



①9



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

①1 Número de publicación: **2 334 856**

⑤1 Int. Cl.:  
**H04S 1/00** (2006.01)

①2

TRADUCCIÓN DE PATENTE EUROPEA

T3

⑨6 Número de solicitud europea: **07803885 .8**

⑨6 Fecha de presentación : **19.06.2007**

⑨7 Número de publicación de la solicitud: **2042001**

⑨7 Fecha de publicación de la solicitud: **01.04.2009**

⑤4 Título: **Espacialización binaural de datos sonoros codificados en compresión.**

③0 Prioridad: **07.07.2006 FR 06 06212**

④5 Fecha de publicación de la mención BOPI:  
**16.03.2010**

④5 Fecha de la publicación del folleto de la patente:  
**16.03.2010**

⑦3 Titular/es: **FRANCE TELECOM**  
**6 place d'Alleray**  
**75015 Paris, FR**

⑦2 Inventor/es: **Virette, David y**  
**Guerin, Alexandre**

⑦4 Agente: **Justo Bailey, Mario de**

ES 2 334 856 T3

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

## DESCRIPCIÓN

Espacialización binaural de datos sonoros codificados en compresión.

La invención se refiere al tratamiento de datos sonoros, con vistas a una restitución espacializada.

La espacialización sonora tridimensional (denominada “efecto de 3D”) de señales de audio comprimidas interviene especialmente durante la descompresión de una señal de audio 3D, por ejemplo, codificada en compresión y representada sobre un cierto número de canales, hacia un número de canales diferentes (dos, por ejemplo, para permitir la restitución de los efectos de audio 3D en unos cascos de escucha).

El término “binaural” se refiere a la restitución sobre unos cascos estereofónicos de una señal sonora con, no obstante, efectos de espacialización. La invención no se limita, sin embargo, a la técnica mencionada anteriormente sino que se aplica, especialmente, a técnicas derivadas de la “binaural”, tales como las técnicas de restitución denominadas TRANSAURAL (marca registrada), es decir, en altavoces distantes. Por tanto, tales técnicas pueden utilizar una “anulación de diafonía” (o “cross-talk cancellation”, en inglés), que consiste en anular los caminos acústicos cruzados, de manera que un sonido así tratado y emitido después por los altavoces, sólo puede percibirse por uno sólo de los dos oídos de un oyente. En lo sucesivo, se designará a estas dos técnicas de restitución, binaural y transaural, de manera conjunta con los mismos términos “restitución binaural”.

Así, más en general, la invención se refiere a la transmisión de señales de audio multicanal y a su conversión para una restitución espacializada (con efecto de 3D) sobre dos vías. El dispositivo de restitución (simples cascos con auriculares, por ejemplo) viene la mayoría de las veces impuesto por el equipo de un usuario. La conversión puede ir dirigida, por ejemplo, al caso de una restitución de una escena sonora inicialmente en el formato multicanal 5.1 (o 7.1, u otro) mediante unos sencillos cascos de escucha de audio (en técnica binaural).

Evidentemente, la invención se refiere también a la restitución, en el marco de un juego o de una grabación de vídeo, por ejemplo, de una o varias muestras sonoras almacenadas en ficheros, con vistas a su espacialización.

Respecto a la técnica anterior, se hace referencia al documento US2005/047618, que muestra un procedimiento de tratamiento de datos sonoros para una restitución espacializada en tres dimensiones sobre dos vías de restitución para el oído izquierdo y derecho de un oyente utilizando una función de transferencia.

Entre las técnicas conocidas en el campo de la espacialización sonora binaural, se han propuesto diferentes enfoques.

En particular, la síntesis binaural bicanal consiste, en referencia a la figura 1, relativa a la técnica anterior, en:

- asociar a cada fuente sonora  $S_i$  (o a cada canal de la señal multicanal) una posición en el espacio,
- filtrar estas fuentes en el dominio de frecuencias mediante las funciones acústicas de transferencia izquierda HRTF-l y derecha HRTF-r correspondientes a la dirección (o la posición) elegida, y definidas por sus coordenadas polares  $(\theta_i, \varphi_i)$ .

Estas funciones de transferencia, denominadas conjuntamente funciones “HRTF” (por “Head Related Transfer Functions”, en inglés), representan las funciones de transferencia acústica entre las posiciones del espacio y el conducto auditivo de cada oído del oyente. Se designa mediante “HRIR” (por “Head Related Impulse Response”, en inglés) su forma temporal o respuesta impulsional. Estas funciones HRIR pueden incluir además un efecto de sala.

Se obtienen, para cada fuente sonora  $S_i$ , dos señales (izquierda y derecha) que se añaden entonces a las señales de izquierda y de derecha resultado de la espacialización de todas las demás fuentes sonoras, para dar finalmente las señales L y R que se difundirán en los oídos izquierdo y derecho del oyente a través de dos altavoces respectivos (auriculares de unos cascos en técnica binaural o altavoces distantes en técnica transaural).

Si N designa el número de fuentes sonoras o de flujo de audio incidentes que van a espacializarse, el número de filtros, o funciones de transferencia, necesarios para la síntesis binaural es  $2 \times N$  para un efecto en espacialización binarual estática y  $4 \times N$  para un efecto en espacialización binaural dinámica (con transiciones de las funciones de transferencia).

El tratamiento descrito anteriormente en referencia a la figura 1 y que pone en práctica las funciones de transferencia HRTF es clásico. Se utiliza a menudo para un efecto de 3D a partir de dos altavoces. Podrá ser la base para una realización puesta en práctica por la presente invención, tal como se verá más adelante. Por este motivo se introduce aquí.

No obstante, la invención parte de otro tipo de técnica anterior.

Existen técnicas de compresión, a menudo en un dominio transformado, de señales en un formato multicanal para poder vehicular estas señales, especialmente a través de redes de telecomunicación, sobre un número restringido de canales, por ejemplo, uno o dos canales solamente. Así, para la emisión de una señal en un formato multicanal que comprende más de dos canales (por ejemplo 5.1, 7.1 u otro), un codificador comprime la señal multicanal sobre únicamente uno o dos canales (normalmente según la capacidad ofrecida sobre la red de telecomunicación) y suministra además información de espacialización. Esta realización se ilustra en la figura 2A en la que, a modo de ejemplo, para una señal en un formato multicanal 5.1, se codifican cinco canales (C para un altavoz central, FL para un altavoz delantero izquierdo, FR para un altavoz delantero derecho, BL para un altavoz trasero izquierdo y BR para un altavoz trasero derecho) en compresión por un módulo COD adecuado para suministrar dos canales comprimidos L Y R, así como información de espacialización ESPAC. Los canales comprimidos L y R, así como la información de espacialización ESPAC se vehiculan a continuación a través de una o varias redes de telecomunicación RED, sobre uno o dos canales según la capacidad ofrecida (figura 2B).

En referencia a la figura 2C, en la recepción de la señal comprimida sobre los dos canales L y R, un decodificador (DECOD) reconstituye la señal original en el formato multicanal inicial gracias a la información de espacialización ESPAC suministrada por el codificador y, en el ejemplo de las figuras 2A y 2C, se recuperan incluso cinco canales, tras la decodificación, que alimentan cinco altavoces (AV-FL, AV-FR, AV-C, AV-BL et AV-BR) para una restitución al formato 5.1.

Numerosos tipos de codificadores/decodificadores paramétricos, especialmente normalizados, ofrecen tales posibilidades.

Los codificadores de audio (AAC, MP3) utilizan representaciones tiempo-frecuencia de las señales para comprimir la información. Estas representaciones se basan en un análisis mediante bancos de filtros o mediante transformación en tiempo-frecuencia de tipo MDCT (por "Modified Discrete Cosine Transform"). En el caso en el que una espacialización binaural deba efectuarse tras una decodificación de audio, las operaciones de filtrado se realizan ventajosamente de una vez en el dominio transformado.

Algunos trabajos recientes sobre el filtrado en el dominio transformado de subbandas han permitido formalizar la arquitectura de filtrado para un banco de filtros utilizado normalmente en los codificadores de audio. Se podrá consultar de manera útil el documento:

"A Generic Framework for Filtering in Subband Domain", A. Benjelloun Touimi, Proceeding IEEE - 9th Workshop on Digital Signal Processing, Hunt, Texas, EE. UU., octubre de 2000.

Una técnica más reciente de filtrado en el dominio transformado de los QMF complejos (por "Quadrature Mirror Filters") se ha propuesto en la norma "MPEG Surround". Esta técnica va dirigida a la conversión de la respuesta impulsional (terminada) del filtro temporal denominado  $h(v)$  en un conjunto de M filtros complejos denominados  $h_m(l)$ , donde M es el número de subbandas de frecuencias. La conversión se realiza mediante análisis del filtro temporal  $h(v)$  mediante un banco de filtros complejos similar al banco de filtros QMF utilizado para el análisis de la señal. En un ejemplo de realización, el filtro prototipo  $q(v)$  utilizado para generar el banco de filtro de conversión puede tener una longitud de 192. Se define una extensión con ceros del filtro temporal mediante la fórmula siguiente:

$$\tilde{h}(v) = \begin{cases} h(v), & v = 0, 1, \dots, N_h - 1; \\ 0, & \text{si no,} \end{cases},$$

donde:

- $N_h$  es la longitud del filtro en el dominio del tiempo,
- $L_q/K_h + 2$ , con  $k_h = [N_h/64]$ , la longitud de los filtros en subbandas (para 64 subbandas).

La conversión por tanto viene dada por la fórmula siguiente:

$$h_m(l) = \sum_{v=0}^{191} \tilde{h}(v + 64(l - 2)) q(v) \exp\left(-j \frac{\pi}{64} \left(m + \frac{1}{2}\right) (v - 95)\right),$$

con:

- $m = 0, 1, \dots, 63$ , correspondiente al índice de la subbanda
- $l = 0, 1, \dots, K_h + 1$ , correspondiente al índice temporal en el dominio diezmado de las subbandas.

## ES 2 334 856 T3

De manera más genérica, se comprenderá que un tratamiento de este tipo, directamente en el dominio transformado, permite pasar de una representación de la señal comprimida sobre dos canales L, R a una representación de la señal sobre dos vías de restitución L-BIN, R-BIN (figura 3) con un ensanchamiento binaural o transaural. Para ello, se proporciona una transcodificación (módulo DECOD BIN de la figura 3) que se basa en un enfoque que consiste en reconstituir, a partir de las señales comprimidas L, R y de información de espacialización ESPAC, las funciones de transferencia, de tipo HRTF, entre un oído de un oyente y cada altavoz (virtual) que se habría alimentado por un canal dado del formato multicanal inicial.

Así, en referencia ahora a la figura 4 que ilustra una restitución “virtual” al formato 5.1, por tanto, a partir de cinco altavoces, la transcodificación puesta en práctica por el módulo DECOD BIN de la figura 3 debe considerar diez funciones de transferencia:

- una para el camino A entre el altavoz delantero izquierdo AV-FL y el oído izquierdo OL del oyente OY,
- una para el camino B entre el altavoz delantero izquierdo AV-FL y el oído derecho OR del oyente OY,
- una para el camino C entre el altavoz trasero izquierdo AV-BL y el oído izquierdo OL del oyente OY,
- una para el camino D entre el altavoz trasero izquierdo AV-BL y el oído derecho OR del oyente OY,
- una para el camino G entre el altavoz delantero derecho AV-FR y el oído derecho OR del oyente OY,
- una para el camino H entre el altavoz delantero derecho AV-FR y el oído izquierdo OL del oyente OY,
- una para el camino F entre el altavoz trasero derecho AV-BR y el oído derecho OR del oyente OY,
- una para el camino E entre el altavoz trasero derecho AV-BR y el oído izquierdo OL del oyente OY,
- una para el camino J entre el altavoz del centro AV-C y el oído izquierdo OL del oyente OY, y
- una para el camino I entre el altavoz del centro AV-C y el oído derecho OR.

Así, los filtros en subbandas en el dominio transformado se calculan para cada oído y para cada una de las cinco posiciones de los altavoces. Esta técnica normalmente se denomina “técnica de los altavoces virtuales”.

Con ayuda de la representación en subbandas de los filtros binaurales determinados tal como se ha descrito anteriormente a partir de las funciones de transferencia HRTF, la espacialización binaural puede por tanto realizarse ventajosamente aplicando estos filtros binaurales directamente en el dominio transformado en el núcleo del decodificador de audio DECOD BIN tal como se representa en la figura 3.

Así, este tipo de decodificador DECOD BIN utiliza una representación monofónica o estereofónica (vías comprimidas L, R) de la escena de audio multicanal, representación a la que están asociados parámetros de espacialización ESPAC (que pueden consistir, por ejemplo, en diferencias de energías entre canales e índices de correlación entre canales). Estos parámetros ESPAC se utilizan en la decodificación para reproducir lo mejor posible la escena sonora multicanal original.

Además, cuando la señal original se codifica mediante un codificador paramétrico (por ejemplo, en el sentido de trabajos recientes en la norma “MPEG Surround”), además de la señal monofónica o estereofónica transmitidas y de la información de espacialización, la decodificación puede utilizar representaciones decorrelacionadas de estas señales L, R (que se obtienen, por ejemplo, mediante la aplicación de filtros de decorrelación pasa todo o de filtros de reverberación). Estas señales se ajustan a continuación en energía gracias a las diferencias de energía entre canales, después se recombinan para obtener la señal multicanal con vistas a la restitución.

En particular, el codificador paramétrico (COD - figura 2A) del formato multicanal hacia dos vías comprimidas (estéreo o mono) según el proyecto de la norma “MPEG Surround” suministra una información de decorrelación entre canales en el formato multicanal inicial y esta información de decorrelación puede retomarse por el decodificador paramétrico homólogo (DECOD-figura 2C) durante la restitución al formato multicanal inicial.

Una descripción de los trabajos preparatorios de esta norma se facilita en la dirección URL:

“<http://www.chiariglione.org/mpeg/technologies/mpd-mps/index.htm>”

y precisiones en cuanto a un codificador de este tipo según este proyecto de norma pueden encontrarse en:

“MPEG Spatial Audio Coding / MPEG Surround: Overview and Current Status”, J. Breebaart *et al.*, en 119th Conv. Aud. Eng. Soc (AES), Nueva York, NY, EE. UU, octubre 2005.

En el caso de un decodificador audio paramétrico para una posible restitución binaural (DECOD BIN - figura 3), es ventajosamente posible simplificar las operaciones de filtrado combinando los filtros delantero y trasero correspondientes a los diferentes altavoces izquierdos (llevando también un tratamiento equivalente los altavoces derechos). Esta combinación se realiza en función de las energías objetivo de los canales de audio, dadas por los parámetros de espacialización. Esta combinación, para el oído izquierdo y los canales delantero izquierdo y trasero izquierdo, se realiza en el dominio transformado según una expresión (1) del tipo:

$$h_{L,L} = g_{L,L} \sigma_{FL} \exp(-j\phi_{FL,BL}^L \sigma_{BL}^2) h_{L,FL} + g_{L,L} \sigma_{BL} \exp(j\phi_{FL,BL}^L \sigma_{FL}^2) h_{L,BL}$$

En esta expresión:

- $h_{L,L}$  es el filtro correspondiente a las contribuciones de los canales delantero y trasero izquierdos,
- $g_{L,L}$  es la ganancia asociada al conjunto de los canales izquierdos,
- $\sigma_{FL}^2$  y  $\sigma_{BL}^2$  son las energías útiles de los canales respectivamente delantero y trasero izquierdos,
- $h_{L,FL}$  y  $h_{L,BL}$  son las funciones de transferencia en el dominio de las subbandas entre el oído izquierdo y los altavoces respectivamente delantero y trasero izquierdos (camino A y C de la figura 4),
- $\Phi_{FL,BL}^L$  es el desfase correspondiente al retardo entre los filtros temporales delantero y trasero izquierdos  $h_{L,FL}$  y  $h_{L,BL}$ .

El objetivo de esta compensación de fase, en función de la energía objetivo de los canales, es evitar un efecto denominado “de coloración” resultado de la adición de dos filtros desplazados en el tiempo (filtrado en peine).

En referencia a la figura 5A, un decodificador recibe los parámetros de espacialización ESPAC que acompañan a las señales comprimidas sobre dos vías L y R en el ejemplo representado, y se ha ilustrado, en esta misma figura 5A, cómo se aplica el filtro  $h_{L,L}$  mencionado anteriormente al canal comprimido L para formar una componente de la señal L-BIN, destinada a la restitución binaural. No obstante, tal como se representa asimismo en la figura 5A, conviene tener en cuenta también la señal comprimida sobre el canal R que debe, a su vez, filtrarse mediante un filtro que hace intervenir funciones de transferencia HRTF (indicadas por  $H_{L,FR}$  et  $H_{L,BR}$ ) relativas a los caminos cruzados H y E de la figura 4, siempre hacia el oído izquierdo. El filtro correspondiente a estos caminos cruzados (indicado por  $h_{L,R}$ ) se calcula en función de las ganancias, energías objetivo y desfases, sacados de los parámetros de espacialización ESPAC, con ayuda de una expresión equivalente a la relación (1) dada anteriormente. Este filtro  $h_{L,R}$  se aplica finalmente a la señal comprimida sobre el canal R.

Conviene tener en cuenta, además, la “contribución” del altavoz central en la construcción de la señal destinada a la restitución binaural L-BIN y, para ello, se aplica un filtro  $h_{L,C}$  (figura 5A) con una combinación (por ejemplo, por adición) de las señales comprimidas de las dos vías L y R para tener en cuenta en este caso el camino J hacia el oído izquierdo OL de la figura 4.

En referencia todavía a la figura 5A, se proporciona un tratamiento equivalente para la construcción de la señal R-BIN destinada a una restitución binaural para el oído derecho OD, con tres contribuciones dadas por:

- la señal comprimida sobre el canal R filtrada por el filtro  $h_{R,R}$  que representa las funciones HRTF de los altavoces de la derecha (camino directo G y F de la figura 4);
- la señal comprimida sobre el canal L filtrada por el filtro  $h_{R,L}$  que representa las funciones HRTF de los altavoces de la izquierda (camino cruzados B y D de la figura 4); y
- una combinación de las señales comprimida L y R filtrada por el filtro  $h_{R,C}$  que representa las funciones HRTF del altavoz del centro (camino directo I de la figura 4).

En la figura 5B, se ha representado otro ejemplo, en el que un decodificador recibe la señal comprimida sobre un único canal M, que acompaña a los parámetros de espacialización ESPAC. En el ejemplo representado, se duplica el canal M en dos canales L y R y el resultado del tratamiento es estrictamente equivalente al tratamiento representado en la figura 5A.

Las dos señales L-BIN y R-BIN que resultan de estos filtrados pueden aplicarse a continuación a dos altavoces destinados respectivamente al oído izquierdo y al oído derecho del oyente tras un paso del dominio transformado al dominio del tiempo.

Sin embargo, un problema relacionado con esta combinación de filtros para una restitución binaural es que no tiene en cuenta una eventual decorrelación entre los canales delantero y trasero. Esta información, utilizada sin embargo en la decodificación de una escena 5.1 de un codificador según el proyecto mencionado anteriormente de la norma MPEG Surround, no se aprovecha en la técnica de decodificación binaural. Así, cuando la escena sonora comprende efectos de decorrelación entre los canales delantero y trasero (por ejemplo para señales reverberadas), esta información no se utiliza en la combinación de los filtros HRTF, lo que conlleva una degradación de la calidad de espacialización y especialmente del efecto envolvente de la escena de audio 3D. Por tanto, la restitución al formato binaural no es óptima.

La presente invención va a mejorar la situación.

Va dirigida, en primer lugar, a un procedimiento de tratamiento de datos sonoros para una restitución espacializada en tres dimensiones sobre dos vías de restitución para los oídos respectivos de un oyente,

estando los datos sonoros inicialmente representados en un formato multicanal, a continuación codificados en compresión sobre un número reducido de canales (por ejemplo, uno o dos canales),

consistiendo dicho formato multicanal en proporcionar más de dos canales susceptibles de alimentar altavoces respectivos, comprendiendo el procedimiento las etapas de:

- obtener, con los datos comprimidos sobre dicho número reducido de canales, parámetros de espacialización,
- para cada vía de restitución asociada a un oído del oyente, formar, a partir de dichos parámetros de espacialización, una combinación de filtros representativos cada uno de funciones de transferencia entre este oído del oyente y altavoces susceptibles de alimentarse mediante canales respectivos del formato multicanal inicial, y
- aplicar a los datos comprimidos la combinación de filtros asociada a cada vía de restitución.

El procedimiento en el sentido de la invención comprende además las etapas de:

- para cada vía de restitución asociada a un oído del oyente, determinar, a partir de dichos parámetros de espacialización, al menos una función de transferencia de un altavoz situado detrás del oído del oyente y representativa de una decorrelación entre los canales del formato multicanal respectivamente asociados al altavoz trasero y al menos un altavoz situado delante del oído del oyente, y
- para cada vía de restitución, integrar dicha función de transferencia representativa de una decorrelación en dicha combinación de filtros asociada a esta vía de restitución.

La restitución espacializada sobre dos vías, en el sentido de la invención, puede ser tanto en el formato binaural como transaural. El formato multicanal inicial puede ser de tipo ambiofónico (o "ambiosonic" en inglés y que se refiere a la descomposición de la señal sonora en una base de armónicos esféricos). Como variante, puede tratarse de un formato de tipo 5.1 o 7.1, o incluso 10.2. Se comprenderá, por tanto, que para estos últimos tipos de formato que ponen en práctica canales destinados a alimentar respectivamente al menos parejas de altavoces delantero-izquierdo/trasero-izquierdo, por una parte, y delantero-derecho/trasero-derecho, por otra parte, la información de decorrelación puede ir dirigida a los canales respectivos de los altavoces delantero/trasero preferiblemente asociados a un mismo oído (izquierdo o derecho).

Según una ventaja que proporciona la invención, debido a que esta información de decorrelación tras una escena en 3D está representada en la restitución binaural o transaural, se obtiene una mejor representación de los ambientes, por ejemplo, de los ruidos de una multitud o una reverberación tras una escena, u otros, al contrario que las realizaciones de la técnica anterior.

En el modo de realización particular, la combinación de filtros comprende una ponderación, según un coeficiente elegido entre:

- una función de transferencia bruta del altavoz situado detrás, y
- una versión de la función de transferencia de este altavoz, representativa de la decorrelación.

Esta ponderación permite ventajosamente favorecer la función de transferencia bruta de este altavoz trasero, o la versión decorrelacionada de esta función de transferencia bruta, según esté la señal en el canal trasero del formato multicanal inicial correlacionada o no con al menos una señal de uno de los canales delanteros.

## ES 2 334 856 T3

Por otro lado, en una realización particular, la combinación de filtros asociada a una vía de restitución comprende al menos un reagrupamiento que forma un filtro a partir de:

- la función de transferencia de un altavoz delantero,
- la función de transferencia de un altavoz trasero, y
- la función de transferencia representativa de una decorrelación entre canales,

y estos altavoces delantero y trasero están situados en un mismo lado con respecto al oyente. Puede tratarse, por ejemplo, de altavoces delantero y trasero situados ambos a la izquierda (o ambo a la derecha) del oyente en el formato 5.1 (como se ilustra en la figura 4). En una realización de este tipo, cuando se proporciona la ponderación entre la versión decorrelacionada y la versión bruta de las funciones de transferencia, podrá ser ventajoso dar prioridad a la versión decorrelacionada en la combinación de filtros de los altavoces de la izquierda para la vía de restitución en el oído derecho (y a la inversa) y dar prioridad a la versión bruta (no decorrelacionada) en la combinación de filtros de los altavoces de la derecha (izquierda) para la vía de restitución en el oído derecho (izquierdo).

De manera ventajosa, la codificación en compresión pone en práctica un codificador paramétrico que proporciona, en el flujo comprimido que incluye los parámetros de espacialización, una información de decorrelación entre canales del formato multicanal, a partir de la cual puede determinarse, de manera dinámica, la ponderación mencionada anteriormente.

Así, en esta realización, para una transcodificación entre un formato multicanal a un formato binaural, la combinación mencionada anteriormente de funciones de transferencia saca provecho de la información ya presente relativa a la correlación entre señales de canales en el formato multicanal, proporcionándose esta información simplemente por el codificador paramétrico, con los parámetros de espacialización mencionados anteriormente.

A modo de ejemplo, se recuerda que el decodificador paramétrico según el proyecto de la norma MPEG Surround proporciona tal información de decorrelación entre canales en el formato multicanal 5.1.

Otras ventajas y características de la invención resultarán evidentes tras la lectura de la descripción detallada dada a continuación a modo de ejemplo, y tras la observación de los dibujos adjuntos, en los que, además de las figuras 1, 2A, 2B, 2C, 3 y 4, 5A y 5B comentadas anteriormente:

- las figuras 6A y 6B ilustran a modo de ejemplo un tratamiento mediante filtrado de datos comprimidos (sobre dos canales en el ejemplo representado), estando determinado el filtrado por la puesta en práctica del procedimiento en el sentido de la invención para suministrar señales L-BIN y R-BIN destinadas a alimentar respectivamente las vías izquierda y derecha de un dispositivo de restitución binaural tal como unos cascos con dos auriculares, teniendo en cuenta una decorrelación delantera/trasera, y

- la figura 7 ilustra esquemáticamente la estructura de un módulo que pone en práctica el procedimiento en el sentido de la invención.

En referencia a la figura 6A, se recupera en primer lugar la señal comprimida, a menudo en el dominio transformado, sobre dos canales L y R en el ejemplo representado, así como los parámetros de espacialización ESPAC que ha proporcionado un codificador tal como el módulo COD de la figura 2A descrita anteriormente. A partir de los parámetros de espacialización ESPAC, se determinan funciones de transferencia para construir una combinación de filtros (signo "+" de la figura 6A), debiendo aplicarse cada filtro a un canal L (filtro  $h_{L,L}$  de la figura 5A), o R (filtro  $h_{L,R}$  de la figura 5A), o a una combinación de estos canales (filtro  $h_{L,C}$  de la figura 5A), para construir una señal que alimenta una de las dos vías de restitución binaural L-BIN. Estas funciones de transferencia, de tipo HRTF, son representativas de las perturbaciones que experimenta una onda acústica sobre un camino entre un altavoz que se habría alimentado por un canal del formato multicanal inicial y un oído del oyente. Por ejemplo, si el contenido de audio está inicialmente en el formato 5.1, tal como se describió anteriormente en referencia a la figura 4, se determinan en total diez funciones de transferencia HRTF, cinco funciones HRTF para el oído derecho (sobre los caminos B, D, G, F e I de la figura 4) y cinco funciones HRTF para el oído izquierdo (sobre los caminos A, C, H, E y J). Se indica que el altavoz del centro se trata por separado en la espacialización binaural y la obtención del filtro correspondiente  $h_{L,C}$  o hic no se describirá aquí, entendiéndose que requiere la intervención, a priori, en el ejemplo descrito, del objeto de la invención.

Así, en términos generales, las funciones HRTF de los altavoces delantero y trasero, situados en un mismo lado del oyente se reagrupan por tanto para construir cada filtro de una combinación de filtros adecuada a una vía de restitución en un oído de un oyente. Un reagrupamiento de funciones HRTF para construir un filtro es, por ejemplo, una adición, por medio de los coeficientes multiplicativos, de lo que se describirá posteriormente un ejemplo.

En el sentido de la invención, a partir de los parámetros de espacialización ESPAC recuperados, se determina además una versión decorrelacionada de las funciones HRTF de los altavoces situados detrás del oyente (caminos C, D, E y F de la figura 4) y se integra esta versión decorrelacionada en cada reagrupamiento para formar un filtro que va a aplicarse a un canal comprimido.

A modo de ejemplo meramente ilustrativo, los datos sonoros iniciales pueden estar en el formato multicanal 5.1 y, en referencia a la figura 6A, un primer reagrupamiento comprende:

- la función HRTF-A (para el altavoz delantero izquierdo según un camino directo hacia el oído izquierdo OL de la figura 4),
- la función HRTF-C (para el altavoz trasero izquierdo según un camino directo hacia el oído izquierdo),
- y la versión decorrelacionada de esta función HRTF-C, indicada por HRTF-C\*, para formar el filtro que va a aplicarse al canal comprimido L.

Un segundo reagrupamiento comprende:

- la función HRTF-H (para el altavoz delantero derecho según un camino cruzado hacia el oído izquierdo),
- la función HRTF-E (para el altavoz trasero derecho según un camino cruzado),
- y la versión decorrelacionada de esta función HRTF-E, indicada por HRTF-E\*, para formar el filtro que va a aplicarse al canal comprimido R.

La adición de las dos señales resultantes de tales filtrados será una componente de la señal que alimenta la vía de restitución binaural L-BIN asociada al oído izquierdo.

Se proporciona un tratamiento similar para construir la señal destinada a alimentar la otra vía de restitución binaural R-BIN de la figura 6B. En este caso, se tienen en cuenta funciones HRTF de los caminos que llevan al oído derecho OD del oyente OY (figura 4). Un primer reagrupamiento comprende las funciones HRTF-G (para el altavoz delantero derecho según un camino directo), HRTF-F (para el altavoz trasero derecho según un camino directo) y la versión decorrelacionada indicada por HRTF-F\* de la función HRTF-F para formar el filtro que va a aplicarse al canal comprimido R.

Un segundo reagrupamiento comprende la función HRTF-B (para el altavoz delantero izquierdo según un camino cruzado), la función HRTF-D (para el altavoz trasero izquierdo según un camino cruzado) y la versión decorrelacionada, indicada por HRTF-D\*, de la función HRTF-D, para formar el filtro que va a aplicarse al canal comprimido L.

Finalmente, las combinaciones de filtros que integran las versiones decorrelacionadas de las funciones HRTF de los altavoces traseros se aplican a los canales comprimidos L y R para suministrar las vías de restitución L-BIN y R-BIN, para una restitución binaural espacializada con efecto de 3D.

En el ejemplo representado en las figuras 6A y 6B, los datos sonoros recibidos se codifican en compresión sobre dos canales estereofónicos L y R, tal como se ilustra en el ejemplo de la figura 5A. Como variante, podrían codificarse en compresión sobre un único canal M, monofónico, tal como se ilustra en la figura 5B, en cuyo caso las combinaciones de filtros se aplican al canal monofónico (duplicado) tal como se ilustra en la figura 5B, para proporcionar todavía dos señales que alimentan respectivamente las dos vías de restitución L-BIN, R-BIN.

En una realización ventajosa, los datos sonoros iniciales están en el formato multicanal 5.1 y se codifican en compresión por un codificador paramétrico según el proyecto de la norma mencionada anteriormente MPEG Surround. Más en particular, durante una codificación de este tipo, es posible obtener, entre los parámetros de espacialización proporcionados, una información de decorrelación entre el canal trasero derecho y el canal delantero derecho (altavoces respectivos AV-BR y AV-FR de la figura 4), así como una información de decorrelación homóloga entre el canal trasero izquierdo y el canal delantero izquierdo (altavoces respectivos AV-FL y AV-BL de la figura 4).

Esta información de decorrelación, en un formato 5.1, van dirigidas a producir la restitución de los altavoces traseros lo más independiente posible de la restitución de los altavoces delanteros, para enriquecer, en el formato 5.1, el efecto envolvente mediante ruidos de reverberación o de público para grabaciones de conciertos, por ejemplo. Se recuerda que este enriquecimiento de la envolvente 3D no se ha propuesto en restitución binaural y una ventaja de la invención es la de sacar provecho de la disponibilidad de la información de decorrelación entre los parámetros de espacialización ESPAC para construir versiones decorrelacionadas de las funciones HRTF que se integran ventajosamente en las combinaciones de filtros para una restitución binaural.

Según otra ventaja, estas combinaciones de filtros pueden calcularse directamente en el dominio transformado, por ejemplo, en el dominio de las subbandas, y los filtros que representan las versiones decorrelacionadas de las funciones HRTF de los altavoces traseros pueden obtenerse, por ejemplo, aplicando a las funciones HRTF iniciales un desfase en función de la subbanda de frecuencias considerada.



## ES 2 334 856 T3

Más en general, los filtros de decorrelación pueden ser filtros de reverberación denominada “natural” (grabada en un entorno acústico particular, como una sala de conciertos, por ejemplo), o “sintética” (creada por la suma de reflexiones múltiples de amplitudes decrecientes en el tiempo). La aplicación de un filtro decorrelacionado puede, por tanto, venir a aplicar a la señal descompuesta en subbandas de frecuencias un desfase diferente en cada una de las subbandas, combinado con la adición de un retardo global. En el caso de un decodificador paramétrico del tipo mencionado anteriormente (fórmula (1) dada anteriormente en la descripción de la técnica anterior), éste viene a multiplicar cada subbanda de frecuencias por una exponencial compleja, de fase diferente en cada subbanda. Estos filtros de decorrelación pueden, por tanto, corresponder a síntesis de filtros pasa todo desfasadores.

Se proporciona ventajosamente una ponderación entre la función de transferencia de un altavoz trasero y su versión decorrelacionada en un mismo reagrupamiento que forma un filtro. Así, retomando la fórmula (1) dada anteriormente para el cálculo de un filtro, por ejemplo,  $h_{L,L}$  para el oído izquierdo, se introducen coeficientes de ponderación  $\alpha$  y  $(1-\alpha)$  y la versión decorrelacionada de una función de transferencia como sigue:

$$h_{L,L} = g_{L,L} \sigma_{FL} \exp(-j\phi_{FL,BL}^L \sigma_{BL}^2) h_{L,FL} + g_{L,L} \sigma_{BL} \exp(j\phi_{FL,BL}^L \sigma_{FL}^2) (\alpha h_{L,BL} + (1-\alpha) h_{L,BL}^{Decor})$$

con las mismas notaciones explicitadas anteriormente y donde  $h_{L,BL}^{Decor}$  representa la versión decorrelacionada de la función de transferencia del altavoz trasero izquierdo. Evidentemente, se proporciona un mismo tipo de relaciones que dan los otros filtros  $h_{L,R}$ ,  $h_{R,R}$  y  $h_{R,L}$  (figuras 5A y 5B). Por ejemplo, para el filtro  $h_{L,R}$  propio de los caminos cruzados hacia el oído izquierdo, se tiene:

$$h_{L,R} = g_{L,R} \sigma_{FR} \exp(-j\phi_{FR,BR}^L \sigma_{BR}^2) h_{L,FR} + g_{L,R} \sigma_{BR} \exp(j\phi_{FR,BR}^L \sigma_{FR}^2) (\alpha h_{L,BR} + (1-\alpha) h_{L,BR}^{Decor})$$

Más específicamente, se proporciona una ponderación mediante coeficientes diferentes,  $\alpha_1$ ,  $(1-\alpha_1)$  y  $\alpha_2$ ,  $(1-\alpha_2)$ , según esté el altavoz trasero en el mismo lado que el oído considerado ( $\alpha=\alpha_1$  da los filtros  $h_{L,L}$  y  $h_{R,R}$ ), o no ( $\alpha=\alpha_2$  da los filtros  $h_{L,R}$  y  $h_{R,L}$ ). Preferiblemente, se da prioridad a la versión decorrelacionada para caminos cruzados (altavoz trasero derecho para el oído izquierdo y altavoz trasero izquierdo para el oído derecho), de modo que en general, el coeficiente  $\alpha_1$ , podrá ser a menudo superior al coeficiente  $\alpha_2$ .

En la práctica, los coeficientes  $\alpha$  ( $\alpha_1$  o  $\alpha_2$ ) vienen dados por funciones de ponderación variables con objeto de favorecer dinámicamente la versión bruta de la función HRTF del altavoz trasero o su versión decorrelacionada según esté la señal trasera correlacionada o no con la señal delantera. Se obtiene así una mejor representación de los ambientes (ruidos de multitud, reverberación u otros) en el efecto de 3D. La función de ponderación  $\alpha$  puede definirse dinámicamente gracias a la información de decorrelación proporcionada con los parámetros de espacialización, a modo de ejemplo no limitativo, de la manera siguiente:

$$\alpha = \text{raíz cuad}(\text{abs}(\text{ICC}_L)), \text{ si } \text{abs}(\text{ICC}_L) > \sigma_{BL}^2$$

$$\alpha = \text{raíz cuad}(\sigma_{BL}^2), \text{ si no,}$$

donde la notación “raíz cuad” es la función “raíz cuadrada”, la notación “abs” se refiere a la función “valor absoluto” y el término  $\text{ICC}_L$  representa la información de decorrelación (denominada de otro modo “índice de correlación”) entre el canal delantero y el canal trasero del mismo lado izquierdo y forma parte de los parámetros de espacialización transmitidos por el codificador según el proyecto de la norma MPEG Surround indicada anteriormente. Como se describió anteriormente, el término  $\sigma_{BL}$  representa la energía objetivo del canal trasero izquierdo cuando se trata de determinar el coeficiente  $\alpha$  para calcular el filtro  $h_{L,L}$  ( $\alpha=\alpha_1$ ). Naturalmente, una expresión equivalente puede aplicarse para calcular el coeficiente de ponderación  $\alpha$  que interviene en el filtro homólogo  $h_{R,R}$  propio de los caminos acústicos directos hacia el oído derecho. No obstante, para los filtros  $h_{L,R}$  y  $h_{R,L}$  propios de los caminos cruzados, por ejemplo, para el filtro  $h_{L,R}$  propio de los caminos cruzados hacia el oído izquierdo, el coeficiente  $\alpha=\alpha_2$  puede preferiblemente escribirse como:

$$\alpha_2 = \text{abs}(\text{ICC}_R), \text{ si } \text{abs}(\text{ICC}_R) > \sigma_{BR}^2$$

$$\alpha_2 = \sigma_{BR}^2, \text{ si no,}$$

representando el término  $\sigma_{BR}$  la energía objetivo del canal trasero derecho y representando el término  $\text{ICC}_R$  el índice de correlación entre el canal delantero derecho y el canal trasero derecho. Debe observarse que la función “raíz cuad” ya no se aplica para los caminos cruzados y para el cálculo del coeficiente correspondiente  $\alpha_2$ , en el ejemplo descrito. En efecto, las energías objetivo y los índices de correlación son términos comprendidos entre 0 y 1 de manera que el coeficiente  $\alpha_2$  es generalmente inferior al coeficiente  $\alpha_1$ .

La combinación de filtros global, para la vía L-BIN, comprende muchos reagrupamientos de funciones HRTF que forman filtros  $h_{L,L}$  y  $h_{L,R}$  obtenidos por las fórmulas dadas anteriormente y, en cada reagrupamiento, intervienen la

## ES 2 334 856 T3

función HRTF de un altavoz delantero, la función HRTF de un altavoz trasero y una versión decorrelacionada de esta última función HRTF, lo que permite representar una decorrelación entre los canales delantero y trasero directamente en la combinación de filtros, y por tanto, directamente en la síntesis binaural.

Se recuerda que, estando los datos sonoros L, R (o M) codificados en comprensión en un dominio transformado, la combinación de filtros puede aplicarse directamente en el dominio transformado en función de las energías objetivo ( $\sigma_{FL}$ ,  $\sigma_{BL}$ ,  $\sigma_{FR}$ ,  $\sigma_{BR}$ ) asociadas a los canales del formato multicanal, determinándose estas energías objetivo a partir de los parámetros de espacialización ESPAC. En esta realización, naturalmente, a continuación se proporciona el paso de nuevo del dominio transformado al dominio del tiempo para la restitución propiamente dicha en contexto binaural (módulos TRANS de las figuras 6A y 6B).

La presente invención va dirigida también a un módulo de decodificación DECOD BIN tal como el representado a modo de ejemplo en la figura 7, para una restitución espacializada en tres dimensiones sobre dos vías de restitución L-BIN y R-BIN, y que comprende en particular medios de tratamiento de datos sonoros (canales comprimidos L, eventualmente R en modo estereofónico y los parámetros de espacialización ESPAC) para la puesta en práctica del procedimiento descrito anteriormente en el presente documento. Estos medios pueden comprender normalmente:

- una entrada E para recibir los canales comprimidos y los parámetros de espacialización,
- una memoria de trabajo MEM y un procesador PROC para construir las combinaciones de filtros a partir de los parámetros ESPAC y aplicar estas combinaciones respectivamente a los canales comprimidos L y R,
- y una salida S para proporcionar las señales comprimidas y filtradas para una restitución binaural espacializada sobre las respectivas vías de restitución L-BIN y R-BIN.

La presente invención va dirigida también a un programa informático, destinado a almacenarse en una memoria de un módulo de decodificación, tal como la memoria MEM del módulo DECOD-BIN de la figura 7, para una restitución espacializada en tres dimensiones sobre dos vías de restitución L-BIN y R-BIN. El programa comprende por tanto instrucciones para la ejecución del procedimiento según la invención y, en particular, para construir las combinaciones de filtros que integran las versiones decorrelacionadas tal como se ilustra en las figuras 6A y 6B descritas anteriormente en el presente documento. A este respecto, una u otra de estas figuras pueden constituir un organigrama que representa el algoritmo básico del programa.

## REIVINDICACIONES

1. Procedimiento de tratamiento de datos sonoros para una restitución espacializada en tres dimensiones sobre dos vías de restitución para los oídos respectivos de un oyente,

estando los datos sonoros inicialmente representados en un formato multicanal, a continuación codificados en compresión (COD) sobre un número reducido de canales (L, R),

consistiendo dicho formato multicanal en proporcionar más de dos canales susceptibles de alimentar dos altavoces respectivos, comprendiendo el procedimiento las etapas de:

- obtener, con los datos comprimidos sobre dicho número reducido de canales, parámetros de espacialización (ESPAC),
- para cada vía de restitución asociada a un oído del oyente, formar, a partir de dichos parámetros de espacialización, una combinación de filtros representativos cada uno de funciones de transferencia (HRTF) entre este oído del oyente y altavoces susceptibles de alimentarse mediante canales respectivos del formato multicanal inicial, y
- aplicar a los datos comprimidos la combinación de filtros ( $h_{L,L}$ ,  $h_{L,R}$ ,  $h_{L,C}$ ,  $h_{R,R}$ ,  $h_{R,L}$ ,  $h_{R,C}$ ) asociada a cada vía de restitución (L-BIN; R-BIN),

**caracterizado** porque el procedimiento comprende las etapas de:

- para cada vía de restitución asociada a un oído del oyente, determinar, a partir de dichos parámetros de espacialización, al menos una función de transferencia de un altavoz situado detrás del oído del oyente y representativa de una decorrelación entre los canales del formato multicanal respectivamente asociados al altavoz trasero y a al menos un altavoz situado delante del oído del oyente, y
- para cada vía de restitución, integrar dicha función de transferencia representativa de una decorrelación en dicha combinación de filtros asociada a esta vía de restitución.

2. Procedimiento según la reivindicación 1, **caracterizado** porque la combinación de filtros asociada a una vía de restitución (L-BIN) comprende al menos un primer reagrupamiento, que forma un primer filtro ( $h_{L,L}$ ), a partir:

- de la función de transferencia de un altavoz delantero (HRTF-A),
- de la función de transferencia de un altavoz trasero (HRTF-C), y
- de una versión (HRTF-C\*) de la función de transferencia del altavoz trasero, representativa de una decorrelación entre canales,

y porque los altavoces delantero y trasero están situados en un mismo primer lado con respecto al oyente.

3. Procedimiento según la reivindicación 2, **caracterizado** porque dicho reagrupamiento comprende una ponderación, según un coeficiente ( $\alpha_1$ ;  $\alpha_2$ ) elegido entre:

- la función de transferencia del altavoz situado detrás, y
- la versión representativa de una decorrelación de esta función de transferencia del altavoz trasero.

4. Procedimiento según la reivindicación 3, **caracterizado** porque la codificación en compresión pone en práctica un codificador (COD) paramétrico que suministra una información de decorrelación entre canales del formato multicanal, y porque el coeficiente de ponderación está representado por una función variable dinámicamente en función de la información de decorrelación ( $ICC_L$ ;  $ICC_R$ ) que suministra el codificador paramétrico.

5. Procedimiento según una de las reivindicaciones 2 a 4, en el que los datos sonoros se codifican en compresión sobre dos canales (L,R),

**caracterizado** porque la combinación de filtros asociada a dicha vía de restitución (L-BIN) comprende, además de dicho primer reagrupamiento que forma el filtro ( $h_{L,L}$ ) de uno de los canales comprimidos (L), un segundo reagrupamiento que forma el filtro ( $h_{L,R}$ ) del otro de los canales comprimidos (R) a partir:

- de la función de transferencia de un altavoz delantero (HRTF-H) situado en un segundo lado, opuesto al primer lado con respecto al oyente,

## ES 2 334 856 T3

- de la función de transferencia de un altavoz trasero (HRTF-E) situado en dicho segundo lado, y
- de una versión (HRTF-E\*) de la función de transferencia de este altavoz trasero, representativa de una decorrelación entre canales.

5

6. Procedimiento según una de las reivindicaciones anteriores, **caracterizado** porque, estando los datos sonoros codificados en comprensión en un dominio transformado, la combinación de filtros se aplica en el dominio transformado en función de energías objetivo asociadas a los canales del formato multicanal, determinándose estas energías objetivo a partir de dichos parámetros de espacialización.

10

7. Procedimiento según una de las reivindicaciones anteriores, **caracterizado** porque dichas funciones de transferencia de los altavoces son de tipo HRTF y representan perturbaciones acústicas sobre caminos entre cada altavoz y un oído por una vía de restitución asociada a este oído.

15

8. Procedimiento según las reivindicaciones 6 y 7, en el que el dominio transformado es el dominio de las subbandas, **caracterizado** porque las versiones decorrelacionadas de las funciones HRTF de los altavoces traseros se obtienen aplicando a las funciones HRTF iniciales de los altavoces traseros un desfase que está en función de cada subbanda de frecuencias.

20

9. Módulo de decodificación (DECOD BIN) para una restitución espacializada en tres dimensiones sobre dos vías de restitución, **caracterizado** porque comprende medios de tratamiento de datos sonoros para la puesta en práctica del procedimiento según una de las reivindicaciones anteriores.

25

10. Programa informático, destinado a almacenarse en una memoria de un módulo de decodificación para una restitución espacializada en tres dimensiones sobre dos vías de restitución, **caracterizado** porque comprende instrucciones para la ejecución del procedimiento según una de las reivindicaciones 1 a 8.

30

35

40

45

50

55

60

65

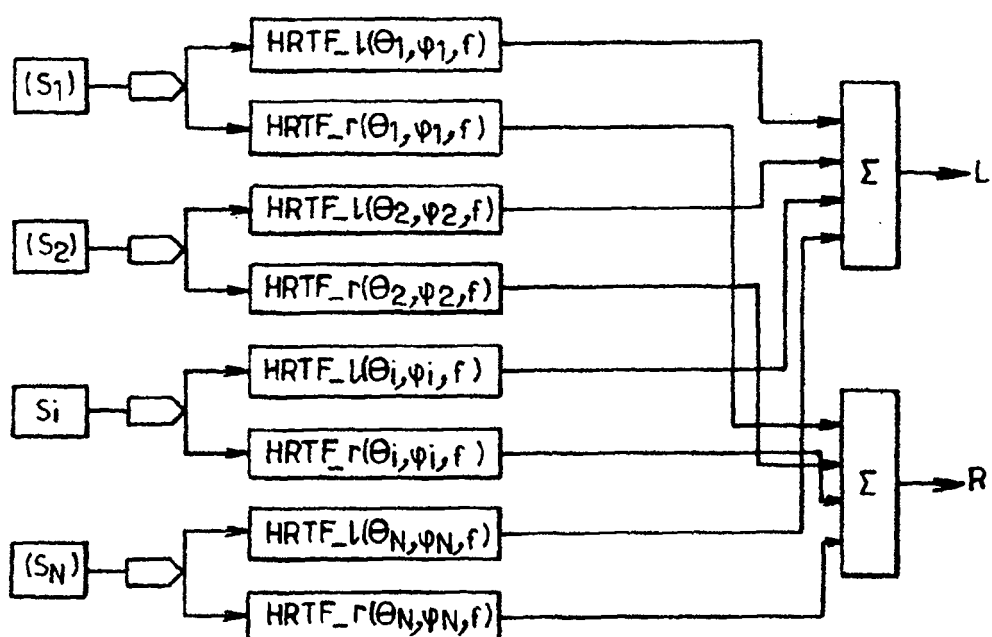
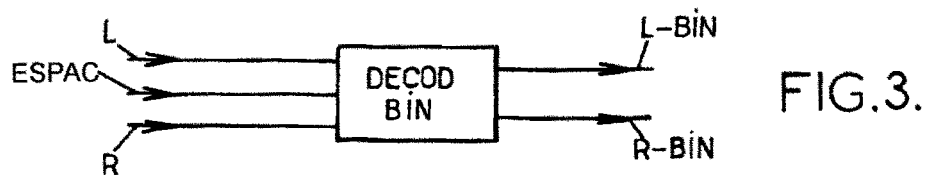
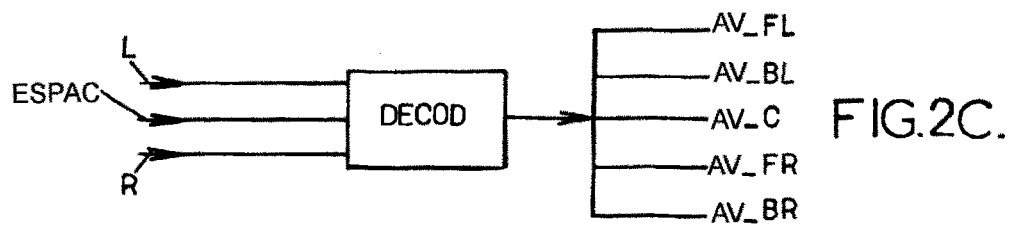
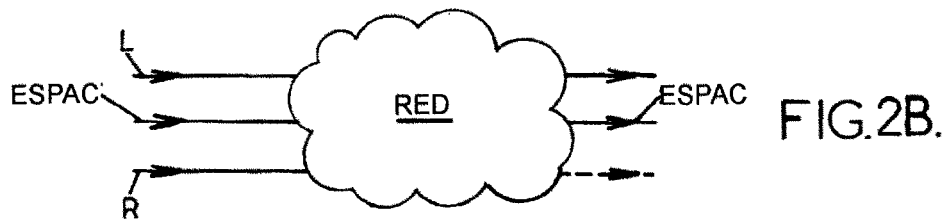
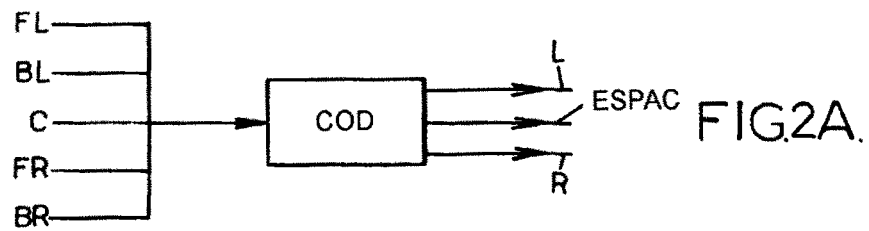
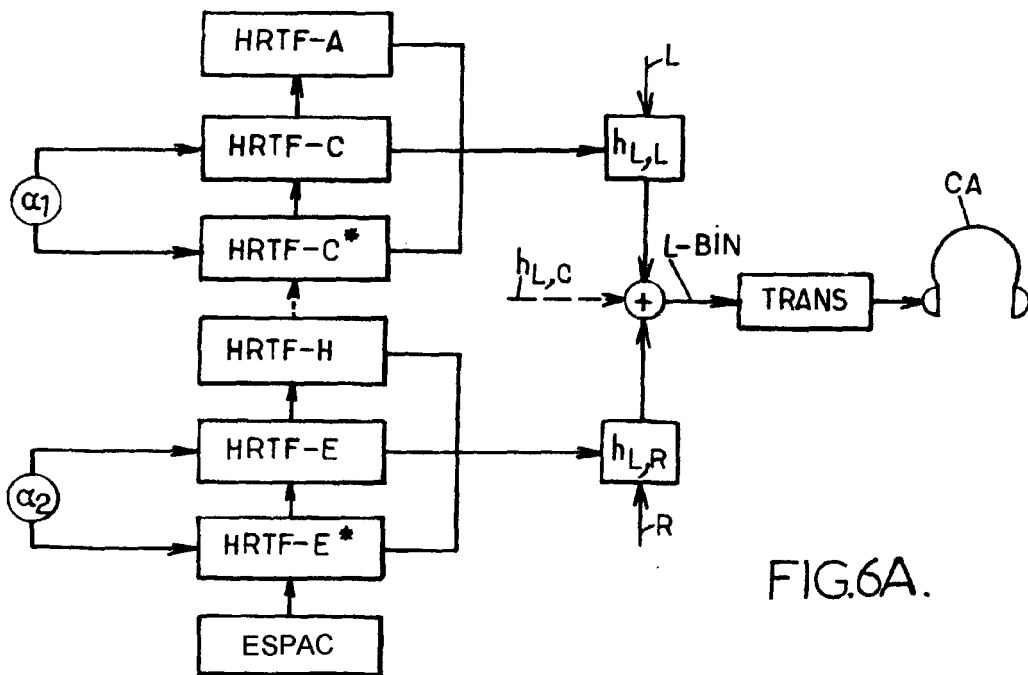
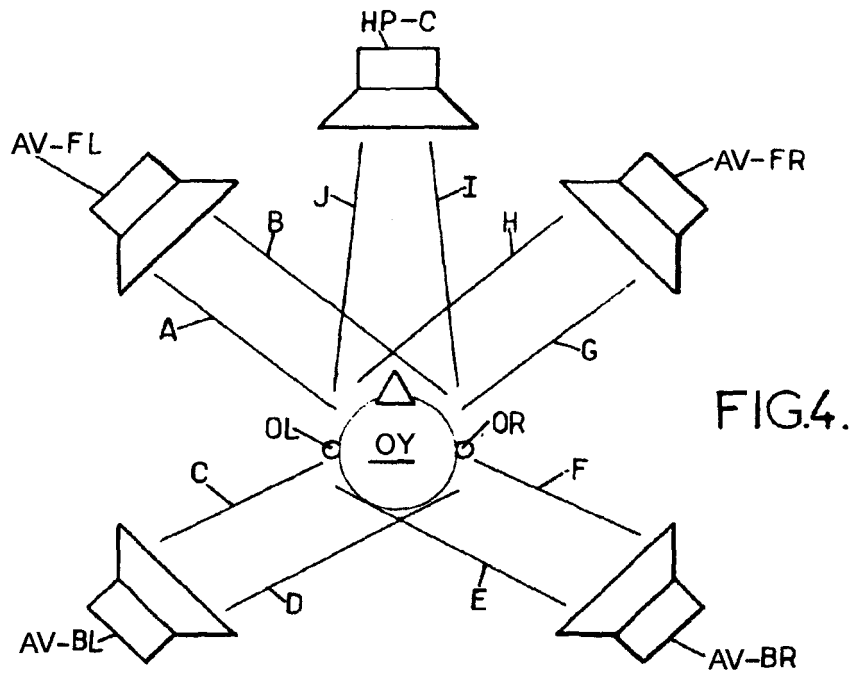


FIG.1.  
(TÉCNICA ANTERIOR)





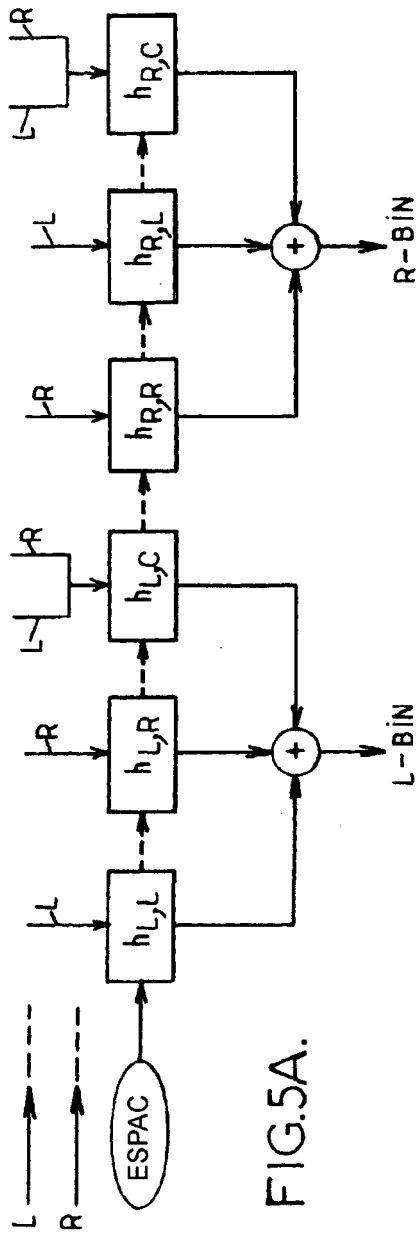


FIG. 5A.

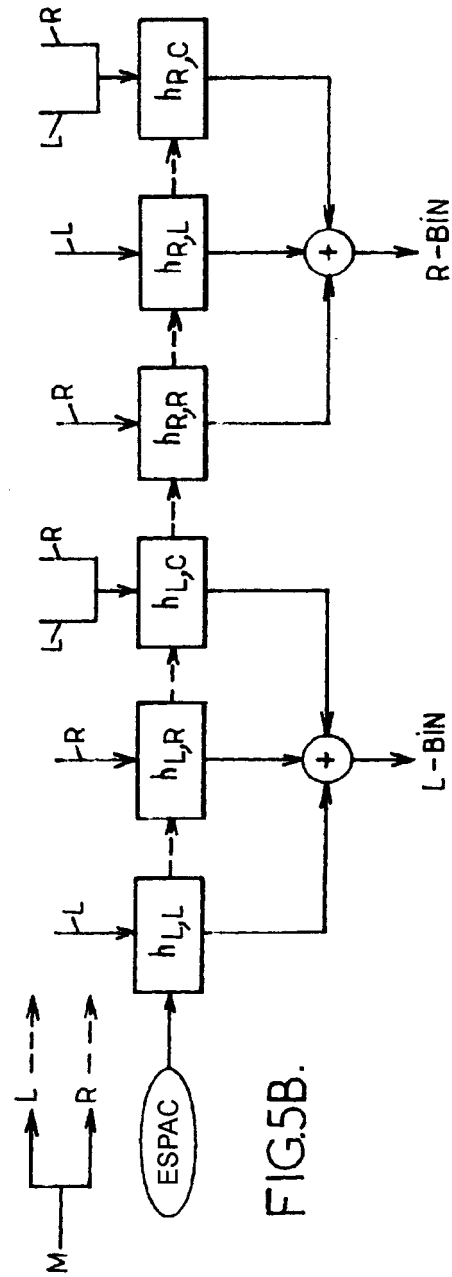


FIG. 5B.



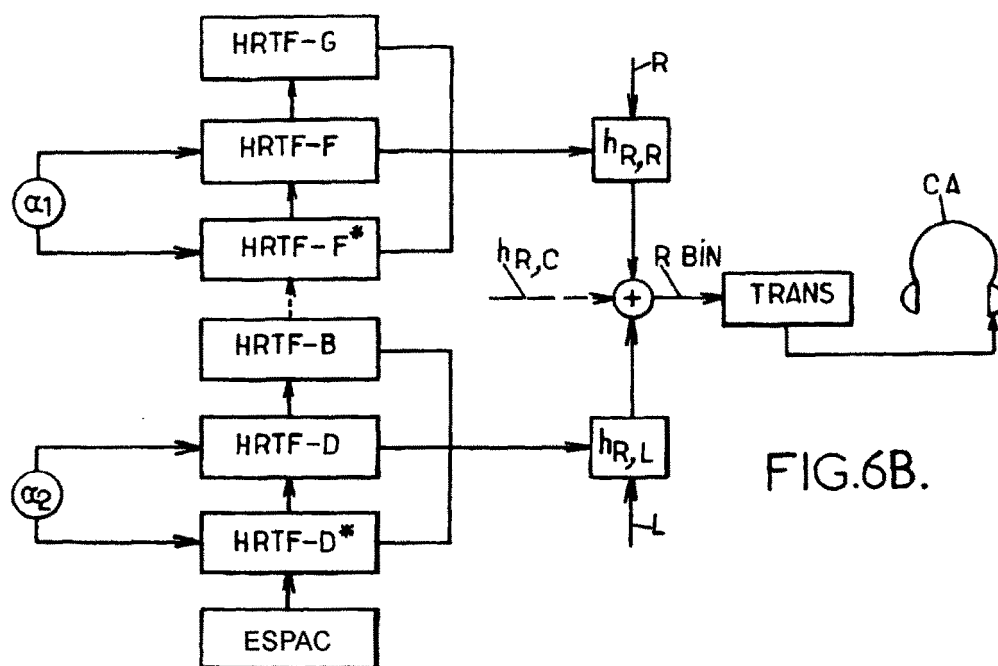


FIG. 6B.

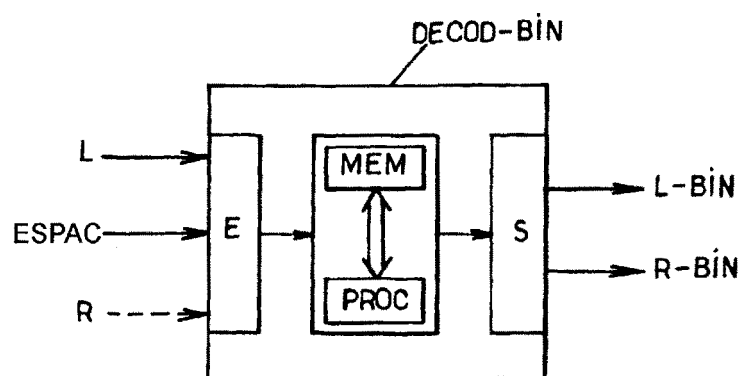


FIG. 7.