

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 797 742**

51 Int. Cl.:

G10L 19/008 (2013.01)

G10L 21/0264 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **27.07.2015 PCT/EP2015/067158**

87 Fecha y número de publicación internacional: **04.02.2016 WO16016189**

96 Fecha de presentación y número de la solicitud europea: **27.07.2015 E 15745433 (1)**

97 Fecha y número de publicación de la concesión europea: **15.04.2020 EP 3175445**

54 Título: **Aparato y método para mejorar un sistema mejorador de sonido de señal de audio**

30 Prioridad:

30.07.2014 EP 14179181

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

03.12.2020

73 Titular/es:

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

72 Inventor/es:

**UHLE, CHRISTIAN;
GAMPP, PATRICK;
HELLMUTH, OLIVER;
VARGA, STEFAN y
SCHARRER, SEBASTIAN**

74 Agente/Representante:

ARIZTI ACHA, Mónica

ES 2 797 742 T3

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

DESCRIPCIÓN

Aparato y método para mejorar un sistema mejorador de sonido de señal de audio

5 La presente solicitud se refiere a procesamiento de señales de audio y en particular a procesamiento de audio de una señal mono o dual-mono.

Se puede modelar una escena auditiva como una mezcla de sonidos directos y ambientales. Los sonidos directos (o direccionales) se emiten por fuentes de sonido por ejemplo, un instrumento musical, un vocalista o un altavoz y llegan en la ruta más corta posible al receptor, por ejemplo el oído del oyente o un micrófono. Cuando se captura un sonido directo utilizando un conjunto de micrófonos separados, las señales recibidas son coherentes. En contraste, los sonidos ambientales (o difusos) se emiten por muchas fuentes de sonido separadas o límites que reflejan sonido que contribuyen a, por ejemplo, reverberación de sala, aplausos o un ruido de voces. Cuando se captura un campo de sonido ambiental utilizando un conjunto de micrófonos separados, las señales recibidas son al menos parcialmente incoherentes.

Se puede considerar apropiada la reproducción de sonido monofónico en algunos escenarios de reproducción (por ejemplo clubes de baile) o algunos tipos de señales (por ejemplo grabaciones de voz), pero la mayoría de grabaciones musicales, sonido de películas y sonido de TV son señales estereofónicas. Las señales estereofónicas pueden crear la sensación de sonidos ambientales (o difusos) y de las direcciones y anchos de fuentes de sonido. Esto se logra por medio de información estereofónica que se codifica por patrones espaciales. Los patrones espaciales más importantes son diferencias de nivel inter-canal (ICLD), diferencias de tiempo inter-canal (ICTD) y coherencia inter-canal (ICC). En consecuencia, las señales estereofónicas y los sistemas de reproducción de sonido correspondientes tienen más de un canal. ICLD e ICTD contribuyen a la sensación de una dirección. La ICC evoca la sensación de ancho de un sonido y, en el caso de sonidos ambientales, que un sonido se percibe como que proviene de todas las direcciones.

Aunque existe reproducción de sonido multicanal en diferentes formatos, la mayoría de grabaciones de audio y sistemas de reproducción de sonido aún tienen dos canales. El sonido estereofónico de dos canales es la norma para sistemas de entretenimiento, y los oyentes lo utilizan. Sin embargo, las señales estereofónicas no se restringen a tener sólo dos señales de canal sino pueden tener más de una señal de canal. De forma similar, las señales monofónicas no se restringen a tener sólo una señal de canal, sino pueden tener múltiples señales de canal idénticas. Por ejemplo, una señal de audio que comprende dos señales de canal idénticas puede ser llamada una señal dual-mono.

El documento WO 01/05187 muestra un sistema para mejorar espacialmente señales dual-mono.

Existen diferentes razones por las cuales están disponibles al oyente las señales monofónicas en lugar de señales estereofónicas. Primero, las grabaciones antiguas son monofónicas debido a que no se utilizaron técnicas estereofónicas en ese momento. En segundo lugar, las restricciones del ancho de banda de una transmisión o medio de almacenamiento pueden conducir a una pérdida de información estereofónica. Un ejemplo prominente es radiodifusión que utiliza modulación de frecuencia (FM). Aquí, las fuentes de interferencia, las distorsiones de múltiples rutas u otros impedimentos de la transmisión pueden conducir a información estereofónica ruidosa, que para la transmisión de señales de dos canales convencionalmente se codifica como la señal de diferencia entre ambos canales. Es una práctica común descartar parcial o completamente la información estereofónica cuando son pobres las condiciones de recepción.

La pérdida de información estereofónica puede conducir a una reducción de la calidad de sonido. En general, una señal de audio que comprende un mayor número de canales puede comprender una mayor calidad de sonido cuando se compara con una señal de audio que comprende un número inferior de canales. Los oyentes pueden preferir escuchar señales de audio que comprenden una alta calidad de sonido. Por razones de eficiencia tal como velocidades de datos transmitidos a través o almacenados en medios a menudo se reduce la calidad de sonido.

Por lo tanto, existe una necesidad de incrementar (mejorar) la calidad de sonido de las señales de audio mono o tipo mono.

55 Un objeto de la presente invención, por lo tanto, es proporcionar un aparato o un método para una mejora de señales de audio y/o para incrementar la sensación de señales de audio reproducidas que son una señal mono o una señal tipo mono.

Este objeto se logra por un aparato para mejorar la señal de audio de acuerdo con la reivindicación 1, un método para mejorar una señal de audio de acuerdo con la reivindicación 14 y un sistema mejorador de sonido de acuerdo con la reivindicación 13 o un programa informático de acuerdo con la reivindicación 15.

La presente invención se basa en el hallazgo de que una señal de audio mono o tipo mono recibida se puede mejorar al generar de forma artificial patrones espaciales al dividir las señales de audio recibidas en al menos dos partes y al

ES 2 797 742 T3

decorrelacionar al menos una de las partes de la señal recibida. Una combinación ponderada de las partes permite recibir una señal de audio percibida como estereofónica y por lo tanto se mejora. El control de la ponderación aplicada permite un grado variante de decorrelación y por lo tanto un grado variante de mejora de tal forma que un nivel de mejora puede ser inferior cuando la decorrelación puede conducir a efectos molestos que reducen la calidad de sonido.

5 Por lo tanto, una señal de audio variante se puede mejorar que comprende porciones o intervalos de tiempo donde se aplica poca o ninguna decorrelación tal como para señales de voz y que comprende porciones o intervalos de tiempo donde se aplica mucha o un alto grado de decorrelación tal como para señales de música.

Una realización de la presente invención proporciona un aparato para mejorar una señal de audio que es una señal mono o una señal tipo mono. El aparato comprende un procesador de señales para procesar la señal de audio a fin de reducir o eliminar porciones transitorias y tonales de la señal procesada. El aparato comprende además un decorrelacionador para generar una primera señal decorrelacionada y una segunda señal decorrelacionada de la señal procesada. El aparato comprende además un combinador y un controlador. El combinador se configura para combinar de forma ponderada la primera señal decorrelacionada, la segunda señal decorrelacionada y la señal de audio o una

10 señal derivada de, la señal de audio por mejora de coherencia utilizando factores de ponderación variantes en el tiempo y para obtener una señal de audio de dos canales. El controlador se configura para controlar los factores de ponderación variantes en el tiempo al analizar la señal de audio de tal forma que diferentes porciones de la señal de audio se multiplican por diferentes factores de ponderación y la señal de audio de dos canales tiene un grado variante en el tiempo de decorrelación.

20 La señal de audio que tiene poca o ninguna información estereofónica (o multicanal), por ejemplo, una señal que tiene un canal o una señal que tiene múltiples pero señales de canal casi idénticas, se puede percibir como una señal multicanal, por ejemplo, una señal estereofónica, después de que se ha aplicado la mejora. Una señal de audio mono o dual-mono recibida se puede procesar de forma diferente en diferentes rutas, en donde en una ruta se reducen o eliminan porciones transitorias y/o tonales de la señal de audio. Una señal procesada de esta forma que se decorrelaciona y la señal decorrelacionada que se combina de factor de ponderación con la segunda ruta que comprende la señal de audio o una señal derivada de la misma permite la obtención de dos canales de señal que pueden comprender un alto factor de decorrelación uno con respecto al otro de tal forma que los dos canales se perciben como una señal estereofónica.

30 Al controlar los factores de ponderación utilizados para combinar de forma ponderada la señal decorrelacionada y la señal de audio (o la señal derivada de la misma) se puede obtener un grado variante en el tiempo de decorrelación de tal forma que en situaciones, en las cuales la mejora de la señal de audio conduciría posiblemente a efectos indeseados, la mejora se puede reducir o evitar. Por ejemplo, es indeseable que una señal de un altavoz de radio u otras señales de fuente de sonido prominente se mejoren ya que la percepción de un altavoz de múltiples ubicaciones de fuentes puede conducir a efectos molestos a un oyente.

40 De acuerdo con la realización, un aparato para mejorar una señal de audio comprende un procesador de señales para procesar la señal de audio a fin de reducir o eliminar las porciones transitorias y tonales de la señal procesada. El aparato comprende además un decorrelacionador, un combinador y un controlador. El decorrelacionador se configura para generar una primera señal decorrelacionada y una segunda señal decorrelacionada de la señal procesada. El combinador se configura para combinar de forma ponderada la primera señal decorrelacionada y la señal de audio o una señal derivada de la señal de audio por mejora de coherencia utilizando factores de ponderación variantes en el tiempo y para obtener una señal de audio de dos canales. El controlador se configura para controlar los factores de ponderación variantes en el tiempo al analizar la señal de audio de tal forma que diferentes porciones de la señal de audio se multiplican por diferentes factores de ponderación y la señal de audio de dos canales tiene un grado variante en el tiempo de decorrelación. Esto permite percibir una señal mono o una señal similar a una señal mono (tal como dual-mono o multi-mono) que es una señal de audio de canal estéreo.

50 Para procesamiento de la señal de audio, el controlador y/o el procesador de señales se puede configurar para procesar una representación de la señal de audio en el dominio de la frecuencia. La representación puede comprender una pluralidad o una multitud de bandas de frecuencia (sub-bandas) cada una que comprende una parte, es decir, una porción de la señal de audio del espectro de la señal de audio respectivamente. Para cada una de las bandas de frecuencia, el controlador se puede configurar para predecir un nivel percibido de decorrelación en la señal de audio de dos canales. El controlador se puede configurar además para incrementar los factores de ponderación para porciones (bandas de frecuencia) de la señal de audio que permite un mayor grado de decorrelación y para disminuir los factores de ponderación para porciones de la señal de audio que permite un menor grado de decorrelación. Por ejemplo, una porción que comprende una señal de fuente de sonido no prominente tal como aplausos o ruido de voces se puede combinar por un factor de ponderación que permite una mayor decorrelación que una porción que comprende una señal de fuente de sonido prominente, en donde la señal de fuente de sonido prominente de corta duración se utiliza para porciones de la señal que se perciben como sonidos directos, por ejemplo voz, un instrumento musical, un vocalista o un altavoz.

60 El procesador se puede configurar para determinar cada una de algunas o todas las bandas de frecuencia, si la banda

ES 2 797 742 T3

de frecuencia comprende componentes transitorios o tonales y para determinar factores de ponderación espectrales que permiten una reducción de las porciones transitorias o tonales. Los factores de pesos espectrales y los factores de modificación de escala cada uno puede comprender una multitud de valores posibles de tal forma que se pueden reducir y/o evitar efectos molestos debido a decisiones binarias.

5 El controlador se puede configurar además para modificar la escala de los factores de ponderación de tal forma que un nivel percibido de decorrelación en la señal de audio de dos canales permanece dentro de un intervalo alrededor de un valor objetivo. El intervalo puede extenderse, por ejemplo a $\pm 20\%$, $\pm 10\%$ o $\pm 5\%$ del valor objetivo. El valor objetivo puede ser, por ejemplo, un valor previamente determinado para una medición de la porción tonal y/o transitoria de tal forma que, por ejemplo, se obtiene la señal de audio que comprende porciones transitorias y tonales variables que modifican el valor objetivo. Esto permite llevar a cabo una poca o incluso ninguna decorrelación cuando se decorrelaciona la señal de audio o no se propone ninguna decorrelación tal como para señales de fuente de sonido prominentes como voz y para una alta decorrelación si no se decorrelaciona la señal y/o se propone la decorrelación. Los factores de ponderación y/o los valores espectrales se pueden determinar y/o ajustar a múltiples valores incluso casi de forma continua.

El decorrelacionador se puede configurar para generar la primera señal decorrelacionada con base en una reverberación o un retraso de la señal de audio. El controlador se puede configurar para generar la señal decorrelacionada de prueba también con base en una reverberación o un retraso de la señal de audio. Una reverberación se puede llevar a cabo al retrasar la señal de audio y al combinar la señal de audio y la versión retrasada de la misma similar a una estructura de filtro de respuesta finita al impulso, en donde la reverberación también se puede implementar como un filtro de respuesta infinita al impulso. Un tiempo de retraso y/o un número de retrasos y combinaciones pueden variar. Un tiempo de retraso que retrasa o reverbera la señal de audio para la señal decorrelacionada de prueba puede ser más corto que un tiempo de retraso, por ejemplo, que resulta de menos coeficientes de filtro del filtro de retraso, para retrasar o reverberar la señal de audio para la primera señal decorrelacionada. Para predecir la intensidad percibida de decorrelación, un menor grado de decorrelación y de esta forma un tiempo de retraso más corto puede ser suficiente de tal forma que al reducir el tiempo de retraso y/o los coeficientes de filtro se puede reducir un esfuerzo computacional y/o una potencia computacional.

30 Posteriormente, se describen realizaciones preferidas de la presente invención con respecto a las figuras anexas, en las cuales:

La Figura 1 muestra un diagrama de bloques esquemático de un aparato para mejorar una señal de audio;

35 La Figura 2 muestra un diagrama de bloques esquemático de un aparato adicional para mejorar la señal de audio;

La Figura 3 muestra una tabla de ejemplo que indica un cálculo de los factores de modificación de escala (factores de ponderación) con base en el nivel de la intensidad percibida, predicha de decorrelación;

40 La Figura 4a muestra un diagrama de flujo esquemático de una parte de un método que se puede ejecutar, para determinar parcialmente factores de ponderación;

La Figura 4b muestra un diagrama de flujo esquemático de pasos adicionales del método de la Figura 4a, que representa un caso, donde la medición del nivel percibido de decorrelación se compara con los valores umbrales;

45 La Figura 5 muestra un diagrama de bloques esquemático de un decorrelacionador que se puede configurar para operar como el decorrelacionador en la Figura 1;

50 La Figura 6a muestra un diagrama esquemático que comprende un espectro de una señal de audio que comprende al menos una porción de señal transitoria (tiempo corto);

La Figura 6b muestra un espectro esquemático de una señal de audio que comprende un componente tonal;

55 La Figura 7a muestra una tabla esquemática que ilustra un procesamiento de transitorios posible llevada a cabo por una etapa de procesamiento de transitorios;

La Figura 7b muestra una tabla de ejemplo que ilustra un procesamiento de tonales posible como que se puede ejecutar por una etapa de procesamiento de tonales.

60 La Figura 8 muestra un diagrama de bloques esquemático de un sistema mejorador de sonido que comprende un aparato para mejorar la señal de audio;

La Figura 9a muestra un diagrama de bloques esquemático de un procesamiento de la señal de entrada de acuerdo con el procesamiento de primer plano/fondo.

ES 2 797 742 T3

La Figura 9b ilustra la separación de la señal de entrada en una señal de primer plano y una de fondo;

5 La Figura 10 muestra un diagrama de bloques esquemático y también un aparato configurado para aplicar pesos espectrales a una señal de entrada;

La Figura 11 muestra un diagrama de flujo esquemático de un método para mejorar una señal de audio;

10 La Figura 12 ilustra un aparato para determinar una medición de un nivel percibido de reverberación/decorrelación en una señal de mezcla que comprende un componente de señal directa o componente de señal seca y un componente de señal de reverberación;

Las Figuras 13a-c muestran implementaciones de un procesador de modelo de intensidad; y

15 La Figura 14 ilustra en implementación del procesador de modelo de intensidad que ya se ha analizado en algunos aspectos con respecto a las Figuras 12, 13a, 13b, 13c.

20 Elementos iguales o equivalentes o elementos con la misma o una función equivalente se denotan en la siguiente descripción por números de referencia iguales o equivalentes incluso si se presentan en diferentes figuras.

25 En la siguiente descripción, se expone una pluralidad de detalles para proporcionar una explicación más extensa de las realizaciones de la presente invención. Sin embargo, será evidente para aquellos expertos en la técnica que se pueden practicar realizaciones de la presente invención sin estos detalles específicos. En otros casos, se muestran estructuras y dispositivos bien conocidos en forma de diagrama de bloques en lugar de en detalle a fin de evitar complicar las realizaciones de la presente invención. Además, las características de las diferentes realizaciones descritas en lo sucesivo se pueden combinar entre sí, a menos que se señale específicamente lo contrario.

30 En lo siguiente, se hará referencia a procesamiento de una señal de audio. Un aparato o un componente del mismo se pueden configurar para recibir, proporcionar, y/o procesar una señal de audio. La señal de audio respectiva se puede recibir, proporcionar o procesar en el dominio del tiempo y/o en el dominio de la frecuencia. Una representación de señal de audio en el dominio del tiempo se puede transformar en una representación de frecuencia de la señal de audio por ejemplo por transformadas de Fourier o similares. La representación de frecuencia se puede obtener, por ejemplo, al utilizar transformada corta de Fourier en el tiempo (STFT), una transformada discreta de coseno y/o una transformada rápida de Fourier (FFT). De forma alternativamente o además, la representación de frecuencia se puede obtener por un banco de filtros que puede comprender filtros de espejo en cuadratura (QMF). Una representación en el dominio de la frecuencia de la señal de audio puede comprender una pluralidad de cuadros cada uno que comprende una pluralidad de sub-bandas como se conoce de las transformadas de Fourier. Cada sub-banda comprende una porción de la señal de audio. Ya que la representación del tiempo y la representación de la frecuencia de la señal de audio se pueden convertir la una en la otra, la siguiente descripción no se debe limitar a la señal de audio que es la representación en el dominio de la frecuencia o la representación en el dominio del tiempo.

40 La Figura 1 muestra un diagrama de bloques esquemático de un aparato 10 para mejorar una señal de audio 102. La señal de audio 102, por ejemplo, es una señal mono o una señal tipo mono, tal como una señal dual-mono, representada en el dominio de la frecuencia o en el dominio del tiempo. El aparato 10 comprende un procesador de señales 110, un decorrelacionador 120, un controlador 130 y un combinador 140. El procesador de señales 110 se configura para recibir la señal de audio 102 y para procesar la señal de audio 102 para obtener una señal procesada 112 a fin de reducir o eliminar las porciones transitorias y tonales de la señal procesada 112 cuando se compara con la señal de audio 102.

50 El decorrelacionador 120 se configura para recibir la señal procesada 112 y para generar una primera señal decorrelacionada 122 y una segunda señal decorrelacionada 124 de la señal procesada 112. El decorrelacionador 120 se puede configurar para generar la primera señal decorrelacionada 122 y la segunda señal decorrelacionada 124 al menos parcialmente al reverberar la señal procesada 112. La primera señal decorrelacionada 122 y la segunda señal decorrelacionada 124 pueden comprender diferentes retrasos de tiempo para la reverberación de tal forma que la primera señal decorrelacionada 122 comprende un retraso de tiempo más corto o más largo (tiempo de reverberación) que la segunda señal decorrelacionada 124. La primera o segunda señal decorrelacionada 122 o 124 también se pueden procesar sin un filtro de retraso o reverberación.

60 El decorrelacionador 120 se configura para proporcionar la primera señal decorrelacionada 122 y la segunda señal decorrelacionada 124 al combinador 140. El controlador 130 se configura para recibir la señal de audio 102 y para controlar los factores de ponderación variantes en el tiempo a y b al analizar la señal de audio 102 de tal forma que se multiplican diferentes porciones de la señal de audio 102 por diferentes factores de ponderación a o b. Por lo tanto, el controlador 130 comprende una unidad controladora 132 configurada para determinar los factores de ponderación a y b. El controlador 130 se puede configurar para operar en el dominio de la frecuencia. La unidad controladora 132 se

ES 2 797 742 T3

- puede configurar para transformar la señal de audio 102 en el dominio de la frecuencia al utilizar una transformada corta de Fourier en el tiempo (STFT), una transformada rápida de Fourier (FFT) y/o una transformada de Fourier regular (FT). Una representación en el dominio de la frecuencia de la señal de audio 102 puede comprender una pluralidad de sub-bandas como se conoce de las transformadas de Fourier. Cada sub-banda comprende una porción de la señal de audio. De forma alternativamente, la señal de audio 102 puede ser una representación de una señal en el dominio de la frecuencia. La unidad controladora 132 se puede configurar para controlar y/o determinar un par de factores de ponderación a y b para cada sub-banda de la representación digital de la señal de audio.
- 5
- El combinador se configura para combinar de forma ponderada la primera señal decorrelacionada 122, la segunda señal decorrelacionada 124, una señal 136 derivada de la señal de audio 102 utilizando los factores de ponderación a y b. la señal 136 derivada de la señal de audio 102 se puede proporcionar por el controlador 130. Por lo tanto, el controlador 130 puede comprender una unidad de derivación opcional 134. La unidad de derivación 134 se puede configurar, por ejemplo, para adaptar, modificar o mejorar porciones de la señal de audio 102. En particular, la unidad de derivación 110 se puede configurar para amplificar porciones de la señal de audio 102 que se atenúan, reducen o eliminan por el procesador de señales 110.
- 10
- 15
- El procesador de señales 110 se puede configurar para operar también en el dominio de la frecuencia y para procesar la señal de audio 102 de tal forma que el procesador de señales 110 reduce o elimina porciones transitorias y tonales para cada sub-banda de un espectro de la señal de audio 102. Esto puede conducir a menos o incluso ningún procesamiento para sub-bandas que comprenden porciones con poco o ningún transitorio o con pocos o ningún tonal (es decir ruidosas). De forma alternativa, el combinador 140 puede recibir la señal de audio 102 en lugar de la señal derivada, es decir, el controlador 130 se puede implementar sin la unidad de derivación 134. Entonces, la señal 136 puede ser igual a la señal de audio 102.
- 20
- Entonces el combinador 140 se configura para recibir una señal de ponderación 138 que comprende los factores de ponderación a y b. El combinador 140 se configura además para tener una señal de audio de salida 142 que comprende un primer canal y_1 y un segundo canal y_2 , es decir, la señal de audio 142 es una señal de audio de dos canales.
- 25
- El procesador de señales 110, el decorrelacionador 120, el controlador 130 y el combinador 140 se pueden configurar para procesar la señal de audio 102, la señal 136 derivada de la misma y/o señales procesadas 112, 122 y/o 124 por cuadros y por sub-banda de tal forma que el procesador de señales 110, el decorrelacionador 120, el combinador 130 y el combinador 140 se pueden configurar para ejecutar operaciones descritas anteriormente a cada sub-banda de frecuencia al procesar una o más bandas de frecuencia (porciones de la señal) a la vez.
- 30
- 35
- La Figura 2 muestra un diagrama de bloques esquemático de un aparato 200 para mejorar la señal de audio 102. El aparato 200 comprende un procesador de señales 210, el decorrelacionador 120, un controlador 230 y un combinador 240. El decorrelacionador 120 se configura para generar la primera señal decorrelacionada 122 indicada como r1 y la segunda señal decorrelacionada 124, indicada como r2.
- 40
- El procesador de señales 210 comprende una etapa de procesamiento de transitorios 211, una etapa de procesamiento de tonales 213 y una etapa de combinación 215. El procesador de señales 210 se configura para procesar una representación de la señal de audio 102 en el dominio de la frecuencia. La representación en el dominio de la frecuencia de la señal de audio 102 comprende una multitud de sub-bandas (bandas de frecuencia), donde la etapa de procesamiento de transitorios 211 y la etapa de procesamiento de tonales 213 se configuran para procesar cada una de las bandas de frecuencia. De forma alternativa, el espectro obtenido por conversión de frecuencia de la señal de audio 102 se puede reducir, es decir, cortar, para excluir ciertos intervalos de frecuencia o bandas de frecuencia de procesamiento adicional, tal como bandas de frecuencia por debajo de 20 Hz, 50 Hz o 100 Hz y/o por encima de 16 kHz, 18 kHz o 22 kHz. Esto puede permitir un esfuerzo computacional reducido y de esta forma un procesamiento más rápido y/o más preciso.
- 45
- 50
- La etapa de procesamiento de transitorios 211 se configura para determinar para cada una de las sub-bandas de frecuencia procesadas, si la banda de frecuencia comprende porciones transitorias. La etapa de procesamiento de tonales 213 se configura para determinar para cada una de las bandas de frecuencia, si la señal de audio 102 comprende porciones tonales en la banda de frecuencia. La etapa de procesamiento de transitorios 211 se configura para determinar al menos para las bandas de frecuencia que comprenden porciones transitorias factores de ponderación espectrales 217, en donde los factores de ponderación espectrales 217 se asocian con la banda de frecuencia respectiva. Como se describirá en las Figuras 6a y 6b, se pueden identificar características transitorias y tonales por procesamiento espectral. Se puede medir un nivel de transitoriedad, y/o tonalidad por la etapa de procesamiento de transitorios 211 y/o la etapa de procesamiento de tonales 213 y convertir a un coeficiente de ponderación espectral. La etapa de procesamiento de tonales 213 se configura para determinar factores de ponderación espectrales 219 al menos para bandas de frecuencia que comprenden las porciones tonales. Los factores de ponderación espectrales 217 y 219 pueden comprender una multitud de valores posibles, la magnitud de los factores de ponderación espectrales 217 y/o 219 que indican una cantidad de porciones transitorias y/o tonales en la
- 55
- 60

ES 2 797 742 T3

banda de frecuencia.

Los factores de ponderación espectrales 217 y 219 pueden comprender un valor absoluto o relativo. Por ejemplo, el valor absoluto puede comprender un valor de energía de sonido transitorio y/o tonal en la banda de frecuencia. De forma alternativa, los factores de ponderación espectrales 217 y/o 219 pueden comprender el valor relativo tal como un valor entre 0 y 1, el valor 0 que indica que la banda de frecuencia comprende ninguna o casi ninguna porción transitoria tonal y el valor 1 que indica que la banda de frecuencia comprende una cantidad alta o porciones completamente transitorias y/o tonales. Los factores de ponderación espectrales pueden comprender una de una multitud de valores tal como un número de 3, 5, 10 o más valores (pasos), por ejemplo, (0, 0.3 y 1), (0.1, 0.2, ..., 1) o similares. Un tamaño de la escala, un número de pasos entre un valor mínimo y un valor máximo pueden ser al menos cero pero de manera preferente al menos uno y de manera más preferente al menos cinco. De manera preferente, la multitud de valores de los pesos espectrales 217 y 219 comprende al menos tres valores que comprenden un valor mínimo, un valor máximo y un valor que está entre el valor mínimo y el valor máximo. Un número mayor de valores entre el valor mínimo y el valor máximo puede permitir una ponderación más continua de cada una de las bandas de frecuencia. El valor máximo y el valor mínimo se pueden modificar en escala a una escala entre 0 y 1 u otros valores. El valor máximo puede indicar un nivel más alto o más bajo de transitoriedad y/o tonalidad.

La etapa de combinación 215 se configura para combinar los pesos espectrales para cada una de las bandas de frecuencia como se describe posteriormente. El procesador de señales 210 se configura para aplicar los pesos espectrales combinados a cada una de las bandas de frecuencia. Por ejemplo los pesos espectrales 217 y/o 219 a un valor derivado de los mismos se puede multiplicar con valores espectrales de la señal de audio 102 en la banda de frecuencia procesada.

El controlador 230 se configura para recibir los factores de ponderación espectrales 217 y 219 o información relacionada con los mismos del procesador de señales 210. La información derivada puede ser, por ejemplo, un número de índice de una tabla, el número de índice que se asocia con los factores de ponderación espectrales. El controlador se configura para mejorar la señal de audio 102 para porciones de señal coherente, es decir, para porciones sin reducir o solamente parcialmente reducidas o eliminadas por la etapa de procesamiento de transitorios 211 y/o la etapa de procesamiento de tonales 213. En términos simples, la unidad de derivación 234 puede amplificar porciones no reducidas o eliminadas por el procesador de señales 210.

La unidad de derivación 234 se configura para proporcionar una señal 236 derivada de la señal de audio 102, indicada como z. El combinador 240 se configura para recibir la señal z (236). El decorrelacionador 120 se configura para recibir una señal procesada 212 indicada como s del procesador de señales 210.

El combinador 240 se configura para combinar las señales decorrelacionadas r1 y r2 con los factores de ponderación (factores de modificación de escala) a y b, para obtener una primera señal de canal y1 y una segunda señal de canal y2. Las señales de canal y1 y y2 se pueden combinar a la señal de salida 242 o se pueden sacar de forma separada.

En otras palabras, la señal de salida 242 es una combinación de una señal correlacionada (convencionalmente) z (236) y una señal decorrelacionada s (r1 o r2, respectivamente). La señal decorrelacionada como se obtiene en dos pasos, primero suprimiendo (reduciendo o eliminando) los componentes transitorios y tonales de la señal y segundo la decorrelación. La supresión de los componentes transitorios de la señal y de los componentes tonales de la señal se realiza por medio de ponderación espectral. La señal se procesa por cuadros en el dominio de la frecuencia. Los pesos espectrales se calculan para cada intervalo de frecuencia (banda de frecuencia) y periodo de tiempo. Por lo tanto la señal de audio se procesa en la banda completa, es decir se procesan todas las porciones que se van a considerar.

La señal de entrada del procesamiento puede ser una señal de canal individual x (102), la señal de salida puede ser una señal de dos canales y = [y1,y2], donde los índices denotan el primer y segundo canal, por ejemplo, el canal izquierdo y el derecho de una señal estéreo. La señal de salida y se puede calcular al combinar de forma lineal una señal de dos canales r = [r1,r2], con una señal de canal individual z con factores de modificación de escala a y b de acuerdo con

$$y1 = a \times z + b \times r1 \quad (1)$$

$$y2 = a \times z + b \times r2 \quad (2)$$

donde "x" se refiere a la operación de multiplicación en las ecuaciones (1) y (2).

Las ecuaciones (1) y (2) se deben interpretar de forma cualitativa, que indica que una parte de las señales z, r1 y r2 se puede controlar (variar) al variar los factores de ponderación. Al formar, por ejemplo, operaciones inversas tal como división por el valor recíproco se pueden obtener los mismos resultados o resultados equivalentes al llevar a cabo diferentes operaciones. De forma alternativa o además, una tabla de búsqueda que comprende los factores de

ES 2 797 742 T3

modificación de escala a y b y/o valores para y1 y/o y2 se puede utilizar para obtener la señal de dos canales y.

5 Los factores de modificación de escala a y/o b se pueden calcular para disminuir de forma monótona con la intensidad percibida de la correlación. El valor escalar predicho para la intensidad percibida se puede utilizar para controlar los factores de modificación de escala.

10 La señal decorrelacionada r que comprende r1 y r2 se puede calcular en dos pasos. Primero, la atenuación de componentes de la señal transitorios y tonales que producen la señal s. Segundo, se puede llevar a cabo la decorrelación de la señal s.

15 La atenuación de los componentes transitorios de la señal y de los componentes tonales de la señal se hace, por ejemplo, por medio de una ponderación espectral. La señal se procesa por cuadros en el dominio de la frecuencia. Se calculan los pesos espectrales para cada intervalo de frecuencia y periodo de tiempo. Un objetivo de la atenuación es doble:

1. Los componentes transitorios o tonales de la señal pertenecen convencionalmente a las llamadas señales de primer plano y como tal su posición dentro de la imagen estéreo a menudo está en el centro.

20 2. La decorrelación de las señales que tienen intensos componentes transitorios de la señal conducen a distorsiones perceptibles. La decorrelación de las señales que tienen intensos componentes tonales de la señal también conduce a distorsiones perceptibles cuando los componentes tonales (es decir sinusoidales) se modulan en frecuencia al menos cuando la modelación en frecuencia es suficientemente lenta para que se perciba como un cambio de la frecuencia y no como un cambio del timbre debido al enriquecimiento de los armónicos del espectro de la señal (posiblemente inarmónicos).

25 La señal correlacionada z se puede obtener al aplicar un procesamiento que mejora los componentes transitorios y tonales de la señal, por ejemplo, de forma cualitativa el inverso de la supresión para calcular la señal s. De forma alternativa, la señal de entrada, por ejemplo, sin procesar, se puede utilizar como tal. Se señala que puede existir el caso donde z también es una señal de dos canales. En realidad, muchos medios de almacenamiento (por ejemplo el disco compacto) utilizan dos canales incluso si la señal es mono. Una señal que tiene dos canales idénticos es llamada "dual-mono". También puede existir el caso donde la señal de entrada z es una señal estéreo, y el objetivo de procesamiento puede ser incrementar el efecto estereofónico.

30 La intensidad percibida de la decorrelación se puede predecir de forma similar a una intensidad percibida, predicha de reverberación tardía utilizando modelos computacionales de intensidad, como se describe en la EP 2 541 542 A1.

La Figura 3 muestra una tabla de ejemplo que indica un cálculo de los factores de modificación de escala (factores de ponderación) a y b con base en el nivel de la intensidad percibida, predicha de decorrelación.

40 Por ejemplo, la intensidad percibida de decorrelación se puede predecir de tal forma que un valor de la misma comprende un valor escalar que puede variar entre un valor de 0, que indica un bajo nivel de decorrelación percibida, ninguna respectivamente y un valor de 10, que indica un alto nivel de decorrelación. Los niveles se pueden determinar, por ejemplo, con base en pruebas de oyentes o simulación predictiva. De forma alternativa, el valor de nivel de decorrelación puede comprender un intervalo entre un valor mínimo y un valor máximo. El valor del nivel percibido de decorrelación se puede configurar para aceptar más del valor mínimo y el valor máximo. De manera preferente, el nivel percibido de la correlación puede aceptar al menos tres valores diferentes y de manera más preferente al menos siete valores diferentes.

50 Los factores de ponderación a y b que se van a aplicar con base en un nivel determinado de decorrelación percibida se pueden almacenar en una memoria y pueden ser accesibles al controlador 130 o 230. Con niveles crecientes de decorrelación percibida también puede incrementar el valor de escala a que se va a multiplicar con la señal de audio a la señal derivada de la misma por el combinador. Un nivel incrementado de decorrelación percibida se puede interpretar como "ya se decorrelacionó la señal (parcialmente)" de tal forma que con niveles crecientes de decorrelación la señal de audio a la señal derivada de la misma comprende una mayor parte en la señal de salida 142 o 242. Con niveles incrementados de decorrelación, el factor de ponderación b se configura para que se disminuye, es decir, las señales r1 y r2 generadas por el decorrelacionador con base en una señal de salida del procesador de señales puede comprender una menor parte cuando se combina en el combinador 140 o 240.

55 Aunque el factor de ponderación a se representa como que comprende un valor escalar de al menos 1 (valor mínimo) y a lo mucho 9 (valor máximo). Aunque el factor de ponderación b se representa como que comprende un valor escalar en el intervalo que comprende un valor mínimo de 2 y un valor máximo de 8, ambos factores de ponderación a y b pueden comprender un valor dentro de un intervalo que comprende un valor mínimo y un valor máximo y de manera preferente al menos un valor entre el valor mínimo y el valor máximo. De forma alternativa a los valores de los factores de ponderación a y b representados en la Figura 3 y con un nivel incrementado de decorrelación percibida, el factor

ES 2 797 742 T3

- de ponderación a puede incrementarse de forma lineal. De forma alternativa o además, el factor de ponderación b puede disminuir de forma lineal con un nivel incrementado de decorrelación percibida. Además, para un nivel de decorrelación percibida, una suma de los factores de ponderación a y b determinada para un cuadro puede ser constante o casi constante. Por ejemplo, el factor de ponderación a puede incrementar de 0 a 10 y el factor de ponderación b puede disminuir de un valor de 10 a un valor de 0 con un nivel creciente de decorrelación percibida. Si ambos factores de ponderación disminuyen o incrementan de forma lineal, por ejemplo con el valor de incremento 1, la suma de los factores de ponderación a y b puede comprender un valor de 10 para cada nivel de decorrelación percibida. Los factores de ponderación a y b que se van a aplicar se pueden determinar por simulación o por experimento.
- La Figura 4a muestra un diagrama de flujo esquemático de una parte de un método 400 que se puede ejecutar, por ejemplo, por el controlador 130 y/o 230. El controlador se configura para determinar una medición para el nivel percibido de una decorrelación en un paso 410 que produce, por ejemplo, en un valor escalar como se representa en la Figura 3. En un paso 420, el controlador se configura para comparar la medición determinada con un valor umbral. Si la medición es mayor que el valor umbral, el controlador se configura para modificar o adaptar los factores de ponderación a y/o b en un paso 430. En el paso 430, el controlador se configura para disminuir el factor de ponderación b, para incrementar el factor de ponderación a o para disminuir el factor de ponderación b y para incrementar el factor de ponderación a con respecto a un valor de referencia para a y b. El umbral puede variar, por ejemplo, dentro de bandas de frecuencia de la señal de audio. Por ejemplo, el umbral puede comprender un bajo valor para bandas de frecuencia que comprende una señal de fuente de sonido prominente que indica que se prefiere o propone un bajo nivel de decorrelación. De forma alternativa o además, el umbral puede comprender un alto valor para bandas de frecuencia que comprenden una señal de fuente de sonido no prominente que indica que se prefiere un alto nivel de decorrelación.
- Puede ser un objetivo incrementar la correlación de bandas de frecuencia que comprenden señales de fuente de sonido no prominentes y limitar la decorrelación para bandas de frecuencia que comprenden señales de fuente de sonido prominentes. Un umbral puede ser, por ejemplo, 20%, 50% o 70% de un intervalo de valores que pueden aceptar los factores de ponderación a y/o b. Por ejemplo, y con referencia a la Figura 3, el valor umbral puede ser menor que 7, menor que 5 o menor que 3 para un intervalo de frecuencia que comprende una señal de fuente de sonido prominente. Si el nivel percibido de decorrelación es demasiado alto, entonces, al ejecutar el paso 430, el nivel percibido de decorrelación se puede disminuir. Los factores de ponderación a y b se pueden variar por separados o ambos a la vez. La tabla representada en la Figura 3 puede ser, por ejemplo, un valor que comprende valores iniciales para los factores de ponderación a y/o b, los valores iniciales que se van a adaptar por el controlador.
- La Figura 4b muestra un diagrama de flujo esquemático de pasos adicionales del método 400, que representa un caso, donde la medición del nivel percibido de decorrelación (determinada en el paso 410) se compara con los valores umbrales, en donde la medición es inferior que el valor umbral (paso 440). El controlador se configura para incrementar b, para disminuir a o para incrementar b y disminuir a con respecto a una referencia para a y b para incrementar el nivel percibido de decorrelación y de tal forma que la medición comprende un valor que es al menos el valor umbral.
- De forma alternativa o adicionalmente, el controlador se puede configurar para modificar la escala de los factores de ponderación a y b de tal forma que un nivel percibido de decorrelación en la señal de audio de dos canales permanece dentro de un intervalo alrededor de un valor objetivo. El valor objetivo puede ser, por ejemplo, el valor umbral, en donde el valor de umbral puede variar con base en el tipo de señal que está comprendida por la banda de frecuencia para la cual se determinan los factores de ponderación y/o los pesos espectrales. El intervalo alrededor del valor objetivo puede extenderse a $\pm 20\%$, $\pm 10\%$, o $\pm 5\%$ del valor objetivo. Se puede permitir detener la adaptación de los factores de ponderación cuando la decorrelación percibida es aproximadamente el valor objetivo (umbral).
- La Figura 5 muestra un diagrama de bloques esquemático de un decorrelacionador 520 que se puede configurar para operar como el decorrelacionador 120. El decorrelacionador 520 comprende un primer filtro decorrelacionador 522 y un segundo filtro decorrelacionador 524. El primer filtro decorrelacionador 526 y el segundo filtro decorrelacionador 528 se configuran ambos para recibir la señal procesada s (512), por ejemplo, del procesador de señales. El decorrelacionador 520 se configura para combinar la señal procesada 512 y una señal de salida 523 del primer filtro decorrelacionador 526 para obtener la primera señal decorrelacionada 522 (r1) y para combinar una señal de salida 525 del segundo filtro decorrelacionador 528 para obtener la segunda señal decorrelacionada 524 (r2). Para la combinación de las señales, el decorrelacionador 520 se puede configurar para aplicar convolución de las señales con respuestas de impulso y/o para multiplicar valores espectrales con valores reales y/o imaginarios. De forma alternativa o además, se pueden ejecutar otras operaciones tal como divisiones, sumas, diferencias o similares.
- Los filtros decorrelacionadores 526 y 528 se pueden configurar para reverberar o retrasar la señal procesada 512. Los filtros decorrelacionadores 526 y 528 pueden comprender un filtro de respuesta finita al impulso (FIR) y/o filtro de respuesta infinita al impulso (IIR). Por ejemplo, los filtros decorrelacionadores 526 y 528 se pueden configurar para aplicar convolución a la señal procesada 512 con una respuesta de impulso obtenida de una señal de ruido que decae o decae exponencialmente con el paso del tiempo y/o frecuencia. Esto permite generar una señal decorrelacionada

ES 2 797 742 T3

- 523 y/o 525 que comprende una reverberación con respecto a la señal 512. Un tiempo de reverberación de la señal de reverberación puede comprender, por ejemplo, un valor entre 50 y 1000 ms, entre 80 y 500 ms y/o entre 120 y 200 ms. El tiempo de reverberación se puede entender como la duración que toma para que la potencia de la reverberación decaiga a un valor pequeño después de que se había excitado por un impulso, por ejemplo decaiga a 60 dB por debajo de la potencia inicial. De manera preferente, los filtros decorrelacionadores 526 y 528 comprenden filtros IIR. Esto permite reducir una cantidad de cálculo cuando al menos algunos de los coeficientes de filtro se ajustan a cero de tal forma que se pueden evitar cálculos para este coeficiente de filtro (cero). Opcionalmente, un filtro decorrelacionador puede comprender más de un filtro, donde los filtros se conectan en serie y/o en paralelo.
- 5
- 10 En otras palabras, la reverberación comprende un efecto de decorrelación. El decorrelacionador se puede configurar no sólo para decorrelacionar, sino también para sólo cambiar ligeramente la sonoridad. Técnicamente, la reverberación se puede considerar como un sistema invariante en el tiempo, lineal (LTI) que se puede caracterizar considerando su respuesta de impulso. Una longitud de la respuesta de impulso a menudo se indica como RT60 para reverberación. Es decir el tiempo después del cual la respuesta de impulso disminuye por 60 dB. La reverberación puede tener una
- 15 longitud de hasta un segundo o incluso hasta unos segundos. El decorrelacionador se puede implementar que comprende una estructura similar como reverberación pero que comprende diferentes ajustes para parámetros que influyen en la longitud de la respuesta de impulso.
- La Figura 6a muestra un diagrama esquemático que comprende un espectro de una señal de audio 602a que comprende al menos una porción de señal transitoria (tiempo corto). Una porción de señal transitoria conduce a un espectro de banda ancha. El espectro se representa como magnitudes $S(f)$ sobre frecuencias f , donde el espectro se sub-divide en una multitud de bandas de frecuencia $b1-3$. La porción de señal transitoria se puede determinar en una o más de las bandas de frecuencia en $b1-3$.
- 20
- 25 La Figura 6b muestra un espectro esquemático de una señal de audio 602b que comprende un componente tonal. Un ejemplo de un espectro se representa en siete bandas de frecuencia $fb1-7$. La banda de frecuencia $fb4$ se arregla en el centro de las bandas de frecuencia $fb1-7$ y comprende una magnitud máxima $S(f)$ cuando se compara con las otras bandas de frecuencia $fb1-3$ y $fb5-7$. Las bandas de frecuencia con distancia creciente con respecto a la frecuencia central (banda de frecuencia $fb5$) comprenden repeticiones armónicas de la señal tonal con magnitudes decrecientes.
- 30 El procesador de señales se puede configurar para determinar el componente tonal, por ejemplo, al evaluar la magnitud $S(f)$. Una magnitud creciente $S(f)$ de un componente tonal se puede incorporar por el procesador de señales por factores de ponderación espectrales, disminuidos. Por lo tanto, entre mayor sea una parte de los componentes transitorios y/o tonales dentro de una banda de frecuencia, menor será la contribución de la banda de frecuencia que puede tener en la señal procesada del procesador de señales. Por ejemplo, el coeficiente de ponderación espectral para la banda de frecuencia $fb4$ puede comprender un valor de cero o cerca de cero u otro valor que indica que la banda de frecuencia $fb4$ se considera con una parte reducida.
- 35
- La Figura 7a muestra una tabla esquemática que ilustra un procesamiento de transitorios posible 211 llevado a cabo por un procesador de señales tal como el procesador de señales 110 y/o 210. El procesador de señales se configura para determinar una cantidad, por ejemplo, una parte, de componentes transitorios en cada una de las bandas de frecuencia de la representación de la señal de audio en el dominio de la frecuencia que se va a considerar. Una evaluación puede comprender una determinación de una cantidad de los componentes transitorios con un valor inicial que comprende al menos un valor mínimo (por ejemplo 1) y a lo mucho un valor máximo (por ejemplo 15), en donde un mayor valor puede indicar una mayor cantidad de componentes transitorios dentro de la banda de frecuencia. Entre mayor sea la cantidad de componentes transitorios en la banda de frecuencia, menor puede ser el coeficiente de ponderación espectral respectivo, por ejemplo el coeficiente de ponderación espectral 217. Por ejemplo, el coeficiente de ponderación espectral puede comprender un valor de al menos un valor mínimo tal como 0 y de a lo mucho un valor máximo tal como 1. El coeficiente de ponderación espectral puede comprender una pluralidad de valores entre el valor mínimo y el valor máximo, en donde el coeficiente de ponderación espectral puede indicar un factor de consideración y/o un factor de consideración de la banda de frecuencia para procesamiento posterior. Por ejemplo, un coeficiente de ponderación espectral de 0 puede indicar que la banda de frecuencia se va a atenuar completamente. De forma alternativa, se pueden implementar otros intervalos de modificación de escala, es decir, la tabla representada en la Figura 7a se puede modificar en escala y/o transformar a tablas con otros valores de incremento con respecto a una evaluación de la banda de frecuencia que es una banda de frecuencia transitoria y/o un valor de incremento del
- 40
- 45
- 50
- 55
- La Figura 7b muestra una tabla de ejemplo que ilustra un procesamiento tonal posible como que se puede ejecutar, por ejemplo, por la etapa de procesamiento de tonales 213. Entre mayor sea una cantidad de componentes tonales dentro de la banda de frecuencia, menor puede ser el coeficiente de ponderación espectral respectivo 219. Por ejemplo, la cantidad de componentes tonales en la banda de frecuencia se puede modificar en escala entre un valor mínimo de 1 y un valor máximo de 8, en donde el valor mínimo indica que la banda de frecuencia no comprende o casi no comprende ningún componente tonal. El valor máximo puede indicar que la banda de frecuencia comprende una gran cantidad de componentes tonales. El coeficiente de ponderación espectral respectivo, tal como el coeficiente de ponderación espectral 219 también puede comprender un valor mínimo y un valor máximo. El valor mínimo, por
- 60

ES 2 797 742 T3

ejemplo, 0.1 puede indicar que la banda de frecuencia se atenúa casi completamente o completamente. El valor máximo puede indicar que la banda de frecuencia casi no se atenúa o completamente no se atenúa. El coeficiente de ponderación espectral 219 puede aceptar uno de una multitud de valores que indican el valor mínimo, el valor máximo y de manera preferente al menos un valor entre el valor mínimo y el valor máximo. De forma alternativa, el coeficiente de ponderación espectral puede disminuir para una parte disminuida de bandas de frecuencia tonales de tal forma que el coeficiente de ponderación espectral es un factor de consideración.

El procesador de señales se puede configurar para combinar el coeficiente de ponderación espectral para procesamiento de transitorios y/o el coeficiente de ponderación espectral para procesamiento de tonales con los valores espectrales de la banda de frecuencia como se describe para el procesador de señales 210. Por ejemplo, para una banda de frecuencia procesada un valor promedio de coeficiente de ponderación espectral 217 y/o 219 se puede determinar por la etapa de combinación 215. Los pesos espectrales de la banda de frecuencia se pueden combinar, por ejemplo multiplicar, con los valores espectrales de la señal de audio 102. De forma alternativa, la etapa de combinación se puede configurar para comparar ambos pesos espectrales 217 y 219 y/o para seleccionar el peso espectral menor o mayor de ambos y para combinar el peso espectral seleccionado con los valores espectrales. De forma alternativa, los pesos espectrales se pueden combinar de forma diferente, por ejemplo como una suma, como una diferencia, como un cociente o como un factor.

Una característica de una señal de audio puede variar con el paso del tiempo. Por ejemplo, una señal de radiodifusión puede comprender primero una señal de voz (señal de fuente de sonido prominente) y después una señal de música (señal de fuente de sonido no prominente) o viceversa. También, pueden presentarse variaciones dentro de una señal de voz y/o una señal de música. Esto puede conducir a cambios rápidos de pesos espectrales y/o factores de ponderación. El procesador de señales y/o el controlador se pueden configurar para adaptar de forma adicional los pesos espectrales y/o los factores de ponderación para disminuir o limitar las variaciones entre dos cuadros, por ejemplo al limitar un valor de incremento máximo entre los dos cuadros de señal. Uno o más cuadros de la señal de audio se pueden sumar en un periodo de tiempo, en donde el procesador de señales y/o el controlador se pueden configurar para comparar pesos espectrales y/o factores de ponderación de un periodo de tiempo anterior, por ejemplo uno o más cuadros anteriores y para determinar si una diferencia pesos espectrales y/o factores de ponderación determinados para un periodo de tiempo real excede un valor umbral. El valor umbral puede representar, por ejemplo, un valor que conduce a efectos molestos para un oyente. El procesador de señales y/o el controlador se pueden configurar para limitar las variaciones de tal forma que estos efectos molestos se reducen o impiden. De forma alternativa, en lugar de la diferencia, también se pueden determinar otras expresiones matemáticas tal como una relación para comparar los pesos espectrales y/o los factores de ponderación del periodo de tiempo anterior y el real.

En otras palabras, se asigna a cada banda de frecuencia una característica que comprende una cantidad de características tonales y/o transitorias.

La Figura 8 muestra un diagrama de bloques esquemático de un sistema mejorador de sonido 800 que comprende un aparato 801 para mejorar la señal de audio 102. El sistema mejorador de sonido 800 comprende una entrada de señal 106 configurada para recibir la señal de audio y para proporcionar la señal de audio al aparato 801. El sistema mejorador de sonido 800 comprende dos altavoces 808a y 808b. El altavoz 808a se configura para recibir la señal y1. El altavoz 808b se configura para recibir la señal y2 de tal forma que por medio de los altavoces 808a y 808b las señales y1 y y2 se pueden transferir a ondas de sonido o señales. La entrada de señal 106 puede ser una entrada de señal alámbrica o inalámbrica, tal como una antena de radio. El aparato 801, por ejemplo, puede ser un aparato 100 y/o 200.

La señal correlación z se obtiene al aplicar un procesamiento que mejora los componentes transitorios y tonales (cualitativamente inverso a la supresión por cálculo de la señal s). La combinación llevada a cabo por el combinador se puede expresar de forma lineal por $y_1/y_2 = \text{factor de modificación de escala } 1 \cdot z + \text{factor de modificación de escala } 2 \cdot \text{factor de modificación de escala } (r_1/r_2)$. Los factores de modificación de escala se pueden obtener al predecir la intensidad percibida de decorrelación.

De forma alternativa, las señales y1 y/o y2 se pueden procesar además antes de que se reciban por un altavoz 808a y/o 808b. Por ejemplo, las señales y1 y/o y2 se pueden amplificar, ecualizar o simular de tal forma que una señal o señales derivadas al procesar la señal y1 y/o y2 se proporcionan a los altavoces 808a y/o 808b.

La reverberación artificial añadida a la señal de audio se puede implementar de tal forma que el nivel de reverberación es audible, pero no demasiado alto (intenso). Los niveles que son audibles o molestos se pueden determinar en pruebas y/o simulaciones. Un nivel que es demasiado alto no suena bien debido a que sufre la claridad, los sonidos de percusión se arrastran en el tiempo, etc. Un nivel objetivo puede depender de la señal de entrada. Si la señal de entrada comprende una cantidad reducida de transitorios y comprende una cantidad reducida de tonos con moderaciones de frecuencia, entonces la reverberación es audible con un menor grado y el nivel se puede incrementar. Lo mismo aplica para una decorrelación ya que el decorrelacionador puede comprender un principio activo similar. Por lo tanto, una intensidad óptima del decorrelacionador puede depender de la señal de entrada. El cálculo puede ser

ES 2 797 742 T3

5 igual, con parámetros modificados. La decorrelación ejecutada en el procesador de señales y en el controlador se puede llevar a cabo con dos decorrelacionadores que pueden ser estructuralmente iguales pero operados con diferentes conjuntos de parámetros. Los procesadores de decorrelación no se limitan a señales estéreo de dos canales sino también se pueden aplicar canales con más de dos señales. La decorrelación se puede cuantificar con una métrica de correlación que puede comprender todos los valores para la decorrelación de todos los pares de señales.

10 Un hallazgo del método inventado es generar patrones espaciales e introducir los patrones espaciales a la señal de tal forma que la señal procesada crea la sensación de una señal estereofónica. El procesamiento se puede considerar como que se diseña de acuerdo con los siguientes criterios:

- 15 1. Las fuentes de sonido directo que tienen alta intensidad (o nivel de intensidad) se coloca en el centro. Estas fuentes de sonido prominentes, por ejemplo un cantante o instrumento ruidoso en una grabación musical.
2. Los sonidos ambientales se perciben como que son difusos.
- 20 3. Se añade difusión a fuente de sonido directo que tienen baja intensidad (es decir, bajos niveles de intensidad), posiblemente a un menor grado que los sonidos ambientales.
4. El procesamiento debe sonar natural y no debe introducir distorsiones.

Los criterios de diseño son consistentes con la práctica común en la producción de grabaciones de audio con características de señal de señales estereofónicas:

- 25 1. Los sonidos directos prominentes se panean convencionalmente al centro, es decir se mezclan con ICLD despreciable e ICTD. Estas señales exhiben una alta coherencia.
2. Los sonidos ambientales exhiben una baja coherencia.
- 30 3. Cuando se graban múltiples fuentes directas en un entorno reverberante, por ejemplo cantantes de ópera con orquesta acompañante la cantidad de difusión de cada sonido directo se relaciona con su distancia a los micrófonos, debido a que la relación entre la señal directa y la reverberación disminuye cuando se incrementa la distancia al micrófono. Por lo tanto, los sonidos que se capturan con poca intensidad son convencionalmente menos coherentes (o viceversa, más difusos) que los sonidos directos prominentes.

35 El procesamiento genera la información espacial por medio de decorrelación. En otras palabras, la ICC de las señales de entrada se disminuye. Sólo en casos extremos la decorrelación conduce a señales completamente no correlacionadas. Convencionalmente, se logra y se desea una decorrelación parcial. El procesamiento no manipula los patrones direccionales (es decir, ICLD e ICTD). La razón de esta restricción es que no está disponible ninguna información acerca de la posición original o propuesta de fuentes de sonido directo.

40 De acuerdo con los criterios de diseño anteriores, la decorrelación se aplica de forma selectiva a los componentes de señal en una señal de mezcla de tal forma que:

- 45 1. Se aplica poca o ninguna decorrelación a componentes de señal como se analiza en el criterio de diseño 1.
2. Se aplica decorrelación a componentes de señal como se analiza en el criterio de diseño 2. Esta decorrelación contribuye ampliamente al ancho percibido de la señal de mezcla que se obtiene a la salida del procesamiento.

50 Se aplica decorrelación a componentes de señal como se analiza en el criterio de diseño 3, pero a un menor grado que los componentes de señal como se analiza en el criterio de diseño 2.

55 Se ilustra este procesamiento por medio de un modelo de señal que representa la señal de entrada x como una mezcla aditiva de una señal de primer plano x_a y una señal de fondo x_b , es decir $x = x_a + x_b$. La señal de primer plano comprende todos los componentes de señal como se analiza en el criterio de diseño 1. La señal de fondo comprende todos los componentes de señal como se analiza en el criterio de diseño 2. Todos los componentes de señal como se analiza en el criterio de diseño 3 no se asignan exclusivamente a cualquiera de uno de los componentes de señal separados sino están contenidos parcialmente en la señal de primer plano y en la señal de fondo.

60 La señal de salida y se calcula como $y = y_a + y_b$, donde y_b se calcula al decorrelacionar de x_b , y $y_a = x_a$ o, alternativamente, y_a se calcula al decorrelacionar de x_a . En otras palabras, la señal de fondo se procesa por medio de decorrelación y la señal de primer plano no se procesa por medio de decorrelación o se procesa por medio de decorrelación, pero a un menor grado que la señal de fondo. La Figura 9b ilustra este procesamiento.

Este enfoque no sólo satisface los criterios de diseño anteriores. Una ventaja adicional es que la señal de primer plano

ES 2 797 742 T3

puede ser propensa a coloración indeseada cuando se aplica decorrelación, mientras que se puede decorrelacionar la señal de fondo sin introducir estas distorsiones audibles. Por lo tanto, el procesamiento descrito produce mejor calidad de sonido en comparación con un procesamiento que aplica decorrelación de forma uniforme a todos los componentes de señal en la mezcla.

5

Hasta ahora, la señal de entrada se descompone en dos señales denotadas como "señal de primer plano" y "señal de fondo" que se procesan de forma separada y combinan a la señal de salida. Se debe señalar que son factibles métodos equivalentes que siguen el mismo razonamiento.

10

La descomposición de señal no es necesariamente un procesamiento que saque señales de audio, es decir señales que se asemejan a la forma de onda con el paso del tiempo. En su lugar, la descomposición de señal puede dar por resultado cualquier otra representación de señal que se puede utilizar como la entrada al procesamiento de decorrelación y transformar posteriormente en una señal de forma de onda. Un ejemplo de esta representación de señal es un espectrograma que se calcula por medio de transformada de Fourier de corta duración. En general, las transformadas invertibles y lineales conducen a representaciones de señal apropiadas.

15

Alternativamente, los patrones espaciales se generan de forma selectiva sin la descomposición de señal precedente al generar la información estereofónica con base en la señal de entrada x . La información estereofónica derivada se pondera con valores de frecuencia selectiva y variantes en el tiempo y se combinan con la señal de entrada. Los factores de ponderación de frecuencia selectiva y variantes en el tiempo se calculan de tal forma que son grandes en regiones de tiempo-frecuencia que están dominadas por la señal de fondo y que son pequeños en regiones de tiempo-frecuencia que están dominadas por la señal de primer plano. Esto se puede formalizar al cuantificar la relación de frecuencia selectiva y variante en el tiempo de la señal de fondo y la señal de primer plano. Los factores de ponderación se pueden calcular a partir de la relación de señal de fondo a primer plano, por ejemplo por medio de funciones monótonamente crecientes.

20

25

Alternativamente, la descomposición de señal precedente puede dar por resultado más de dos señales separadas.

30

Las Figuras 9a y 9b ilustran la separación de la señal de entrada en una señal de primer plano y una de fondo, por ejemplo, al suprimir (reducir o eliminar) las porciones transitorias tonales en una de las señales.

35

Un procesamiento simplificado se deriva al utilizar la suposición de que la señal de entrada es una mezcla aditiva de la señal de primer plano y la señal de fondo. La Figura 9b ilustra esto. Aquí, la separación 1 denota la separación de cualquiera de la señal de primer plano o de la señal de fondo. Si se separa la señal de primer plano, la salida 1 denota la señal de primer plano y la salida 2 es la señal de fondo. Si se separa la señal de fondo, la salida 1 denota la señal de fondo y la salida 2 es la señal de primer plano.

40

El diseño e implementación del método de separación de señal se basa en el hallazgo de que las señales de primer plano y las señales de fondo tienen distintas características. Sin embargo, las desviaciones de una separación ideal, es decir pérdida de componentes de señal de las fuentes de sonido directo prominentes en la señal de fondo o pérdida de componente de señal ambientales en la señal de primer plano, son aceptables y no necesariamente perjudican la calidad de sonido del resultado final.

45

Para características temporales, en general se puede observar que las envolventes temporales de las señales de sub-banda de señales de primer plano ofrecen modulaciones de amplitud más intensas que las envolventes temporales de señales de sub-bandas de señales de fondo. En contraste, las señales de fondo convencionalmente son menos transitorias (o percutoras, es decir más sostenidas) que las señales de primer plano.

50

Para características espectrales, en general se puede observar que las señales de primer plano pueden ser más tonales. En contraste, las señales de fondo convencionalmente son más ruidosas que las señales de primer plano.

55

Para características de fase, en general se puede observar que la información de fase de las señales de fondo es más ruidosa que de las señales de primer plano. La información de fase para muchos ejemplos de señales de primer plano es congruente a través de múltiples bandas de frecuencia.

60

Las señales que ofrecen características que son similares a señales de fuente de sonido prominentes es más probable que sean señales de primer plano que señales de fondo. Las señales de fuente de sonido prominentes están caracterizadas por transiciones entre componentes de señal tonales y ruidosos, donde los componentes de señal tonales son trenes de pulsos filtrados variantes en el tiempo cuya frecuencia fundamental se modula fuertemente. El procesamiento espectral puede estar basado en estas características, la descomposición se puede implementar por medio de sustracción espectral o ponderación espectral.

Se lleva a cabo sustracción espectral, por ejemplo, en el dominio de la frecuencia, donde los espectros de cuadros cortos de porciones sucesivas (que posiblemente se superponen) de la señal de entrada se procesan. El principio

básico es sustraer una estimación del espectro de magnitud de una señal de interferencia de los espectros de magnitud de las señales de entrada que se asume que es una mezcla aditiva de una señal deseada y de una señal de interferencia. Para la separación de la señal de primer plano, la señal deseada es la señal de primer plano y la señal de interferencia es la señal de fondo. Para la separación de la señal de fondo, la señal deseada es la señal de fondo y la señal de interferencia es la señal de primer plano.

La ponderación espectral (o atenuación espectral de corta duración) sigue el mismo principio y atenúa la señal de interferencia al modificar la escala de la representación de señal de entrada. La señal de entrada $x(t)$ se transforma utilizando una transformada de Fourier de corta duración (STFT), un banco de filtros o cualquier otro medio para derivar una representación de señal con múltiples bandas de frecuencia $X(n,k)$, con índice de banda de frecuencia n e índice de tiempo k . Las representaciones en el dominio de la frecuencia de las señales de entrada se procesan de tal forma que las señales de sub-banda se modifican en escala con coeficientes de ponderación variantes en el tiempo $G(n,k)$,

$$Y(n, k) = G(n, k)X(n, k) \quad (3)$$

El resultado de la operación de ponderación $Y(n,k)$ es la representación en el dominio de la frecuencia de la señal de salida. La señal de tiempo de salida $y(t)$ se calcula utilizando el procesamiento inverso de la transformada en el dominio de la frecuencia, por ejemplo la STFT inversa. La Figura 10 ilustra la ponderación espectral.

La decorrelación se refiere a un procesamiento de una o más señales de entrada idénticas de tal forma que se obtienen múltiples señales de salida que están mutuamente no correlacionadas (parcial o completamente), pero que suenan similar a la señal de entrada. La correlación entre dos señales se puede medir por medio del coeficiente de correlación o coeficiente de correlación normalizado. El coeficiente de correlación normalizado NCC en bandas de frecuencia para dos señales $X_1(n,k)$ y $X_2(n,k)$ se define como

$$NCC(n, k) = \frac{|\phi_{1,2}(n, k)|}{\sqrt{\phi_{1,1}(n, k)\phi_{2,2}(n, k)}}, \quad (4)$$

donde $\phi_{1,1}$ y $\phi_{2,2}$ son las densidades de potencia auto-espectrales (PSD) de la primera y segunda señales de entrada, respectivamente, y $\phi_{1,2}$ es la PSD transversal, dada por

$$\phi_{i,j}(n, k) = \mathcal{E}\{X_i(n, k)X_j^*(n, k)\}, \quad i, j = 1, 2, \quad (5)$$

donde $\mathcal{E}\{\cdot\}$ es la operación de expectativa y X^* denota el conjugado complejo de X .

La decorrelación se puede implementar al utilizar filtros decorrelacionadores o al manipular la fase de las señales de entrada en el dominio de la frecuencia. Un ejemplo de filtros decorrelacionadores es el filtro de todo paso, que por definición no cambia el espectro de magnitud de las señales de entrada sino sólo su fase. Esto conduce a un sonido neutro de las señales de salida en el sentido de que las señales de salida suenan similar a las señales de entrada. Otro ejemplo es reverberación, que se también se puede modelar como un filtro o un sistema invariante en el tiempo, lineal. En general, la decorrelación se puede lograr al añadir múltiples copias retrasadas (y posiblemente filtradas) de la señal de entrada a la señal de entrada. En términos matemáticos, la reverberación de distorsión se puede implementar como convolución de la señal de entrada con la respuesta de impulso del sistema de reverberación (o decorrelación). Cuando es pequeño el tiempo de retraso, por ejemplo menor que 50 ms, las copias retrasadas de la señal no se perciben como señales separadas (ecos). El valor exacto del tiempo de retraso que conduce a la sensación de ecos es el umbral de eco y depende de las características temporales y espectrales de la señal. Por ejemplo es más pequeño para sonidos tipo impulso que para sonidos cuyo envolvente se eleva lentamente. Para el problema de interés se desea utilizar tiempos de retraso que sean más pequeños que el umbral de eco.

En el caso general, la decorrelación procesa una señal de entrada que tiene n canales y saca una señal que tiene m canales de tal forma que las señales de canal de la salida no están mutuamente correlacionadas (parcialmente o completamente).

En muchos escenarios de aplicación para el método descrito no es apropiado procesar de forma constante la señal de entrada sino activarla y controlar su impacto con base en un análisis de la señal de entrada. Un ejemplo es difusión FM, donde el método descrito se aplica sólo cuando los impedimentos de la transmisión conducen a una pérdida completa o parcial de la información estereofónica. Otro ejemplo es la escucha de una colección de grabaciones musicales, donde un subconjunto de las grabaciones son monofónicas y otros subconjuntos son grabaciones estéreo. Ambos escenarios están caracterizados por una cantidad que varían en el tiempo de información estereofónica de las señales de audio. Esto requiere un control de la activación y el impacto de la mejora estereofónica, es decir un control del algoritmo.

ES 2 797 742 T3

El control se implementa por medio de un análisis de las señales de audio que estima los patrones espaciales (ICLD, ICTD e ICC, o un subconjunto de las mismas) de las señales de audio. La estimación se puede llevar a cabo de una forma de frecuencia selectiva. La salida de la estimación se mapea a un valor escalar que controla la activación o el impacto del procesamiento. El análisis de señal procesa la señal de entrada o, alternativamente, la señal de fondo separada.

Una forma directa de controlar el impacto de procesamiento es disminuir su impacto al añadir una copia (posiblemente modificada en escala) de la señal de entrada a la señal de salida (posiblemente modificada en escala) de la mejora estereofónica. Las transiciones suaves del control se obtienen por filtración de paso bajo de la señal de control con el paso del tiempo.

La Figura 9a muestra un diagrama de flujo esquemático de un procesamiento 900 de la señal de entrada 102 de acuerdo con un procesamiento de primer plano/fondo. La señal de entrada 102 se separa de tal forma que se puede procesar una señal de primer plano 914. En un paso 916 se lleva a cabo decorrelación a la señal de primer plano 914. El paso 916 es opcional. De forma alternativa, la señal de primer plano 914 puede permanecer sin procesar, es decir sin decorrelación. En un paso 922 de una ruta de procesamiento 920, se extrae una señal de fondo 924, es decir se filtra. En un paso 926 se decorrelaciona de la señal de fondo 924. En un paso 904 una señal de primer plano decorrelacionada 918 (alternativamente la señal de primer plano 914) y una señal de fondo decorrelacionada 928 se mezclan de tal forma que se obtiene una señal de salida 906. En otras palabras, la Figura 9a muestra un diagrama de bloques de la mejora estereofónica. Se calcula una señal de primer plano y una señal de fondo. La señal de fondo se procesa por decorrelación. Opcionalmente, la señal de primer plano se puede procesar por decorrelación, pero a un menor grado que la señal de fondo. Las señales procesadas se combinan a la señal de salida.

La Figura 9b ilustra un diagrama de bloques esquemático de un procesamiento 900' que comprende un paso de separación 912' de la señal de entrada 102. El paso de separación 912' se puede llevar a cabo como se describió anteriormente. Se obtiene una señal de primer plano (señal de salida 1) 914' por el paso de separación 912'. Se obtiene una señal de fondo 928' al combinar la señal de primer plano 914', los factores de ponderación a y/o b y la señal de entrada 102 en un paso de combinación 926'. Se obtiene una señal de fondo (señal de salida 2) 928' por el paso de combinación 926'.

La Figura 10 muestra un diagrama de bloques esquemático y también un aparato 1000 configurado para aplicar pesos espectrales a una señal de entrada 1002 que puede ser, por ejemplo, la señal de entrada 1002. La señal de entrada 1002 en el dominio del tiempo se divide en sub-bandas $X(1,k) \dots X(n,k)$ en el dominio de la frecuencia. Un banco de filtros 1004 se configura para dividir la señal de entrada 1002 en N sub-bandas. El aparato 1000 comprende N instancias de computación configuradas para determinar el coeficiente de ponderación espectral transitorio y/o el coeficiente de ponderación espectral tonal $G(1,k) \dots G(n,k)$ para cada una de las N sub-bandas en la instancia (cuadro) de tiempo k. Los pesos espectrales $G(1,k) \dots G(n,k)$ se combinan con la señal de sub-banda $X(1,k) \dots X(n,k)$, de tal forma que se obtienen señales de sub-banda ponderadas $Y(1,k) \dots Y(n,k)$. El aparato 1000 comprende una unidad de procesamiento inverso 1008 configurada para combinar las señales de sub-banda ponderadas para obtener una señal de salida filtrada 1012 indicada como $Y(t)$ en el dominio del tiempo. El aparato 1000 puede ser una parte del procesador de señales 110 o 210. En otras palabras, la Figura 10 ilustra la descomposición de una señal de entrada en una señal de primer plano y una señal de fondo.

La Figura 11 muestra un diagrama de flujo esquemático de un método 1100 para mejorar una señal de audio. El método 1100 comprende el primer paso 1110 en la cual la señal de audio se procesa a fin de reducir o eliminar porciones transitorias y tonales de la señal procesada. El método 1100 comprende un segundo paso 1120 en el cual se genera una primera señal decorrelacionada y una segunda señal decorrelacionada de la señal procesada. En un paso 1130 del método 1100 se combinan de forma ponderada la primera señal decorrelacionada, la segunda señal decorrelacionada y la señal de audio o una señal derivada de la señal de audio por mejora de coherencia al utilizar factores de ponderación variantes en el tiempo para obtener una señal de audio de dos canales. En un paso 1140 del método 1100 los factores de ponderación variante en el tiempo se controlan al analizar la señal de audio de tal forma que se multiplican diferentes porciones de la señal de audio por diferentes factores de ponderación y la señal de audio de dos canales tiene un grado variante en el tiempo de una decorrelación.

En lo siguiente se expondrán detalles para ilustrar la posibilidad de determinar el nivel percibido de decorrelación con base en una medición de intensidad. Como se mostrará, una medición de intensidad puede permitir predecir un nivel percibido de reverberación. Como se indicó anteriormente, la reverberación también se refiere a la decorrelación de tal forma que el nivel percibido de reverberación también se puede considerar como un nivel percibido de decorrelación, en donde para una decorrelación, la reverberación puede ser más corta que un segundo, por ejemplo más corta que 500 ms, más corta que 250 ms o más corta que 200 ms.

La Figura 12 ilustra un aparato para determinar una medición para un nivel percibido de reverberación en una señal de mezcla que comprende un componente de señal directa o componente de señal seca 1201 y un componente de

ES 2 797 742 T3

señal de reverberación 102. El componente de señal seca 1201 y el componente de señal de reverberación 1202 se introduce en un procesador de modelo de intensidad 1204. El procesador de modelo de intensidad se configura para recibir el componente de señal directa 1201 y el componente de señal de reverberación 1202 y comprende además una etapa de filtro perceptual 1204a y una calculadora de intensidad conectada de forma posterior 1204b como se ilustra en la Figura 13a. El procesador de modelo de intensidad genera, como su salida, una primera medición de intensidad 1206 y una segunda medición de intensidad 1208. Ambas mediciones de intensidad se introducen en un combinador 1210 para combinar la primera medición de intensidad 1206 y la segunda medición de intensidad 1208 para obtener finalmente una medición 1212 para el nivel percibido de reverberación. Dependiendo de la implementación, la medición para el nivel percibido 1212 se puede introducir a un predictor 1214 para predecir el nivel percibido de reverberación con base en un valor promedio de al menos dos mediciones para la intensidad percibida de diferentes cuadros de señal. Sin embargo, el predictor 1214 en la Figura 12 es opcional y transforma realmente la medición para el nivel percibido en un cierto intervalo de valores o intervalo de unidades tal como el intervalo de unidades Sone que es útil para dar valores cuantitativos relacionados con la intensidad. Sin embargo, otros usos para la medición del nivel percibido 1212 que no se procesa por el predictor 1214 también se puede utilizar, por ejemplo, en el controlador, que no necesariamente tiene que depender de un valor sacado por el predictor 1214, sino que también puede procesar directamente la medición para el nivel percibido 1212, ya sea en una forma directa o de manera preferente en una clase de forma suavizada donde la suavización con el paso del tiempo es preferible a fin de no tener que cambiar intensamente las correcciones del nivel de la señal reverberada o de un factor de ganancia g.

De forma particular, la etapa de filtro perceptual se configura para filtrar el componente de señal directa, el componente de señal de reverberación o el componente de señal de mezcla, en donde la etapa de filtro perceptual se configura para modelar un mecanismo de percepción auditivo de una entidad tal como un ser humano para obtener una señal directa filtrada, una señal de reverberación filtrada o una señal de mezcla filtrada. Dependiendo de la implementación, la etapa de filtro perceptual puede comprender dos filtros que operan en paralelo o puede comprender un almacenamiento y un filtro individual ya que una y el mismo filtro pueden utilizarse realmente para filtrar cada una de las tres señales, es decir, la señal de reverberación, la señal de mezcla y la señal directa. En este contexto, sin embargo, se va a señalar que, aunque la Figura 13a ilustra n filtros que modelan el mecanismo de percepción auditivo, realmente serán suficientes dos filtros o un solo filtro que filtra dos señales del grupo que comprende el componente de señal de reverberación, el componente de señal de mezcla y el componente de señal directa.

La calculadora de intensidad 1204b o estimador de intensidad se configura para estimar la primera medición relacionada con la intensidad utilizando la señal directa filtrada o para estimar la segunda medición de intensidad utilizando la señal de reverberación filtrada o la señal de mezcla filtrada, en donde la señal de mezcla se deriva de una súper posición del componente de señal directa y el componente de señal de reverberación.

La Figura 13c ilustra cuatro modelos preferidos para calcular la medición del nivel percibido de reverberación. Una implementación se basa en la intensidad parcial donde se utiliza tanto el componente de señal directa x como el componente de señal de reverberación r en el procesador de modelo de intensidad, pero donde, a fin de determinar la primera medición EST1, la señal de reverberación se utiliza como el estímulo y la señal directa se utiliza como el ruido. Para determinar la segunda medición de intensidad EST2, se cambia la situación, y el componente de señal directa se utiliza como un estímulo y el componente de señal de reverberación se utiliza como el ruido. Entonces, la medición para el nivel percibido de correlación generada por el combinador es una diferencia entre la primera medición de intensidad EST1 y la segunda medición de intensidad EST2.

Sin embargo, existen de forma adicional otras realizaciones computacionalmente eficientes que se indican en las líneas 2, 3, y 4 en la Figura 13c. Estas mediciones más computacionalmente eficientes se basan en el cálculo de la intensidad total de tres señales que comprende la señal de mezcla m , la señal directa x y la señal de reverberación n . Dependiendo del cálculo requerido llevado a cabo por el combinador indicado en la última columna de la Figura 13c, la primera medición de intensidad EST1 es la intensidad total de la señal de mezcla o la señal de reverberación y la segunda medición de intensidad EST2 es la intensidad total del componente de señal directa x o el componente de señal de mezcla m , donde las combinaciones reales son como se ilustra en la Figura 13c.

La Figura 14 ilustra en implementación del procesador de modelo de intensidad que ya se ha analizado en algunos aspectos con respecto a las Figuras 12, 13a, 13b, 13c. Particularmente, la etapa de filtro perceptual 1204a comprende un convertidor de tiempo-frecuencia 1401 para cada derivación, donde, en la realización de la Figura 3, $x[k]$ indica el estímulo y $n[k]$ indica el ruido. La señal convertida en tiempo/frecuencia se retransmite en un bloque de función de transferencia auricular 1402 (por favor nótese que la función de transferencia auricular se puede calcular alternativamente antes del convertidor del tiempo-frecuencia con resultados similares, pero a mayor carga computacional) y la salida de este bloque 1402 se introduce a un bloque de cálculo de patrón de excitación 1404 seguido por un bloque de integración temporal 1406. Entonces, en el bloque 1408, la intensidad específica en esta realización se calcula, donde el bloque 1408 corresponde al bloque calculador de intensidad 1204b en la Figura 13a. Posteriormente, se lleva a cabo una integración sobre la frecuencia en el bloque 1410, donde el bloque 1410 corresponde al sumador ya descrito como 1204c y 1204d en la Figura 13b. También se va a señalar que el bloque 1410 genera la primera medición para un primer conjunto de estímulo y ruido y la segunda medición para un segundo

ES 2 797 742 T3

conjunto de estímulo y ruido. Particularmente, cuando se considera la Figura 13b, el estímulo para calcular la primera medición es la señal de reverberación y el ruido es la señal directa en tanto que, para calcular la segunda medición, se cambia la situación y el estímulo es el componente de señal directa y el ruido es el componente de señal de reverberación. Por lo tanto, para generar dos mediciones de intensidad diferentes, se ha llevado a cabo dos veces el procedimiento ilustrado en la Figura 14. Sin embargo, sólo se presentan cambios en el cálculo en el bloque 1408 que opera de forma diferente, de tal forma que los pasos ilustrados por los bloques 1401 a 1406 sólo se tienen que llevar a cabo una vez, y el resultado de bloque de integración temporal 1406 se puede almacenar a fin de calcular la primera intensidad estimada y la segunda intensidad estimada para la implementación representada en la Figura 13c. Se va a señalar que, para la otra implementación, el bloque 1408 se puede reemplazar por un bloque individual "calcular intensidad total" para cada derivación, donde, en esta implementación es indiferente, si se considera que una señal va a ser un estímulo o un ruido.

Aunque se han descrito algunos aspectos en el contexto de un aparato, es claro que estos aspectos también representan una descripción del método correspondiente, donde un bloque o dispositivo corresponde a un paso de método o una característica de un paso de método. De forma análoga, los aspectos descritos en el contexto de un paso de método también representan una descripción de un bloque o elemento o característica correspondiente de un aparato correspondiente.

Dependiendo de ciertos requerimientos de implementación, las realizaciones de la invención se pueden implementar en hardware o software. La implementación se puede llevar a cabo utilizando un medio de almacenamiento digital, por ejemplo un disco flexible, un DVD, un CD, una ROM, una PROM, una EPROM, una EEPROM o una memoria FLASH, que tiene señales de control eléctricamente leíbles almacenadas en la misma, que cooperan (o son capaces de cooperar) con un sistema de ordenador programable de tal forma que se lleve a cabo el método respectivo.

Algunas realizaciones de acuerdo con la invención comprenden un portador de datos que tiene señales de control electrónicamente leíbles, que son capaces de cooperar con un sistema de ordenador programable, de tal forma que se lleve a cabo uno de los métodos descritos en la presente.

En general, las realizaciones de la presente invención se pueden implementar como un producto de programa informático con un código de programa, el código de programa que es operativo para llevar a cabo uno de los métodos cuando el producto de programa informático se ejecuta en un ordenador. El código de programa se puede almacenar, por ejemplo en un portador leíble por máquina.

Otras realizaciones comprenden el programa informático para realizar uno de los métodos descritos en la presente, almacenado en un portador leíble por máquina.

En otras palabras, una realización del método inventivo es, por lo tanto, un programa informático que tiene un código de programa para realizar uno de los métodos descritos en la presente, cuando el programa informático se ejecuta en un ordenador.

Una realización adicional de los métodos inventivos es, por lo tanto, un portador de datos (o un medio de almacenamiento digital, o un medio leíble por ordenador) que comprende, grabado en el mismo, el programa informático para realizar uno de los métodos descritos en la presente.

Una realización adicional del método inventivo es, por lo tanto un flujo de datos o una secuencia de señales que representan el programa informático para realizar uno de los métodos descritos en la presente. El flujo de datos o la secuencia de señales se puede configurar, por ejemplo para que se transfiera a través de una conexión de comunicación de datos, por ejemplo a través de internet.

Una realización adicional comprende un medio de procesamiento, por ejemplo un ordenador, o un dispositivo lógico programable, configurado para o adaptado para llevar a cabo uno de los métodos descritos en la presente.

Una realización adicional comprende un ordenador que tiene instalada en la misma el programa informático para realizar uno de los métodos descritos en la presente.

Las realizaciones descritas anteriormente son simplemente ilustrativas para los principios de la presente invención. Se entiende que serán evidentes modificaciones y variaciones de los arreglos y los detalles descritos en la presente para otros expertos en la técnica. Se propone que, por lo tanto, se limite solamente por el alcance de las reivindicaciones de patente que siguen y no por los detalles específicos presentados a manera de descripción y explicación de las realizaciones en la presente.

ES 2 797 742 T3

REIVINDICACIONES

1. Aparato (100; 200) para mejorar una señal de audio (102) que es una señal mono o una señal tipo mono, que comprende:
5 un procesador de señales (110; 210) para procesar la señal de audio (102) a fin de reducir o eliminar porciones transitorias y tonales de la señal procesada (112; 212);
10 un decorrelacionador (120; 520) para generar una primera señal decorrelacionada (124; r2) y una segunda señal decorrelacionada de la señal procesada (112; 212);
15 un combinador (140; 240) para combinar de forma ponderada la primera señal decorrelacionada (122; 522, r1), la segunda señal decorrelacionada (124; r2) y la señal de audio o una señal derivada de la señal de audio (102) por mejora de coherencia utilizando factores de ponderación variantes en el tiempo (a, b) y para obtener una señal de audio de dos canales (142; 242); y
20 un controlador (130; 230) para controlar los factores de ponderación variante en el tiempo (a, b) al analizar la señal de audio (122) de tal forma que se multiplican diferentes porciones (fb1-fb7) de la señal de audio por diferentes factores de ponderación (a, b) y la señal de audio de dos canales (142; 242) tiene un grado variante en el tiempo de decorrelación.
2. Aparato de acuerdo con la reivindicación 1, en donde el controlador (130; 230) se configura para incrementar los factores de ponderación (a, b) para porciones (fb1-fb7) de la señal de audio (102) que permiten un mayor grado de decorrelación y para disminuir los factores de ponderación (a, b) para porciones (fb1-fb7) de la señal de audio (102) que permiten un menor grado de decorrelación.
3. Aparato de acuerdo con la reivindicación 1 o 2, en donde el controlador (130; 230) se configura para modificar la escala de los factores de ponderación (a, b) de tal forma que un nivel percibido de decorrelación en la señal de audio de dos canales (142; 242) permanece dentro de un intervalo alrededor de un valor objetivo, el intervalo que se extiende a $\pm 20\%$ del valor objetivo.
4. Aparato de acuerdo con la reivindicación 3, en donde el controlador (130; 230) se configura para determinar el valor objetivo al reverberar la señal de audio (102) para obtener una señal de audio reverberada y al comparar la señal de audio (102) reverberada con la señal de audio para obtener un resultado de la comparación, en donde el controlador se configura para determinar el nivel percibido de decorrelación (232) con base en el resultado de la comparación.
5. Aparato de acuerdo con una de las reivindicaciones precedentes, en donde el controlador (130; 230) se configura para determinar una porción de señal de fuente de sonido prominente en la señal de audio (102) y para disminuir los factores de ponderación (a, b) para la porción de señal de fuente de sonido prominente comparada con una porción de la señal de audio (102) que no comprende una señal de fuente de sonido prominente; y
40 en donde el controlador (130; 230) se configura para determinar una porción de señal de fuente de sonido no prominente en la señal de audio (102) y para incrementar los factores de ponderación (a, b) para la porción de señal de fuente de sonido no prominente comparada con una porción de la señal de audio (102) que no comprende una señal de fuente de sonido no prominente.
6. Aparato de acuerdo con una de las reivindicaciones anteriores, en donde el controlador (130; 230) se configura para:
50 generar una señal decorrelacionada de prueba de una porción de la señal de audio (102);
55 derivar una medición para un nivel percibido de decorrelación de la porción de la señal de audio y la señal decorrelacionada de prueba; y
para derivar los factores de ponderación de la medición (a, b) para el nivel percibido de decorrelación.
7. Aparato de acuerdo con la reivindicación 6, en donde el decorrelacionador (120, 520) se configura para generar la primera señal decorrelacionada (122; r1) con base en una reverberación de la señal de audio (102) con un primer tiempo de reverberación en donde el controlador (130; 230) se configura para generar la señal decorrelacionada de prueba con base en una reverberación de la señal de audio (102) con un segundo tiempo de reverberación, en donde el segundo tiempo de reverberación es más corto que el primer tiempo de reverberación.

ES 2 797 742 T3

8. Aparato de acuerdo con una de las reivindicaciones anteriores, en donde
- 5 el controlador (130; 230) se configura para controlar los factores de ponderación (a, b) de tal forma que los factores de ponderación (a, b) cada uno comprende un valor de una primera multitud de valores posibles la primera multitud que comprende al menos tres valores que comprenden un valor mínimo, un valor máximo y un valor entre el valor mínimo y el valor máximo; y en el que
- 10 el procesador de señales (110; 210) se configura para determinar pesos espectrales (217, 219) para una segunda multitud de bandas o frecuencia cada una que representa una porción de la señal de audio (102) en el dominio de la frecuencia, en donde los pesos espectrales (217, 219) cada uno comprende un valor de una tercera multitud de valores posibles, la tercera multitud que comprende al menos tres valores que comprenden un valor mínimo, un valor máximo y un valor entre el valor mínimo y el valor máximo.
- 15 9. Aparato de acuerdo con una de las reivindicaciones anteriores en donde el procesador de señales (110; 210) se configura para:
- 20 procesar la señal de audio (102) de tal forma que la señal de audio (102) se transfiere en el dominio de la frecuencia y de tal forma que una segunda multitud de bandas de frecuencia (fb1-fb7) representa la segunda multitud de porciones de la señal de audio (102) en el dominio de la frecuencia;
- 25 para determinar para cada banda de frecuencia (fb1-fb7) un primer coeficiente de ponderación espectral (217) que representa un valor de procesamiento para procesamiento (211) de transitorios de la señal de audio (102);
- 30 para determinar para cada banda de frecuencia (fb1-fb7) un segundo coeficiente de ponderación espectral (219) que representa un valor de procesamiento para procesamiento (213) de tonales de la señal de audio (102); y
- 35 para aplicar para cada banda de frecuencia (fb1-fb7) al menos uno del primer coeficiente de ponderación espectral (217) y el segundo coeficiente de ponderación espectral (219) a valores espectrales de la señal de audio (102) en la banda de frecuencia (fb1-fb7);
- 40 en donde los primeros pesos espectrales (217) y los segundos pesos espectrales (219) cada uno comprende un valor de una tercera multitud de valores posibles, la tercera multitud que comprende al menos tres valores que comprenden un valor mínimo, un valor máximo y un valor entre el valor mínimo y el valor máximo.
- 45 10. Aparato de acuerdo con la reivindicación 9, en donde para cada una de la segunda multitud de bandas de frecuencia (fb1-fb7) el procesador de señales (110; 210) se configura para comparar el primer coeficiente de ponderación espectral (217) y el segundo coeficiente de ponderación espectral (219) determinados para la banda de frecuencia (fb1-fb7), para determinar, si uno de los dos valores comprende un valor más pequeño y para aplicar el coeficiente de ponderación espectral (217, 219) que comprende el valor más pequeño a los valores espectrales de la señal de audio (102) en la banda de frecuencia(fb1-fb7).
- 50 11. Aparato de acuerdo con una de las reivindicaciones anteriores, en donde el decorrelacionador (520) comprende un primer filtro decorrelacionador (526) configurado para filtrar la señal de audio procesada (512, s) para obtener la primera señal decorrelacionada (522, r1) y un segundo filtro decorrelacionado (528) configurado para filtrar la señal de audio procesada (512, s) para obtener una segunda señal decorrelacionada (524, r2), en donde el combinador (140; 240) se configura para combinar de forma ponderada la primera señal decorrelacionada (522, r1), la segunda señal decorrelacionada (524, r2) y la señal de audio (102) o la señal (136; 236) derivada de la señal de audio (102) para obtener la señal de audio de dos canales(142; 242).
- 55 12. Aparato de acuerdo con una de las reivindicaciones anteriores, en donde para una segunda pluralidad de bandas de frecuencia (fb1-fb7), cada una de las bandas de frecuencia (fb1-fb7) que comprende una porción de la señal de audio (102) representada en el dominio de la frecuencia y con un primer periodo de tiempo
- 60 el controlador (130; 230) se configura para controlar los factores de ponderación (a, b) de tal forma que los factores de ponderación (a, b) cada uno comprende un valor de una primera multitud de valores posibles la primera multitud que comprende al menos tres valores que comprenden un valor mínimo, un valor máximo, y un valor entre el valor mínimo y el valor máximo y para adaptar los factores de ponderación (a, b) determinados para un periodo de tiempo real si una relación o una diferencia basada en un valor de los factores de ponderación (a, b) determinados para el periodo de tiempo real y un valor de los factores de ponderación (a, b) determinados para un periodo de tiempo anterior es mayor que o igual a un valor umbral

ES 2 797 742 T3

de tal forma que se reduce un valor de la relación o la diferencia; y

5 el procesador de señales (110; 210) se configura para determinar los pesos espectrales (217, 219) cada uno que comprende un valor de una tercera multitud de valores posibles, la tercera multitud que comprende al menos tres valores que comprenden un valor mínimo, un valor máximo y un valor entre el valor mínimo y el valor máximo.

13. Sistema mejorador de sonido (800) que comprende

10 un aparato (801) para mejorar una señal de audio de acuerdo con una de las reivindicaciones anteriores;

una entrada de señal (106) configurada para recibir la señal de audio (102);

15 al menos dos altavoces (808a, 808b) configurados para recibir la señal de audio de dos canales (y_1/y_2) o una señal derivada de la señal de audio de dos canales (y_1/y_2) y para generar señales acústicas de la señal de audio de dos canales (y_1/y_2) o la señal derivada de la señal de audio de dos canales (y_1/y_2).

14. Método (1100) para mejorar una señal de audio (102) que es una señal mono o una señal tipo mono, que comprende:

20

procesar (1110) la señal de audio (102) a fin de reducir o eliminar porciones transitorias y tonales de la señal procesada (112; 212);

25

generar (1120) una primera señal decorrelacionada (122, r_1) y una segunda señal decorrelacionada (124, r_2) de la señal procesada (112, 212);

30

combinar de forma ponderada (1130) la primera señal decorrelacionada (122, r_1), la segunda señal decorrelacionada (124, r_2) y la señal de audio (102) o una señal (136; 236) derivada de la señal de audio (102) por mejora de coherencia utilizando factores de ponderación variantes en el tiempo (a, b) y para obtener una señal de audio de dos canales (142; 242); y

35

controlar (1140) los factores de ponderación variantes (a, b) en el tiempo al analizar la señal de audio (102) de tal forma que se multiplican diferentes porciones de la señal de audio por diferentes factores de ponderación (a, b) y la señal de audio de dos canales (142; 242) tiene un grado variante en el tiempo de decorrelación.

15. Medio de almacenamiento no transitorio que tiene almacenado en el mismo un programa informático que tiene un código de programa para realizar, cuando se ejecuta en un ordenador, un método para mejorar una señal de audio de acuerdo con la reivindicación 14.

40

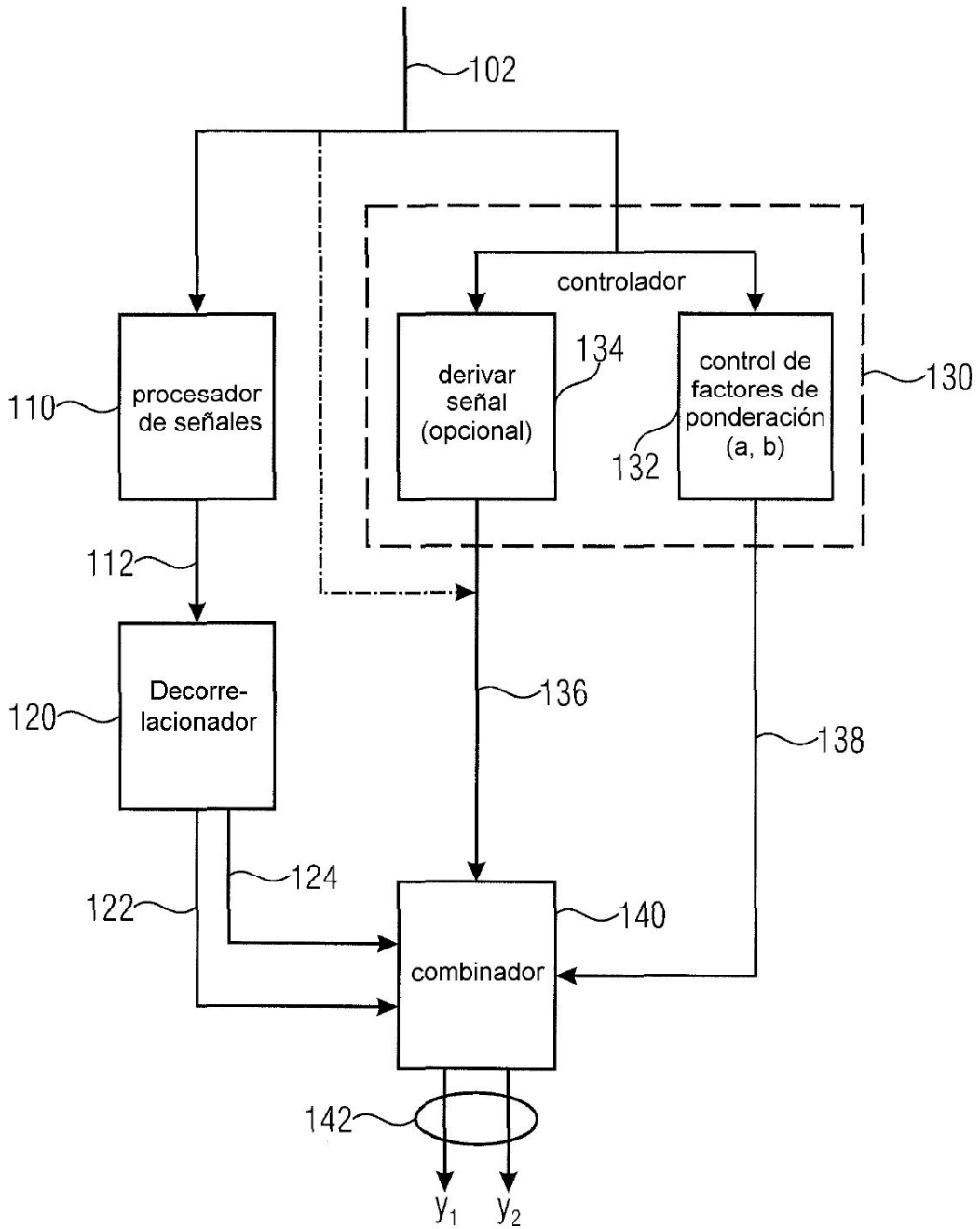


FIGURA 1

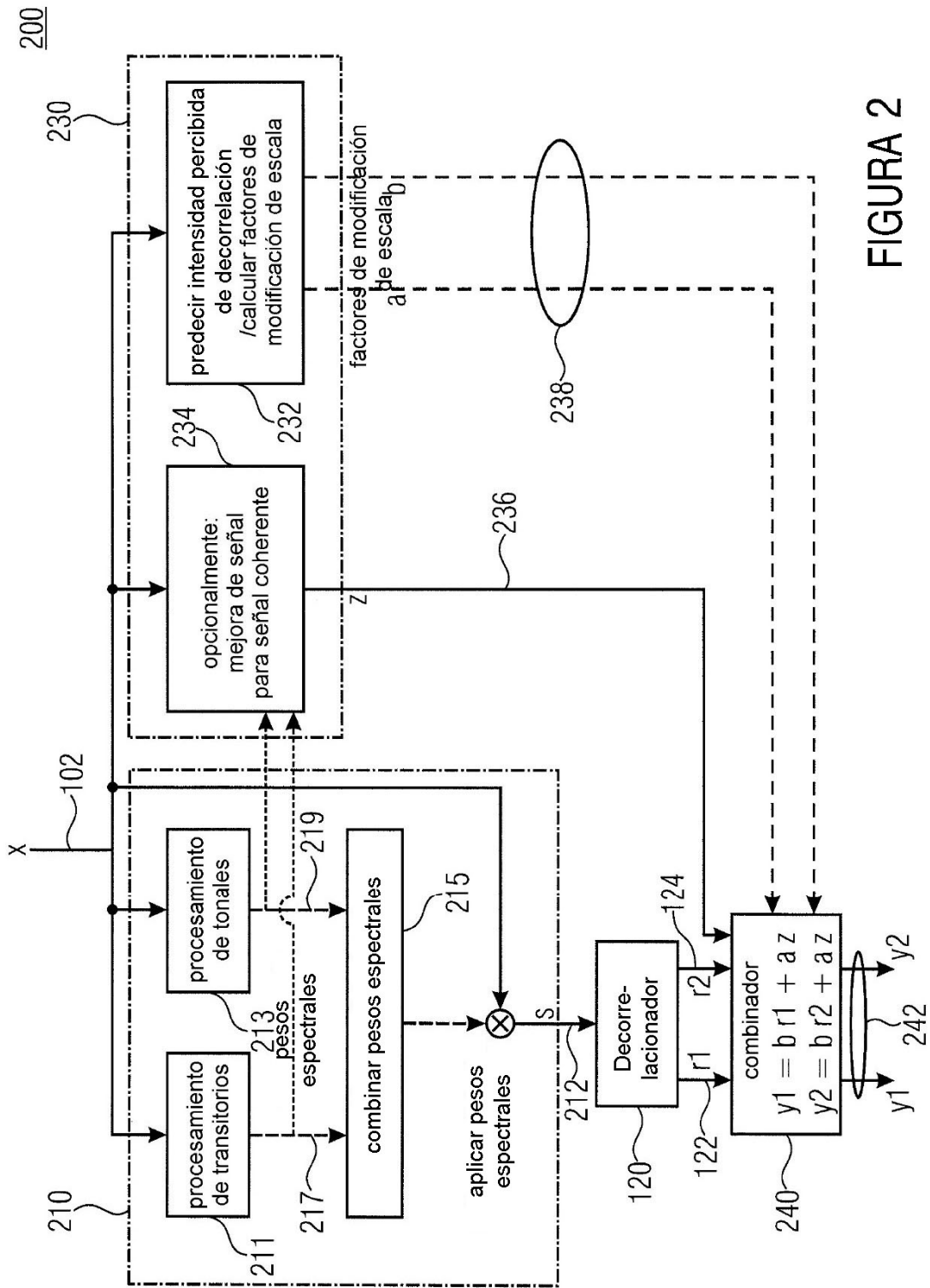


FIGURA 2

nivel de decorrelación	a	b
10	9	2
9	8	2
8	7	2
7	6	2
6	5.4	3
5	4	3.8
4	3.5	4
3	3.5	5
2	2.8	6
1	2	7
0	1	8

FIGURA 3

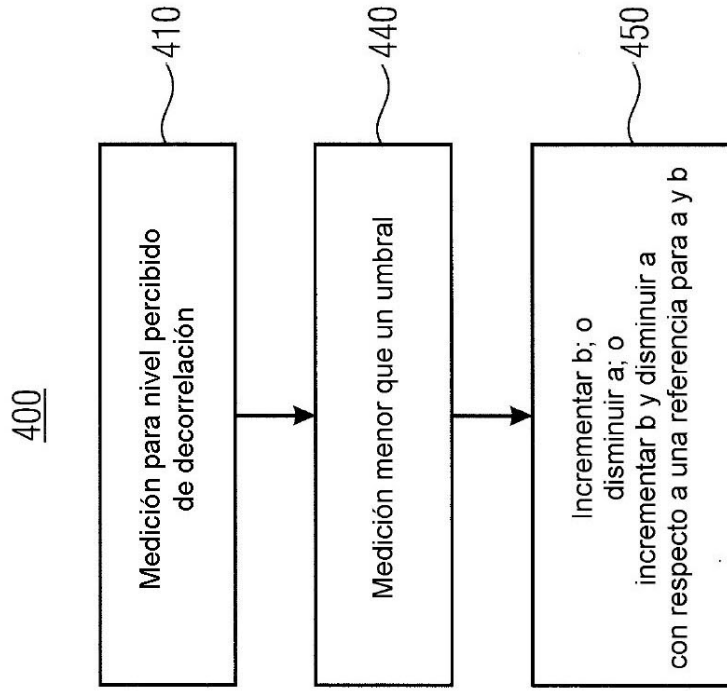


FIGURA 4B

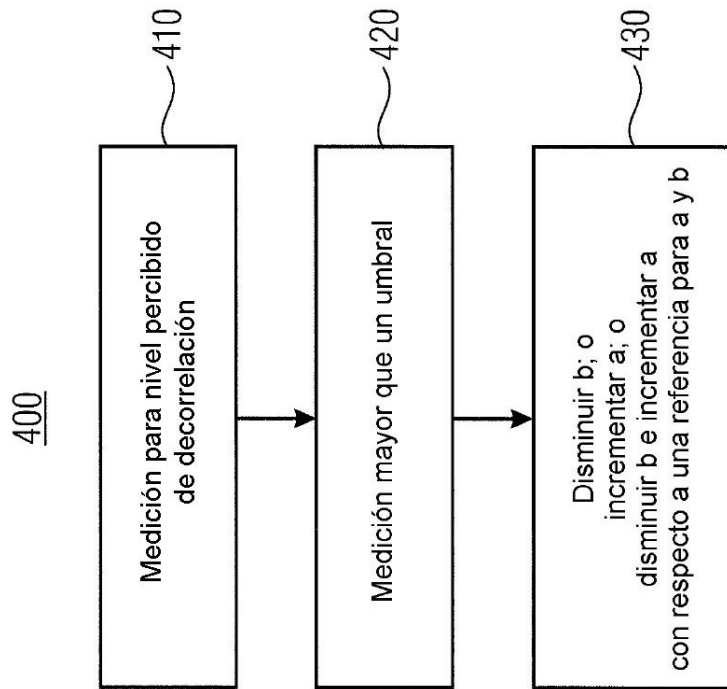


FIGURA 4A

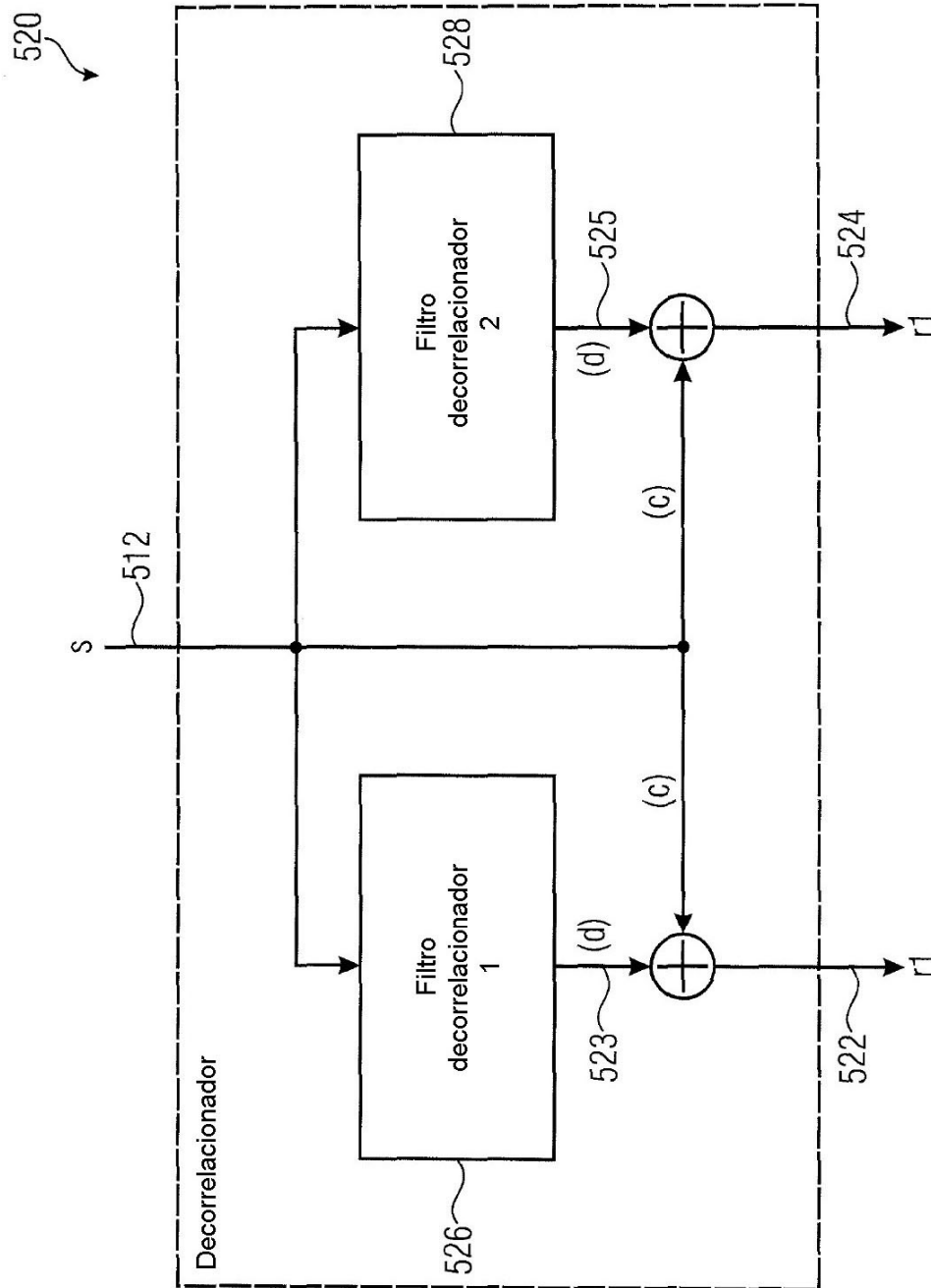


FIGURA 5

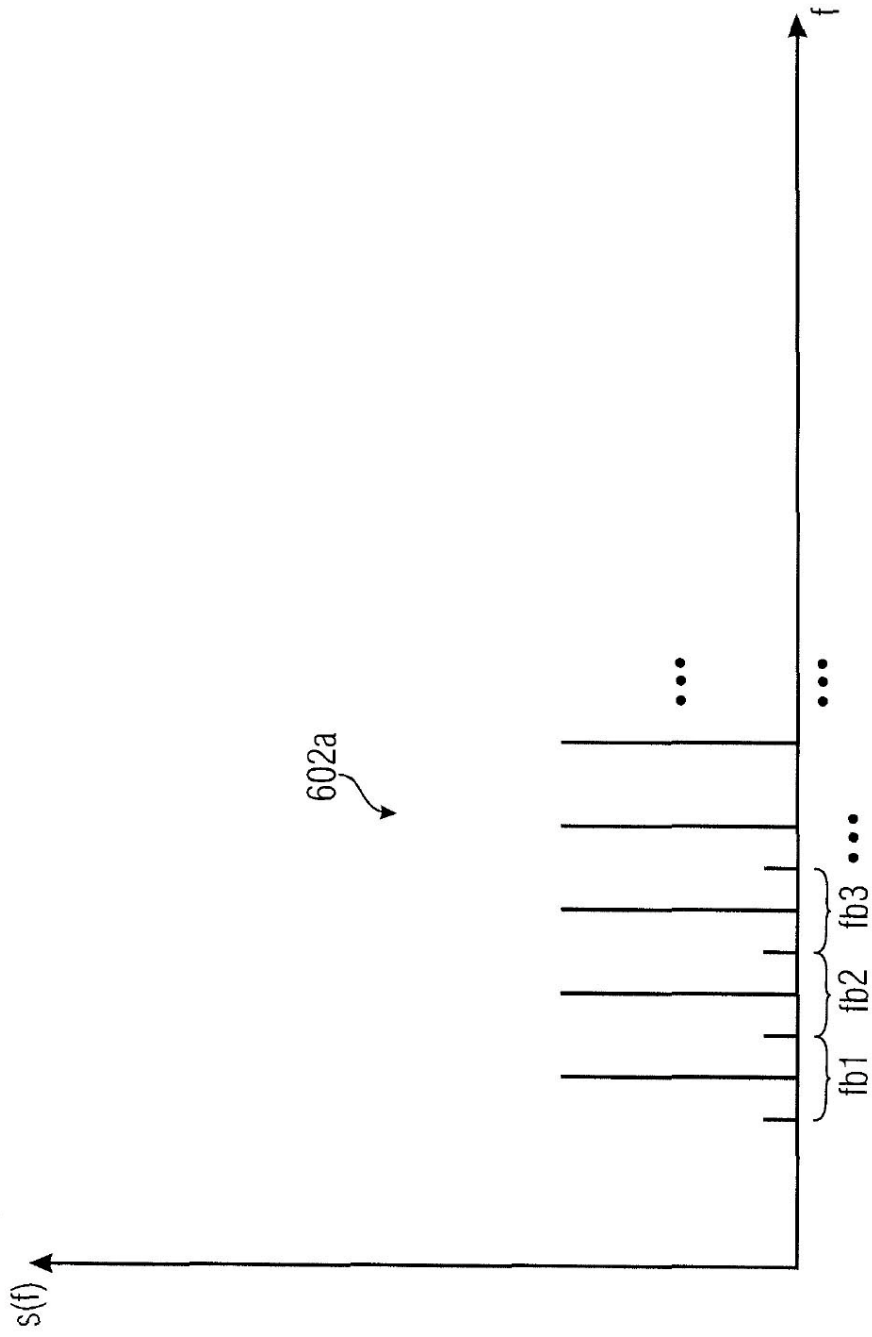


FIGURA 6A

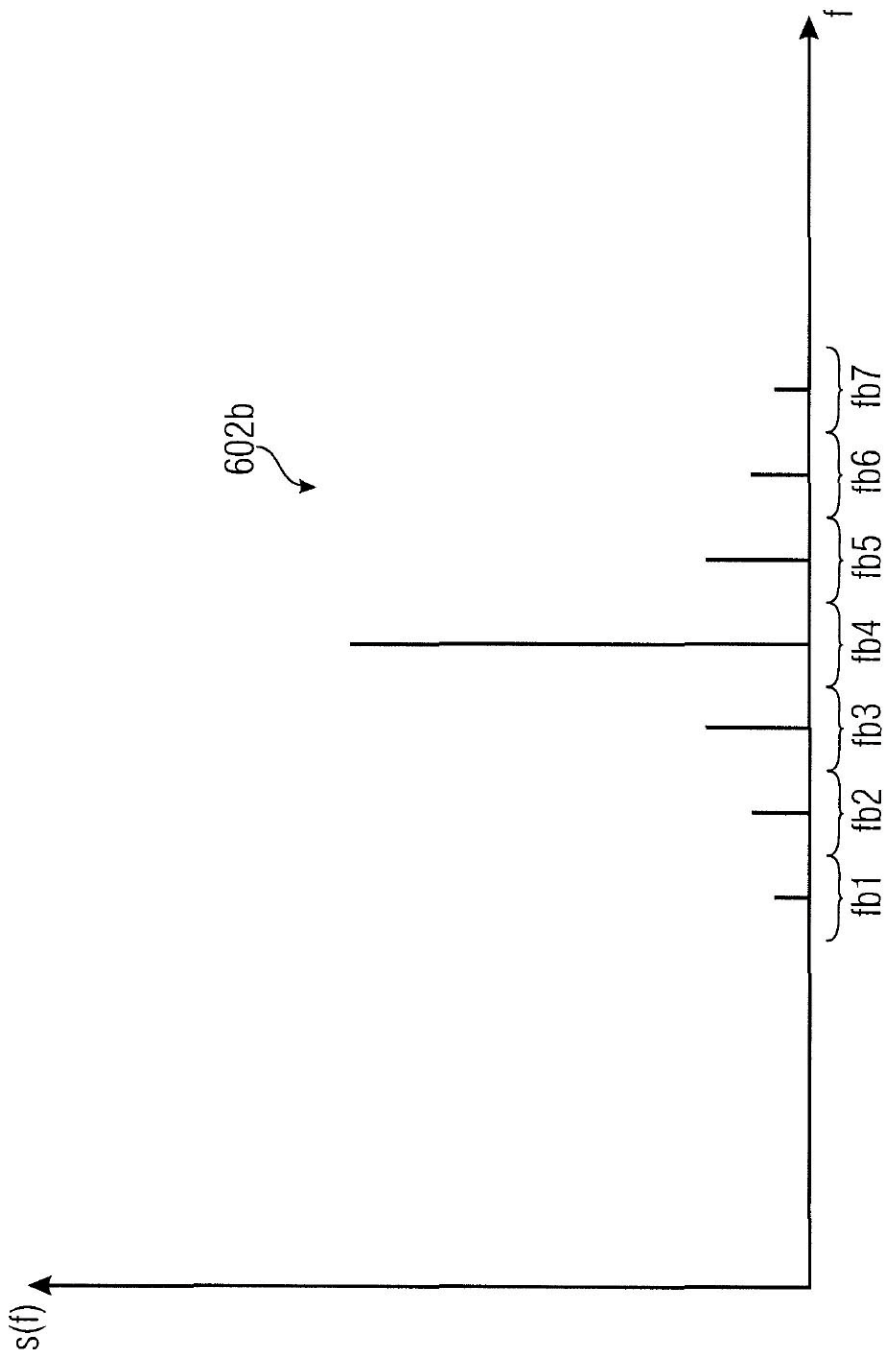


FIGURA 6B

banda de frecuencia transitoria	coeficiente de ponderación espectral (219)
8	0.1
••	••
5	0.55
4	0.6
3	0.75
2	0.9
1	1

FIGURA 7B

banda de frecuencia transitoria	coeficiente de ponderación espectral (217)
15	0
••	••
5	0.3
4	0.7
3	0.8
2	0.85
1	1

FIGURA 7A

800

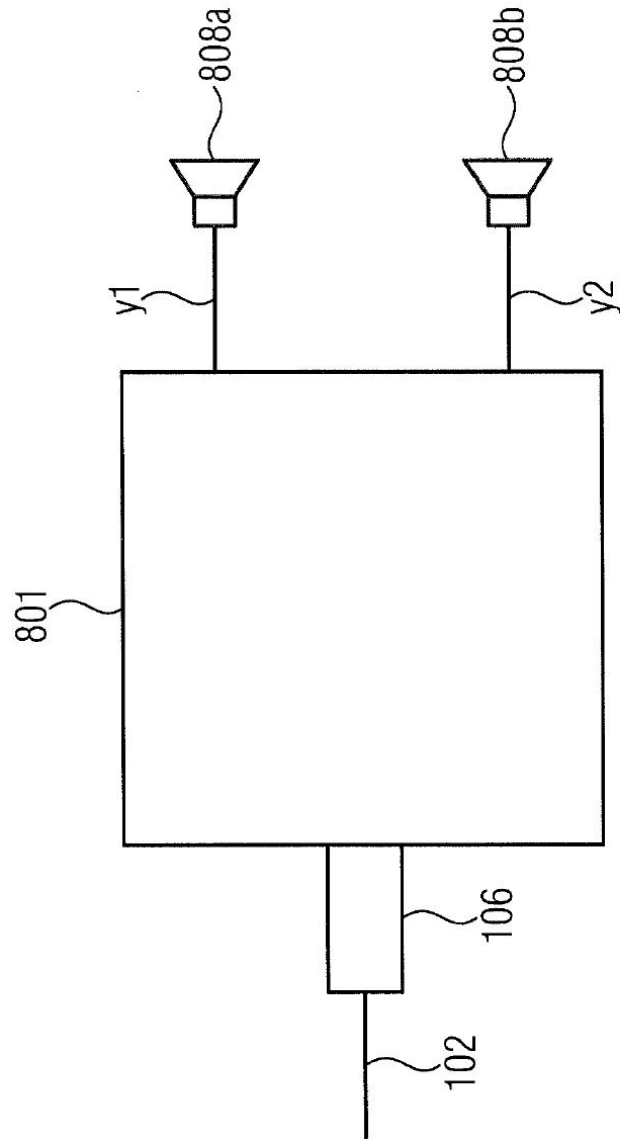


FIGURA 8

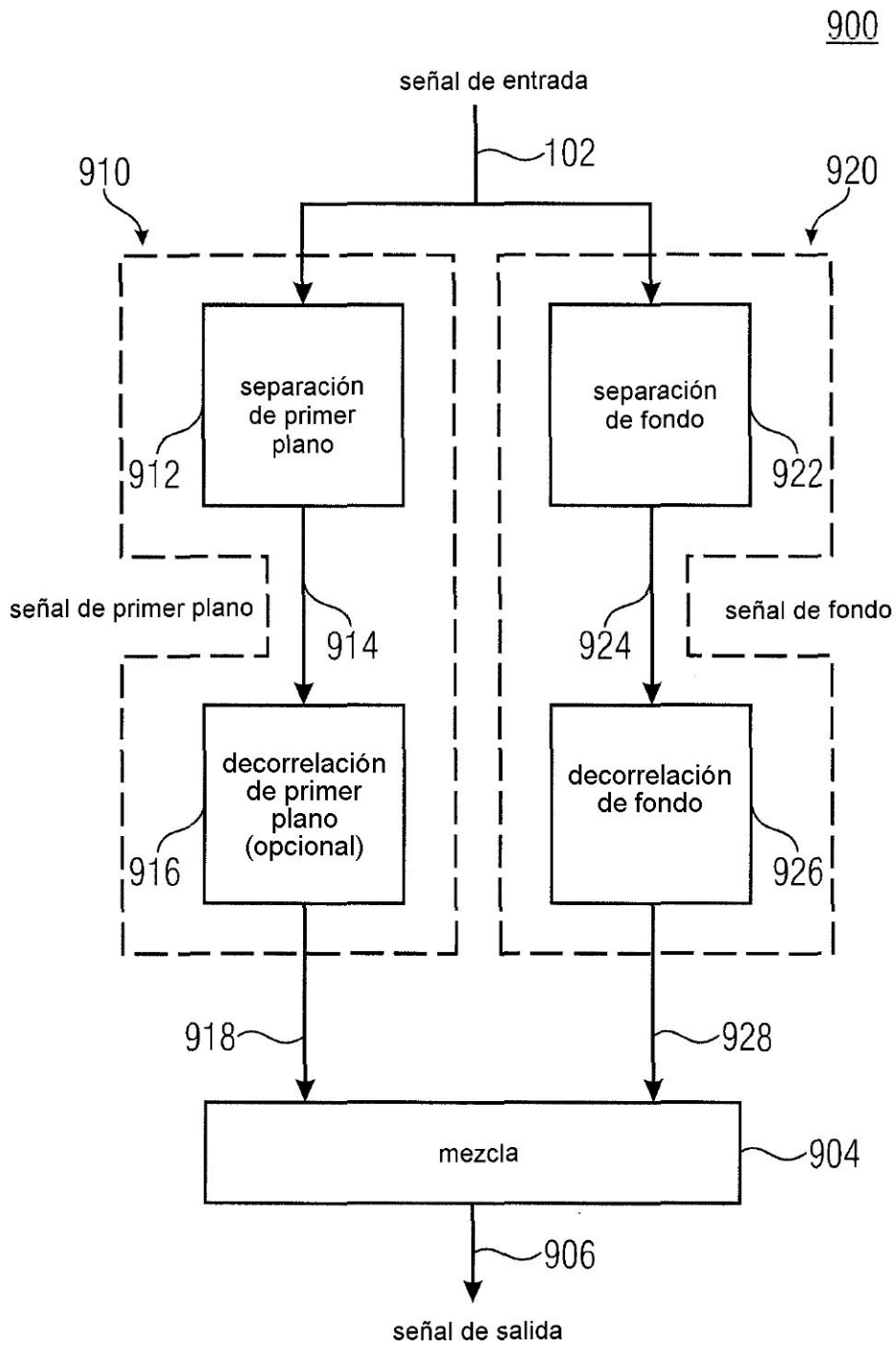


FIGURA 9A

900'

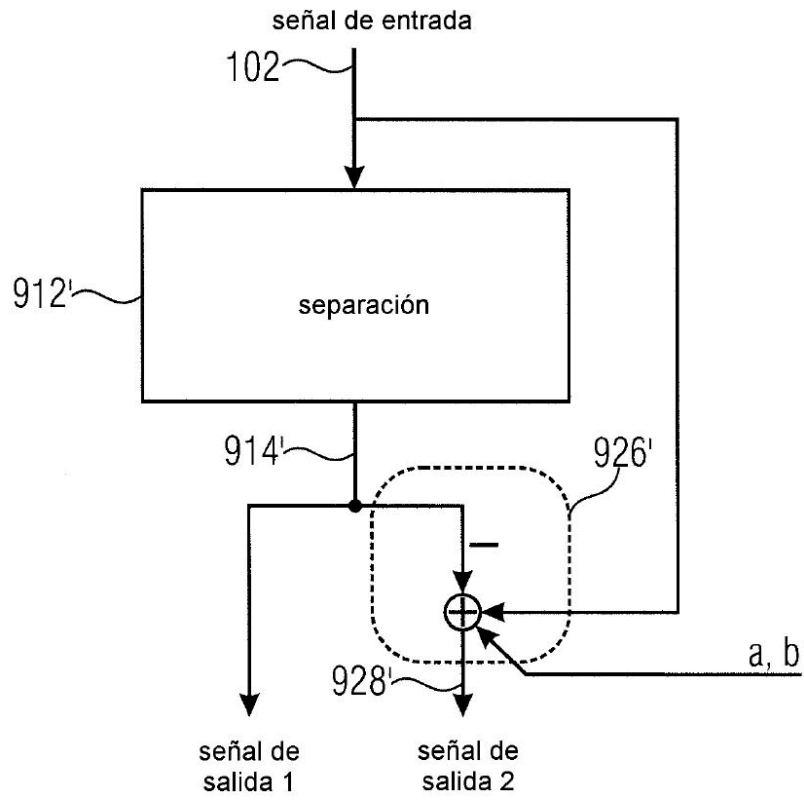


FIGURA 9B

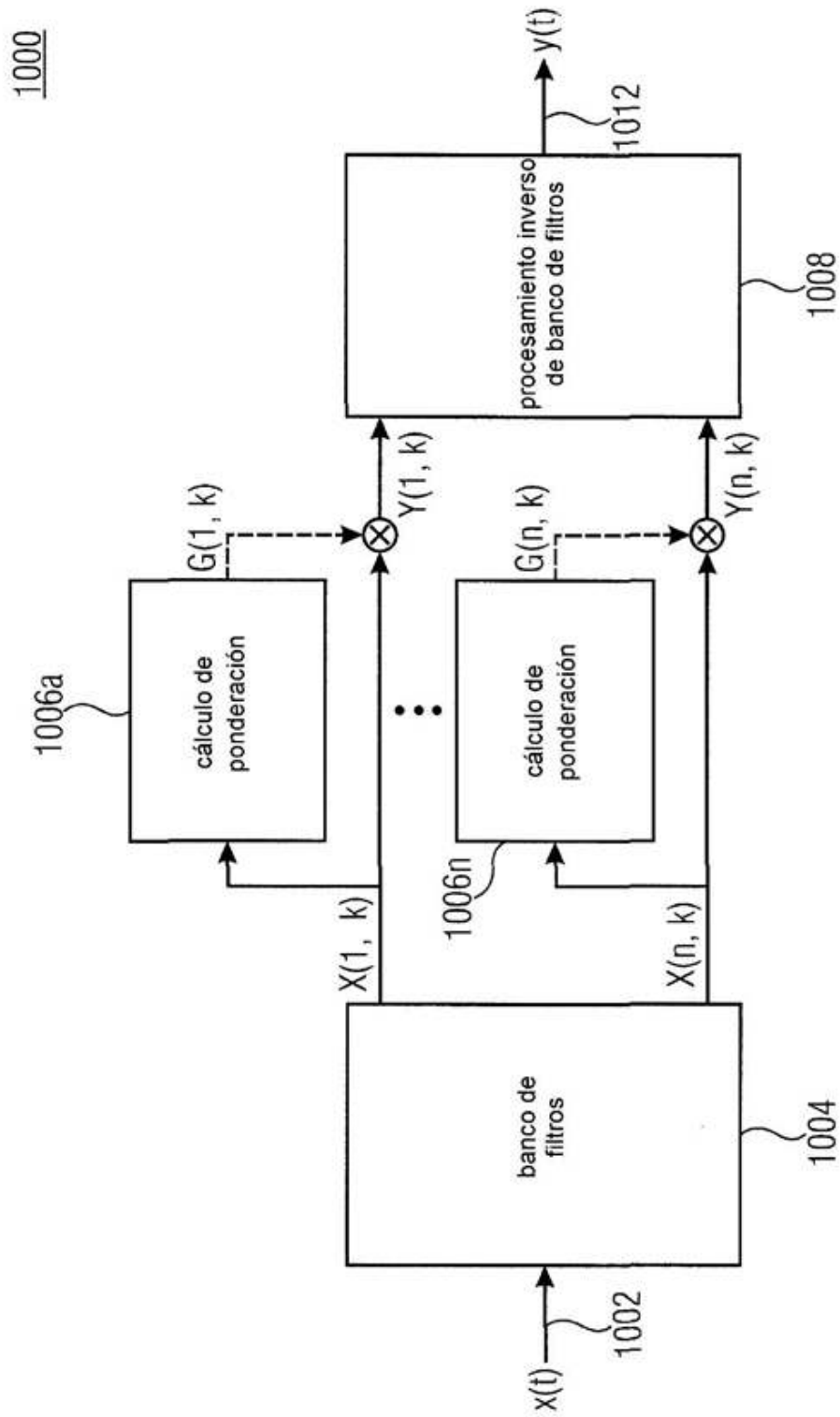


FIGURA 10

1100

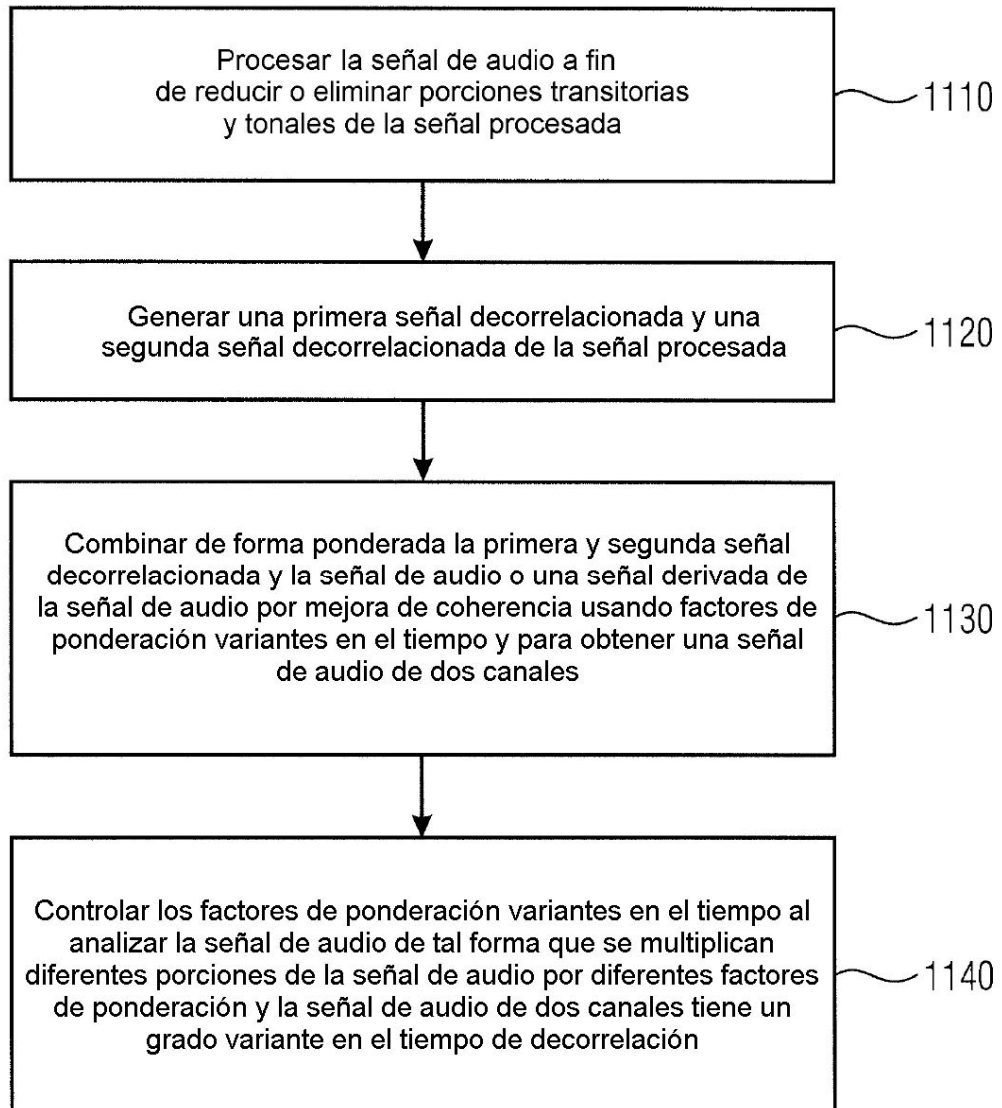


FIGURA 11

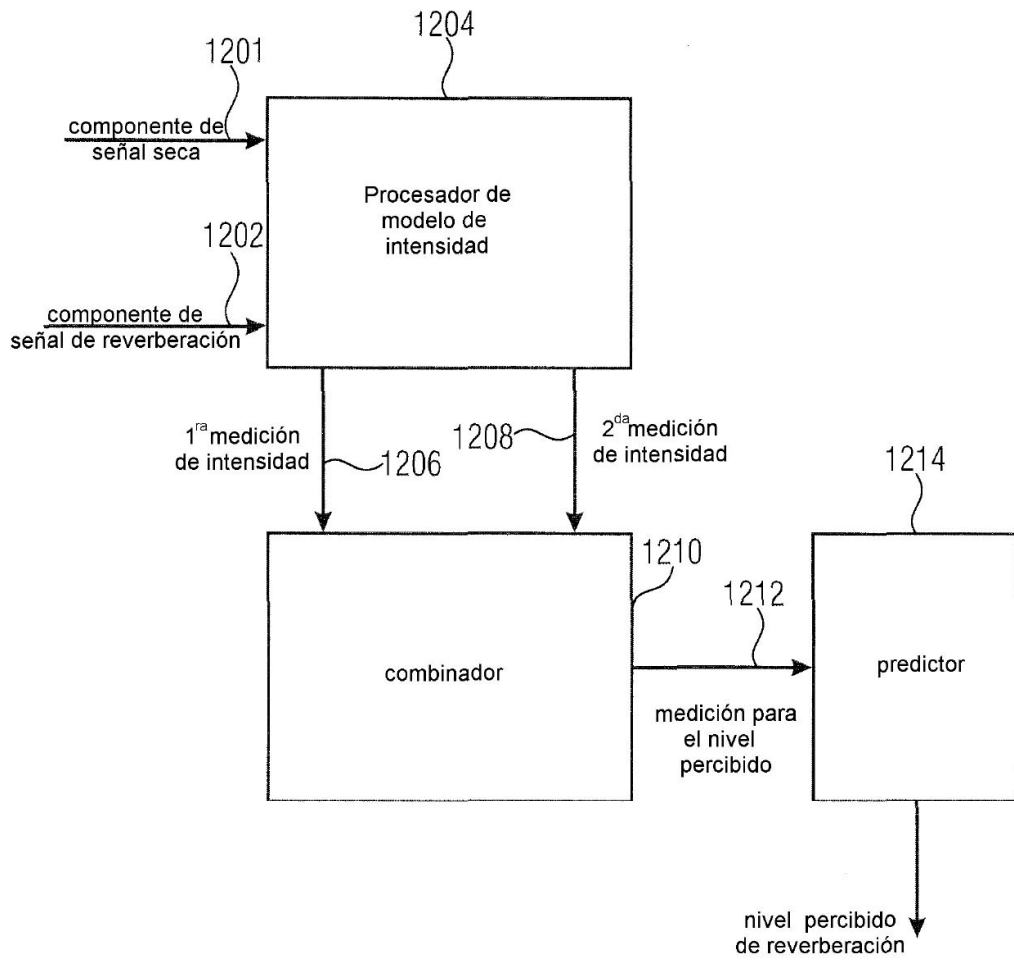


FIGURA 12

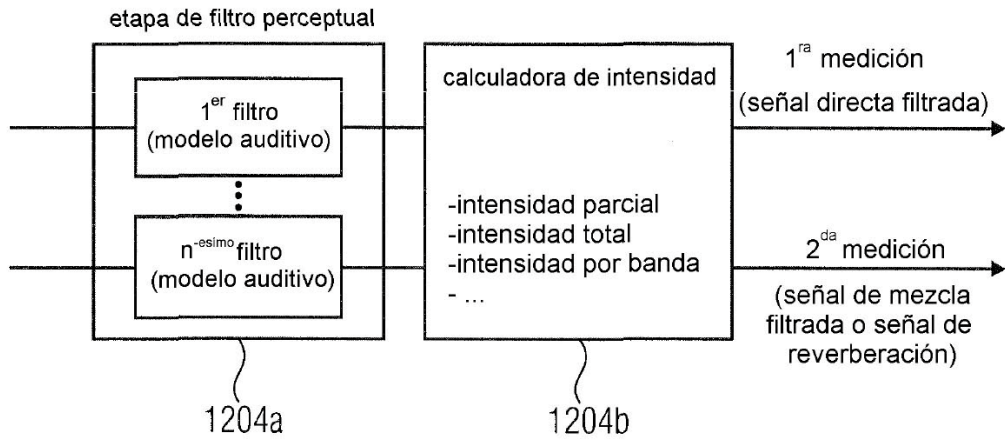


FIGURA 13A

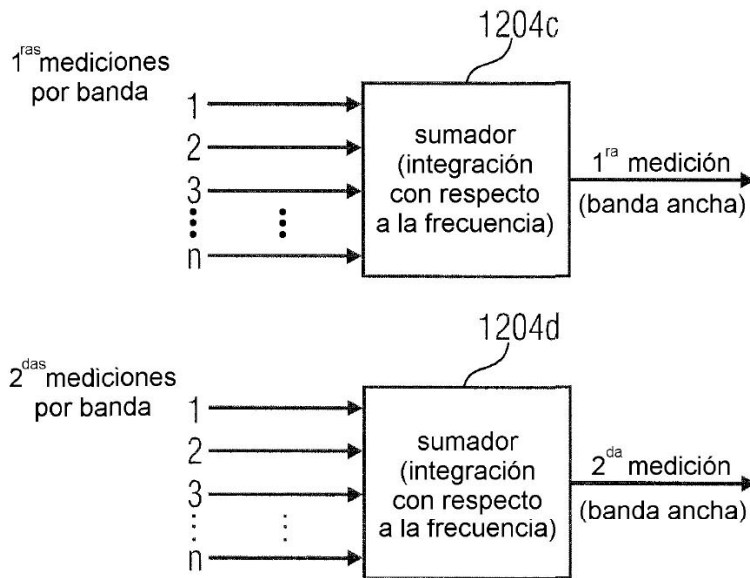


FIGURA 13B

	modelo	EST ₁	EST ₂	medición
1	parcial	r, x	x, r	$N_{r,x} - N_{x,r}$
2	total	m	x	$N_m - N_x$
3	total	r	m	$N_r - N_m$
4	total	r	x	$N_r - N_x$

FIGURA 13C

PROCESADOR DE MODELO DE INTENSIDAD

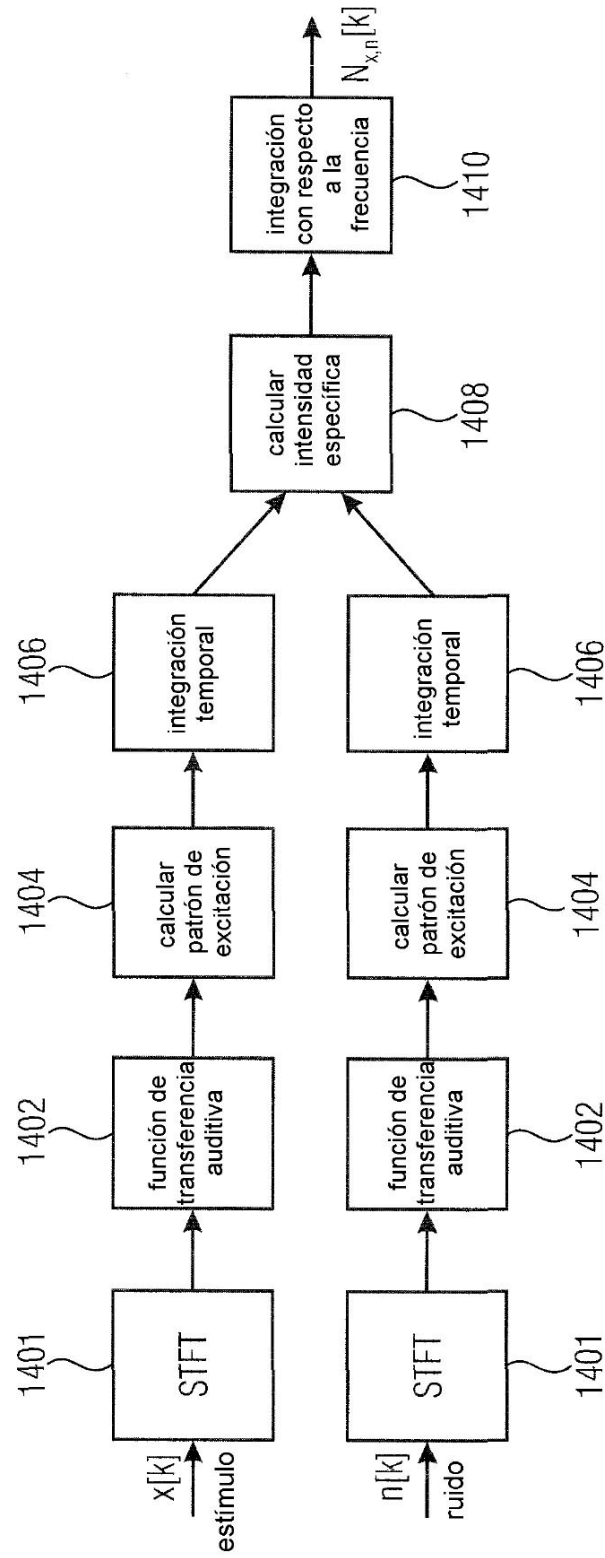


FIGURA 14