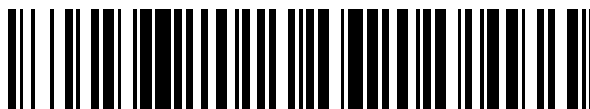


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 531 422**

51 Int. Cl.:

H04S 3/00 (2006.01)

H04S 7/00 (2006.01)

G10K 15/12 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **30.07.2009** **E 11168514 (5)**

97 Fecha y número de publicación de la concesión europea: **10.09.2014** **EP 2384029**

54 Título: **Generación de señales para señales binaurales**

30 Prioridad:

31.07.2008 US 85286

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
13.03.2015

73 Titular/es:

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

72 Inventor/es:

**PLOGSTIES, JAN;
MUNDT, HARALD;
NEUEBAUER, BERNHARD;
HILPERT, JOHANNES y
SILZLE, ANDREAS**

74 Agente/Representante:

PONTI SALES, Adelaida

ES 2 531 422 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

Generación de señales para señales binaurales

5 **[0001]** La presente invención se refiere a la generación de una contribución relacionada con la reflexión y/o reverberación de un espacio de una señal binaural, la generación de una señal binaural en sí, y la formación de un conjunto decreciente de intersimilitud de funciones de transferencia relacionadas con la cabeza.

[0002] El sistema auditivo humano es capaz de determinar la dirección o las direcciones de las que proceden
 10 los sonidos percibidos. Con este fin, el sistema auditivo humano evalúa ciertas diferencias entre el sonido recibido en el oído derecho y el sonido recibido en el oído izquierdo. Esta información comprende, por ejemplo, las denominadas pistas interaurales que pueden, a su vez, referirse a la diferencia de señal sonora entre los oídos. Las pistas interaurales son los medios más importantes para la localización. La diferencia de nivel de presión entre los oídos, concretamente la diferencia de nivel interaural (ILD) es la pista individual más importante para la localización.
 15 Cuando el sonido llega del plano horizontal con un acimut distinto de cero, tiene un nivel diferente en cada oído. El oído en sombra tiene una imagen sonora suprimida naturalmente, comparado con el oído no en sombra. Otra propiedad muy importante que se ocupa de la localización es la diferencia de tiempo interaural (ITD). El oído en sombra tiene una distancia más larga hasta la fuente de sonido, y por tanto recibe el frente de onda sonora después que el oído no en sombra. El significado de la ITD se enfatiza en las bajas frecuencias que no se atenúan mucho
 20 cuando llegan al oído en sombra comparado con el oído no en sombra. La ITD es menos importante en las frecuencias más altas porque la longitud de onda del sonido se acerca a la distancia entre los oídos. De ahí que, en otras palabras, la localización se aprovecha del hecho de que el sonido está sometido a diferentes interacciones con la cabeza, los oídos, y los hombros del oyente que se desplaza de la fuente de sonido al oído izquierdo y derecho, respectivamente.

25 **[0003]** Se producen problemas cuando una persona escucha una señal estéreo que está pensada para ser reproducida por un sistema de altavoces a través de auriculares. Es muy probable que el oyente considere el sonido como poco natural, molesto y perturbador ya que el oyente siente que la fuente de sonido está localizada en la cabeza. Este fenómeno a menudo se denomina en la bibliografía como localización "en la cabeza". La escucha
 30 prolongada de sonido "en la cabeza" puede conducir a fatiga auditiva. Se produce porque la información en la que se basa el sistema auditivo humano, cuando posiciona las fuentes de sonido, es decir, las pistas interaurales, falta o es ambigua.

[0004] Con el fin de interpretar señales estéreo, o incluso señales multicanal con más de dos canales para
 35 reproducción por auriculares, pueden usarse filtros direccionales con el fin de modelizar estas interacciones. Por ejemplo, la generación de una salida de auriculares a partir de una señal multicanal descodificada puede comprender filtrar cada señal después la descodificación por medio de un par de filtros direccionales. Estos filtros modelan típicamente la transmisión acústica desde una fuente de sonido virtual en un espacio hasta el conducto auditivo de un oyente, la denominada función de transferencia binaural de un espacio (BRTF). La BRTF realiza
 40 modificaciones de tiempo, de nivel y espectrales, y modeliza las reflexiones y la reverberación de un espacio. Los filtros direccionales pueden implementarse en el dominio del tiempo o la frecuencia.

[0005] Sin embargo, como se requieren muchos filtros, concretamente $N \times 2$ siendo N el número de canales descodificados, estos filtros direccionales son bastante largos, tales como 20000 derivaciones de filtro a 44,1 kHz, y
 45 el proceso de filtrado es computacionalmente exigente. Por lo tanto, los filtros direccionales a veces se reducen a un mínimo. Las denominadas funciones de transferencia relacionadas con la cabeza (HRTF) contienen la información direccional que incluye las pistas interaurales. Se usa un bloque de procesamiento común para modelizar las reflexiones y la reverberación de un espacio. El módulo de procesamiento de espacio puede ser un algoritmo de reverberación en el dominio del tiempo o la frecuencia, y puede operar sobre una señal de entrada de uno o dos
 50 canales de la señal de entrada multicanal. Tal estructura se describe, por ejemplo, en el documento WO99/14983A1. Tal como se acaba de describir, el bloque de procesamiento de espacio implementa las reflexiones y/o la reverberación de un espacio. Las reflexiones y la reverberación de un espacio son esenciales para los sonidos localizados, especialmente con respecto a la distancia y los sonidos que significan externalización se perciben fuera de la cabeza del oyente. El documento anteriormente mencionado también sugiere implementar los filtros
 55 direccionales como un conjunto de filtros FIR que operan sobre versiones retardadas de manera diferente del canal respectivo, para modelizar el recorrido directo desde la fuente de sonido hasta el oído respectivo y distintas reflexiones. Por otra parte, al describir varias medidas para proporcionar una experiencia auditiva más agradable a través de un par de auriculares, este documento también sugiere retardar una mezcla del canal central y el canal delantero izquierdo, y el canal central y el canal delantero derecho, respectivamente, en relación con una suma y

una diferencia de los canales trasero izquierdo y trasero derecho, respectivamente.

[0006] Sin embargo, los resultados auditivos conseguidos hasta ahora todavía carecen en gran medida de una anchura espacial reducida de la señal de salida binaural y carecen de externalización. Además, se ha caído en la cuenta de que, a pesar de las medidas anteriormente mencionadas para interpretar señales multicanal para reproducción por auriculares, porciones de voz en los diálogos de películas y la música a menudo se perciben reverberantes de manera poco natural y espectralmente desiguales.

[0007] El documento US5371799 describe un sistema para procesar una señal de audio para reproducción por auriculares en el cual la fuente de sonido aparate está localizada fuera de la cabeza del oyente procesa la señal de entrada como si estuviera constituida por una porción de onda directa, una porción de reflexiones iniciales, y una porción de reverberaciones. La porción de onda directa de la señal se procesa en filtros cuyos coeficientes de filtro se escogen basándose en el acimut deseado de la localización de la fuente de sonido virtual. La porción de reflexión inicial se pasa a través de un banco de filtros conectados en paralelo cuyos coeficientes se escogen basándose en cada acimut de reflexión. Las salidas de estos filtros se pasan a través de escalares para ajustar la amplitud para simular un intervalo deseado de la fuente de sonido virtual. La porción de reverberación se procesa sin ninguna información de localización de fuente de sonido, usando un generador de números aleatorios, por ejemplo, y la salida es atenuada en un atenuador exponencial para que se atenúe gradualmente. Todas las salidas de los escalares y los atenuadores se suman entonces para producir señales de auriculares izquierda y derecha para reproducción por los transductores de auriculares respectivos.

[0008] Por lo tanto, el objeto de la presente invención es proporcionar un esquema para generación de señales binaurales, que produzca una reproducción por auriculares más estable y agradable.

[0009] Este objeto se consigue mediante un dispositivo según la reivindicación 1 y un procedimiento según la reivindicación 7.

[0010] Una idea subyacente a la presente solicitud es que algunas porciones en películas y música tienen como resultado una reproducción por auriculares percibida de manera más natural, cuando la mezcla descendente mono o estéreo de los canales de la señal multicanal que ha de someterse al procesador de espacio para generar la contribución relacionada con las reflexiones/la reverberación de un espacio de la señal binaural, se forma de manera que la pluralidad de canales contribuyen a la mezcla descendente mono o estéreo a un nivel que difiere entre al menos dos canales de la señal multicanal. Por ejemplo, los inventores se dieron cuenta de que las voces en los diálogos de películas y la música típicamente se mezclan principalmente con el canal central de una señal multicanal, y que la señal del canal central, cuando se suministra al módulo de procesamiento de espacio, tiene como resultado una salida a menudo percibida reverberante poco natural y espectralmente desigual. Los inventores descubrieron, sin embargo, que estas deficiencias pueden superarse suministrando el canal central al módulo de procesamiento de espacio con una reducción de nivel tal como, por ejemplo, una atenuación de 3-12 dB o, específicamente, 6 dB.

[0011] Además, puede conseguirse una señal binaural más estable y agradable para reproducción por auriculares procesando de manera diferente, y reduciendo así la similitud entre, al menos uno de un canal izquierdo y uno derecho de la pluralidad de canales de entrada, un canal delantero y uno trasero de la pluralidad de canales de entrada, y un canal central y uno no central de la pluralidad de canales, obteniendo así un conjunto reducido de intersimilitud de canales. Este conjunto reducido de intersimilitud de canales se suministra luego a una pluralidad de filtros direccionales seguidos por mezcladores respectivos para el oído izquierdo y el derecho, respectivamente. Reduciendo la intersimilitud de canales de la señal de entrada multicanal, puede aumentarse la anchura espacial de la señal de salida binaural y puede mejorarse la externalización.

[0012] Además, puede conseguirse una señal binaural más estable y agradable para reproducción por auriculares realizando – en un sentido variable espectralmente – una modificación de fase y/o magnitud de manera diferente entre al menos dos canales de la pluralidad de canales, obteniendo así el conjunto reducido de intersimilitud de canales que, a su vez, puede suministrarse luego a una pluralidad de filtros direccionales seguidos por mezcladores respectivos para el oído izquierdo y el derecho, respectivamente. De nuevo, reduciendo la intersimilitud de canales de la señal de entrada multicanal, puede aumentarse la anchura espacial de la señal de salida binaural y puede mejorarse la externalización.

[0013] Puede formarse además un conjunto decreciente de intersimilitud de funciones de transferencia relacionadas con la cabeza haciendo que las respuestas impulsionales de una pluralidad original de funciones de

transferencia relacionadas con la cabeza se retarden unas en relación con otras, o – en un sentido variable espectralmente – las respuestas de fase y/o magnitud de la pluralidad original de funciones de transferencia relacionadas con la cabeza de manera diferente unas en relación con otras. La formación puede hacerse fuera de línea como una etapa de diseño, o en línea durante la generación de señales binaurales, usando las funciones de transferencia relacionadas con la cabeza como filtros direccionales tales como, por ejemplo, sensibles a una indicación de localizaciones de fuentes de sonido virtuales que han de usarse.

[0014] En lo que viene a continuación, las realizaciones preferidas se describen con más detalle con respecto a las figuras, entre las cuales:

La fig. 1 muestra un diagrama de bloques de un dispositivo para generar una señal binaural;

la fig. 2 muestra un diagrama de bloques de un dispositivo para formar un conjunto decreciente de intersimilitud de funciones de transferencia relacionadas con la cabeza;

la fig. 3 muestra un dispositivo para generar una contribución relacionada con la reflexión y/o reverberación de un espacio de una señal binaural según una realización adicional;

las figs. 4a y 4b muestran diagramas de bloques del procesador de espacio de la fig. 3 según distintas realizaciones;

la fig. 5 muestra un diagrama de bloques del generador de mezcla descendente de la fig. 3 según una realización;

la fig. 6 muestra un diagrama esquemático que ilustra una representación de una señal multicanal que usa codificación de audio espacial según una realización;

la fig. 7 muestra un generador de señales de salida binaurales según una realización;

la fig. 8 muestra un diagrama de bloques de un generador de señales de salida binaurales según una realización adicional;

la fig. 9 muestra un diagrama de bloques de un generador de señales de salida binaurales según una realización adicional más;

la fig. 10 muestra un diagrama de bloques de un generador de señales de salida binaurales según una realización adicional;

la fig. 11 muestra un diagrama de bloques de un generador de señales de salida binaurales según una realización adicional;

la fig. 12 muestra un diagrama de bloques del decodificador de audio espacial binaural de la fig. 11 según una realización; y

la fig. 13 muestra un diagrama de bloques del decodificador de audio espacial modificado de la fig. 11 según una realización.

[0015] La fig. 1 muestra un dispositivo para generar una señal binaural pensada, por ejemplo, para reproducción por auriculares basándose en una señal multicanal que representa una pluralidad de canales y pensada para reproducción mediante una configuración de altavoces que tiene una posición de fuente de sonido virtual asociada a cada canal. El dispositivo que se indica en general con el signo de referencia 10, comprende un reductor de similitud 12, una pluralidad 14 de filtros direccionales 14a-14h, un primer mezclador 16a y un segundo mezclador 16b.

[0016] El reductor de similitud 12 está configurado para convertir la señal multicanal 18 que representa la pluralidad de canales 18a-18d, en un conjunto reducido de intersimilitud 20 de canales 20a-20d. El número de canales 18a-18d representado por la señal multicanal 18 puede ser dos o más. Únicamente con fines de ilustración, en la fig. 1 se han mostrado explícitamente cuatro canales 18a-18d. La pluralidad 18 de canales puede, por ejemplo, comprender un canal central, un canal delantero izquierdo, un canal delantero derecho, un canal trasero izquierdo, y un canal trasero derecho. Los canales 18a-18d han sido mezclados, por ejemplo, por un diseñador de sonido a partir de una pluralidad de señales de audio individuales que representan, por ejemplo, instrumentos individuales, voces, u otras fuentes de sonido individuales, suponiendo que o con la intención de que los canales 18a-18d sean

reproducidos por un sistema de altavoces (no mostrado en la fig. 1), que tiene los altavoces colocados en posiciones de fuentes de sonido virtuales asociadas a cada canal 18a-18d.

[0017] Según la fig. 1, la pluralidad de canales 18a-18d comprende, al menos, un par de un canal izquierdo y uno derecho, un par de un canal delantero y uno trasero, o un par de un canal central y uno no central. Por supuesto, pueden estar presentes más de uno de los pares recién mencionados dentro de la pluralidad 18 de canales 18a-18d. El reductor de similitud 12 está configurado para procesar de manera diferente, y reducir así una similitud entre canales de la pluralidad de canales, con el fin de obtener el conjunto reducido de intersimilitud 20 de canales 20a-20d. Según un primer aspecto, la similitud entre al menos uno de un canal izquierdo y uno derecho de la pluralidad 18 de canales, un canal delantero y uno trasero de una pluralidad 18 de canales, y un canal central y uno no central de la pluralidad 18 de canales puede reducirse mediante el reductor de similitud 12, con el fin de obtener el conjunto reducido de intersimilitud 20 de canales 20a-20d. Según un segundo aspecto, el reductor de similitud (12) puede – adicional o alternativamente – realizar – en un sentido variable espectralmente – una modificación de fase y/o magnitud de manera diferente entre al menos dos canales de la pluralidad de canales, con el fin de obtener el conjunto reducido de intersimilitud 20 de canales.

[0018] Tal como se explicará resumidamente con más detalle más adelante, el reductor de similitud 12 puede, por ejemplo, conseguir el diferente procesamiento haciendo que los pares respectivos se retarden unos en relación con otros, o sometiendo a los pares de canales respectivos a retardos de diferentes cantidades en, por ejemplo, cada una de una pluralidad de bandas de frecuencia, obteniendo así un conjunto reducido de intersimilitud 20 de canales. Por supuesto, existen otras posibilidades con el fin de disminuir la correlación entre los canales. Incluso en otras palabras, el reductor de correlación 12 puede tener una función de transferencia según la cual la distribución de energía espectral de cada canal permanece igual, es decir, la función de transferencia como una magnitud de uno a lo largo del intervalo del espectro de audio pertinente en el que, sin embargo, el reductor de similitud 12 modifica de manera diferente las fases de las subbandas o las componentes de frecuencia de las mismas. Por ejemplo, el reductor de correlación 12 podría estar configurado de manera que el mismo cause una modificación de fase en todos, o uno o varios de, los canales 18 de manera que una señal de un primer canal para una cierta banda de frecuencia se retarde en relación con otro de los canales en al menos una muestra. Además, el reductor de correlación 12 podría estar configurado de manera que el mismo cause la modificación de fase de manera que los retardos de grupo de un primer canal en relación con otro de los canales para una pluralidad de bandas de frecuencia, muestren una desviación típica de al menos un octavo de una muestra. Las bandas de frecuencia consideradas podrían ser las bandas de Bark o un subconjunto de las mismas o cualquier otra subdivisión de banda de frecuencia.

[0019] Reducir la correlación no es el único modo de impedir la localización en la cabeza del sistema auditivo humano. Más bien, la correlación es una de varias medidas posibles mediante cuyo uso el sistema auditivo humano mide la similitud del sonido que llega a ambos oídos, y de ese modo, la dirección de llegada del sonido. Por consiguiente, el reductor de similitud 12 también puede conseguir el diferente procesamiento sometiendo a los pares de canales respectivos a reducciones de nivel de diferentes cantidades en, por ejemplo, cada una de una pluralidad de bandas de frecuencia, obteniendo así un conjunto reducido de intersimilitud 20 de canales de un modo formado espectralmente. La formación espectral puede, por ejemplo, exagerar la reducción formada espectralmente relativa que se produce, por ejemplo, para el sonido del canal trasero en relación con el sonido del canal delantero debido al ensombrecimiento por el lóbulo. Por consiguiente, el reductor de similitud 12 puede someter al (los) canal(es) trasero(s) a unas reducciones de nivel variables espectralmente en relación con otros canales. En esta formación espectral, el reductor de similitud 12 puede tener una respuesta de fase que es constante a lo largo del intervalo del espectro de audio pertinente en el que, sin embargo, el reductor de similitud 12 modifica de manera diferente las magnitudes de las subbandas o las componentes de frecuencia de las mismas.

[0020] El modo en el que la señal multicanal 18 representa una pluralidad de canales 18a-18d no está, en principio, restringido a ninguna representación específica. Por ejemplo, la señal multicanal 18 podría representar la pluralidad de canales 18a-18d de una manera comprimida, usando codificación de audio espacial. Según la codificación de audio espacial, la pluralidad de canales 18a-18d podría estar representada por medio de una señal de mezcla descendente hasta la cual se realiza la mezcla descendente de los canales, acompañada por información de mezcla descendente que revela la proporción de mezcla según la cual los canales individuales 18a-18d han sido mezclados dentro del canal de mezcla descendente o los canales de mezcla descendente, y parámetros espaciales que describen la imagen espacial de la señal multicanal por medio de, por ejemplo, diferencias de nivel/intensidad, diferencias de fase, diferencias de tiempo y/o medidas de correlación/coherencia entre los canales individuales 18a-18d. La salida del reductor de correlación 12 se divide en los canales individuales 20a-20d. Estos canales pueden, por ejemplo, producirse como salida como señales de tiempo o como espectrogramas tales como, por ejemplo,

descompuestos espectralmente en subbandas.

[0021] Los filtros direccionales 14a-14h están configurados para modelizar una transmisión acústica de un canal respectivo de los canales 20a-20d desde una posición de fuente de sonido virtual asociada con el canal respectivo hasta un conducto auditivo respectivo del oyente. En la fig. 1, los filtros direccionales 14a-14d modelizan la transmisión acústica hasta, por ejemplo, el conducto auditivo izquierdo, mientras que los filtros direccionales 14e-14h modelizan la transmisión acústica hasta el conducto auditivo derecho. Los filtros direccionales pueden modelizar la transmisión acústica desde una posición de fuente de sonido virtual en un espacio hasta un conducto auditivo del oyente y pueden realizar esta modelización realizando modificaciones de tiempo, de nivel y espectrales, y opcionalmente, modelizando las reflexiones y la reverberación de un espacio. Los filtros direccionales 18a-18h pueden implementarse en el dominio del tiempo o de la frecuencia. Es decir, los filtros direccionales pueden ser filtros del dominio del tiempo tales como filtros, filtros FIR, o pueden operar en el dominio de la frecuencia multiplicando los valores de muestra de la función de transferencia respectivos por los valores espectrales respectivos de los canales 20a-20d. En particular, los filtros direccionales 14a-14h pueden seleccionarse para modelizar la función de transferencia relacionada con la cabeza respectiva que describe la interacción de la señal del canal respectivo 20a-20d desde la posición de la fuente de sonido virtual respectiva hasta el conducto auditivo respectivo, incluyendo, por ejemplo, las interacciones con la cabeza, los oídos, y los hombros de una persona humana. El primer mezclador 16a está configurado para mezclar las salidas de los filtros direccionales 14a-14d que modelizan la transmisión acústica hasta el conducto auditivo izquierdo del oyente para obtener una señal 22a pensada para contribuir a, o incluso ser el canal izquierdo de la señal de salida binaural, mientras que el segundo mezclador 16b está configurado para mezclar las salidas de los filtros direccionales 14e-14h que modelizan la transmisión acústica hasta el conducto auditivo derecho del oyente para obtener una señal 22b, y pensada para contribuir o incluso ser el canal derecho de la señal de salida binaural.

[0022] Tal como se describe con más detalle más adelante con las realizaciones respectivas, pueden añadirse contribuciones adicionales a las señales 22a y 22b, con el fin de tener en cuenta las reflexiones y/o la reverberación de un espacio. Mediante esta medida, puede reducirse la complejidad de los filtros direccionales 14a-14h.

[0023] En el dispositivo de la fig. 1, el reductor de similitud 12 contrarresta los efectos secundarios negativos de la suma de las señales correlacionadas introducidas en los mezcladores 16a y 16b, respectivamente, según lo cual resulta una anchura espacial más reducida de la señal de salida binaural 22a y 22b y una falta de externalización. La descorrelación conseguida mediante el reductor de similitud 12 reduce estos efectos secundarios negativos.

[0024] La fig. 1 muestra, en otras palabras, un flujo de señales para la generación de una salida de auriculares a partir de, por ejemplo, una señal multicanal descodificada. Cada señal es filtrada por un par de pares de filtros direccionales. Por ejemplo, el canal 18a es filtrado por el par de filtros direccionales 14a-14e. Por desgracia, entre los canales 18a-18d existe una cantidad significativa de similitud tal como correlación en las producciones de sonido multicanal típicas. Esto afectaría negativamente a la señal de salida binaural. Concretamente, después de procesar las señales multicanal con un filtro direccional 14a-14h, las señales intermedias producidas como salida por los filtros direccionales 14a-14h se suman en el mezclador 16a y 16b para formar la señal de salida de auriculares 20a y 20b. La suma de señales de salida similares/correlacionadas tendría como resultado una anchura espacial más reducida de la señal de salida 20a y 20b, y una falta de externalización. Esto es particularmente problemático para la similitud/correlación de la señal izquierda y derecha y el canal central. Por consiguiente, el reductor de similitud 12 ha de reducir la similitud entre estas señales tanto como sea posible.

[0025] Cabe destacar que la mayoría de las medidas realizadas por el reductor de similitud 12 para reducir la similitud entre canales de la pluralidad 18 de canales 18a-18d también podría conseguirse eliminando el reductor de similitud 12 modificando simultáneamente los filtros direccionales para realizar no sólo la modelización anteriormente mencionada de la transmisión acústica, sino también conseguir la disimilitud tal como la descorrelación recién mencionada. Por consiguiente, los filtros direccionales no modelizarían, por lo tanto, por ejemplo, HRTF, sino funciones de transferencia relacionadas con la cabeza modificadas.

[0026] La fig. 2, por ejemplo, muestra un dispositivo para formar un conjunto decreciente de intersimilitud de funciones de transferencia relacionadas con la cabeza para modelizar una transmisión acústica de un conjunto de canales desde una posición de fuente de sonido virtual asociada con el canal respectivo hasta los conductos auditivos de un oyente. El dispositivo, el cual se indica en general por 30, comprende un proveedor de HRTF 32, así como un procesador de HRTF 34.

[0027] El proveedor de HRTF 32 está configurado para proporcionar una pluralidad original de HRTF. La etapa 32

puede comprender mediciones usando una cabeza maniquí estándar, con el fin de medir las funciones de transferencia relacionadas con la cabeza desde ciertas posiciones de sonido hasta los conductos auditivos de un oyente maniquí estándar. Igualmente, el proveedor de HRTF 32 puede estar configurado para simplemente buscar o cargar las HRTF originales desde una memoria. Incluso alternativamente, el proveedor de HRTF 32 puede estar configurado para calcular las HRTF según una fórmula predeterminada, dependiendo de, por ejemplo, las posiciones de fuentes de sonido virtuales de interés. Por consiguiente, el proveedor de HRTF 32 puede estar configurado para operar en un entorno de diseño para diseñar un generador de señales de salida binaurales, o puede ser parte de tal señal de generador de señales de salida binaurales en sí, con el fin de proporcionar las HRTF originales en línea tales como, por ejemplo, sensibles a una selección o cambio de las posiciones de fuentes de sonido virtuales. Por ejemplo, el dispositivo 30 puede ser parte de un generador de señales de salida binaurales que es capaz de admitir señales multicanal que están pensadas para diferentes configuraciones de altavoces que tienen diferentes posiciones de fuentes de sonido virtuales asociadas con sus canales. En este caso, el proveedor de HRTF 32 puede estar configurado para proporcionar las HRTF originales de un modo adaptado a las posiciones de fuentes de sonido virtuales pensadas actualmente.

[0028] El procesador de HRTF 34, a su vez, está configurado para hacer que las respuestas impulsionales de al menos un par de las HRTF se desplacen unas en relación con otras o modifiquen – en un sentido variable espectralmente – las respuestas de fase y/o magnitud de las mismas de manera diferente unas en relación con otras. El par de HRTF puede modelizar la transmisión acústica desde uno de los canales izquierdo y derecho, los canales delantero y trasero, y los canales central y no central. En efecto, esto puede conseguirse mediante una o una combinación de las siguientes técnicas aplicadas a uno o varios canales de la señal multicanal, concretamente retardando la HRTF de un canal respectivo, modificando la respuesta de fase de una HRTF respectiva y/o aplicando un filtro de descorrelación tal como un filtro de todo paso a la HRTF respectiva, obteniendo así un conjunto reducido de intercorrelación de HRTF, y/o modificando – en un sentido de modificando espectralmente – la respuesta de magnitud de una HRTF respectiva, obteniendo así un, al menos, conjunto reducido de intersimilitud de HRTF. En cualquier caso, la descorrelación/disimilitud resultante entre los canales respectivos puede tener soporte para el sistema auditivo humano al localizar externamente la fuente de sonido a impedir así que se produzca la localización en la cabeza. Por ejemplo, el procesador de HRTF 34 podría estar configurado de manera que el mismo cause una modificación de la respuesta de fase de todos, o de uno o varios de, las HRTF de los canales de manera que se introduzca un retardo de grupo de una primera HRTF para una cierta banda de frecuencia – o se retarde una cierta banda de frecuencia de una primera HRTF – en relación con otra de las HRTF en al menos una muestra. Además, el procesador de HRTF 34 podría estar configurado de manera que el mismo cause la modificación de la respuesta de fase de manera que los retardos de grupo de una primera HRTF en relación con otra de las HRTF para una pluralidad de bandas de frecuencia, muestren una desviación típica de al menos un octavo de una muestra. Las bandas de frecuencia consideradas podrían ser las bandas de Bark o un subconjunto de las mismas o cualquier otra subdivisión de banda de frecuencia.

[0029] El conjunto decreciente de interesimilitud de HRTF que resulta del procesador de HRTF 34 puede usarse para establecer las HRTF de los filtros direccionales 14a-14h del dispositivo de la fig. 1, en el que el reductor de similitud 12 puede estar presente o ausente. Debido a la propiedad de disimilitud de las HRTF modificadas, las ventajas anteriormente mencionadas con respecto a la anchura espacial de la señal de salida binaural y la externalización mejorada se consiguen igualmente incluso cuando falta el reductor de similitud 12.

[0030] Como ya se describió anteriormente, el dispositivo de la fig. 1 puede estar acompañado por un paso adicional configurado para obtener las contribuciones relacionadas con la reflexión y/o la reverberación de un espacio de la señal de salida binaural basándose en una mezcla descendente de al menos algunos de los canales de entrada 18a-18d. Esto alivia la complejidad planteada por los filtros direccionales 14a-14h. En la fig. 3 se muestra un dispositivo para generar tal contribución relacionada con la reflexión de un espacio y/o la reverberación de un espacio de una señal de salida binaural. El dispositivo 40 comprende el generador de mezcla descendente 42 y un procesador de espacio 44 conectado en serie entre sí con el procesador de espacio 44 después del generador de mezcla descendente 42. El dispositivo 40 puede estar conectado entre la entrada del dispositivo de la fig. 1 en la cual se introduce la señal multicanal 18, y la salida de la señal de salida binaural donde la contribución del canal izquierdo 46a del procesador de espacio 44 se suma a la salida 22a, y la salida del canal derecho 46b del procesador de espacio 44 se suma a la salida 22b. El generador de mezcla descendente 42 forma una mezcla descendente mono o estéreo 48 a partir de los canales de la señal multicanal 18, y el procesador 44 está configurado para generar el canal izquierdo 46a y el canal derecho 46b de las contribuciones relacionadas con la reflexión y/o la reverberación de un espacio de la señal binaural modelizando la reflexión y/o la reverberación de un espacio basándose en la señal mono o estéreo 48.

[0031] La idea subyacente al procesador de espacio 44 es que la reflexión/reverberación de un espacio que se produce, por ejemplo, en un espacio, puede ser modelizada de una manera transparente para el oyente, basándose en una mezcla descendente tal como una simple suma de los canales de la señal multicanal 18. Como las reflexiones/la reverberación de un espacio se producen más tarde que los sonidos que se desplazan a lo largo del recorrido directo o la línea de visión desde la fuente de sonido hasta los conductos auditivos, la respuesta impulsional del procesador de espacio es representativa de, y sustituye a, la cola de las respuestas impulsionales de los filtros direccionales mostrados en la fig. 1. Las respuestas impulsionales de los filtros direccionales pueden, a su vez, restringirse a modelizar el recorrido directo y la reflexión y las atenuaciones que se producen en la cabeza, los oídos, y los hombros del oyente, permitiendo así acortar las respuestas impulsionales de los filtros direccionales. Por supuesto, la frontera entre lo que es modelizado por el filtro direccional y lo que es modelizado por el procesador de espacio 44 puede variarse libremente de manera que el filtro direccional también puede, por ejemplo, modelizar las primeras reflexiones/reverberación de un espacio.

[0032] Las figs. 4a y 4b muestran posibles implementaciones para la estructura interna del procesador de espacio. Según la fig. 1a, al procesador de espacio 44 se le suministra una señal de mezcla descendente mono 48 y comprende dos filtros de reverberación 50a y 50b. Análogamente a los filtros direccionales, los filtros de reverberación 50a y 50b pueden implementarse para que operen en el dominio del tiempo o el dominio de la frecuencia. Las entradas de ambos reciben la señal de mezcla descendente mono 48. La salida del filtro de reverberación 50a proporciona la salida de contribución del canal izquierdo 46a, mientras que el filtro de reverberación 50b produce como salida la señal de contribución del canal derecho 46b. La fig. 4b muestra un ejemplo de la estructura interna del procesador de espacio 44, en el caso de que al procesador de espacio 44 se le proporciona una señal de mezcla descendente estéreo 48. En este caso, el procesador de espacio comprende cuatro filtros de reverberación 50a-50d. Las entradas de los filtros de reverberación 50a y 50b están conectadas a un primer canal 48a de la mezcla descendente estéreo 48, mientras que la entrada de los filtros de reverberación 50c y 50d están conectadas al otro canal 48b de la mezcla descendente estéreo 48. Las salidas de los filtros de reverberación 50a y 50c están conectadas a la entrada de un sumador 52a, cuya salida proporciona la contribución del canal izquierdo 46a. La salida de los filtros de reverberación 50b y 50d están conectadas a las entradas de un sumador adicional 52b, cuya salida proporciona la contribución del canal derecho 46b.

[0033] El generador de mezcla descendente 42 de la realización de la fig. 3 está configurado para formar la mezcla descendente mono o estéreo 48, de manera que la pluralidad de canales contribuye a la mezcla descendente mono o estéreo en un nivel que difiere entre al menos dos canales de la señal multicanal 18. Mediante esta medida, puede impedirse o fomentarse que ciertos contenidos de las señales multicanal tal como la voz o la música de fondo que se mezclan en un canal específico o canales específicos de la señal multicanal sea sometido al procesamiento de espacio, evitando así un sonido poco natural.

[0034] Por ejemplo, el generador de mezcla descendente 42 de la fig. 3 puede estar configurado para formar la mezcla descendente mono o estéreo 48 de manera que un canal central de la pluralidad de canales de la señal multicanal 18 contribuye a la señal de mezcla descendente mono o estéreo 48 de una manera de nivel reducido en relación con los otros canales de la señal multicanal 18. Por ejemplo, la cantidad de reducción de nivel puede ser entre 3 dB y 12 dB. La reducción de nivel puede extenderse uniformemente por el intervalo espectral efectivo de los canales de la señal multicanal 18, o puede ser dependiente de la frecuencia tal como concentrado en una porción espectral específica, tal como la porción espectral ocupada típicamente por señales de voz. La cantidad de reducción de nivel en relación con los otros canales puede ser la misma para todos los otros canales. Es decir, los otros canales pueden mezclarse en la señal de mezcla descendente 48 al mismo nivel. Alternativamente, los otros canales pueden mezclarse en la señal de mezcla descendente 48 a un nivel desigual. Después, puede medirse la cantidad de reducción de nivel en relación con los otros canales frente al valor medio de los otros canales o el valor medio de todos los canales incluyendo el reducido. De ser así, la desviación típica de las ponderaciones de mezcla de los otros canales o la desviación típica de las ponderaciones de mezcla de todos los canales puede ser menor que el 66 % de la reducción de nivel de la ponderación de mezcla del canal de nivel reducido en relación con el valor medio recién mencionado.

[0035] El efecto de la reducción de nivel con respecto al canal central es que la señal de salida binaural obtenida por medio de las contribuciones 56a y 56b es – al menos en algunas circunstancias que se analizan con más detalle más adelante – percibida de manera más natural por los oyentes que sin la reducción de nivel. En otras palabras, el generador de mezcla descendente 42 forma una suma ponderada de los canales de los canales de la señal multicanal 18, con el valor de ponderación asociado con el canal central estando reducido en relación con los valores de ponderación de los otros canales.

- [0036]** La reducción de nivel del canal central es especialmente ventajosa durante las porciones de voz de diálogos de películas o música. La mejora de la impresión de audio obtenida durante estas porciones de voz compensa con creces las penalizaciones de poca importancia debidas a la reducción de nivel en las fases no de voz. Sin embargo, según una realización alternativa, la reducción de nivel no es constante. En cambio, el generador de mezcla descendente 42 puede estar configurado para conmutar entre un modo en el que la reducción de nivel está desconectada, y un modo en el que la reducción de nivel está conectada. En otras palabras, el generador de mezcla descendente 42 puede estar configurado para variar la cantidad de reducción de nivel de una manera variable con el tiempo. La variación puede ser de una naturaleza binaria o analógica, entre cero y un valor máximo. El generador de mezcla descendente 42 puede estar configurado para realizar la conmutación de modo o la variación de cantidad de reducción de nivel dependiendo de la información contenida dentro de la señal multicanal 18. Por ejemplo, el generador de mezcla descendente 42 puede estar configurado para detectar fases de voz o distinguir estas fases de voz de las fases no de voz, o puede asignar una medida de contenido de voz que mide el contenido de voz, que es de al menos escala ordinal, a tramas consecutivas del canal central. Por ejemplo, el generador de mezcla descendente 42 detecta la presencia de voz en el canal central por medio de un filtro de voz y determina en cuanto a si el nivel de salida de este filtro excede el umbral de suma. Sin embargo, la detección de fases de voz dentro del canal central mediante el generador de mezcla descendente 42 no es el único modo de hacer la conmutación de modo anteriormente mencionada de variación de cantidad de reducción de nivel dependiente del tiempo. Por ejemplo, la señal multicanal 18 podría tener información secundaria asociada con ella, lo cual está pensado especialmente para distinguir entre fases de voz y fases no de voz, o medir cuantitativamente el contenido de voz. En este caso, el generador de mezcla descendente 42 operaría en respuesta a esta información secundaria. Otra probabilidad sería que el generador de mezcla descendente 42 realizara la conmutación de modo anteriormente mencionada o las variaciones de cantidad de reducción de nivel dependiendo de una comparación entre, por ejemplo, los niveles actuales del canal central, el canal izquierdo, y el canal derecho. En caso de que el canal central sea mayor que los canales izquierdo y derecho, ya sea individualmente o en relación con la suma de los mismos, en más de una cierta proporción umbral, entonces el generador de mezcla descendente 42 puede suponer que una fase de voz está actualmente presente y actuar en consecuencia, es decir, realizando la reducción de nivel. Igualmente, el generador de mezcla descendente 42 puede usar las diferencias de nivel entre los canales central, izquierdo y derecho con el fin de producir las dependencias anteriormente mencionadas.
- [0037]** Además de esto, el generador de mezcla descendente 42 puede ser sensible a parámetros espaciales usados para describir la imagen espacial de los múltiples canales de la señal multicanal 18. Esto se muestra en la fig. 5. La fig. 5 muestra un ejemplo del generador de mezcla descendente 42 en caso de que la señal multicanal 18 represente una pluralidad de canales mediante el uso de codificación de audio especial, es decir usando una señal de mezcla descendente 62 dentro de la cual se ha realizado la mezcla descendente de la pluralidad de canales y parámetros espaciales 64 que describen la imagen espacial de la pluralidad de canales. Opcionalmente, la señal multicanal 18 también puede comprender la información que describe las proporciones en las cuales han sido mezclados los canales individuales dentro de la señal de mezcla descendente 62, o los canales individuales de la señal de mezcla descendente 62, ya que el canal de mezcla descendente 62 puede ser, por ejemplo, una señal de mezcla descendente normal 62 o una señal de mezcla descendente estéreo 62. El generador de mezcla descendente 42 de la fig. 5 comprende un descodificador 64 y un mezclador 66. El descodificador 64 descodifica, según una descodificación de audio espacial, la señal multicanal 18 con el fin de obtener la pluralidad de canales incluyendo, entre otros, el canal central 66, y otros canales 68. El mezclador 66 está configurado para mezclar el canal central 66 y los otros canales no centrales 68 para obtener la señal mono o estéreo 48 realizando la reducción de nivel anteriormente mencionada. Tal como se indica por la línea discontinua 70, el mezclador 66 puede estar configurado para usar el parámetro espacial 64 con el fin de conmutar entre el modo de reducción de nivel y el modo de no reducción de nivel de la cantidad variada de reducción de nivel, tal como se mencionó anteriormente. El parámetro espacial 64 usado por el mezclador 66 puede, por ejemplo, ser los coeficientes de predicción de canal que describen cómo el canal central 66, un canal izquierdo o el canal derecho pueden obtenerse a partir de la señal de mezcla descendente 62, en el que el mezclador 66 puede usar además parámetros de coherencia/correlación cruzada entre canales que representan la coherencia o la correlación cruzada entre los canales izquierdo y derecho recién mencionados los cuales, a su vez, pueden ser mezclas descendentes de los canales delantero izquierdo y trasero izquierdo, y los canales delantero derecho y trasero derecho, respectivamente. Por ejemplo, el canal central puede mezclarse en una proporción fija dentro del canal izquierdo anteriormente mencionado y el canal derecho de la señal de mezcla descendente estéreo 62. En este caso, son suficientes dos coeficientes de predicción de canal con el fin de determinar cómo pueden obtenerse los canales central, izquierdo y derecho a partir de una combinación lineal respectiva de los dos canales de la señal de mezcla descendente estéreo 62. Por ejemplo, el mezclador 66 puede usar una proporción entre una suma y una diferencia de los coeficientes de predicción de canal con el fin de diferenciar entre las fases de voz y las fases no de voz.

[0038] Aunque la reducción de nivel con respecto al canal central se ha descrito con el fin de ejemplificar la suma ponderada de la pluralidad de canales de manera que los mismos contribuyan a la mezcla descendente mono o estéreo a un nivel que difiere entre al menos dos canales de la señal multicanal 18, también hay otros ejemplos en los que, ventajosamente, se reduce el nivel o se amplifica el nivel de otros canales en relación con otro canal u otros canales porque algún contenido de la fuente de sonido presente en este o estos canales ha/han de, o no ha/han de ser, sometido al procesamiento de espacio al mismo nivel que otros contenidos en la señal multicanal pero a un nivel reducido/aumentado.

[0039] La fig. 5 se explicó de manera más bien general con respecto a una posibilidad de representar la pluralidad de canales de entrada por medio de una señal de mezcla descendente 62 y parámetros espaciales 64. Con respecto a la fig. 6, esta descripción se intensifica. La descripción con respecto a la fig. 6 también se usa para la comprensión de las siguientes realizaciones descritas con respecto a las figs. 10 a 13. La fig. 6 muestra la señal de mezcla descendente 62 descompuesta espectralmente en una pluralidad de subbandas 82. En la fig. 6, las subbandas 82 se muestran de manera ejemplar extendiéndose horizontalmente estando las subbandas 82 dispuestas con la frecuencia de subbanda aumentando de abajo a arriba tal como se indica por la flecha del dominio de frecuencia 84. La extensión a lo largo de la dirección horizontal denotará el eje de tiempo 86. Por ejemplo, la señal de mezcla descendente 62 comprende una secuencia de valores espectrales 88 por subbanda 82. La resolución de tiempo a la que las subbandas 82 son muestreadas por los valores de muestra 88 puede definirse mediante intervalos de bancos de filtros 90. Por lo tanto, los intervalos de tiempo 90 y las subbandas 82 definen alguna resolución o cuadrícula de tiempo/frecuencia. Se define una cuadrícula de tiempo/frecuencia más gruesa uniendo valores de muestra adyacentes 88 a baldosas de tiempo/frecuencia 92 tal como se indica por las líneas discontinuas en la fig. 6, definiendo estas baldosas la resolución o cuadrícula de parámetros de tiempo/frecuencia. Los parámetros espaciales anteriormente mencionados 62 están definidos en esa resolución de parámetros de tiempo/frecuencia 92. La resolución de parámetros de tiempo/frecuencia 92 puede cambiar en el tiempo. Con este fin, la señal multicanal 62 puede dividirse en tramas consecutivas 94. Para cada trama, la cuadrícula de resolución de tiempo/frecuencia 92 puede establecerse individualmente. En caso de que el descodificador 64 reciba la señal de mezcla descendente 62 en el dominio del tiempo, el descodificador 64 puede estar comprendido por un banco de filtros de análisis interno con el fin de obtener la representación de la señal de mezcla descendente 62 tal como se muestra en la fig. 6. Alternativamente, la señal de mezcla descendente 62 entra en el descodificador 64 en la forma tal como se muestra en la fig. 6, en cuyo caso no es necesario ningún banco de filtros de análisis en el descodificador 64. Tal como ya se mencionó en la fig. 5, para cada baldosa 92 pueden estar presentes dos coeficientes de predicción de canal que revelan cómo, con respecto a la baldosa de tiempo/frecuencia respectiva 92, pueden obtenerse los canales derecho e izquierdo a partir de los canales izquierdo y derecho de la señal de mezcla descendente estéreo 62. Además, puede estar presente un parámetro de coherencia/correlación cruzada entre canales (ICC) para la baldosa 92 que indica las similitudes de ICC entre el canal izquierdo y el derecho que han de obtenerse a partir de la señal de mezcla descendente estéreo 62, en el que un canal ha sido mezclado completamente dentro de un canal de la señal de mezcla descendente estéreo 62, mientras que el otro ha sido mezclado completamente dentro del otro canal de la señal de mezcla descendente estéreo 62. Sin embargo, además puede estar presente un parámetro de diferencia de nivel de canal (CLD) para cada baldosa 92 que indica la diferencia de nivel entre los canales izquierdo y derecho recién mencionados. Puede aplicarse una cuantificación no uniforme en una escala logarítmica a los parámetros de CLD, donde la cuantificación tiene una elevada exactitud cercana a cero dB y una resolución más gruesa cuando hay una gran diferencia de nivel entre los canales. Además, pueden estar presentes parámetros adicionales dentro del parámetro espacial 64. Estos parámetros pueden, entre otras cosas, definir la CLD y la ICC relacionadas con los canales que servían para formar, mediante mezcla, los canales izquierdo y derecho recién mencionados, tales como los canales trasero izquierdo, delantero izquierdo, trasero derecho y delantero derecho.

[0040] Cabe destacar que las realizaciones anteriormente mencionadas pueden combinarse entre sí. Algunas posibilidades de combinación ya han sido mencionadas anteriormente. En lo que viene a continuación se mencionarán más posibilidades con respecto a las realizaciones de las figs. 7 a 13. Además, las realizaciones anteriormente mencionadas de las figs. 1 y 5 suponían que los canales intermedios 20, 66 y 68, respectivamente, están realmente presentes dentro del dispositivo. Sin embargo, este no es necesariamente el caso. Por ejemplo, las HRTF modificadas tal como se obtienen mediante el dispositivo de la fig. 2 pueden usarse para definir los filtros direccionales de la fig. 1 omitiendo el reductor de similitud 12, y en este caso, el dispositivo de la fig. 1 puede operar sobre una señal de mezcla descendente tal como la señal de mezcla descendente 62 mostrada en la fig. 5, que representa la pluralidad de canales 18a-18d, combinando de manera adecuada los parámetros espaciales y las HRTF modificadas en la resolución de parámetros de tiempo/frecuencia 92, y aplicando los coeficientes de combinación lineal obtenidos en consecuencia con el fin de formar las señales binaurales 22a y 22b.

[0041] Igualmente, el generador de mezcla descendente 42 puede estar configurado para combinar de manera

adecuada los parámetros espaciales 64 y la cantidad de reducción de nivel que ha de conseguirse para el canal central con el fin de obtener la mezcla descendente mono o estéreo 48 pensada para el procesador de espacio 44. La fig. 7 muestra un generador de señales de salida binaurales según una realización. Un generador que se indica en general con el signo de referencia 100 comprende un descodificador multicanal 102, una salida binaural 104, y dos recorridos que se extienden entre la salida del descodificador multicanal 102 y la salida binaural 104, respectivamente, concretamente un recorrido directo 106 y un recorrido de reverberación 108. En el recorrido directo, los filtros direccionales 110 están conectados a la salida del descodificador multicanal 102. El recorrido directo comprende además un primer grupo de sumadores 112 y un segundo grupo de sumadores 114. Los sumadores 112 suman la señal de salida de una primera mitad de los filtros direccionales 110 y los segundos sumadores 114 suman la señal de salida de una segunda mitad de los filtros direccionales 110. Las salidas sumadas de los primeros y segundos sumadores 112 y 114 representan la contribución del recorrido directo anteriormente mencionado de la señal de salida binaural 22a y 22b. Los sumadores 116 y 118 están provistos con el fin de combinar las señales de contribución 22a y 22b con las señales de contribución binaurales provistas por el recorrido de reverberación 108, es decir, las señales 46a y 46b. En el recorrido de reverberación 108, un mezclador 120 y un procesador de espacio 122 están conectados en serie entre la salida del descodificador multicanal 102 y la entrada respectiva de los sumadores 116 y 118, cuyas salidas definen la señal de salida binaural producida como salida en la salida 104.

[0042] Con el fin de facilitar la comprensión de la siguiente descripción del dispositivo de la fig. 7, los signos de referencia usados en las figs. 1 a 6 se han usado parcialmente con el fin de denotar los elementos de la fig. 7, los cuales corresponden a aquellos, o asumir la responsabilidad de la funcionalidad de los elementos que se produce en las figs. 1 a 6. La descripción correspondiente resultará más clara en la siguiente descripción. Sin embargo, se observa que, con el fin de facilitar la siguiente descripción, las siguientes realizaciones se han descrito con la suposición de que el reductor de similitud realiza una reducción de correlación. Por consiguiente, en lo que viene a continuación, éste se denota como un reductor de correlación. Sin embargo, como resultó claro de lo anterior, las realizaciones explicadas resumidamente más adelante son fácilmente transferibles a los casos en los que el reductor de similitud realiza una reducción de similitud aparte de en cuanto a correlación. Además, las realizaciones explicadas resumidamente más adelante se han redactado suponiendo que el mezclador para generar la mezcla descendente para el procesamiento de espacio genera una reducción de nivel del canal central aunque, tal como se describió anteriormente, se podría conseguir fácilmente una transferencia a realizaciones alternativas.

[0043] El dispositivo de la fig. 7 usa un flujo de señales para la generación de una salida de auriculares en la salida 104 a partir de una señal multicanal descodificada 124. El multicanal descodificado 124 se obtiene mediante el descodificador multicanal 102 a partir de un tren de bits introducido en una entrada de tren de bits 126, tal como, por ejemplo, mediante descodificación de audio espacial. Después de la descodificación, cada señal o canal de la señal multicanal descodificada 124 es filtrado por un par de filtros direccionales 110. Por ejemplo, el primer canal (superior) de la señal multicanal descodificada 124 es filtrado por los filtros direccionales 20 DirFilter(1,L) y DirFilter(1,R), y una segunda señal (segunda desde arriba) o canal es filtrado por el filtro direccional DirFilter(2,L) y DirFilter(2,R), y así sucesivamente. Estos filtros 110 pueden modelizar la transmisión acústica desde una fuente de sonido virtual en un espacio hasta el conducto auditivo de un oyente, una función denominada de transferencia de espacio binaural (BRTF). Pueden realizar modificaciones de tiempo, de nivel y espectrales, y también pueden modelizar parcialmente la reflexión y la reverberación de un espacio. Los filtros direccionales 110 pueden implementarse en los dominios del tiempo o de la frecuencia. Como se requieren muchos filtros 110 ($N \times 2$, siendo N el número de canales descodificados), estos filtros direccionales podrían, si debieran modelizar la reflexión y la reverberación de un espacio completamente, ser bastante largos, es decir, 20000 derivaciones de filtro a 44,1 kHz, en cuyo caso el procedimiento de filtrado sería exigente computacionalmente. Los filtros direccionales 110 se reducen ventajosamente al mínimo, las denominadas funciones de transferencia relacionadas con la cabeza (HRTF) y el bloque de procesamiento común 122 se usa para modelizar las reflexiones y reverberaciones de un espacio. El módulo de procesamiento de espacio 122 puede implementar un algoritmo de reverberación en un dominio del tiempo o de la frecuencia y puede operar desde una señal de entrada de uno o dos canales 48, la cual se calcula a partir de la señal de entrada multicanal descodificada 124 mediante una matriz de mezclado dentro del mezclador 120. El bloque de procesamiento de espacio implementa las reflexiones y/ la reverberación de un espacio. Las reflexiones y la reverberación de un espacio son esenciales para localizar los sonidos, especialmente con respecto a la distancia y los sonidos que significan externalización son percibidos fuera de la cabeza del oyente.

[0044] Típicamente, el sonido multicanal se produce de manera que la energía del sonido dominante está contenida en los canales delanteros, es decir, delantero izquierdo, delantero derecho, central. Las voces en los diálogos de películas y la música típicamente se mezclan principalmente en el canal central. Si las señales del canal central se suministran al módulo de procesamiento de espacio 122, la salida resultante a menudo es percibida

reverberante de manera poco natural y espectralmente desigual. Por lo tanto, según la realización de la fig. 7, el canal central se suministra al módulo de procesamiento de espacio 122 con una reducción de nivel significativa, tal como atenuada en 6 dB, reducción de nivel que se realiza, como ya se denotó anteriormente, dentro del mezclador 120. Hasta ahora, la realización de la fig. 7 comprende una configuración según las figs. 3 y 5, en la que los signos de referencia 102, 124, 120 y 122 de la fig. 7 corresponden a los signos de referencia 18, 64, la combinación de los signos de referencia 66 y 68, el signo de referencia 66 y el signo de referencia 44 de las figs. 3 y 5, respectivamente.

[0045] La fig. 8 muestra otro generador de señales de salida binaurales según una realización adicional. El generador se indica en general con el signo de referencia 140. Con el fin de facilitar la descripción de la fig. 8, se han usado los mismos signos de referencia que en la fig. 7. Con el fin de denotar que el mezclador 120 no tiene necesariamente la funcionalidad tal como se indica con las realizaciones de las figs. 3, 5 y 7, concretamente la realización de la reducción de nivel con respecto al canal central, se ha usado el signo de referencia 40' con el fin de denotar la disposición de los bloques 102, 120 y 122, respectivamente. En otras palabras, la reducción de nivel dentro del mezclador 122 es opcional en el caso de la fig. 8. A diferencia de la fig. 7, sin embargo, los descorrelacionadores están conectados entre cada par de filtros direccionales 110 y la salida del descodificador 102 para el canal asociado de la señal multicanal descodificada 124, respectivamente. Los descorrelacionadores se indican con los signos de referencia 142₁, 142₂, y así sucesivamente. Los descorrelacionadores 142₁ - 142₄ actúan como el reductor de correlación 12 indicado en la fig. 1. Aunque se muestra en la fig. 8, no es necesario que esté provisto un descorrelacionador 142₁ - 142₄ para cada uno de los canales de la señal multicanal descodificada 124. En cambio, sería suficiente un descorrelacionador. Los descorrelacionadores 142 podrían ser simplemente un retardo. Con preferencia, la cantidad de retardo causada por cada uno de los retardos 142₁ - 142₄ sería diferente entre sí. Otra posibilidad sería que los descorrelacionadores 142₁ - 142₄ sean filtros de todo paso, es decir, filtros que tienen una función de transferencia de una magnitud de ser constantemente uno, sin embargo, cambiando las fases de las componentes espectrales del canal respectivo. Las modificaciones de fase causadas por los descorrelacionadores 142₁ - 142₄ serían, con preferencia, diferentes para cada uno de los canales. Por supuesto, también existirían otras posibilidades. Por ejemplo, el descorrelacionador 142₁ - 142₄ podría implementarse como filtros FIR, o similares.

[0046] Por lo tanto, según la realización de la fig. 8, los elementos 142₁ - 142₄, 110, 112 y 114 actúan de acuerdo con el dispositivo 10 de la fig. 1.

[0047] Igualmente a la fig. 8, la fig. 9 muestra una variación del generador de señales de salida binaurales de la fig. 7. Por lo tanto, la fig. 9 también se explica más adelante usando los mismos signos de referencia que se usan en la fig. 7. Igualmente a la realización de la fig. 8, la reducción de nivel del mezclador 122 es meramente opcional en el caso de la fig. 9 y, por lo tanto, el signo de referencia 40' ha estado en la fig. 9 en lugar de '40, como era el caso en la fig. 7. La realización de la fig. 9 se ocupa del problema de que existe una correlación significativa entre todos los canales en las producciones de sonido multicanal. Después del procesamiento de las señales multicanal con los filtros direccionales 110, las señales intermedias de dos canales de cada par de filtros se suman mediante los sumadores 112 y 114, para formar la señal de salida de auriculares en la salida 104. La suma de las señales de salida correlacionadas por los sumadores 112 y 114 tiene como resultado una anchura espacial reducida en gran medida de la señal de salida en la salida 104, y una falta de una externalización. Esto es particularmente problemático para la correlación de la señal izquierda y derecha y el canal central dentro de la señal multicanal descodificada 124. Según la realización de la fig. 9, los filtros direccionales están configurados para que tengan una salida descorrelacionada en la medida de lo posible. Con este fin, el dispositivo de la fig. 9 comprende el dispositivo 30 para formar un conjunto decreciente de intercorrelación de HRTF que ha ser usado por los filtros direccionales 110 basándose en algún conjunto original de HRTF. Tal como se describió anteriormente, el dispositivo 30 puede usar una, o una combinación de, las siguientes técnicas con respecto a las HRTF del par de filtros direccionales asociado con uno o varios canales de la señal multicanal descodificada 124:

retardar el filtro direccional o el par de filtros direccionales respectivos tal como, por ejemplo, desplazando la respuesta impulsional de los mismos, lo cual podría hacerse, por ejemplo, desplazando las derivaciones de filtro;

modificar la respuesta de fase de los filtros direccionales respectivos; y

aplicar un filtro de descorrelación tal como un filtro de todo paso a los filtros direccionales respectivos del canal respectivo. Tal filtro de todo paso podría implementarse como un filtro FIR.

[0048] Tal como se describió anteriormente, el dispositivo 30 podría operar en respuesta al cambio en la configuración del altavoz para el cual está pensado el tren de bits en la entrada de tren de bits 126.

[0049] Las realizaciones de las figs. 7 a 9 concernían a una señal multicanal descodificada. Las siguientes realizaciones se ocupan de la descodificación multicanal paramétrica para auriculares.

5 **[0050]** En general, la codificación de audio espacial es una técnica de compresión multicanal que aprovecha la irrelevancia perceptiva entre canales en las señales de audio multicanal para conseguir tasas de compresión más elevadas. Esto puede captarse en cuanto a pistas espaciales o parámetros espaciales, es decir, parámetros que describen la imagen espacial de una señal de audio multicanal. Las pistas espaciales incluyen típicamente diferencias de nivel/intensidad, diferencias de fase y medidas de las correlaciones/la coherencia entre canales, y
 10 pueden representarse de una manera extremadamente compacta. El concepto de codificación de audio espacial ha sido adoptado por el MPEG con el resultado de la norma de sonido envolvente MPEG, es decir, la ISO/11EC23003-1. Los parámetros espaciales tales como los empleados en la codificación de audio espacial también pueden emplearse para describir filtros direccionales. De hacerlo así, la etapa de descodificar datos de audio espacial y aplicar filtros direccionales puede combinarse para descodificar eficientemente e interpretar audio multicanal para
 15 reproducción por auriculares.

[0051] La estructura general de un descodificador de audio espacial para salida por auriculares se ofrece en la fig. 10. El descodificador de la fig. 10 se indica en general con el signo de referencia 200, y comprende un modificador de subbanda espacial binaural 202 que comprende una entrada para una señal de mezcla descendente estéreo o
 20 mono 204, otra entrada para parámetros espaciales 206, y una salida para la señal de salida binaural 208. La señal de mezcla descendente junto con los parámetros espaciales 206 forman la señal multicanal anteriormente mencionada 18 y representan la pluralidad de canales de la misma.

[0052] Internamente, el modificador de subbanda 202 comprende un banco de filtros de análisis 208, una unidad de matrizado o combinador lineal 210 y un banco de filtros de síntesis 212 conectados en el orden mencionado entre la entrada de señal de mezcla descendente y la salida del modificador de subbanda 202. Además, el modificador de subbanda 202 comprende un convertidor de parámetros 214 al que se le suministran los parámetros espaciales 206 y un conjunto modificado de HRTF tal como se obtiene mediante el dispositivo 30.

30 **[0053]** En la fig. 10, se supone que la señal de mezcla descendente ya ha sido descodificada de antemano, incluyendo, por ejemplo, codificación de entropía. Al descodificador de audio espacial binaural se le suministra la señal de mezcla descendente 204. El convertidor de parámetros 214 usa los parámetros espaciales 206 y la descripción paramétrica de los filtros direccionales en forma del parámetro de HRTF modificado 216 para formar los parámetros binaurales 218. Estos parámetros 218 son aplicados por la unidad de matrizado 210 en forma de una matriz de dos por dos (en caso de una señal de mezcla descendente estéreo) y en forma de una matriz de uno por dos (en caso de una señal de mezcla descendente mono 204), en el dominio de la frecuencia, a los valores espectrales 88 producidos como salida por el banco de filtros de análisis 208 (véase la fig. 6). En otras palabras, los parámetros binaurales 218 varían en la resolución de parámetros de tiempo/frecuencia 92 mostrada en la fig. 6 y se aplican a cada valor de muestra 88. Puede usarse interpolación para alisar los coeficientes de la matriz y los
 40 parámetros binaurales 218, respectivamente, del dominio de parámetros de tiempo/frecuencia más gruesos 92 a la resolución de tiempo/frecuencia del banco de filtros de análisis 208. Es decir, en el caso de una mezcla descendente estéreo 204, el matrizado realizado por la unidad 210 tiene como resultado dos valores de muestra por par de valor de muestra del canal izquierdo de la señal de mezcla descendente 204 y el valor de muestra correspondiente del canal derecho de la señal de mezcla descendente 204. Los dos valores de muestra resultantes son parte de los canales izquierdo y derecho de la señal de salida binaural 208, respectivamente. En caso de una señal de mezcla descendente mono 204, el matrizado mediante la unidad 210 tiene como resultado dos valores de muestra por valor de muestra de la señal de mezcla descendente mono 204, concretamente uno para el canal izquierdo y uno para el canal derecho de la señal de salida binaural 208. Los parámetros binaurales 218 definen la operación de matrizado que conduce del uno o dos valores de muestra de la señal de mezcla descendente 204 a los valores de muestra del canal izquierdo y derecho respectivos de la señal de salida binaural 208. Los parámetros binaurales 218 ya reflejan los parámetros de HRTF modificados. Por lo tanto, descorrelacionan los canales de entrada de la señal multicanal 18 tal como se indicó anteriormente.

55 **[0054]** Por lo tanto, la salida de la unidad de matrizado 210 es un espectrograma modificado tal como se muestra en la fig. 6. El banco de filtros de síntesis 212 reconstruye a partir del mismo la señal de salida binaural 208. En otras palabras, el banco de filtros de síntesis 212 convierte la señal de dos canales resultada producida como salida por la unidad de matrizado 210 al dominio del tiempo. Por supuesto, esto es opcional.

[0055] En el caso de la fig. 10, los efectos de reflexión y reverberación de un espacio no se trataron por separado.

Si acaso alguna vez, estos efectos han de tenerse en cuenta en las HRTF 216. La fig. 11 muestra un generador de señales de salida binaurales que combina un descodificador de audio espacial 200' con procesamiento separado de reflexión/reverberación de un espacio. El ' del signo de referencia 200' en la fig. 11 denotará que el descodificador de audio espacial binaural 200' de la fig. 11 puede usar HRTF sin modificar, es decir, las HRTF originales tal como se indican en la fig. 2. Opcionalmente, sin embargo, el descodificador de audio espacial binaural 200' de la fig. 11 puede ser el mostrado en la fig. 10. En cualquier caso, el generador de señales de salida binaurales de la fig. 11 que se indica en general con el signo de referencia 230 comprende, además del descodificador espacial binaural 200', un descodificador de audio de mezcla descendente 232, un modificador de subbanda de audio espacial modificado 234, un procesador de espacio 122, y dos sumadores 116 y 118. El descodificador de audio de mezcla descendente 232 está conectado entre una entrada de tren de bits 126 y un modificador de subbanda de audio espacial binaural 202 del descodificador de audio espacial binaural 200'. El descodificador de audio de mezcla descendente 232 está configurado para descodificar la entrada de tren de bits en la entrada 126 para obtener la señal de mezcla descendente 214 y los parámetros espaciales 206. Tanto al modificador de subbanda de audio espacial binaural 202, como al modificador de subbanda de audio espacial modificado 234 se les proporciona una señal de subbanda 204 además de los parámetros espaciales 206. El modificador de subbanda de audio espacial modificado 234 calcula a partir de la señal de mezcla descendente 204 – mediante el uso de los parámetros espaciales 206 así como los parámetros modificados 236 que reflejan la cantidad anteriormente mencionada de reducción de nivel del canal central – la mezcla descendente mono o estéreo 48 que sirve como entrada para el procesador de espacio 122. Las contribuciones producidas como salida tanto por el modificador de subbanda de audio espacial modificado 202 como el procesador de espacio 122, respectivamente, se suman en lo que respecta a los canales en los sumadores 116 y 118 para obtener como resultado la señal de salida binaural en la salida 238.

[0056] La fig. 12 muestra un diagrama de bloques que ilustra la funcionalidad del descodificador de audio binaural 200' de la fig. 11. Cabe destacar que la fig. 12 no muestra la estructura interna real del descodificador de audio espacial binaural 200' de la fig. 11, sino que ilustra las modificaciones de señal obtenidas por el descodificador de audio espacial binaural 200'. Se recuerda que la estructura interna del descodificador de audio espacial binaural 200' obedece en general a la estructura mostrada en la fig. 10, con la excepción de que el dispositivo 30 puede omitirse en el caso de que el mismo esté operando con las HRTF originales. Además, la fig. 12 muestra la funcionalidad del descodificador de audio espacial binaural 200' a modo de ejemplo para el caso de que sólo se usen tres canales representados por la señal multicanal 18 por parte del descodificador de audio espacial binaural 200' con el fin de formar la señal de salida binaural 208. En particular, se usa una caja de "2 a 3", es decir TTT, para obtener un canal central 242, un canal derecho 244, y un canal izquierdo 246 a partir de los dos canales de la mezcla descendente estéreo 204. En otras palabras, la fig. 12 supone a modo de ejemplo que la mezcla descendente 204 es una mezcla descendente estéreo. Los parámetros espaciales 206 usados por la caja TTT 248 comprenden los coeficientes de predicción de canal anteriormente mencionados. La reducción de correlación se consigue mediante tres descorrelacionadores, denotados como DelayL, DelayR, y DelayC en la fig. 12. Corresponden a la descorrelación introducida en el caso de, por ejemplo, las figs. 1 y 7. Sin embargo, se recuerda de nuevo que la fig. 12 simplemente muestra las modificaciones de señal conseguidas por el descodificador de audio espacial binaural 200', aunque la estructura real corresponde a la mostrada en la fig. 10. Por lo tanto, aunque los retardos que forma el reductor de correlación 12 se muestran como características separadas en relación con las HRTF que forman los filtros direccionales 14, la existencia de los retardos en el reductor de correlación 12 puede verse como una modificación de los parámetros de HRTF que forman las HRTF originales de los filtros direccionales 14 de la fig. 12. En primer lugar, la fig. 12 muestra simplemente que el descodificador de audio espacial binaural 200' descorrelaciona los canales para reproducción por auriculares. La descorrelación se consigue por medios sencillos, concretamente, sumando un bloque de retardo en el procesamiento paramétrico para la matriz M y el descodificador de audio espacial binaural 200'. Por lo tanto, el descodificador de audio espacial binaural 200' puede aplicar las siguientes modificaciones a los canales individuales, concretamente

- retardar el canal central con preferencia al menos una muestra,
- retardar el canal central diferentes intervalos en cada banda de frecuencia,
- retardar los canales izquierdo y derecho con preferencia al menos una muestra y/o
- retardar los canales izquierdo y derecho diferentes intervalos en cada banda de frecuencia.

[0057] La fig. 13 muestra un ejemplo para una estructura del modificador de subbanda de audio espacial modificado de la fig. 11. El modificador de subbanda 234 de la fig. 13 comprende una caja de dos a tres o TTT 262, etapas de ponderación 264a-264e, primeros sumadores 266a y 266b, segundos sumadores 268a y 268b, una

entrada para la mezcla descendente estéreo 204, una entrada para los parámetros espaciales 206, una entrada adicional para una señal residual 270 y una salida para la mezcla descendente 48 pensada para ser procesada por el procesador de espacio, y que es, de acuerdo con la fig. 13, una señal estéreo.

5 **[0058]** Como la fig. 13 define en un sentido estructural una realización para el modificador de subbanda de audio espacial modificado 234, la caja TTT 262 de la fig. 13 simplemente reconstruye el canal central, el canal derecho 244, y el canal izquierdo 246 a partir de la mezcla descendente estéreo 204 usando los parámetros espaciales 206. Se recuerda una vez más que en el caso de la fig. 12, los canales 242-246 no se calculan realmente. En cambio, el modificador de subbanda de audio espacial binaural modifica la matriz M de tal manera que la señal de mezcla
10 descendente estéreo 204 se convierte directamente en la contribución binaural que refleja las HRTF. La caja TTT 262 de la fig. 13, sin embargo, realiza realmente la reconstrucción. Opcionalmente, tal como se muestra en la fig. 13, la caja TTT 262 puede usar una señal residual 270 que refleja el residuo de predicción cuando se reconstruyen los canales 246-246 basándose en la mezcla descendente estéreo 204 y los parámetros espaciales 206, los cuales como se denotó anteriormente, comprenden los coeficientes de predicción de canal y, opcionalmente, los valores
15 ICC. Los primeros sumadores 266a están configurados para sumar los canales 242-246 para formar el canal izquierdo de la mezcla descendente estéreo 48. En particular, se forma una suma ponderada mediante los sumadores 266a y 266b, en la que los valores de ponderación se definen por las etapas de ponderación 264a, 264b, 264c, y 264e que podrían aplicarse al canal respectivo 246 a 242, un valor de ponderación respectivo EQ^{LL} , EQ^{RL} y EQ^{CL} . Igualmente, los sumadores 268a y 268b forman una suma ponderada de los canales 246 a 242 con las etapas
20 de ponderación 264b, 264d, y 264e que forman los valores de ponderación, formando la suma ponderada el canal derecho de la mezcla descendente estéreo 48

[0059] Los parámetros 270 para las etapas de ponderación 264a-264e se seleccionan, tal como se describió anteriormente, de manera que se consigue la reducción de nivel del canal central descrita anteriormente en la
25 mezcla descendente estéreo 48, con el resultado, tal como se describió anteriormente, de las ventajas con respecto a percepción de sonido natural.

[0060] Por lo tanto, en otras palabras, la fig. 13 muestra un módulo de procesamiento de espacio que puede aplicarse en combinación con el descodificador paramétrico binaural 200' de la fig. 12. En la fig. 13, la señal de
30 mezcla descendente 204 se usa para suministrar al módulo. La señal de mezcla descendente 204 contiene todas las señales de la señal multicanal para poder proporcionar compatibilidad estéreo. Tal como se mencionó anteriormente, es deseable suministrar al módulo de procesamiento de espacio una señal que contenga sólo una señal central reducida. El modificador de subbanda de audio espacial modificado de la fig. 13 sirve para realizar esta reducción de nivel. En particular, según la fig. 13, puede usarse una señal residual 270 con el fin de reconstruir
35 canales central, izquierdo y derecho 242-246. La señal residual de los canales central y el izquierdo y derecho 242-246 puede ser descodificada por el descodificador de audio de mezcla descendente 232, aunque no se muestre en la fig. 11. Los parámetros de EQ o valores de ponderación aplicados por las etapas de ponderación 264a-264e pueden ser valores reales para los canales izquierdo, derecho y central 242-246. Puede almacenarse y aplicarse un único conjunto de parámetros para el canal central 242, y el canal central se mezcla igualmente a modo de ejemplo,
40 según la fig. 13, en ambas salidas izquierda y derecha de la mezcla descendente estéreo 48.

[0061] Los parámetros de EQ 270 suministrados al modificador de subbanda de audio espacial modificado 234 pueden tener las siguientes propiedades. En primer lugar, la señal del canal central puede atenuarse, con
45 preferencia, en al menos 6 dB. Además, la señal del canal central puede tener una característica de paso bajo. Aún más, la señal de diferencia de los canales restantes puede intensificarse a bajas frecuencias. Con el fin de compensar el nivel más bajo del canal central 242 en relación con los otros canales 244 y 246, la ganancia de los parámetros de HRTF para el canal central usados en el modificador de subbanda de audio espacial binaural 202 deberían aumentarse en consecuencia.

50 **[0062]** El objetivo principal del ajuste de los parámetros de EQ es la reducción de la señal del canal central en la salida para el módulo de procesamiento de espacio. Sin embargo, el canal central sólo debería suprimirse hasta cierto límite: la señal del canal central se resta de los canales de mezcla descendente izquierdo y derecho dentro de la caja TTT. Si se reduce el nivel central, pueden resultar audibles artefactos en el canal izquierdo y derecho. Por lo tanto, la reducción de nivel central en la etapa de EQ es una solución de compromiso entre la supresión y los
55 artefactos. Encontrar un ajuste fijo de los parámetros de EQ es posible, pero puede no ser óptimo para todas las señales. Por consiguiente, según una realización, puede usarse un algoritmo o módulo adaptativo 274 para controlar la cantidad de reducción de nivel central por parte de uno, o una combinación de los siguientes parámetros:

[0063] Los parámetros espaciales 206 usados para descodificar el canal central 242 a partir del canal de mezcla

descendente izquierdo y derecho 204 dentro de la caja TTT 262 pueden usarse como se indica por la línea discontinua 276.

5 **[0064]** El nivel de los canales central, izquierdo y derecho puede usarse como se indica por la línea discontinua 278.

[0065] Las diferencias de nivel entre los canales central, izquierdo y derecho 242-246 pueden usarse como también se indica por la línea discontinua 278.

10 **[0066]** La salida de un algoritmo de detección de un solo tipo, tal como un detector de actividad vocal, puede usarse como también se indica por la línea discontinua 278.

15 **[0067]** Por último, puede usarse estática de metadatos dinámicos que describen el contenido de audio con el fin de determinar la cantidad de reducción de nivel central como se indica por la línea discontinua 280.

15 **[0068]** Aunque algunos aspectos se han descrito en el contexto de un aparato, resulta claro que estos aspectos también representan una descripción del procedimiento correspondiente, en el que un bloque o dispositivo corresponde a una etapa de procedimiento o una característica de una etapa de procedimiento. Análogamente, los aspectos descritos en el contexto de una etapa de procedimiento también representan una descripción de un bloque
20 o elemento o característica correspondiente de un aparato correspondiente tal como una parte de un ASIC, una subrutina de un código de programa o una parte de una lógica programable programada.

[0069] La señal de acudo codificada inventiva puede almacenarse en un medio de almacenamiento digital o puede transmitirse por un medio de transmisión tal como un medio de transmisión inalámbrico o un medio de transmisión
25 por cable tal como Internet.

[0070] Dependiendo de ciertos requisitos de implementación, las realizaciones de la invención pueden implementarse en hardware o en software. La implementación puede realizarse usando un medio de almacenamiento digital, por ejemplo un disco flexible, un DVD, un CD, una ROM, una PROM, una EPROM, una
30 EEPROM o una memoria FLASH, que tenga señales de control legibles electrónicamente almacenadas en el mismo, que coopere (o sea capaz de cooperar) con un sistema informático programable de manera que se realice el procedimiento respectivo.

[0071] Algunas realizaciones según la invención comprenden un soporte de datos que tiene señales de control
35 legibles electrónicamente, que son capaces de cooperar con un sistema informático programable, de manera que se realice uno de los procedimientos descritos en este documento.

[0072] Generalmente, las realizaciones de la presente invención pueden implementarse como un producto de programa informático con un código de programa, siendo el código de programa operativo para realizar uno de los
40 procedimientos cuando el producto de programa informático se ejecuta en un ordenador. El código de programa puede almacenarse, por ejemplo, en un soporte legible por una máquina.

[0073] Otras realizaciones comprenden el programa informático para realizar uno de los procedimientos descritos en este documento, almacenado en un soporte legible por una máquina.

45 **[0074]** En otras palabras, una realización del procedimiento inventivo es, por ejemplo, un programa informático que tiene un código de programa para realizar uno de los procedimientos descritos en este documento, cuando el programa informático se ejecuta en un ordenador.

50 **[0075]** Una realización adicional de los procedimientos inventivos es, por lo tanto, un soporte de datos (o un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, grabado en el mismo, el programa informático para realizar uno de los procedimientos descritos en este documento.

[0076] Una realización adicional del procedimiento inventivo es, por lo tanto, un tren de datos o una secuencia de
55 señales que representa el programa informático para realizar uno de los procedimientos descritos en este documento. El tren de datos o la secuencia de señales puede estar configurado, por ejemplo, para ser transferido a través de una conexión de comunicación de datos, por ejemplo a través de Internet.

[0077] Una realización adicional comprende un medio de procesamiento, por ejemplo un ordenador, o un

dispositivo lógico programable, configurado o adaptado para realizar uno de los procedimientos descritos en este documento.

[0078] Una realización adicional comprende un ordenador que tiene instalado en el mismo el programa informático para realizar uno de los procedimientos descritos en este documento.

[0079] En algunas realizaciones, puede usarse un dispositivo lógico programable (por ejemplo una matriz de puertas programable in situ) para realizar alguna o todas las funcionalidades de los procedimientos descritos en este documento. En algunas realizaciones, una matriz de puertas programable in situ puede cooperar con un microprocesador con el fin de realizar uno de los procedimientos descritos en este documento. Generalmente, los procedimientos son realizados, con preferencia, por algún aparato de hardware.

[0080] Las realizaciones anteriormente descritas son meramente ilustrativas de los principios de la presente invención. Se comprende que las modificaciones y variaciones de las disposiciones y los detalles descritos en este documento resultarán evidentes para otros expertos en la materia. Por lo tanto, la intención es que esté limitada sólo por el alcance de las reivindicaciones de la patente inminente y no por los detalles específicos presentados a modo de descripción y explicación de las realizaciones de este documento.

[0081] Por lo tanto, las realizaciones anteriores, entre otras cosas, describían un dispositivo para generar una señal binaural basándose en una señal multicanal que representa una pluralidad de canales y pensada para reproducción por una configuración de altavoces que tiene una posición de fuente de sonido binaural asociada a cada canal, que comprende un reductor de similitud 12 para procesar de manera diferente, y reducir así una similitud entre, al menos uno de un canal izquierdo y uno derecho de la pluralidad de canales, un canal delantero y uno trasero de la pluralidad de canales, y un canal central y uno no central de la pluralidad de canales, con el fin de obtener un conjunto reducido de intersimilitud 20 de canales; una pluralidad 14 de filtros direccionales para modelizar una transmisión acústica de uno respectivo del conjunto reducido de intersimilitud 20 de canales desde una posición de fuente de sonido virtual asociada con el canal respectivo del conjunto reducido de intersimilitud de canales hasta un conducto auditivo respectivo de un oyente; un primer mezclador 16a para mezclar las salidas de los filtros direccionales que modelizan la transmisión acústica hasta el primer conducto auditivo del oyente para obtener un primer canal 22a de la señal binaural; y un segundo mezclador 16b para mezclar las salidas de los filtros direccionales que modelizan la transmisión acústica hasta el segundo conducto auditivo del oyente para obtener un segundo canal 22b de la señal binaural. En este sentido, el reductor de correlación 12 puede estar configurado para realizar el diferente procesamiento causando un retardo relativo entre, o realizando – en un sentido variable espectralmente – una modificación de fase de manera diferente entre, el al menos uno de los canales izquierdo y derecho de la pluralidad de canales, los canales delantero y trasero de la pluralidad de canales, y los canales central y no central de la pluralidad de canales, y/o realizar – en un sentido variable espectralmente – una modificación de magnitud de manera diferente entre, el al menos uno de los canales izquierdo y derecho de la pluralidad de canales, los canales delantero y trasero de la pluralidad de canales, y los canales central y no central de la pluralidad de canales.

[0082] Las realizaciones anteriores también describían, entre otras cosas, un dispositivo para generar una señal binaural basándose en una señal multicanal que representa una pluralidad de canales y pensada para reproducción por una configuración de altavoces que tiene una posición de fuente de sonido virtual asociada a cada canal, que comprende un reductor de similitud 12 para realizar – en un sentido variable espectralmente – una modificación de fase y/o magnitud de manera diferente entre al menos dos canales de la pluralidad de canales, con el fin de obtener un conjunto reducido de intersimilitud 20 de canales; una pluralidad 14 de filtros direccionales para modelizar una transmisión acústica de uno respectivo del conjunto reducido de intersimilitud 20 de canales desde una posición de fuente de sonido virtual asociada con el canal respectivo del conjunto reducido de intersimilitud de canales hasta un conducto auditivo respectivo de un oyente; un primer mezclador 16a para mezclar las salidas de los filtros direccionales que modelizan la transmisión acústica hasta el primer conducto auditivo del oyente para obtener un primer canal 22a de la señal binaural; y un segundo mezclador 16b para mezclar las salidas de los filtros direccionales que modelizan la transmisión acústica hasta el segundo conducto auditivo del oyente para obtener un segundo canal 22b de la señal binaural.

[0083] Además, las realizaciones anteriores, entre otras cosas, describían un dispositivo para formar un conjunto decreciente de intersimilitud de funciones de transferencia relacionadas con la cabeza para modelizar una transmisión acústica de una pluralidad de canales desde una posición de fuente de sonido virtual asociada con el canal respectivo hasta los conductos auditivos de un oyente, comprendiendo el dispositivo un proveedor de HRTF 32 para proporcionar una pluralidad original de HRTF; y un procesador de HRTF 34 para hacer que las respuestas

impulsionales de las HRTF que modelizan las transmisiones acústicas de un par predeterminado de canales se retarden unas en relación con otras, o modificar de manera diferente – en un sentido variable espectralmente – las respuestas de fase y/o magnitud de las mismas, siendo el par de canales uno de un canal izquierdo y un derecho de la pluralidad de canales, un canal delantero y uno trasero de la pluralidad de canales, y un canal central y uno no central de la pluralidad de canales. El proveedor de HRTF 32 puede estar configurado para proporcionar la pluralidad original de HRTF basándose en las posiciones de fuentes de sonido virtuales y los parámetros de HRTF. El procesador de HRTF 34 puede estar configurado para filtrar con filtro de todo paso de manera diferente las respuestas impulsionales del par de canales predeterminado.

- 10 **[0084]** Aún más, las realizaciones anteriores también describían, entre otras cosas, un dispositivo para generar una contribución relacionada con la reflexión/reverberación de un espacio de una señal binaural basándose en una señal multicanal que representa una pluralidad de canales y que está pensada para reproducción por una configuración de altavoces que tiene una posición de fuente de sonido virtual asociada a cada canal, que comprende: un generador de mezcla descendente que forma una mezcla descendente mono o estéreo de los
- 15 canales de la señal multicanal; y un procesador de espacio para generar la contribución relacionada con las reflexiones/la reverberación de un espacio basándose en la señal mono o estéreo, en el que el generador de mezcla descendente está configurado para formar la mezcla descendente mono o estéreo de manera que la pluralidad de canales contribuyen a la mezcla descendente mono o estéreo a un nivel que difiere entre al menos dos canales de la señal multicanal. El dispositivo puede comprender además un detector de tipo de señal para detectar las fases de
- 20 voz y no de voz dentro de la señal multicanal, en el que el generador de mezcla descendente está configurado para realizar la formación de manera que una cantidad de reducción de nivel es más alta durante las fases de voz que durante las fases no de voz.

- [0085]** Por consiguiente, las realizaciones anteriormente descritas también describían procedimientos respectivos
- 25 para generar una señal binaural basándose en una señal multicanal que representa una pluralidad de canales y pensada para reproducción por una configuración de altavoces que tiene una posición de fuente de sonido virtual asociada a cada canal, y un procedimiento respectivo para formar un conjunto decreciente de intersimilitud de funciones de transferencia relacionadas con la cabeza para modelizar una transmisión acústica de una pluralidad de canales desde una posición de fuente de sonido virtual asociada con el canal respectivo hasta los conductos
- 30 auditivos de un oyente.

REIVINDICACIONES

1. Dispositivo para generar una contribución relacionada con la reflexión/reverberación de un espacio (46a, 46b) de una señal binaural basándose en una señal multicanal (18) que representa una pluralidad de canales y
5 que está pensada para reproducción por una configuración de altavoces que tiene una posición de fuente de sonido virtual asociada a cada canal, que comprende:

un generador de mezcla descendente (42) que forma una mezcla descendente mono o estéreo (48) de los canales de la señal multicanal; y un
10 procesador de espacio (44) para generar la contribución relacionada con las reflexiones/la reverberación de un espacio (46a, 46b) de la señal binaural modelizando las reflexiones/reverberaciones de un espacio basándose en la señal mono o estéreo,

15 en el que el generador de mezcla descendente (42) está configurado para formar la mezcla descendente mono o estéreo (48) de manera que la pluralidad de canales contribuyen a la mezcla descendente mono o estéreo a un nivel que difiere entre al menos dos canales de la señal multicanal (18),

en el que el generador de mezcla descendente (42) está configurado para reconstruir, mediante codificación de
20 audio espacial, la pluralidad de canales a partir de una señal de mezcla descendente (18) y de parámetros espaciales asociados (64) que describen diferencias de nivel, diferencias de fase, diferencias de tiempo y/o medidas de correlación entre las pluralidades de canales, y

en el que el generador de mezcla descendente (42) está configurado para realizar la formación de la mezcla
25 descendente mono o estéreo (48) de manera que una cantidad de reducción de nivel de un primero de los al menos dos canales en relación con un segundo de los al menos dos canales depende de los parámetros espaciales (64).

2. Dispositivo según la reivindicación 1, en el que el generador de mezcla descendente (42) está configurado para formar la mezcla descendente mono o estéreo de manera que un canal central de la pluralidad de
30 canales contribuye a la mezcla descendente mono o estéreo de una manera de nivel reducido en relación con los otros canales de la señal multicanal.

3. Dispositivo según la reivindicación 1, en el que el generador de mezcla descendente (42) está configurado para reconstruir, mediante codificación de audio espacial, la pluralidad de canales a partir de una señal
35 de mezcla descendente estéreo, describiendo los coeficientes de predicción de canal cómo los canales de la señal de mezcla descendente estéreo han de ser combinados linealmente para predecir un triplete de canales central, derecho e izquierdo, y una señal residual (270) que refleja un residuo de predicción cuando se predice el triplete.

4. Dispositivo según cualquiera de las reivindicaciones 1 a 3, en el que el generador de mezcla descendente (42) está configurado para realizar la formación de manera que una cantidad de reducción de nivel de
40 un primero de los al menos dos canales en relación con un segundo de los al menos dos canales depende de una diferencia de nivel y/o una correlación entre canales individuales de la pluralidad de canales.

5. Dispositivo según la reivindicación 4, en el que el generador de mezcla descendente (42) está configurado para adquirir la diferencia de nivel y/o la correlación entre canales individuales de la pluralidad de
45 canales basándose en parámetros espaciales que acompañan a una señal de mezcla descendente que representan juntos la pluralidad de canales.

6. Dispositivo según cualquiera de las reivindicaciones 1 a 3, en el que el generador de mezcla descendente (42) está configurado para realizar la formación de manera que una cantidad de reducción de nivel de
50 un primero de los al menos dos canales en relación con un segundo de los al menos dos canales varía con el tiempo tal como se indica por un indicador variable con el tiempo transmitido dentro de información secundaria de la señal multicanal.

55 7. Procedimiento para generar una contribución relacionada con la reflexión/reverberación de un espacio de una señal binaural basándose en una señal multicanal que representa una pluralidad de canales y que está pensada para reproducción por una configuración de altavoces que tiene una posición de fuente de sonido virtual asociada a cada canal, que comprende:

formar una mezcla descendente mono o estéreo de los canales de la señal multicanal; y

generar la contribución relacionada con las reflexiones/la reverberación de un espacio de la señal binaural modelizando las reflexiones/reverberaciones de un espacio basándose en la señal mono o estéreo,

5

en el que la mezcla descendente mono o estéreo se forma de manera que la pluralidad de canales contribuyen a la mezcla descendente mono o estéreo a un nivel que difiere entre al menos dos canales de la señal multicanal,

10 en el que el procedimiento comprende además reconstruir, mediante codificación de audio espacial, la pluralidad de canales a partir de una señal de mezcla descendente y de parámetros espaciales asociados que describen diferencias de nivel, diferencias de fase, diferencias de tiempo y/o medidas de correlación entre las pluralidades de canales, y

15 la formación de la mezcla descendente mono o estéreo de manera que una cantidad de reducción de nivel de un primero de los al menos dos canales en relación con un segundo de los al menos dos canales depende de los parámetros espaciales.

8. Programa informático que tiene instrucciones para realizar, cuando se ejecuta en un ordenador, un procedimiento según la reivindicación 7.

20

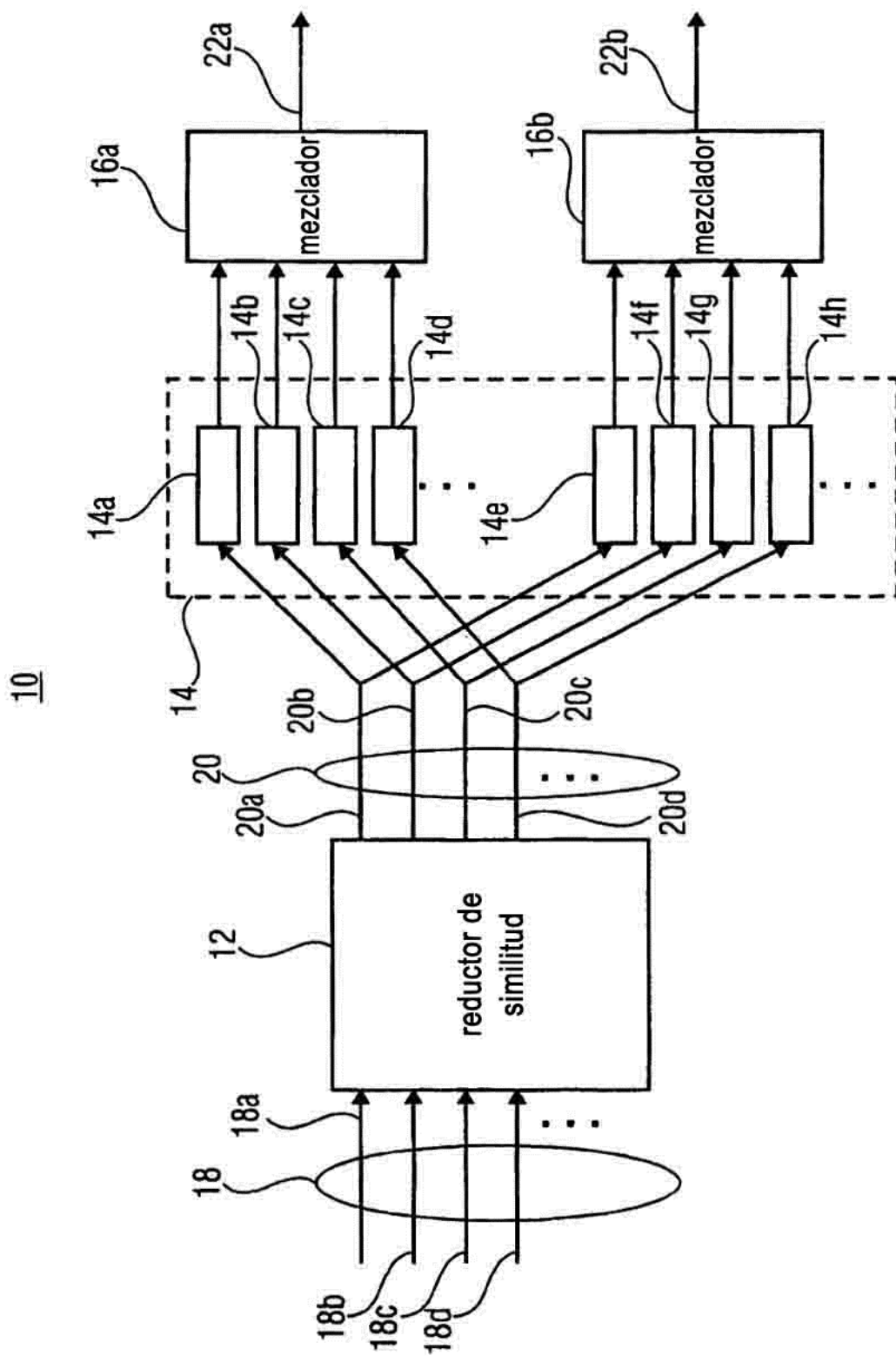


FIG 1

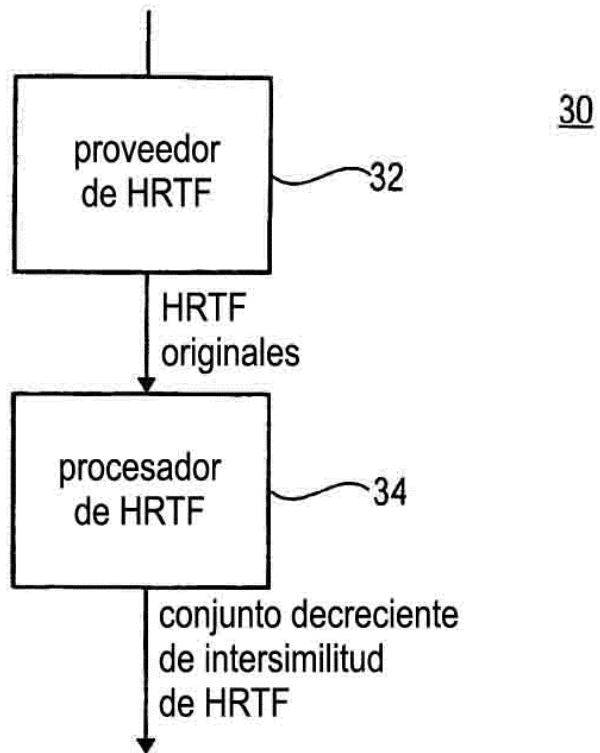


FIG 2

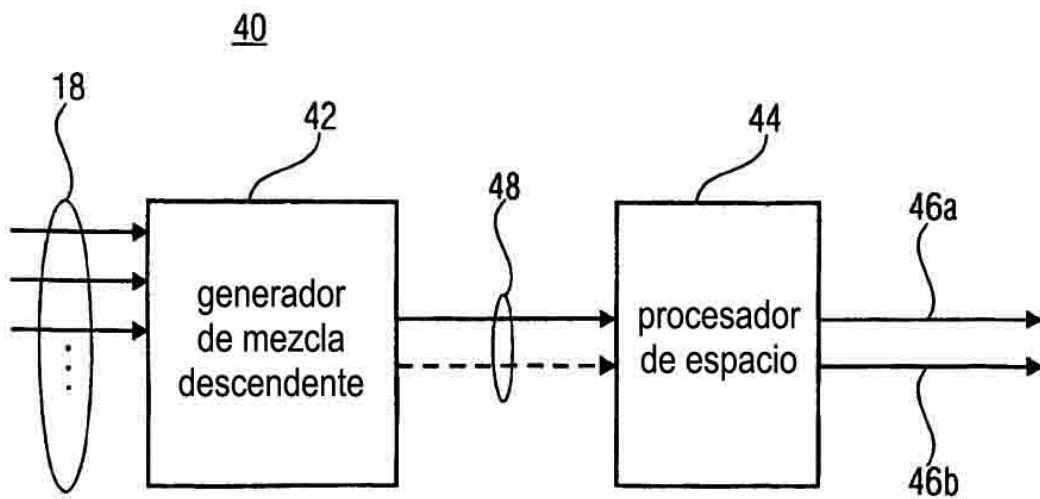


FIG 3

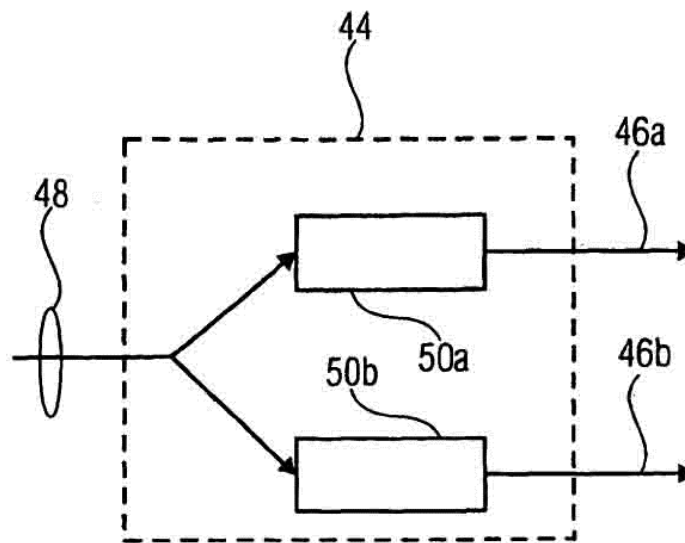


FIG 4A

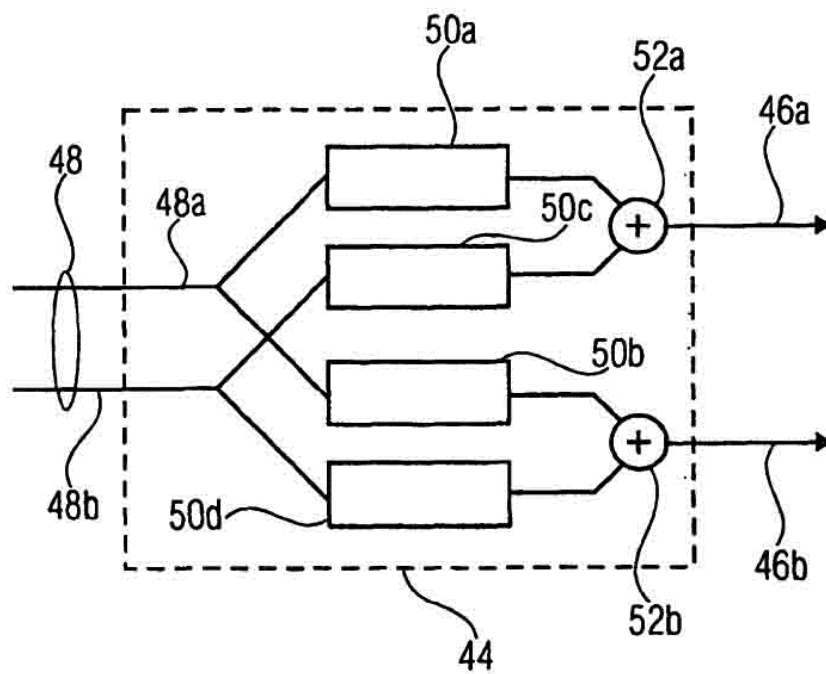


FIG 4B

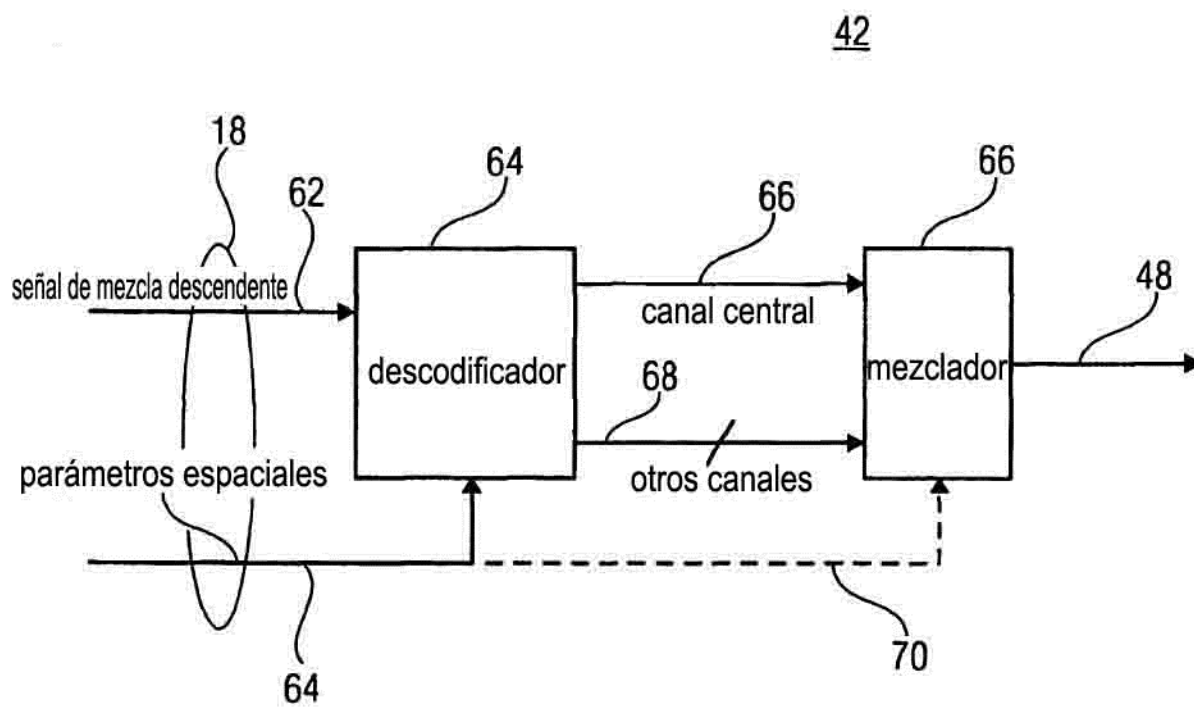


FIG 5

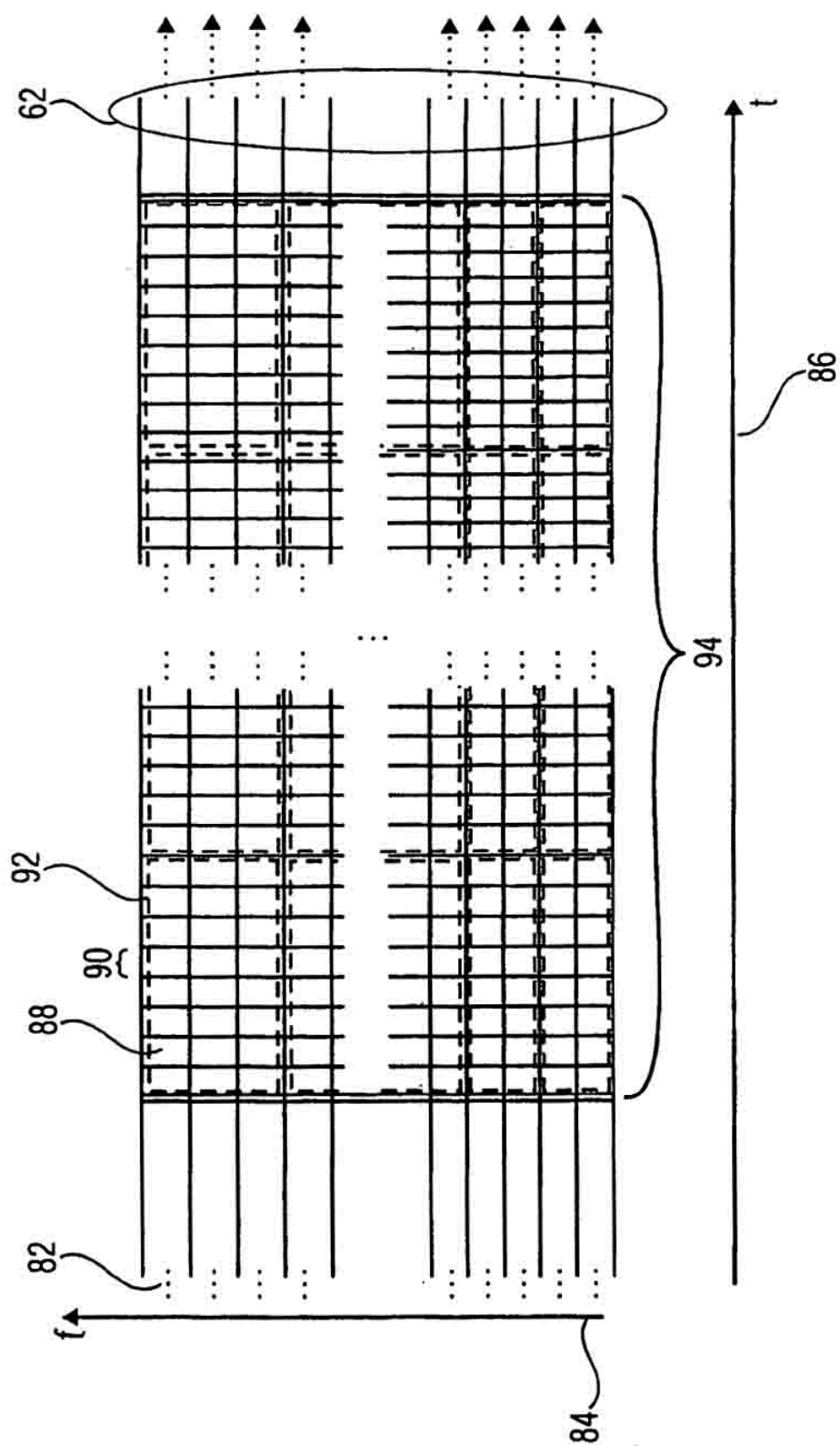


FIG 6

FIG 7

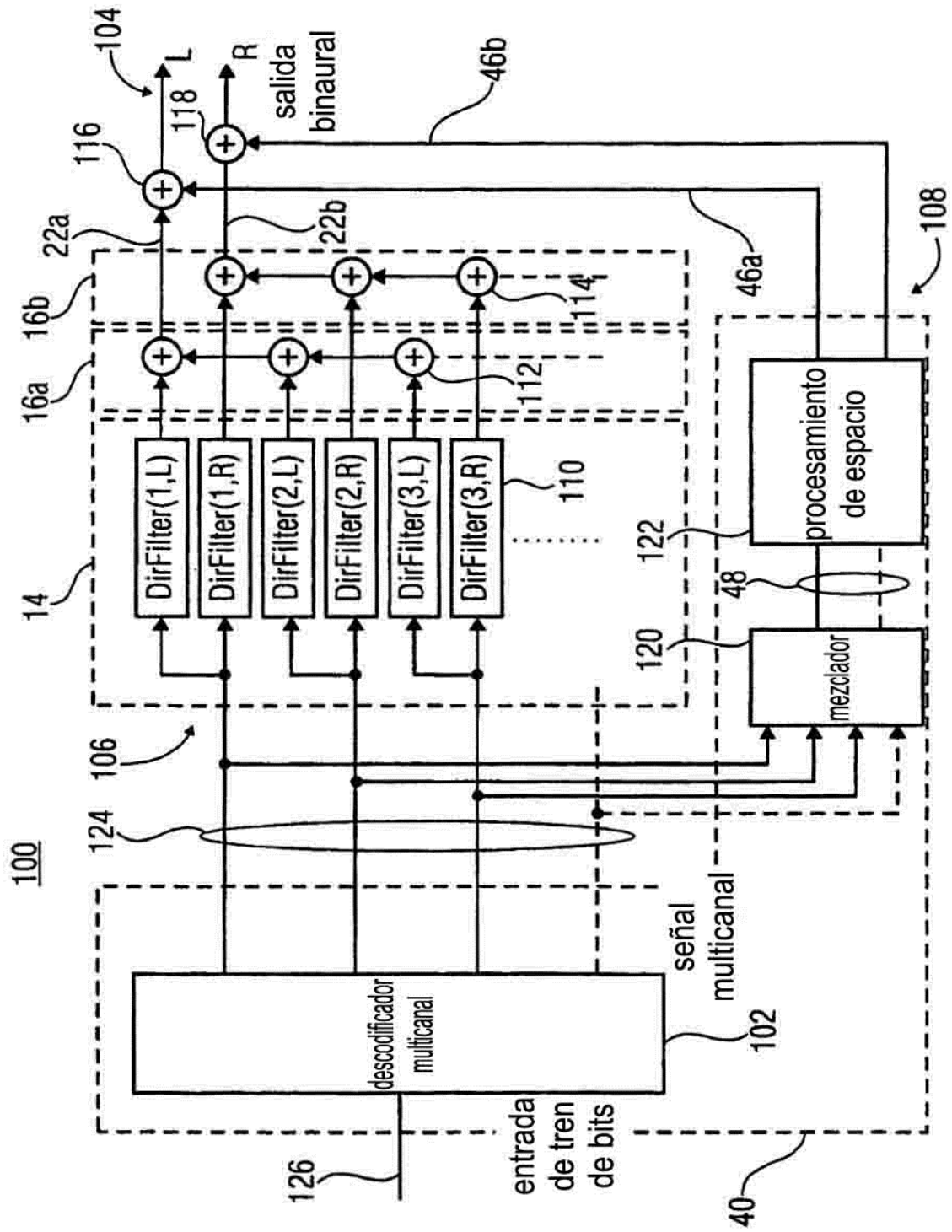


FIG 8

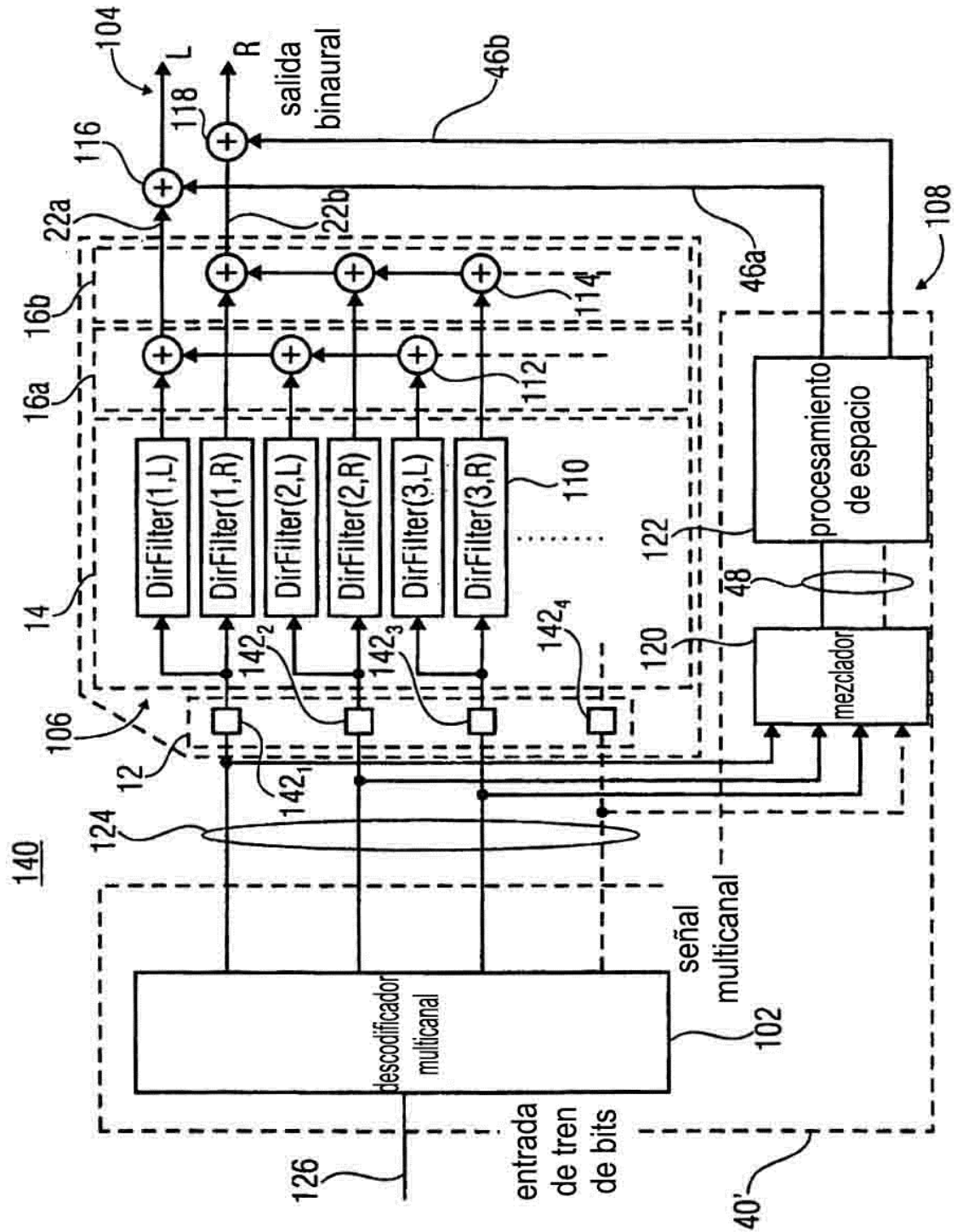
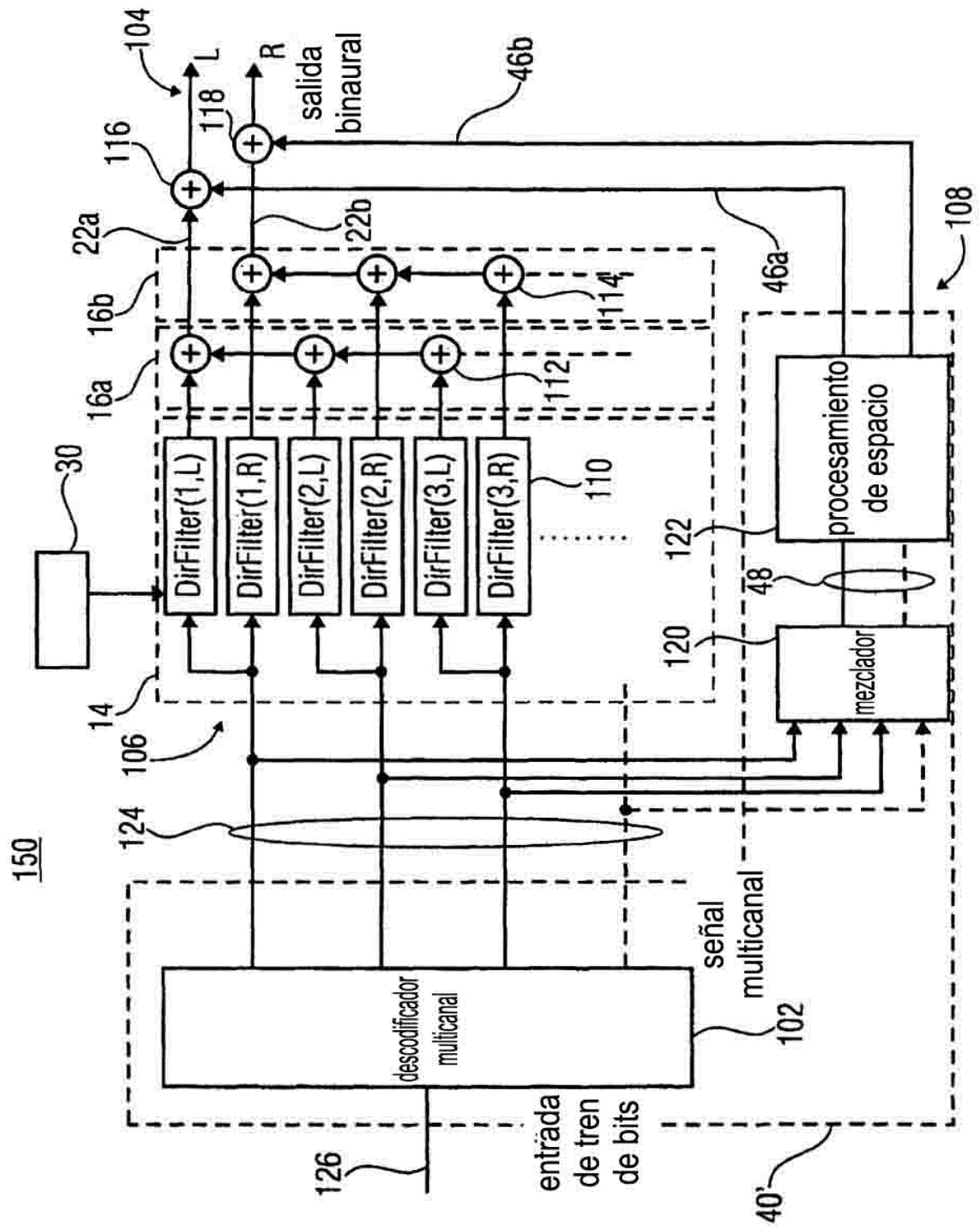


FIG 9



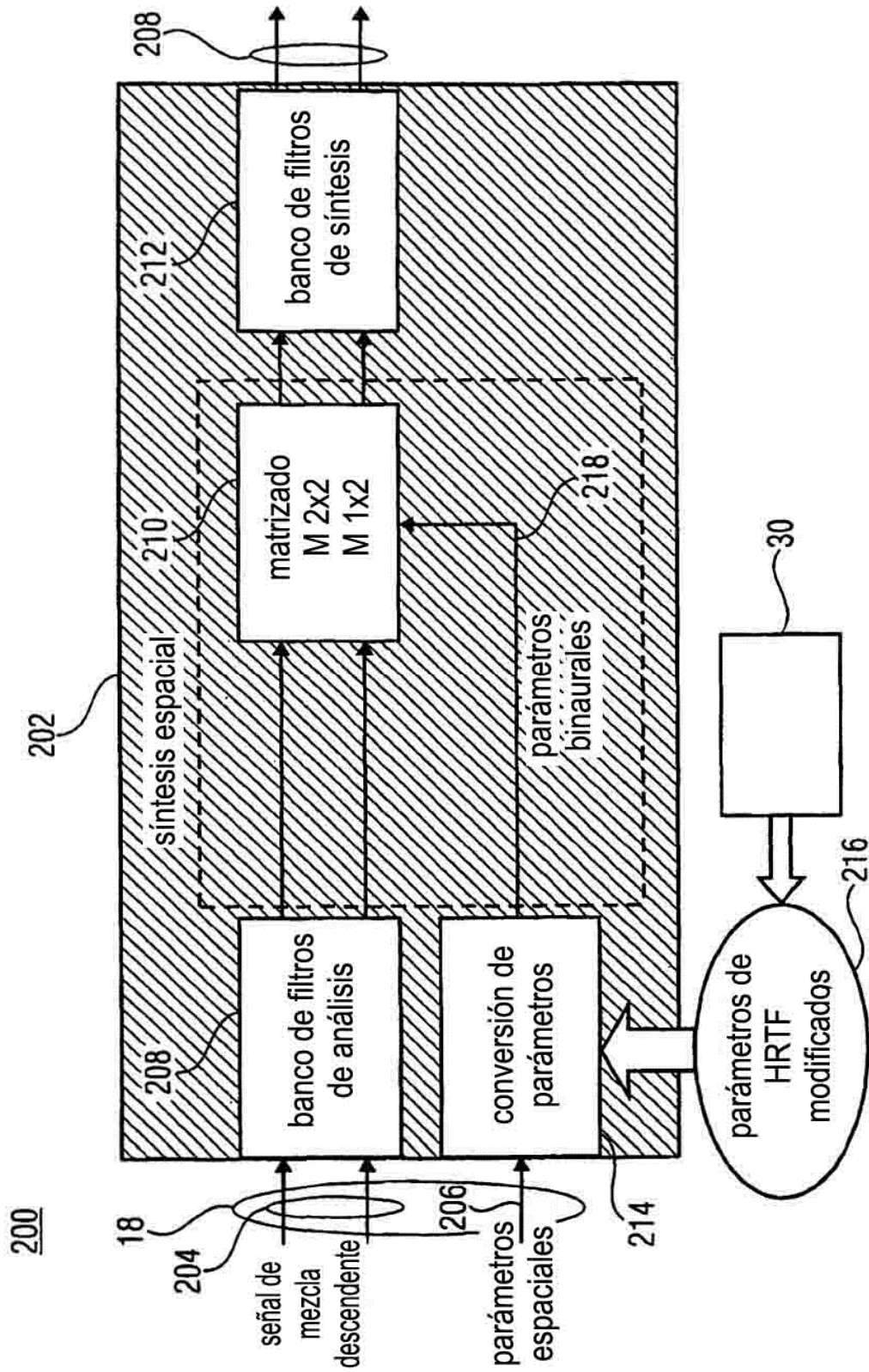


FIG 10

FIG 11

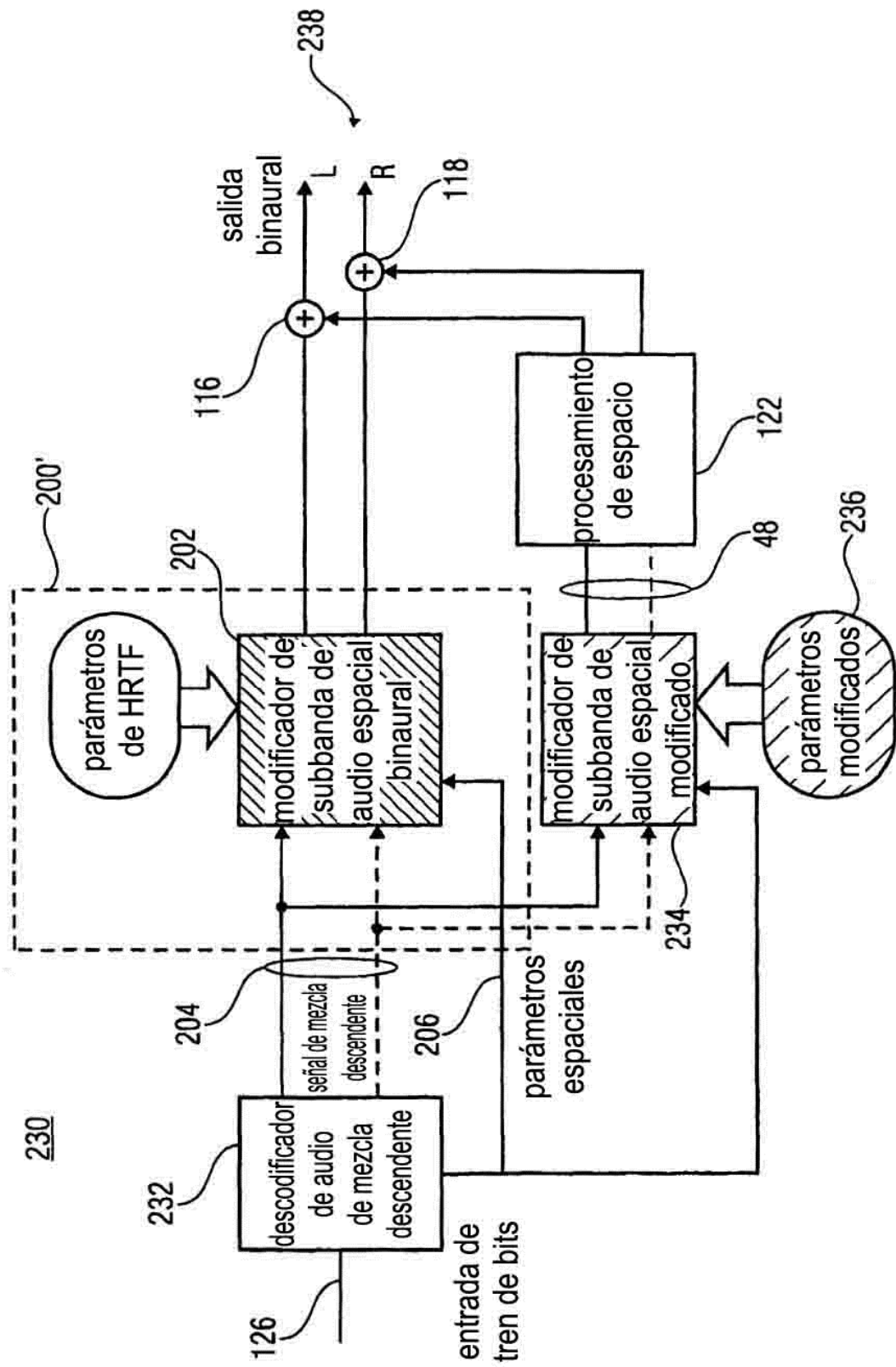


FIG 12

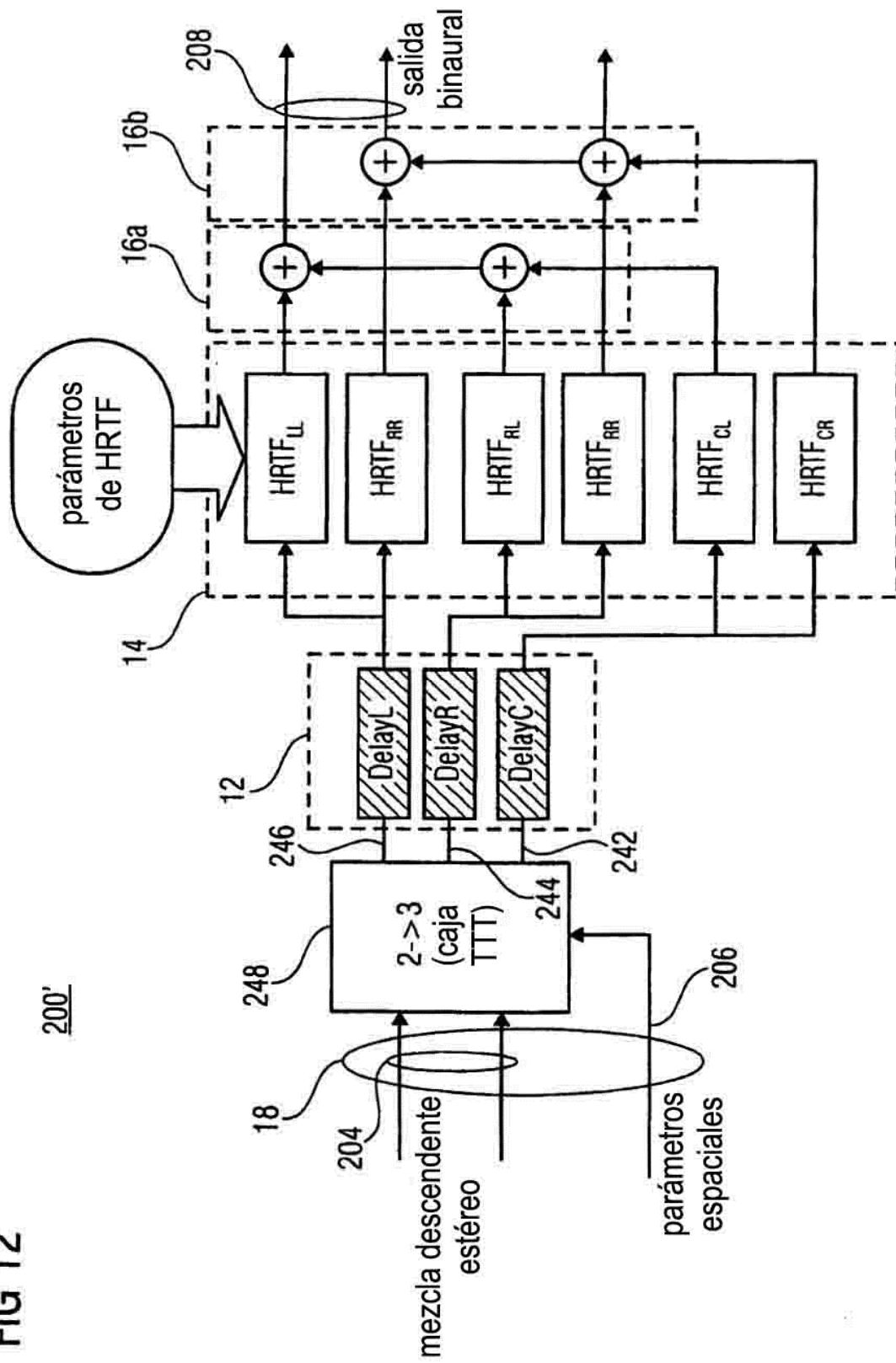


FIG 13

