es:

$$E \left[|A_{Ldmx}|^2 \right] = E \left[|A_L|^2 \right] + E \left[|A_{Ls}|^2 \right] + \frac{E \left[|A_C|^2 \right] + E \left[|A_{LF}|^2 \right]}{2}$$

5 y, de la misma manera, para la parte directa y la parte directa y del ambiente del canal derecho. Cabe destacar que lo anterior es solo una regla de mezclado descendente. Pero también pueden existir otras reglas de mezclado descendente.

Quinta etapa (cálculo de las ponderaciones para la extracción del ambiente en los canales de mezcla descendente)

10 La proporción DDT de mezcla descendente izquierda es

$$DTT_{Ldmx} = 1 - \frac{E \left[|A_{Ldmx}|^2 \right]}{E \left[|L_{dmx}|^2 \right]}$$

15 Los factores de ponderación pueden calcularse de la manera que se describe en la forma de realización de la Figura 5 (usando el enfoque de sqrt(DTT) o sqrt(1-DDT), o bien como se indica en la forma de realización de la Figura 6 (es decir, usando un método de matriz de mezcla cruzada).

Básicamente, el proceso de ejemplo que se describió anteriormente se refiere a los parámetros CPC, ICC, y CLD en la corriente MPS respecto a las proporciones del ambiente de los canales de mezcla descendente.

20 Según otras formas de realización, típicamente existen otros medios de lograr metas similares, como así también otras condiciones. Por ejemplo, puede haber otras reglas para el mezclado descendente, otros diseños de altavoces, otros métodos de decodificación y otras maneras de hacer la estimación del ambiente de múltiples canales fuera de las que se han descrito previamente, donde un canal específico se compara con el resto de los canales.

25 Si bien se ha descrito la presente invención en el contexto de los diagramas de bloques donde los bloques representan componentes de hardware reales o lógicos, la presente invención también puede implementarse a través de un método implementado por ordenador. En este último caso, los bloques representan las correspondientes etapas del método donde estas etapas representan las funcionalidades que se implementan por los bloques de hardware lógico o físico correspondientes.

30 Las formas de realización que se describieron son meramente ilustrativas de los principios de la presente invención. Cabe comprender que se pueden introducir modificaciones y variantes a las disposiciones, y los especialistas en la técnica podrán apreciar los detalles que se describen en la presente. Por lo tanto, la invención sólo se verá limitada por las reivindicaciones de patente que acompañan y no por los detalles específicos que se presentan a modo de descripción y explicación de las formas de realización del presente documento.

35 Según los requisitos de ciertas implementaciones de los métodos de la invención, los métodos de la invención pueden implementarse en hardware o en software. La implementación puede realizarse usando un medio de almacenamiento digital, en particular un disco, un DVD o un CD que tenga almacenadas señales de control que se puedan leer electrónicamente, que cooperan con los sistemas informáticos programables, tales como los que se proponen en los métodos de la invención. Por lo tanto, en términos generales la invención puede implementarse como un producto de un programa informático con el código del programa almacenado en un portador legible en máquina, el código del programa es operativo para poner en práctica los métodos de la invención cuando el producto del programa informático se ejecuta en un ordenador. En otras palabras, los métodos inventivos son, por lo tanto, un programa informático que tiene un código de programa para llevar a cabo, por lo menos, uno de los métodos inventivos cuando el programa informático funciona en un ordenador. La señal de audio codificada de la invención se puede almacenar en cualquier medio de almacenamiento legible por máquina, tal como un medio de almacenamiento digital.

50 Una ventaja del concepto y de la técnica novedosos es que las formas de realización anteriormente mencionadas, es decir el aparato, método o programa informático, que se describen en esta solicitud permiten estimar y extraer los componentes directo y/o del ambiente de una señal de audio con la ayuda de la información paramétrica espacial. En particular, el procesamiento novedoso de la presente invención funciona en bandas de frecuencia, como típicamente en el campo de la extracción de la señal del ambiente. El concepto que se presenta es pertinente al procesamiento de señal audio, puesto que hay un número de aplicaciones que requieren la separación de

55 componentes directo y del ambiente de una señal de audio.

A diferencia de los métodos de extracción del ambiente de la técnica anterior, el actual concepto no se basa en señales de entrada estéreo solamente y puede también aplicarse a las situaciones de mezcla descendente mono. Para una mezcla descendente de un único canal, en general, ninguna diferencia entre canales puede calcularse. Sin embargo, teniendo en cuenta la información lateral espacial, la extracción de la señal del ambiente llega a ser posible en este caso también.

La presente invención es ventajosa porque utiliza los parámetros espaciales para estimar los niveles del ambiente de la señal "original". Se basa en el concepto de que los parámetros espaciales ya contienen la información sobre las diferencias entre canales de la señal estéreo "original" o de múltiples canales.

Una vez que se estiman los niveles estéreos originales o de múltiples canales del ambiente, también se pueden derivar los niveles de la directa y del ambiente en el o los canales mezcla descendente que se proporcionan. Esto puede hacerse por las combinaciones lineales (es decir, la suma ponderada) de las energías del ambiente para la parte del ambiente, y dirigir las energías o las amplitudes para la parte directa. Por lo tanto, las formas de realización de la presente invención proporcionan la estimación y la extracción del ambiente con ayuda de la información lateral espacial.

Se pueden alcanzar las siguientes propiedades beneficiosas o ventajas a partir de este concepto de procesamiento basado en información lateral.

Las formas de realización de la presente invención proporcionan la estimación de del ambiente con la ayuda de la información lateral espacial y de los canales de mezcla descendente que se proporcionan. Dicha estimación del ambiente es importante en los casos en que hay más de un canal de mezcla descendente proporcionado junto con la información lateral. La información lateral, y la información que se mide a partir de los canales de mezcla descendente, se pueden utilizar juntas en la estimación del ambiente. En MPEG Surround con una mezcla descendente estéreo, estas dos fuentes de información juntas proporcionan la información completa de las relaciones entre canales del sonido de múltiples canales originales, y la estimación del ambiente se basa en estas relaciones.

Las formas de realización de la presente invención también proporcionan mezclado descendente de las energías directas y del ambiente. En la situación que se describe respecto a la extracción del ambiente basada en información lateral, hay una etapa intermedia de estimar el ambiente en un número de canales superiores a los canales de mezcla descendente proporcionados. Por lo tanto, esta información del ambiente tiene que mapearse en el número de los canales de audio de mezcla descendente de una manera válida. Se puede hacer referencia a este proceso como mezcla descendente debido a su correspondencia con la mezcla descendente del canal de audio. Esto se puede realizar de manera muy directa combinando la energía directa y del ambiente de la misma manera en que se mezclaron de manera descendente los canales de mezcla descendente proporcionados.

La regla de mezclado descendente no tiene una solución ideal, sino que es probable que dependa de la aplicación. Por ejemplo, en MPEG Surround puede ser beneficioso tratar los canales de manera diferente (centro, altavoces delanteros, altavoces traseros) debido a su contenido de señal típicamente diferente. Además, las formas de realización proporcionan una estimación de múltiples canales del ambiente independientemente en cada canal respecto a los otros canales. Este enfoque/propiedad permite utilizar simplemente la fórmula estéreo de la estimación del ambiente a cada canal concerniente al resto de los canales. Por esta medida, no es necesario suponer un nivel igual del ambiente en todos los canales. El enfoque presentado se basa en la suposición acerca de la percepción espacial de que el componente del ambiente en cada canal es que el componente que tiene una contraparte incoherente en algunos de todos los otros canales. Un ejemplo que sugiere la validez de esta suposición es que uno de dos canales que emiten ruido (ambiente) puede dividirse en dos canales con la mitad de la energía cada uno, sin afectar en gran medida la escena del sonido percibido.

En términos del proceso de señal, es ventajoso que la estimación real de la proporción entre señal directa/del ambiente se produzca aplicando la fórmula de la estimación del ambiente que se presenta a cada canal respecto de la combinación lineal del resto de los canales.

Finalmente, las formas de realización proporcionan una aplicación de las energías directas estimadas del ambiente para extraer las señales reales. Una vez que se conocen los niveles de ambiente en los canales de mezcla descendente, se pueden aplicar dos métodos inventivos para obtener las señales del ambiente. El primer método se basa en una multiplicación simple, donde las partes directa y del ambiente para cada canal de mezcla descendente pueden generarse multiplicando la señal con $\sqrt{\text{proporción de energía directa a total}}$ y $\sqrt{\text{proporción de energía ambiente a total}}$. Esto proporciona dos señales para cada canal de mezcla descendente que son coherentes entre sí, pero tienen las energías que se estimaron para la parte directa y del ambiente.

El segundo método se basa en una solución de mínimos cuadrados promediados con mezcla cruzada de los

canales, donde la mezcla cruzada de los canales (también es posible con signos negativos) permite una mejor estimación de las señales directas y del ambiente que las soluciones indicadas previamente. En contraste a una solución de mínimos cuadrados promediados para ingreso estéreo y niveles de ambiente iguales en los canales previstos en "Multiple-loudspeaker playback of stereo signals", C. Faller, Journal of the AES, octubre de 2007, y en la solicitud de patente titulada "Method to Generate Multi-Channel Audio Signal from Stereo Signals", inventores: Christof Faller, agentes: FISH & RICHARDSON P.C., cesionaria: LG ELECTRONICS, INC., origen: Minneapolis, MN, Estados Unidos, Clase IPC8: AH04R500FI, Clase USPC: 381 1, la presente invención proporciona una solución de mínimos cuadrados promediados que no requiere niveles de ambiente iguales y también se puede extender a cualquier número de canales.

Las características adicionales del procesamiento novedoso son las que se describen a continuación. En el procesamiento del ambiente para producción binaural, el ambiente puede procesarse con un filtro que tenga la propiedad de suministrar coherencia inter-aural en bandas de frecuencia que sea similar a la coherencia inter-aural en campos de sonidos difusos reales, donde el filtro también puede incluir el efecto de la habitación. En el procesamiento de la parte directa para la producción binaural, la parte directa puede alimentarse a través de funciones de transferencia relativas a la cabeza (HRTF) con la posible adición del efecto de la habitación, tal como reflexiones y/o reverberaciones tempranas.

Además de esto, un control del "nivel de separación" que corresponde a un control de seco/húmedo se puede observar en otras formas de realización. En particular, la separación completa puede no ser deseable en muchas aplicaciones, ya que puede conducir a defectos audibles, tal como cambios abruptos, efectos de modulación, y otros. Por lo tanto, todas las partes relevantes de los procesos que se describen se pueden poner en práctica con un control del "nivel de separación" para controlar la cantidad de separación deseada y útil. Con respecto a la Figura 11, tal control del nivel de separación es indicado por una entrada del control 1105 de recuadro discontinuo para controlar la separación directa/del ambiente 1120 y/o los dispositivos de producción binaural 910, 1010, respectivamente. Este control puede funcionar similar a un control de seco/húmedo en el procesamiento de efectos de audio.

Las ventajas principales de la solución presentada son las siguientes. El sistema funciona en todas las situaciones, también con estéreo paramétrico y MPEG Surround con mezcla descendente mono, a diferencia de las soluciones previas que dependen sólo de la información de mezcla descendente. El sistema puede además utilizar la información lateral espacial transportada junto con la señal audio en corrientes de bits de audio espaciales para estimar más exactamente las energías directa y de ambiente que con el análisis entre canales simple de los canales del mezcla descendente. Por lo tanto, muchas aplicaciones, tales como el procesamiento binaural, pueden beneficiarse aplicando diversos procesos para las partes directa y del ambiente del sonido.

Las formas de realización se basan en las siguientes suposiciones psicoacústicas. Los sistemas auditivos humanos localizan las fuentes basadas en señales inter-aurales en los mosaicos de tiempo-frecuencia (áreas restringidas en cierto rango de la frecuencia y del tiempo). Si dos o más fuentes concurrentes incoherentes que se superponen en tiempo y frecuencia se presentan simultáneamente en diferentes ubicaciones, el sistema auditivo no puede percibir la ubicación de las fuentes. Esto es porque la suma de estas fuentes no produce señales inter-aurales confiables en el oyente. Se puede describir el sistema auditivo de modo que tome desde la escena de audio mosaicos de frecuencia y tiempo cerrados que proporcionen información sobre localización confiable y trate al resto como no localizable. Por estos medios el sistema auditivo puede localizar fuentes en entornos de sonido complejos. Las fuentes coherentes simultáneas tienen un diverso efecto, ellas forman aproximadamente las mismas señales inter-aurales que formaría una sola fuente entre las fuentes coherentes.

Esta es también la propiedad aprovechada por las formas de realización. El nivel del sonido localizable (directo) y no localizable (del ambiente) puede estimarse y se extraen estos componentes. El procesamiento de la señal de espacialización se aplica únicamente a la parte localizable/directa, mientras que el procesamiento de difusión/espacio/envolvente se aplica a la parte no localizable/del ambiente. Esto aporta una ventaja significativa en el diseño de un sistema de proceso binaural, puesto que muchos procesos se pueden aplicar solamente donde sean necesarios, dejando la señal restante sin afectar. Todo el proceso se produce en bandas de frecuencia que se aproximan a la resolución de la frecuencia auditiva del ser humano.

Las formas de realización se basan en una descomposición de la señal para maximizar la calidad perceptual, pero reducen al mínimo los problemas percibidos. Por tal descomposición, es posible obtener por separado el componente directo y del ambiente de una señal audio. Los dos componentes se pueden procesar adicionalmente para alcanzar un efecto o una representación deseado.

Específicamente, las formas de realización de la presente invención permiten la estimación del ambiente con la ayuda de la información lateral espacial en el dominio codificado.

La presente invención es, también, ventajosa en que los problemas típicos de la reproducción en auricular de señales audio pueden reducirse separando las señales en una señal directa y del ambiente. Las formas de realización permiten mejorar los métodos existentes de extracción de la directa/del ambiente que se aplicarán al sonido binaural que se produce para la reproducción en auricular.

- 5 El caso del uso principal del procesamiento basado en información lateral espacial es MPEG Surround natural y estéreo paramétrico (y las técnicas de codificación paramétricas similares). Las aplicaciones típicas que se benefician a partir de la extracción de la señal del ambiente son la reproducción binaural debido a la capacidad de aplicar un diverso grado del efecto de habitación a diferentes partes del sonido, y de mezclar a un número más alto
- 10 de canales debido a la capacidad de colocar y de procesar diversos componentes del sonido de manera diferente. También pueden existir aplicaciones en las cuales el usuario requiera la modificación del nivel de la señal directa/del ambiente, por ejemplo, con el propósito de mejorar la inteligibilidad del habla.

REIVINDICACIONES

1. Un aparato (100) para extraer una señal directa y/o del ambiente (125-1, 125-2) de una señal de mezcla descendente (115) e información paramétrica espacial (105), la señal de mezcla descendente (115) y la información paramétrica espacial (105) representan una señal de audio de múltiples canales (101) que tiene más canales ($Ch_1 \dots Ch_N$) que la señal de mezcla descendente (115), en el que la información paramétrica espacial (105) comprende relaciones inter-canal de la señal de audio de múltiples canales (101), comprendiendo el aparato (100):
- 5 un estimador de directo/del ambiente (110) para estimar una información de nivel directa (113) de una porción directa de la señal de audio de múltiples canales (101) y/o para estimar una información de nivel del ambiente (113) de una porción del ambiente de la señal de audio de múltiples canales (101) basándose en la información paramétrica espacial (105); y
- 10 un extractor directo/del ambiente (420) para extraer una porción de la señal directa (125-1) y/o una porción de la señal del ambiente (125-2) desde la señal de mezcla descendente (115) basándose en la información de nivel directa estimada (113) de la porción directa o basándose en la información de nivel del ambiente estimada (113) de la porción del ambiente,
- 15 en el que el extractor directo/del ambiente (420) está configurado para mezclar de manera descendente la información de nivel directa estimada (113) de la porción directa o la información de nivel del ambiente estimada (113) de la porción del ambiente para obtener información de nivel con mezcla descendente de la porción directa o de la porción del ambiente y extraer la porción de señal directa (125-1) o la porción de señal de ambiente (125-2) de la señal de mezcla descendente (115) basándose en la información de nivel de mezcla descendente.
- 20
2. El aparato según la reivindicación 1, en el que el extractor de directo/del ambiente (420) está configurado para mezclar de manera descendente la información de nivel directa estimada (113) de la porción directa combinando la
- 25 información de nivel estimada de la porción directa con la suma coherente, y en el que el extractor directo/del ambiente (420) está configurado para mezclar de manera descendente la información de nivel de ambiente estimada (113) de la porción del ambiente combinando la información de nivel estimada de la porción de ambiente con la suma incoherente.
- 30
3. El aparato según la reivindicación 2, en el que el extractor directo/del ambiente (420) también está configurado para realizar una mezcla de manera descendente de la información de nivel directa estimada (113) de la porción directa o la información de nivel del ambiente estimada (113) de la porción del ambiente combinando la información de nivel directa estimada (113) de la porción directa con la suma coherente y la información de nivel del ambiente estimada (113) de la porción del ambiente con suma incoherente.
- 35
4. El aparato según la reivindicación 2 o 3, en el que el extractor directo/del ambiente (520) está también configurado para derivar parámetros de ganancia (565-1, 565-2) de la información de nivel de mezcla descendente (555-1, 555-2) de la porción directa o de la porción del ambiente y aplicar los parámetros de ganancia derivados (565-1, 565-2) a la señal de mezcla descendente (115) para obtener la porción de la señal directa (125-1) o la porción de la señal del ambiente (125-2).
- 40
5. El aparato según la reivindicación 4, en el que el extractor directo/del ambiente (520) también está configurado para determinar una proporción de energía directa a total (DTT) o ambiente a total (ATT) de la información de nivel mezclada de manera descendente (555-1, 555-2) de la porción directa o de la porción del ambiente y el uso como los parámetros de extracción de los parámetros de ganancia (565-1, 565-2) basándose en la proporción de energía de DTT o de ATT determinada.
- 45
6. El aparato según una de las reivindicaciones 1 a 5, en el que el extractor directo/del ambiente (520) está configurado para extraer la porción de señal directa (125-1) o la porción de la señal del ambiente (125-2) aplicando una matriz de extracción cuadrática M por M a la señal de mezcla descendente (115), en el que un tamaño (M) de la matriz de extracción cuadrática M por M corresponde a un número (M) de los canales de mezcla descendente ($Ch_1 \dots Ch_M$), en el que la matriz de extracción cuadrática M por M tiene M columnas y M filas.
- 50
7. El aparato según la reivindicación 6, en el que el extractor directo/del ambiente (520) está también configurado para aplicar una primera pluralidad de parámetros de extracción a la señal de mezcla descendente (115) para obtener la porción de señal directa (125-1) y una segunda pluralidad de parámetros de extracción a la señal de mezcla descendente (115) para obtener la porción de señal del ambiente (125-2); las pluralidades primera y segunda de los parámetros de extracción constituyen una matriz diagonal.
- 55
8. El aparato según una de las reivindicaciones 1 a 7, en el que el estimador directo/del ambiente (110) está configurado para estimar la información de nivel directa (113) de la porción directa de la señal de audio de múltiples canales (101) o para estimar la información de nivel de ambiente (113) de la porción de ambiente de la señal de audio de múltiples canales (101) basándose en la información paramétrica espacial (105) y al menos dos canales de
- 60

mezcla descendente (825) de la señal de mezcla descendente (115) recibida por el estimador directo/del ambiente (110).

- 5 9. El aparato según una de las reivindicaciones 1 a 8, en el que el estimador directo/del ambiente (710) está configurado para aplicar una fórmula de estimación del ambiente estéreo usando la información paramétrica espacial (105) para cada canal (Ch_i) de la señal de audio de múltiples canales (101), en el que la fórmula de estimación del ambiente estéreo está dada por

$$DTT_i = f_{DTT} [\sigma_i(Ch_i, R), ICC_i(Ch_i, R)],$$

$$ATT_i = 1 - DTT_i$$

- 10 dependiendo de una diferencia de nivel de canal (CLD_i), que es un valor en decibeles de σ_i , y un parámetro de coherencia inter-canal (ICC_i) del canal Ch_i , y en el que R es una combinación lineal de los canales restantes, en el que DTT es una proporción de energía directa a total y ATT es una proporción de energía ambiente a total.

- 15 10. El aparato según una de las reivindicaciones 1 a 9, en el que el extractor directo/del ambiente (620) está configurado para extraer la porción de la señal directa (125-1) mediante una solución de mínimos cuadrados promediados (LMS) con mezclado cruzado de canales basándose en un modelo de señal proporcionado mediante:

$$Ch_i = a_i D + A_i,$$

- 20 en el que Ch_i es el canal i, D es una parte directa, A es una parte de ambiente, y a_i es una ganancia de un sonido directo en el canal i.

- 25 11. El aparato según la reivindicación 10, en el que el extractor directo/del ambiente (620) está configurado para derivar la solución de LMS suponiendo un modelo de señal, de modo que la solución de LMS no se restrinja a una señal de mezcla descendente del canal estéreo.

12. El aparato según una de las reivindicaciones 1 a 11, comprendiendo adicionalmente el aparato:

- 30 un dispositivo que genera sonido directo binaural (910) para procesar la porción de la señal directa (125-1) para obtener una primera señal de salida binaural (915);
 un dispositivo que genera sonido del ambiente binaural (1010) para procesar la porción de la señal del ambiente (125-2) para obtener una segunda señal de salida binaural (1015); y
 un combinador (1130) para combinar la primera (915) y la segunda (1015) señal de salida binaural para obtener
 35 una señal binaural combinada (1135).

13. El aparato según la reivindicación 12, en el que el dispositivo que produce sonido del ambiente binaural (1010) está configurado para aplicar el efecto de habitación y/o un filtro a la porción de la señal del ambiente (125-2) para proporcionar la segunda señal de salida binaural (1015), estando adaptada la segunda señal de salida binaural
 40 (1015) a la coherencia inter-aural de los campos de sonidos difusos reales.

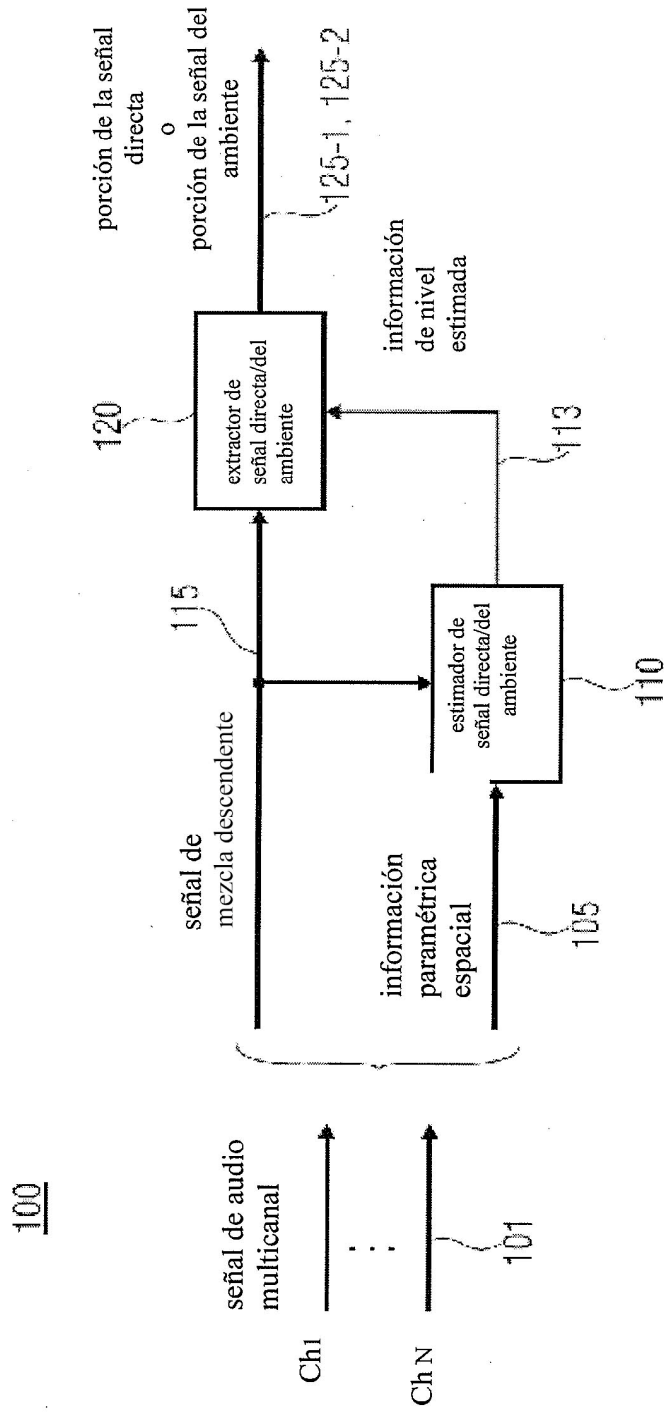
14. El aparato según la reivindicación 12 o 13, en el que el dispositivo que produce sonido directo binaural (910) está configurado para alimentar la porción de señal directa (125-1) a través de filtros basados en funciones de
 45 transferencia relativas a la cabeza (HRTF) para obtener la primera señal de salida binaural (915).

15. Un método (100) para extraer una señal directa/del ambiente (125-1, 125-2) de una señal de mezcla descendente (115) e información paramétrica espacial (105), representando la señal de mezcla descendente (115) y la información paramétrica espacial (105) una señal de audio de múltiples canales (101) que tiene más canales (Ch_1
 ... Ch_N) que la señal de mezcla descendente (115), en el que la información paramétrica espacial (105) comprende
 50 relaciones inter-canal de la señal de audio de múltiples canales (101), comprendiendo el método (100):

- estimar (110) una información de nivel directa (113) de una porción directa de la señal de audio de múltiples canales (101) y/o estimar (110) una información de nivel del ambiente (113) de una porción del ambiente de la
 55 señal de audio de múltiples canales (101) basándose en la información paramétrica espacial (105); y
 extraer (420) una porción de la señal directa (125-1) y/o una porción de la señal del ambiente (125-2) desde la señal de mezcla descendente (115) basándose en la información de nivel directa estimada (113) de la porción directa o basándose en la información de nivel de ambiente estimada (113) de la porción del ambiente,
 en el que la extracción (420) comprende mezclar de manera descendente la información de nivel directa
 60 estimada (113) de la porción directa o la información de nivel del ambiente estimada (113) de la porción de ambiente para obtener información de nivel mezclada de manera descendente de la porción directa o de la

porción de ambiente y extraer la porción de señal directa (125-1) o la porción de señal de ambiente (125-2) de la señal de mezcla descendente (115) basándose en la información de nivel mezclada de manera descendente.

16. Un programa informático que tiene un código de programa para realizar el método (100) de la reivindicación 15
5 cuando el programa informático se ejecuta en un ordenador.



(APARATO PARA EXTRAER UNA SEÑAL DIRECTA/DEL AMBIENTE)

FIG. 1

200

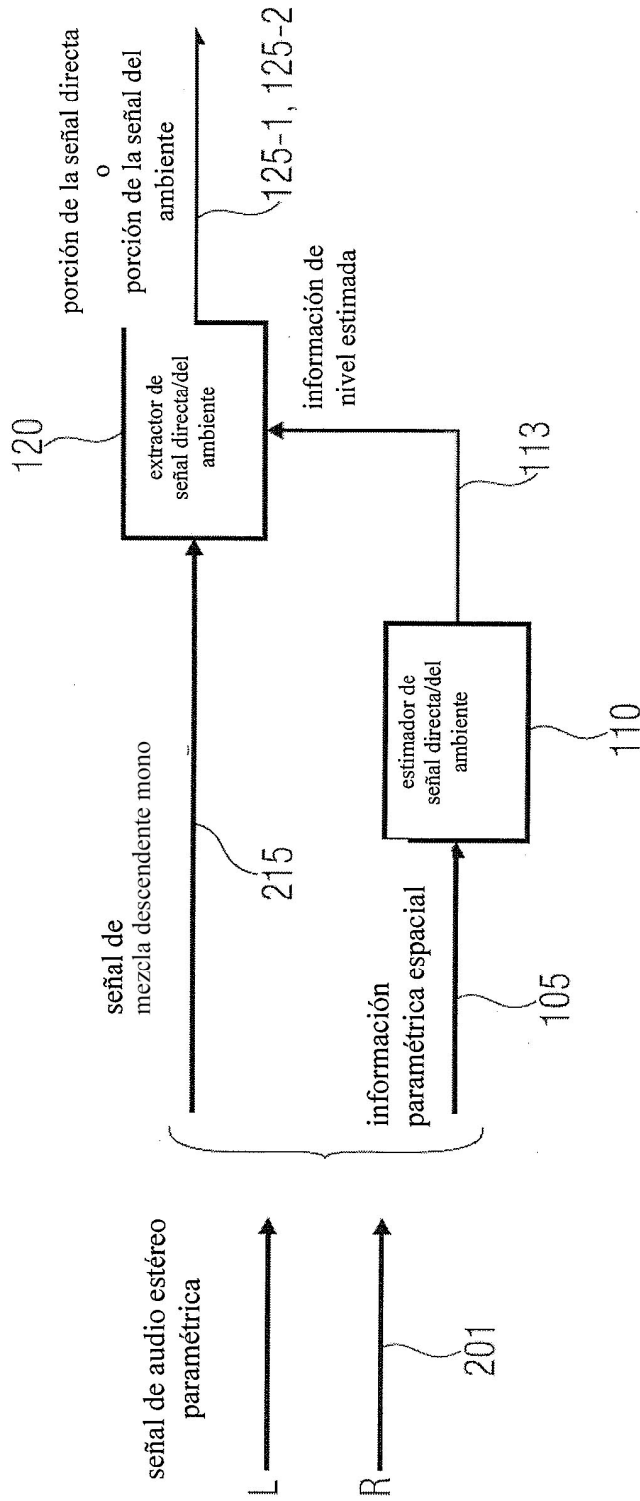


FIG. 2

(CORRIENTE ESTÉREO PARAMÉTRICA)

300

descomposición espectral de Ch_i (canal inspeccionado), $i=1...N$

descomposición espectral de R (suma del resto de los canales)

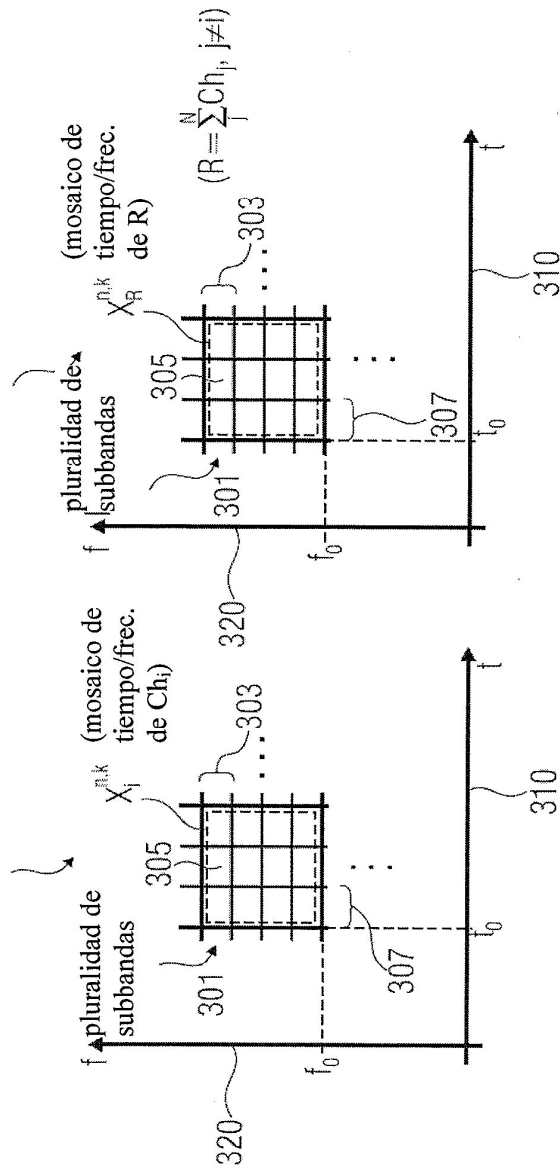


FIG. 3 A

(DESCOMPOSICIÓN ESPECTRAL DE LA SEÑAL DE AUDIO MULTICANAL $Ch_1...Ch_N$)

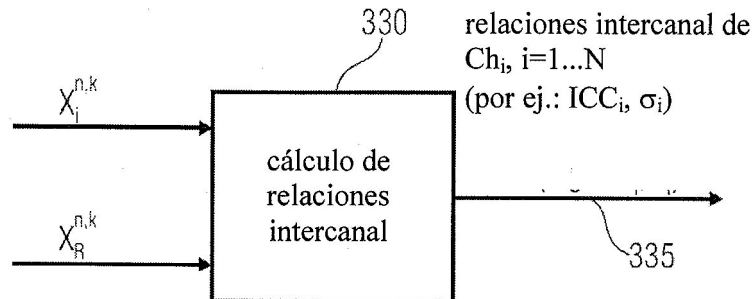


FIG. 3 B

(RELACIONES INTERCANAL)

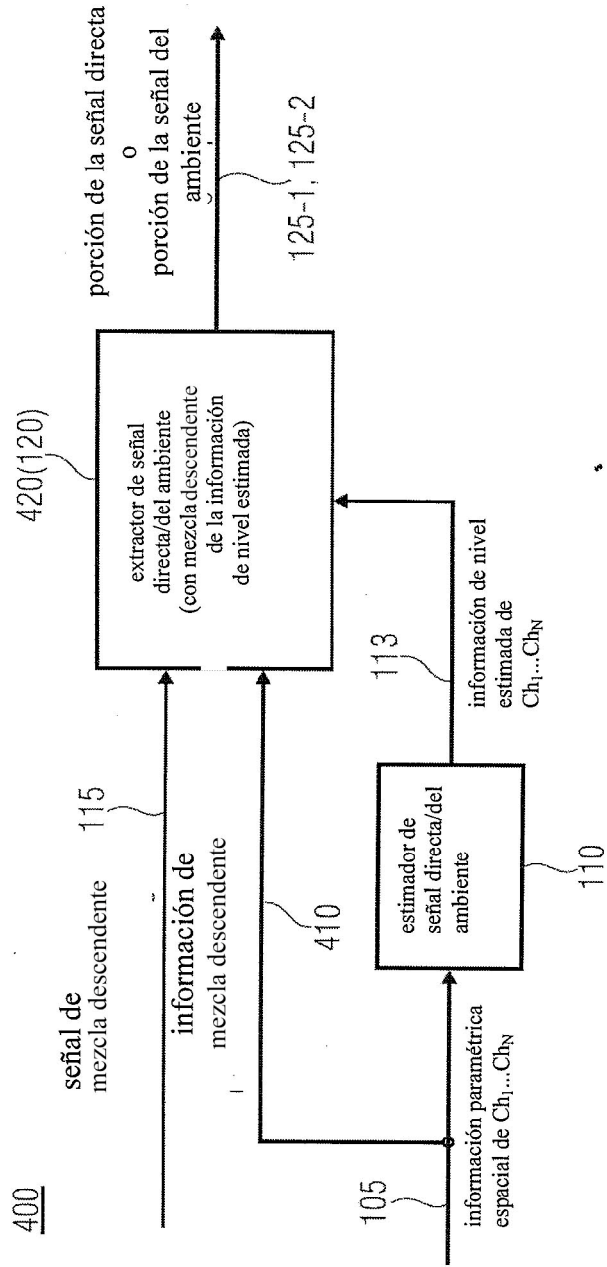


FIG. 4
 (EXTRACCIÓN DE LA SEÑAL DIRECTA/DEL AMBIENTE CON
 MEZCLA DESCENDENTE DE LA INFORMACIÓN DE NIVEL
 ESTIMADA)

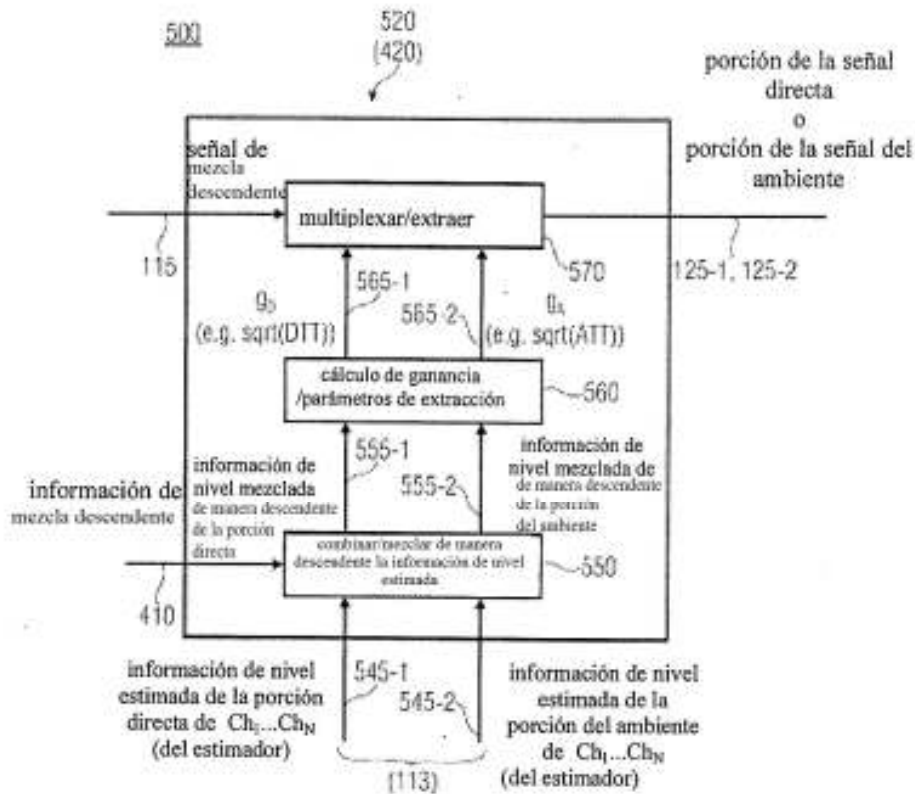


FIG. 5

(EXTRACCIÓN DE LA SEÑAL DIRECTA/DEL AMBIENTE APLICANDO PARÁMETROS DE GANANCIA)

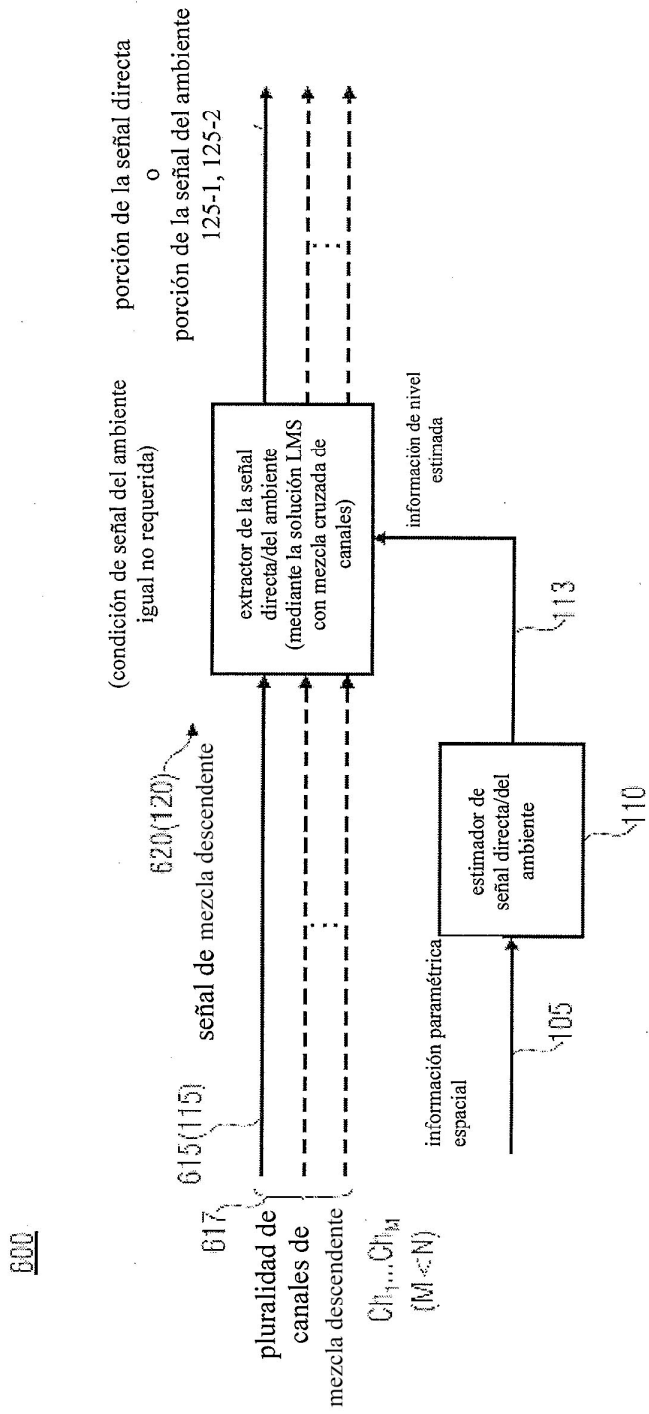


FIG. 6

(EXTRACCIÓN DE LA SEÑAL DIRECTA/DEL AMBIENTE
MEDIANTE LA SOLUCIÓN LMS)

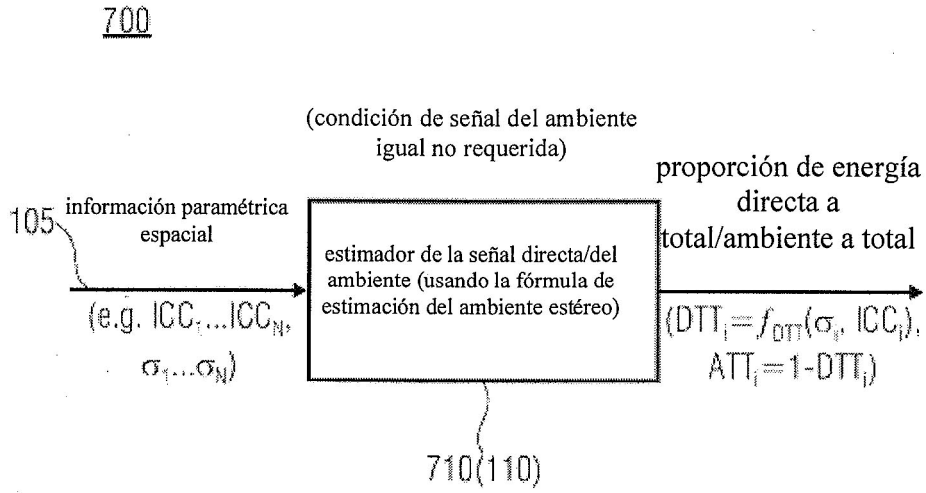
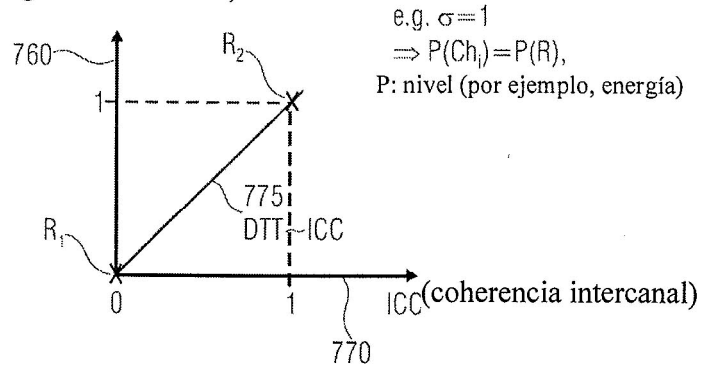


FIG. 7 A

(ESTIMACIÓN DE LA SEÑAL DIRECTA/DEL AMBIENTE USANDO LA FÓRMULA DE ESTIMACIÓN DEL AMBIENTE ESTÉREO)

750

DTT (proporción de energía directa a total)



R_1 : ICC=0 (totalmente incoherente)
 \Rightarrow DTT=0 (totalmente del ambiente)

R_2 : ICC=1 (totalmente coherente)
 \Rightarrow DTT=1 (totalmente directa)

FIG. 7 B
 (PROPORCIÓN DE ENERGÍA DIRECTA A TOTAL
 RESPECTO DE COHERENCIA INTERCANAL)

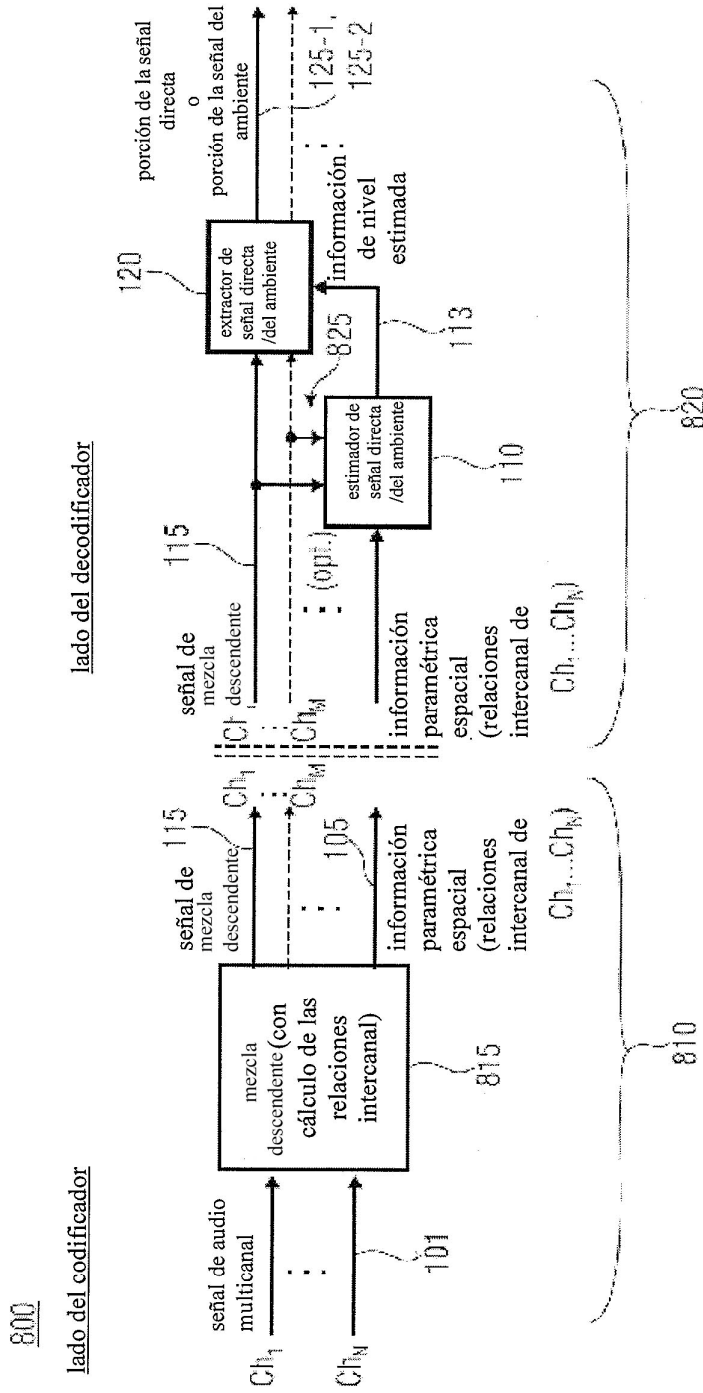


FIG. 8
(CODIFICADOR/DECODIFICADOR)

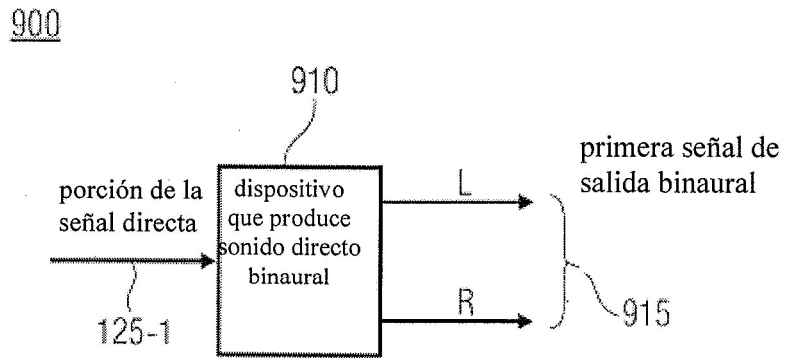
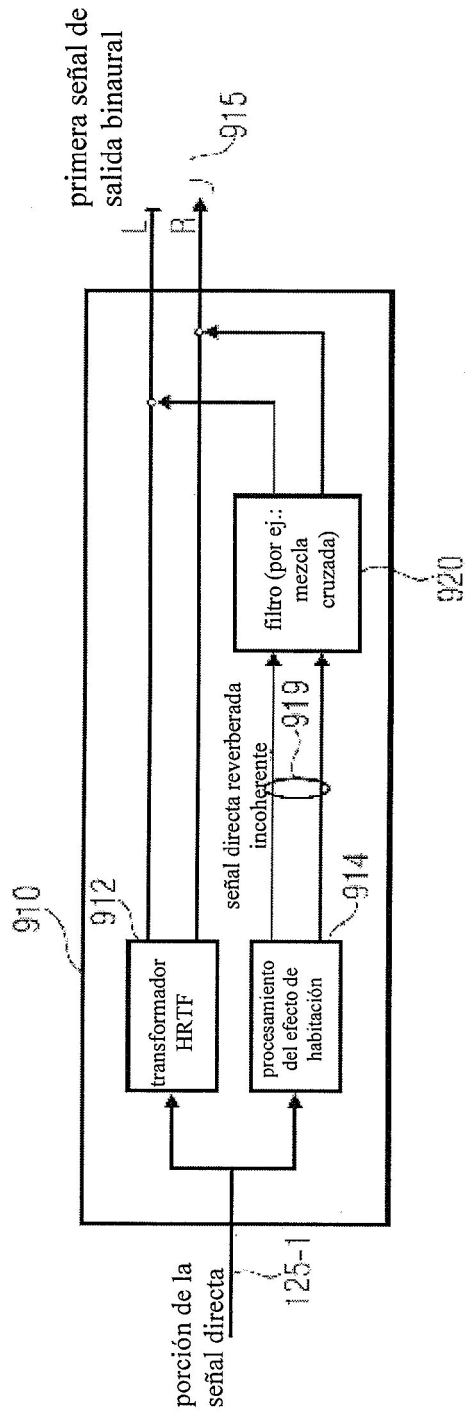


FIG. 9 A

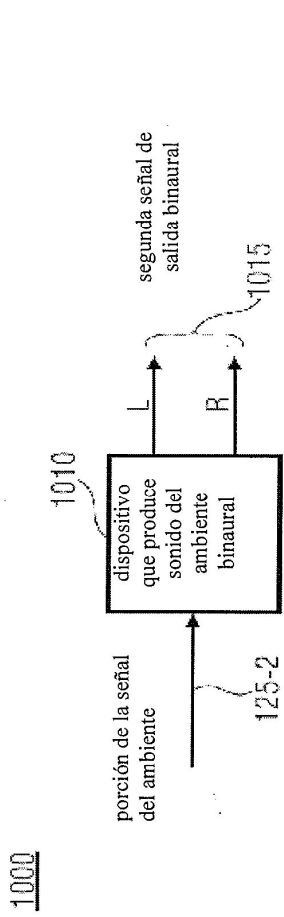
(VISTA GENERAL DE LA PRODUCCIÓN DE SONIDO DIRECTO BINAURAL)

905



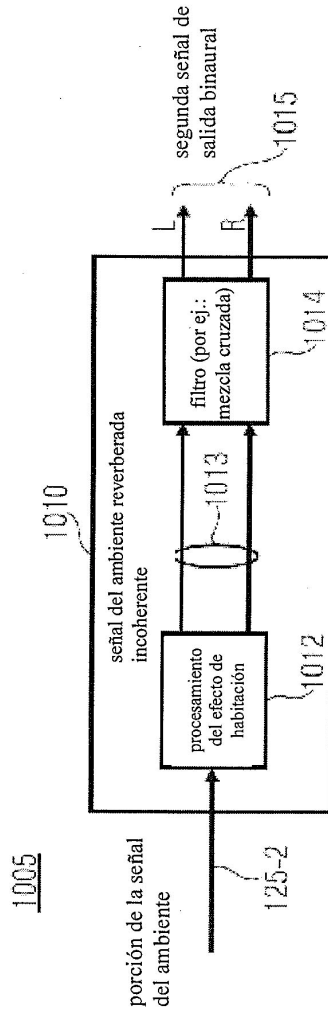
(DETALLES DE LA PRODUCCIÓN DE SONIDO DIRECTO BINAURAL)

FIG. 9 B



(VISTA GENERAL DE LA PRODUCCIÓN DE SONIDO DEL AMBIENTE BINAURAL)

FIG. 10 A



(DETALLES DE LA PRODUCCIÓN DE SONIDO DEL AMBIENTE BINAURAL)

FIG. 10 B

1100

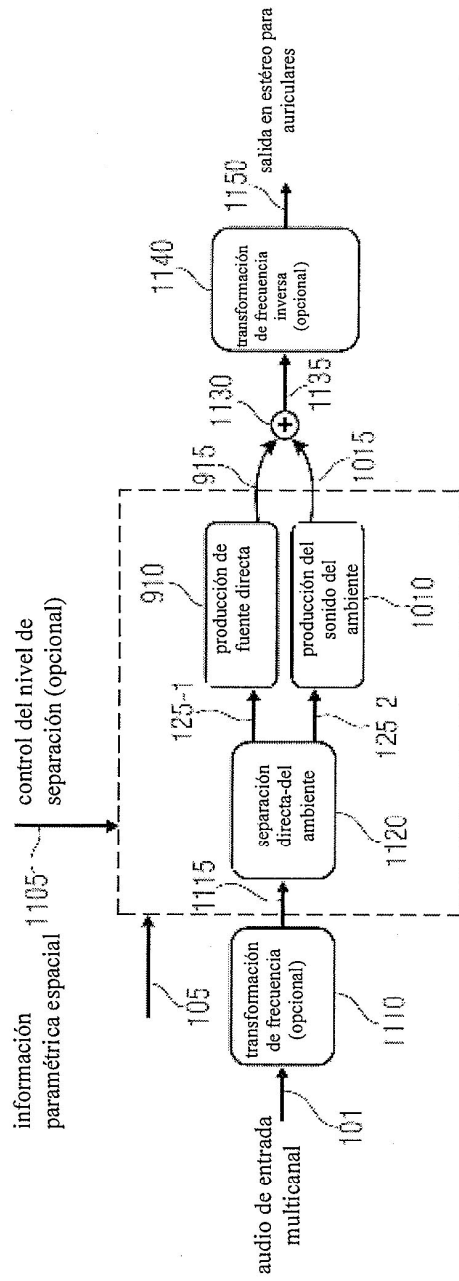
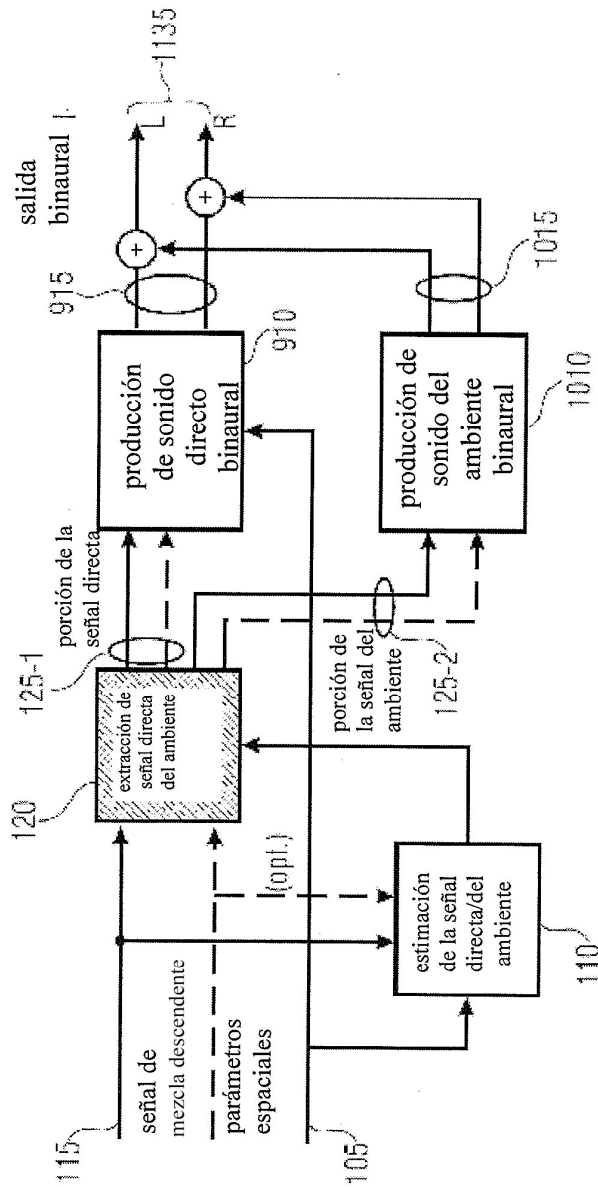


FIG. 11

(DIAGRAMA DE BLOQUES CONCEPTUAL DE LA REPRODUCCIÓN BINAURAL PROPUESTA)

1200



(DIAGRAMA DE BLOQUES GENERAL QUE INCLUYE EL USO DEL CASO DE REPRODUCCIÓN BINAURAL)

FIG. 12

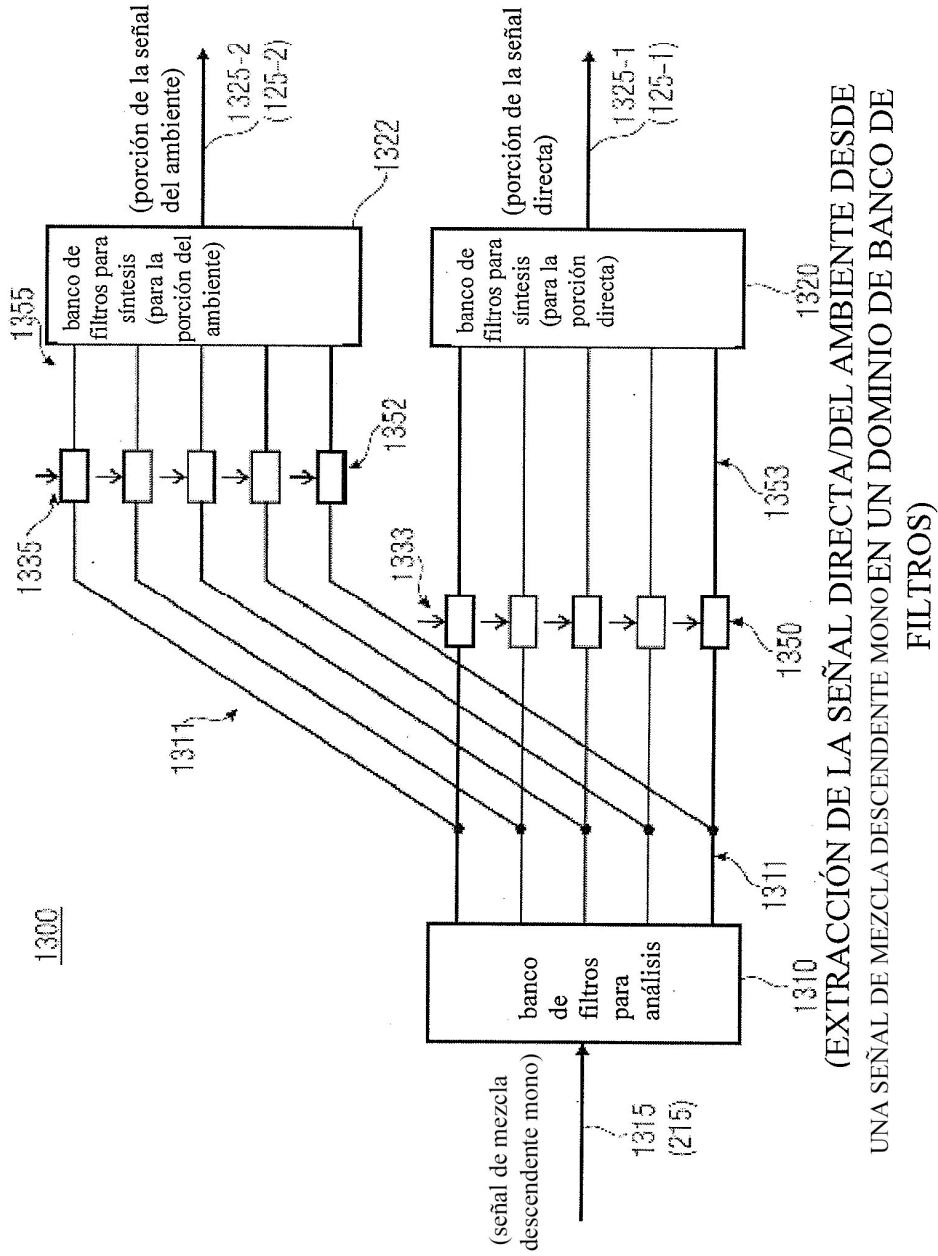
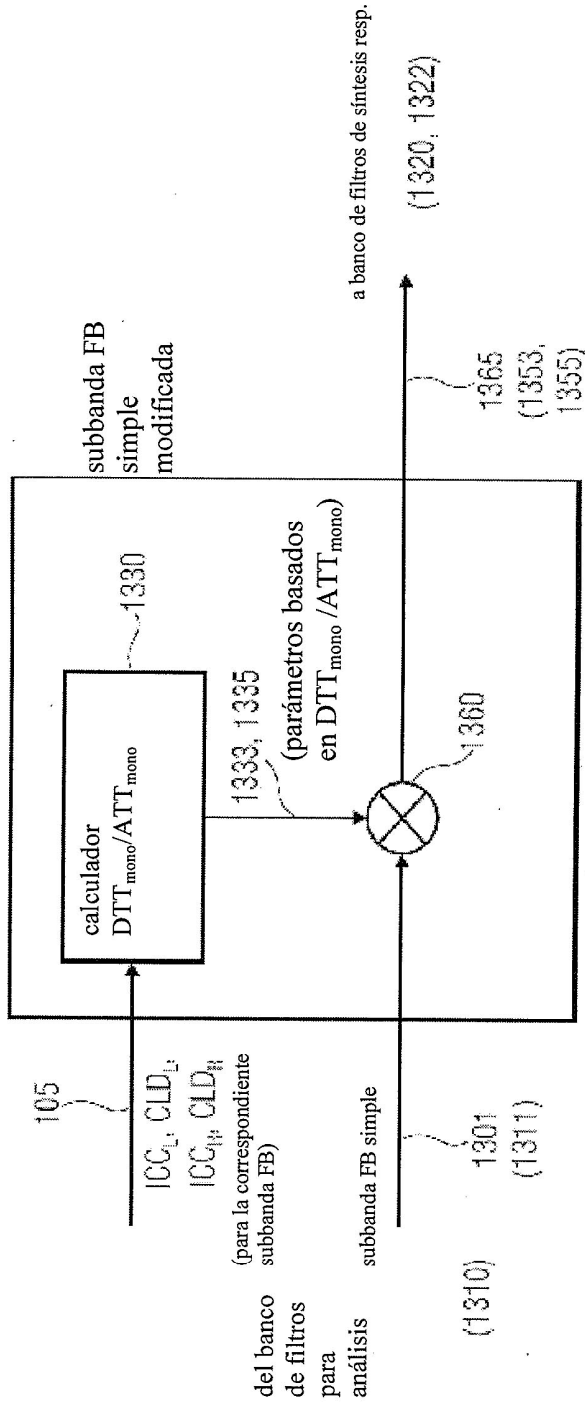


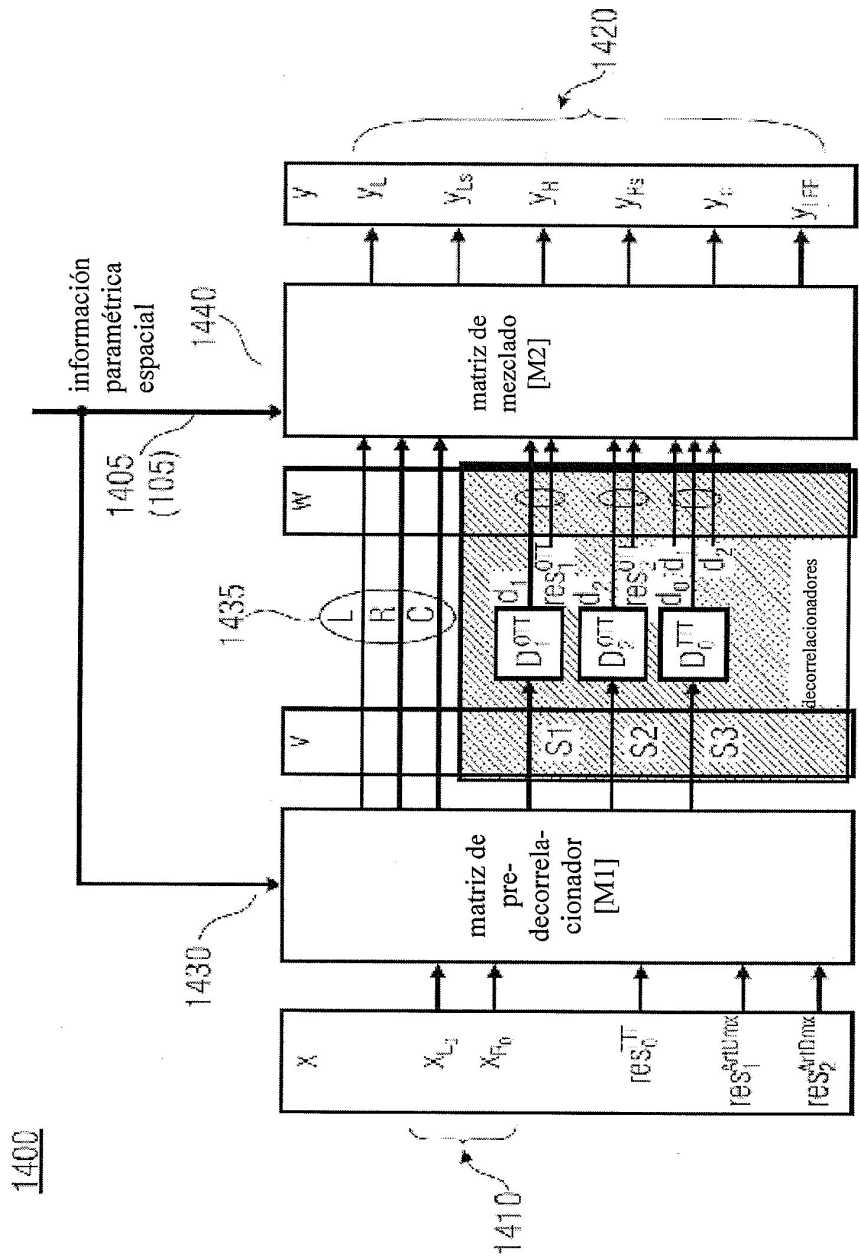
FIG. 13 A

1380



(BLOQUES DE EXTRACCIÓN DE LA SEÑAL DIRECTA/DEL AMBIENTE)

FIG. 13 B



(ESQUEMA DE DECODIFICACIÓN MPEG SURROUND)

FIG. 14