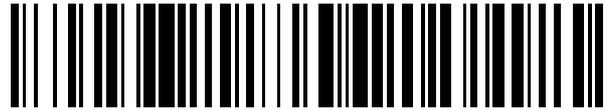


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 435 792**

51 Int. Cl.:

**G10L 19/00** (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **11.12.2009 E 09803839 (1)**

97 Fecha y número de publicación de la concesión europea: **29.05.2013 EP 2374124**

54 Título: **Codificación perfeccionada de señales digitales de audio multicanal**

30 Prioridad:

**15.12.2008 FR 0858563**

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

**23.12.2013**

73 Titular/es:

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

72 Inventor/es:

**JAILLET, FLORENT y  
VIRETTE, DAVID**

74 Agente/Representante:

**PÉREZ BARQUÍN, Eliana**

**ES 2 435 792 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**

Codificación perfeccionada de señales digitales de audio multicanal

5 La presente invención se refiere al campo de la codificación/decodificación de señales digitales de audio multicanal.  
De manera más particular, la presente invención se refiere a la codificación/decodificación paramétrica de señales de audio multicanal.

10 Este tipo de codificación/decodificación se basa en la extracción de parámetros de espacialización para que se pueda reconstituir en la decodificación la percepción espacial del oyente.

Este tipo de técnica de decodificación se conoce con el nombre de "Binaural Cue Coding" en inglés (BCC) que busca, por una parte, extraer y a continuación codificar los índices de espacialización auditiva y, por otra parte, codificar una señal monofónica o estereofónica resultante de un matizado de la señal multicanal original.

15 Este enfoque paramétrico es una codificación de baja velocidad. La principal ventaja de este enfoque de codificación es que permite una tasa de compresión mejor que los métodos clásicos de compresión de señales digitales de audio multicanal garantizando al mismo tiempo la compatibilidad con versiones anteriores del formato comprimido que se obtiene con los formatos de codificación y los sistemas de difusiones ya existentes.

20 La norma MPEG Surround que se describe en el documento de la norma MPEG ISO/IEC 23003-1:2007 y en el documento de "Breebaart, J., Hotho, G., Koppens, J., Schuijers, E., Oomen, W. y van de Par, S.", titulado "Background, concept and architecture for the recent MPEG surround standard on multichannel audio compression" en el Journal of the Audio Engineering Society, 55-5 (2007), págs. 331-351, describe una estructura de codificación paramétrica como la que se representa en la figura 1.

30 Otro ejemplo de estructura de codificación paramétrica se describe en el documento de Bin Cheng, Christian Ritz e Ian Burnett titulado "Encoding Independent Sources in Spatially Squeezed Surround Audio Coding" en Advances in Multimedia Information Processing - PCM2a7, Lecture Notes in Computer Science Volume, 4.810, 2007, págs. 804-813.

35 Así pues, la figura 1 describe dicho sistema de codificación/decodificación en el cual el codificador 100 construye una señal compuesta ("downmix" en inglés)  $S_c$  mediante el matizado en 110 de los canales de la señal multicanal original  $S$  y suministra a través de un módulo de extracción de parámetros 120, un conjunto reducido de parámetros  $P$  que caracterizan el contenido espacial de la señal multicanal original.

40 En el decodificador 150, la señal multicanal se reconstruye ( $S'$ ) mediante un módulo de síntesis 160 que tiene en cuenta a la vez la señal compuesta y los parámetros  $P$  transmitidos.

45 La señal compuesta comprende un número reducido de canales. Estos canales se pueden codificar mediante un codificador de audio clásico antes de su transmisión o almacenamiento. Tradicionalmente, la señal compuesta comprende dos canales y es compatible con una difusión estéreo clásica. Antes de su transmisión o almacenamiento, esta señal compuesta puede de este modo codificarse mediante cualquier codificador estéreo clásico. De este modo la señal codificada es entonces compatible con los dispositivos que comprende el decodificador correspondiente que reconstruyen la señal compuesta ignorando los datos espaciales.

50 Cuando se realiza este tipo de codificación mediante el matizado de una señal multicanal para obtener una señal compuesta, tras la transformación en el espacio frecuencial de la señal multicanal, pueden aparecer problemas de reconstrucción de la señal multicanal.

55 En efecto, en este caso, no hay necesariamente coherencia espacial entre la señal compuesta y el sistema de restitución en el cual se puede reproducir la señal. Por ejemplo, cuando la señal compuesta contiene dos canales, una restitución estereofónica debe permitir respetar la posición relativa de las fuentes sonoras en el espacio sonoro reconstruido. La posición izquierda/derecha de las fuentes sonoras debe poder respetarse.

Además, tras el matizado por banda de frecuencia, la señal compuesta resultante se transmite a continuación al decodificador en forma de una señal temporal.

60 El paso del espacio tiempo-frecuencia al espacio temporal implica unas interacciones entre las bandas de frecuencia y las tramas temporales próximas que introducen fallos y artefactos molestos.

65 Existe, por lo tanto, la necesidad de una técnica de codificación/decodificación paramétrica por banda de frecuencia que permita limitar los defectos introducidos por los pasos de las señales del campo tiempo-frecuencia al campo temporal y controlar la coherencia espacial entre la señal de audio multicanal y la señal compuesta resultante de un matizado de fuentes sonoras.

La presente invención viene a mejorar la situación.

5 Para ello, propone un procedimiento de codificación de una señal de audio multicanal que representa una escena sonora que comprende una multitud de fuentes sonoras. El procedimiento es tal que comprende una etapa de descomposición de la señal multicanal en bandas de frecuencia y las siguientes etapas por banda de frecuencia:

- obtención de datos representativos de la dirección de las fuentes sonoras de la escena sonora;

10 - selección de un conjunto de fuentes sonoras de la escena sonora constituyendo las fuentes principales;

- adaptación de los datos representativos de la dirección de las fuentes principales seleccionadas, en función de las características de restitución de la señal multicanal, mediante la modificación de la posición de las fuentes para obtener una distancia mínima entre dos fuentes;

15 - determinación de una matriz de mezcla de las fuentes principales en función de los datos adaptados;

- matizado de las fuentes principales mediante la matriz determinada para obtener una señal compuesta con un número reducido de canales;

20 - codificación de los datos representativos de la dirección de las fuentes sonoras y formación de un flujo binario que comprende los datos codificados, adaptándose el flujo binario para transmitirse en paralelo a la señal compuesta.

25 De este modo, para la obtención de la señal compuesta, la matriz de mezcla tiene en cuenta los datos de informaciones de dirección de las fuentes. Esto permite adaptar la señal compuesta resultante, para una buena restitución del sonido en el espacio durante la reconstrucción de esta señal en el decodificador. La señal compuesta se adapta entonces a las características de restitución de la señal multicanal y a los eventuales solapamientos de las posiciones de las fuentes sonoras. De este modo se respeta la coherencia espacial entre la señal compuesta y la señal multicanal.

30 La adaptación de los datos que modifican la posición de las fuentes para obtener una distancia mínima entre dos fuentes permite, de este modo, que las dos fuentes que estarían tras la restitución sonora demasiado próximas entre sí, se separen para que la restitución de la señal permita que el oyente diferencie la posición de estas fuentes.

35 Al codificar por separado los datos de dirección y las fuentes sonoras por banda de frecuencia, se saca partido al hecho de que el número de fuentes activas en una banda de frecuencia es por lo general bajo, lo que aumenta los rendimientos de codificación.

40 No es necesario transmitir otros datos de reconstrucción de la matriz de mezcla al decodificador ya que esta se determinará a partir de los datos de direcciones codificadas.

45 Los diferentes modos particulares de realización que se exponen a continuación se pueden añadir de forma independiente o combinándose unos con otros, en las etapas del procedimiento de codificación definido con anterioridad.

En un modo de realización, los datos representativos de la dirección son informaciones de directividades representativas de la distribución de las fuentes sonoras en la escena sonora.

50 Las informaciones de directividad asociadas a una fuente dan no solo la dirección de la fuente, sino también la forma, o la distribución espacial, de la fuente, es decir la interacción que puede tener esta fuente con las demás fuentes de la escena sonora.

55 El conocimiento de estas informaciones de directividades asociado a la señal compuesta va a permitir que el decodificador obtenga una señal de mejor calidad que tiene en cuenta las redundancias entre los canales de manera global y las probables oposiciones de fase entre canales.

En un modo particular de realización, la codificación de las informaciones de directividades se realiza mediante un método de representación paramétrica.

60 Este método tiene una baja complejidad y se adapta en particular al caso de las escenas sonoras de síntesis que representan una situación de codificación ideal.

65 En otro modo de realización, la codificación de las informaciones de directividad se lleva a cabo mediante un método de análisis de componentes principales que emite unos vectores de directividad de base asociados a unas ganancias que permiten la reconstrucción de las directividades iniciales.

Así pues, esto permite codificar las directividades de escenas sonoras complejas cuya codificación no se puede representar fácilmente mediante un modelo.

5 En otro modo de realización la codificación de las informaciones de directividad se lleva a cabo mediante la combinación de un método de análisis de componentes principales y de un método de representación paramétrica.

De este modo, se puede, por ejemplo, realizar en paralelo la codificación mediante los dos métodos y seleccionar aquella que responde a un criterio de optimización de la velocidad de codificación, por ejemplo.

10 También se puede realizar estos dos métodos en cascada de tal modo que se codifique simplemente una parte de las directividades mediante el método de codificación paramétrica y para aquellas que no están modelizadas llevar a cabo una codificación mediante el método de análisis de componentes principales, de tal modo que se representen de la mejor manera posible todas las directividades. La distribución del flujo entre los dos modelos de codificación de las directividades se puede seleccionar de acuerdo con un criterio de minimización del error de reconstrucción de las  
15 directividades.

En un modo de realización de la invención, el procedimiento comprende, además, la codificación de fuentes secundarias entre las fuentes no seleccionadas de la escena sonora y de inserción de informaciones de codificación de las fuentes secundarias en el flujo binario.  
20

La codificación de las fuentes secundarias va a permitir, de este modo, aportar una precisión adicional sobre la señal decodificada, en particular para las señales complejas de tipo, por ejemplo, ambiofónica.

25 La presente invención también se refiere a un procedimiento de decodificación de una señal de audio multicanal que representa una escena sonora que comprende una multitud de fuentes sonoras, a partir de un flujo binario y de una señal compuesta. El procedimiento es tal que comprende las siguientes etapas.

30 - extracción en el flujo binario y decodificación de datos representativos de la dirección de las fuentes sonoras en la escena sonora;

- adaptación de al menos una parte de los datos de dirección en función de características de restitución de la señal multicanal, mediante la modificación de la posición de las fuentes obtenidas por los datos de dirección, para obtener una distancia mínima entre dos fuentes;

35 - determinación de una matriz de mezcla de la señal compuesta en función de los datos adaptados y cálculo de la inversa de la matriz de mezcla;

40 - desmatrizado de la señal compuesta mediante la inversa de la matriz de mezcla para obtener un conjunto de fuentes principales;

- reconstrucción de la señal de audio multicanal mediante la espacialización al menos de las fuentes principales con los datos extraídos decodificados.

45 Los datos de direcciones decodificados van a permitir de este modo encontrar la inversa de la matriz de mezcla utilizada por el codificador. Esta matriz de mezcla permite encontrar a partir de la señal compuesta, las fuentes principales que se restituirán en el espacio con una buena coherencia espacial.

50 La etapa de adaptación permite de este modo encontrar las direcciones de las fuentes que hay que espacializar de tal modo que se obtenga una restitución del sonido que sea coherente con el sistema de restitución.

La señal reconstruida se adapta bien por tanto a las características de restitución de la señal multicanal evitando los eventuales solapamientos de las posiciones de las fuentes sonoras.

55 De este modo, dos fuentes demasiado próximas se separan para restituirse de tal modo que un oyente pueda diferenciarlas.

En un modo de realización, el procedimiento de decodificación comprende, además, las siguientes etapas:

60 - extracción del flujo binario, de informaciones de codificación de fuentes secundarias codificadas;

- decodificación de las fuentes secundarias a partir de las informaciones de codificación extraídas;

- agrupamiento de las fuentes secundarias con las fuentes principales para la espacialización.

65 La decodificación de fuentes secundarias aporta por tanto más precisión en la escena sonora.

La presente invención también se refiere a un codificador de una señal de audio multicanal que representa una escena sonora que comprende una multitud de fuentes sonoras. El codificador es tal que comprende:

- 5 - un módulo de descomposición de la señal multicanal en bandas de frecuencia;
- un módulo de obtención de datos representativos de la dirección de las fuentes sonoras de la escena sonora;
- 10 - un módulo de selección de un conjunto de fuentes sonoras de la escena sonora que constituyen las fuentes principales;
- un módulo de adaptación de los datos representativos de la dirección de las fuentes principales seleccionadas, en función de características de restitución de la señal multicanal, mediante unos medios de modificación de la posición de las fuentes para obtener una distancia mínima entre dos fuentes;
- 15 - un módulo de determinación de una matriz de mezcla de las fuentes principales en función de los datos resultantes del módulo de adaptación;
- un módulo de matizado de las fuentes principales seleccionadas mediante la matriz determinada para obtener una señal compuesta con un número reducido de canales;
- 20 - un módulo de codificación de los datos representativos de la dirección de las fuentes sonoras; y
- un módulo de formación de un flujo binario que comprende los datos codificados, adaptándose el flujo binario para transmitirse en paralelo a la señal compuesta.
- 25 Esta se refiere también a un decodificador de una señal de audio multicanal que representa una escena sonora que comprende una multitud de fuentes sonoras, recibiendo en la entrada un flujo binario y una señal compuesta. El decodificador es tal que comprende:
- 30 - un módulo de extracción y de decodificación de datos representativos de la dirección de las fuentes sonoras en la escena sonora;
- un módulo de adaptación de al menos una parte de los datos de dirección en función de las características de restitución de la señal multicanal, mediante unos medios de modificación de la posición de las fuentes que se obtienen mediante los datos de dirección, para obtener una distancia mínima entre dos fuentes;
- 35 - un módulo de determinación de una matriz de mezcla de la señal compuesta en función de los datos resultantes del módulo de adaptación y de cálculo de la inversa de la matriz de mezcla;
- 40 - un módulo de desmatizado de la señal compuesta mediante la inversa de la matriz de mezcla para obtener un conjunto de fuentes principales;
- un módulo de reconstrucción de la señal de audio multicanal mediante la espacialización al menos de las fuentes principales con los datos extraídos decodificados.
- 45 Esta también se refiere a un programa informático que comprende unas instrucciones de código para la aplicación de las etapas de un procedimiento de codificación como el que se ha descrito y/o de un procedimiento de decodificación como se ha descrito, cuando estas instrucciones las ejecuta un procesador.
- 50 De manera más general, un medio de almacenamiento, legible por un ordenador o un procesador, integrado o no en el codificador, eventualmente amovible, memoriza un programa informático que aplica un procedimiento de codificación y/o un procedimiento de decodificación de acuerdo con la invención.
- 55 Se mostrarán de manera más clara otras características y ventajas de la invención en la lectura de la siguiente descripción, que se da únicamente a título de ejemplo no excluyente, y que se hace en referencia a los dibujos adjuntos, en los que:
- 60 - la figura 1 ilustra un sistema de codificación/decodificación del estado de la técnica de tipo sistema normalizado MPEG Surround;
- la figura 2 ilustra un codificador y un procedimiento de codificación de acuerdo con un modo de realización de la invención;
- 65 - la figura 3a ilustra un primer modo de realización de la codificación de las directividades de acuerdo con la invención;

- la figura 3b ilustra un segundo modo de realización de la codificación de las directividades de acuerdo con la invención;

5 - la figura 4 ilustra un organigrama que representa las etapas de la determinación de una matriz de mezcla de acuerdo con un modo de realización de la invención;

- la figura 5a ilustra un ejemplo de distribución de fuentes sonoras alrededor de un oyente;

10 - la figura 5b ilustra la adaptación de la distribución de fuentes sonoras alrededor de un oyente para adaptar los datos de dirección de las fuentes sonoras de acuerdo con un modo de realización de la invención;

- la figura 6 ilustra un decodificador y un procedimiento de decodificación de acuerdo con un modo de realización de la invención; y

15 - las figuras 7a y 7b representan respectivamente un ejemplo de dispositivo que comprende un codificador y un ejemplo de dispositivo que comprende un decodificador de acuerdo con la invención.

20 La figura 2 ilustra en forma de diagrama de bloques, un codificador de acuerdo con un modo de realización de la invención así como las etapas de un procedimiento de codificación de acuerdo con un modo de realización de la invención.

25 Todos los tratamientos en este codificador se llevan a cabo por trama temporal. En aras de la simplificación, la representación y la descripción del codificador tal y como se representa en la figura 2 se hace considerando el tratamiento realizado en una trama temporal fija, sin que aparezca la dependencia temporal en las anotaciones.

Sin embargo, un mismo tratamiento se aplica de forma sucesiva a todas las tramas temporales de la señal.

30 Así pues, el codificador que se ilustra comprende un módulo de transformada tiempo-frecuencia 210 que recibe en la entrada una señal multicanal original que representa una escena sonora que comprende una multitud de fuentes sonoras.

Este módulo lleva a cabo una etapa T de cálculo de la transformada tiempo-frecuencia de la señal multicanal original  $S_m$ . Esta transformada se realiza, por ejemplo, mediante una transformada de Fourier a corto plazo.

35 Para ello, cada uno de los  $n_x$  canales de la señal original se visualiza en la trama temporal actual, a continuación se calcula la transformada de Fourier F de la señal visualizada mediante un algoritmo de cálculo rápido en  $n_{FFT}$  puntos. De este modo se obtiene una matriz X compleja de tamaño  $n_{FFT} \times n_x$  que contiene los coeficientes de la señal multicanal original en el espacio frecuencial.

40 A continuación los tratamientos que lleva a cabo el codificador se realizan por banda de frecuencia. Para ello, se corta la matriz de los coeficientes X en un conjunto de sub-matrices  $X_j$  conteniendo cada una los coeficientes de frecuencia en la j-ésima banda.

45 Existen diferentes opciones para el corte de frecuencia de las bandas. Con el fin de garantizar que el tratamiento se aplique en señales reales, se selecciona unas bandas simétricas con respecto a la frecuencia nula en la transformada de Fourier a corto plazo. Además, con el fin de optimizar la eficacia de codificación, se da preferencia a la elección de bandas de frecuencia que se aproximan a las escalas de frecuencia aconsejadas, por ejemplo seleccionado unos anchos de banda constantes en las escalas ERB (por "Equivalent Rectangular Bandwidth" en inglés) o Bark.

50 En aras de la simplificación, la descripción de las etapas de codificación que lleva a cabo el codificador se hará para una banda de frecuencia dada. Por supuesto, las etapas se realizan para cada una de las bandas de frecuencia que hay que tratar.

55 A la salida del módulo 210, se obtiene, por lo tanto, la señal para una banda de frecuencia dada  $S_{ij}$ .

60 Un módulo de obtención de datos de direcciones de las fuentes sonoras 220 permite determinar mediante una etapa OBT, por una parte, los datos de dirección asociados a cada una de las fuentes de la escena sonora y, por otra parte, determinar las fuentes de la escena sonora para la banda de frecuencia dada.

Los datos de direcciones pueden ser, por ejemplo, unos datos de dirección de llegada de una fuente que corresponden a la posición de la fuente.

65 Los datos de este tipo se describen, por ejemplo, en el documento de M. Goodwin y J.-M. Jot, "Analysis and synthesis for universal spatial audio coding", 121<sup>o</sup> AES Convention, octubre de 2006.

En otro modo de realización, los datos de las direcciones son unos datos de diferencias de intensidad entre las fuentes sonoras. Estas diferencias de intensidad permiten definir unas posiciones medias de las fuentes. Toman, por ejemplo, la apelación CLD (por "Channel Level Differences" en inglés) para el codificador normalizado MPEG Surround.

5 En el modo de realización que se describe aquí en detalle, los datos representativos de las direcciones de las fuentes son informaciones de directividades.

10 Las informaciones de directividades son representativas de la distribución espacial de las fuentes sonoras en la escena sonora.

Las directividades son unos vectores con una dimensión igual al número  $n_s$  de canales de la señal multicanal  $S_m$ .

15 Cada fuente está asociada a un vector de directividad.

Para una señal multicanal, el vector de directividad asociado a una fuente corresponde a la función de ponderación que hay que aplicar a esta fuente antes de reproducirla por un altavoz, de tal modo que se reproduzca lo mejor posible una dirección de llegada y un ancho de fuente.

20 Se entiende fácilmente que para un número muy elevado de altavoces separados de forma regular, el vector de directividad permite representar fielmente la radiación de una fuente sonora. En presencia de una señal ambiofónica, el vector de directividad se obtiene mediante la aplicación de una transformada de Fourier inversa esférica sobre los componentes de las órdenes ambiofónicas. En efecto, las señales ambiofónicas corresponden a una descomposición en armónicos esféricos, de ahí la correspondencia directa con las directividad de las fuentes.

25 El conjunto de los vectores de directividad constituye, por lo tanto, una importante cantidad de datos que sería demasiado costoso de transmitir directamente para unas aplicaciones de baja velocidad de codificación. Para reducir la cantidad de información que hay que transmitir, se pueden utilizar por ejemplo dos métodos de representación de las directividades.

30 De este modo el módulo 230 de codificación Cod.Di de las informaciones de directividades puede aplicar uno de los dos métodos que se describen a continuación o incluso una combinación de los dos métodos.

35 Un primer método es un método de modelización paramétrica que permite sacar partido de los conocimientos *a priori* sobre el formato de señal utilizado. Esta consiste en transmitir únicamente un número muy reducido de parámetros y en reconstruir las directividades en función de modelos de codificación conocidos.

40 Por ejemplo, se trata de sacar provecho de los conocimientos sobre la codificación de las ondas planas para unas señales de tipo ambiofónica para transmitir únicamente el valor de la dirección (azimut y elevación) de la fuente. Con estas informaciones, es entonces posible reconstruir la directividad correspondiente a una onda plana procedente de esta dirección.

45 Por ejemplo, para una orden ambiofónica definida, se conoce la directividad asociada en función de la dirección de llegada de la fuente sonora. Existen numerosos métodos de estimación de los parámetros del modelo. Así pues, una búsqueda de picos en el diagrama de directividad (por analogía con el análisis sinusoidal, tal y como se explica, por ejemplo, en el documento "*Modélisation informatique du son musical (analyse, transformation, synthèse)*" de Sylvain Marchand, PhD thesis, Université Bordeaux 1, permite detectar de una manera relativamente fiel la dirección de llegada.

50 En este contexto también se pueden utilizar otros métodos como el "matching pursuit", tal y como se presenta en S. Mallat, Z. Zhang, Matching pursuit with time-frequency dictionaries, IEEE Transactions on Signal Processing 41 (1993), págs. 3.397-3.415, o el análisis espectral paramétrico.

55 Una representación paramétrica también puede utilizar un diccionario de forma simple para representar las directividades. En la codificación de las directividades, se asocia a un elemento del diccionario un dato, por ejemplo el azimut correspondiente, y una ganancia que permite jugar con la amplitud de este vector de directividad del diccionario. De este modo se puede, a partir de un diccionario de forma de directividad, deducir la mejor forma o la combinación de formas que permitirá reconstruir lo mejor posible la directividad inicial.

60 Para la aplicación de este primer método, el módulo 230 de codificación de las directividades comprende un módulo de modelización paramétrica que da en la salida unos parámetros de directividades P. Estos parámetros se cuantifican a continuación mediante el módulo de cuantificación 240.

65 Este primer método permite obtener un muy buen nivel de compresión cuando la escena corresponde a una codificación ideal. Este será el caso en particular en las escenas sonoras de síntesis.

Sin embargo, para las escenas complejas o procedentes de tomas de sonido microfónicas, es necesario utilizar unos modelos de codificación más genéricos, que implican la transmisión de una mayor cantidad de informaciones.

5 El segundo método que se describe a continuación permite superar este inconveniente. En este segundo método, la representación de las informaciones de directividad se lleva a cabo en forma de combinación lineal de un número limitado de directividades de base. Este método se basa en el hecho de que el conjunto de las directividades en un instante dado tiene por lo general una dimensión reducida. En efecto, solo un número reducido de fuentes se activa en un instante dado y la directividad para cada fuente varía poco con la frecuencia.

10 De este modo se puede representar el conjunto de las directividades en un grupo de bandas de frecuencia a partir de un número más reducido de directividades de base bien seleccionadas. Los parámetros que se transmiten son entonces los vectores de directividad de base para el grupo de bandas consideradas, y para cada directividad que hay que codificar, los coeficientes que hay que aplicar a las directividades de base para reconstruir la directividad considerada.

15 Este método se basa en un método de análisis de componentes principales (ACP o PCA en inglés por "Principal Component Analysis"). Esta herramienta la desarrolla ampliamente I. T. Jolliffe en "Principal Component Analysis", Springer, 2002. La aplicación del análisis de componentes principales en la codificación de las directividades se lleva a cabo de la siguiente manera: en primer lugar, se forma una matriz de las directividades iniciales  $D_i$ , cuyo número de líneas corresponde al número total de fuentes de la escena sonora, y el número de columnas corresponde al número de canales de la señal multicanal original. A continuación, se lleva a cabo el análisis propiamente dicho de componentes principales que corresponde a la diagonalización de la matriz de covarianza, y que da la matriz de vectores propios. Por último, se seleccionan los vectores propios portadores de la parte de información más importante y que corresponden a los valores propios de valor más alto. El número de vectores propios que hay que conservar puede ser fijo o variable en el tiempo en función de la velocidad disponible. Esta nueva base da, por lo tanto, la matriz  $D_B^T$ . Los coeficientes de ganancia asociados a esta base se calculan fácilmente a partir de  $G_D = D_i \cdot D_B^T$ .

20 En este modo de realización, la representación de las directividades se lleva a cabo, por lo tanto, a partir de directividades de base. La matriz de las directividades  $D_i$  se escribe como la combinación lineal de estas directividades de base. De este modo, se puede escribir  $D_i = G_D D_B$ , en la que  $D_B$  es la matriz de las directividades de base para el conjunto de las bandas y  $G_D$  la matriz de las ganancias asociadas. El número de líneas de esta matriz representa el número total de fuentes de la escena sonora y el número de columnas representa el número de vectores de directividad de base.

30 En una variante de este modo de realización, las directividades de base se envían por grupo de bandas consideradas, con el fin de representar de manera más fiel las directividades. Se puede, por ejemplo, proporcionar dos grupos de directividad de base: una para las bajas frecuencias y una para las altas frecuencias. El límite entre estos dos grupos se puede seleccionar, por ejemplo, entre 5 y 7 kHz.

35 De este modo, para cada banda de frecuencia, se transmite el vector de ganancia asociado a las directividades de base.

40 Para este modo de realización, el módulo de codificación 230 comprende un módulo de análisis de componentes principales que suministra unos vectores de directividades de base  $D_B$  y unos coeficientes o vectores de ganancia asociados  $G_D$ .

45 De este modo, tras el ACP, se codificará y se transmitirá un número limitado de vectores de directividad. Para ello, se utiliza una cuantificación escalar que lleva a cabo el módulo de cuantificación 240, de los coeficientes y de los vectores de directividades de base. El número de vectores de base que hay que transmitir se podrá fijar, o bien seleccionarse en el codificador utilizando, por ejemplo, un umbral sobre el error cuadrático medio entre la directividad original y la directividad reconstruida. De este modo, si el error es inferior al umbral, el o los vector(es) de base hasta entonces seleccionado(s) es (son) suficiente(s), no es por tanto necesario codificar un vector de base adicional.

50 En unas variantes de realización, la codificación de las directividades se realiza mediante una combinación de dos representaciones que se exponen a continuación. La figura 3a ilustra, de forma detallada, el bloque de codificación de directividades 230, en una primera variante de realización.

55 Este modo de codificación utiliza los dos esquemas de representación de las directividades. De este modo, un módulo 310 lleva a cabo una modelización paramétrica tal y como se ha explicado con anterioridad para suministrar unos parámetros de directividades (P).

60 Un módulo 320 lleva a cabo un análisis de componentes principales para suministrar a la vez unos vectores de directividades de base ( $D_B$ ) y unos coeficientes asociados ( $G_D$ ).

En esta variante un módulo de selección 330 selecciona, banda de frecuencia por banda de frecuencia, el mejor modo de codificación para la directividad seleccionando el mejor equilibrio reconstrucción de las directividades/velocidad.

- 5 Para cada directividad, la elección de la representación escogida (representación paramétrica o por combinación lineal de directividades de base) se hace de tal modo que se optimice la eficacia de la compresión.

10 Un criterio de selección es, por ejemplo, la minimización del error cuadrático medio. Una ponderación perceptual se puede utilizar eventualmente para la elección del modo de codificación de la directividad. Esta ponderación tiene como objetivo, por ejemplo, favorecer la reconstrucción de las directividades en la zona frontal, para la cual el oído es más sensible. En este caso, la función de error que hay que minimizar en el caso del modelo de codificación mediante el ACP se puede poner de la siguiente forma:

$$15 \quad E = (W(D_i - G_D D_B))^2$$

siendo  $D_i$  las directividades originales y  $W$ , la función de ponderación perceptual.

20 Los parámetros de directividades resultantes del módulo de selección se cuantifican a continuación mediante el módulo de cuantificación 240 de la figura 2.

25 En una segunda variante del bloque de codificación 230, los dos modos se siguen en cascada. La figura 3b ilustra en detalle este bloque de codificación. De este modo, en esta variante de realización, un módulo 340 de modelización paramétrica lleva a cabo una modelización para un determinado número de directividades y suministra en la salida a la vez unos parámetros de directividades ( $P$ ) para las directividades modelizadas y unas directividades no modelizadas o directividades residuales  $D_iR$ .

Estas directividades residuales ( $D_iR$ ) se codifican mediante un módulo de análisis de componentes principales 350 que suministra en la salida unos vectores de directividades de base ( $D_B$ ) y unos coeficientes asociados ( $G_D$ ).

- 30 Los parámetros de directividades, los vectores de directividades así como los coeficientes se suministran en la entrada del módulo de cuantificación 240 de la figura 2.

35 La cuantificación  $C$  se lleva a cabo reduciendo la precisión en función de datos sobre la percepción, y a continuación aplicando una codificación entrópica. También las posibilidades de aprovechamiento de la redundancia entre bandas de frecuencia o entre tramas sucesivas pueden permitir reducir la velocidad. Se pueden utilizar, por lo tanto, unas predicciones intra-tramas o inter-tramas. De manera general, se podrán utilizar los métodos clásicos de cuantificación. Por otra parte, al ser ortonormales los vectores que hay que cuantificar, esta propiedad se podrá utilizar en la cuantificación escalar de las componentes del vector. En efecto, para un vector de dimensión  $N$ , solo deberán cuantificarse  $N-1$  componentes, pudiendo recalcularse la última componente.

40 De este modo, a la salida del módulo de codificación 230 de los datos de direcciones  $D_i$  de la figura 2, los parámetros destinados al decodificador se decodifican mediante el módulo interno de decodificación 235 para encontrar las mismas informaciones que tendrá el decodificador tras la recepción de los datos de direcciones codificadas para las fuentes principales seleccionadas por el módulo 260 que se describe más adelante. De este modo se obtienen las direcciones principales.

45 Cuando se trata de datos de direcciones en forma de dirección de llegada de las fuentes, la información se puede tener en cuenta tal cual.

- 50 Cuando los datos están en forma de diferencia de intensidad entre las fuentes, se realiza una etapa de cálculo de la posición media de las fuentes para utilizar esta información en el módulo de determinación de la matriz de mezcla 275.

55 Por último, cuando los datos son unas informaciones de directividades, el módulo 235 determina una posición única por fuente llevando a cabo una media de las directividades. Esta media se puede calcular, por ejemplo, como el baricentro del vector de directividad. Estas posiciones únicas o direcciones principales las utiliza a continuación el módulo 275.

60 Esto determina en un primer momento, las direcciones de las fuentes principales y las adapta en función del criterio de coherencia espacial, al conocer el sistema de restitución de la señal multicanal.

En el caso de una restitución estereofónica, por ejemplo, la restitución se lleva a cabo por dos altavoces situados por delante del oyente.

- 65 En este caso, las etapas que aplica el módulo 275 se describen en referencia a la figura 4.

De este modo, a partir de las informaciones sobre la posición de las fuentes, así como el conocimiento de las características de restitución, las fuentes situadas detrás del oyente se llevan hacia delante en la etapa E30 de la figura 4.

5 En referencia a las figuras 5a y 5b, se ilustran las etapas de adaptación de la posición de las fuentes. De este modo, la figura 5a representa una escena sonora original con 4 fuentes sonoras (A, B, C y D) distribuidas alrededor del oyente.

10 Las fuentes C y D se sitúan detrás del oyente centrado en el centro del círculo. Las fuentes C y D se llevan a la parte de delante de la escena por simetría.

La figura 5b ilustra, en forma de flechas, esta operación.

15 La etapa E31 de la figura 4 lleva a cabo una prueba para saber si la operación anterior genera un solapamiento de las posiciones de las fuentes en el espacio. En el ejemplo de la figura 5b, este es, por ejemplo, el caso para las fuentes B y D que tras la operación de la etapa E30 se sitúan a una distancia que no permite diferenciarlas.

20 Si existen fuentes en esta situación (prueba positiva de la etapa E31), la etapa E32 modifica la posición de una de las dos fuentes en cuestión para situarla a una distancia mínima  $e_{min}$  que permite que el oyente diferencie estos interlocutores. La separación se hace de forma simétrica con respecto al punto equidistante de las dos fuentes para minimizar el desplazamiento de cada una. Si las fuentes se sitúan demasiado cerca del límite de la imagen sonora (extremo izquierdo o derecho), se sitúa la fuente más próxima a este límite en esta posición límite, y se coloca la otra fuente con la distancia mínima con respecto a la primera fuente.

25 En el ejemplo que se ilustra en la figura 5b, es la fuente B la que se desplaza de tal modo que la distancia  $e_{min}$  separa las fuentes B y D.

30 Si la prueba de la etapa E31 es negativa, las posiciones de las fuentes se mantienen y se aplica la etapa E33. Esta etapa consiste en construir una matriz de mezcla a partir de las informaciones de posiciones de las fuentes así definidas en las etapas anteriores.

35 En el caso de una restitución de la señal mediante un sistema de tipo 5.1, los altavoces se distribuyen alrededor del oyente. No es, por tanto, necesario aplicar la etapa E30 que desplaza las fuentes situadas detrás del oyente hacia delante.

Por el contrario, se puede realizar la etapa E32 de modificación de las distancias entre dos fuentes. En efecto, cuando se quiere situar una fuente sonora entre dos altavoces del sistema de restitución 5.1, puede suceder que se sitúen dos fuentes a una distancia que no permite que el oyente las diferencie.

40 Las direcciones de las fuentes se modifican, por lo tanto, para obtener una distancia mínima entre dos fuentes, tal y como se ha explicado con anterioridad.

45 La matriz de mezcla se determina, por lo tanto, en la etapa E33, en función de las direcciones obtenidas tras o sin modificaciones.

50 Esta matriz se construye de tal modo que garantice la coherencia espacial de la señal compuesta, es decir si esta se restituye sola, la señal compuesta permite ya obtener una escena sonora en la que se respeta la posición relativa de las fuentes sonoras: una fuente frontal en la escena original se percibirá bien frente al oyente, una fuente a la izquierda se percibirá a la izquierda, una fuente más a la izquierda también se percibirá más a la izquierda, lo mismo sucede a la derecha.

Con estos nuevos valores de ángulo, se construye una matriz invertible.

55 Las diferentes variantes de opción de matriz de mezcla están vinculadas a las diferentes leyes de distribución espacial o "panning" en inglés (ley seno, tangente, etc.) presentadas en "Spatial sound generation and perception by amplitude panning techniques", PhD thesis, Helsinki University of Technology, Espoo, Finlandia, 2001, V. Pulkki.

60 Se puede, por ejemplo, de manera ventajosa elegir representar los canales de la derecha con una forma de seno y los canales de la izquierda con una forma de coseno, de tal modo que esta matriz sea reversible.

Por otra parte, para que las posiciones extremas ( $-45^\circ$  y  $45^\circ$ ) estén bien representadas, se pueden seleccionar por ejemplo unos coeficientes de ponderación puestos en 1 para el canal de la izquierda y en 0 para el canal de la derecha para representar la señal en la posición  $-45^\circ$  y a la inversa para representar la señal en  $45^\circ$ .

65 Para que la posición central en  $0^\circ$  esté bien representada, los coeficientes de matrizado para el canal de la izquierda y para el canal de la derecha deben ser iguales.

A continuación se explica un ejemplo de determinación de la matriz de mezcla.

5 Al escoger la ley de “panning” como una ley tangente, se calculan las ganancias asociadas a una fuente para una señal compuesta estereofónica (2 canales) de la siguiente manera:

$$g_{Gs1} = \cos \theta_{s1}$$

10  $g_{Ds2} = \sin \theta_{s1}$

siendo  $\theta_{s1}$  el ángulo entre la fuente 1 y el altavoz izquierdo, considerando una apertura entre los altavoces de 90°.

La señal compuesta  $S_{sfi}$  se obtiene, por lo tanto, mediante la siguiente operación:

15  $S_{sfi} = S_{princ} M$

con

$$M = \begin{bmatrix} g_{Gs1} & g_{Ds1} \\ g_{Gs2} & g_{Ds2} \end{bmatrix}$$

20 Volviendo a la descripción de la figura 2, el codificador tal y como se describe en este caso comprende, además, un módulo de selección 260 adaptado para seleccionar en la etapa Selec unas fuentes principales ( $S_{princ}$ ) entre las fuentes de la escena sonora que hay que codificar ( $S_{tot}$ ).

25 Para ello, un modo particular de realización utiliza un método de análisis de componentes principales, ACP, en cada banda de frecuencia en el bloque 220 para extraer todas las fuentes de la escena sonora ( $S_{tot}$ ). Este análisis permite clasificar las fuentes en sub-bandas por orden de importancia de acuerdo con el nivel de energía, por ejemplo.

30 Las fuentes más importantes (por lo tanto con mayor energía) se seleccionan entonces mediante el módulo 260 para constituir las fuentes principales ( $S_{princ}$ ), que matriz a continuación el módulo 270 mediante la matriz  $M$  tal y como la define el módulo 275, para construir una señal compuesta ( $S_{cfi}$ ) (o “downmix” en inglés).

35 Esta señal compuesta por banda de frecuencia se somete a una transformada tiempo-frecuencia  $T^{-1}$  mediante el módulo de transformada inversa 290 con el fin de suministrar una señal compuesta temporal ( $S_c$ ). Esta señal compuesta se codifica a continuación mediante un codificador de palabras o un codificador de audio del estado de la técnica (por ejemplo: G.729.1 o MPEG-4 AAC).

40 Se pueden codificar las fuentes secundarias ( $S_{sec}$ ) mediante un módulo de codificación 280 y añadirse al flujo binario en el módulo de construcción de flujo binario 250.

45 Para estas fuentes secundarias, es decir las fuentes que no se transmiten directamente en la señal compuesta, existen diferentes alternativas de tratamientos.

Al considerarse estas fuentes como no esenciales en la escena sonora, pueden no transmitirse.

50 Sin embargo, es posible codificar una parte o la totalidad de estas fuentes secundarias mediante el módulo de codificación 280 que puede, en un modo de realización, ser un módulo de codificación por transformada de Fourier a corto plazo. Estas fuentes se pueden codificar a continuación por separado utilizando los codificadores de audio o de palabra mencionados con anterioridad.

En una variante de esta codificación, se pueden codificar directamente los coeficientes de la transformada de estas fuentes secundarias solo en las bandas en las que se considera que son importantes.

55 Las fuentes secundarias se pueden codificar mediante unas representaciones paramétricas, estas representaciones pueden ser en forma de envolvente espectral o de envolvente temporal.

60 Estas representaciones se codifican en la etapa Cod. $S_{sec}$  del módulo 280 y se insertan en la etapa Con.Fb del módulo 250, en el flujo binario con las informaciones de directividades codificadas cuantificadas. Estas representaciones paramétricas constituyen entonces unas informaciones de codificación de las fuentes secundarias.

En el caso de algunas señales multicanal, en particular de tipo ambiofónico, el codificador tal y como se ha descrito aplica una etapa adicional de pre-tratamiento P mediante un módulo de pre-tratamiento 215.

Este módulo lleva a cabo una etapa de cambio de base con el fin de expresar la escena sonora utilizando la descomposición en ondas planas del campo acústico.

5 La señal ambiofónica original se ve como la transformada de Fourier angular de un campo sonoro. De este modo, las diferentes componentes representan los valores para las diferentes frecuencias angulares. La primera operación de descomposición en ondas planas corresponde, por lo tanto, a considerar que la componente omnidireccional de la señal ambiofónica representa la frecuencia angular nula (esta componente es, por lo tanto, una componente real). A continuación, las componentes ambiofónicas siguientes (orden 1, 2, 3, etc.) se combinan para obtener los coeficientes complejos de la transformada de Fourier angular.

10 Para una descripción más precisa del formato ambiofónico, se podrá remitir a la tesis de Jérôme Daniel, titulada "Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia", 2011, Paris 6.

15 Así pues, para cada orden ambiofónica superior a 1 (en 2 dimensiones), la primera componente representa la parte real, y la segunda componente representa la parte imaginaria. Para una representación bidimensional, para una orden O, se obtienen O+1 componentes complejas. Se aplica a continuación una Transformada de Fourier a Corto Plazo (en la dimensión temporal) para obtener las transformadas de Fourier (en el campo frecuencial) de cada armónico angular, esta etapa integra entonces la etapa de transformación T del módulo 210. A continuación, se construye la transformada angular completa recreando los armónicos de frecuencias negativas por simetría hermitiana. Por último, se realiza una transformada de Fourier inversa en la dimensión de las frecuencias angulares para pasar al campo de las directividades.

20 Esta etapa de pre-tratamiento P permite que el codificador trabaje en un espacio de señales cuya interpretación física y perceptiva se simplifica, lo que permite sacar provecho de manera más eficaz de los conocimientos sobre la percepción auditiva espacial y de este modo mejorar los rendimientos de codificación. La codificación de las señales ambiofónicas sigue siendo, no obstante, posible sin esta etapa de pre-tratamiento.

30 Para las señales no-resultantes de las técnicas ambiofónicas, esta etapa no es necesaria. Para estas señales, el conocimiento del sistema de captación o de restitución asociado a la señal permite interpretar directamente las señales como una descomposición en ondas planas del campo acústico.

35 La figura 6 describe a continuación un decodificador y un procedimiento de decodificación en un modo de realización de la invención.

Este decodificador recibe en la entrada el flujo binario  $F_b$  tal y como se construye mediante el codificador descrito con anterioridad así como la señal compuesta  $S_c$ .

40 De la misma forma que para el codificador, todos los tratamientos se llevan a cabo por trama temporal. Para simplificar las anotaciones, la siguiente descripción del decodificador únicamente describe el tratamiento que se realiza en una trama temporal fija y no muestra la dependencia temporal en las anotaciones. En el codificador, este tratamiento se aplica, sin embargo, de forma sucesiva a todas las tramas temporales de la señal.

45 El codificador así descrito comprende un módulo 650 de decodificación Decod.Fb de las informaciones contenidas en el flujo binario  $F_b$  recibido.

Las informaciones de direcciones y, de manera más particular en este caso, de directividades se extraen, por lo tanto, del flujo binario.

50 Las salidas posibles de este módulo de decodificación del flujo binario dependen de los métodos de codificación de las directividades utilizadas en la codificación. Esta pueden ser en forma de vectores de directividades de base  $D_B$  y de coeficientes asociados  $G_D$  y/o de los parámetros de modelización P.

55 Estos datos se transmiten entonces a un módulo de reconstrucción de las informaciones de directividades 660 que lleva a cabo la decodificación de las informaciones de directividades mediante unas operaciones inversas a las que se realizan en la codificación.

El número de directividad que hay que reconstruir es igual al número  $n_{tot}$  de fuentes en la banda de frecuencia considerada, asociándose cada fuente a un vector de directividad.

60 En el caso de la representación de las directividades a partir de la directividad de base, la matriz de las directividades  $D_i$  se escribe como la combinación lineal de estas directividades de base. De este modo, se puede escribir  $D_i = G_D D_B$ , en la que  $D_B$  es la matriz de las directividades de base para el conjunto de las bandas y  $G_D$  la matriz de las ganancias asociadas. Esta matriz de ganancia tiene un número de líneas igual al número total de fuentes  $n_{tot}$ , y un número de columnas igual al número de vectores de directividad de base.

65

- 5 En una variante de este modo de realización, las directividades de base se decodifican por grupo de bandas de frecuencia consideradas, con el fin de representar de forma más fiel las directividades. Tal y como se ha explicado para la codificación, se pueden proporcionar, por ejemplo, dos grupos de directividades de base: uno para las bajas frecuencias y uno para las altas frecuencias. A continuación se decodifica un vector de ganancias asociadas a las directividades de base para cada banda.
- 10 Al final se reconstruyen tantas directividades como fuentes. Estas directividades se agrupan en una matriz  $D_i$  en la que las líneas corresponden a los valores de ángulo (tantos valores de ángulo como canales en la señal multicanal que hay que reconstruir), y cada columna corresponde a la directividad de la fuente correspondiente, es decir que la columna  $r$  de  $D_i$  da la directividad de la fuente que está en la columna  $r$  de  $S$ .
- 15 Un módulo 690 de definición de las direcciones principales de las fuentes y de determinación de la matriz  $N$  de mezcla recibe estas informaciones de direcciones o de directividades decodificadas.
- Este módulo calcula, en primer lugar, las direcciones principales realizando, por ejemplo, una media de las directividades recibidas para encontrar las direcciones. En función de estas direcciones, se determina una matriz de mezcla, inversa a la que se ha utilizado para la codificación.
- 20 Al conocer las leyes de "panning" utilizadas para la matriz de mezcla en el codificador, el decodificador es capaz de reconstruir la inversa de la matriz de mezcla con las informaciones de direcciones que corresponden a las direcciones de las fuentes principales.
- 25 La información de directividad se transmite por separado para cada fuente. De este modo, en el flujo binario, se identifican bien las directividades relativas a las fuentes principales y las directividades de las fuentes secundarias.
- Hay que señalar que este decodificador no precisa otras informaciones para calcular esta matriz ya que depende de las informaciones de direcciones que se reciben en el flujo binario.
- 30 El mismo algoritmo que se ha descrito en referencia a la figura 4 se aplica entonces en el módulo 690 para encontrar la matriz de mezcla que se adapta a la restitución prevista para la señal compuesta.
- El número de líneas de la matriz  $N$  corresponde al número de canales de la señal compuesta, y el número de columnas corresponde al número de fuentes principales transmitidas.
- 35 La matriz inversa  $N$  tal y como se ha definido la utiliza a continuación el módulo de desmatrizado 620.
- 40 El decodificador recibe, por lo tanto, en paralelo al flujo binario, la señal compuesta  $S_c$ . Esta se somete a una primera etapa de transformada tiempo-frecuencia  $T$  mediante el módulo de transformada 610 para obtener una señal compuesta por banda de frecuencia  $S_{cf}$ .
- 45 Esta transformada se realiza utilizando, por ejemplo, la transformada de Fourier a corto plazo. Hay que señalar que también se pueden utilizar otras transformadas o bancos de filtros  $y$ , en particular, unos bancos de filtros no uniformes de acuerdo con una escala de percepción (p. ej. *Bark*). Hay que señalar que para evitar discontinuidades en la reconstrucción de la señal a partir de esta transformada, se utiliza un método de solapamiento y suma.
- 50 Para la trama temporal considerada, la etapa de cálculo de la transformada de Fourier a corto plazo consiste en mostrar cada uno de los  $n_f$  canales de la señal compuesta  $S_c$  por medio de una ventana  $w$  con una longitud superior a la trama temporal, a continuación en calcular la transformada de Fourier de la señal mostrada por medio de un algoritmo de cálculo rápido en  $n_{FFT}$  puntos. De este modo se obtiene una matriz  $F$  compleja de tamaño  $n_{FFT} \times n_f$  que contiene los coeficientes de la señal compuesta en el espacio frecuencial.
- 55 De aquí en adelante, el conjunto del tratamiento se lleva a cabo por bandas de frecuencia. Para ello, se divide la matriz de los coeficientes  $F$  en un conjunto de sub-matrices  $F_j$  conteniendo cada una los coeficientes de frecuencia en la  $j$ -ésima banda. Existen diferentes opciones para la división de frecuencias de las bandas. Con el fin de garantizar que se aplica el tratamiento en las señales reales, se seleccionan unas bandas simétricas con respecto a la frecuencia nula en la transformada de Fourier a corto plazo. Además, con el fin de optimizar la eficacia de decodificación, se da preferencia a la elección de bandas de frecuencias que se aproximan a las escalas de frecuencia recomendadas, por ejemplo seleccionando unos anchos de banda constante en las escalas ERB o Bark.
- 60 En aras de la simplificación, la descripción de las etapas de decodificación que lleva a cabo el decodificador se realizará para una banda de frecuencia dada. Se sobreentiende que las etapas se llevan a cabo para cada una de las bandas de frecuencia que hay que tratar.
- 65 Los coeficientes de frecuencia de la transformada de la señal compuesta de la banda de frecuencia considerada se matrizan mediante el módulo 620 por la matriz  $N$  determinada de acuerdo con la etapa de determinación descrita con anterioridad de tal modo que se encuentren las fuentes principales de la escena sonora.

De manera más precisa, la matriz  $S_{\text{princ}}$  de los coeficientes de frecuencia para la banda de frecuencia actual de las  $n_{\text{princ}}$  fuentes principales se obtiene de acuerdo con la relación:

- 5  $S_{\text{princ}} = BN$ , en la que  $N$  tiene un tamaño  $n_f \times n_{\text{princ}}$  y  $B$  es una matriz de tamaño  $n_{\text{bin}} \times n_f$ , en la que  $n_{\text{bin}}$  es el número de componentes (o *bins*) de frecuencia seleccionados en la banda de frecuencia considerada.

10 Las líneas de  $B$  son las componentes de frecuencia en la banda de frecuencia actual, las columnas corresponden a los canales de la señal compuesta. Las líneas de  $S_{\text{princ}}$  son las componentes de frecuencia en la banda de frecuencia actual, y cada columna corresponde a una fuente principal.

15 Cuando la escena es compleja, puede suceder que el número de fuentes que hay que reconstruir en la banda de frecuencia actual para obtener una reconstrucción satisfactoria de la escena sea superior al número de canales de la señal compuesta.

En este caso, se codifican unas fuentes adicionales o secundarias y luego se decodifican a partir del flujo binario para la banda actual mediante el módulo 650 de decodificación del flujo binario.

20 Este módulo de decodificación decodifica entonces, además de las informaciones de directividades, las fuentes secundarias.

La decodificación de las fuentes secundarias se lleva a cabo mediante las operaciones inversas a las que se han realizado en la codificación.

25 Sea cual sea el método de codificación que se ha escogido para las fuentes secundarias, si se han transmitido en el flujo binario para la banda actual unos datos de reconstrucción de las fuentes secundarias, los datos correspondientes se decodifican para reconstruir la matriz  $S_{\text{sec}}$  de los coeficientes de frecuencia en la banda actual de las  $n_{\text{sec}}$  fuentes secundarias. La forma de la matriz  $S_{\text{sec}}$  es similar a la matriz  $S_{\text{princ}}$ , es decir que las líneas son las componentes de frecuencia en la banda de frecuencia actual y cada columna corresponde a una fuente secundaria.

30 De este modo se puede construir la matriz completa  $S$  en 680, los coeficientes de frecuencia del conjunto de las  $n_{\text{tot}} = n_{\text{princ}} + n_{\text{sec}}$  fuentes necesarias para la reconstrucción de la señal multicanal en la banda considerada, obtenida al agrupar las dos matrices  $S_{\text{princ}}$  y  $S_{\text{supp}}$  de acuerdo con la relación  $S = (S_{\text{princ}} S_{\text{supp}})$ .  $S$  es, por lo tanto, una matriz de tamaño  $n_{\text{bin}} \times n_{\text{tot}}$ . También la forma es idéntica a las matrices  $S_{\text{princ}}$  y  $S_{\text{supp}}$ : las líneas son las componentes de frecuencia en la banda de frecuencia actual, cada columna es una fuente, con  $n_{\text{tot}}$  fuentes en total.

40 A partir de la matriz  $S$  de los coeficientes de las fuentes y de la matriz  $D_i$  de las directividades asociadas, se calculan en el módulo de espacialización 630 los coeficientes de frecuencia de la señal multicanal reconstruida en la banda, de acuerdo con la relación:

40  $Y = SD^T$ , en la que  $Y$  es la señal reconstruida en la banda. Las líneas de la matriz  $Y$  son las componentes de frecuencia en la banda de frecuencia actual, y cada columna corresponde a una canal de la señal multicanal que hay que reconstruir.

45 Al reproducir el mismo tratamiento en cada una de las bandas de frecuencia, se reconstruyen las transformadas de Fourier completas de los canales de la señal que hay que reconstruir para la trama temporal actual. Las señales temporales correspondientes se obtienen entonces mediante transformada de Fourier inversa  $T^{-1}$ , por medio de un algoritmo rápido que aplica el módulo de transformada inversa 640.

50 De este modo se obtiene la señal multicanal  $S_m$  en la trama temporal actual. Las diferentes tramas temporales se combinan a continuación mediante el método clásico de solapamiento y suma (u "overlap-add" en inglés) para reconstruir la señal multicanal completa.

55 De manera general, se podrán utilizar unos suavizados temporales o frecuenciales de los parámetros tanto en el análisis como en la síntesis para garantizar unas transiciones suaves en la escena sonora. Se podrá reservar una señalización de cambio brusco de la escena sonora en el flujo binario para evitar los suavizados del decodificador en caso de una detección de un cambio rápido de la composición de la escena sonora. Por otra parte, se pueden utilizar los métodos clásicos de adaptación de la resolución del análisis tiempo-frecuencia (cambio de tamaño de las ventanas de análisis y de síntesis a lo largo del tiempo).

60 De la misma manera que en el codificador, un módulo de cambio de base puede llevar a cabo un pre-tratamiento para obtener una descomposición en ondas planas de las señales, un módulo de cambio de base 670 lleva a cabo la operación inversa  $P^{-1}$  a partir de las señales en ondas planas para encontrar la señal multicanal original.

65 Los codificadores y decodificadores tal y como se ha descrito en referencia a las figuras 2 y 6 se pueden integrar en un equipo multimedia de tipo decodificador de salón, ordenador o incluso equipo de comunicación como un teléfono

móvil o una agenda personal electrónica.

La figura 7a representa un ejemplo de un equipo multimedia de este tipo o dispositivo de codificación que comprende un codificador de acuerdo con la invención. Este dispositivo comprende un procesador PROC que coopera con un bloque de memoria BM que comprende una memoria de almacenamiento y/o de trabajo MEM.

El dispositivo comprende un módulo de entrada adaptado para recibir una señal multicanal que representa una escena sonora, bien mediante una red de comunicación, o bien mediante la lectura de un contenido almacenado en un soporte de almacenamiento. Este equipo multimedia también puede comprender unos medios de captura de esta señal multicanal.

El bloque de memoria BM puede comprender de manera ventajosa un programa informático que comprende unas instrucciones de codificación para la aplicación de las etapas del procedimiento de codificación en el sentido de la invención, cuando estas instrucciones las ejecuta el procesador PROC, y en particular las etapas de descomposición de la señal multicanal en bandas de frecuencia y las siguientes etapas por banda de frecuencia:

- obtención de datos representativos de la dirección de las fuentes sonoras de la escena sonora;
- selección de un conjunto de fuentes sonoras de la escena sonora que constituyen las fuentes principales;
- adaptación de datos representativos de la dirección de las fuentes principales seleccionadas, en función de las características de restitución de la señal multicanal;
- determinación de una matriz de mezcla de las fuentes principales en función de los datos adaptados;
- matizado de las fuentes principales mediante la matriz determinada para obtener una señal compuesta con un número reducido de canales;
- codificación de los datos representativos de la dirección de las fuentes sonoras y formación de un flujo binario que comprende los datos codificados, estando el flujo binario adaptado para transmitirse en paralelo a la señal compuesta.

Tradicionalmente, la descripción de la figura 2 retoma las etapas de un algoritmo de dicho programa informático. El programa informático también se puede almacenar en un soporte de memoria legible por un lector del dispositivo o descargable en el espacio de memoria del equipo.

El dispositivo comprende un módulo de salida adaptado para transmitir un flujo binario Fb y una señal compuesta Sc resultantes de la codificación de la señal multicanal.

De la misma forma, la figura 7b ilustra un ejemplo de equipo multimedia o dispositivo de decodificación que comprende un decodificador de acuerdo con la invención.

Este dispositivo comprende un procesador PROC que coopera con un bloque de memoria BM que comprende una memoria de almacenamiento y/o de trabajo MEM.

El dispositivo comprende un módulo de entrada adaptado para recibir un flujo binario Fb y una señal compuesta Sc procedente, por ejemplo, de una red de comunicación. Estas señales de entrada pueden provenir de una lectura en un soporte de almacenamiento.

El bloque de memoria puede comprender de manera ventajosa un programa informático que comprende unas instrucciones de código para la aplicación de las etapas del procedimiento de decodificación en el sentido de la invención, cuando estas instrucciones las ejecuta el procesador PROC, y en particular las etapas de extracción en el flujo binario y de decodificación de los datos representativos de la dirección de las fuentes sonoras en la escena sonora:

- de adaptación de al menos una parte de los datos de dirección en función de las características de restitución de la señal multicanal;
- de determinación de una matriz de mezcla de la señal compuesta en función de los datos adaptados y de cálculo de la inversa de la matriz de mezcla;
- de desmatizado de la señal compuesta mediante la matriz inversa para obtener un conjunto de fuentes principales;
- de reconstrucción de la señal de audio multicanal mediante la espacialización al menos de las fuentes principales con los datos extraídos decodificados.

Tradicionalmente, la descripción de la figura 6 retoma las etapas de un algoritmo de dicho programa informático. El programa informático también se puede almacenar en un soporte de memoria legible por un lector del dispositivo o descargable en el espacio de memoria del equipo.

- 5 El dispositivo comprende un módulo de salida adaptado para transmitir una señal multicanal decodificada mediante el procedimiento de decodificación que aplica el equipo.

Este equipo multimedia también puede comprender unos medios de restitución de tipo altavoz o unos medios de comunicación adaptado para transmitir esta señal multicanal.

- 10 Por supuesto, un equipo multimedia de este tipo puede comprender a la vez el codificador y el decodificador de acuerdo con la invención. Siendo entonces la señal de entrada la señal multicanal original y la señal de salida, la señal multicanal decodificada.

**REIVINDICACIONES**

1. Procedimiento de codificación de una señal de audio multicanal que representa una escena sonora que comprende una multitud de fuentes sonoras, caracterizado por que comprende una etapa de descomposición (T) de la señal multicanal en bandas de frecuencia y las siguientes etapas por banda de frecuencia:
- 5 - obtención (OBT) de datos representativos de la dirección de las fuentes sonoras de la escena sonora;
  - 10 - selección (Selec) de un conjunto de fuentes sonoras de la escena sonora que constituyen las fuentes principales;
  - 15 - adaptación (DiA\_M) de los datos representativos de la dirección de las fuentes principales seleccionadas, en función de las características de restitución de la señal multicanal, mediante la modificación de la posición de las fuentes para obtener una distancia mínima entre dos fuentes;
  - 20 - determinación (DiA\_M) de una matriz de mezcla de las fuentes principales en función de los datos adaptados;
  - matizado (M) de las fuentes principales mediante la matriz determinada para obtener una señal compuesta con un número reducido de canales;
  - 20 - codificación (Cod.Di) de los datos representativos de la dirección de las fuentes sonoras y formación de un flujo binario que comprende los datos codificados, estando el flujo binario adaptado para transmitirse en paralelo a la señal compuesta.
2. Procedimiento de acuerdo con la reivindicación 1, caracterizado por que los datos representativos de la dirección son informaciones de directividades representativas de la distribución de las fuentes sonoras en la escena sonora.
3. Procedimiento de acuerdo con la reivindicación 2, caracterizado por que la codificación de las informaciones de directividades se lleva a cabo mediante un método de representación paramétrica.
- 30 4. Procedimiento de acuerdo con la reivindicación 2, caracterizado por que la codificación de las informaciones de directividad se lleva a cabo mediante un método de análisis de componentes principales que proporciona unos vectores de directividad de base asociados a unas ganancias que permiten la reconstrucción de las directividades iniciales.
- 35 5. Procedimiento de acuerdo con la reivindicación 2, caracterizado por que la codificación de las informaciones de directividad se lleva a cabo mediante una combinación de un método de análisis de componentes principales y de un método de representación paramétrica.
- 40 6. Procedimiento de acuerdo con la reivindicación 1, caracterizado por que comprende, además, la codificación de fuentes secundarias entre las fuentes no seleccionadas de la escena sonora y de inserción de informaciones de codificación de las fuentes secundarias en el flujo binario.
- 45 7. Procedimiento de decodificación de una señal de audio multicanal que representa una escena sonora que comprende una multitud de fuentes sonoras, a partir de un flujo binario y de una señal compuesta, caracterizado por que comprende las siguientes etapas.
- extracción (Decod. Fb) en el flujo binario y decodificación de datos representativos de la dirección de las fuentes sonoras en la escena sonora;
  - 50 - adaptación (DiA\_N) de al menos una parte de los datos de dirección en función de las características de restitución de la señal multicanal, mediante la modificación de la posición de las fuentes obtenidas por los datos de dirección, para obtener una distancia mínima entre dos fuentes;
  - 55 - determinación (DiA\_N) de una matriz de mezcla de la señal compuesta en función de los datos adaptados y cálculo de la inversa de la matriz de mezcla;
  - desmatizado (N) de la señal compuesta mediante la inversa de la matriz de mezcla para obtener un conjunto de fuentes principales;
  - 60 - reconstrucción (SPAC.) de la señal de audio multicanal mediante la espacialización al menos de las fuentes principales con los datos extraídos decodificados.
8. Procedimiento de decodificación acuerdo con la reivindicación 7, caracterizado por que comprende, además, las siguientes etapas:
- 65 - extracción del flujo binario, de informaciones de codificación de fuentes secundarias codificadas;

- decodificación de las fuentes secundarias a partir de las informaciones de codificación extraídas;

- agrupamiento de las fuentes secundarias con las fuentes principales para la espacialización.

5 9. Codificador de una señal de audio multicanal que representa una escena sonora que comprende una multitud de fuentes sonoras, caracterizado por que comprende:

10 - un módulo (210) de descomposición de la señal multicanal en banda de frecuencia;

- un módulo (220) de obtención de datos representativos de la dirección de las fuentes sonoras de la escena sonora;

15 - un módulo (260) de selección de un conjunto de fuentes sonoras de la escena sonora que constituyen las fuentes principales;

- un módulo (275) de adaptación de los datos representativos de la dirección de las fuentes principales seleccionadas, en función de las características de restitución de la señal multicanal, mediante unos medios de modificación de la posición de las fuentes para obtener una distancia mínima entre dos fuentes;

20 - un módulo (275) de determinación de una matriz de mezcla de las fuentes principales en función de los datos resultantes del módulo de adaptación;

- un módulo (270) de matrizado de las fuentes principales seleccionadas mediante la matriz determinada para obtener una señal compuesta con un número reducido de canales;

25 - un módulo (230) de codificación de los datos representativos de la dirección de las fuentes sonoras; y

- un módulo (250) de formación de un flujo binario que comprende los datos codificados, estando el flujo binario adaptado para transmitirse en paralelo a la señal compuesta.

30 10. Decodificador de una señal de audio multicanal que representa una escena sonora que comprende una multitud de fuentes sonoras, que recibe en la entrada un flujo binario y una señal compuesta, caracterizado por que comprende:

35 - un módulo (650) de extracción y de decodificación de datos representativos de la dirección de las fuentes sonoras en la escena sonora;

- un módulo (690) de adaptación de al menos una parte de los datos de dirección en función de las características de restitución de la señal multicanal, mediante unos medios de modificación de la posición de las fuentes obtenidos mediante los datos de dirección, para obtener una distancia mínima entre dos fuentes;

40 - un módulo (690) de determinación de una matriz de mezcla de la señal compuesta en función de los datos resultantes del módulo de adaptación y de cálculo de la inversa de la matriz de mezcla;

45 - un módulo (620) de desmatrizado de la señal compuesta mediante la inversa de la matriz de mezcla para obtener un conjunto de fuentes principales;

50 - un módulo (630) de reconstrucción de la señal de audio multicanal mediante la espacialización al menos de las fuentes principales con los datos extraídos decodificados.

11. Programa informático que comprende unas instrucciones de código para la aplicación de las etapas de un procedimiento de codificación de acuerdo con una de las reivindicaciones 1 a 6 y/o de un procedimiento de decodificación de acuerdo con una de las reivindicaciones 7 a 8, cuando estas instrucciones las ejecuta un procesador.

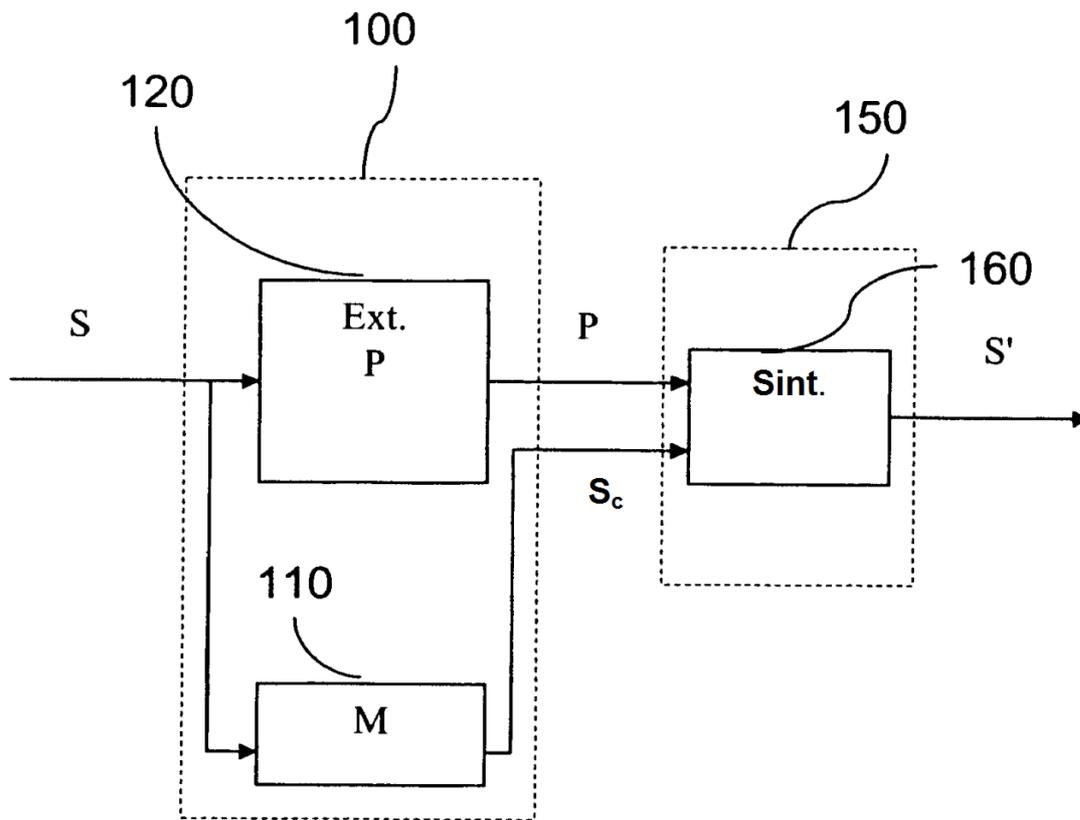


Figura 1 (Estado de la técnica)

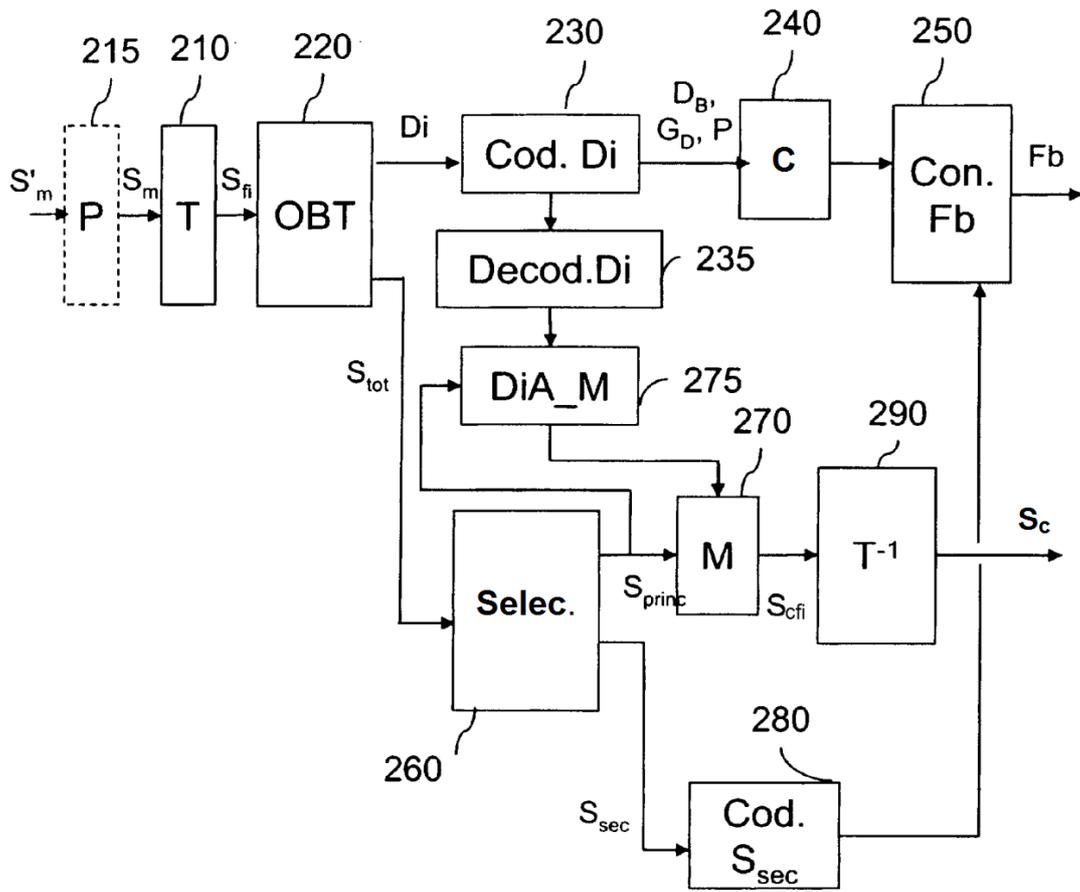


Fig.2

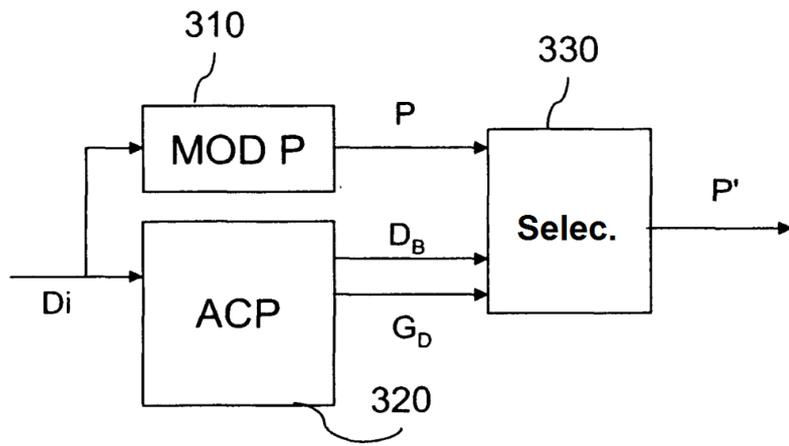


Fig.3a

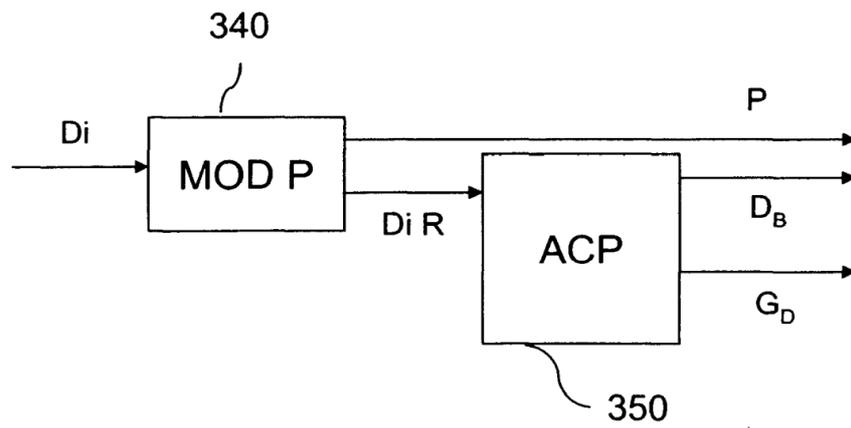


Fig.3b

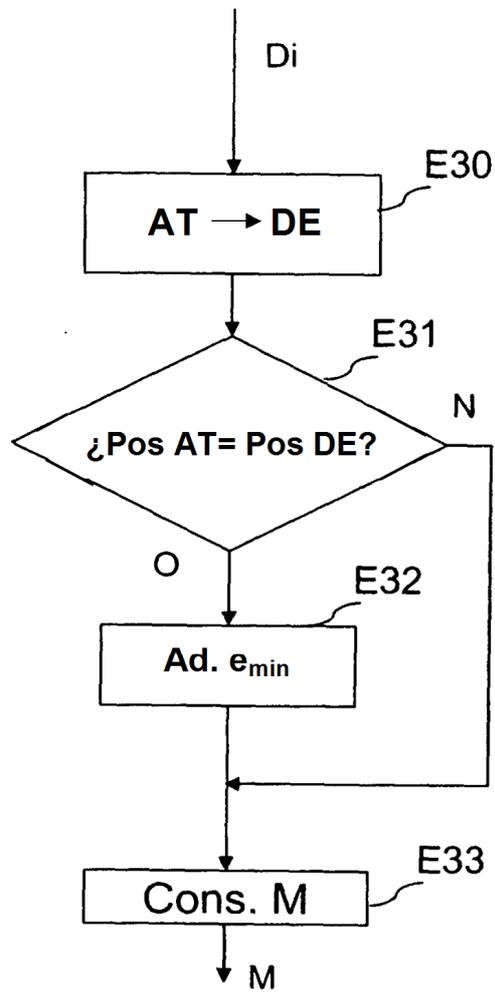


Fig.4

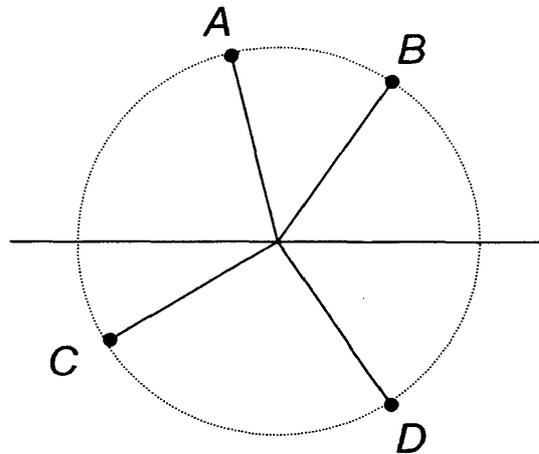


Fig.5a

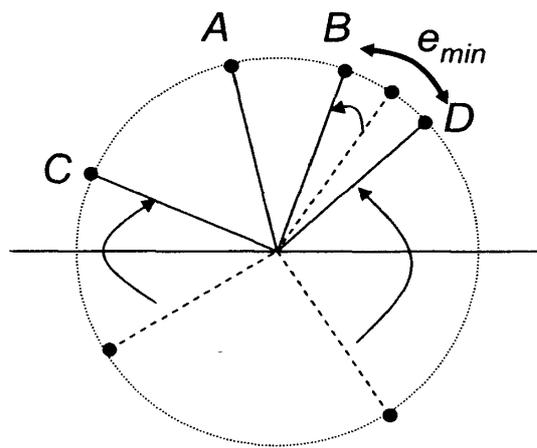


Fig.5b

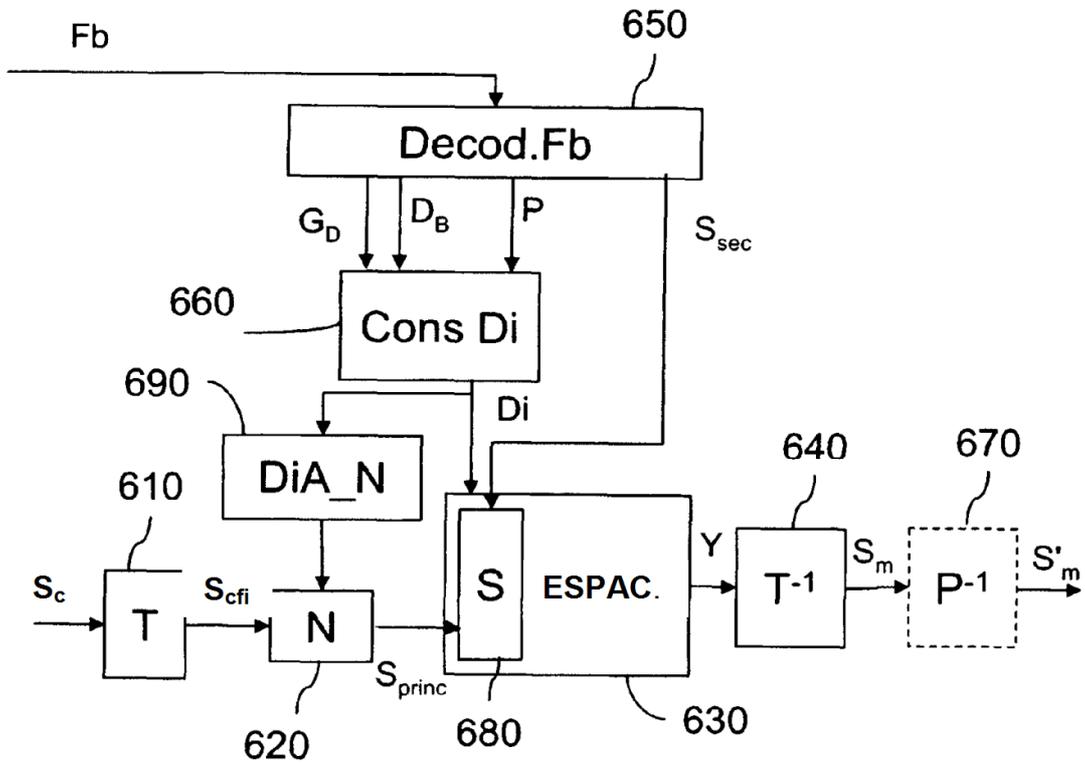


Fig.6

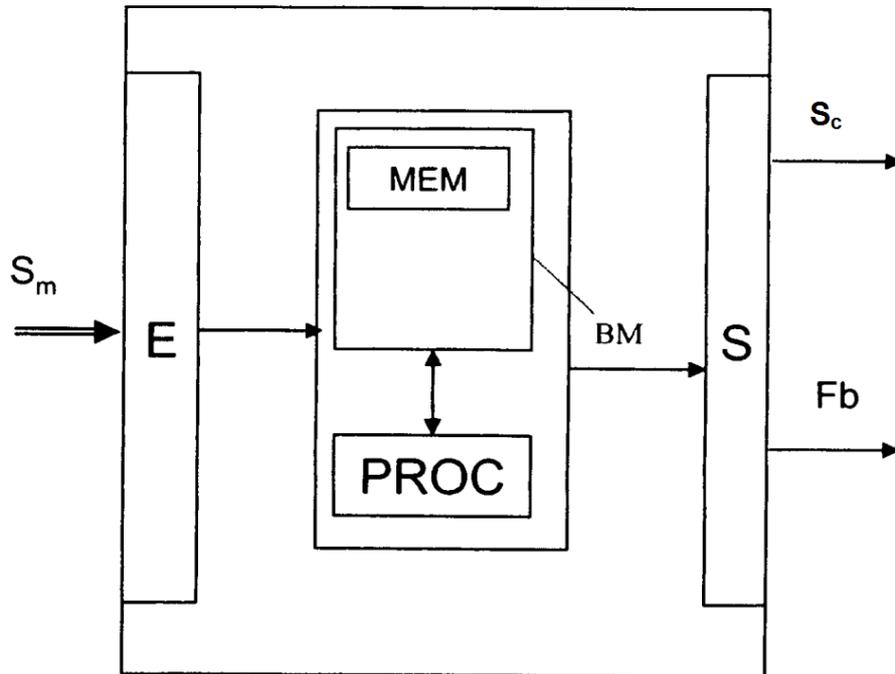


Fig.7a

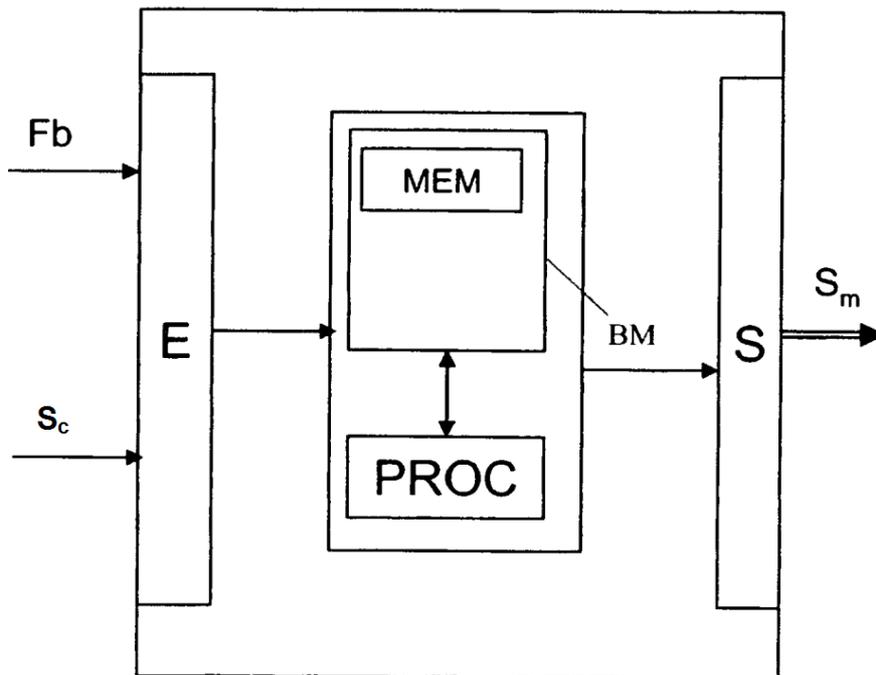


Fig.7b