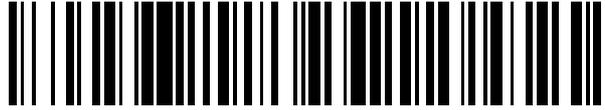


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 472 456**

51 Int. Cl.:

**H04S 3/02** (2006.01)

**G10L 19/00** (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **25.03.2011 E 11709968 (9)**

97 Fecha y número de publicación de la concesión europea: **07.05.2014 EP 2553947**

54 Título: **Método y dispositivo para decodificar una representación de un campo acústico de audio para reproducción de audio**

30 Prioridad:

**26.03.2010 EP 10305316**

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

**01.07.2014**

73 Titular/es:

**THOMSON LICENSING (100.0%)  
1 rue Jeanne d'Arc  
92443 Issy-les-Moulineaux, FR**

72 Inventor/es:

**BATKE, JOHANN-MARKUS;  
KEILER, FLORIAN y  
BOEHM, JOHANNES**

74 Agente/Representante:

**DE ELZABURU MÁRQUEZ, Alberto**

**ES 2 472 456 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

Método y dispositivo para decodificar una representación de un campo acústico de audio para reproducción de audio

**Campo de la invención**

5 Esta invención se refiere a un método y un dispositivo para decodificar una representación de un campo acústico de audio, y en particular una representación de audio formateada Ambisonics, para una reproducción de sonido.

**Antecedentes**

10 Esta sección tiene como objeto introducir al lector en los diversos aspectos de la técnica, que pueden estar relacionados con diversos aspectos de la presente invención que se describen y/o reivindican más adelante. Se cree que la discusión es útil para proveer al lector de una información sobre los antecedentes para facilitar una mejor comprensión de los diversos aspectos de la presente invención. Por consiguiente, se debería comprender que estas exposiciones han de ser leídas en este sentido, y no como supuestos de la técnica anterior, a menos que una fuente sea mencionada expresamente.

15 La localización exacta es un objetivo clave para cualquier sistema de reproducción de sonido. Tales sistemas de reproducción son altamente aplicables a sistemas de conferencias, juegos, u otros entornos virtuales que se benefician del sonido en 3D. Las escenas de sonido en 3D pueden ser sintetizadas o captadas como un campo acústico natural. Las señales del campo acústico tales como por ejemplo Ambisonics llevan una representación de un campo acústico deseado. El formato Ambisonics se basa en la descomposición armónica esférica del campo acústico. En tanto que el formato Ambisonics básico o formato-B utiliza armónicos esféricos de orden cero y uno, el denominado Ambisonics de Orden Superior (HOA) también utiliza armónicos esféricos adicionales de al menos 2º orden. Se requiere un proceso de decodificación para obtener las señales de altavoz individuales. Para sintetizar escenas de audio, se requieren funciones de encuadre que se refieren a la configuración espacial de los altavoces para obtener una localización espacial de la fuente de sonido dada. Si se tuviera que grabar un campo acústico natural se requiere un conjunto de micrófonos para captar la información espacial. El conocido enfoque Ambisonics es una herramienta muy apropiada para llevarlo a cabo. Las señales formateadas Ambisonics llevan una representación del campo acústico deseado. Se requiere un proceso de decodificación para obtener las señales de altavoz individuales a partir de las señales formateadas Ambisonics. Como también en este caso las funciones de encuadre pueden ser obtenidas a partir de las funciones de decodificación, las funciones de encuadre son el tema clave para describir la tarea de la localización espacial. La configuración espacial de los altavoces se denomina aquí disposición de altavoces.

25 Las disposiciones de altavoces comúnmente utilizadas son la disposición estereofónica, la cual emplea dos altavoces, la disposición envolvente normal que usa cinco altavoces, y las ampliaciones de la disposición envolvente que usa más de cinco altavoces. Estas disposiciones son bien conocidas. No obstante, están limitadas a dos dimensiones (2D), por ejemplo no se reproduce información sobre la altura.

35 Las disposiciones de altavoces para reproducción tridimensional (3D) están descritas por ejemplo en "Área de escucha amplia con una calidad de sonido espacial excepcional de un sistema multicanal de sonido 22,2", K. Hamasaki, T. Nishiguchi, R. Okumaura, e Y. Nakayama en Audio Engineering Society Preprints, Viena, Austria, Mayo 2007, que es una propuesta de la TV de ultra alta definición NHK con formato 22.2, o la disposición 2+2+2 de Dabringhaus (mdg – musikproduktion dabringhaus und Grimm, [www.mdg.de](http://www.mdg.de)) y una disposición 10.2 en "Sound for Film and Television", T. Holman en 2ª edición, Boston: Focal Press, 2002. Uno de los pocos sistemas conocidos que se refiere a la reproducción espacial y a las estrategias de encuadre es el enfoque de encuadre de amplitud de base vectorial (VBAP) en "Ubicación virtual de la fuente acústica que usa el encuadre de amplitud de base vectorial", Journal of Audio Engineering Society, volumen 45, nº 6, pp. 456-466, Junio 1997, aquí Pulkki. El VBAP (Encuadre de Amplitud de Base Vectorial (VBAP) ha sido usado por Pulkki para reproducir fuentes acústicas virtuales con una disposición de altavoces arbitraria. Para colocar una fuente virtual en un plano en 2D se requiere un par de altavoces, en tanto que en el caso de 3D se requieren tríos de altavoces. Para cada fuente virtual se suministra una señal monofónica con ganancias diferentes (dependientes de la posición de la fuente virtual) a los altavoces seleccionados desde la disposición total. A continuación se suman las señales de altavoz de todas las fuentes virtuales. El VBAP aplica un método geométrico para calcular las ganancias de las señales de los altavoces para el encuadre entre los altavoces.

45 Una disposición de altavoces en 3D a modo de ejemplo considerada y aquí recientemente propuesta tiene 16 altavoces, los cuales están colocados como se muestra en la Figura 2. La colocación se eligió debido a consideraciones prácticas, que tienen cuatro columnas con tres altavoces cada una y unos altavoces adicionales entre estas columnas. Con más detalle, ocho de los altavoces, que abarcan unos ángulos de 45 grados, están igualmente distribuidos en un círculo alrededor de la cabeza de la persona que escucha. Cuatro altavoces adicionales, que abarcan unos ángulos de 90 grados, están situados en la parte de arriba y en el fondo. Con respecto a Ambisonics, esta disposición es irregular y da lugar a problemas en el diseño del decodificador, como se ha mencionado en "Un formato ambisonics para distribuciones flexibles de la reproducción", por H. Pomberger y F. Zotter en Proceedings of the 1<sup>st</sup> Ambisonics Symposium, Graz, Austria, Julio 2009.

La decodificación Ambisonics convencional, como está descrita en el documento EP 2.094.032 y en "Sistemas acústicos envolventes tridimensionales basados en armónicos esféricos" por M. Poletti en J. Audio Eng. Soc., volumen 53, nº 11, pp 1.004-1.025, Noviembre 2005, emplea el comúnmente conocido proceso de adaptación de modos. Los modos son descritos por vectores de modos que contienen valores de los armónicos esféricos para una dirección de incidencia distinta. La combinación de todas las direcciones dadas por los altavoces individuales lleva a la matriz de modos de la disposición de altavoces, de modo que la matriz de modos representa las posiciones de los altavoces. Para reproducir el modo de una señal de fuente distinta se ponderan los modos de los altavoces de tal forma que los modos superpuestos de los altavoces individuales alcancen el modo deseado. Para obtener las ponderaciones necesarias es preciso calcular una representación de la matriz inversa de la matriz de modos de los altavoces. Desde el punto de vista de la decodificación de la señal las ponderaciones forman la señal piloto de los altavoces, y la matriz de modos inversa de los altavoces, a la que se hace referencia como "matriz de decodificación", la cual se aplica para decodificar una representación de la señal formateada Ambisonics. En particular, para muchas disposiciones de altavoces, por ejemplo la disposición mostrada en la Figura 2, es difícil de obtener la inversa de la matriz de modos.

Como se ha mencionado antes, las disposiciones comúnmente usadas están limitadas a 2D, es decir no se reproduce información sobre la altura. La decodificación de una representación de campo acústico en una disposición de altavoces con una distribución espacial matemáticamente no regular lleva a problemas de localización y coloración con las técnicas comúnmente conocidas. Para decodificar una señal Ambisonics se usa una matriz de decodificación (es decir, una matriz de coeficientes de decodificación). En la decodificación convencional de las señales Ambisonics, y particularmente de las señales HOA, se presentan por lo menos dos problemas. Primero, para la correcta decodificación es necesario conocer las direcciones de la fuente de la señal para obtener la matriz de decodificación. Segundo, el establecimiento de una correspondencia con una disposición de altavoces existente es sistemáticamente erróneo debido al siguiente problema matemático: una decodificación matemáticamente correcta dará lugar a no sólo una positiva, sino a también algunas amplitudes negativas de los altavoces. No obstante, éstas son reproducidas erróneamente como señales positivas, lo que da lugar a los problemas antes mencionados.

#### Compendio de la invención

La presente invención describe un método para decodificar una representación de un campo acústico para distribuciones espaciales no regulares con unas propiedades de localización y coloración altamente mejoradas. Representa otra forma de obtener la matriz de decodificación para datos del campo acústico, por ejemplo en formato Ambisonics, y utiliza un proceso en una forma de estimación del sistema. Considerando un conjunto de posibles direcciones de incidencia se calculan las funciones de encuadre relacionadas con los altavoces deseados. Las funciones de encuadre se toman como el resultado de un proceso de decodificación Ambisonics. La señal de entrada requerida es la matriz de modos de todas las direcciones consideradas. Por lo tanto, como se muestra más adelante, la matriz de decodificación se obtiene por multiplicación directa de la matriz de ponderación por una versión inversa de la matriz de modos de las señales de entrada.

Con relación al segundo problema antes mencionado, se ha visto que también es posible obtener la matriz de decodificación a partir de la inversa de la denominada matriz de modos, la cual representa las posiciones de los altavoces, y las funciones  $W$  de ponderación dependientes de la posición ("funciones de encuadre"). Un aspecto de la invención es que estas funciones de encuadre  $W$  pueden obtenerse mediante un método diferente del comúnmente usado. Ventajosamente, se usa un método geométrico sencillo. Tal método no requiere conocimiento alguno de ninguna dirección de la fuente de la señal, lo que resuelve el problema antes mencionado. Tal método es conocido como "Encuadre de Amplitud de Base Vectorial" (VBAP). De acuerdo con la invención, el VBAP se usa para calcular las funciones de encuadre requeridas, las cuales son a continuación usadas para calcular la matriz de decodificación Ambisonics. Otro problema se presenta debido a que se requiere la inversa de la matriz de modos (que representa la disposición de altavoces). Sin embargo, es difícil de obtener la inversa exacta, lo que también lleva a una reproducción de audio errónea. De este modo, un aspecto adicional consiste en que para obtener la matriz de decodificación se calcula una matriz de modos seudoinvertida, la cual es mucho más fácil de obtener.

La invención usa un enfoque de dos pasos. El primer paso es una obtención de las funciones de encuadre que dependen de la disposición de altavoces usada para la reproducción. En el segundo paso se calcula una matriz de decodificación Ambisonics a partir de estas funciones de encuadre para todos los altavoces.

Una ventaja de la invención es que no se requiere una descripción paramétrica de la fuente acústica; en vez de ello, se puede usar una descripción del campo acústico tal como Ambisonics.

De acuerdo con la invención, un método para decodificar la representación de un campo acústico de audio para una reproducción de audio comprende los pasos de calcular, para cada uno de una pluralidad de altavoces, una función de encuadre mediante un método geométrico basado en las posiciones de los altavoces y una pluralidad de direcciones de la fuente mediante el cálculo de una matriz de modos a partir de las direcciones de la fuente, mediante el cálculo de una matriz de modos seudoinvertida de la matriz de modos, y mediante la decodificación de una representación del campo acústico de audio, en donde la decodificación está basada en una matriz de decodificación que se obtiene a partir de al menos la función de encuadre y de la matriz de modos seudoinvertida.

De acuerdo con otro aspecto, un dispositivo para decodificar una representación de audio de un campo acústico para reproducción de audio comprende unos primeros medios de cálculo para calcular, para cada uno de una pluralidad de altavoces, una función de encuadre mediante un método geométrico basado en las posiciones de los altavoces y en una pluralidad de direcciones de la fuente, unos segundos medios de cálculo para calcular una matriz de modos a partir de las direcciones de la fuente, unos terceros medios de cálculo para calcular una matriz de modos pseudoinversa de la matriz de modos, y unos medios de decodificador para decodificar la representación del campo acústico, en donde la decodificación está basada en una matriz de decodificación y los medios de decodificador utilizan al menos la función de encuadre y la segunda matriz de modos pseudoinversa para obtener la matriz de decodificación. Los primero, segundo y tercer medios de cálculo pueden ser un único procesador o dos o más procesadores independientes.

De acuerdo con otro aspecto más, un medio de ordenador, que puede ser leído tiene almacenadas en él unas instrucciones ejecutables para hacer que un ordenador ejecute un método para decodificar una representación de campo acústico de audio para reproducción de audio comprende los pasos de calcular, para cada uno de una pluralidad de altavoces, una función de encuadre mediante un método geométrico basado en las posiciones de los altavoces y en una pluralidad de direcciones de la fuente, calcular una matriz de modos a partir de las direcciones de la fuente, calcular la pseudoinversa de la matriz de modos, y decodificar la representación del campo acústico de audio, en donde la decodificación está basada en una matriz de decodificación que se obtiene a partir de al menos la función de encuadre y la matriz de modos pseudoinversa.

Las realizaciones ventajosas de la invención se describen en las reivindicaciones dependientes, en la siguiente descripción y en las figuras.

#### Breve descripción de los dibujos

Las realizaciones a modo de ejemplo de la invención se describen con referencia a los dibujos que se acompañan, las cuales se muestran en

la Figura 1 es un diagrama de flujos del método;

la Figura 2 es una disposición a modo de ejemplo en 3D con 16 altavoces;

la Figura 3 es un patrón de haz que resulta de la decodificación mediante una adaptación de modos no regularizados;

la Figura 4 es un patrón de haz que resulta de la decodificación mediante una matriz de modos regularizados;

la Figura 5 es un patrón de haz que resulta de la decodificación mediante una matriz obtenida a partir de VBAP;

la Figura 6 resulta de una prueba de escucha; y

la Figura 7 es un diagrama de bloques de un dispositivo.

#### Descripción detallada de la invención

Como se muestra en la Figura 1, un método para decodificar una representación SF<sub>c</sub> del campo acústico de audio para reproducción de audio comprende los pasos de calcular 110, para cada uno de una pluralidad de altavoces, una función de encuadre W mediante un método geométrico basado en las posiciones 102 de los altavoces (L es el número de altavoces) y una pluralidad de direcciones 103 de la fuente (S es el número de direcciones de la fuente), calcular 120 una matriz de modos  $\Xi$  a partir de las direcciones de la fuente y de un orden dado N de la representación de un campo acústico, calcular 130 una matriz de modos pseudoinversa  $\Xi^+$  de la matriz de modos  $\Xi$ , y decodificar 135, 140 la representación SF<sub>c</sub> del campo acústico de audio, en donde se obtienen los datos de sonido decodificados AU<sub>dec</sub>. La decodificación está basada en una matriz D de decodificación que se obtiene 135 a partir de al menos la función W de encuadre y la matriz de modos pseudoinversa  $\Xi^+$ . En una realización la matriz de modos pseudoinversa se obtiene de acuerdo con  $\Xi^+ = \Xi^H [\Xi \Xi^H]^{-1}$ . El orden N de la representación de un campo acústico puede ser predefinido, o puede ser extraído 105 a partir de la señal de entrada SF<sub>c</sub>.

Como se ha mostrado en la Figura 7, un dispositivo para decodificar una representación de audio de un campo acústico para una reproducción de audio comprende unos primeros medios de cálculo 210 para calcular, para cada uno de una pluralidad de altavoces, una función W de encuadre mediante un método geométrico basado en las posiciones 102 de los altavoces y en una pluralidad de direcciones 103 de la fuente, unos segundos medios de cálculo 220 para calcular una matriz  $\Xi$  de modos a partir de las direcciones de la fuente, unos terceros medios de cálculo 230 para calcular una matriz  $\Xi^+$  de modos pseudoinversa de la matriz de modos  $\Xi$ , y unos medios de decodificador 240 para decodificar la representación de un campo acústico. La decodificación está basada en una matriz D de modos pseudoinversa, que se obtiene a partir de al menos la función de encuadre W y de la matriz  $\Xi^+$  de modos pseudoinversa por unos medios de cálculo 235 de matriz de decodificación (por ejemplo un multiplicador). Los medios de decodificador 240 usan la matriz D de decodificación para obtener una señal de audio decodificada AU<sub>dec</sub>. Los medios de cálculo primero, segundo y tercero 220, 230, 240 pueden ser un único procesador, o dos o más

procesadores independientes. El orden N de la representación de un campo acústico puede ser predefinido, o puede ser obtenido por un medio 205 para extraer el orden a partir de la señal de entrada SF.

Una disposición de altavoces en 3D particularmente útil tiene 16 altavoces. Como se ha mostrado en la Figura 2, hay cuatro columnas con tres altavoces cada una, y unos altavoces adicionales entre estas columnas. Ocho altavoces están igualmente distribuidos en un círculo alrededor de la cabeza del oyente, que abarcan unos ángulos de 45 grados. Cuatro altavoces adicionales están situados en la parte de arriba y en el fondo, que abarcan unos ángulos de azimut de 90 grados. Con respecto a Ambisonics, esta disposición es irregular y usualmente da lugar a problemas en el diseño del decodificador.

En lo que sigue se describe con detalle el método de encuadre de amplitud de base vectorial (VBAP). En una realización, el VBAP se usa aquí para colocar unas fuentes acústicas virtuales con una disposición de altavoces arbitraria en la que se supone la misma distancia de los altavoces desde la posición de escucha. El VBAP usa tres altavoces para colocar una fuente virtual en el espacio de 3D. Para cada fuente virtual se alimenta a los altavoces que van a ser usados una señal monofónica con ganancias diferentes. Las ganancias de los diferentes altavoces dependen de la posición de la fuente virtual. El VBAP es un enfoque geométrico para calcular las ganancias de las señales de los altavoces para el encuadre entre los altavoces. En el caso de 3D tres altavoces configurados en un triángulo crean una base vectorial. Cada base vectorial está identificada por los números del altavoz k, m, n, y los vectores  $l_k, l_m, l_n$ , de las posiciones de los altavoces están dados en coordenadas cartesianas normalizadas con la longitud unidad. La base vectorial de los altavoces k, m, n está definida por

$$L_{kmn} = \{l_k, l_m, l_n\} \quad (1)$$

La dirección deseada  $\Omega = (\theta, \phi)$  de la fuente virtual tiene que ser dada como un ángulo de azimut  $\theta$  y un ángulo de inclinación  $\phi$ . El vector de posición de longitud unidad  $p(\Omega)$  de la fuente virtual en coordenadas cartesianas está por lo tanto definido por

$$p(\Omega) = \{\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta\}^T \quad (2)$$

Una posición virtual de la fuente puede ser representada con la base vectorial y los factores de ganancia

$$g(\Omega) = (\tilde{g}_k, \tilde{g}_m, \tilde{g}_n)^T \text{ por } p(\Omega) = L_{kmn} g(\Omega) = \tilde{g}_k l_k, \tilde{g}_m l_m, \tilde{g}_n l_n \quad (3)$$

Invirtiendo la matriz de la base vectorial los factores de ganancia requeridos pueden ser calculados por

$$g(\Omega) = L_{mn}^{-1} p(\Omega) \quad (4)$$

La base vectorial que se ha de usar se determina de acuerdo con el documento de Pulkki: Primero se calculan las ganancias de acuerdo con Pulkki para todas las bases vectoriales. A continuación para cada base vectorial se evalúa el mínimo sobre los factores de ganancia por  $\tilde{g}_{\min} = \min\{\tilde{g}_k, \tilde{g}_m, \tilde{g}_n\}$ . Finalmente se usa la base vectorial en la que  $\tilde{g}_{\min}$  tiene el valor más alto. Los factores de ganancia resultantes no deben ser negativos. Dependiendo de las condiciones acústicas del salón de escucha los factores de ganancia pueden ser normalizados para la conservación de la energía.

En lo que sigue se describe el formato Ambisonics, que es un formato de un campo acústico a título de ejemplo. La representación Ambisonics es un método de descripción de un campo acústico que emplea una aproximación matemática del campo acústico en un lugar determinado. Mediante el uso de un sistema de coordenadas esféricas, la presión en el punto  $r = (r, \theta, \phi)$  es descrita por medio de la transformación de Fourier esférica:

$$p(r, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n A_n^m(k) j_n(kr) Y_n^m(\vartheta, \varphi) \quad (5)$$

en donde k es el número de ondas. Normalmente n va hasta un orden finito M. Los coeficientes  $A_n^m(k)$  de la serie describen el campo acústico (que supone las fuentes fuera de la región de validez),  $j_n(kr)$  es la función de Bessel esférica de primera clase e  $Y_n^m(\theta, \phi)$  representa los armónicos esféricos. Los coeficientes  $A_n^m(k)$  son considerados como coeficientes Ambisonics en este contexto. Los armónicos esféricos  $Y_n^m(\theta, \phi)$  sólo dependen de los ángulos de inclinación y de azimut y describen una función sobre la esfera unidad.

Por motivos de simplicidad a menudo se suponen ondas planas para la reproducción de un campo acústico. Los coeficientes Ambisonics que describen una onda plana como una fuente acústica de dirección  $\Omega_s$  son

$$A_n^m, \text{ plane}(\Omega_s) = 4\pi i^n Y_n^m(\Omega_s)^* \quad (6)$$

Su dependencia del número k de ondas disminuye hasta una mera dependencia direccional en este caso especial. Para un orden limitado M los coeficientes forman un vector A que puede ser dispuesto como

$$A(\Omega_s) = [A_0^0 \ A_1^{-1} \ A_1^0 \ A_1^1] \dots [A_M^M]^T \quad (7)$$

que mantiene  $O=(M+1)^2$  elementos. La misma disposición se usa para los coeficientes armónicos esféricos que producen un vector

$$Y(\Omega_s)^* = [Y_0^0 \ Y_1^{-1} \ Y_1^0 \ Y_1^1 \dots A_M^M]^H$$

5 en donde el superíndice H representa la traspuesta conjugada compleja.

Para calcular las señales de los altavoces a partir de una representación Ambisonics de un campo acústico la adaptación de modos es un enfoque usado comúnmente. La idea básica es expresar una descripción del campo acústico  $A(\Omega_s)$  mediante una suma ponderada de las descripciones  $A(\Omega_l)$  de los campos acústicos de los altavoces.

$$A(\Omega_s) = \sum_{l=1}^L W_l A(\Omega_l) \quad (8)$$

10 en donde  $\Omega_l$  representa las direcciones de los altavoces, W son ponderaciones, y L es el número de altavoces. Para obtener las funciones de encuadre a partir de la ecuación (8), suponemos una dirección conocida de incidencia  $\Omega_s$ . Si los campos acústicos de la fuente y del que habla son ambos ondas planas, el factor  $4\pi^n$  (véase la ecuación (6)) puede omitirse, y la ecuación (8) sólo depende de los conjugados complejos de los vectores armónicos esféricos, a los que también se hace referencia como "modos". Mediante la notación matricial, esto se escribe como

$$15 \quad Y(\Omega_s)^* = \Psi W(\Omega_s) \quad (9)$$

$$\Psi = [Y(\Omega_1)^* \ Y(\Omega_2)^* \dots Y(\Omega_L)^*] \quad (10)$$

con  $O \times L$  elementos. Para obtener el vector de ponderación  $w$  deseado se conocen diversas estrategias para llevarlo a cabo. Si se elige  $M=3$ ,  $\Psi$  es cuadrada y puede ser invertible. A pesar de todo, debido a la disposición irregular de los altavoces la matriz está mal graduada. En tal caso, a menudo se elige la matriz seudoinvertida, y

$$20 \quad D = [\Psi^H \ \Psi]^{-1} \Psi^H \quad (11)$$

produce una matriz de decodificación  $D$   $L \times O$ . Finalmente podemos escribir

$$w(\Omega_s) = D Y(\Omega_s)^* \quad (12)$$

en donde las ponderaciones ( $\Omega_s$ ) son la solución de energía mínima para la ecuación (9). A continuación se describen las consecuencias de usar la seudoinvertida.

25 Lo que sigue describe el enlace entre las funciones de encuadre y la matriz de decodificación Ambisonics. En el comienzo con Ambisonics las funciones de encuadre de los altavoces individuales se pueden calcular mediante la ecuación (12). Supongamos que

$$\Xi = [Y(\Omega_1)^* \ Y(\Omega_2)^* \dots Y(\Omega_s)^*] \quad (13)$$

30 es la matriz de modos de las direcciones ( $\Omega_s$ ) de la señal de entrada S, por ejemplo una rejilla esférica con un ángulo de inclinación que va por pasos de un grado de 1 a  $180^\circ$  y un ángulo de azimut de 1 a  $360^\circ$  respectivamente. Esta matriz de modos tiene  $O \times S$  elementos. Mediante la ecuación (12), la matriz W que resulta tiene  $L \times S$  elementos, la fila l mantiene las ponderaciones de encuadre S para el altavoz respectivo:

$$W = D \Xi \quad (14)$$

35 Como ejemplo representativo, en la Figura 3 se muestra como patrón de haz la función de encuadre de un único altavoz 2. En este ejemplo la matriz de decodificación D es de orden  $M=3$ . Como se puede ver, los valores de la función de encuadre no se refieren en modo alguno a la ubicación física del altavoz. Esto se debe a la colocación matemática irregular de los altavoces, que no es suficiente como un esquema de muestreo espacial para el orden escogido. Por lo tanto, se hace referencia a la matriz de decodificación como a una matriz de modos no regularizada. Este problema puede ser superado mediante la regularización de la matriz de modos  $\Psi$  de los altavoces en la ecuación (11). Esta solución funciona a costa de la resolución espacial de la matriz de decodificación, que a su vez puede ser expresada como un orden Ambisonics inferior. La Figura 4 muestra a modo de ejemplo el patrón de haz que resulta de la decodificación mediante una matriz de modos regularizada, y particularmente mediante la media de los valores propios de la matriz de modos para la regularización. En comparación con la Figura 3, se puede reconocer ahora claramente la dirección del altavoz considerado.

45

Como se ha esbozado en la introducción, es posible otro modo para obtener una matriz de decodificación D para la reproducción de las señales Ambisonics cuando ya son conocidas las funciones de encuadre. Las funciones de encuadre W son consideradas como la señal deseada definida en un conjunto de direcciones  $\Omega$  de la fuente virtual, y la matriz de modos  $\Xi$  de estas direcciones sirve como señal de entrada. Entonces, la matriz de decodificación puede ser calculada mediante

$$D = W \Xi [\Xi \Xi^H]^{-1} = W \Xi^+ \quad (15)$$

en donde  $\Xi^+ [\Xi \Xi^H]^{-1}$ , o simplemente  $\Xi^+$ , es la pseudoinversa de la matriz de modos  $\Xi$ . En el nuevo método tomamos las funciones de encuadre en W a partir del VBAP y calculamos una matriz de decodificación Ambisonics a partir de éste.

Las funciones de encuadre de W se toman como valores de ganancia  $g(\Omega)$  calculados mediante la ecuación (4), en la que  $\Omega$  se elige de acuerdo con la ecuación (13). La matriz de decodificación resultante mediante la ecuación (15) es una matriz de decodificación Ambisonics que facilita las funciones de encuadre del VBAP. En la Figura 5 está representado un ejemplo que muestra un patrón de haz que resulta de la decodificación mediante una matriz de decodificación obtenida a partir del VBAP. Ventajosamente, los lóbulos laterales SL son significativamente menores que los lóbulos laterales  $SL_{reg}$  del resultado de la adaptación de modos regularizados de la Figura 4. Por otra parte, el patrón de haz obtenido del VBAP para cada uno de los altavoces individuales sigue la geometría de la disposición de altavoces cuando las funciones de encuadre del VBAP dependen de la base vectorial de la dirección considerada. Como consecuencia, el nuevo método de acuerdo con la invención produce mejores resultados en todas las direcciones de la disposición de altavoces.

Las direcciones 103 de la fuente pueden ser definidas más libremente. Una condición del número de direcciones de la fuente S es que debe ser al menos  $(N+1)^2$ . De este modo, con un orden N dado de la señal  $SF_c$  del campo acústico es posible definir S de acuerdo con  $S \geq (N+1)^2$ , y distribuir las direcciones de la fuente S uniformemente en una esfera unidad. Como se ha mencionado antes, el resultado puede ser una rejilla esférica con un ángulo de inclinación  $\theta$  que va por pasos constantes de x (por ejemplo,  $x = 1$  a 5 ó  $x = 10, 20$ , etc) grados de 1 a  $180^\circ$ , y un ángulo de azimut  $\phi$  de 1 a  $360^\circ$  respectivamente, en donde cada dirección  $\Omega = (\theta, \phi)$  de la fuente puede ser dada por el ángulo de azimut  $\phi$  y el ángulo de inclinación  $\theta$ .

El efecto ventajoso ha sido confirmado en una prueba de escucha. Para la evaluación de la localización de una única fuente, se compara una fuente virtual con una fuente real de referencia. Para la fuente real se usa un altavoz en la posición deseada. Los métodos de reproducción usados son el VBAP, la decodificación de adaptación de modos Ambisonics, y la decodificación Ambisonics propuesta, y la decodificación Ambisonics recientemente propuesta que usa las funciones de encuadre del VBAP de acuerdo con la presente invención. En los dos últimos métodos, para cada posición ensayada y cada señal de entrada ensayada se genera una señal Ambisonics de tercer orden. Esta señal Ambisonics sintética es después decodificada mediante las correspondientes matrices de decodificación. Las señales de prueba usadas son un ruido rosado de banda ancha y una señal de conversación masculina. Las posiciones ensayadas se colocan en la región frontal con las direcciones

$$\Omega_1 = (76,1^\circ, -23,2^\circ), \quad \Omega_2 = (63,3^\circ, -4,3^\circ) \quad (16)$$

La prueba de escucha fue realizada en una habitación acústica con un tiempo de reverberación medio de aproximadamente 0,2 s. Nueve personas participaron en la prueba de escucha. Se preguntó a los miembros de la prueba que evaluaran la realización de la reproducción espacial de todos los métodos de reproducción comparados con la referencia. Se tenía que encontrar un único valor de evaluación para representar la localización de la fuente virtual y las alteraciones de timbre. La Figura 5 muestra una lista de los resultados de la prueba.

Como muestran los resultados, la decodificación de adaptación de modos Ambisonics no regularizados está evaluada perceptualmente peor que los otros métodos ensayados. Este resultado corresponde al de la Figura 3. El método de adaptación de modos Ambisonics sirve como referencia en esta prueba de escucha. Otra ventaja es que los intervalos de confianza para la señal de ruido son mayores para el VBAP que para los otros métodos. Los valores medios muestran los valores más altos para la decodificación Ambisonics mediante las funciones de encuadre del VBAP. De este modo, aunque la resolución espacial es reducida – debido al orden de Ambisonics usado - este método muestra ventajas sobre el enfoque del VBAP paramétrico. En comparación con el VBAP, las decodificaciones Ambisonics que decodifican con funciones de encuadre robustas y el VBAP tienen la ventaja de que no se usan sólo tres altavoces para representar la fuente virtual. En el VBAP los altavoces únicos pueden ser dominantes si la posición de la fuente virtual está cerca de una de las posiciones físicas de los altavoces. La mayor parte de los participantes informó de menos alteraciones del timbre en el VBAP impulsado por Ambisonics que en el VBAP aplicado directamente. El problema de las alteraciones de timbre en el VBAP es ya conocido a partir de Pulkki. Al contrario que el VBAP, el método recientemente propuesto usa más de tres altavoces para la reproducción de una fuente virtual, pero sorprendentemente produce una menor coloración.

Como conclusión, se describe un nuevo modo de obtener una matriz de decodificación Ambisonics a partir de las funciones de encuadre del VBAP. Para disposiciones diferentes de altavoces este método es ventajoso en comparación con las matrices del enfoque de adaptación de modos. Las propiedades y consecuencias de estas

5 matrices de decodificación se han discutido anteriormente. En resumen, la decodificación Ambisonics recientemente propuesta con las funciones de encuadre del VBAP evita los típicos problemas del bien conocido enfoque de adaptación de modos bien conocido. Una prueba de escucha ha mostrado que la decodificación Ambisonics obtenida a partir del VBAP puede producir una calidad de reproducción espacial mejor que la que puede producir el uso directo del VBAP. El método propuesto requiere sólo una descripción del campo acústico en tanto que el VBAP requiere una descripción paramétrica de las fuentes virtuales que han de ser presentadas.

10 En tanto que se han mostrado, descrito, y señalado las características novedosas fundamentales de la presente invención como aplicadas a las realizaciones preferidas de ella, se comprenderá que los expertos en la técnica pueden realizar diversas omisiones y sustituciones y cambios en el aparato y método descritos, en la forma y detalles de los dispositivos descritos, y en su funcionamiento, sin apartarse del espíritu de la presente invención. Se pretende expresamente que todas las combinaciones de los elementos que llevan a cabo sustancialmente la misma función en sustancialmente la misma forma para conseguir los mismos resultados estén dentro del alcance de la invención. Las sustituciones de los elementos de una realización descrita a otra son también totalmente intencionadas y contempladas. Se comprenderá que se pueden realizar modificaciones de detalles sin apartarse del alcance de la invención. Cada característica descrita en la descripción y (en donde proceda) las reivindicaciones y dibujos pueden ser proporcionados independientemente o en cualquier combinación apropiada. Donde proceda, las características pueden ser puestas en práctica en el soporte físico, en el soporte lógico, o en una combinación de los dos. Los números de referencia que aparecen en las reivindicaciones son solamente a modo ilustrativo y no tendrán efecto limitativo alguno sobre el alcance de las reivindicaciones.

20

**REIVINDICACIONES**

1. Un método de acuerdo con una representación de un campo acústico de audio para reproducción de audio, que comprende los pasos de:
  - 5 - calcular (110) para cada uno de una pluralidad de altavoces una función de encuadre ( $W$ ) que usa un método geométrico basado en las posiciones de los altavoces y en una pluralidad de direcciones de la fuente;
  - calcular (120) una matriz de modos ( $\Xi_N$ ) a partir de las direcciones de la fuente;
  - calcular (130) una matriz de modos pseudoinversa ( $\Xi^+$ ) de la matriz de modos ( $\Xi$ ); y
  - decodificar (140) la representación de un campo acústico de audio, en donde la decodificación está basada en una matriz de decodificación ( $D$ ) que se obtiene a partir de la función de encuadre ( $W$ ) y la matriz de modos pseudoinversa ( $\Xi^+$ ).
- 10 2. Método de acuerdo con la reivindicación 1, en donde el método geométrico usado en el paso de cálculo de una función de encuadre es el Encuadre de Amplitud de la Base Vectorial (VBAP).
3. Método de acuerdo con la reivindicación 1 ó 2, en donde la representación de un campo acústico de audio es un formato Ambisonics de al menos 2º orden.
- 15 4. Método de acuerdo con cualquiera de las reivindicaciones 1-3, en donde la matriz de modos pseudoinversa ( $\Xi^+$ ) se obtiene de acuerdo con  $\Xi^H [\Xi \Xi^H]^{-1}$ , en donde  $\Xi$  es la matriz de modos de la pluralidad de direcciones de la fuente.
5. Método de acuerdo con la reivindicación 4, en donde la matriz de decodificación ( $D_N$ ) se obtiene (135) de acuerdo con  $D = W \Xi^H [\Xi \Xi^H]^{-1} = W \Xi^+$ , en donde  $W$  es el conjunto de funciones de encuadre para cada altavoz.
- 20 6. Un dispositivo para decodificar una representación de reproducción de audio, que comprende:
  - unos primeros medios de cálculo (210) para calcular para cada uno de una pluralidad de altavoces una función de encuadre ( $W$ ) que usa un método geométrico basado en las posiciones de los altavoces y en una pluralidad de direcciones de la fuente;
  - 25 - unos segundos medios de cálculo (220) para calcular una matriz de modos ( $\Xi$ ) a partir de las direcciones de la fuente;
  - unos terceros medios de cálculo (230) para calcular una matriz de modos pseudoinversa ( $\Xi^+$ ) de la matriz de modos ( $\Xi$ ); y
  - unos medios (240) de decodificador para decodificar la representación de un campo acústico, en donde la decodificación está basada en una matriz de decodificación ( $D$ ) y los medios de decodificador usan al menos la función de encuadre ( $W$ ) y la matriz de modos pseudoinversa ( $\Xi^+$ ) para obtener la matriz de decodificación ( $D$ ).
- 30 7. Dispositivo de acuerdo con la reivindicación 6, en donde el dispositivo para decodificación comprende además unos medios (235) para calcular la matriz de decodificación ( $D$ ) a partir de la función de encuadre ( $W$ ) y de la matriz de modos pseudoinversa ( $\Xi^+$ ).
- 35 8. Dispositivo de acuerdo con la reivindicación 6 ó 7, en donde el método geométrico usado en el paso de cálculo de una función de encuadre es el Encuadre de Amplitud de la Base Vectorial (VBAP).
9. Dispositivo de acuerdo con cualquiera de las reivindicaciones 6-8, en donde la representación de un campo acústico es un formato Ambisonics de al menos 2º orden.
10. Dispositivo de acuerdo con cualquiera de las reivindicaciones 6-9, en donde la matriz de modos pseudoinversa  $\Xi^+$  se obtiene de acuerdo con  $\Xi^+ = \Xi^H [\Xi \Xi^H]^{-1}$ , en donde  $\Xi$  es la matriz de modos de la pluralidad de direcciones de la fuente.
- 40 11. Método de acuerdo con la reivindicación 10, en donde la matriz de decodificación ( $D_N$ ) se obtiene en un medio (245) para calcular una matriz de decodificación, de acuerdo con  $D = W \Xi^H [\Xi \Xi^H]^{-1} = W \Xi^+$ , en donde  $W$  es el conjunto de funciones de encuadre para cada altavoz.
- 45 12. Medio que puede ser leído por un ordenador que tiene almacenado en él unas instrucciones ejecutables para hacer que un ordenador ejecute un método para decodificar una representación de un sonido acústico para reproducción de audio, en donde el método comprende los pasos de
  - calcular (110) para cada uno de una pluralidad de altavoces una función de encuadre ( $W$ ) que usa un método geométrico basado en las posiciones de los altavoces y en una pluralidad de direcciones de la fuente;

- calcular (120) una matriz de modos ( $\Xi$ ) a partir de las direcciones de la fuente;
- calcular (130) una matriz de modos pseudoinversa ( $\Xi^+$ ) de la matriz de modos ( $\Xi$ ); y
- decodificar (140) la representación de un campo acústico de audio, en donde la decodificación está basada en una matriz de decodificación (D) que se obtiene a partir de la función de encuadre (W) y la matriz de modos pseudoinversa ( $\Xi^+$ ).

5

13. Un medio que puede ser leído por un ordenador, de acuerdo con la reivindicación 12, en donde el método geométrico usado en el paso de calcular una función de encuadre es un Encuadre de Amplitud de Base Vectorial (VBAP).

10 14. Un medio que puede ser leído por un ordenador, de acuerdo con la reivindicación 12 ó 13, en donde la representación de un campo acústico es un formato Ambisonics de al menos 2º orden.

15. Un medio que puede ser leído por un ordenador, de acuerdo con cualquiera de las reivindicaciones 12-14, en donde la matriz de modos pseudoinversa  $\Xi^+$  se obtiene de acuerdo con  $\Xi^+ = \Xi^H [\Xi \Xi^H]^{-1}$ , en donde  $\Xi$  es la matriz de modos de la pluralidad de direcciones de la fuente.

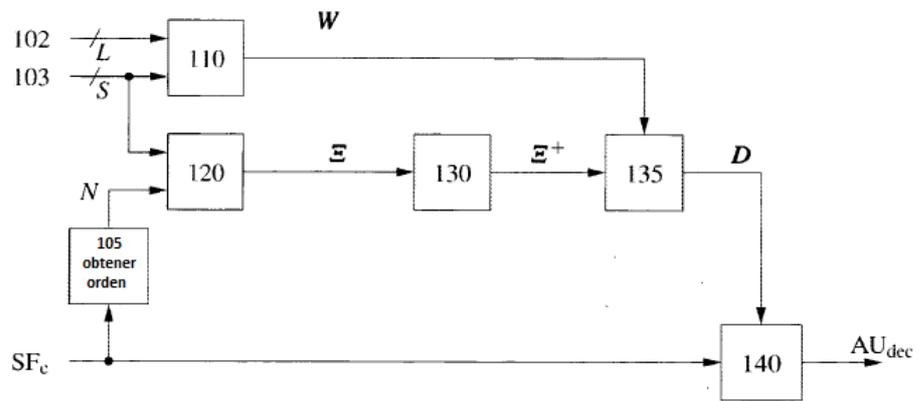


Fig.1

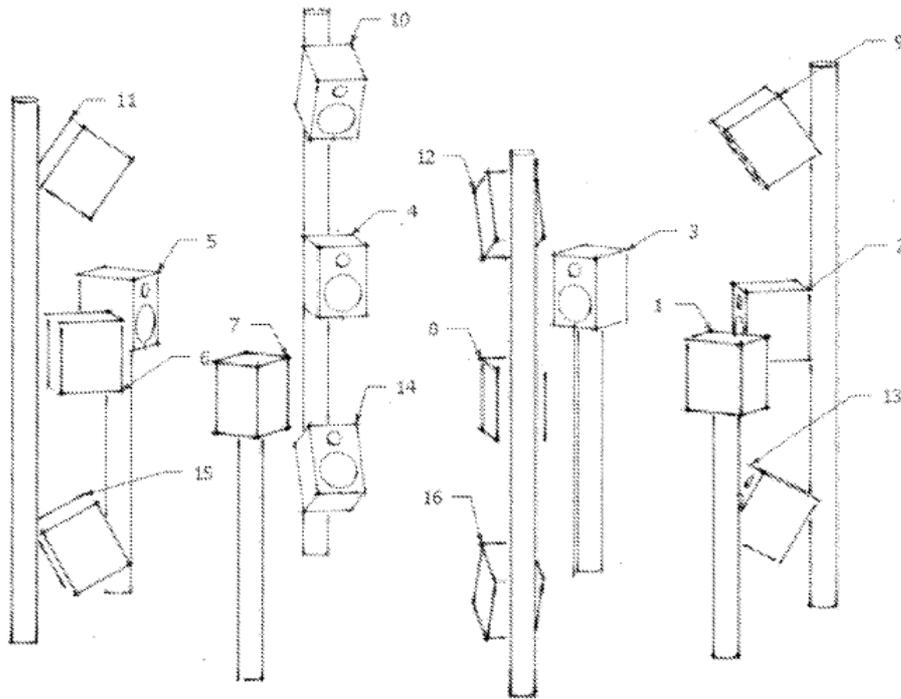


Fig.2

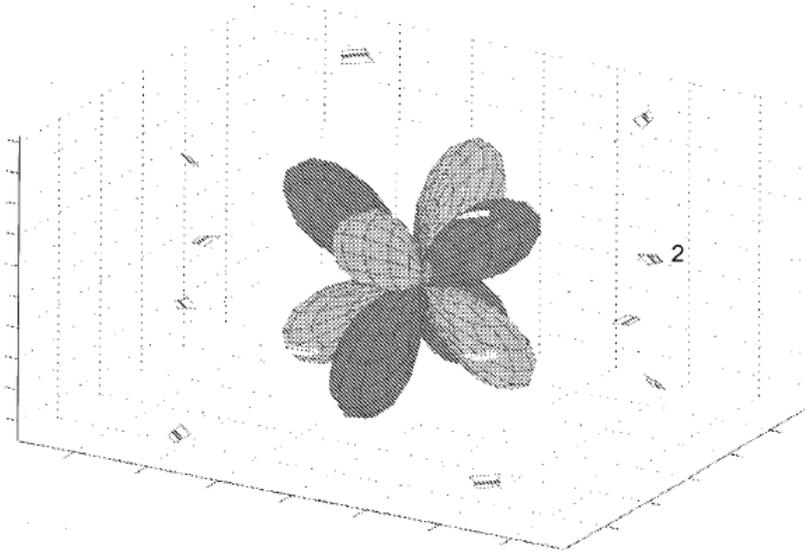


Fig.3

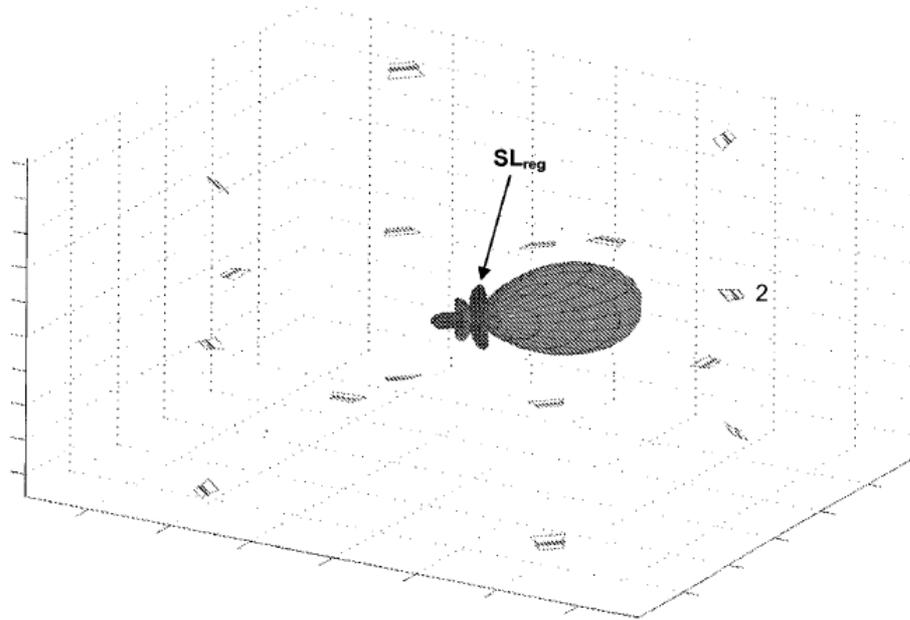


Fig.4

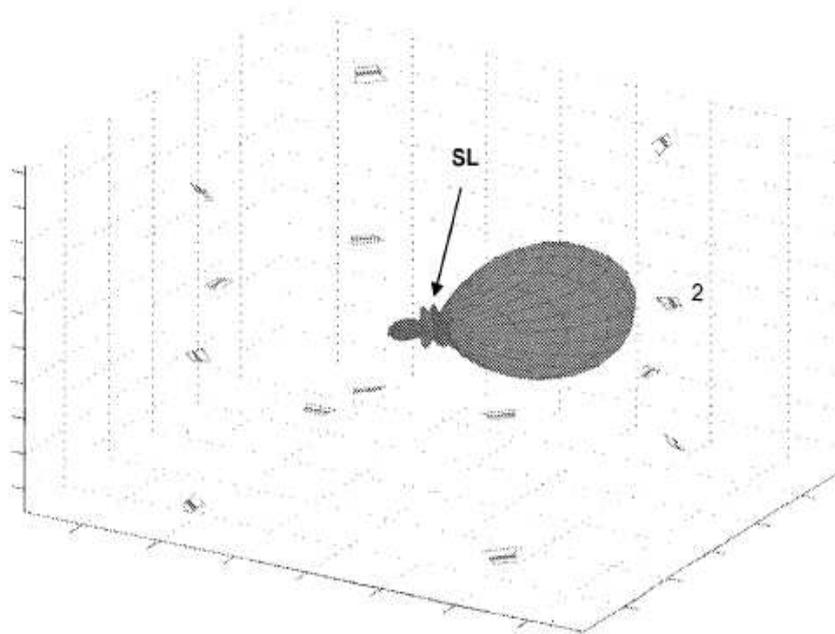


Fig.5

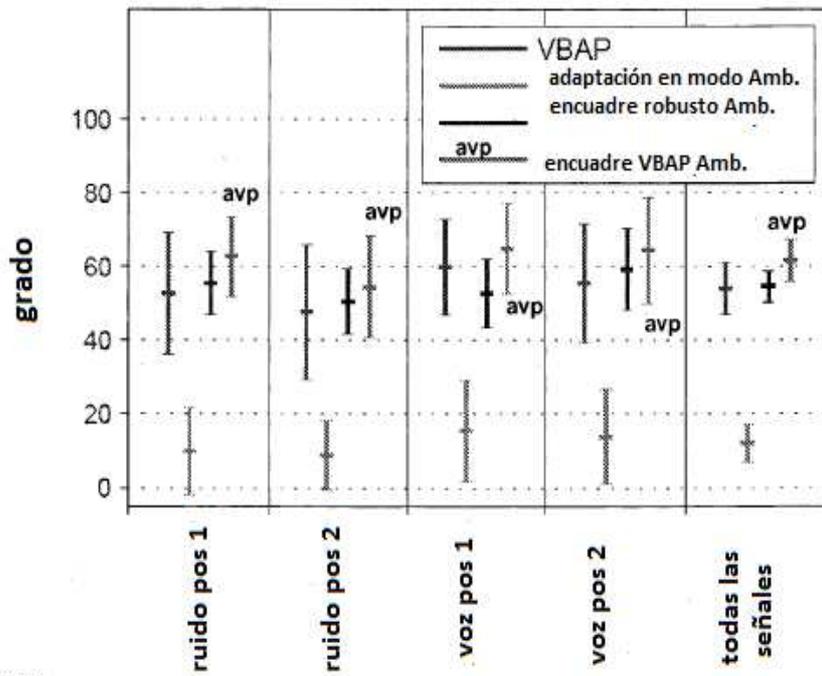


Fig.6

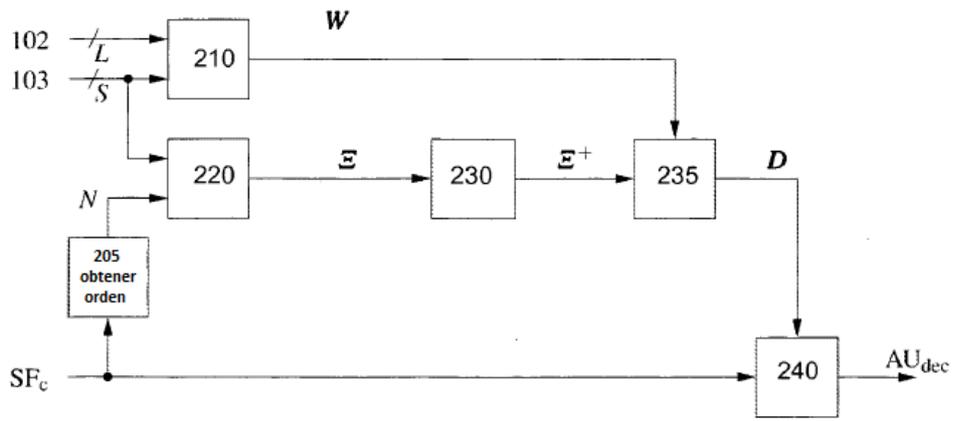


Fig.7