

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 982 054**

51 Int. Cl.:

G10K 15/12 (2006.01)

H04S 1/00 (2006.01)

H04S 3/00 (2006.01)

H04S 7/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **14.10.2014** **E 22211949 (7)**

97 Fecha y número de publicación de la concesión europea: **28.02.2024** **EP 4184505**

54 Título: **Espacialización del sonido con efecto de sala, optimizada para la complejidad**

30 Prioridad:

18.10.2013 FR 1360185

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
14.10.2024

73 Titular/es:

ORANGE (100.0%)
78 Rue Olivier de Serres
75015 Paris, FR

72 Inventor/es:

PALLONE, GRÉGOR Y
EMERIT, MARC

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 982 054 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

Espacialización del sonido con efecto de sala, optimizada para la complejidad

5 Esta invención hace referencia a la espacialización del sonido con efecto de sala.

La invención tiene una aplicación ventajosa pero no restrictiva en el procesamiento de señales sonoras procedentes respectivamente de L canales asociados a altavoces virtuales (por ejemplo, en una representación multicanal, o incluso en una representación de sonido envolvente, del sonido a reproducir), para una reproducción espacializada en altavoces reales (por ejemplo dos auriculares de unos cascos en reproducción binaural, o dos altavoces separados en reproducción transaural).

Por ejemplo, la señal de uno de estos canales se puede procesar para tener una primera contribución en el auricular izquierdo y una segunda contribución en el auricular derecho, en reproducción binaural, en particular aplicando a cada una de estas contribuciones una función de transferencia con efecto de sala. La aplicación de estas funciones de transferencia con efecto de sala contribuye entonces a dar al oyente una sensación de inmersión, permitiéndole prácticamente "localizar en el espacio" el altavoz virtual asociado a este canal.

En una forma de realización particular, descrita en particular en el documento FR1357299, se aplica una función de transferencia con efecto de sala a cada señal de sonido de un canal correspondiente, en el dominio del tiempo en forma de una respuesta al impulso de tipo BRIR (por "Binaural Room Impulse Response" o respuesta al impulso binaural de sala). En particular, en este documento, incorporado aquí por referencia, esta función de transferencia BRIR se construye como la combinación:

- 25 • una primera función de transferencia, específica para cada señal, y
- una segunda función de transferencia global, común a todas las señales y que caracteriza en particular un campo difuso, cuya presencia suele producirse en una sala al cabo de cierto tiempo, normalmente después de las primeras reflexiones de una onda de sonido.

Una forma de realización de este tipo permite aplicar de forma ventajosa un procesamiento común a todas las señales, que corresponde, en una realidad física, a una "mezcla" de las ondas acústicas a medida que avanzan las reverberaciones, por tanto, más allá de una duración dada (que caracteriza el inicio de la presencia del campo difuso). Una forma de realización de este tipo permite reducir la complejidad de los procesamientos de espacialización con efectos de sala en varios canales iniciales.

Además, el documento US2011/170721 A1 describe un método de procesamiento de al menos una señal de entrada a través de un conjunto de filtros binaurales de tal manera que las salidas se pueden reproducir en auriculares para proporcionar una sensación auditiva de sonido en una sala de escucha por medio de uno o más altavoces virtuales.

Sin embargo, en los módulos de espacialización anteriores a la reproducción, el objetivo sigue siendo reducir al máximo la complejidad del procesamiento de la espacialización. En efecto, por ejemplo (pero sin limitación), las señales de canal se reciben codificadas por un descodificador de compresión. Una vez descodificadas, este descodificador envía las señales de canal a un módulo de espacialización para una reproducción del sonido con efecto de sala, utilizando dos altavoces. Esta etapa de espacialización (que sigue a la descodificación de las señales recibidas) debe tener una complejidad de procesamiento reducida para no retrasar el conjunto de las etapas de descodificación y espacialización en la recepción de las señales antes de la reproducción.

La presente invención mejora la situación.

Para ello, la invención propone reducir la complejidad de la aplicación de la función de transferencia de efecto de sala, en particular reduciendo esta complejidad en el dominio espectral. En efecto, en el dominio espectral, la convolución por una función de transferencia se convierte en una multiplicación de componentes espectrales de la señal, por una parte, y de un filtro que representa la función de transferencia, por otra parte (la figura 1 comentada se analiza en detalle más adelante).

La invención parte de la ventajosa observación de acuerdo con la cual, tras su propagación directa, una onda de sonido tiende a atenuarse en las frecuencias altas debido a las reflexiones progresivas en las superficies (normalmente paredes, la cara del oyente, etc.) que absorben la onda, en particular en las frecuencias altas. Además, el propio aire absorbe las componentes espectrales de las frecuencias más altas del sonido a medida que éste se propaga. Este fenómeno es tanto más pronunciado para el campo de sonido difuso, por ejemplo, para el que no es necesario disponer de una representación frecuencial para las frecuencias muy altas (por ejemplo, por encima de una frecuencia en la gama de 5 a 15 kHz).

De este modo, es posible reducir la complejidad del procesamiento de la aplicación de la función de transferencia con efecto sala, en el dominio espectral, simplemente no teniendo en cuenta, para efectuar las multiplicaciones de

componentes espectrales mencionadas anteriormente, las componentes asociadas a frecuencias por encima de una frecuencia de corte predeterminada (por ejemplo, por encima de 5 a 15 kHz).

Más allá del alcance de las reivindicaciones, esta invención presenta un método de espacialización del sonido, que comprende la aplicación de al menos una función de transferencia de efecto de sala a al menos una señal de sonido, consistiendo dicha aplicación en multiplicar, en el dominio espectral, componentes espectrales de la señal de sonido por las componentes espectrales de un filtro correspondiente a la función de transferencia mencionada anteriormente. Cada componente espectral del filtro tiene una evolución temporal en una representación tiempo-frecuencia (según se detalla a continuación con referencia a la figura 3).

En particular, estas componentes espectrales del filtro se ignoran, para las multiplicaciones de componentes mencionadas anteriormente, más allá de una frecuencia umbral y después de al menos un instante dado en dicha representación tiempo-frecuencia. De este modo, después de este instante dado, las componentes espectrales del filtro se tienen en cuenta hasta una frecuencia de corte que se puede elegir, por ejemplo, entre 5 y 15 kHz (en función del efecto de sala que se deba aplicar y/o de la señal que se deba espacializar, según se describe a continuación). Más allá de la frecuencia de corte, ni siquiera se realiza la multiplicación, lo que matemáticamente equivale a multiplicar la señal por cero.

Este instante dado representa normalmente el momento en que una onda de sonido comienza a sufrir reverberaciones (por reflexiones sucesivas o, más tarde, por la presencia de un campo de sonido difuso). De este modo, en términos generales, en una forma de realización en la que la función de transferencia tiene en cuenta las reverberaciones en el efecto de sala (teniendo en cuenta, por ejemplo, un campo de sonido difuso), el instante dado mencionado anteriormente se puede elegir en función de dichas reverberaciones. Por ejemplo, el instante dado mencionado anteriormente puede ser posterior, en el efecto de sala, a la propagación de sonido directa con las primeras reflexiones, y corresponder entonces al inicio de la presencia de un campo de sonido difuso.

Además, se puede prever una forma de realización en la que la frecuencia umbral mencionada anteriormente disminuya en función del tiempo en dicha representación tiempo-frecuencia. Por ejemplo, si la señal se muestrea en varios bloques sucesivos, se puede prever, a modo de ejemplo, conservar las componentes espectrales presentes en la señal, en la multiplicación de las componentes, para un primer bloque y, a continuación, ignorarlas más allá de una primera frecuencia umbral para un segundo bloque que sigue al primer bloque y, a continuación, ignorarlas más allá de una segunda frecuencia umbral para un tercer bloque que sigue al segundo bloque, etc., siendo la segunda frecuencia umbral inferior a la primera.

De este modo, en términos más genéricos, en una forma de realización en la que la señal se muestrea en varios bloques sucesivos, las componentes espectrales del filtro se pueden ignorar, para la multiplicación de componentes:

- más allá de un primer umbral de frecuencia para un bloque dado,
- a continuación, más allá de un segundo umbral de frecuencia, para un bloque siguiente al bloque dado,

siendo la segunda frecuencia umbral inferior a la primera frecuencia umbral.

El bloque dado mencionado anteriormente puede incluir, por ejemplo, muestras localizadas temporalmente en instantes que corresponden a momentos en los que una onda de sonido ha sufrido una o más reflexiones, con incluso un inicio de presencia de campo de sonido difuso. El bloque que sigue a este bloque dado (inmediatamente o algunos bloques después) puede incluir, por ejemplo, muestras situadas temporalmente después o a partir del inicio de la presencia de un campo de sonido difuso.

Una forma de realización de este tipo, permite por ejemplo limitar cualquier artefacto de limitación de señal audible en las frecuencias altas para las reverberaciones, completándose esta forma de realización progresivamente en varios bloques. También permite considerar varias formas de funciones de transferencia (en lo sucesivo identificadas como $B_{mean}^k(m)$ siendo m un índice de bloque) que caracteriza un campo de sonido difuso. En efecto, es posible, por ejemplo, aplicar una función de transferencia B_{mean}^k a un bloque dado mencionado anteriormente, y aplicar una ventana de corte temporalmente progresiva (del tipo "fade out") a esta función de transferencia B_{mean}^k , para el bloque siguiente, para "terminar" la presencia del campo de sonido difuso.

Los cálculos de multiplicación también se pueden limitar más allá de un primer umbral de frecuencia, a partir del primer bloque o bloques de muestras, en función de las características de la señal (por ejemplo, su frecuencia de muestreo, o la frecuencia más alta representada en las componentes espectrales de la señal) o en función de las características de espacialización aplicadas (con, por ejemplo, una limitación de las componentes de alta frecuencia para una trayectoria acústica contralateral, según se detalla a continuación).

En este caso, la señal resultante de las reverberaciones (tras las reflexiones o en el campo difuso) no tiene normalmente componentes espectrales a una frecuencia más alta que la señal inicial. De este modo, la frecuencia umbral mencionada anteriormente no puede ser superior a esta frecuencia más alta.

5 De este modo, en términos más genéricos, en una forma de realización, se obtiene información del componente espectral de frecuencia más alta en la señal de sonido, y dicha frecuencia umbral mencionada anteriormente se elige como el mínimo entre una frecuencia umbral predeterminada (por ejemplo, entre 5 y 15 kHz) y dicha frecuencia más alta.

10 Normalmente, en una forma de realización en la que la señal de sonido se emite desde un decodificador de compresión, la información del componente espectral de frecuencia más alta puede ser proporcionada por el decodificador.

15 Del mismo modo, si la espacialización se lleva a cabo en un módulo capaz de soportar diferentes formatos de señales, en particular en lo que hace referencia a la frecuencia de muestreo de dichas señales, la frecuencia más alta mencionada anteriormente no puede ser superior a la mitad de la frecuencia de muestreo, y por lo tanto la frecuencia umbral para la implementación de la invención también se puede elegir en función de esta frecuencia de muestreo.

20 En una forma de realización en la que la señal de sonido se espacializa en al menos un primer y un segundo altavoces virtuales, asociados respectivamente a un primer y un segundo canal, se aplican respectivamente a estos primer y segundo canales unas primeras y segundas funciones de transferencia con efecto de sala, según se explicó anteriormente en la introducción (por ejemplo, adaptando las señales de los canales de sonido envolvente para pasar a una reproducción binaural o transaural). En particular, en el caso en que una de las funciones de transferencia primera y segunda aplique un efecto de trayectoria acústica ipsilateral, mientras que la otra de las funciones de transferencia primera y segunda aplique un efecto de trayectoria acústica contralateral, se puede prever la eliminación de las componentes espectrales de la señal de sonido más allá de una determinada frecuencia de apantallamiento dada. Esta frecuencia de "apantallamiento" se debe al hecho de que para un trayecto contra-lateral entre un altavoz virtual y un oído de un oyente en cuestión, la cabeza del oyente enmascara la trayectoria acústica y absorbe los tonos más altos de la onda acústica (eliminando por lo tanto las componentes espectrales asociadas a las frecuencias más altas de la onda acústica). De este modo, la frecuencia umbral mencionada anteriormente, para la función de transferencia que aplica un efecto de trayectoria contralateral, se puede elegir como un mínimo entre una frecuencia umbral predeterminada (por ejemplo, elegida entre 5 y 15 kHz) y esta frecuencia de apantallamiento. Esta forma de realización tiene la ventaja de que se puede aplicar ya al primer bloque de muestras. Por otra parte, no excluye la posibilidad de aumentar de nuevo la frecuencia umbral para el bloque siguiente, para simular una primera reflexión en una pared situada frente al oído en cuestión, siendo esta primera reflexión recibida en este auricular por una trayectoria ipsi-lateral.

35 En una forma de realización donde el método se lleva a cabo mediante un módulo de espacialización de sonido que recibe varias señales de entrada y que entrega al menos dos señales de salida, para entregar cada señal de salida, aplica una función de transferencia de efecto de sala a cada señal de entrada mediante:

- multiplicación en el dominio espectral de las componentes espectrales de cada señal de entrada con componentes espectrales del filtro correspondiente a una función de transferencia asociada con esta señal de entrada, hasta una frecuencia de corte para cada señal de entrada más allá de un instante dado asociado con la función de transferencia de esta señal de entrada, y
- suma, sobre todas las señales de entrada, de las componentes espectrales multiplicadas de este modo, frecuencia por frecuencia.

50 Esta forma de realización se describirá en detalle más adelante con referencia a las figuras 2 y 5, en particular

En cualquier caso, se entiende que la frecuencia de corte se puede elegir para que sea común a todas las señales, en una posible forma de realización, después de un instante dado que corresponde, por ejemplo, a la presencia del campo difuso.

55 De este modo, la forma de realización descrita en el documento FR1357299 introducida anteriormente puede ser ventajosa en el contexto de la invención, y en particular si cada función de transferencia aplicada a una señal tiene:

- una función de transferencia específica para esta señal, añadida a
- una función de transferencia global, común a todas las señales y representativa de la presencia de un campo difuso,

60 entonces el instante dado mencionado anteriormente puede ser común a todas las señales y corresponder, por ejemplo, al inicio de la presencia de un campo de sonido difuso.

65

En una forma de realización en la que las señales comprenden bloques sucesivos de muestras, de los mismos tamaños entre las señales, se proporciona al menos un instante determinado para limitar la toma en consideración de las componentes de frecuencia hasta una frecuencia de corte, estando este instante dado situado temporalmente al comienzo de un bloque diferente de un primer bloque en una sucesión de bloques. Este instante dado se produce por lo tanto después de una propagación directa, y en el momento de las reflexiones de sonido o de la presencia de un campo difuso.

En una forma de realización que repite lo anterior según una fórmula ventajosa pero no restrictiva, cada una de las señales de salida mencionadas anteriormente puede estar dada por la aplicación de una relación del tipo:

$$O^k = \sum_{l=1}^L (I(l) *_{|0;\dots;f^k(l)} A^k(l)) + \sum_{m=1}^M \left(z^{-iDDm} \cdot G(I(l)) \cdot \sum_{l=1}^L \left(\frac{1}{W^k(l)} \cdot I(l) \right) \right) *_{|0;\dots;f^k(m)} B_{mean}^k(m)$$

- siendo O^k una señal de salida, y siendo k el índice relativo a una señal de salida,
- siendo $l \in [1; L]$, el índice relativo a una de dichas señales de entrada, siendo L el número de señales de entrada, y siendo $I(l)$ una de dichas señales de entrada,
- siendo $A^k(l)$ una función de transferencia con efecto de sala específica para una señal de entrada,
- siendo z^{-iDDm} una aplicación de retardo, contada en número de bloques de muestras, correspondiente a una diferencia de tiempo entre una emisión de sonido en una sala correspondiente al efecto de sala, y un inicio de presencia de campo difuso en esta sala, correspondiendo el índice m a un número de bloques de muestras de duración correspondiente a este retardo, siendo M el número total de bloques que dura una función de transferencia en una representación tiempo-frecuencia,
- siendo $B_{mean}^k(m)$ una función de transferencia global, con efecto de sala, común a las señales de entrada; esta función puede variar con el índice de bloque m por ejemplo para aplicar una disminución de ganancia (del tipo ventana "fade out") al último bloque M.
- siendo $W^k(l)$ un peso de ponderación seleccionado y $G(I(l))$ una ganancia de compensación de energía predeterminada, estos parámetros se pueden establecer en el valor 1 o por el contrario, permiten gestionar las ponderaciones de presencia en el campo difuso dependiendo de las señales de entrada
- el signo "." indica la multiplicación,
- designando el signo " $*_{|0;\dots;f^k(l)}$ " el operador de convolución en un número limitado de frecuencias y que va de una frecuencia inferior a una frecuencia máxima $f^k(l)$ que es función al menos de la señal de entrada del índice l, y
- designando el signo " $*_{|0;\dots;f^k(m)}$ " el operador de convolución en un número limitado de frecuencias y que va de una frecuencia inferior a una frecuencia $f^k(m)$ que es función del bloque de muestras de índice m

Esta forma de realización se describirá con más detalle a continuación con referencia a la figura 5, que también ilustra, en un ejemplo de forma de realización, un posible algoritmo de un programa informático que sería ejecutado por un procesador de un módulo de espacialización que opera el método en el sentido de la invención. A este respecto, la presente descripción también tiene por objetivo, más allá del alcance de las reivindicaciones, en general un programa de ordenador que comprende instrucciones para la implementación del método anterior, cuando son ejecutadas por un procesador.

La presente invención tiene por objetivo un módulo de espacialización del sonido, que comprende medios de cálculo para aplicar al menos una función de transferencia de efecto de sala a al menos una señal de sonido de entrada, equivaliendo dicha aplicación a multiplicar, en el dominio espectral, componentes espectrales de la señal de sonido por las componentes espectrales de un filtro correspondiente a dicha función de transferencia, comprendiendo cada componente espectral del filtro una evolución temporal en una representación de tiempo-frecuencia. En particular, dichos medios de cálculo se configuran para ignorar dichas componentes espectrales del filtro para dichas multiplicaciones de componentes, más allá de una frecuencia umbral y después de al menos un instante dado en dicha representación tiempo-frecuencia.

Este módulo se puede integrar en un dispositivo de decodificación por compresión o, de forma más general, en un sistema de reproducción.

La figura 6 muestra dicho módulo de espacialización SPAT de este tipo, junto con un dispositivo de decodificación DECOD que recibe señales codificadas por compresión $I(l)$ (donde $l = 1, \dots, L$) de una red RES, en el ejemplo

mostrado, y las descodifica antes de la reproducción transmitiendo las señales descodificadas $l(l)$ (donde $l = 1, \dots, L$) al módulo de espacialización. En el ejemplo representado, este último comprende una interfaz de entrada IN para recibir las señales descodificadas, así como medios de cálculo tales como un procesador PROC y una memoria de trabajo MEM que cooperan con las interfaces IN/OUT para espacializar las señales $l(l)$ y suministrar a través de la interfaz de salida OUT únicamente dos señales O^d y O^g que tiene por objetivo alimentar los respectivos auriculares de unos cascos CAS.

Otras características y ventajas de la invención se desprenderán con el examen de la descripción detallada que figura a continuación y de los dibujos adjuntos en los que:

- la figura 1 ilustra una forma de realización general del método utilizado por el módulo de acuerdo con la invención;
- la figura 2 ilustra un ejemplo de aplicación del método de acuerdo con una forma de realización en la que las funciones de transferencia tienen la forma de una combinación de dos funciones de transferencia, una de las cuales se aplica con un retardo a la señal que se va a procesar;
- la figura 3 muestra un ejemplo de representación tiempo-frecuencia de una función de transferencia con frecuencias de corte (o "frecuencias umbral" mencionadas anteriormente) que varían, en particular, en función del tiempo;
- la figura 4 ilustra un organigrama correspondiente a un posible algoritmo general del programa informático mencionado anteriormente,
- la figura 5 muestra una forma de realización particular del modo mostrado en la Figura 2, pero a lo largo de más de dos bloques de tiempo sucesivos, con una evolución de la función de transferencia $D_{msaan}^k(m)$ que representa el campo difuso, en función de los bloques m ;
- la figura 6 ilustra un ejemplo de módulo de espacialización en el sentido de la invención;
- la figura 7 ilustra esquemáticamente los altavoces virtuales y el efecto de sala por aplicar una función de transferencia adecuada, con limitación de las componentes de frecuencia de esta función de transferencia hasta una frecuencia de corte adecuada.

Antes de describir la figura 1 y el principio general de la invención, se hace referencia a la figura 7 para explicar los fenómenos físicos subyacentes a la presente invención.

Varios altavoces virtuales rodean, en el ejemplo mostrado, la cabeza TE de un oyente. A cada uno de los altavoces virtuales HPV se le suministra inicialmente una señal $l(l)$ con $l \in [1;L]$ por ejemplo previamente descodificada según se indicó anteriormente con referencia a la figura 6. La disposición de los altavoces virtuales puede referirse a una representación multicanal o también de sonido envolvente de las señales $l(l)$ a procesar para reproducirlas conjuntamente de forma espacializada con un efecto de sala en unos cascos con auriculares CAS (figura 6). Para ello, se suele aplicar a cada señal una función de transferencia de efecto de sala para cada señal de auricular que se va a suministrar O^k , donde $k = d$ (para la derecha), g (para la izquierda). De este modo, con referencia a la figura 7, para cada altavoz virtual HPV se considera la trayectoria acústica (TIL ipsilateral en el ejemplo mostrado) desde el altavoz HPV hasta el oído izquierdo OG, así como la trayectoria acústica (TCL contralateral en el ejemplo mostrado) desde el altavoz HPV hasta el oído derecho OD, así como las reflexiones en las paredes MUR (trayectoria RIL) y, por último, un campo difuso tras varias reflexiones. Con cada reflexión, se supone que la onda acústica se atenúa en las frecuencias más altas.

De este modo, con referencia a la figura 3 relativa a una representación tiempo-frecuencia de una función de transferencia adaptada para el altavoz virtual HPV mostrado en la figura 7, ya es evidente que la cabeza del oyente enmascara de forma natural la trayectoria contra-lateral y las frecuencias más altas que se deben considerar para la función de transferencia específica del oído derecho OD son inferiores a las que se deben considerar para la función de transferencia específica del oído izquierdo OG (que se enfrenta al altavoz virtual HPV a lo largo de un trayecto ipsilateral). De este modo, considerando un primer bloque de tiempo de 0 a N-1, identificado como $m=0$, la frecuencia máxima $F_c^d(0)$ de un filtro que represente la función de transferencia específica del oído derecho puede ser inferior a la frecuencia máxima $F_c^g(0)$ de un filtro que represente la función de transferencia específica del oído izquierdo. Un diseñador de un filtro de este tipo puede de este modo limitar las componentes de su filtro para el oído derecho hasta la frecuencia de corte $F_c^d(0)$ (correspondiente a una frecuencia de apantallamiento de la cabeza), aunque la señal que se debe procesar $l(l)$ pueda tener componentes espectrales más elevados y hasta la frecuencia $F_c^g(0)$ como mínimo.

Acto seguido, después de las reflexiones, la onda acústica tiende a atenuarse en las frecuencias altas, lo que respeta tanto la representación tiempo-frecuencia de la función de transferencia para el oído izquierdo, como para el oído

derecho, para los instantes N a $2N-1$, correspondientes al bloque siguiente identificado como $m=1$. De este modo, un diseñador de filtros que representen estas funciones de transferencia puede prever limitar las componentes del filtro para el oído derecho hasta la frecuencia de corte $F_c^d(1)$ y para el oído izquierdo hasta la frecuencia de corte $F_c^g(1)$. En una forma de realización ilustrada en particular en la figura 5, se puede considerar que en el bloque $m=1$, la función de transferencia caracteriza normalmente un campo difuso tanto para el oído derecho como para el oído izquierdo y, por lo tanto, se puede establecer (posiblemente pero no de forma restrictiva) que $F_c^d(1)=F_c^g(1)$.

Acto seguido, en presencia de un campo difuso con atenuación global del sonido ("fade out"), la onda acústica tiende a atenuarse más en las frecuencias altas, lo que de nuevo respeta tanto la representación tiempo-frecuencia de la función de transferencia para el oído izquierdo, como para el oído derecho, en la figura 3, para los instantes $2N$ a $3N-1$, correspondientes al bloque identificado como $m=2$. De este modo, un diseñador de filtros que representen estas funciones de transferencia puede prever limitar las componentes del filtro para el oído derecho hasta la frecuencia de corte $F_c^d(2)$ y para el oído izquierdo hasta la frecuencia de corte $F_c^g(2)$.

Obsérvese que bloques más cortos permitirían hacer variar con más precisión la frecuencia más alta a considerar, por ejemplo para tener en cuenta un reflejo RIL inicial para el que la frecuencia más alta aumenta para el oído derecho (líneas discontinuas alrededor de $F_c^d(0)$ en la figura 3) en los primeros instantes del bloque $m=0$.

De este modo, es posible no tener en cuenta todas las componentes espectrales de un filtro que representa una función de transferencia, en particular más allá de una frecuencia de corte F_c . Por lo tanto, es ventajoso procesar la aplicación de la función de transferencia en el dominio espectral. En efecto, la convolución de una señal $I(f)$ por una función de transferencia se convierte, en el dominio espectral, en una multiplicación de las componentes espectrales de la señal $I(f)$ por las componentes espectrales del filtro que representa la función de transferencia en el dominio espectral, y, en particular, esta multiplicación sólo se puede efectuar hasta una frecuencia de corte, que es función de un bloque dado, por ejemplo, y de la señal que se deba procesar.

De este modo, con referencia a la figura 1, L señales de entrada $1(1)$, $I(2)$, ..., $I(L)$ se transforman en el dominio de la frecuencia en las etapas TF11, TF12, ..., TF1L respectivamente. Alternativamente, señales de entrada de este tipo pueden estar ya disponibles en forma de frecuencia (por ejemplo, a partir del descodificador).

En la etapa BA11, se almacena en memoria una respuesta al impulso de espacialización completa (normalmente del tipo BRIR por "Binaural Room Impulse Response") en forma temporal correspondiente a la señal $1(1)$ del canal 1. En la etapa TFA11, esta respuesta al impulso se transforma en forma de frecuencia para obtener un filtro correspondiente en el dominio espectral. En una forma de realización ventajosa, el filtro se almacena en su forma espectral para evitar repetir el cálculo de la transformada. Acto seguido, este filtro se multiplica por la señal de entrada en forma de frecuencia del canal 1 (lo que equivale a una convolución en el dominio del tiempo). Se obtiene de este modo la señal espacializada del canal 1 para la señal $1(1)$.

Se realizan las mismas operaciones para los otros $L-1$ canales. Se obtiene de este modo un total de L canales espacializados. Acto seguido, estos canales se suman para obtener una única señal de salida que representa los L canales, y volvemos al dominio del tiempo en la etapa ITF11, para suministrar una de las señales O^k (con $k=d, g$) que alimentan un auricular. Se realiza un procesamiento similar para el otro auricular. En una forma de realización descrita en detalle a continuación con referencia a las figuras 2 y 5, los L canales espacializados no son accesibles de forma independiente antes de la suma: la señal de salida única se construye sumando progresivamente cada canal espacializado con la señal de salida anterior.

Estas operaciones se realizan para cada señal de salida O^k que se desee construir. Normalmente, en el caso de la reproducción binaural, estas etapas se realizan dos veces, una para la señal de salida que tiene por objetivo alimentar el auricular izquierdo de unos cascos y otra para la señal de salida que tiene por objetivo alimentar el auricular derecho de los cascos. De este modo, el resultado final son dos señales espacializadas, O^d y O^g , cada una correspondiente a un oído.

Las L señales de entrada pueden corresponder normalmente a los L canales de un contenido de audio multicanal destinado a alimentar altavoces ("virtuales"). Las L señales de entrada pueden corresponder, por ejemplo, a las L señales de sonido envolvente de un contenido de audio con representación envolvente.

Con referencia ahora a la figura 2 que ilustra una implementación en el sentido de la invención, se repite el principio de una espacialización de L canales como se muestra en la figura 1. Sin embargo, la presentación en la figura 2 se simplifica en el sentido de que las L señales de entrada se combinan en un único canal $I(f)$. De este modo, las L señales de entrada $1(1)$, $I(2)$, ..., $I(L)$ se transforman en el dominio de la frecuencia en la etapa S21. Según se indicó anteriormente, señales de entrada de este tipo pueden alternativamente estar ya disponibles en forma de frecuencia. En la etapa S22, una respuesta al impulso de espacialización $A^k(f)$ (normalmente del tipo BRIR) correspondiente a la señal $I(f)$ del canal f se transforma al dominio espectral para obtener un filtro de frecuencia. Esta respuesta al impulso $A^k(f)$ está incompleta en la representación de la figura 2 porque corresponde a un primer bloque temporal de muestras $m=0$. Como se indicó anteriormente, esta respuesta al impulso puede estar ya disponible en forma frecuencial. Acto seguido, los componentes de este filtro se multiplican por la señal espectral del canal correspondiente $I(f)$. Esta

multiplicación se parametriza (según se indica a continuación con referencia a la figura 4) de modo que se ignoren determinados componentes de frecuencia, en el sentido de la invención. Normalmente, las componentes de frecuencia más altas se ignorarán para limitar la complejidad de los cálculos. En las figuras 2 y 5, la multiplicación de las componentes limitadas a una frecuencia de corte se indica mediante el signo: χ

5 Se define una frecuencia de corte $f_{cA(l)}$, por encima de la cual se ignoran las componentes de frecuencia (por ejemplo, la frecuencia máxima representada en la señal del canal $l(l)$, o la mitad de su frecuencia de muestreo). Además, esta frecuencia de corte es específica para cada filtro y para cada bloque (disminuye, por ejemplo, para los bloques $m=1$, $m=2$). Como, en este caso, los filtros son específicos para cada señal de entrada y cada oído, la frecuencia de corte es específica para una señal de entrada, un oído (y, por tanto, una señal de salida) y un bloque temporal.

10 La señal espacializada para el canal l está entonces disponible para este primer bloque temporal. Estas operaciones se realizan para todos los L canales: $l = 1, \dots, L$. Se obtienen de este modo L canales espacializados. Acto seguido, estos canales se suman en la etapa S23 para obtener una única señal que representa los L canales en el primer bloque temporal.

15 En la práctica, la suma se realiza de manera especial, ya que tiene en cuenta un retardo en los canales para caracterizar las reverberaciones (reflexiones y campo difuso), según se detalla a continuación. En efecto, en una forma de realización, los L canales espacializados no son accesibles de forma independiente antes de la suma: la señal de salida única se construye sumando progresivamente cada canal espacializado con la señal de salida anterior. Para ello, en la etapa DBD, las señales de entrada $l(l)$ se retrasan un cierto retardo dado por $z^{-iDD.m}$ específico para cada bloque $m = 1, \dots, M$. Nótese que para el primer bloque, el retardo m es cero. En el caso de una representación frecuencial, este retardo corresponde generalmente al tamaño de una trama de señal procesada para el primer bloque, y se puede interpretar como la toma del bloque de entrada anterior en su forma frecuencial.

20 En la etapa S24, una respuesta al impulso de espacialización incompleta $B^k_m(l)$ (normalmente del tipo BRIR) correspondiente a la señal $l(l)$ del canal 1 se transforma en el dominio espectral para obtener un filtro de frecuencia. Esta respuesta al impulso $B^k_m(l)$ está incompleta porque corresponde a un segundo bloque temporal de muestras (y, a continuación, un tercer bloque y así sucesivamente, para $m = 1, \dots, M$). Según se indicó anteriormente, esta respuesta al impulso puede estar ya disponible en forma de frecuencia. Aplicando el principio descrito en el documento FR1357299, es posible reducir la complejidad del procesamiento planteando $B^k_m(1) = \dots = B^k_m(l) = \dots = B^k_m(L) = B^k_{mean}(m)$ y hacer finalmente que esta función de transferencia dependa únicamente del bloque m considerado (campo difuso principal, o campo difuso secundario con atenuación "fade out") y del oído k . Del mismo modo, el campo difuso no depende de los canales y es posible fijar la frecuencia de corte f_c como idéntica para cada canal (pero que se puede reducir aún más de un bloque al siguiente, como vimos anteriormente con referencia a la figura 3). Esta forma de realización se muestra en la figura 5.

25 Con referencia de nuevo a la figura 2, este filtro $B^k_m(l)$ se multiplica acto seguido por la señal $l(l)$ del canal l . Las frecuencias de corte son diferentes para este segundo bloque temporal. Según se muestra con referencia a la figura 3, las mediciones muestran que las frecuencias altas se atenúan más en los bloques de tiempo distantes (correspondientes a sonidos difusos y reverberaciones múltiples). Por lo tanto, las frecuencias de corte de estos bloques distantes pueden ser más bajas que las de los primeros bloques. Sin embargo, cuanto más baja sea la frecuencia de corte, más limitado será el número de operaciones. De este modo, la complejidad de los cálculos se reduce de forma ventajosa.

30 Se realizan las mismas operaciones para los L canales y se repiten las operaciones de multiplicación del filtro en las señales espectrales progresivamente retardadas sumando las contribuciones en la etapa S25 repetida para cada retardo m hasta que se obtenga una única señal que representa los L canales sobre el conjunto de los bloques de tiempo m considerados. La señal de salida única se construye sumando progresivamente cada canal espacializado con la señal de salida anterior, según se verá ahora con referencia a la figura 4.

35 Por último, volvemos al dominio temporal en la etapa S26 para obtener una señal de salida que tiene por objetivo alimentar uno de los auriculares.

40 Con referencia a la figura 4, se describe ahora un método de espacialización para un bloque temporal dado (por ejemplo, para el bloque que representa el campo de sonido directo con valores en el intervalo temporal $[0; N-1]$) y para una señal correspondiente, por ejemplo, al oído derecho. Por supuesto, se aplica el mismo método para la señal correspondiente al oído izquierdo. La distinción entre los dos oídos se introduce mediante la aplicación de filtros específicos a cada uno de ellos.

45 En la etapa S40, la señal de salida S se pone a 0. Esta señal de salida se expresa en el dominio de la frecuencia. Tiene un tamaño limitado de una longitud mayor que la frecuencia de corte $f_c(l)$. Por ejemplo, esta señal se define en $[0; fs(l)/2]$, siendo $fs(l)$ la frecuencia de muestreo de esta señal $l(l)$. Una primera variable de conteo i también se inicializa en 1. Esta primera variable de conteo identifica una de las señales de canal $1(1), l(2), \dots, l(l), \dots, l(L)$ sobre el bloque temporal $[0; N-1]$ para el oído derecho. En la etapa S41, una segunda variable de conteo j se inicializa en 0. Esta

segunda variable de conteo identifica una componente de frecuencia de una señal $I(l)$ sobre el bloque temporal $[0; N-1]$ para el oído derecho.

5 En la etapa S42, se almacena en memoria el coeficiente $c_{BRIR}(j; l)$. Este coeficiente corresponde a la componente de frecuencia j del filtro $BRIR(l)$ sobre el bloque temporal $[0; N-1]$ para el oído derecho. Del mismo modo, se almacena en memoria el coeficiente $c_i(j; l)$. Este coeficiente corresponde a la componente de frecuencia j de la señal $I(l)$ sobre el bloque temporal $[0; N-1]$ para el oído derecho. De este modo, los coeficientes $c_{BRIR}(j; l)$ y $c_i(j; l)$ corresponden a la misma componente de frecuencia (identificada por la variable j) y se podrán multiplicar posteriormente término a término (etapa S44).

10 En la prueba T47, comprobamos que la frecuencia correspondiente a la variable j es inferior (por ejemplo estrictamente) a la frecuencia de corte $fc(l)$. Esta frecuencia de corte corresponde a la frecuencia de corte de la señal $I(l)$ para el bloque temporal $[0; N-1]$ para el oído derecho. Si la frecuencia j es menor que la frecuencia de corte $fc(l)$, se realiza la etapa S44.

15 En la etapa S44, se calcula un valor $MULT(j)$ correspondiente a la multiplicación de los coeficientes $c_{BRIR}(j; l)$ y $c_i(j; l)$. Estos coeficientes se multiplican término a término porque corresponden al mismo componente de frecuencia j (para el mismo canal, en el mismo bloque y para el mismo oído).

20 En la etapa S45, este valor $MULT(j)$ se incrementa a la señal S en la posición de frecuencia j .

25 De este modo, se construye paso a paso una señal S que comprende (al final del bucle de longitud $fc(l)$) todas las componentes de frecuencia hasta la frecuencia de corte $fc(l)$ (para esta señal $I(l)$ en el bloque $[0; N-1]$ y para un oído derecho). Como al principio del bucle de la figura 4, todas las componentes ya se han inicializado a 0, al final del bucle, finalmente hemos llenado un búfer (inicialmente cero) hasta la frecuencia de corte para construir sucesivamente la señal S . De este modo, cada multiplicación $MULT(j)$ de coeficientes se añade paso a paso a la señal S que se está construyendo.

30 En la etapa S46, se incrementa la variable j y se repite la etapa S42. Si la variable j es mayor que (por ejemplo o igual a) la frecuencia de corte $fc(l)$, se realiza la prueba T48. De este modo, la señal S se ha llenado en el intervalo $[0; fc(l)]$.

35 Como mencionó anteriormente, esta señal se puede definir en un intervalo mayor que $[0; fc(l)]$ (por ejemplo $[0; fs(l)/2]$). Además, esta señal se había inicializado a 0 en todo su intervalo de definición. Por lo tanto, es cero en el resto del intervalo que no se ha llenado (por ejemplo $[fc(l); fs(l)/2]$). Por lo tanto, en este caso se mejora la complejidad porque no se han realizado etapas de llenado de la señal S , lo que reduce el número de cálculos necesarios.

40 En la prueba T48, se comprueba que la variable de recuento l correspondiente a la señal $I(l)$ del canal l es inferior (por ejemplo, estrictamente) al número L de canales. Si la variable l es menor o igual que L , se incrementa la variable l en la etapa S49 y se repite el método en la etapa S41. Si la variable l es mayor que L , la señal S correspondiente a la señal espacializada para el bloque temporal $[0; N-1]$ para el oído derecho está disponible en la etapa S50.

45 Esta señal S correspondiente al bloque temporal $[0; N-1]$ se suma acto seguido con las demás señales generadas de manera similar para otros bloques de tiempo $[N; 2N-1]$, $[2N; 3N-1]$, etc., (y a las que se ha aplicado un retardo adecuado de acuerdo con la etapa DBD anterior de la Figura 2, por ejemplo).

Normalmente, para construir el bloque $[N; 2N-1]$, se aplica un filtro en el dominio de la frecuencia correspondiente a una función de transferencia común a todas las señales de entrada $I(l)$, que representa el campo difuso, con una frecuencia de corte fc en la multiplicación de las componentes espectrales que corresponde al mínimo entre:

- 50 • una frecuencia máxima del campo difuso F_c (difusa), según se ilustra en la figura 3 descrita anteriormente (elegida, por ejemplo, entre 10 y 15 kHz para el bloque $m=1$ y entre 5 y 10 kHz para el bloque $m=2$), y
- 55 • la frecuencia máxima f_{max} representada en cada señal de entrada (por ejemplo, su frecuencia de muestreo o la frecuencia máxima cuya componente espectral no es cero, dándose normalmente este valor por un descodificador de compresión).

60 Hay que señalar que la multiplicación de frecuencias por parada en una frecuencia de corte dada (lo que matemáticamente equivale a multiplicar por 0 más allá de dicha frecuencia) no es trivial para el experto en la técnica. En efecto, en el contexto del filtrado de una señal de audio, este tipo de filtro pasa bajo muy violento produce generalmente artefactos audibles (denominados "aliasing"), debidos a fenómenos de eco o pre-eco resultantes del plegamiento temporal generado por la convolución circular, que generalmente es deseable evitar. Sin embargo, en el contexto de la invención, este filtro pasa bajo no se aplica a la señal de audio, sino al filtro $BRIR$ (que a su vez está convolucionado con la señal de audio), que ya está compuesto por múltiples reflexiones; por lo tanto, los artefactos producidos se percibirán, en el peor de los casos, como reflexiones adicionales del filtro $BRIR$ original, y en la práctica rara vez serán perceptibles. Sin embargo, es posible atenuar estos artefactos modificando ligeramente las frecuencias

del filtro que preceden a la frecuencia de corte (por ejemplo, mediante una atenuación suave a través de la aplicación de una media ventana de Hanning (tipo fade out)).

5 En general, con referencia a la figura 4, se observará que se realizan dos operaciones en la misma instancia de bucle (normalmente una señal de reloj): la multiplicación MULT(k) y su adición a la señal de salida S. Esto permite implementar este método en procesadores que puedan realizar varias operaciones en la misma instancia de bucle (normalmente una señal de reloj), reduciendo de este modo el tiempo necesario para los cálculos.

10 La figura 5 ilustra una forma algorítmica completa de procesamiento, de acuerdo con la fórmula que da una señal de salida O^k presentada anteriormente:

$$O^k = \sum_{l=1}^L (I(l) *_{[0, \dots, f^k(l)]} A^k(l)) + \sum_{m=1}^M \left(z^{-iD^m} \cdot G(I(l)) \cdot \sum_{l=1}^L \left(\frac{1}{W^k(l)} \cdot I(l) \right) \right) *_{[0, \dots, f^k(m)]} B_{mean}^k(m)$$

15 Como ya se indicó anteriormente, las ponderaciones $W^k(l)$ y las ganancias $G(I(l))$ se pueden fijar en 1. Las ganancias $G(I(l))$ no se muestran en la figura 5 porque esta figura se debe leer como una integración de las ganancias con las ponderaciones $1/W^k(l)$. Cuando se diseñan los filtros, estos dos parámetros se determinan, se fijan y se multiplican entre sí de una vez por todas.

REIVINDICACIONES

1. Módulo de espacialización del sonido, que comprende medios de cálculo para aplicar al menos una función de transferencia de efecto de sala a al menos una señal de sonido de entrada, consistiendo dicha aplicación en multiplicar, en el dominio espectral, componentes espectrales de la señal de sonido por las componentes espectrales de un filtro correspondiente a dicha función de transferencia, incluyendo cada componente espectral del filtro una evolución temporal en una representación de tiempo-frecuencia,

en el que los medios de cálculo se configuran para ignorar dichas componentes espectrales del filtro, para dichas multiplicaciones de componentes, más allá de una frecuencia umbral y después de al menos un instante dado en dicha representación tiempo-frecuencia y en el que el módulo de espacialización del sonido que recibe varias señales de entrada, suministra al menos dos señales de salida, estando configurados los medios de cálculo para aplicar una función de transferencia de efecto de sala a cada señal de entrada,y

15 **caracterizándose** el módulo **por que** cada una de dichas señales de salida viene dada por la aplicación de la siguiente fórmula:

$$O^k = \sum_{l=1}^L (I(l) *_{[0;...;f^k(l)]} A^k(l)) + \sum_{m=1}^M \left(z^{-iDDm} \cdot G(I(l)) \cdot \sum_{l=1}^L \left(\frac{1}{W^k(l)} \cdot I(l) \right) \right) *_{[0;...;f^k(m)]} B_{mean}^k(m)$$

20 - siendo O^k una señal de salida, y siendo k el índice relativo a una señal de salida,
 - siendo $l \in [1; L]$, el índice relativo a una de dichas señales de entrada, siendo L el número de señales de entrada, y siendo $I(l)$ una de dichas señales de entrada,

25 - siendo $A^k(l)$ una función de transferencia con efecto de sala específica para una señal de entrada, $B_{mean}^k(m)$ siendo una función de transferencia global, con efecto de sala, común a las señales de entrada,

- siendo $W^k(l)$ un peso seleccionado y siendo $G(I(l))$ una ganancia de compensación de energía predeterminada,

30 - siendo z^{-iDDm} una aplicación de retardo, contada en número de bloques de muestras, correspondiente a una diferencia de tiempo entre una emisión de sonido en una sala correspondiente al efecto de sala, y un inicio de presencia de campo difuso en esta sala, correspondiendo el índice m a un número de bloques de muestras de duración correspondiente a este retardo, siendo M el número total de bloques que dura una función de transferencia en una representación tiempo-frecuencia,

35 - el signo "." indica la multiplicación,

- designando el signo "[0[#](l)]" el operador de convolución sobre un número limitado de frecuencias que van de una frecuencia mínima a una frecuencia máxima [#](l) que es función de al menos la señal de entrada de índice l, y

40 - designando el signo "[0;...;[#](m)]" el operador de convolución sobre un número limitado de frecuencias, que van de una frecuencia mínima a una ; ; frecuencia [#](m) que es función del bloque de muestras de índice m

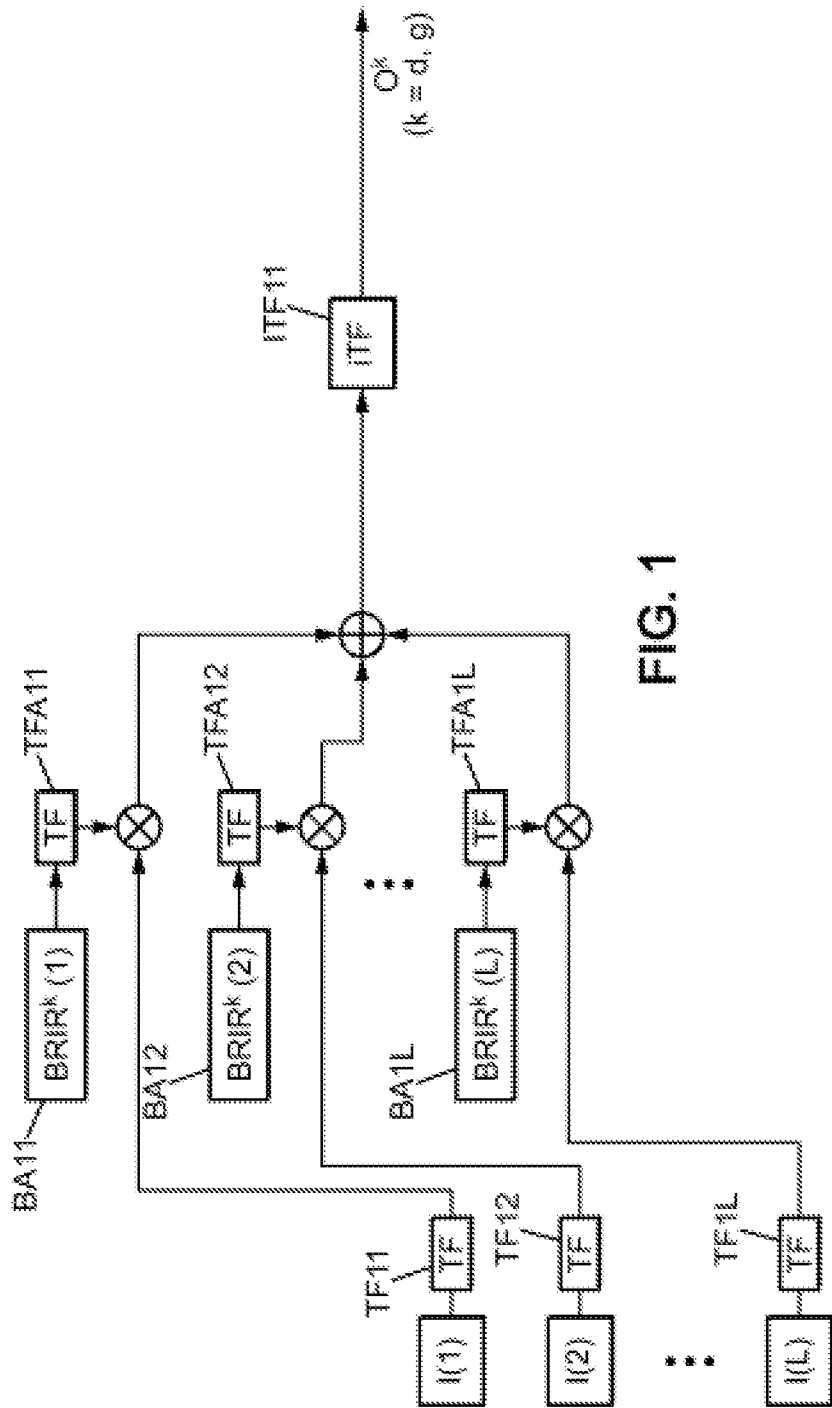


FIG. 1

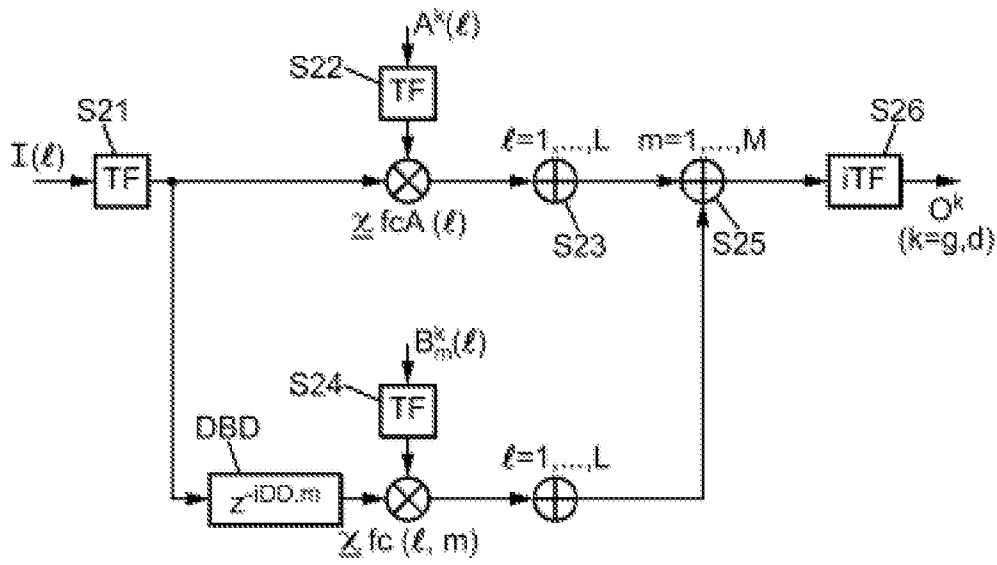


FIG. 2

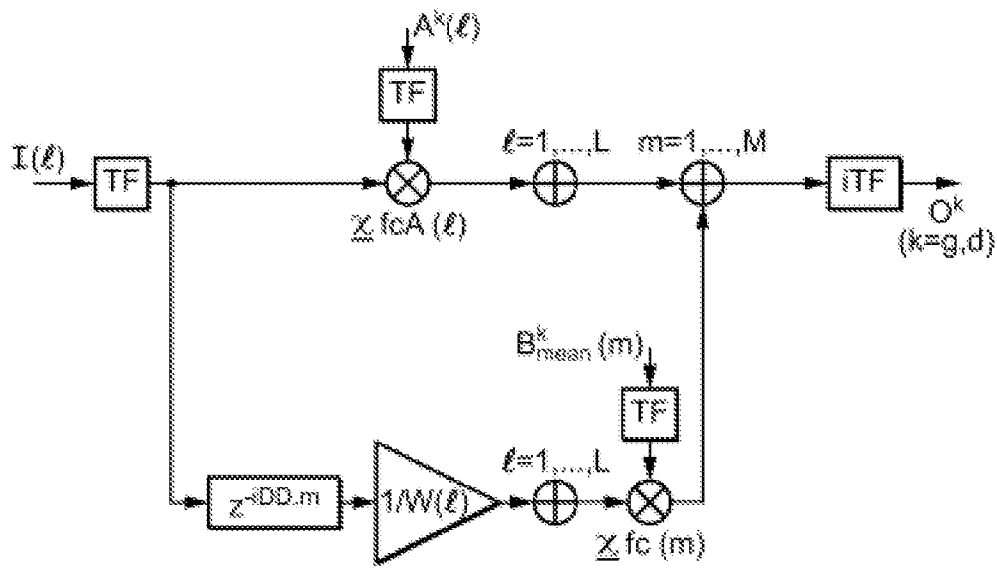


FIG. 5

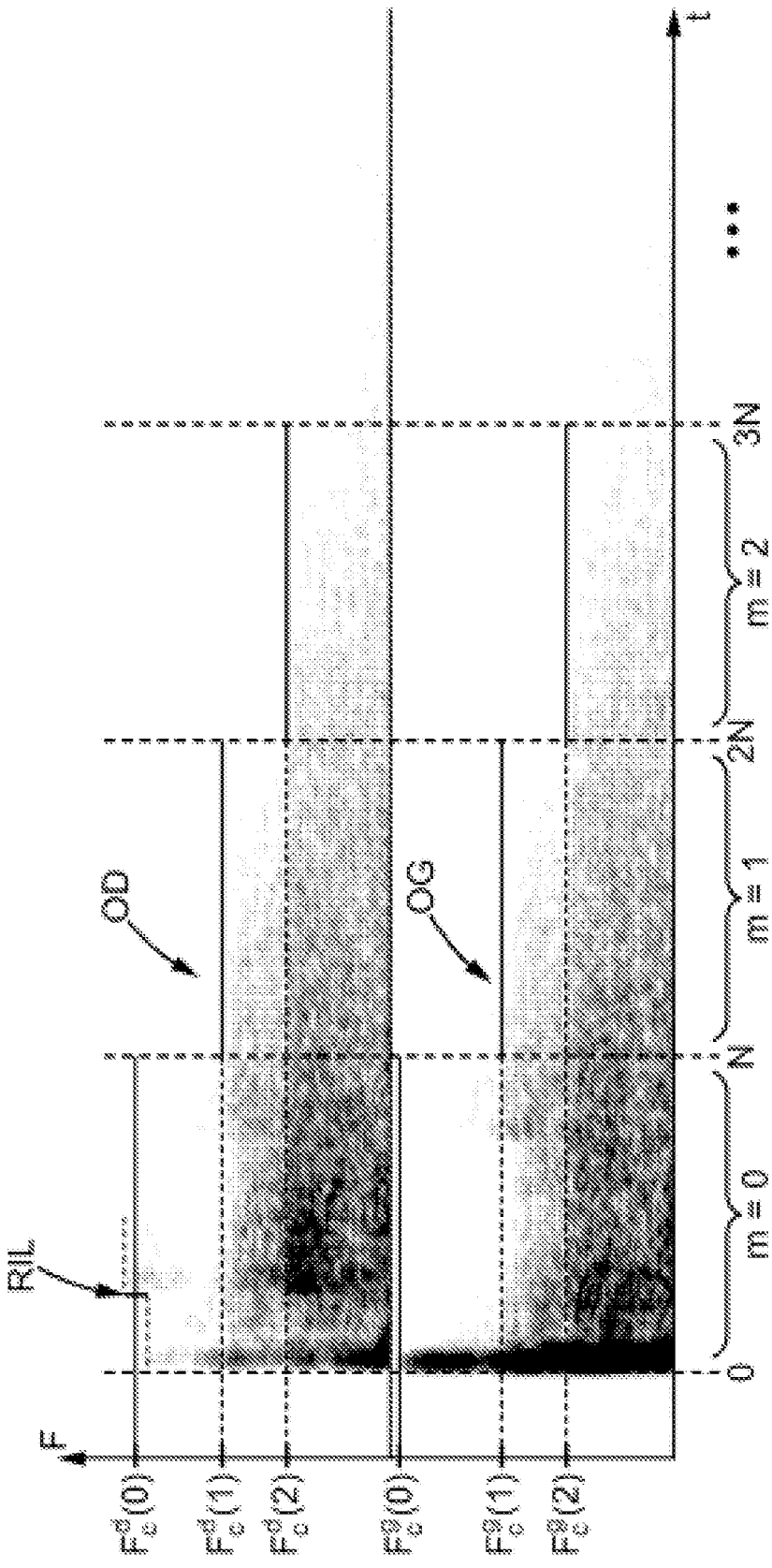


FIG. 3

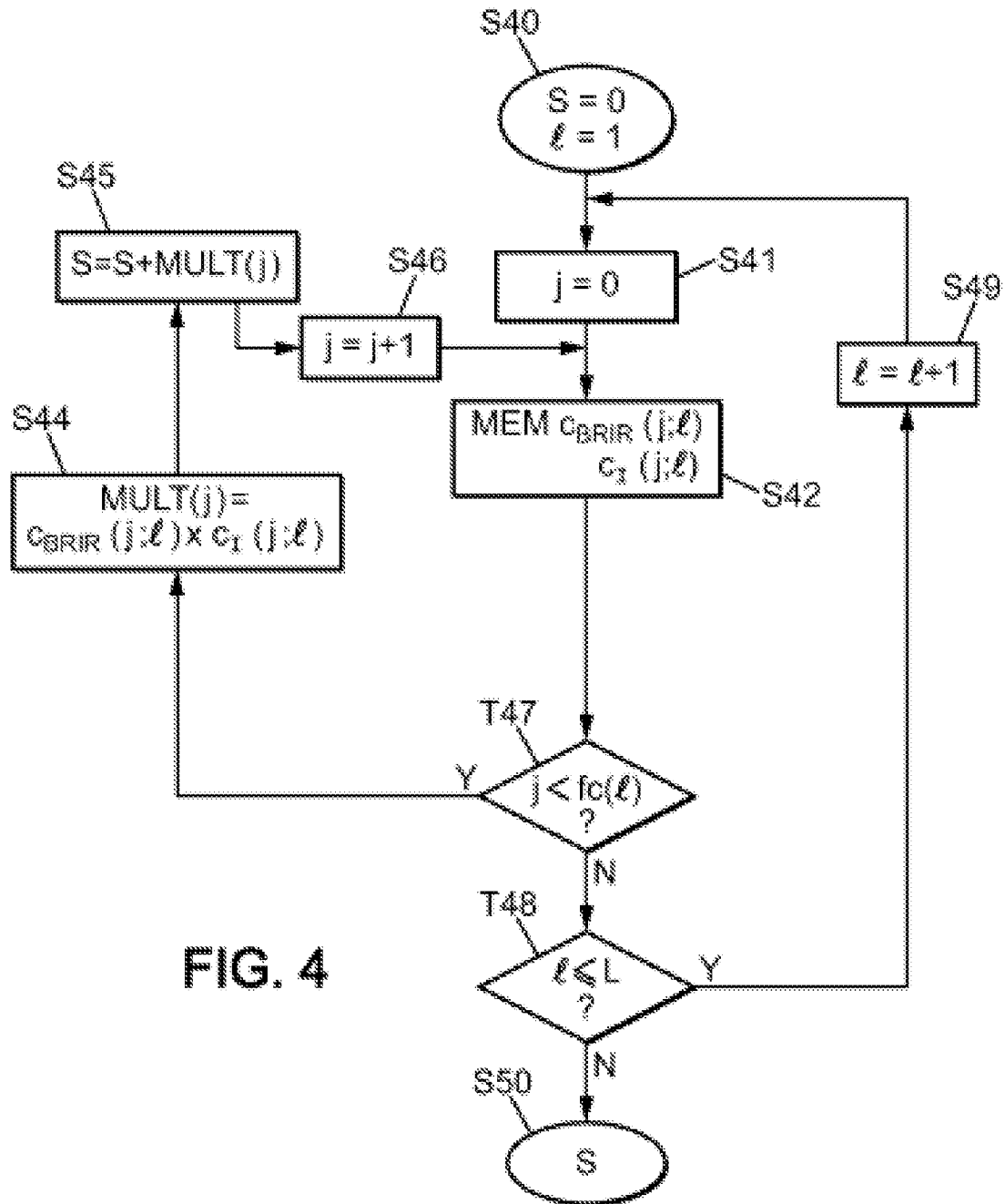


FIG. 4

