

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 974 440**

51 Int. Cl.:

H04S 3/02 (2006.01)

G10L 19/008 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **22.06.2015** **E 21159478 (3)**

97 Fecha y número de publicación de la concesión europea: **21.02.2024** **EP 3860154**

54 Título: **Método para decodificar una representación de tramas de datos HOA comprimida de un campo de sonido**

30 Prioridad:

27.06.2014 EP 14306024

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

27.06.2024

73 Titular/es:

DOLBY INTERNATIONAL AB (100.0%)
77 Sir John Rogerson's Quay, Block C, Grand Canal Docklands
Dublin, D02 VK60, IE

72 Inventor/es:

KRUEGER, ALEXANDER y
KORDON, SVEN

74 Agente/Representante:

LINAGE GONZÁLEZ, Rafael

ES 2 974 440 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 para decodificar una representación de tramas de datos HOA comprimida de un campo de sonido

5 **Campo técnico**

La invención se refiere a un método para decodificar una representación de tramas de datos HOA comprimida de un sonido o de un campo de sonido.

10 **Antecedentes**

La ambisónica de orden superior denominada HOA ofrece una posibilidad de representar sonido tridimensional. Otras técnicas son la síntesis de campo de ondas (WFS) o enfoques basados en canales como 22.2. A diferencia de los métodos basados en canales, la representación HOA ofrece la ventaja de ser independiente de una configuración de altavoz específica. Sin embargo, esta flexibilidad se produce a expensas de un proceso de decodificación que se requiere para la reproducción de la representación HOA en una configuración de altavoz particular. En comparación con el enfoque WFS, en el que la cantidad de altavoces necesarios es habitualmente muy grande, HOA se puede también renderizar en configuraciones que constan de sólo unos pocos altavoces. Una ventaja adicional de HOA es que la misma representación se puede también emplear sin ninguna modificación para la renderización binaural para auriculares.

HOA se basa en la representación de la densidad espacial de amplitudes de ondas planas armónicas complejas mediante una expansión de Armónicos Esféricos (SH) truncada. Cada coeficiente de expansión es una función de la frecuencia angular, lo que puede representarse de manera equivalente mediante una función en el dominio tiempo. Por consiguiente, sin pérdida de generalidad, se puede asumir que la representación completa de campo de sonido HOA consta en realidad de 0 funciones en el dominio tiempo, donde 0 denota el número de coeficientes de expansión. Estas funciones en el dominio tiempo se denominarán en lo sucesivo, de manera equivalente, secuencias de coeficientes HOA o canales HOA.

La resolución espacial de la representación HOA mejora con un orden máximo N creciente de la expansión. Desafortunadamente, el número de coeficientes O de expansión crece cuadráticamente con el orden N, en particular $O = (N + 1)^2$. Por ejemplo, las representaciones típicas de HOA que utilizan el orden $N = 4$ requieren $O = 25$ coeficientes HOA (expansión). La velocidad de bits total para la transmisión de la representación HOA, dada una velocidad de muestreo de un solo canal f_s deseada y un número de bits N_b por muestra, se determina mediante $O \cdot f_s \cdot N_b$. La transmisión de una representación HOA de orden $N = 4$, con una velocidad de muestreo de $f_s = 48$ kHz empleando $N_b = 16$ bits por muestra, da como resultado una velocidad de bits de 19,2 MBits/s, que es muy alta para muchas aplicaciones prácticas, por ejemplo, para transmisión ininterrumpida. De este modo, la compresión de las representaciones de HOA es muy conveniente.

Anteriormente, la compresión de las representaciones HOA de campo de sonido se propuso en los documentos EP 2665208 A1, EP 2743922 A1, EP 2800401 A1, cf. ISO/IEC JTC1/SC29/WG11, N14264, WD1-HOA Text of MPEG-H 3D Audio, enero de 2014. Estos enfoques tienen en común que realizan un análisis de campo de sonido y descomponen la representación HOA dada en un componente direccional y un componente ambiental residual. Por un lado, se asume que la representación comprimida final consta de una serie de señales cuantificadas, resultantes de la codificación perceptual de señales direccionales y basadas en vectores, así como de secuencias de coeficientes relevantes del componente HOA ambiental. Por otro lado, comprende información secundaria adicional relacionada con las señales cuantificadas, información secundaria que se requiere para la reconstrucción de la representación HOA a partir de su versión comprimida.

Antes de pasar al codificador perceptual, se requiere que estas señales intermedias en el dominio tiempo tengan una amplitud máxima dentro del intervalo de valores $[-1, 1]$, que es un requisito que surge de la implantación de los codificadores perceptuales actualmente disponibles. Con el fin de satisfacer este requisito al comprimir representaciones HOA, se utiliza una unidad de procesamiento de control de ganancia (véase el documento EP 2824661 A1 y el documento ISO/IEC JTC1/SC29/WG11 N14264 mencionado anteriormente) delante de los codificadores perceptuales, que atenúa o amplifica suavemente las señales de entrada. Se asume que la modificación de la señal resultante es invertible y que se aplica en cada trama, asumiéndose que, en particular, el cambio de las amplitudes de la señal entre tramas sucesivas es una potencia de '2'. Para facilitar la inversión de esta modificación de señal en el descompresor HOA, se incluye la información secundaria de normalización correspondiente en la información secundaria total. Esta información secundaria de normalización puede consistir en exponentes en base '2', exponentes que describen el cambio de amplitud relativa entre dos tramas sucesivas. Estos exponentes se codifican usando un código de longitud de ejecución de acuerdo con el documento ISO/IEC JTC1/SC29/WG11 N14264 mencionado anteriormente, ya que los cambios de amplitud menores entre tramas sucesivas son más probables que los mayores.

Sumario de la invención

Es factible utilizar cambios de amplitud codificados diferencialmente para reconstruir las amplitudes de la señal original en la descompresión HOA, por ejemplo, en caso de que un solo archivo se descomprima desde el principio hasta el final sin saltos temporales. Sin embargo, para facilitar el acceso aleatorio, deben estar presentes unidades de acceso independientes en la representación codificada (que normalmente es una corriente de bits) con el fin de permitir el inicio de la descompresión desde una posición deseada (o al menos en las proximidades de ella), independientemente de la información de tramas anteriores. Tal unidad de acceso independiente tiene que contener el cambio de amplitud absoluta total (es decir, un valor de ganancia no diferencial) originado por la unidad de procesamiento de control de ganancia desde la primera trama hasta la trama actual. Suponiendo que los cambios de amplitud entre dos tramas sucesivas son una potencia de '2', es suficiente describir también el cambio de amplitud absoluta total mediante un exponente de base '2'. Para una codificación eficiente de este exponente, es fundamental conocer las ganancias máximas potenciales de las señales antes de la aplicación de la unidad de procesamiento de control de ganancia. Sin embargo, este conocimiento depende en gran medida de la especificación de restricciones en el intervalo de valores de las representaciones HOA que se van a comprimir. Desafortunadamente, el documento de audio MPEG-H 3D ISO/IEC JTC1/SC29/WG11 N14264 proporciona sólo una descripción del formato para la representación HOA de entrada, sin establecer ninguna restricción en los intervalos de valores.

La invención establece una interrelación entre el intervalo de valores de la representación HOA de entrada y las ganancias máximas potenciales de las señales antes de la aplicación de la unidad de procesamiento de control de ganancia dentro del compresor HOA. En base a esa interrelación, se determina la cantidad de bits necesarios - para una especificación dada para el intervalo de valores de una representación HOA de entrada - para una codificación eficiente de los exponentes de base '2' para describir, dentro de una unidad de acceso, los cambios de amplitud absolutos totales (es decir, un valor de ganancia no diferencial) de las señales modificadas originados por la unidad de procesamiento de control de ganancia desde la primera trama hasta la trama actual.

Además, una vez fijada la regla para computar la cantidad de bits requeridos para la codificación del exponente, la invención utiliza un procesamiento para verificar si una representación HOA dada satisface las restricciones de intervalo de valores requeridas de tal manera que pueda comprimirse correctamente.

Breve descripción de los dibujos

Se describen realizaciones ejemplares de la invención con referencia a los dibujos que se acompañan, los cuales muestran:

Figura 1 Compresor HOA;

Figura 2 Descompresor HOA;

Figura 3 Valores K de escala para direcciones virtuales $\Omega_j^{(N)}$, $1 \leq j \leq O$, para órdenes HOA $N = 1, \dots, 29$;

Figura 4 Normas euclidianas de matrices de modo inverso Ψ^{-1} para direcciones virtuales $\Omega_{\text{MIN},d}$, $d = 1, \dots, O_{\text{MIN}}$ para órdenes HOA $N_{\text{MIN}} = 1, \dots, 9$;

Figura 5 Determinación de la magnitud máxima permitida γ_{dB} de señales de altavoces virtuales en posiciones $\Omega_j^{(N)}$, $1 \leq j \leq O$, donde $O = (N + 1)^2$;

Figura 6 Sistema de coordenadas esféricas.

Descripción de realizaciones

A continuación se presenta el principio de compresión y descompresión HOA con el fin de proporcionar un contexto más detallado en el que se produce el problema mencionado anteriormente. La base de esta presentación es el procesamiento descrito en el documento MPEG-H 3D de audio ISO/IEC JTC1/SC29/WG11 N14264, véanse también los documentos EP 2665208 A1, EP 2800401 A1 y EP 2743922 A1. En el documento N14264, el 'componente direccional' se extiende a un 'componente de sonido predominante'. Como componente direccional, se asume que el componente de sonido predominante está representado en parte por señales direccionales, es decir, señales monoaurales con su dirección correspondiente, desde la cual se asume que inciden en el oyente, junto con algunos parámetros de predicción para predecir porciones de la representación HOA original a partir de las señales direccionales. Además, se asume que el componente de sonido predominante está representado por 'señales basadas en vectores', es decir, señales monoaurales con un vector correspondiente que define la distribución direccional de las señales basadas en vectores.

Compresión HOA

La arquitectura general del compresor HOA descrito en el documento EP 2800401 A1 se ilustra en la figura 1. Tiene una parte de codificación HOA espacial representada en la figura 1A y una parte de codificación de fuente

y percepción representada en la figura 1B. El codificador espacial HOA proporciona una primera representación HOA comprimida que consta de I señales junto con información secundaria que describe cómo crear una representación HOA de las mismas. En los codificadores de fuentes de información perceptual y secundaria, las I señales se codifican perceptualmente, y la información secundaria se somete a codificación de fuente, antes de que se multiplexen las dos representaciones codificadas.

Codificación espacial HOA

En un primer paso, se introduce una k -ésima trama $C(k)$ actual de la representación HOA original en un paso o una etapa 11 de procesamiento de estimación de dirección y vector, que se asume que proporciona los conjuntos $M_{DIR(k)}$ y $M_{VEC(k)}$ de tuplas. El conjunto $M_{DIR(k)}$ de tuplas consta de tuplas de las cuales el primer elemento indica el índice de una señal direccional y el segundo elemento indica la dirección cuantificada respectiva. El conjunto de tuplas $M_{VEC(k)}$ consta de tuplas de las cuales el primer elemento indica el índice de una señal basada en vectores y el segundo elemento denota el vector que define la distribución direccional de las señales, es decir, cómo se computa la representación HOA de la señal basada en vectores.

Utilizando ambos conjuntos $M_{DIR(k)}$ y $M_{VEC(k)}$ de tuplas, la trama HOA inicial $C(k)$ se descompone en un paso o una etapa 12 de descomposición HOA en la trama $X_{PS}(k-1)$ de todas las señales de sonido predominantes (es decir, direccionales y basadas en vectores) y la trama $C_{AMB}(k-1)$ del componente HOA ambiental. Obsérvese el retardo de una trama, que se debe al procesamiento de superposición y adición con el fin de evitar artefactos de bloqueo. Además, se asume que el paso o la etapa 12 de descomposición HOA genera algunos parámetros $\zeta(k-1)$ de predicción que describen cómo predecir porciones de la representación HOA original a partir de las señales direccionales, con el fin de enriquecer el componente HOA de sonido predominante. Además, se asume que se proporciona un vector $v_{A,T}(k-1)$ de asignación objetivo que contiene información sobre la asignación de señales de sonido predominantes, que se determinaron en el paso o la etapa 12 del procesamiento de descomposición HOA, a los I canales disponibles. Se puede asumir que los canales afectados están ocupados, es decir, no están disponibles para transportar ninguna secuencia de coeficientes del componente HOA ambiental en el intervalo de tiempo respectivo.

En el paso o la etapa 13 de procesamiento de modificación de componente ambiental, la trama $C_{AMB}(k-1)$ del componente HOA ambiental se modifica de acuerdo con la información proporcionada por el vector de asignación objetivo $v_{A,T}(k-1)$. En particular, se determina qué secuencias de coeficientes del componente HOA ambiental deben transmitirse en los canales I dados, dependiendo (entre otros aspectos) de la información (contenida en el vector $v_{A,T}(k-1)$ de asignación objetivo) sobre qué canales están disponibles y no están ocupados por las señales de sonido predominantes. Además, se realiza una aparición y desaparición gradual de secuencias de coeficientes si los índices de las secuencias de coeficientes elegidas varían entre tramas sucesivas.

Además, se asume que las primeras secuencias de coeficientes O_{MIN} del componente HOA ambiental $C_{AMB}(k-2)$ siempre se eligen para codificarse y transmitirse perceptualmente, donde $O_{MIN} = (N_{MIN} + 1)^2$, siendo típicamente $N_{MIN} \leq N$ un orden más pequeño que el de la representación original de HOA. Con el fin de descorrelacionar estas secuencias de coeficientes HOA, se pueden transformar en el paso/la etapa 13 en señales direccionales (es decir, funciones de onda plana generales) que inciden desde algunas direcciones predefinidas $\Omega_{MIN, d}$, $d = 1, \dots, O_{MIN}$.

Junto con el componente HOA ambiental modificado $C_{M,A}(k-1)$ se computa en el paso/la etapa 13 un componente HOA ambiental modificado $C_{P,M,A}(k-1)$ predicho temporalmente y se utiliza en los pasos o las etapas 15, 151 de procesamiento de control de ganancia con el fin de permitir una anticipación razonable, estando la información sobre la modificación del componente HOA ambiental directamente relacionada con la asignación de todos los tipos posibles de señales a los canales disponibles en el paso o la etapa 14 de asignación de canales. Se asume que la información final sobre esa asignación está contenida en el vector de asignación final $v_A(k-2)$. Con el fin de computar este vector en el paso/la etapa 13, se explota la información contenida en el vector $v_{A,T}(k-1)$ de asignación objetivo.

La asignación de canal en el paso/la etapa 14 asigna, con la información proporcionada por el vector de asignación $v_A(k-2)$, las señales apropiadas contenidas en la trama $X_{PS}(k-2)$ y la contenida en la trama $C_{M,A}(k-2)$ a los I canales disponibles, lo que produce las tramas de señal $y_i(k-2)$, $i = 1, \dots, I$. Además, las señales apropiadas contenidas en la trama $X_{PS}(k-1)$ y en la trama $C_{P,M,A}(k-1)$ también se asignan a los I canales disponibles, lo que produce las tramas de señal predichas $y_{P,i}(k-1)$, $i = 1, \dots, I$.

Cada una de las tramas $y_i(k-2)$, $i = 1, \dots, I$ de señal es finalmente procesada por el control 15, 151 de ganancia dando como resultado exponentes $e_i(k-2)$ e indicadores $\beta_i(k-2)$, $i = 1, \dots, I$ de excepción y señales $z_i(k-2)$, $i = 1, \dots, I$, en los que la ganancia de la señal se modifica suavemente tal como para conseguir un intervalo de valores adecuado para los pasos o las etapas 16 de codificador perceptual. Los pasos/las etapas 16 emiten las tramas de señal codificadas correspondientes $\tilde{z}_i(k-2)$, $i = 1, \dots, I$. Las tramas de señal predichas $y_{P,i}(k-1)$, $i = 1, \dots, I$

= 1, ..., I permiten una especie de anticipación con el fin de evitar cambios severos de ganancia entre bloques sucesivos. Los datos de información secundaria $M_{DIR(k-1)}$, $M_{VEC(k-1)}$, $e_i(k-2)$, $\beta_i(k-2)$, $\zeta_i(k-1)$ y $v_A(k-2)$ están codificados en fuente en el paso o la etapa 17 del codificador de fuente de información secundaria, dando como resultado la trama de información secundaria codificada $\tilde{I}(k-2)$. En un multiplexor 18, las señales codificadas $\tilde{z}_i(k-2)$ de la trama (k - 2) y los datos de información secundaria codificados $\tilde{I}(k-2)$ para esta trama se combinan, lo que da como resultado la trama $\tilde{B}(k-2)$ de salida.

En un decodificador HOA espacial, se asume que las modificaciones de ganancia en los pasos/las etapas 15, 151 se revierten utilizando la información secundaria de control de ganancia, que consta de los exponentes $e_i(k-2)$ y de los indicadores de excepción $\beta_i(k-2)$, $i = 1, \dots, I$.

Descompresión HOA

La arquitectura general del descompresor HOA descrita en el documento EP 2800401 A1 se ilustra en la figura 2. Consta de las contrapartes de los componentes de compresor HOA, que están dispuestos en orden inverso e incluyen una parte de percepción y decodificación de fuente representada en la figura 2A y una parte de decodificación HOA espacial representada en la figura 2B.

En la parte de decodificación de fuente y perceptual (que representa un decodificador de fuente de información secundaria y perceptual) un paso o una etapa de demultiplexación 21 recibe la trama $\tilde{B}(k)$ de entrada de la corriente de bits y proporciona la representación codificada perceptualmente $\tilde{z}_i(k)$, $i = 1, \dots, I$ de las señales I y los datos $\tilde{I}(k)$ de información secundaria codificados describiendo cómo crear una representación de HOA de los mismos. Las $\tilde{z}_i(k)$ señales se decodifican perceptualmente en un paso o una etapa 22 del decodificador perceptual, dando como resultado señales decodificadas $\hat{z}_i(k)$, $i = 1, \dots, I$. Los datos $\tilde{I}(k)$ de información secundaria codificados se decodifican en un paso o una etapa 23 de decodificador de fuente de información secundaria, dando como resultado conjuntos de datos $M_{DIR(k+1)}$, $M_{VEC(k+1)}$, exponentes $e_i(k)$, indicadores $\beta_i(k)$ de excepción, parámetros $\zeta(k+1)$ de predicción y un vector $v_{AMB,ASSIGN}(k)$ de asignación. Con respecto a la diferencia entre v_A y $v_{AMB,ASSIGN}(k)$, véase el documento N14264 mencionado anteriormente de MPEG.

Decodificación espacial HOA

En la parte de decodificación HOA espacial, cada una de las señales decodificadas perceptualmente $\hat{z}_i(k)$, $i = 1, \dots, I$, se introduce en un paso o una etapa 24, 241 de procesamiento de control de ganancia inversa junto con su exponente $e_i(k)$ de corrección de ganancia y su indicador $\beta_i(k)$ de excepción de corrección de ganancia asociados. El i-ésimo paso/la i-ésima etapa de procesamiento de control de ganancia inversa proporciona una trama $\hat{y}_i(k)$ de señal corregida de ganancia.

Todas las tramas I de señal corregidas de ganancia $\hat{y}_i(k)$, $i = 1, \dots, I$ se alimentan junto con el vector $v_{AMB,ASSIGN}(k)$ de asignación y los conjuntos de tuplas $M_{DIR(k+1)}$ y $M_{VEC(k+1)}$ a un paso o una etapa de reasignación de canal 25, cf. la definición descrita anteriormente de los conjuntos de tuplas $M_{DIR(k+1)}$ y $M_{VEC(k+1)}$. El vector $v_{AMB,ASSIGN}(k)$ de asignación consta de I componentes que indican para cada canal de transmisión si contiene una secuencia de coeficientes del componente HOA ambiental y cuál contiene. En el paso/la etapa 25 de reasignación de canal, las tramas $\hat{y}_i(k)$ de señal corregidas de ganancia se redistribuyen con el fin de reconstruir la trama $\hat{X}_{PS}(k)$ de todas las señales de sonido predominantes (es decir, todas las señales direccionales y basadas en vectores) y la trama $C_{I,AMB}(k)$ de una representación intermedia del componente HOA ambiental. Además, se proporcionan el conjunto $J_{AMB,ACT}(k)$ de índices de secuencias de coeficientes del componente HOA ambiental activo en la k-ésima trama, y los conjuntos $J_E(k-1)$, $J_D(k-1)$ y $J_U(k-1)$ de datos de índices de coeficientes del componente HOA ambiental, que deben habilitarse, deshabilitarse y permanecer activos en la (k - 1)-ésima trama.

En un paso o una etapa 26 de síntesis de sonido predominante, la representación HOA del componente $\hat{C}_{PS}(k-1)$ de sonido predominante se computa a partir de la trama $\hat{X}_{PS}(k)$ de todas las señales de sonido predominantes usando el conjunto $M_{DIR(k+1)}$ de tuplas, el conjunto $\zeta(k+1)$ de parámetros de predicción, el conjunto $M_{VEC(k+1)}$ de tuplas y los conjuntos $J_E(k-1)$, $J_D(k-1)$ y $J_U(k-1)$ de datos.

En un paso o una etapa 27 de síntesis de ambiente, la trama de componente HOA ambiental $\hat{C}_{AMB}(k-1)$ se crea a partir de la trama $C_{I,AMB}(k)$ de la representación intermedia del componente HOA ambiental, usando el conjunto $J_{AMB,ACT}(k)$ de índices de secuencias de coeficientes del componente HOA ambiental que están activos en la k-ésima trama. El retardo de una trama se introduce debido a la sincronización con el componente HOA de sonido predominante. Finalmente, en un paso o una etapa 28 de composición HOA, la trama $\hat{C}_{AMB}(k-1)$ de componente HOA ambiental y la trama $\hat{C}_{PS}(k-1)$ de componente HOA de sonido predominante se superponen como para proporcionar la trama HOA decodificada $\hat{C}(k-1)$.

Más tarde, el decodificador HOA espacial crea, a partir de las I señales y de la información secundaria, la representación HOA reconstruida.

En caso de que en el lado de codificación el componente HOA ambiental se haya transformado en señales direccionales, esa transformación se invierte en el lado del decodificador en el paso/la etapa 27.

Las ganancias máximas potenciales de las señales antes de las etapas/etapas 15, 151 de procesamiento de control de ganancia dentro del compresor HOA dependen en gran medida del intervalo de valores de la representación HOA de entrada. Por lo tanto, al principio se define un intervalo de valores significativo para la representación HOA de entrada, seguido de una conclusión sobre las ganancias máximas potenciales de las señales antes de ingresar a los pasos/las etapas de procesamiento de control de ganancia.

Normalización de la representación HOA de entrada

Para utilizar el procesamiento inventivo se debe llevar a cabo antes una normalización de la señal de representación HOA de entrada (total). Para la compresión HOA se realiza un procesamiento por tramas, donde la k-ésima trama $C(k)$ de la representación HOA de entrada original se define con respecto al vector $c(t)$ de secuencias de coeficientes HOA continuas en el tiempo especificadas en la ecuación (54) en la sección de conceptos básicos de la ambisónica de orden superior como

$$C(k) := [c((kL + 1)T_s) \quad c((kL + 2)T_s) \quad \dots \quad c((k + 1)LT_s)] \in \mathbb{R}^{O \times L}, \quad (1)$$

donde k denota el índice de la trama, L la longitud de la trama (en muestras), $0 = (N + 1)^2$ el número de secuencias de coeficientes HOA y T_s indica el período de muestreo.

Como se menciona en el documento EP 2824661 A1, no se consigue una normalización significativa de una representación HOA vista desde una perspectiva práctica imponiendo restricciones en el intervalo de valores de las secuencias $c_n^m(t)$ de coeficientes HOA individuales, ya que estas funciones en el dominio tiempo no son las señales que realmente reproducen los altavoces después de la renderización. En cambio, es más conveniente considerar la 'representación de dominio espacial equivalente', que se obtiene renderizando la representación HOA en O señales de altavoces virtuales $w_j(t)$, $1 \leq j \leq O$. Se asume que las respectivas posiciones de los altavoces virtuales están expresadas mediante un sistema de coordenadas esféricas, asumiéndose que cada posición se encuentra en la esfera unitaria y tiene un radio de '1'. Por consiguiente, las posiciones pueden expresarse de manera equivalente mediante direcciones $\Omega_j^{(N)} = (\theta_j^{(N)}, \phi_j^{(N)})$ dependientes de orden, $1 \leq j \leq O$, donde $\theta_j^{(N)}$ y $\phi_j^{(N)}$ denotan las inclinaciones y acimuts, respectivamente (véase también la figura 6 y su descripción para la definición del sistema de coordenadas esféricas). Estas direcciones deben distribuirse lo más uniformemente posible en la esfera unitaria, véase, por ejemplo, J. Fliege, U. Maier, "A two-stage approach for computing cubature formulae for the sphere", Technical report, Fachbereich Mathematik, Universidad de Dortmund, 1999. Los números de nodo se encuentran en <http://www.mathematik.uni-dortmund.de/lx/research/projects/fliege/nodes/nodes.html> para el cálculo de direcciones específicas. Estas posiciones dependen en general del tipo de definición de "distribución uniforme en la esfera" y, por tanto, no son inequívocas.

La ventaja de definir intervalos de valores para señales de altavoces virtuales sobre definir intervalos de valores para secuencias de coeficientes HOA es que el intervalo de valores para las primeras se puede establecer intuitivamente igualmente al intervalo $[-1, 1]$ como es el caso para señales de altavoces convencionales que asumen representación PCM. Esto conduce a un error de cuantificación distribuido espacialmente de manera uniforme, de tal modo que, ventajosamente, la cuantificación se aplica en un dominio que es relevante con respecto a la escucha real. Un aspecto importante en este contexto es que se puede elegir que el número de bits por muestra sea tan bajo como suele ser para señales de altavoces convencionales, es decir, 16, lo que aumenta la eficiencia en comparación con la cuantificación directa de secuencias de coeficientes HOA, en la que normalmente se requiere un mayor número de bits (por ejemplo, 24 o incluso 32) por muestra.

Para describir en detalle el proceso de normalización en el dominio espacial, todas las señales de altavoces virtuales se resumen en un vector como

$$w(t) := [w_1(t) \quad \dots \quad w_O(t)]^T, \quad (2)$$

donde $(\cdot)^T$ denota transposición. Denotando la matriz modal con respecto a las direcciones virtuales $\Omega_j^{(N)}$, $1 \leq j \leq O$, por Ψ , que está definida por

$$\Psi := [S_1 \quad \dots \quad S_O] \in \mathbb{R}^{O \times O} \quad (3)$$

$$\text{con } \mathbf{S}_j := \begin{pmatrix} S_0^0(\boldsymbol{\Omega}_j^{(N)}) & S_1^{-1}(\boldsymbol{\Omega}_j^{(N)}) & S_1^0(\boldsymbol{\Omega}_j^{(N)}) & S_1^1(\boldsymbol{\Omega}_j^{(N)}) & \dots & S_N^{N-1}(\boldsymbol{\Omega}_j^{(N)}) & S_N^N(\boldsymbol{\Omega}_j^{(N)}) \end{pmatrix}^T,$$

el proceso de renderización se puede formular como una multiplicación de matrices

$$\mathbf{w}(t) = (\boldsymbol{\Psi})^{-1} \cdot \mathbf{c}(t). \quad (5)$$

Utilizando estas definiciones, un requisito razonable para las señales de los altavoces virtuales es:

$$\|\mathbf{w}(lT_s)\|_\infty = \max_{1 \leq j \leq O} |w_j(lT_s)| \leq 1 \quad \forall l, \quad (6)$$

lo que significa que se requiere que la magnitud de cada señal de altavoz virtual esté dentro del intervalo $[-1, 1]$. Un instante de tiempo t está representado por un índice l de muestra y por un periodo T_s de muestra de los valores de muestra de dichas tramas de datos HOA.

Por consiguiente, la potencia total de las señales de los altavoces satisface la condición

$$\|\mathbf{w}(lT_s)\|_2^2 = \sum_{j=1}^O |w_j(lT_s)|^2 \leq O \quad \forall l. \quad (7)$$

La renderización y la normalización de la representación de la trama de datos HOA se lleva a cabo aguas arriba de la entrada $C(k)$ de la figura 1A.

Consecuencias para el intervalo de valores de la señal antes del control de ganancia

Asumiendo que la normalización de la representación HOA de entrada se realiza de acuerdo con la descripción en la sección Normalización de la representación HOA de entrada, el intervalo de valores de las señales y_i , $i = 1, \dots, I$, que se introduce en la unidad 15, 151 de procesamiento de control de ganancia en el compresor HOA, se considera a continuación. Estas señales se crean mediante la asignación a los canales I disponibles de una o más de las secuencias de coeficientes HOA, o señales de sonido predominantes $x_{PS,d}$, $d = 1, \dots, D$, y/o secuencias de coeficientes particulares del componente HOA ambiental $c_{AMB,n}$, $n = 1, \dots, O$, a parte de las cuales se aplica una transformación espacial. Por consiguiente, es necesario analizar el posible intervalo de valores de estos diferentes tipos de señales mencionados bajo la asunción de normalización de la ecuación (6). Dado que todo tipo de señales se computan intermedidamente a partir de las secuencias de coeficientes HOA originales, se echa una ojeada a sus posibles intervalos de valores.

El caso en el que sólo una o más secuencias de coeficientes HOA estén contenidas en los canales I no se representa en las figuras 1A y 2B, es decir, en tal caso no se requieren la descomposición de HOA, la modificación de los componentes ambientales y los bloques de síntesis correspondientes.

Consecuencias para el intervalo de valores de la representación HOA La representación HOA continua en el tiempo se obtiene a partir de las señales de altavoz virtuales mediante

$$\mathbf{c}(t) = \boldsymbol{\Psi} \mathbf{w}(t), \quad (8)$$

que es la operación inversa a la de la ecuación (5). Por consiguiente, la potencia total de todas las secuencias de coeficientes HOA está limitada de la siguiente manera:

$$\|\mathbf{c}(lT_s)\|_2^2 \leq \|\boldsymbol{\Psi}\|_2^2 \cdot \|\mathbf{w}(lT_s)\|_2^2 \leq \|\boldsymbol{\Psi}\|_2^2 \cdot O, \quad (9)$$

utilizando las ecuaciones (8) y (7).

Bajo la asunción de normalización de N3D de las funciones armónicas esféricas, la norma euclidiana al cuadrado de la matriz modal se puede escribir mediante

$$\|\boldsymbol{\Psi}\|_2^2 = K \cdot O, \quad (10a)$$

donde

$$K = \frac{\|\Psi\|_2^2}{O} \quad (10b)$$

denota la relación entre la norma euclidiana al cuadrado de la matriz modal y el número O de secuencias de coeficientes HOA.

Esta relación depende del orden N de HOA específico y de las direcciones específicas $\Omega_j^{(N)}$ de altavoz virtuales, $1 \leq j \leq O$, que se puede expresar añadiendo a la relación la lista de parámetros respectiva de la siguiente manera:

$$K = K(N, \Omega_1^{(N)}, \dots, \Omega_O^{(N)}) \quad (10c)$$

La figura 3 muestra los valores de K para direcciones virtuales $\Omega_j^{(N)}$, $1 \leq j \leq O$, de acuerdo con el artículo de Fliege et al. anteriormente mencionado para órdenes $N = 1, \dots, 29$ de HOA.

La combinación de todos los argumentos y consideraciones anteriores proporciona un límite superior para la magnitud de las secuencias de coeficientes HOA de la siguiente manera:

$$\|c(lT_s)\|_\infty \leq \|c(lT_s)\|_2 \leq \sqrt{K} \cdot O \quad (11)$$

donde la primera desigualdad procede directamente de las definiciones de las normas.

Es importante señalar que la condición de la ecuación (6) implica la condición de la ecuación (11), pero que no se sostiene lo contrario, es decir, que la ecuación (11) no implica la ecuación (6).

Un aspecto importante adicional es que, bajo la asunción de posiciones de altavoces virtuales distribuidas casi uniformemente, los vectores de columna de la matriz Ψ de modos, que representan los vectores de modos con respecto a las posiciones de los altavoces virtuales, son casi ortogonales entre sí y tienen una norma euclidiana de $N + 1$ cada uno. Esta propiedad significa que la transformación espacial conserva casi la norma euclidiana excepto por una constante multiplicativa, es decir, que

$$\|c(lT_s)\|_2 \approx (N + 1) \|w(lT_s)\|_2 \quad (12)$$

La verdadera norma $\|c(lT_s)\|_2$ difiere tanto más de la aproximación en la ecuación (12) cuanto más se viola la asunción de ortogonalidad en los vectores modales.

Consecuencias para el intervalo de valores de las señales de sonido predominantes

Ambos tipos de señales de sonido predominantes (direccionales y basadas en vectores) tienen en común que su contribución a la representación HOA se describe mediante un único vector $v_1 \in \mathbb{R}^D$ con norma euclidiana de $N + 1$, es decir, que

$$\|v_1\|_2 = N + 1 \quad (13)$$

En el caso de la señal direccional, este vector corresponde al vector de modos con respecto a una determinada dirección $\Omega_{s,1}$ de la fuente de señal, es decir, que

$$v_1 = S(\Omega_{s,1}) \quad (14)$$

$$:= [S_0^0(\Omega_{s,1}) \ S_1^{-1}(\Omega_{s,1}) \ S_1^0(\Omega_{s,1}) \ S_1^1(\Omega_{s,1}) \ \dots \ S_N^{N-1}(\Omega_{s,1}) \ S_N^N(\Omega_{s,1})]^T \quad (15)$$

Este vector describe mediante una representación HOA un haz direccional en la dirección $\Omega_{s,1}$ de la fuente de señal. En el caso de una señal basada en vectores, el vector v_1 no está obligado a ser un vector de modo con respecto a ninguna dirección, y, por consiguiente, puede describir una distribución direccional más general de la señal basada en vectores monoaurales.

A continuación se considera el caso general de D señales $x_d(t)$, $d = 1, \dots, D$ de sonido predominantes, que pueden recogerse en el vector $x(t)$ de acuerdo con

$$\mathbf{x}(t) = [x_1(t) \ x_2(t) \ \dots \ x_D(t)]^T. \quad (16)$$

Estas señales tienen que determinarse en base a la matriz

$$\mathbf{V} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_D] \quad (17)$$

que está formada por todos los vectores \mathbf{v}_d , $d = 1, \dots, D$, que representan la distribución direccional de las señales de sonido monoaurales predominantes $x_d(t)$, $d = 1, \dots, D$.

10 Para una extracción significativa de las señales de sonido predominantes $\mathbf{x}(t)$, se formulan las siguientes restricciones:

a) Cada señal de sonido predominante se obtiene como una combinación lineal de las secuencias de coeficientes de la representación HOA original, es decir, que

$$\mathbf{x}(t) = \mathbf{A} \cdot \mathbf{c}(t), \quad (18)$$

donde $\mathbf{A} \in \mathbb{R}^{D \times O}$ denota la matriz de mezcla.

20 b) La matriz \mathbf{A} de mezcla debe elegirse de tal manera que su norma euclidiana no exceda el valor de '1', es decir, que

$$\|\mathbf{A}\|_2 \leq 1, \quad (19)$$

25 y de tal manera que la norma euclidiana al cuadrado (o potencia equivalente) del residual entre la representación HOA original y la de las señales de sonido predominantes no sea mayor que la norma euclidiana al cuadrado (o potencia equivalente) de la representación HOA original, es decir, que

$$\|\mathbf{c}(t) - \mathbf{V} \cdot \mathbf{x}(t)\|_2^2 \leq \|\mathbf{c}(t)\|_2^2. \quad (20)$$

30 Insertando la ecuación (18) en la ecuación (20) se puede ver que la ecuación (20) es equivalente a la restricción

$$\|\mathbf{I} - \mathbf{V} \cdot \mathbf{A}\|_2 \leq 1, \quad (21)$$

35 donde \mathbf{I} denota la matriz identidad.

A partir de las restricciones en la ecuación (18) y en (19) y de la compatibilidad de la matriz euclidiana y las normas vectoriales, se encuentra un límite superior para las magnitudes de las señales de sonido predominantes mediante

$$\|\mathbf{x}(lT_s)\|_\infty \leq \|\mathbf{x}(lT_s)\|_2 \quad (22)$$

$$\leq \|\mathbf{A}\|_2 \|\mathbf{c}(lT_s)\|_2 \quad (23)$$

$$\leq \sqrt{K} \cdot O, \quad (24)$$

utilizando las ecuaciones (18), (19) y (11). Por consiguiente, se garantiza que las señales de sonido predominantes permanezcan en el mismo intervalo que las secuencias de coeficientes HOA originales (compárese con la ecuación (11)), es decir, que

$$\boxed{\|\mathbf{x}(lT_s)\|_\infty \leq \sqrt{K} \cdot O} \quad (25)$$

. (25)

55 *Ejemplo de elección de matriz de mezcla*

Un ejemplo de cómo determinar la matriz de mezcla que satisface la restricción (20) se obtiene computando las señales de sonido predominantes de tal manera que se minimice la norma euclidiana del residuo después de la extracción, es decir, que

$$\mathbf{x}(t) = \operatorname{argmin}_{\mathbf{x}(t)} \|\mathbf{V} \cdot \mathbf{x}(t) - \mathbf{c}(t)\|_2 \quad (26)$$

La solución al problema de minimización en la ecuación (26) viene dada por

$$\mathbf{x}(t) = \mathbf{V}^+ \mathbf{c}(t), \quad (27)$$

donde $(\cdot)^+$ indica la pseudoinversa de Moore-Penrose. Comparando la ecuación (27) con la ecuación (18) se deduce que, en este caso, la matriz de mezcla es igual a la pseudo inversa de Moore-Penrose de la matriz \mathbf{V} , es decir, que $\mathbf{A} = \mathbf{V}^+$. Sin embargo, todavía hay que elegir la matriz \mathbf{V} para satisfacer la restricción (19), es decir, que

$$\|\mathbf{V}^+\|_2 \leq 1 \quad (28)$$

En el caso de señales únicamente direccionales, donde la matriz \mathbf{V} es la matriz de modo con respecto a algunas direcciones de la señal fuente

$$\boldsymbol{\Omega}_{S,d}, \quad d = 1, \dots, D, \quad \text{i.e.} \quad \mathbf{V} = [\mathbf{s}(\boldsymbol{\Omega}_{S,1}) \quad \mathbf{s}(\boldsymbol{\Omega}_{S,2}) \quad \dots \quad \mathbf{s}(\boldsymbol{\Omega}_{S,D})] \quad (29)$$

la restricción (28) puede satisfacerse eligiendo las direcciones de la señal fuente $\boldsymbol{\Omega}_{S,d}$, $d = 1, \dots, D$, de tal modo que la distancia de dos direcciones vecinas cualesquiera no sea demasiado pequeña.

Consecuencias para el intervalo de valores de las secuencias de coeficientes del componente HOA ambiental

El componente HOA ambiental se computa restando de la representación HOA original la representación HOA de las señales de sonido predominantes, es decir, que

$$\mathbf{c}_{\text{AMB}}(t) = \mathbf{c}(t) - \mathbf{V} \cdot \mathbf{x}(t). \quad (30)$$

Si el vector de señales de sonido predominantes $\mathbf{x}(t)$ se determina de acuerdo con el criterio (20), se puede concluir que

$$\|\mathbf{c}_{\text{AMB}}(lT_S)\|_\infty \leq \|\mathbf{c}_{\text{AMB}}(lT_S)\|_2 \quad (31)$$

$$\stackrel{(30)}{=} \|\mathbf{c}(lT_S) - \mathbf{V} \cdot \mathbf{x}(lT_S)\|_2 \quad (32)$$

$$\stackrel{(20)}{\leq} \|\mathbf{c}(lT_S)\|_2 \quad (33)$$

$$\stackrel{(11)}{=} \sqrt{K} \cdot O \quad (34)$$

Rango de valores de secuencias de coeficientes transformadas espacialmente del componente HOA ambiental

Un aspecto adicional en el procesamiento de compresión HOA propuesto en el documento EP 2743922 A1 y en el documento N14264 de MPEG mencionado anteriormente es que las primeras secuencias de coeficientes O_{MIN} del componente HOA ambiental siempre se eligen para asignarse a los canales de transporte, donde $O_{\text{MIN}} = (N_{\text{MIN}} + 1)^2$ siendo $N_{\text{MIN}} \leq N$ típicamente un orden más pequeño que el de la representación HOA original. Con el fin de decorrelacionar estas secuencias de coeficientes HOA, se pueden transformar en señales de altavoces virtuales que inciden desde algunas direcciones predefinidas $N_{\text{MIN}}, d, d = 1, \dots, O_{\text{MIN}}$ (en analogía con el concepto descrito en la sección Normalización de la representación HOA de entrada). Definiendo el vector de todas las secuencias de coeficientes del componente HOA ambiental con índice $n \leq N_{\text{MIN}}$ de orden por $\mathbf{c}_{\text{AMB},\text{MIN}}(t)$ y la matriz de modos con respecto a las direcciones virtuales $\boldsymbol{\Omega}_{\text{MIN}, d}, d = 1, \dots, O_{\text{MIN}}$, por $\boldsymbol{\Psi}_{\text{MIN}}$, el vector de todas las señales de altavoces virtuales (definidas por) $\mathbf{w}_{\text{MIN}}(t)$ se obtiene mediante

$$\mathbf{w}_{\text{MIN}}(t) = \boldsymbol{\Psi}_{\text{MIN}}^{-1} \cdot \mathbf{c}_{\text{AMB},\text{MIN}}(t) \quad (35)$$

Por consiguiente, utilizando la compatibilidad de la matriz euclidiana y las normas vectoriales,

$$\|\mathbf{w}_{\text{MIN}}(lT_s)\|_{\infty} \leq \|\mathbf{w}_{\text{MIN}}(lT_s)\|_2 \quad (36)$$

$$\stackrel{(35)}{\leq} \|\Psi_{\text{MIN}}^{-1}\|_2 \cdot \|\mathbf{c}_{\text{AMB,MIN}}(lT_s)\|_2 \quad (37)$$

$$\stackrel{(34)}{\leq} \|\Psi_{\text{MIN}}^{-1}\|_2 \cdot \sqrt{K} \cdot O \quad (38)$$

En el documento N14264 de MPEG antes mencionado, las direcciones virtuales $\Omega_{\text{MIN}, d}$, $d = 1, \dots, O_{\text{MIN}}$, se eligen de acuerdo con el artículo de Fliege et al.. Las respectivas normas euclidianas de la inversa de las matrices modales Ψ_{MIN} se ilustran en la figura 4 para órdenes $N_{\text{MIN}} = 1, \dots, 9$. Puede observarse que

$$\|\Psi_{\text{MIN}}^{-1}\|_2 < 1 \text{ for } N_{\text{MIN}} = 1, \dots, 9 \quad (39)$$

Sin embargo, esto en general no es válido para $N_{\text{MIN}} > 9$, donde los valores de $\|\Psi_{\text{MIN}}^{-1}\|_2$ son típicamente mucho mayores que '1'. Sin embargo, al menos para $1 \leq N_{\text{MIN}} \leq 9$ las amplitudes de las señales de los altavoces virtuales están limitadas por

$$\|\mathbf{w}_{\text{MIN}}(lT_s)\|_{\infty} \stackrel{(38), \text{Fig. 4}}{\leq} \sqrt{K} \cdot O \text{ for } 1 \leq N_{\text{MIN}} \leq 9 \quad (40)$$

Al restringir la representación HOA de entrada para satisfacer la condición (6), que requiere que las amplitudes de las señales de altavoz virtuales creadas a partir de esta representación HOA no superen un valor de '1', se puede garantizar que las amplitudes de las señales antes del control de ganancia no excederán el valor $\sqrt{K} \cdot O$ (véanse ecuaciones (25), (34) y (40)) bajo las siguientes condiciones:

a) El vector de todas las señales $x(t)$ de sonido predominantes se computa de acuerdo con las ecuaciones/restricciones (18), (19) y (20);

b) El orden mínimo N_{MIN} , que determina el número O_{MIN} de primeras secuencias de coeficientes del componente HOA ambiental al que se aplica una transformada espacial, tiene que ser inferior a '9', si se usan como posiciones de altavoces virtuales las definidas en el artículo antes mencionado de Fliege et al.

Se puede concluir adicionalmente que las amplitudes de las señales antes del control de ganancia no excederán el valor $\sqrt{K_{\text{MAX}}} \cdot O$ para cualquier orden N hasta un orden máximo N_{MAX} de interés, es decir, que $1 \leq N \leq N_{\text{MAX}}$, donde

$$K_{\text{MAX}} = \max_{1 \leq N \leq N_{\text{MAX}}} K\left(N, \Omega_1^{(N)}, \dots, \Omega_O^{(N)}\right) \quad (41a)$$

En particular, se puede concluir de la figura 3 que, si se asume que se eligen las direcciones $\Omega_j^{(N)}$ de altavoz virtual, $1 \leq j \leq O$, para la transformación espacial inicial, de acuerdo con la distribución en el artículo de Fliege et al., y si además se asume que el orden máximo de interés va a ser $N_{\text{MAX}} = 29$ (como, por ejemplo, en el documento N14264 de MPEG), entonces las amplitudes de las señales antes del control de ganancia no excederán el valor 1,5 O , ya que $\sqrt{K_{\text{MAX}}} < 1.5$ en este caso especial. Es decir, que se puede seleccionar $\sqrt{K_{\text{MAX}}} = 1.5$.

K_{MAX} depende del orden N_{MAX} de interés máximo y de las direcciones $\Omega_j^{(N)}$ de los altavoces virtuales, $1 \leq j \leq O$, que puede expresarse mediante

$$K_{\text{MAX}} = K_{\text{MAX}}\left(\left\{\Omega_1^{(N)}, \dots, \Omega_O^{(N)} \mid 1 \leq N \leq N_{\text{MAX}}\right\}\right) \quad (41b)$$

Por consiguiente, la ganancia mínima aplicada por el control de ganancia para garantizar que las señales antes de la codificación perceptual se encuentren dentro del intervalo $[-1, 1]$ viene dada por $2^{e_{\text{MIN}}}$, donde

$$e_{\text{MIN}} = -\lceil \log_2(\sqrt{K_{\text{MAX}}} \cdot O) \rceil < 0 \quad (41c)$$

En caso de que las amplitudes de las señales antes del control de ganancia sean demasiado pequeñas, se propone, en el documento N14264 de MPEG, que es posible amplificarlas suavemente con un factor de hasta $2^{e_{\text{MAX}}}$, donde $e_{\text{MAX}} \geq 0$ se transmite como información secundaria dentro la representación codificada de HOA.

De este modo, cada exponente de base '2', que describe el cambio de amplitud absoluta total dentro de una unidad de acceso de una señal modificada causada por la unidad de procesamiento de control de ganancia desde la primera trama hasta la trama actual, puede asumir cualquier valor entero dentro del intervalo $[e_{\text{MIN}}, e_{\text{MAX}}]$. En consecuencia, el número (entero más bajo) β_e de bits necesarios para codificarlo viene dado por

$$\beta_e = \lceil \log_2(|e_{\text{MIN}}| + e_{\text{MAX}} + 1) \rceil = \lceil \log_2(\lceil \log_2(\sqrt{K_{\text{MAX}}} \cdot O) \rceil + e_{\text{MAX}} + 1) \rceil \quad (42)$$

En caso de que las amplitudes de las señales antes del control de ganancia no sean demasiado pequeñas, la ecuación (42) se puede simplificar:

$$\beta_e = \lceil \log_2(|e_{\text{MIN}}| + 1) \rceil = \lceil \log_2(\lceil \log_2(\sqrt{K_{\text{MAX}}} \cdot O) \rceil + 1) \rceil \quad (42a)$$

Este número β_e de bits se puede calcular en la entrada de los pasos/las etapas 15, ..., 151 de control de ganancia.

El uso de este número β_e de bits para el exponente garantiza que todos los posibles cambios de amplitud absoluta originados por las unidades 15, ..., 151 de procesamiento de control de ganancia del compresor HOA puedan capturarse, permitiendo el inicio de la descompresión en algunos puntos de entrada predefinidos. dentro de la representación comprimida.

Al iniciar la descompresión de la representación HOA comprimida en el descompresor HOA, los valores de ganancia no diferencial que representan los cambios de amplitud absoluta total asignados a la información secundaria para algunas tramas de datos y recibidos desde el demultiplexor 21 fuera de la corriente \tilde{B} de datos recibida se utilizan en los pasos o las etapas 24, ..., 241 de control de ganancia inverso para aplicar un control de ganancia correcto, de manera inversa al procesamiento que se llevó a cabo en los pasos/las etapas 15, ..., 151 de control de ganancia.

Realización adicional

Al implantar un sistema de compresión/descompresión HOA particular como se describe en las secciones Compresión HOA, Codificación HOA espacial, Descompresión HOA y Decodificación HOA espacial, la cantidad β_e de bits para la codificación del exponente tiene que establecerse de acuerdo con la ecuación (42) en dependencia de un factor $K_{\text{MAX,DES}}$ de escala, el cual, a su vez, depende de un orden máximo deseado $N_{\text{MAX,DES}}$ de las representaciones HOA a comprimir y de ciertas direcciones $\Omega_{\text{DES},1}^{(N)}, \dots, \Omega_{\text{DES},O}^{(N)}$ de los altavoces virtuales, $1 \leq N \leq N_{\text{MAX}}$.

Por ejemplo, al asumir $N_{\text{MAX,DES}} = 29$ y elegir las direcciones virtuales de altavoces de acuerdo con el artículo de Fliege et al., una elección razonable sería $\sqrt{K_{\text{MAX,DES}}} = 1.5$. En esa situación, se garantiza la compresión correcta para las representaciones HOA de orden N con $1 \leq N \leq N_{\text{MAX}}$ que están normalizadas de acuerdo con la sección Normalización de la representación HOA de entrada utilizando las mismas direcciones $\Omega_{\text{DES},1}^{(N)}, \dots, \Omega_{\text{DES},O}^{(N)}$ de altavoces virtuales. Sin embargo, esta garantía no se puede dar en el caso de una representación HOA que también (por razones de eficiencia) esté representada de manera equivalente por señales de altavoces virtuales en formato PCM, pero donde las direcciones $\Omega_j^{(N)}$, $1 \leq j \leq O$, de los altavoces virtuales se eligen para que sean diferentes a las direcciones $\Omega_{\text{DES},1}^{(N)}, \dots, \Omega_{\text{DES},O}^{(N)}$ de los altavoces virtuales, asumidas en la etapa de diseño de sistema.

Debido a esta elección diferente de posiciones de los altavoces virtuales, aunque las amplitudes de estas señales de los altavoces virtuales se encuentran dentro del intervalo $[1, 1]$, ya no se puede garantizar que las amplitudes de las señales antes del control de ganancia no superen el valor $\sqrt{K_{\text{MAX,DES}}} \cdot O$. Y, por consiguiente, no se puede garantizar que esta representación HOA tenga la normalización adecuada para la compresión de acuerdo con el procesamiento descrito en el documento N14264 de MPEG.

En esta situación es ventajoso tener un sistema que proporcione, en base al conocimiento de las posiciones de los altavoces virtuales, la amplitud máxima permitida de las señales de los altavoces virtuales con el fin de garantizar que la representación HOA respectiva sea adecuada para la compresión de acuerdo con el procesamiento descrito en el documento N14264 de MPEG.

En la figura 5 se ilustra tal sistema. Toma como entrada las posiciones $\Omega_j^{(N)}$ de los altavoces virtuales, $1 \leq j \leq O$, donde $O = (N + 1)^2$ con $N \in \mathbb{N}_0$, y proporciona como salida la amplitud máxima permitida γ_{dB} (medida en decibeles) de las señales de altavoces virtuales. En el paso o la etapa 51, la matriz Ψ de modo con respecto a las posiciones de altavoces virtuales se computa de acuerdo con la ecuación (3). En el paso o la etapa 52

siguiente se computa la norma euclidiana $\|\Psi\|_2$ de la matriz modal. En un tercer paso o tercera etapa 53, la amplitud γ se computa como el mínimo de '1' y el cociente entre el producto de la raíz cuadrada del número de posiciones de altavoces virtuales y $K_{\text{MAX,DES}}$ y la norma euclidiana de la matriz de modos. es decir, que

$$\gamma = \min \left(1, \frac{\sqrt{O} \cdot \sqrt{K_{\text{MAX,DES}}}}{\|\Psi\|_2} \right) \quad (43)$$

El valor en decibelios se obtiene mediante

$$\gamma_{\text{dB}} = 20 \log_{10}(\gamma) \quad (44)$$

Para explicación: de las derivaciones anteriores se puede ver que si la magnitud de las secuencias de coeficientes HOA no excede un valor $\sqrt{K_{\text{MAX,DES}}} \cdot O$, es decir, que si

$$\|c(lT_s)\|_{\infty} \leq \sqrt{K_{\text{MAX,DES}}} \cdot O \quad (45)$$

todas las señales antes de las unidades 15, 151 de procesamiento de control de ganancia no excederán, en consecuencia, este valor, que es el requisito para una adecuada compresión HOA.

A partir de la ecuación (9) se encuentra que la magnitud de las secuencias de coeficientes HOA está limitada por

$$\|c(lT_s)\|_{\infty} \leq \|c(lT_s)\|_2 \leq \|\Psi\|_2 \cdot \|w(lT_s)\|_2 \quad (46)$$

En consecuencia, si γ se establece de acuerdo con la ecuación (43) y las señales de altavoces virtuales en formato PCM satisfacen que

$$\|w(lT_s)\|_{\infty} \leq \gamma \quad (47)$$

de la ecuación (7) se deduce que

$$\|w(lT_s)\|_2 \leq \gamma \cdot \sqrt{O} \quad (48)$$

y que se satisface el requisito (45). Es decir, que el valor de magnitud máxima de '1' en la ecuación (6) se reemplaza por el valor de magnitud máxima γ en la ecuación (47).

35 *Conceptos básicos de la ambisónica de orden superior*

La ambisónica de orden superior (HOA) se basa en la descripción de un campo de sonido dentro de un área compacta de interés, que se asume que está libre de fuentes de sonido. En ese caso, el comportamiento espaciotemporal de la presión $p(t, x)$ de sonido en el tiempo t y en la posición x dentro del área de interés está físicamente determinado en su totalidad por la ecuación de onda homogénea. A continuación se asume un sistema de coordenadas esférico como el que se muestra en la figura 6. En el sistema de coordenadas utilizado, el eje x apunta a la posición delantera, el eje y apunta a la izquierda, y el eje z apunta a la parte superior. Una posición $x = (r, \theta, \phi)^T$ en el espacio está representada por un radio $r > 0$ (es decir, la distancia al origen de las coordenadas), un ángulo $\theta \in [0, \pi]$ de inclinación medido desde el eje polar z , y un ángulo $\phi \in [0, 2\pi[$ de acimut medido en sentido antihorario en el plano $x - y$ desde el eje x . Además, $(\cdot)^T$ denota la transposición.

Entonces, se puede demostrar a partir del libro de texto "Fourier Acoustics" que la transformada de Fourier de la presión de sonido con respecto al tiempo denotada por $\mathcal{F}_t(\cdot)$, es decir, que

$$P(\omega, x) = \mathcal{F}_t(p(t, x)) = \int_{-\infty}^{\infty} p(t, x) e^{-i\omega t} dt \quad (49)$$

donde ω denota la frecuencia angular e i indica la unidad imaginaria, se puede expandir a la serie de armónicos esféricos de acuerdo con

$$P(\omega = kc_s, r, \theta, \phi) = \sum_{n=0}^N \sum_{m=-n}^n A_n^m(k) j_n(kr) S_n^m(\theta, \phi) \quad (50)$$

donde c_s denota la velocidad del sonido y k denota el número de onda angular, que está relacionado con la frecuencia angular ω por $k = \omega/c_s$. Además, $j_n(\cdot)$ denota las funciones esféricas de Bessel de primer tipo y $S_n^m(\theta, \phi)$ denota los armónicos esféricos de valor real de orden n y grado m , que se definen en la sección

Definición de Armónicos Esféricos de valor real. Los coeficientes $A_n^m(k)$ de expansión sólo dependen del número k de onda angular. Obsérvese que se ha asumido implícitamente que la presión de sonido está limitada espacialmente por una banda. De este modo, la serie se trunca con respecto al índice n de orden en un límite superior N , que se denomina orden de la representación HOA.

5

Si el campo de sonido está representado por una superposición de un número infinito de ondas planas armónicas de diferentes frecuencias angulares ω que llegan desde todas las direcciones posibles especificadas por la tupla (θ, ϕ) de ángulos, se puede mostrar (véase B. Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution", J. Acoust. Soc. Am., vol. 4(116), páginas 2149-2157, octubre de 2004) que la respectiva función $C(\omega, \theta, \phi)$ de amplitud compleja de onda plana se puede expresar mediante la siguiente expansión de armónicos esféricos

10

$$C(\omega = kc_s, \theta, \phi) = \sum_{n=0}^N \sum_{m=-n}^n C_n^m(k) S_n^m(\theta, \phi) , \quad (51)$$

15

donde los coeficientes $C_n^m(k)$ de expansión están relacionados con los coeficientes $A_n^m(k)$ de expansión por

$$A_n^m(k) = i^n C_n^m(k) . \quad (52)$$

20

Asumiendo que los coeficientes individuales $C_n^m(k = \omega/c_s)$ sean funciones de la frecuencia angular ω , la aplicación de la transformada inversa de Fourier (denotada por $\mathcal{F}^{-1}(\cdot)$) proporciona funciones en el dominio tiempo

$$c_n^m(t) = \mathcal{F}_t^{-1}(C_n^m(\omega/c_s)) = \frac{1}{2\pi} \int_{-\infty}^{\infty} C_n^m\left(\frac{\omega}{c_s}\right) e^{i\omega t} d\omega \quad (53)$$

25

para cada orden n y grado m . Estas funciones en el dominio tiempo se denominan aquí secuencias de coeficientes HOA de tiempo continuo, que se pueden recopilar en un único vector $c(t)$ mediante

$$c(t) = \begin{bmatrix} c_0^0(t) & c_1^{-1}(t) & c_1^0(t) & c_1^1(t) & c_2^{-2}(t) & c_2^{-1}(t) & c_2^0(t) & c_2^1(t) & c_2^2(t) & \dots & c_N^{N-1}(t) & c_N^N(t) \end{bmatrix}^T \quad (54)$$

30

El índice de posición de una secuencia $c_n^m(t)$ de coeficientes HOA dentro del vector $c(t)$ viene dado por $n(n+1) + 1 + m$. El número global de elementos en el vector $c(t)$ viene dado por $O = (N+1)^2$. El formato final de ambisónica proporciona la versión muestreada de $c(t)$ utilizando una frecuencia f_s de muestreo como

$$\{c(lT_s)\}_{l \in \mathbb{N}} = \{c(T_s), c(2T_s), c(3T_s), c(4T_s), \dots\} \quad (55)$$

35

donde $T_s = 1/f_s$ denota el período de muestreo. Los elementos de $c(lT_s)$ se denominan secuencias de coeficientes HOA de tiempo discreto, y se puede demostrar que tienen siempre valores reales. Esta propiedad también se sostiene para las versiones $c_n^m(t)$ de tiempo continuo.

40

Definición de armónicos esféricos de valor real

Los armónicos esféricos $S_n^m(\theta, \phi)$ de valor real (asumiendo la normalización SN3D de acuerdo con J. Daniel, "Représentation de champs acoustiques, application à la Transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia", tesis doctoral, Université Paris, 6, 2001, capítulo 3.1) vienen dados por

45

$$S_n^m(\theta, \phi) = \sqrt{(2n+1) \frac{(n-|m|)!}{(n+|m|)!}} P_{n,|m|}(\cos\theta) \text{trg}_m(\phi) \quad (56)$$

siendo

$$\text{trg}_m(\phi) = \begin{cases} \sqrt{2} \cos(m\phi) & m > 0 \\ 1 & m = 0 \\ -\sqrt{2} \sin(m\phi) & m < 0 \end{cases} . \quad (57)$$

50

Las funciones $P_{n,m}(x)$ de Legendre asociadas se definen como

$$P_{n,m}(x) = (1 - x^2)^{m/2} \frac{d^m}{dx^m} P_n(x), m \geq 0 \quad (58)$$

5 con el polinomio $P_n(x)$ de Legendre, y, a diferencia de E.G. Williams, "Fourier Acoustics", vol. 93 de Applied Mathematical Sciences, Academic Press, 1999, sin el término $(-1)^m$ de fase de Condon-Shortley.

10 El procesamiento inventivo se puede llevar a cabo mediante un único procesador o circuito electrónico, o mediante varios procesadores o circuitos electrónicos que funcionen en paralelo y/o que funcionen en diferentes partes del procesamiento inventivo.

Las instrucciones para hacer funcionar el procesador o los procesadores pueden almacenarse en una o más memorias. El objeto para el cual se busca protección está definido de manera única por el juego de reivindicaciones adjunto.

REIVINDICACIONES

1. Un método para decodificar una representación de sonido de ambisónica de orden superior (HOA) comprimida de un sonido o campo de sonido, comprendiendo el método:

recibir y demultiplexar una corriente de bits que contiene la representación HOA comprimida para proporcionar una representación codificada perceptualmente de señales I y datos de información secundaria codificados que describen cómo crear una representación HOA de la misma; decodificar, en un decodificador perceptual, la representación codificada perceptualmente de las señales I para determinar, para una trama de datos HOA actual (k), señales decodificadas perceptualmente $\hat{z}_i(k)$, $i = 1, \dots, I$, y decodificar, en un decodificador de fuente de información secundaria, los datos de información secundaria codificados para determinar, para la trama de datos HOA actual (k), datos de información secundaria decodificados que dan como resultado el exponente $e_i(k)$ de corrección de ganancia asociado, el indicador $\beta_i(k)$ de excepción de corrección de ganancia, conjuntos $M_{DIR(k+1)}$ y $M_{VEC(k+1)}$ de datos de tupla, los parámetros $\zeta(k+1)$ de predicción y un vector $V_{AMB, ASSIGN}(k)$ de asignación, y valores de ganancia no diferenciales, representados como un exponente en base dos y que representan cambios de amplitud absoluta total, desde una primera trama HOA hasta una trama HOA actual, asignados a la información secundaria de la trama HOA actual para aplicar un control de ganancia correcto para las señales de canal de la trama HOA actual, donde cada uno de los valores de ganancia no diferenciales está representado por un número entero más bajo de bits;

proporcionar tramas $\hat{y}_i(k)$, $i = 1, \dots, I$ de señal corregidas de ganancia, realizando un procesamiento de control de ganancia inverso para las señales perceptualmente decodificadas $\hat{z}_i(k)$, $i = 1, \dots, I$, en base a los valores de ganancia no diferenciales, el exponente $e_i(k)$ de corrección de ganancia asociado y el indicador $\beta_i(k)$ de excepción de corrección de ganancia,

redistribuir las tramas $\hat{y}_i(k)$, $i = 1, \dots, I$ de señal corregidas de ganancia, durante la reasignación de canales utilizando los conjuntos $M_{DIR(k+1)}$ y $M_{VEC(k+1)}$ de datos de tupla, y el vector $V_{AMB, ASSIGN}(k)$ de asignación, con el fin de reconstruir una trama $\hat{X}_{PS}(k)$ de todas las señales de sonido predominantes y una trama $C_{I, AMB}(k)$ de una representación intermedia de un componente HOA ambiental,

proporcionar un conjunto $J_{AMB, ACT}(k)$ de índices de secuencias de coeficientes del componente HOA ambiental activo en la k -ésima trama, y conjuntos $J_E(k-1)$, $J_D(k-1)$ y $J_U(k-1)$ de datos de los índices de coeficientes del componente ambiental HOA,

computar una representación HOA del componente de sonido predominante $\hat{C}_{PS}(k-1)$ a partir de la trama $\hat{X}_{PS}(k)$ de todas las señales de sonido predominantes utilizando el conjunto $M_{DIR(k+1)}$ de tuplas, el conjunto $\zeta(k+1)$ de parámetros de predicción, el conjunto $M_{VEC(k+1)}$ de tuplas y los conjuntos $J_E(k-1)$, $J_D(k-1)$ y $J_U(k-1)$ de datos,

crear una trama $\hat{C}_{AMB}(k-1)$ de componente HOA ambiental a partir de la trama $C_{I, AMB}(k)$ de la representación intermedia del componente HOA ambiental, utilizando el conjunto $J_{AMB, ACT}(k)$ de índices de secuencias de coeficientes del componente HOA ambiental que están activos en la k -ésima trama, introduciendo un retardo de una trama debido a la sincronización con el componente de sonido predominante,

superponer la trama $\hat{C}_{AMB}(k-1)$ de componente HOA ambiental y la trama $\hat{C}_{PS}(k-1)$ del componente HOA de sonido predominante para proporcionar la trama HOA decodificada $\hat{C}(k-1)$, y

crear a partir de las señales I y de la información secundaria la representación HOA reconstruida,

donde el número entero más bajo β_e de bits aplicado para representar los valores de ganancia no diferencial se puede determinar mediante

$$\beta_e = \lceil \log_2(\lceil \log_2(\sqrt{K_{MAX}} \cdot O) \rceil + 1) \rceil ,$$

donde $K_{MAX} = \max_{1 \leq N \leq N_{MAX}} K(N, \Omega_1^{(N)}, \dots, \Omega_O^{(N)})$, N es el orden, N_{MAX} es el orden máximo de interés, $\Omega_1^{(N)}, \dots, \Omega_O^{(N)}$ son direcciones de altavoces virtuales, $O = (N+1)^2$ es el número de secuencias de coeficientes HOA, y K es una relación entre la norma euclidiana al cuadrado $\|\Psi\|_2^2$ de una matriz de modos y O , donde Ψ es una matriz de modos con respecto a dichas direcciones, y donde $\sqrt{K_{MAX}} = 1.5$.

2. Producto de programa de ordenador que comprende instrucciones que, cuando se llevan a cabo en un ordenador, realizan el método de la reivindicación 1.

3. Medio de almacenamiento que almacena el producto de programa de ordenador de la reivindicación 2.

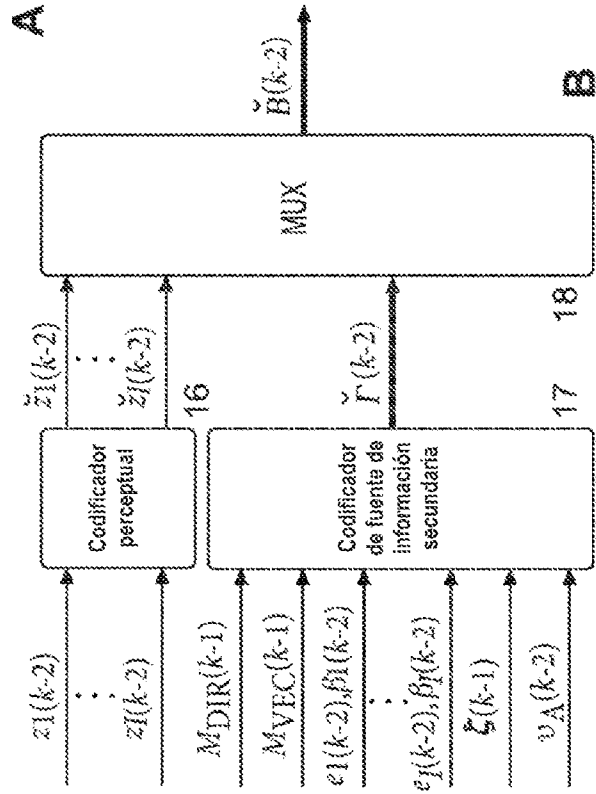
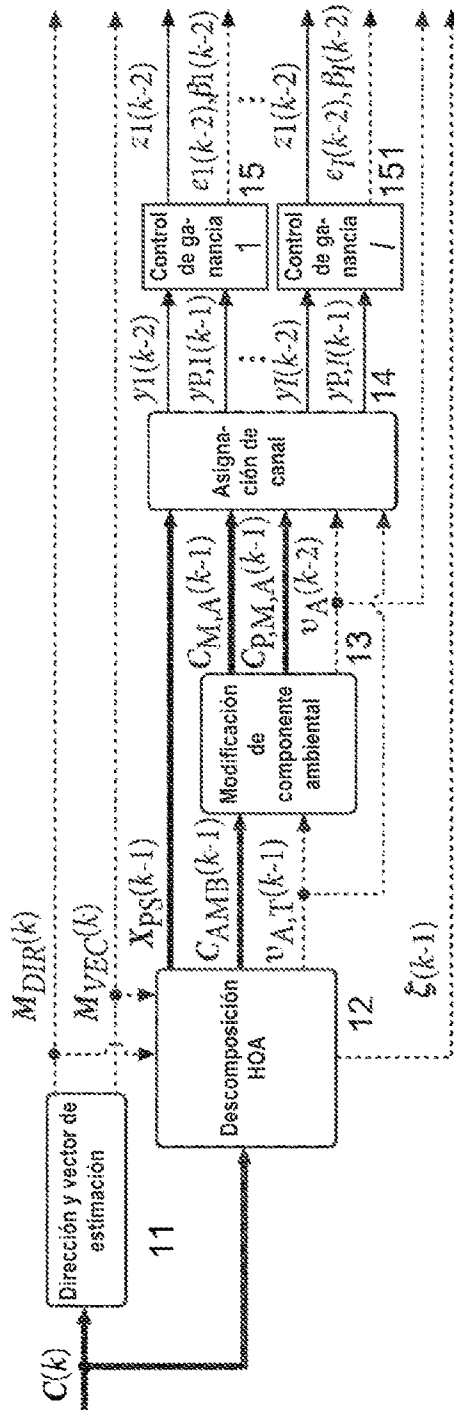


Fig. 1

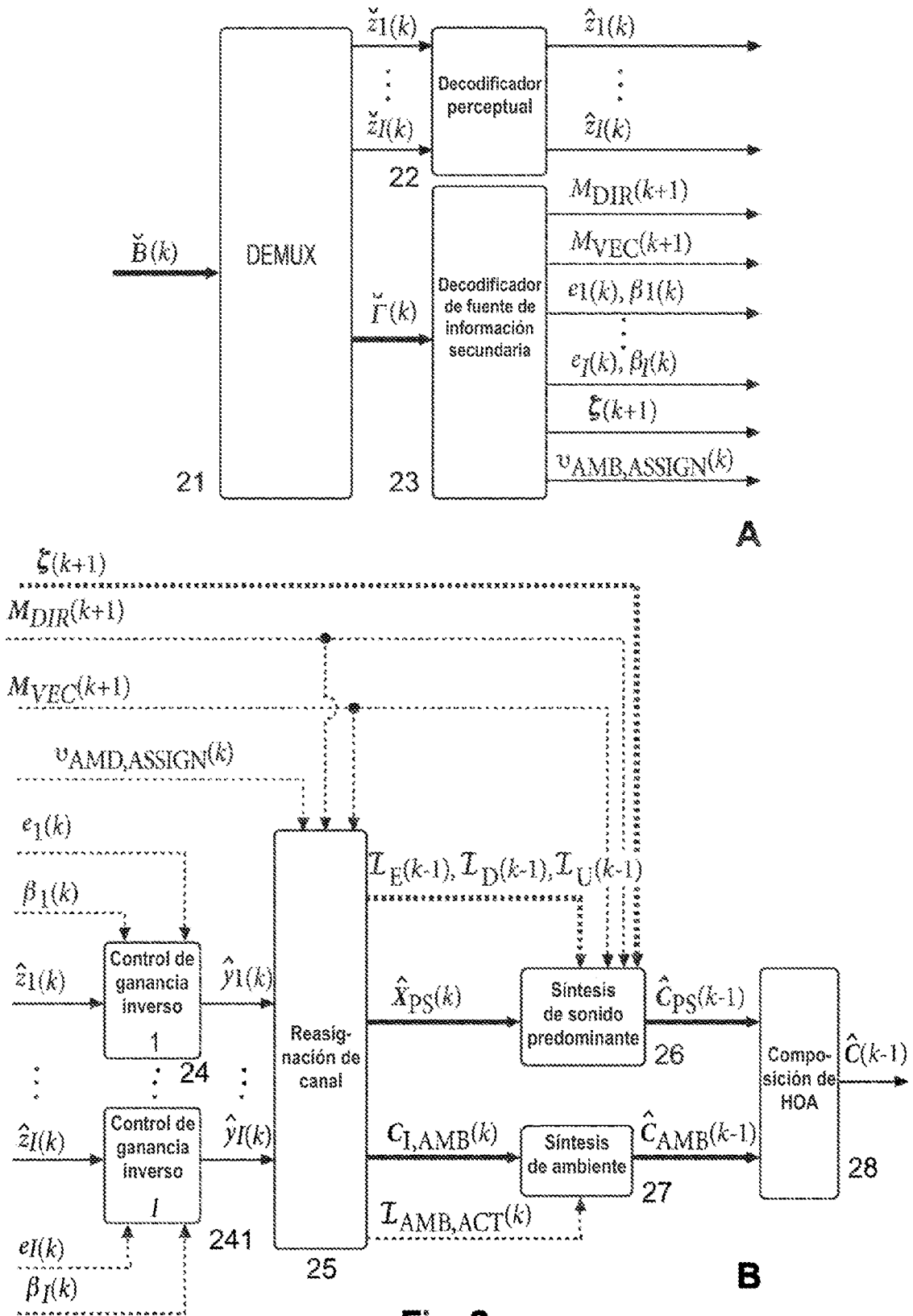


Fig. 2

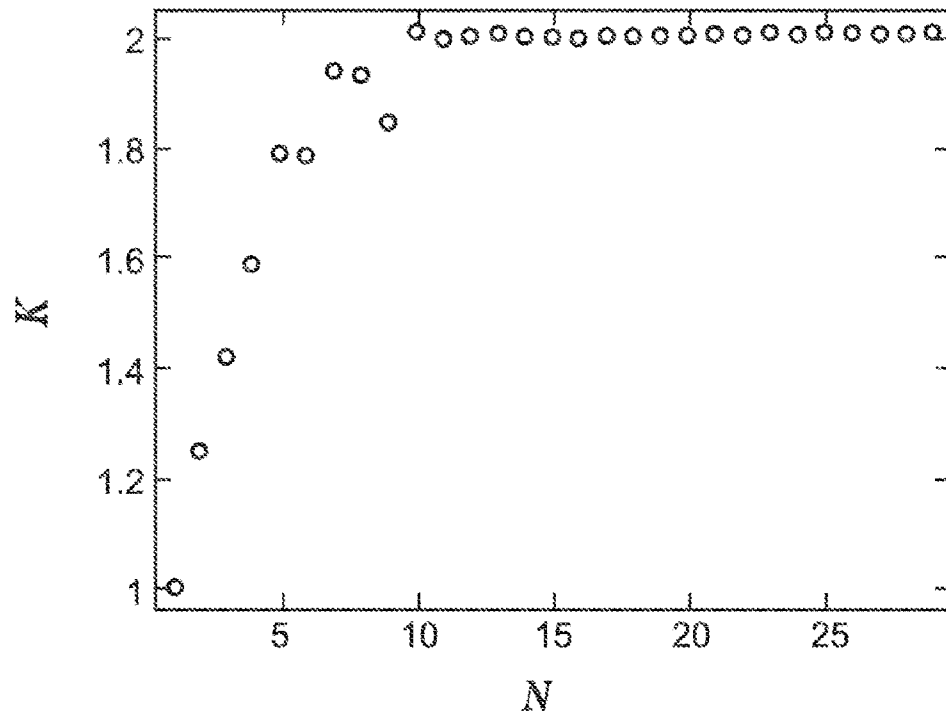


Fig. 3

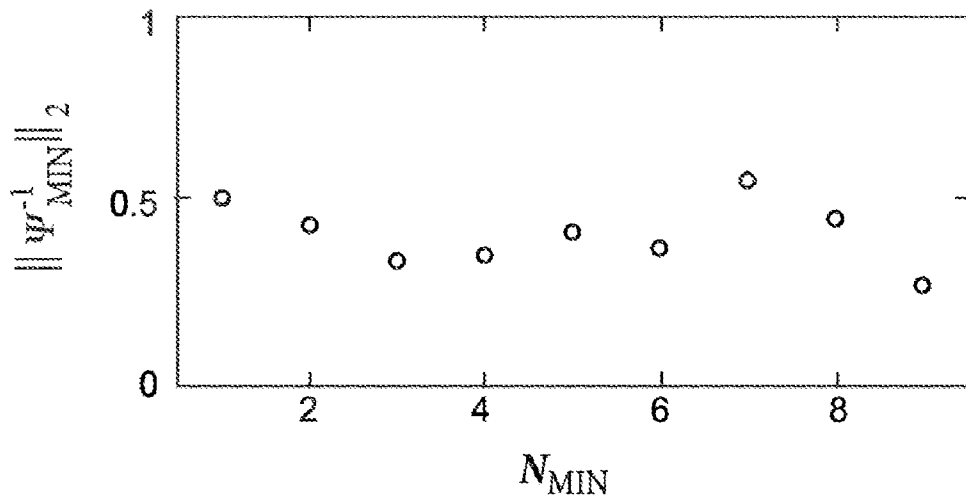


Fig. 4

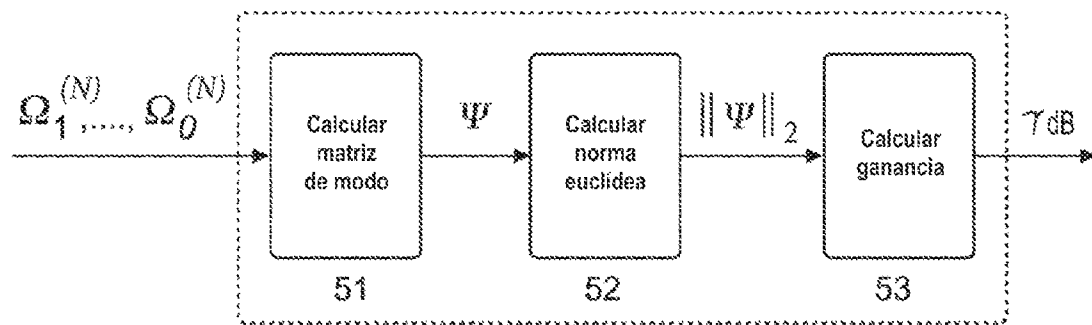


Fig. 5

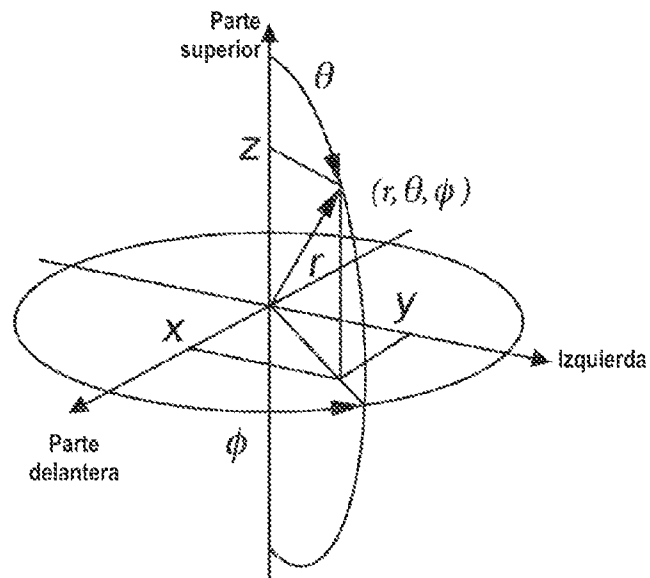


Fig. 6