



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 305 852**

51 Int. Cl.:
G10L 19/14 (2006.01)
G10L 19/00 (2006.01)
H04N 7/26 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Número de solicitud europea: **04775644 .0**
86 Fecha de presentación : **06.10.2004**
87 Número de publicación de la solicitud: **1673764**
87 Fecha de publicación de la solicitud: **28.06.2006**

54 Título: **Procedimiento de codificación de una señal digital en un flujo binario escalable, procedimiento para la decodificación de un flujo binario escalable.**

30 Prioridad: **10.10.2003 US 510629 P**

45 Fecha de publicación de la mención BOPI:
01.11.2008

45 Fecha de la publicación del folleto de la patente:
01.11.2008

73 Titular/es:
Agency for Science, Technology and Research
20 Biopolis Way, No. 07-01 Centros
Singapore 138668, SG

72 Inventor/es: **Yu, Rongshan;**
Lin, Xiao y
Rahardja, Susanto

74 Agente: **Carpintero López, Mario**

ES 2 305 852 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

Procedimiento de codificación de una señal digital en un flujo binario escalable, procedimiento para la descodificación de un flujo binario escalable.

5

Antecedentes de la invención

Recientemente, con los avances en los ordenadores, el funcionamiento en red y las comunicaciones que portan flujos con contenidos de audio sobre redes tales como Internet, redes de área local sin hilos, redes domésticas y sistemas de telefonía celular comercial se está convirtiendo en un medio predominante para ofrecer servicios de audio. Se piensa que con el progreso de las infraestructuras de red de banda ancha, incluyendo xDSL, las fibras ópticas y el acceso sin hilos de banda ancha, las velocidades binarias para estos canales se aproximarán de una manera rápida a aquéllas para la entrega de señales de audio sin pérdidas de alta velocidad de muestreo y alta resolución de amplitud (por ejemplo, 96 kHz, 24 bit/muestra). Por otra parte, aún existen áreas de aplicación en las que se requieren los formatos de audio digital de alta compresión tales como MPEG-4 AAC (descritos en [1]). Como resultado de esto, son altamente demandadas soluciones interfuncionales y que sirven de puente entre los canales actuales y los canales de banda ancha que emergen con rapidez. Además, incluso cuando los canales de banda ancha se encuentren ampliamente disponibles y las restricciones de ancho de banda se eliminan por última vez, aún es altamente favorable un sistema de codificación de velocidad binaria escalable que sea capaz de producir un flujo binario jerárquico cuyas velocidades binarias se puedan cambiar de manera dinámica durante la transmisión. Por ejemplo, para aplicaciones en las que ocurren pérdidas de paquetes de manera ocasional debido a accidentes o debido a requisitos de compartición de recursos, las representaciones de la forma de onda de banda ancha actuales tales como PCM (Modulación de Impulsos Codificados), (del inglés *Pulse Code Modulation*) y los formatos de codificación sin pérdidas pueden sufrir serias distorsiones en una situación de flujo binario. Sin embargo, este problema se puede resolver si uno pudiese fijar las prioridades de paquete en el caso de que los recursos de red se cambien de manera dinámica. Finalmente, un sistema de codificación de velocidad binaria escalable proporciona también las ventajas del servidor para los servicios de flujo binario de audio, en los que se podría conseguir una degradación de la QoS digna si llega un número excesivo de peticiones desde los sitios de los clientes.

Se han propuesto con anterioridad muchos algoritmos de codificación de audio sin pérdidas (véase [2] - [8]). La mayoría de las aproximaciones dependen de un filtro de predicción para eliminar la redundancia de las señales de audio originales mientras que los residuos son codificados en entropía (como se describe en [5] - [12]). Debido a la existencia de los filtros predictivos, los flujos binarios generados por estas aproximaciones basadas en la predicción son difíciles y no son eficientes en (véase [5], [6]), si no imposible, para ser escalado para conseguir un escalado de la velocidad binaria. Otras aproximaciones tales como las que se describen en [3], construyen el codificador de audio sin pérdidas a través de una aproximación de dos capas en la que las señales de audio originales se codifican en primer lugar con un codificador con pérdidas y su residuo se codifica después sin pérdidas con un codificador residual. Aunque este diseño de dos capas proporciona alguna clase de escalado de la velocidad binaria, su granularidad es demasiado basta como para ser apreciada por aplicaciones de flujos binarios de audio. Los *códecs* de audio que proporcionan el escalado de grano fino sobre la velocidad binaria fueron propuestos previamente en [4] y en [18], sin embargo, a diferencia del sistema que se trata en este documento, estos *códecs* no proporcionan la compatibilidad hacia atrás en que los flujos binarios con pérdidas producidos por ambos *códecs* son incompatibles para cualquier *códec* de audio existente.

En [21], [22], [23] se describen modelos perceptuales. El *códec* de vídeo como el de [25] proporciona un escalado de grano fino así como compatibilidad hacia atrás, en donde la codificación de la señal residual se base en consideraciones independientes de la señal.

El objeto de la invención es proporcionar un procedimiento de codificación de una señal digital en un flujo binario escalable en el que la compatibilidad hacia atrás se pueda mantener y se pueda mejorar la calidad perceptual.

50

Sumario de la invención

Se proporciona un procedimiento para codificar una señal digital dentro de un flujo binario escalable, que comprende: la cuantificación de la señal digital y la codificación de la señal cuantificada para formar un flujo binario de capa de núcleo; la realización de una correspondencia de errores en base a la señal digital y al flujo binario de capa de núcleo para eliminar la información que haya sido codificada en el flujo binario de capa de núcleo, dando como resultado una señal de error; la codificación de plano binario de la señal de error en base a la información perceptual de la señal digital, dando como resultado un flujo binario de capa de mejora, en el que la información perceptual de la señal digital se determina usando un modelo perceptual; y

60

- multiplexar el flujo binario de capa de núcleo y el flujo binario de capa de mejora, generando por lo tanto el flujo binario escalable.

Además, se proporcionan un codificador para codificar una señal digital en un flujo binario escalable, un medio que sea legible por medio de un ordenador, un elemento de programa de ordenador, un procedimiento para la descodificación del flujo binario escalable en una señal digital, un descodificador para descodificar un flujo binario escalable en una señal digital, un medio adicional legible por un ordenador y un elemento adicional de programa de ordenador de acuerdo con el procedimiento descrito con anterioridad.

ES 2 305 852 T3

En una realización, se presenta el *códec* de audio sin pérdidas que consigue el escalado de la velocidad binaria de grano fino (FGBS, del inglés *Fine Grain Bit-Rate Scalability*) con las siguientes características:

- 5 - Compatibilidad hacia atrás: se incorpora en el flujo binario sin pérdidas un flujo binario de capa de núcleo de alta compresión, tal como un flujo binario MPEG-4 AAC.
- Flujo binario sin pérdidas perceptualmente incorporado: el flujo binario sin pérdidas se puede truncar a cualquier velocidad con pérdidas sin perder en la optimización perceptual en el audio reconstruido.
- 10 - Baja complejidad: añade solamente un cálculo muy limitado sobre AAC (*códec* aritmético binario) así como una memoria muy limitada.

15 La abundante funcionalidad proporcionada por el *códec* de audio presentado sugiere su capacidad para servir como un formato de audio “universal” para cumplir con los varios requisitos de velocidad/calidad para los diferentes flujos binarios de audio o aplicaciones de almacenamiento. Por ejemplo, un flujo binario que cumpla con MPEG-4 AAC que se use como el flujo binario de capa de núcleo se puede extraer fácilmente del flujo binario generado usando el *códec* para servicios de audio MPEG-4 AAC convencionales. Por otra parte, la comprensión sin pérdidas también es proporcionada por el *códec* para aplicaciones de edición de audio o para aplicaciones de almacenamiento con requisito de reconstrucción sin pérdidas. En aplicaciones de flujo binario de audio, en las que se necesite el FGBS, el flujo binario sin pérdidas del *códec* se puede truncar de manera adicional a velocidades binarias más bajas en el codificador/descodificador o en el canal de comunicaciones para cualquier restricción de la velocidad/fidelidad/complejidad que pueda surgir en sistemas prácticos.

25 En una realización, se proporciona un procedimiento de codificación de una señal digital para formar un flujo binario escalable, en el que el flujo binario escalable se puede truncar en cualquier punto para producir una señal de calidad más baja (con pérdidas) cuando sea descodificada por un descodificador. El procedimiento se puede usar para la codificación de cualquier tipo de señal digital, tal como señales de audio, señales de imagen o señales de vídeo. La señal digital, que corresponde a una señal física medida, puede ser generada por medio de la exploración de al menos un rasgo característico de una señal analógica correspondiente (por ejemplo, los valores de luminancia y de crominancia de una señal de vídeo, la amplitud de una señal de sonido analógica o la señal de detección analógica proveniente de un sensor). Por ejemplo, se puede usar un micrófono para capturar una señal de audio analógica, que después se convierte en una señal de audio digital por medio del muestreo y de la cuantificación de la señal de audio analógica capturada. Se puede usar una cámara de vídeo para capturar la señal de vídeo analógica, que después se convierte en una señal de vídeo digital usando un convertidor analógico a digital adecuado. De manera alternativa, se puede usar una cámara digital para capturar directamente la imagen o la señal de vídeo sobre un sensor de imágenes (CMOS o CCD) como señales digitales.

40 La señal digital se cuantifica y se codifica para formar un flujo binario de capa de núcleo. El flujo binario de capa de núcleo forma la velocidad binaria/calidad mínimas del flujo binario escalable.

45 Se usa un flujo binario de capa de mejora para proporcionar una velocidad binaria/calidad adicionales del flujo binario escalable. El flujo binario de capa de mejora se forma de acuerdo con la invención por medio de la realización de una correspondencia de errores en base a la señal transformada y al flujo binario de capa de núcleo para generar una señal de error. El propósito de realizar una correspondencia de errores es eliminar la información que ya ha sido codificada en el flujo binario de capa de núcleo.

50 La señal de error se codifica en plano binario para formar el flujo binario de capa de mejora. La codificación en plano binario de la señal de error se realiza en base a la información perceptual, es decir, la importancia percibida o perceptual de la señal digital. La información perceptual usada en esta presente invención se refiere a la información que está relacionada con el sistema sensorial humano, por ejemplo, el sistema visual humano (es decir, el ojo humano) y el sistema auditivo humano (es decir, el oído humano). Dicha información perceptual para la señal digital (vídeo o audio) se obtiene usando un modelo perceptual, por ejemplo el Modelo Psicoacústico I o II en el audio MPEG-1 (descrito en [21]), para las señales de audio, y el Modelo del Sistema Visual Humano para la imagen (descrito en [22]), y el Modelo Espacio-Temporal usado en vídeo (descrito en [23]).

55 El modelo psicoacústico se basa en el efecto de que el oído humano solamente es capaz de recoger sonidos dentro de una cierta banda de frecuencias dependiendo de varias condiciones ambientales. De manera similar, el HVM (modelo visual humano) (del inglés, *Human Visual Model*) se basa en el efecto de que el ojo humano está más atento a cierto movimiento, colores y contrastes.

60 El flujo binario de capa de núcleo y el flujo binario de capa de mejora se multiplexan para formar el flujo binario escalable.

65 El flujo binario escalable se puede descodificar para reconstruir sin pérdidas la señal digital. Como se ha mencionado con anterioridad, el flujo binario de capa de núcleo es un flujo binario incorporado que forma la mínima velocidad binaria/calidad del flujo binario escalable, y el flujo binario de capa de mejora forma la pérdida para la parte sin pérdidas del flujo binario escalable. Como el flujo binario de capa de mejora está de manera perceptual codificado

ES 2 305 852 T3

plano binario, el flujo binario de capa de mejora se puede truncar de una manera tal que los datos del flujo binario de capa de mejora que sean menos importantes desde el punto de vista perceptual se trunquen primero, para proporcionar el escalado perceptual del flujo binario escalable. En otras palabras, el flujo binario escalable se puede escalar por medio del truncado del flujo binario de capa de mejora, de forma que el flujo binario de capa de mejora y así el flujo binario escalable se puedan optimizar desde el punto de vista perceptual incluso cuando se trunquen a una velocidad binaria/calidad más bajas.

El procedimiento de acuerdo con la invención se puede usar como un codificador sin pérdidas para la señal digital, tal como una señal de imagen, de vídeo o de audio, en sistemas de alto ancho de banda o en sistemas de alta fidelidad. Cuando cambia el requisito de ancho de banda, la velocidad binaria del flujo binario generado por el codificador se puede cambiar de acuerdo con esto para cumplir con el cambio de los requisitos de ancho de banda. Dicho procedimiento se puede implementar en muchas aplicaciones y sistemas tales como audio MPEG, compresión de imagen y de vídeo de JPEG 2000.

De acuerdo con una realización de la invención, la señal digital se transforma en un dominio adecuado antes de ser cuantificada para formar la señal cuantificada. La señal digital se puede transformar dentro del mismo dominio o desde un dominio a otro dominio con el fin de representar mejor la señal digital, y por lo tanto permitir una cuantificación y una codificación fáciles y eficientes de la señal digital para formar el flujo binario de capa de núcleo. Dicho dominio puede incluir, pero no se limita a, el dominio en el tiempo, el dominio en la frecuencia y un híbrido de los dominios en el tiempo y en la frecuencia. La transformación de la señal digital se puede llevar a cabo incluso por medio de una matriz unitaria, I.

En una realización, la señal digital se transforma en una señal transformada usando una Transformada de Coseno Discreta Modificada entera (intMDCT). La intMDCT es una aproximación reversible al filtro de bancos de Transformada de Coseno Discreta Modificada (MDCT) que es comúnmente usada en un codificador MPEG-4 AAC. Se pueden usar otras transformadas para transformar la señal digital en un dominio adecuado para el procesamiento adicional, incluyendo, pero no limitándose a, Transformada de Coseno Discreta, Transformada de Seno Discreta, Transformada Rápida de Fourier y Transformada Discreta de Ondas Pequeñas.

Cuando se usa intMDCT para transformar la señal digital en la señal transformada, la señal transformada (de manera específica los coeficientes intMDCT que describen la señal transformada) está de manera preferible normalizada o escalada para aproximar la salida de un filtro de bancos MDCT. La normalización de la señal transformada intMDCT puede ser útil en el caso cuando un cuantificador para cuantificar la señal transformada, por ejemplo, un cuantificador AAC, tiene un filtro de bancos MDCT con una ganancia global diferente de la ganancia global del filtro de bancos intMDCT. Dicho proceso de normalización aproxima la señal transformada intMDCT al filtro de bancos MDCT de forma que es adecuada para ser cuantificada y codificada de manera directa por el cuantificador para formar el flujo binario de capa de núcleo.

Para codificar una señal digital de audio, la señal digital/transformada es de manera preferible cuantificada y codificada de acuerdo con la especificación MPEG AAC para generar el flujo binario de capa de núcleo. Esto es porque AAC es uno de los algoritmos de codificación de audio perceptuales más eficientes para generar una baja velocidad binaria pero un alto flujo binario de audio de alta calidad. Por lo tanto, el flujo binario de capa de núcleo generado usando AAC (al que se hace referencia como flujo binario AAC) tiene una baja velocidad binaria, e incluso cuando se trunca el flujo binario escalable al flujo binario de capa de núcleo, la calidad perceptual del flujo binario truncado aún sigue siendo alta. Se debería notar que también se pueden usar otros algoritmos/procedimientos de cuantificación y de codificación, por ejemplo, MPEG 1 Capa de Audio 3 (MP3) u otros procedimientos propietarios de codificación/cuantificación para generar el flujo binario de capa de núcleo.

La correspondencia de errores que elimina información que ya ha sido codificada en el flujo binario de capa de núcleo y que genera una señal residual (o señal de error) se realiza mediante la resta de umbral de cuantificación más bajo (más cercano a cero) de cada uno de los valores cuantificados de la señal cuantificada de la señal transformada. Dicho procedimiento de correspondencia de errores en base al umbral de cuantificación tiene la ventaja de que el los valores de la señal residual siempre son positivos, y la amplitud de la señal residual es independiente del umbral de cuantificación. Esto permite una baja complejidad y un esquema de codificación incorporado eficiente para su implementación. Si embargo, también es posible restar una señal transformada reconstruida de la señal transformada para generar la señal residual.

Para determinar la información perceptual de la señal digital para la codificación de plano binario de la señal de error, se puede usar un modelo psicoacústico como el modelo perceptual. El modelo psicoacústico puede estar basado en el Modelo Psicoacústico I o II usados en audio MPEG-1 (como se describe en [21]), o el Modelo Psicoacústico en audio MPEG-4 (como se describe en [19]). Cuando un cuantificador perceptual tal como el usado de acuerdo con AAC, se usa para cuantificar y para codificar la señal digital/transformada, el modelo perceptual usado en el cuantificador perceptual también se puede usar para determinar la información perceptual para la codificación de plano binario de la señal de error. En otras palabras, no se necesita un modelo perceptual independiente en este caso para proporcionar la información perceptual para la codificación de plano binario de la señal de error.

La información perceptual para la codificación de plano binario de la señal de error también es de manera preferible multiplexada con los flujos binarios de capa de núcleo y de capa de mejora para formar el flujo binario escalable

ES 2 305 852 T3

como información lateral. La información lateral se puede usar para reconstruir la señal de error por medio de un decodificador.

5 La señal de error está dispuesta en una pluralidad de planos binarios, teniendo cada uno de los planos binarios una pluralidad de símbolos de plano binario.

10 En una realización de la invención, la disposición o el orden de los planos binarios de la señal de error se cambia o se desplaza y los planos binarios son posteriormente explorados y codificados de una manera secuencial y consecutiva. Los planos binarios se desplazan de una manera tal que cuando se realiza la codificación del plano binario en los planos binarios desplazados, los planos binarios que comprenden los símbolos de plano binario más importantes desde el punto de vista perceptual, son explorados y codificados los primeros. En esta realización, todos los símbolos de plano binario de un plano binario están codificados antes de la codificación de los símbolos de plano binario de un plano binario adyacente posterior.

15 En otra realización de la invención, los símbolos de plano binario de los planos binarios son explorados y codificados en una secuencia basada en la información perceptual. En otras palabras, no todos los símbolos de plano binario en un plano binario se codifican antes de la codificación de los símbolos de plano binario de otro plano binario. La secuencia de exploración y de codificación de los símbolos de plano binario de la pluralidad de planos binarios se determina en base a la información perceptual de forma que los símbolos de plano binario que sean los más importantes desde el punto de vista perceptual son los que se codifican los primeros.

20 La información perceptual de la señal digital determinada por el modelo perceptual puede incluir el primer (o máximo) plano binario $M(s)$ (es decir, un número (índice) que especifica el primer plano binario) de la pluralidad de planos binarios para la codificación del plano binario de la señal de error, y/o el nivel de la Distorsión Apenas Perceptible (JND) (del inglés, *Just Noticeable Distortion*) de la señal digital. Se debería notar que la información perceptual se refiere a la señal digital para cada una de las diferentes características de dominio, (por ejemplo, la frecuencia, el tiempo, la amplitud de la señal, etc.) o un margen de características de dominio. Por ejemplo, cuando la señal digital se transforme al dominio de la frecuencia, los valores de la información perceptual de la señal digital en cada frecuencia o en una banda de frecuencias (banda de frecuencia s , o de manera más general, banda de dominio s) pueden ser diferentes indicando que la señal puede ser más importante desde el punto de vista perceptual a ciertas frecuencias.

25 En una realización de la invención, la significancia perceptual $P(s)$ de la señal digital correspondiente a cada banda de frecuencias s se determina como la información perceptual. En esta realización, se determina el nivel JND $\tau(s)$ de la señal digital correspondiente al plano binario de la señal de error. El plano binario correspondiente al nivel JND $\tau(s)$ se resta posteriormente del primer plano binario de la pluralidad de planos binarios para la codificación de plano binario de la señal de error $M(s)$ para dar como resultado la significancia perceptual $P_s(s)$. La significancia perceptual $P_s(s)$ se puede usar para controlar el desplazamiento de los planos binarios, de forma que los planos binarios que comprenden los símbolos de plano binario más importantes desde el punto de vista perceptual son explorados y codificados los primeros. De manera más ventajosa, la significancia perceptual $P_s(s)$ se puede usar para controlar la secuencia de exploración y de codificación de los símbolos de plano binario de una pluralidad de planos binarios de forma que los símbolos de plano binario que sean los más importantes desde el punto de vista perceptual son los que se codifican los primeros.

30 En una realización adicional de la invención, la significancia perceptual $P_s(s)$ se normaliza para formar una significancia perceptual $P_s'(s)$. En esta realización, se define una significancia perceptual común $P_{s_común}$ de la señal digital en base a una función de la significancia perceptual $P_s(s)$. Ejemplos de dicha función de la significancia perceptual $P_s(s)$ incluyen el valor promedio, el valor máximo, el valor mínimo o un valor normalizado de la significancia perceptual $P_s(s)$. La significancia perceptual común $P_{s_común}$ se resta de la significancia perceptual $P_s(s)$ para dar como resultado la significancia perceptual normalizada $P_s'(s)$ para cada banda de frecuencias s . Cuando la banda de frecuencias s contiene al menos una señal cuantificada de valor distinto de cero, la banda de frecuencias s es una banda significativa.

35 En cualquier otro caso, la banda de frecuencias s es una banda insignificante s . Para la banda significativa, el valor de la significancia perceptual $P_s(s)$ se fija al valor de la significancia perceptual común $P_{s_común}$. Para la banda insignificante, la significancia perceptual normalizada correspondiente $P_s'(s)$ se multiplexa con el flujo binario de capa de núcleo y el flujo binario de capa de mejora para generar el flujo binario escalable para la transmisión. La significancia perceptual normalizada $P_s'(s)$ se transmite en el flujo binario escalable como información lateral para la decodificación del flujo binario escalable en un decodificador.

40 La normalización de la significancia perceptual $P_s(s)$ por medio de la definición de una significancia perceptual común $P_{s_común}$ tiene la ventaja de reducir la cantidad de información perceptual que se ha de transmitir en el flujo binario escalable mediante la utilización de la información obtenida cuando se cuantificó la señal digital/transformada para generar el flujo binario de capa de núcleo. Por lo tanto, la información perceptual, en particular la significancia perceptual normalizada $P_s'(s)$ solamente se necesita transmitir al lado del decodificador para la banda no significativa, como tal información perceptual para la banda significativa se puede regenerar fácilmente por medio del decodificador.

ES 2 305 852 T3

El índice del primer (máximo) plano binario de la pluralidad de planos binarios para la codificación del plano binario de la señal de error $M(s)$, que es parte de la información perceptual de la señal digital, se puede determinar a partir del intervalo máximo de cuantificación usado para cuantificar la señal digital/transformada. Para la banda significativa, se determina el intervalo máximo de cuantificación (la diferencia entre el umbral superior y el umbral inferior de cuantificación correspondiente a cada uno de los valores cuantificados de la señal cuantificada), y el mencionado primer plano binario (especificado por $M(s)$) se determina de acuerdo con esto. Dicho intervalo máximo de cuantificación también se puede determinar en el lado del descodificador, y de esta forma, no se necesita transmitir el mencionado primer plano binario (especificado por $M(s)$) como parte del flujo binario escalable en este caso (para la banda significativa).

Aunque se describe la codificación de una señal digital en un flujo binario escalable, se debe entender que la invención incluye también la descodificación del flujo binario escalable en una señal digital descodificada por el inverso del procedimiento descrito con anterioridad.

En una realización de la invención, se proporciona un procedimiento para la descodificación del flujo binario escalable en la señal digital que incluye la demultiplexación del flujo binario escalable en un flujo binario de capa de núcleo y en un flujo binario de capa de mejora, la descodificación y la descuantificación del flujo binario de capa de núcleo para generar una señal de capa de núcleo, la descodificación del plano binario de la capa de mejora en base a la información perceptual de la señal digital, la realización de una correspondencia de errores en base a la señal de capa de mejora descodificada del plano binario y la señal de capa de núcleo descuantificada para generar una señal transformada reconstruida, en la que la señal transformada reconstruida es la señal digital. Se debería notar que el procedimiento para la descodificación del flujo binario escalable se puede usar en combinación con, pero también de manera independiente del procedimiento de codificación de una señal digital en el flujo binario escalable como se ha descrito con anterioridad.

La señal transformada reconstruida se puede transformar para generar la señal digital, si la señal digital está en un dominio diferente de la señal transformada reconstruida.

La implementación exacta de la descodificación del flujo binario escalable para generar la señal digital depende de cómo se codifique el flujo binario escalable por medio del codificador. En un ejemplo, la señal transformada reconstruida se puede transformar usando intMDCT para generar la señal digital. El flujo binario de capa de núcleo se puede descodificar y se puede descuantificar de acuerdo con la aplicación MPEG AAC. La correspondencia de errores se realiza mediante la suma del umbral de cuantificación inferior usado para la descuantificación de la señal transformada y el flujo binario de capa de mejora descodificado de plano binario para generar la señal transformada reconstruida. Las ventajas y otras implementaciones del descodificador son similares al codificador, que ya se ha descrito con anterioridad.

La información perceptual de la señal digital se puede obtener por medio de la demultiplexación del flujo binario escalable, si la información perceptual se ha multiplexado en el flujo binario escalable como información lateral. De manera alternativa, si el flujo binario de capa de núcleo se codifica de manera perceptual, la información perceptual obtenida por medio de la descodificación y de la descuantificación del flujo binario de capa de núcleo se puede usar para la descodificación de plano binario del flujo binario de capa de mejora.

En una realización de la invención, el flujo binario de capa de mejora es descodificado en plano binario en una secuencia consecutiva para generar una pluralidad de planos binarios que comprenden una pluralidad de símbolos de plano binario, y los planos binarios son desplazados en base a la información perceptual de la señal digital para generar el flujo binario de capa de mejora descodificado de plano binario.

En otra realización de la invención, el flujo binario de capa de mejora está descodificado en plano binario en una secuencia en base a la información perceptual de la señal digital para generar una pluralidad de planos binarios que comprenden una pluralidad de símbolos de plano binario, generando por lo tanto el flujo binario de capa de mejora descodificado de plano binario.

La información perceptual de la señal digital puede ser al menos una de las siguientes:

- el plano binario que corresponde al flujo binario de capa de mejora cuando comienza la descodificación de plano binario del flujo binario de capa de mejora $M(s)$; y
- el nivel de la Distorsión Apenas Perceptible (JND) de la señal digital, en donde s corresponde a una banda de frecuencias de la señal digital.

El plano binario corresponde al flujo binario de capa de mejora cuando se inicia la descodificación de plano binario del flujo binario de capa de mejora $M(s)$ se determina a partir del intervalo de cuantificación máximo usado para la descuantificación del flujo binario de capa de núcleo.

El segundo aspecto de la invención so se refiere solamente a un procedimiento para la descodificación de un flujo binario escalable en una señal digital, sino que también incluye un programa de ordenador, un medio legible por un ordenador y un dispositivo para implementar el mencionado procedimiento.

Descripción detallada de la invención

Ahora se describirán varias realizaciones e implementaciones de la invención con detalle con referencia a las figuras, en las que:

5

La figura 1 muestra un codificador de acuerdo con una realización de la invención.

La figura 2 muestra un decodificador de acuerdo con una realización de la invención.

10

La figura 3 ilustra una estructura de un proceso de codificación de plano binario.

La figura 4 muestra un codificador de acuerdo con una realización de la invención.

15

La figura 5 muestra un decodificador de acuerdo con una realización de la invención.

La figura 6 muestra un codificador de acuerdo con una realización de la invención.

La figura 7 muestra un decodificador de acuerdo con una realización de la invención.

20

La figura 1 muestra un codificador 100 de acuerdo con una realización de la invención.

El codificador 100 sirve para generar un flujo binario escalable, y comprende dos capas distintas, a saber, una capa de núcleo que genera el flujo binario de capa de núcleo, y una capa de Mejora sin Pérdidas (LLE) (del inglés, *Lossless Enhancement*) que genera el flujo binario de capa de mejora.

25

El codificador comprende un transformador de dominio 101, un cuantificador 102, una unidad de correspondencia de errores 103, un codificador de plano binario perceptual 104 y un multiplexor 105.

30

En el codificador 100, la señal digital se transforma primero por medio del transformador de dominio 101 a un dominio adecuado, tal como el dominio de la frecuencia, dando como resultado una señal transformada. Los coeficientes de la señal transformada son cuantificados por medio del cuantificador 102 y codificados para generar el flujo binario de capa de núcleo. La correspondencia de errores se realiza por medio de la unidad de correspondencia de errores 103, que corresponde a la capa LLE, para eliminar la información de los coeficientes de la señal transformada que haya sido usada o haya sido codificada en la capa de núcleo para formar el flujo binario de capa de núcleo. El residuo resultante o la señal de error resultante, de manera específica los coeficientes de error, con codificados en plano binario por medio del codificador de plano binario 104 para generar el flujo binario LLE incorporado. Este flujo binario incorporado se puede trunca de manera adicional a velocidades binarias más bajas en el codificador 100 o en el correspondiente decodificador (como el decodificador 200 que se muestra en la figura 2 y que se describe a continuación), o en el canal de comunicación para cumplir con los requisitos de velocidad/fidelidad. Se usa un modelo perceptual 106 para controlar la codificación de plano binario de los coeficientes de error, de forma que los bits de los coeficientes de error que sean más significativos desde el punto de vista perceptual son los que se codifican primero.

35

40

Finalmente, el flujo binario de capa LLE resultante se multiplexa con el flujo binario de capa de núcleo por medio del multiplexor 105 para generar el flujo binario escalable. Además, la información perceptual para controlar la codificación de plano binario de los coeficientes de error también se puede transmitir como una información lateral de forma que un correspondiente decodificador de plano binario sea capaz de reconstruir los coeficientes de error en un orden correcto.

45

50

Cuando el flujo binario LLE se trunque a velocidades más bajas, la señal decodificada sería una versión con pérdidas de la señal de entrada original.

La figura 2 muestra un decodificador 200 de acuerdo con una realización de la invención.

55

El decodificador 200 decodifica un flujo binario escalable generado por el codificador 100 para reconstruir la señal digital que se codificó por medio del codificador 100.

El decodificador 200 comprende un transformador de dominio 201, un descuantificador 202, una unidad de correspondencia de errores 203, un decodificador de plano binario perceptual 204 y un demultiplexor 205.

60

65

El demultiplexor 205 recibe el flujo binario escalable como entrada y divide el flujo binario escalable en el flujo binario de capa de núcleo y en el flujo binario de capa de mejora como se generó en el codificador 100. El flujo binario de capa de núcleo se decodifica y se descuantifica por medio del descuantificador 202 para formar la señal de capa de núcleo. El flujo binario de capa de mejora se decodifica en plano binario de manera perceptual en base a la información perceptual dada por un modelo perceptual 206 por el decodificador de plano binario perceptual 204, y es posteriormente sometida a correspondencia de errores por parte de la unidad de correspondencia de errores 203 con la señal de capa de núcleo para generar una señal de capa de mejora. La señal de capa de mejora se transforma finalmente de nuevo al dominio de la señal digital por medio del transformador de dominio 201, dando como resultado una señal transformada de capa de mejora que es la señal digital reconstruida.

ES 2 305 852 T3

El procesamiento llevado a cabo por el codificador 100 y por el decodificador 200 se explica con detalle a continuación.

La señal de entrada se transforma normalmente al dominio de la frecuencia por el transformador de dominio 101 antes de que sea cuantificada por medio del cuantificador 102 (que es parte del codificador de capa de núcleo) para generar el flujo binario de capa de núcleo. Se pueden usar varias funciones de transformadas para transformar la señal de entrada al dominio de la frecuencia, tales como la Transformada de Coseno Discreta (DCT), la Transformada de Coseno Discreta Modificada (MDCT), MDCT entera (intMDCT) o la Transformada Rápida de Fourier (FFT).

Cuando se usa un codificador MPEG-4 AAC como el codificador de capa de núcleo (para la señal de audio), se usa de manera común MDCT para transformar la señal de audio de entrada al dominio de la frecuencia, como se describe en [1]. En [13] la MDCT entera (intMDCT) se propone como una aproximación reversible al banco de filtros de la Transformada de Coseno Discreta Modificada (MDCT) usado con el codificador MPEG-4 AAC. Una manera generalmente usada para implementar la intMDCT es factorizar el banco de filtros MDCT en una cascada de rotaciones de Givens en la forma de:

$$\begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix},$$

que es factorizada de manera adicional en tres pasos de elevación

$$\begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} = \begin{pmatrix} 1 & \frac{\cos \alpha - 1}{\sin \alpha} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \sin \alpha & 1 \end{pmatrix} \begin{pmatrix} 1 & \frac{\cos \alpha - 1}{\sin \alpha} \\ 0 & 1 \end{pmatrix}.$$

Cada paso de elevación se puede aproximar por medio de un entero reversible a un entero con correspondencia con el redondeo a la operación de entero más cercana $r: R \rightarrow Z$. Por ejemplo, el último paso de elevación se aproxima por medio de:

$$(x_1, x_2) \rightarrow \left(x_1 + r \left(\frac{\cos \alpha - 1}{\sin \alpha} x_2 \right), x_2 \right),$$

que se puede invertir de manera sin pérdidas por medio de:

$$(x'_1, x'_2) \rightarrow \left(x'_1 - r \left(\frac{\cos \alpha - 1}{\sin \alpha} x'_2 \right), x'_2 \right).$$

Se obtiene de esta manera intMDCT por medio de la implementación de todas las rotaciones de Givens con los enteros reversibles a los que se ha sometido a la correspondencia como se ha descrito anteriormente.

En el decodificador, intMDCT se puede usar de nuevo por el transformador de dominio 201 para transformar la señal de capa de mejora en la señal digital (reconstruida).

En la capa de núcleo, los coeficientes $c(k)$ de la señal transformada, donde $k = 1, \dots, 1024$, es la longitud de una trama del flujo binario de capa de núcleo, se cuantifican por medio del cuantificador 102 y se codifican en el flujo binario de capa de núcleo. En el contexto de una señal de audio de entrada, los coeficientes de señal transformada se pueden cuantificar de acuerdo con los valores de cuantificación de un codificador MPEG-4 AAC, un codificador MPEG-1 de Capa 3 (MP3) o cualquier codificador de audio propietario.

Cuando se use el codificador MPEG-4 AAC junto con la intMDCT, los coeficientes de la señal transformada (también conocidos como los coeficientes intMDCT), $c(k)$, se normalizan primero como:

$$c'(k) = \alpha \cdot c(k)$$

ES 2 305 852 T3

para aproximar las salidas normalizadas a las salidas del filtro de bancos MDCT. Los coeficientes normalizados intMDCT, $c'(k)$ se cuantifican y se codifican después, por ejemplo, de acuerdo con un cuantificador AAC (véase [19]) que se da de la siguiente manera:

$$i(k) = \text{sgn}[c'(k)] \left[\left(\frac{|c'(k)|}{\sqrt[4]{2} \text{factor_escala}(s)} \right)^{3/4} + 0,4054 \right].$$

Aquí, $[\cdot]$ denota la operación de fondo que trunca un operando flotante a un entero, $i(k)$ es los coeficientes cuantificados AAC y $\text{factor_escala}(s)$ es un factor de escala de una banda s de factor de escala a la que pertenece el coeficiente $c(k)$. Los factores de escala se pueden ajustar de manera adaptable por medio de un procedimiento de conformación de ruido de forma que el ruido de cuantificación se enmascare mejor por medio del umbral de enmascaramiento del sistema auditivo humano. Una aproximación ampliamente adoptada para este procedimiento de conformación del ruido es el bucle anidado de cuantificación y de codificación como se describe con detalle en [1].

Los coeficientes cuantificados $i(k)$ son codificados sin ruido (en este ejemplo por medio del cuantificador 102), por ejemplo, usando el código de Huffman o el Código Aritmético de bit partido (BSAC) (del inglés, *Bit Sliced Arithmetic Code*) como se describe en [17]. El BSAC es el preferido si se requiere de manera adicional un escalado de la velocidad binaria en el flujo binario de capa de núcleo. Los factores de escala se codifican de manera diferencial, por ejemplo, por medio del proceso de codificación DPCM descrito en [1], o usando el código de Huffman. El flujo binario de capa de núcleo se puede generar entonces por medio de la multiplexación de toda la información codificada de acuerdo con la sintaxis de flujo binario AAC.

Se puede encontrar una descripción más completa acerca de MPEG AAC en [1] o en el documento de normativa internacional sobre MPEG AAC [19].

Se debería notar que aunque se describe el mecanismo de incorporación del flujo binario que cumple con el estándar MPEG-4 AAC, también es posible usar flujos binarios que cumplan con otros codificadores tales como MPEG 1/2 de Capa I, II, III (MP3), Dolby AC3 o los codificadores propietarios ATRAC de SONY como se describe en [20].

Cuando el cuantificador 102 funciona de acuerdo con el codificador MPEG AAC, el descuantificador 202 de manera preferible funciona de acuerdo con un descodificador MPEG AAC para descodificar y descuantificar el flujo binario de capa de núcleo en el descodificador 200. De manera específica, del descuantificador 202 se usa para generar la señal de capa de núcleo que es usada con posterioridad para la correspondencia de errores por parte de la unidad de correspondencia de errores 203 en el descodificador 200 para generar la señal de capa de mejora como se describirá con posterioridad.

Sin embargo, se debería notar que se pueden usar en el descodificador 200 los descuantificadores de acuerdo con otras especificaciones tales como MP3 u otros descodificadores propietarios.

En la capa LLE, se emplea un procedimiento de correspondencia de errores para eliminar la información que ya ha sido codificada en el flujo binario de capa de núcleo. Una posible aproximación para construir dicho procedimiento de correspondencia de errores es por medio de la resta del umbral de cuantificación inferior (más cercano a cero) de cada coeficiente cuantificado del coeficiente de señal de entrada transformada correspondiente.

Esto se puede ilustrar como:

$$e(k) = c(k) - \text{thr}(k),$$

donde $\text{thr}(k)$ es el umbral de cuantificación inferior (más cercano a cero) para $c(k)$, y $e(k)$ es el coeficiente de error que representa la señal de error.

Cuando se usa el codificador MPEG-4 AAC como el cuantificador:

$$\text{thr}(k) = \text{sgn}[c(k)] \left[\frac{\sqrt[4]{2} \text{factor_escala}(s) [i(k) - 0,4054]^{4/3}}{\alpha} \right].$$

ES 2 305 852 T3

En aplicaciones prácticas, para asegurar la reconstrucción robusta, se puede realizar una correspondencia desde el entero $i(k)$ al entero $thr(k)$ usando una tabla de consulta. Como se puede ver claramente de la fórmula anterior, se requieren un total de 4 tablas para diferentes valores de factores_escalas (mientras que la misma tabla se puede compartir entre diferentes valores de factores_escalas si tienen un desplazamiento módulo 4 por bit), en la que cada una de las tablas contiene la correspondencia entre todos los posibles valores de $i(k)$ y el $thr(k)$ correspondiente para cualquier factor_escalas del conjunto de aquéllos con módulo 4.

También es posible realizar un procedimiento de correspondencia de errores mediante la resta de un coeficiente reconstruido del coeficiente de la señal de entrada transformada del coeficiente de la señal transformada como se describe en [3], que se puede ilustrar como:

$$e(k) = c(k) - \hat{c}(k)$$

15

donde $\hat{c}(k)$ es el coeficiente de señal transformada reconstruido.

En general, también es posible realizar el procedimiento de correspondencia de errores en base a usar:

20

$$e(k) = c(k) - f(k)$$

25

en donde $f(k)$ es cualquier función que corresponde a $c(k)$, tal como

30

$$f(k) = \frac{1}{2} (\text{thr}(k+1) - \text{thr}(k)).$$

Claramente, para $c(k)$ que ya ha sido significativo en la capa de núcleo ($thr(k) \neq 0$), el signo de $e(k)$ residual intMDCT se puede determinar a partir de la reconstrucción de la capa de núcleo y así solamente se necesita codificar su amplitud en la capa LLE. Además, se sabe bien que para la mayoría de las señales de audio, $c(k)$ puede aproximarse por variables aleatorias laplacianas con la función de densidad de probabilidad (pdf):

40

$$f(c(k)) = e^{-|c(k)|\sqrt{2/\sigma^2}} / 2\sigma^2,$$

45

donde σ es la varianza de $c(k)$. A partir de la propiedad de "sin memoria" de una pdf laplaciana, es fácil verificar que la amplitud de $e(k)$ está distribuida de manera geométrica como,

50

$$f(|e(k)|) = \beta \cdot \theta(k)^{|e(k)|}, \quad (1)$$

donde el parámetro de distribución $\theta(k)$ está determinado por la varianza $c(k)$ y el tamaño del paso del cuantificador de capa de núcleo. Esta propiedad hace posible un esquema de codificación de plano binario muy eficiente, tal como el código Golomb de plano binario (BPGC) 0 para la codificación de la señal de error que se vaya a aplicar.

En el decodificador 200, los coeficientes de la señal transformada se pueden reconstruir por medio del procedimiento de correspondencia de errores realizado por la unidad de correspondencia de errores 203 de acuerdo con la siguiente ecuación:

65

$$c(k) = e'(k) + \text{thr}(k)$$

ES 2 305 852 T3

donde $e'(k)$ son los coeficientes de error descodificados que describen el flujo binario de capa de mejora descodificado de plano binario, que corresponde con los coeficientes de error $e(k)$ en el codificador 100. Así, se puede ver que con los coeficientes de señal transformada, $c(k)$ se puede regenerar a partir de los coeficientes de error descodificados $e'(k)$ (es posible una versión con pérdidas si el flujo binario LLE se trunca a velocidades inferiores) y el umbral de cuantificación $thr(k)$ generado de la misma manera en el codificador con el índice de cuantificación $i(k)$ contenido en el flujo binario de capa de núcleo incorporado (AAC).

Similar al codificador 100, los coeficientes de señal transformados $c(k)$ en el descodificador 200 también se pueden generar usando (sumando) los coeficientes de error descodificados $e'(k)$ y los coeficientes reconstruidos del flujo binario de capa de núcleo. También, los coeficientes de la señal transformados $c(k)$ se pueden generar usando (sumando) los coeficiente de error descodificados $e'(k)$ y una función de $c(k)$.

Para producir el escalado a la parte sin pérdidas del flujo binario sin pérdidas incorporado final, la señal de residuo o de error se codifica de manera adicional por medio del codificador de plano binario perceptual 104 usando codificación de plano binario, una tecnología de codificación incorporada que ha sido ampliamente adoptada en la codificación de audio [3] o en la codificación de imagen [5], en la capa LLE.

En [4] y en [15] se puede encontrar una descripción de un procedimiento de codificación de plano binario general. Considérese un vector de datos de dimensión n , $\underline{x}_n = \{x_1, \dots, x_n\}$ donde x_i se extrae de algunas fuentes aleatorias de algún alfabeto $A \subset \mathfrak{R}$. Claramente, x_i se puede representar en formato binario

$$x_i = (2s_i - 1) \cdot \sum_{j=-\infty}^{\infty} b_{i,j} \cdot 2^j, \quad i = 1, \dots, k$$

por medio de la configuración en cascada de símbolos binarios de plano binario que comprenden un símbolo de signo

$$s_i = \begin{cases} 1 & x_i \geq 0 \\ 0 & x_i < 0 \end{cases}$$

y símbolos de amplitud $b_{i,j} \in \{0, 1\}$. En la práctica, la codificación de plano binario se podría iniciar desde e máximo plano binario M del vector \underline{x}_n donde M es un entero que satisface

$$2^{M-1} \leq \max \{|x_i|\} < 2^M, \quad i = 1, \dots, k,$$

y se detuvo en el 0 de plano binario si \underline{x}_n es un vector entero.

El proceso de codificación y de descodificación de plano binario de acuerdo con una realización de la invención y como por ejemplo realizados por el codificador de plano binario perceptual 104 y el descodificador de plano binario perceptual 204 se explica en lo que sigue con referencia a la figura 3.

La figura 3 ilustra una estructura de un proceso de codificación de plano binario (BPC) anterior en el que cada vector de entrada se descompone primero en el signo binario y en los símbolos de amplitud, que son después explorados en un orden deseado por medio de una unidad de exploración de plano binario 301 y son codificados por medio de un codificador de entropía 302 (por ejemplo, como código aritmético, como código de Huffman o como código de longitud de fiabilidad no acreditada). Además, un modelo estadístico 303 por ejemplo basado en una distribución laplaciana de la señal de entrada, se usa por lo general para determinar la asignación de probabilidad para cada símbolo binario que se vaya a codificar. En el descodificador correspondiente, el flujo de datos se invierte, es decir, la salida del codificador de entropía 302 se descodifica por medio de un descodificador de entropía 303 usando un modelo estadístico correspondiente 304 y el resultado se usa por una unidad de reconstrucción de plano binario 304 para reconstruir el plano binario, donde el signo y los símbolos de amplitud que están descodificados para reconstruir el plano binario del vector de datos sigue el mismo orden de exploración en el codificador.

ES 2 305 852 T3

La ventaja más significativa de tener un sistema de codificación de plano binario como el anterior es que la compresión resultado del flujo binario se puede truncar fácilmente a cualquier velocidad deseada, en donde se puede obtener un vector de datos de reproducción \hat{x} por medio de planos binarios parciales reconstruidos descodificados a partir de este flujo binario truncado. Para un mejor funcionamiento de la codificación, se adopta por lo general un principio incorporado (véase [24]) en BPC, de acuerdo con el que se codifican los símbolos de plano binario en el orden de pendiente decreciente de velocidad - distorsión de manera que los símbolos que la contribución de significación mayor para la distorsión final por velocidad unidad siempre se codifiquen los primeros.

La selección del orden de la exploración del plano binario depende de la medida de la distorsión deseada. Cuando el error cuadrático medio (MSE) o la expectativa sobre la función de error cuadrático se usa como la medida de la distorsión como se muestra:

$$d(\underline{x}_n, \underline{\hat{x}}_n) = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2$$

en donde $d(\underline{x}_n, \underline{\hat{x}}_n)$ es el valor de distorsión, x_n es el vector de datos original, y \hat{x}_n es el vector reconstruido de x_n en el descodificador. Los resultados de [24] muestran que el principio incorporado se satisface bien por medio de una exploración secuencial de plano binario y un procedimiento de codificación para la mayoría de las fuentes, excepto aquéllas con muy poca desviación de la distribución de símbolos de plano binario.

Un ejemplo de un procedimiento de exploración y de codificación secuencial de plano binario comprende los siguientes pasos:

1. Iniciar desde el plano binario más significativo $j = M - 1$;
2. Codificar solamente $b_{i,j}$ con $b_{i,M-1} = b_{i,M-2} = \dots = b_{i,j+1} = 0$. Si $b_{i,j} = 1$ en la exploración significativa, codificar si (paso de significancia);
3. Codificar $b_{i,j}$ que no estén codificados en el paso de significancia (Paso de refinamiento);
4. Progresar al plano binario $j - 1$.

Lista 1

40 *Procedimiento de exploración y de codificación de plano binario*

El anterior procedimiento se itera hasta que se alcanza cierto criterio de terminación, que generalmente es una restricción de la velocidad/distorsión predefinida. Además, puede que se requiera un ajuste adicional de la secuencia de codificación en un paso de significancia si se encuentra que los símbolos de plano binario tienen distribuciones desiguales.

Se ilustra un ejemplo del procedimiento de codificación secuencial anterior considerando un vector de datos x con dimensión 4, digamos $\{9, -7, 14, 2\}$. Así, se codifica en plano binario desde su plano binario más significativo 4. El paso de significancia se comienza desde que todos los elementos son aún insignificantes. (X denota los símbolos de punteo). El signo se codifica de la siguiente manera: el positivo se codifica como 1, y el negativo se codifica como 0.

(Tabla pasa a página siguiente)

ES 2 305 852 T3

5	Vector de datos	9	-7	14	1
10	1 ^{er} paso significativo (signo)	1 (signo: 1)	0	1 (signo: 1)	0
15	1 ^{er} paso de refinamiento	X	X	X	X
20	2 ^o paso significativo (signo)	X	1 (signo: 0)	X	0
25	2 ^o paso de refinamiento	0	X	1	X
30	3 ^{er} paso significativo (signo)	X	X	X	0
35	3 ^{er} paso de refinamiento	0	1	1	X
40	4 ^o paso significativo (signo)	X	X	X	1 (signo: 1)
45	4 ^o paso de refinamiento	1	1	0	X

De esta manera, el flujo binario de salida es 11011010001001111110, que es la entropía codificada y enviada al codificador. En el codificador, se reconstruye la estructura de plano binario del vector de datos original. Si el decodificador recibe el flujo binario en su integridad, el plano binario del vector de datos original se puede restaurar, y de esta manera se obtiene una reconstrucción sin pérdidas del vector de datos original. Si solamente se recibe un subconjunto (la parte más significativa) del flujo binario, el decodificador es aún capaz de restaurar un plano binario parcial del vector de datos original, de forma que es una versión de reconstrucción basta (cuantificada) del vector de datos original.

Lo anterior solamente es un ejemplo sencillo de un procedimiento de exploración y de codificación de plano binario. En la práctica, el paso significativo se puede fraccionar de manera adicional para explorar la correlación estadística de elementos en el vector de datos, tal como el proceso de codificación del plano binario en JPEG2000, o el del codificador de audio incorporado (EAC) descrito en [4].

El anterior procedimiento de exploración y de codificación de plano binario secuencial solamente proporciona un esfuerzo para optimizar el funcionamiento MSE. En el área de audio, imagen o codificación de vídeo, la minimización de la distorsión perceptual en lugar de MSE es normalmente un procedimiento de codificación más eficiente para obtener una calidad perceptual óptima en la señal de audio, imagen o de vídeo reconstruida. Por lo tanto, la codificación secuencial de plano binario de la señal de error es definitivamente una opción subóptima.

En el codificador 100, los coeficientes de error están agrupados de manera preferible en bandas de frecuencia de manera que cada banda de frecuencia s contiene un número de coeficientes de error en orden consecutivo. (La agrupación de banda de factor de escala se puede basar en la agrupación de banda adoptada en el cuantificador 102 si se usa un codificador perceptual como el cuantificador 102. Sin embargo, también es posible otra agrupación de banda).

ES 2 305 852 T3

Se dice que una banda de frecuencia s es significativa si existe un coeficiente de error en la banda de frecuencia s de forma que el coeficiente cuantificado $\text{thr}(k)$ del cuantificador no sea cero. En otras palabras, si $e(k)$ es un coeficiente de error en la banda de frecuencia s :

5

$$e(k) = c(k) - \text{thr}(k),$$

10 la banda de frecuencia s es significativa si $\text{thr}(k) \neq 0$ ($\text{thr}(k) = 0$ cuando $i(k) = 0$), y así $e(k) = c(k)$, luego la banda de frecuencia s se considera insignificante.

15 La significancia perceptual de los bits de los coeficientes de error se puede determinar por el nivel de la Distorsión Apenas Perceptible (JNP) en una localización de frecuencia i . Este nivel de JND, T_i se puede determinar a partir de un modelo perceptual tal como el modelo psicoacústico (I o II) o cualquiera de los modelos perceptuales propietarios. Cuando se usa un cuantificador perceptual para formar el flujo binario de capa de núcleo, el modelo perceptual usado en el cuantificador también se puede usar para generar la JND para la codificación de plano binario perceptual de los coeficientes de error.

20 Por razones de simplicidad, la significancia perceptual de los bits de los coeficientes de error en una misma banda de frecuencia s se puede fijar al mismo valor.

En lo que sigue, se explica con referencia a la figura 4 una posible implementación de codificación de plano binario perceptual.

25

La figura 4 muestra un codificador 400 de acuerdo con una realización de la invención.

30 De manera análoga al codificador 100, el codificador 400 comprende un transformador de dominio 401, un cuantificador 402, una unidad de correspondencia de errores 403, un codificador de plano binario perceptual 404 (que usa un modelo perceptual 406) y un multiplexor 405.

El bloque BPC perceptual, es decir, el codificador de plano binario perceptual 404 comprende un bloque de desplazamiento de plano binario 407 y un bloque BPC convencional 408.

35 En el bloque de desplazamiento de plano binario 407, los planos binarios son desplazados de manera perceptual y los planos binarios desplazados de manera perceptual son codificados en el bloque BPC 408 de una manera de exploración y de codificación secuencial convencional.

Considérese la siguiente medida de la distorsión ponderada de manera perceptual (modificada)

40

$$d(\underline{x}_n, \hat{\underline{x}}_n) = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2 w_i(x_i).$$

45

En el contexto de la codificación de audio perceptual, la señal de audio generalmente se cuantifica y se codifica en el dominio de la frecuencia de forma que el vector de datos \underline{x}_n es la señal de audio transformada y la función de ponderación $w_i(x_i)$ es la importancia de x_i en diferentes localizaciones de frecuencia i , es decir,

50

$$w_i(x_i) = \frac{1}{T_i}.$$

55

La anterior función de distorsión ponderada de manera perceptual se puede reescribir de la siguiente manera:

60

$$\begin{aligned} d(\underline{x}_n, \hat{\underline{x}}_n) &= \frac{1}{n} \sum_{i=1}^n \frac{1}{T_i} (x_i - \hat{x}_i)^2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{\sqrt{T_i}} x_i - \frac{1}{\sqrt{T_i}} \hat{x}_i \right)^2 \\ &= \frac{1}{n} \sum_{i=1}^n (x'_i - \hat{x}'_i)^2, \end{aligned}$$

65

ES 2 305 852 T3

donde

$$x'_i = \frac{\Delta}{\sqrt{T_i}} x_i, i = 1, \dots, n.$$

Así, la función de error cuadrático de ponderación pasa a ser ahora una función de error cuadrático sobre el vector escalado $\underline{x}'_n = \{x'_1, \dots, x'_n\}$. Por lo tanto, la codificación optimizada de manera perceptual de \underline{x}_n se puede conseguir simplemente por medio de la realización de la codificación de plano binario secuencial sobre \underline{x}'_n . En el descodificador correspondiente, cada uno de los elementos del vector de datos descodificado de plano binario $\underline{\hat{x}}'_n$ se puede escalar hacia atrás para obtener un vector de datos reconstruido $\underline{\hat{x}}$ de la siguiente manera.

$$\hat{x}_i = \sqrt{T_i} \cdot \hat{x}'_i, i = 1, \dots, n.$$

Claramente, los pesos T_i son de manera preferible transmitidos al descodificador como información lateral si son desconocidos en el descodificador.

w_i es cuantificado de manera adicional a un entero par potencia de dos de forma que pasa a ser:

$$\hat{T}_i = 2^{2\tau_i},$$

donde

$$\tau_i = \frac{1}{2} \log_2 T_i$$

y el vector de datos escalado se puede obtener de esta forma por medio del desplazamiento binario de cada elemento en el vector de datos original de la siguiente manera

$$x'_i = 2^{-\tau_i} x_i.$$

que es fácilmente obtenible por medio de la realización de la operación de desplazamiento a la derecha sobre x_i por τ_i . Por ejemplo, si $x_i = 00010011$ y $\tau_i = -2$, el elemento de vector de datos escalado x'_i es entonces 01001100; si $\tau_i = 2$, pasará a ser 0000010011.

De esta manera, los planos binarios de los coeficientes de error son desplazados de manera perceptual de una manera tal que cuando se realiza una codificación de plano binario secuencial sobre el plano binario desplazado, los bits que sean más significativos desde el punto de vista perceptual (en lugar de tener el MSE más alto) se puede codificar los primeros.

Claramente, si cada uno de los elementos en el vector de datos original es entero con longitud de palabra limitada, por ejemplo, si cada uno de los elementos en \underline{x} tiene un plano binario máximo de L , la codificación sin pérdidas de \underline{x} se puede conseguir si cada x'_i en el vector escalado está codificado en plano binario a partir del plano binario $-\tau_i \sim L - \tau_i$.

Como se ha mencionado anteriormente, la información acerca de la significancia perceptual tal como el nivel de JND se puede proporcionar al bloque de desplazamiento de plano binario desde un modelo perceptual.

En el proceso de codificación de plano binario, se puede usar un plano binario máximo $M(s)$ para especificar el plano binario de inicio en el que deberían comenzar la exploración y la codificación del plano binario. El plano binario máximo $M(s)$ y τ_i se deberían transmitir de manera preferible como información lateral en el flujo binario escalable al descodificador correspondiente con el fin de que el descodificador sea capaz de descodificar el flujo binario de manera correcta. Para reducir la cantidad de información lateral, $M(s)$ y τ_i pueden estar restringidos al mismo valor para la misma banda s de factor de escala en el codificador.

ES 2 305 852 T3

El valor del plano binario máximo $M(s)$ en cada una de las bandas de frecuencia s se puede determinar a partir de los coeficientes de error $e(k)$ usando la siguiente expresión:

$$2^{M(s)-1} \leq \max (|e(k)|) < 2^{M(s)}, k \in s.$$

Además, el valor absoluto máximo de los coeficientes de error $\max(|e(k)|)$ en cada una de las bandas de frecuencia s significativas está limitado por el intervalo de cuantificador del cuantificador perceptual:

$$\max (|e(k)|) \leq \text{thr}(i(k) + 1) - \text{thr}(i(k)).$$

Por lo tanto, esto da como resultado el plano binario máximo $M(s)$ para cada banda de frecuencia s significativa que se ha de determinar a partir de la siguiente expresión:

$$2^{M(s)-1} \leq \max (|\text{thr}(i(k) + 1) - \text{thr}(i(k))|) < 2^{M(s)}, k \in s.$$

Como los coeficientes cuantificados del cuantificador perceptual $i(k)$ son conocidos por el descodificador, de esta manera no es necesario transmitir el máximo plano binario $M(s)$ como información lateral al descodificador para la banda de frecuencia s significativa.

El valor del plano binario máximo $M(s)$ se puede predefinir también en el codificador y en el descodificador, y de esta manera, no se necesitan transmitir como información lateral.

La figura 5 muestra un decodificador 500 de acuerdo con una realización de la invención.

El descodificador 500 implementa un descodificador de plano binario perceptual que comprende el desplazamiento de plano binario y la codificación de plano binario convencional (secuencial).

De manera análoga al descodificador 200, el descodificador 500 comprende un transformador de dominio 501, un descuantificador 502, una unidad de correspondencia de errores 503, un descodificador de plano binario perceptual 504 (usando un modelo perceptual 506) y un demultiplexor 505.

De manera similar al codificador de plano binario perceptual 404, el descodificador de plano binario perceptual 504 comprende un bloque de desplazamiento de plano binario 507 y un bloque BPC convencional 508.

El flujo binario de capa de mejora generado por el codificador 400 es descodificado en plano binario por el descodificador 500 de la manera secuencia consecutiva (el mismo procedimiento de exploración de plano binario secuencial que en el codificador 400) para reconstruir los planos binarios. Los planos binarios reconstruidos se desplazan de la manera inversa en el codificador 400, en base al valor recibido o regenerado de τ_i para generar los coeficientes de error descodificados $e'(k)$ que describe el flujo binario de capa de mejora descodificado de plano binario.

La figura 6 muestra un codificador 600 de acuerdo con una realización de la invención.

El descodificador 600 usa codificación de plano binario perceptual.

El codificador 600 comprende un transformador de dominio (intMDCT) 601, un cuantificador (cuantificador y codificador AAC) 602, una unidad de correspondencia de errores 603, una unidad de cálculo de la significancia perceptual 604 (que usa un modelo psicoacústico 605), una unidad de codificación de plano binario perceptual 606 y un multiplexor 607.

En esta implementación, el orden de exploración de los planos binarios y de los símbolos de plano binario no necesita ser secuencial, sino basado en la importancia perceptual de los símbolos de plano binario correspondientes a las diferentes bandas de frecuencia. La importancia perceptual de los símbolos de plano binario se determina por medio del cálculo de los parámetros relacionados con la información perceptual, tal como la significancia perceptual y el primer (máximo) plano binario para la descodificación de plano binario. El cálculo de los parámetros de la información perceptual está representado como el bloque de cálculo de la significancia perceptual 604 conectado a la unidad de codificación perceptual de plano binario 606.

Existen numerosas maneras para determinar la importancia perceptual, o de manera específica la significancia perceptual, de los símbolos de plano binario correspondientes a diferentes bandas de frecuencia. Una ampliamente

ES 2 305 852 T3

adoptada es por medio del uso del modelo psicoacústico, tal como el Modelo Psicoacústico 2 descrito en [19], de la señal digital de entrada. EL nivel de la distorsión apenas perceptible (JND) $T(s)$ para cada banda de frecuencias determinado usando el modelo psicoacústico se puede convertir a la unidad de nivel de plano binario $\tau(s)$ de la siguiente manera:

$$\tau(s) = \frac{1}{2} \log_2 (T(s))$$

Sin embargo, la invención no restringe el procedimiento sobre cómo se pueden obtener $T(s)$ o $\tau(s)$.

Ahora, sea $Ps(s)$ que represente la significancia perceptual de la banda de frecuencia s , que se puede determinar por medio de la distancia desde $M(s)$ a $\tau(s)$ como,

$$Ps(s) = M(s) - \tau(s).$$

Se puede notar de manera adicional que el nivel de ruido, o el nivel de los coeficientes de error $e(k)$ intMDCT tendería a ser plano con respecto al nivel JND para las bandas significativas (como resultado del mecanismo de conformación del ruido en el codificador de núcleo). En otras palabras, el valor de $Ps(s)$ debería estar muy cercano, si no idéntico, para las bandas de frecuencia significativas. Este hecho se puede explorar en el procedimiento de acuerdo con la invención por medio de la compartición de un factor común $Ps_común$ para todas las bandas significativas. Las posibles selecciones de $Ps_común$ pueden ser el valor promedio, el valor máximo, el valor mínimo o cualquier otra función razonable de $Ps(s)$ para todas las s que sean significativas. La $Ps(s)$ se puede normalizar entonces de la siguiente manera:

$$Ps'(s) = Ps(s) - Ps_común,$$

Como se sabe que para la banda significativa s , $Ps(s)$ debería ser cero, y que por lo tanto, no se necesita transmitir al descodificador. En cualquier otro caso, la banda no significativa s , $Ps'(s)$ se debería transmitir de manera preferible al descodificador correspondiente como información lateral.

En algunos otros ejemplos cuando no existe banda significativa, $Ps_común$ se puede fijar a cero.

También es posible usar el procedimiento de conformación del ruido en el codificador de núcleo para atender a las necesidades de la codificación perceptual. Así, no existe la necesidad de implementar de manera adicional ninguna conformación del ruido, o identificación significativa perceptual en la capa de mejora. En dichos casos, se puede fijar $Ps'(s) = 0$ para todas las s . Generalmente, no se necesitan transmitir al descodificador si el descodificador sabe que todas ellas son cero.

Una posible implementación del mecanismo de codificación de plano binario perceptual se puede describir usando el siguiente pseudocódigo. Aquí, el número total de la banda de frecuencia se denota como s_total .

1. Encontrar la banda de frecuencia s con el mayor $Ps'(s)$
2. Codificar los símbolos de plano binario del plano binario $M(s)$ para cada $e(k)$ de la banda s
- 3.

$$M(s) = M(s) - 1; \quad Ps'(s) = Ps'(s) - 1$$

4. Si existe banda s para la que $M(s) \geq 0$, ir a 1.

Se describe aquí un procedimiento para obtener el máximo plano binario $M(s)$.

Para la banda de significancia, $M(s)$ se puede determinar a partir del intervalo de cuantificación máximo del cuantificador si se usa un cuantificador perceptual tal como un cuantificador AAC. De manera específica, $M(s)$ es un entero que satisface:

$$2^{M(s)-1} \leq \max \left(\left| \text{thr}(i(k) + 1) - \text{thr}(i(k)) \right| \right) < 2^{M(s)}, \quad k \in s.$$

En este caso, $M(s)$ no necesita transmitirse al descodificador mientras que $i(k)$ debería ser conocido para el descodificador.

ES 2 305 852 T3

Para las bandas no significativas $M(s)$ se puede calcular a partir de $e(k)$ de la siguiente manera:

$$2^{M(s)-1} \leq \max (|e(k)|) < 2^{M(s)}, k \in s,$$

y para esas bandas, $M(s)$ se debería enviar de manera preferible al descodificador como información lateral tal como una información que no esté contenida en el flujo binario de capa de núcleo.

El valor del plano binario máximo $M(s)$ se puede predefinir también en el codificador 600 y en el correspondiente descodificador, y así, no se necesitan transmitir como información lateral.

También son posibles otras aproximaciones alternativas para explorar el parámetro $Ps(s)$ en una aproximación de codificación de plano binario, hacia algunos objetivos de conformación de ruido deseados. En general, $Ps(s)$ también se puede obtener por medio de cualquier función de $M(s)$ y de $\tau(s)$, por ejemplo, la siguiente:

$$Ps(s) = M(s) - 2\tau(s),$$

o

$$Ps(s) = \frac{M(s) - \tau(s)}{2}.$$

La figura 7 muestra un codificador 700 de acuerdo con una realización de la invención.

El descodificador 700 es el descodificador correspondiente del codificador 600, en el que la descodificación de plano binario perceptual se implementa usando el procedimiento de exploración de plano binario perceptual como se ha descrito con anterioridad.

El descodificador 700 de acuerdo con esto, comprende un transformador de dominio (intMDCT inversa) 701, un descuantificador (descuantificador y descodificador AAC) 702, una unidad de correspondencia de errores 703, una unidad de cálculo de significancia perceptual 704, una unidad de descodificación de plano binario perceptual 706 y un demultiplexor 707.

En el descodificador 700, para la banda significativa, $Ps'(s)$ se fija a cero, y $M(s)$ se puede calcular a partir del índice de cuantificación AAC $i(k)$ de la misma manera que en el codificador, es decir:

$$2^{M(s)-1} \leq \max (|\text{thr}(i(k) + 1) - \text{thr}(i(k))|) < 2^{M(s)}, k \in s.$$

Para la banda no significativa, $Ps(s)$ y $M(s)$ se pueden recuperar simplemente a partir de la información lateral transmitida. Una vez que se han recuperado $Ps(s)$ y $M(s)$ para todas las bandas de frecuencia, los coeficientes de error intMDCT $\hat{e}(k)$ se pueden reconstruir fácilmente por medio de la descodificación del flujo binario recibido y de la reconstrucción de sus símbolos de plano binario en un orden que es exacto al mismo orden que en el codificador 700. Por ejemplo, el proceso de descodificación para el ejemplo de codificación dado con anterioridad sería:

1. Encontrar la banda de frecuencia s con el mayor $Ps'(s)$
2. Descodificar los símbolos de plano binario del plano binario $M(s)$ para $\hat{e}(k)$ en la banda
- 3.

$$M(s) = M(s) - 1; Ps'(s) = Ps'(s) - 1$$

4. Si existe la banda s para la que $M(s) \geq 0$, ir a 1.

Determinar el plano binario máximo para la codificación de plano binario de los coeficientes de error.

ES 2 305 852 T3

Para una banda significativa s (es decir, el coeficiente de error $e(k) \neq c(k)$ ó $\exists k \in s, i(k) \neq 0$), el máximo valor absoluto de $e(k)$ está limitado por el intervalo del cuantificador en el cuantificador AAC como:

$$\max (|e(k)|) \leq \text{thr} (i(k) + 1) - \text{thr} (i(k)).$$

Por lo tanto, el plano binario máximo $M(k)$ se puede determinar usando:

$$2^{M(k)-1} \leq \max (|\text{thr} (i(k) + 1) - \text{thr} (i(k))|) < 2^{M(k)}, k \in s.$$

Como $i(k)$ ya es conocido por el codificador, $M(k)$ no necesita ser transmitido al decodificador ya que el decodificador es capaz de regenerar $\text{thr}(k)$ y así, $M(k)$ a partir de $i(k)$ para la banda s significativa.

Para la banda no significativa, $M(k)$ se puede calcular a partir de $e(k)$ de la siguiente manera:

$$2^{M(s)-1} \leq \max (|e(k)|) < 2^{M(s)}, k \in s,$$

y el $M(s)$ calculado se transmite de manera preferible con el flujo binario de capa de mejora como información lateral para el flujo binario de capa de mejora para ser decodificado en plano binario de manera correcta.

Para reducir la cantidad de información lateral, $M(k)$ se puede restringir de manera adicional para tener el mismo valor para k para la misma banda s de factor de escala en el cuantificador de capa de núcleo. Por lo tanto, $M(k)$ también se puede denotar como $M(s)$.

En el decodificador 700, los coeficientes de error correspondientes a la señal de error se pueden reconstruir por medio de la descodificación de plano binario del flujo binario de capa de mejora usando el mismo procedimiento de exploración de plano binario que en el codificador basado en $M(s)$. Para la banda significativa, $M(s)$ se puede regenerar usando la siguiente expresión:

$$2^{M(k)-1} \leq \max (|\text{thr} (i(k) + 1) - \text{thr} (i(k))|) < 2^{M(k)}, k \in s.$$

Para la banda no significativa, el decodificador hace uso del $M(s)$ que es transmitido por el codificador como información lateral.

Referencias

[1] M. Bosi y colaboradores, "ISO/IEC Mpeg-2 Advanced Audio Coding", *J. Audio Eng. Soc.*, Vol. 45, Nº. 10, páginas 789 a 814, octubre de 1997.

[2] Jr. Stuart y colaboradores, "MLP lossless compression", AES 9ª Convención Regional, Tokyo.

[3] R. Geiger, J. Herre, J. Koller, y K. Brandenburg, "INTMDCT - A link between perceptual and lossless audio coding", *IEEE Proc. ICASSP 2002*.

[4] J. Li, "Embedded audio coding (EAC) with implicit auditory masking", *ACM Multimedia 2002*, Niza, Francia, diciembre de 2002.

[5] T. Moriya, N. Iwakami, T. Mori, y A. Jin, "A design of lossy and lossless scalable audio coding", *IEEE Proc. ICASSP 2000*.

[6] T. Moriya y colaboradores, "Lossless Scalable Audio Coder and Quality Enhancement", *Proceeding of ICASSP 2002*.

[7] M. Hans y R.W. Schafer, "Lossless Compression of Digital Audio", *IEEE Signal processing magazine*. Vol. 18 Nº. 4, páginas 21 a 32, 2001.

ES 2 305 852 T3

- [8] Lin **Xiao**, Li **Gang**, Li **Zhengguo**, Chia **Thien King**, Yoh **Ai Ling**, “A Novel Prediction Scheme for Lossless Compression of Audio Waveform”, *Proc. IEEE ICME 2001*, agosto, Japón.
- [9] Shorten: <http://www.softsound.com/Shorten.html>
- 5 [10] WaveZip: <http://www.gadgetlabs.com/wavezip.html>
- [11] LPAC: <http://www-ft.ee.tu-berlin.de/liebchen/>
- 10 [12] Wave Archiver: www.ecf.utoronto.ca/~denlee/wavarc.html
- [13] R. **Geiger**, T. **Sporer**, J. **Koller**, y K. **Brandenburg**, “Audio Coding based on Integer Transforms”, 111ª Convención AES, septiembre de 2001.
- 15 [14] J. **Johnston**, “Estimation of Perceptual Entropy”, *Proc. ICASSP 1988*.
- [15] R. **Yu**, C. C. **Ko**, X. **Lin** y S. **Rahardja**, “Bit-plane Golomb code for sources with Laplacian distributions”, *proceeding of ICASSP 2003*.
- 20 [16] Monkey’s Audio, <http://www.monkeysaudio.com>
- [17] S. H. **Park** y colaboradores, “Multi-Layer Bit-Sliced Bit Rate Scalable MPEG-4 Audio Coder”, presentado en la 103ª Convención de la AES, Nueva York, septiembre de 1997 (preimpresión 4520)
- 25 [18] Ralf **Geiger** y colaboradores, “FINE GRAIN SCALABLE PERCEPTUAL AND LOSSLESS AUDIO CODING BASED ON INTMDCT”, *Proceeding of ICASSP 2003*.
- [19] ISO/IEC 14496-3 Subparte 4, Information Technology - Coding of Audiovisual Objects, Parte 3. Audio, Subparte 4 Time/Frequency Coding, ISQ/JTC 1/SC 29/WG11, 1998
- 30 [20] T. **Painter**, A. **Spanias**, “Perceptual Coding of Digital Audio”, *IEEE Proceedings*, vol. 88, nº. 4, abril de 2000.
- [21] ISO/IEC 11172-3, “CODING OF MOVING PICTURES AND ASSOCIATED AUDIO FOR DIGITAL STORAGE MEDIA AT UP TO ABOUT 1,5 MBIT/s”, Parte 3 AUDIO.
- 35 [22] **Westen**, S.J.P., R.L. **Lagendijk**, y J. **Biemond**, “Optimization of JPEG color image coding using a human visual system model”, Conferencia SPIE sobre Visión Humana e Imagen Electrónica.
- 40 [23] **Westen**, S.J.P., R.L. **Lagendijk**, y J. **Biemond**, “Spatio-Temporal Model of Human Vision For Digital Video Compression”, SPIE Proceeding of Electronic Imaging 97.
- [24] J. **Li** y S. **Lie**, “An embedded still image coder with rate-distortion optimization”, *IEEE Trans. On Image Processing*, vol. 8, nº. 7, páginas 913 a 924, julio de 1999.
- 45 [25] W. **Li**, “Overview of Fine Granularity Scalability in MPEG-4 Video Standard”, *IEEE Trans. On Circuits and Systems for Video Technology*, vol. 11, nº. 3, páginas 301 a 317, marzo de 2001.
- 50
- 55
- 60
- 65

ES 2 305 852 T3

REIVINDICACIONES

1. Un procedimiento de codificación de una señal digital en un flujo binario escalable, que comprende:
 - 5 - cuantificar (102) la señal digital y codificar la señal cuantificada para formar un flujo binario de capa de núcleo;
 - realizar una correspondencia de errores (103) en base a la señal digital y al flujo binario de capa de núcleo para eliminar información que haya sido codificada en el flujo binario de capa de núcleo, dando como resultado una señal de error;
 - 10 - codificar en plano binario (104) la señal de error en base a la información perceptual de la señal digital, dando como resultado un flujo binario de capa de mejora, en el que la información perceptual de la señal digital se determina usando un modelo perceptual; y
 - 15 - multiplexar (105) el flujo binario de capa de núcleo y el flujo binario de capa de mejora, generando por lo tanto el flujo binario escalable.
2. El procedimiento de la reivindicación 1, comprendiendo de manera adicional:
 - 20 - transformar (101) la señal digital en un dominio adecuado;
 - en el que la señal transformada está cuantificada para formar la señal cuantificada antes de codificar la señal cuantificada.
- 25 3. El procedimiento de la reivindicación 1 ó 2,
 - en el que la información perceptual de la señal digital es además multiplexada (105) con el flujo binario de capa de núcleo y el flujo binario de capa de mejora para generar el flujo binario escalable.
- 30 4. El procedimiento de la reivindicación 2,
 - en el que la señal digital se transforma (101) en una señal digital transformada usando una Transformada de Coseno Discreta Modificada entera.
- 35 5. El procedimiento de la reivindicación 5,
 - en el que la señal transformada está normalizada para aproximar la salida de un banco de filtros MDCT.
- 40 6. El procedimiento de cualquiera de las reivindicaciones 1 a la 5,
 - en el que la señal digital o la señal digital transformada se cuantifica (102) y se codifica de acuerdo con la especificación de Codificación de Audio Avanzada (AAC) del Grupo de Expertos de Imágenes en Movimiento (MPEG).
- 45 7. El procedimiento de cualquiera de las reivindicaciones 1 a la 6,
 - en el que la correspondencia de errores (103) se realiza por medio de la resta del umbral de cuantificación más bajo correspondiente a cada uno de los valores cuantificados de la señal cuantificada de la señal digital o de la señal digital transformada, generando por lo tanto la señal de error.
- 50 8. El procedimiento de cualquiera de las reivindicaciones 1 a la 7,
 - en el que se usa un modelo psicoacústico como el modelo perceptual (106) para determinar la información perceptual de la señal digital.
- 55 9. El procedimiento de cualquiera de las reivindicaciones de la 1 a la 8,
 - en el que la señal de error está representada por planos binarios que comprenden una pluralidad de símbolos de plano binario, y en el que los planos binarios son desplazados en base a la información perceptual de la señal digital, de forma que los planos binarios que sean perceptualmente importantes son los que se codifican los primeros cuando los planos binarios sean explorados y sean codificados en una secuencia consecutiva durante la codificación de plano binario (104) de la señal de error.
- 60 10. El procedimiento de cualquiera de las reivindicaciones de la 1 a la 8,
 - 65 en el que la señal de error está representada por planos binarios que comprenden una pluralidad de símbolos de plano binario, y en el que los planos binarios y los símbolos de plano binario son explorados y

ES 2 305 852 T3

codificados durante la codificación de plano binario (104) de la señal de error en una secuencia en base a la información perceptual de la señal digital, de forma que los símbolos de plano binario de los planos binarios que sean más importantes perceptualmente son los que se codifican los primeros.

5 11. El procedimiento de la reivindicación 9 ó 10

en el que se determina al menos una de las siguientes informaciones como la información perceptual de la señal digital por medio del modelo perceptual (106):

- 10 - el plano binario de la señal de error con el que comienza la codificación de plano binario de la señal de error (Ms); y
- el nivel de la distorsión apenas perceptible (JND) de la señal digital, en donde s corresponde a la banda de frecuencia de la señal digital de la de señal digital transformada.

15

12. El procedimiento de la reivindicación 11,

en el que la significancia perceptual $P_s(s)$ de la señal digital se determina de manera adicional como la información perceptual, la significancia perceptual se determina por medio de:

20

- determinar el plano binario de la señal de error correspondiente al nivel JND $\tau(s)$ de la señal digital;
- restar el plano binario de la señal de error correspondiente al nivel JND $\tau(s)$ de la señal digital del plano binario de la señal de error con el que comienza la codificación de plano binario de la señal de error, determinando por lo tanto la significancia perceptual $P_s(s)$, en donde la significancia perceptual $P_s(s)$ se usa para controlar la secuencia de exploración y de codificación de al menos los planos binarios o de los símbolos de plano binario de los planos binarios.

25

13. El procedimiento de la reivindicación 12,

30

en el que la significancia perceptual $P_s(s)$ es normalizada por medio de

- definir una significancia perceptual común $P_s(s)_{\text{común}}$ en base a una función de la significancia perceptual $P_s(s)$; y
- restar la significancia perceptual común $P_s(s)_{\text{común}}$ de la significancia perceptual $P_s(s)$, generando por lo tanto la significancia perceptual normalizada $P_s'(s)$, en el que para la banda de frecuencia s para la que los valores cuantificados no son todos cero, el valor de la significancia perceptual $P_s(s)$ se fija al valor de la significancia perceptual común $P_s(s)_{\text{común}}$, y

35

40

en el que para la banda de frecuencia s para la que los valores cuantificados no son todos cero, la significancia perceptual normalizada $P_s'(s)$ se multiplexa con el flujo binario de capa de núcleo y con el flujo binario de capa de mejora para generar el flujo binario escalable.

45

14. El procedimiento de la reivindicación 11,

en el que el plano binario de la señal de error con la que empieza la codificación de plano binario de la señal de error se determina a partir del intervalo de cuantificación máximo usado en la banda de frecuencia s para cuantificar la señal digital o la señal transformada.

50

15. Un codificador para codificar una señal digital en un flujo binario escalable, que comprende:

- 55 - una unidad de cuantificación (102) para cuantificar la señal digital, y codificar la señal cuantificada para formar un flujo binario de capa de núcleo;
- una unidad de correspondencia de errores (103) para realizar una correspondencia de errores en base a la señal digital y al flujo binario de capa de núcleo para eliminar la información que se haya codificado en el flujo binario de capa de núcleo, dando como resultado una señal de error;
- 60 - una unidad de codificación de plano binario perceptual (104) para la codificación de plano binario de la señal de error en base a la información perceptual de la señal digital, dando como resultado un flujo binario de capa de mejora, en el que la información perceptual de la señal digital se determina usando un modelo perceptual (106); y
- 65 - una unidad de multiplexación (105) para multiplexar el flujo binario de capa de núcleo y el flujo binario de capa de mejora, generando por lo tanto el flujo binario escalable.

ES 2 305 852 T3

16. Un medio legible por un ordenador, con un programa grabado en el mismo, en el que el programa cuando se ejecuta por parte del ordenador, hace que el ordenador realice un procedimiento para codificar una señal digital en un flujo binario escalable, comprendiendo del procedimiento:

- 5 - cuantificar (102) la señal digital y codificar la señal cuantificada para formar un flujo binario de capa de núcleo;
- realizar una correspondencia de errores (103) en base a la señal digital y al flujo binario de capa de núcleo para eliminar la información que se haya codificado en el flujo binario de capa de núcleo, dando como
10 resultado una señal de error;
- codificación de plano binario (104) de la señal de error en base a la información perceptual de la señal digital, dando como resultado un flujo binario de capa de mejora, en el que la información perceptual de la señal digital se determina usando un modelo perceptual; y
- 15 - multiplexar (105) el flujo binario de capa de núcleo y el flujo binario de capa de mejora, generando por lo tanto el flujo binario escalable.

17. Un elemento de programa de ordenador que, cuando es ejecutado por un ordenador, hace que el ordenador realice un procedimiento para codificar una señal digital en un flujo binario escalable, comprendiendo el procedimiento:

- 20 - cuantificar (102) la señal digital, y codificar la señal cuantificada para formar un flujo binario de capa de núcleo;
- 25 - realizar una correspondencia de errores (103) en base a la señal digital y al flujo binario de capa de núcleo para eliminar la información que se haya codificado en el flujo binario de capa de núcleo, dando como resultado un error en la señal;
- 30 - codificación de plano binario (104) de la señal de error en base a la información perceptual de la señal digital, dando como resultado un flujo binario de capa de mejora, en el que la información perceptual de la señal digital se determina usando un modelo perceptual (106); y
- multiplexar (105) el flujo binario de capa de núcleo y el flujo binario de capa de mejora, generando por lo tanto el flujo binario escalable.

35 18. Un procedimiento para descodificar el flujo binario escalable en una señal digital, que comprende:

- demultiplexar (205) el flujo binario escalable en un flujo binario de capa de núcleo y en un flujo binario de capa de mejora;
- 40 - descodificar (204) y descuantificar (202) el flujo binario de capa de núcleo para generar la señal de capa de núcleo;
- descodificación de plano binario del flujo binario de capa de mejora en base a la información perceptual de la señal digital; y
- 45 - realizar una correspondencia de errores (203) en base al flujo binario de capa de mejora descodificado en plano binario y a la señal de capa de núcleo descuantificada, dando como resultado una señal transformada reconstruida, en la que la señal transformada reconstruida es la señal digital.

50 19. El procedimiento de la reivindicación 18, comprendiendo de manera adicional:

 transformar (201) la señal transformada reconstruida en una señal reconstruida, en la que la señal reconstruida es la señal digital.

55 20. El procedimiento de la reivindicación 18 ó de la reivindicación 19,

 en el que la información perceptual de la señal digital se obtiene a partir de la demultiplexación (205) del flujo binario escalable.

60 21. El procedimiento de la reivindicación 19 ó de la reivindicación 20,

 en el que la señal de capa de núcleo y la señal de capa de mejora son transformadas usando una Transformada de Coseno Discreta Modificada entera MDCT).

65 22. El procedimiento de cualquiera de las reivindicaciones 18 a la 21,

 en el que el flujo binario de capa de núcleo se descodifica (204) y se descuantifica (202) de acuerdo con la especificación de Codificación de Audio Avanzada (AAC) del Grupo de Expertos de Imágenes en Movimiento.

ES 2 305 852 T3

23. El procedimiento de cualquiera de las reivindicaciones 18 a la 22,

en el que la correspondencia de errores (203) se realiza por medio de la suma del umbral de cuantificación más bajo usado para la descuantificación (202) de la señal transformada y el flujo binario de capa de mejora descodificado de plano binario, generando por lo tanto la señal de capa de mejora.

24. El procedimiento de cualquiera de las reivindicaciones 18 a la 23,

en el que el flujo binario de capa de mejora está descodificado en plano binario para generar una pluralidad de planos binarios comprendiendo una pluralidad de símbolos de plano binario en una secuencia consecutiva, y los planos binarios se desplazan en base a la información perceptual de la señal digital para generar el flujo binario de capa de mejora descodificado de plano binario.

25. El procedimiento de cualquiera de las reivindicaciones 18 a la 23,

en el que el flujo binario de capa de mejora está descodificado en plano binario para generar una pluralidad de planos binarios comprendiendo una pluralidad de símbolos de plano binario en una secuencia basada en la información perceptual de la señal digital, generando por lo tanto el flujo binario de capa de mejora descodificado de plano binario.

26. El procedimiento de la reivindicación 24 ó 25,

en el que se recibe al menos una de la siguiente información como la información perceptual de la señal digital:

- el plano binario que corresponde al flujo binario de capa de mejora cuando comienza la descodificación del plano binario del flujo binario de capa de mejora, cuyo plano binario está especificado por un número $M(s)$; y
- el nivel de distorsión apenas perceptible (JND) de la señal digital, en el que s corresponde a una banda de frecuencias de la señal digital.

27. El procedimiento de la reivindicación 26,

en el que el plano binario que corresponde al flujo binario de capa de mejora cuando se inicia la descodificación de plano binario del flujo binario de capa de mejora $M(s)$ se determina a partir del intervalo máximo de cuantificación usado en la banda de frecuencia s para la descuantificación (202) del flujo binario de capa de núcleo.

28. Un descodificador para descodificar un flujo binario escalable en una señal digital comprendiendo:

- una unidad de demultiplexación (205) para demultiplexar el flujo binario escalable en un flujo binario de capa de núcleo y en un flujo binario de capa de mejora;
- una unidad de descuantificación (202) para descodificar y para descuantificar el flujo binario de capa de núcleo para generar una señal de capa de núcleo;
- una unidad de descodificación de plano binario (204) para la descodificación de plano binario del flujo binario de capa de mejora en base a la información perceptual de la señal digital; y
- una unidad de correspondencia de errores (203) para realizar una correspondencia de errores en base al flujo binario de capa de mejora descodificado de plano binario y a la señal de capa de núcleo descuantificada, dando como resultado una señal transformada reconstruida, en la que la señal transformada reconstruida es la señal digital.

29. Un medio legible por un ordenador que tiene un programa grabado dentro, en donde el programa cuando es ejecutado por un ordenador, hace que el ordenador realice un procedimiento para descodificar un flujo binario escalable en una señal digital, comprendiendo el procedimiento:

- demultiplexar (205) el flujo binario escalable en un flujo binario de capa de núcleo y un flujo binario de capa de mejora;
- descodificar y descuantificar el flujo binario de capa de núcleo para generar una señal de capa de núcleo;
- descodificar en plano binario (204) el flujo binario de capa de mejora en base a la información perceptual de la señal digital; y

ES 2 305 852 T3

- realizar una correspondencia de errores en base al flujo binario de capa de mejora descodificado de plano binario y a la señal de capa de núcleo descuantificada, dando como resultado una señal transformada reconstruida, en la que la señal transformada reconstruida es la señal digital.

5 30. Un elemento de programa de ordenador que, cuando es ejecutado por un ordenador, hace que el ordenador realice un procedimiento para descodificar un flujo binario escalable en una señal digital, comprendiendo el procedimiento:

10 - demultiplexar (205) del flujo binario escalable en un flujo binario de capa de núcleo y en un flujo binario de capa mejorada;

 - descodificar y descuantificar (202) el flujo binario de capa de núcleo para generar una señal de capa de núcleo;

15 - descodificar en plano binario (204) el flujo binario de capa de mejora en base a la información perceptual de la señal digital; y

20 - realizar una correspondencia de error (203) en base al flujo binario de capa mejorada descodificado de plano binario y a la señal de capa de núcleo descuantificada, dando como resultado una señal transformada reconstruida, en la que la señal transformada reconstruida es la señal digital.

25

30

35

40

45

50

55

60

65

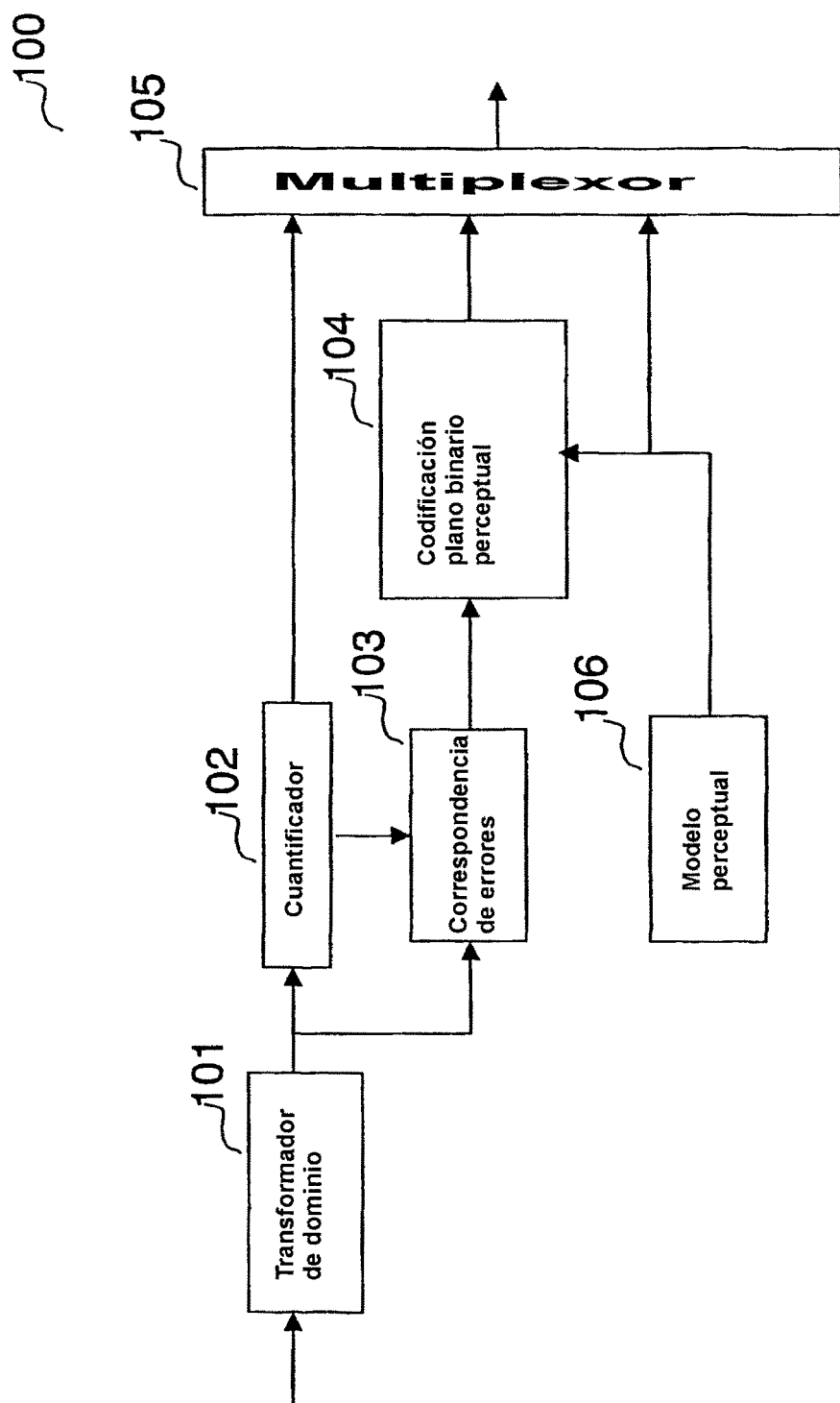


Fig. 1

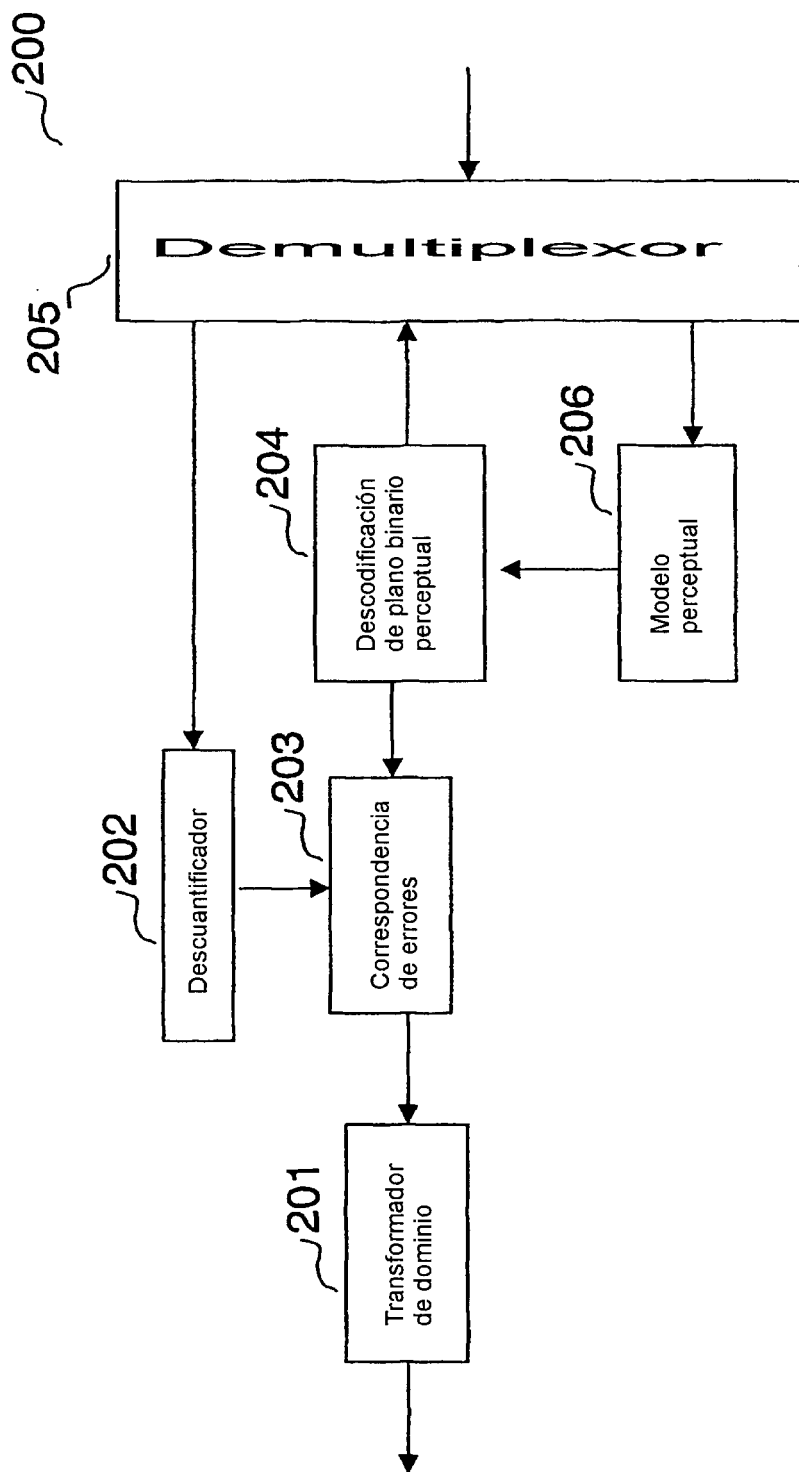


Fig. 2

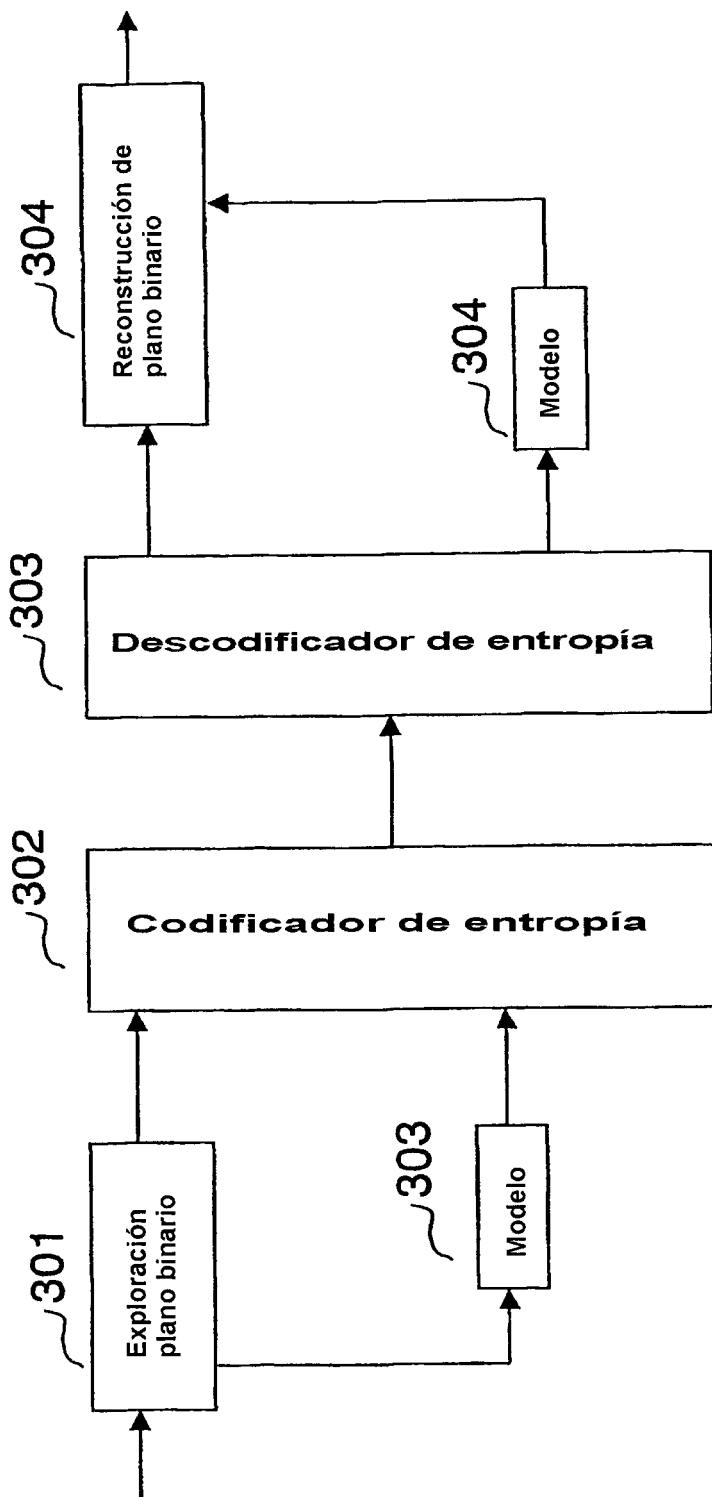


Fig. 3

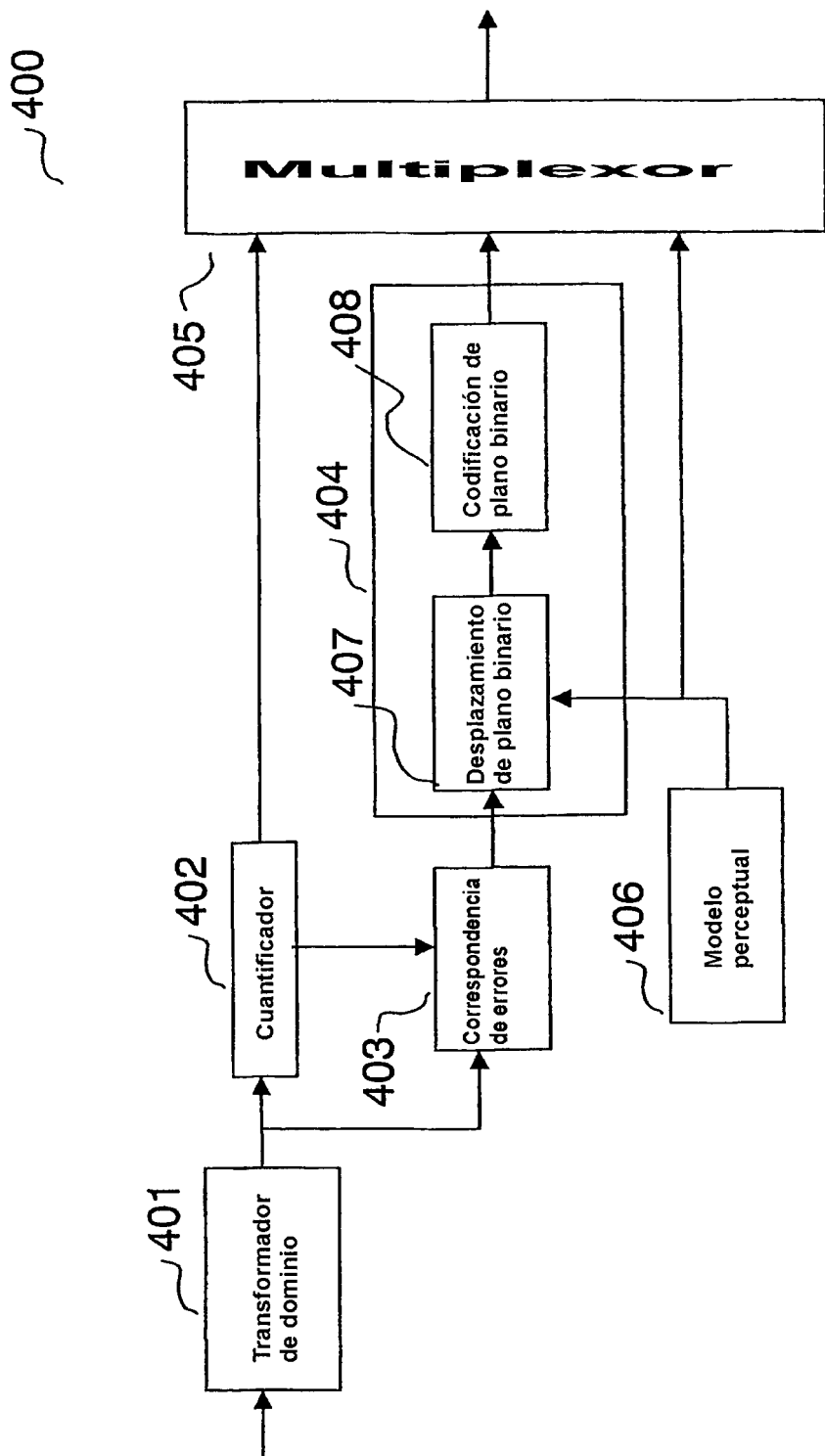


Fig. 4

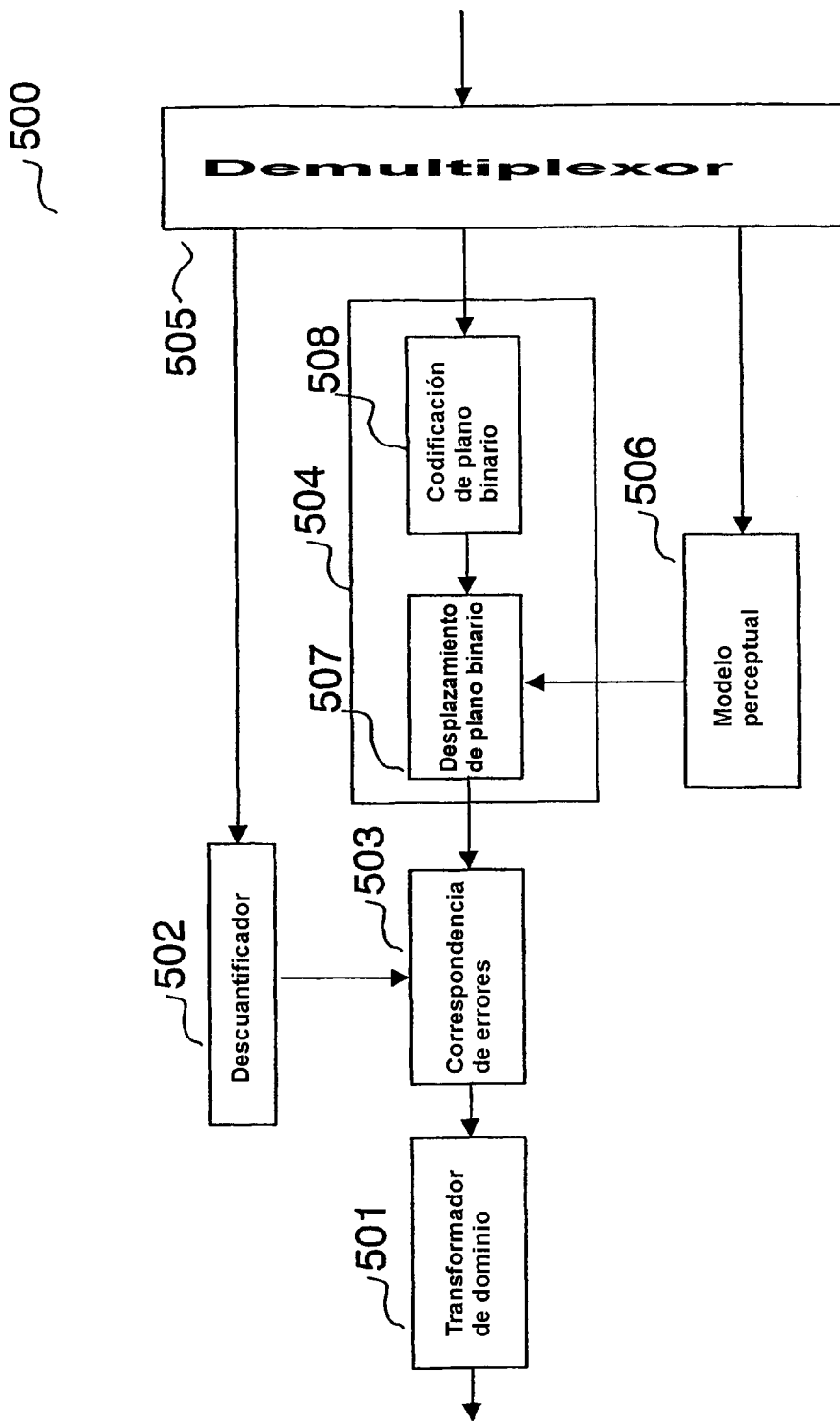


Fig. 5

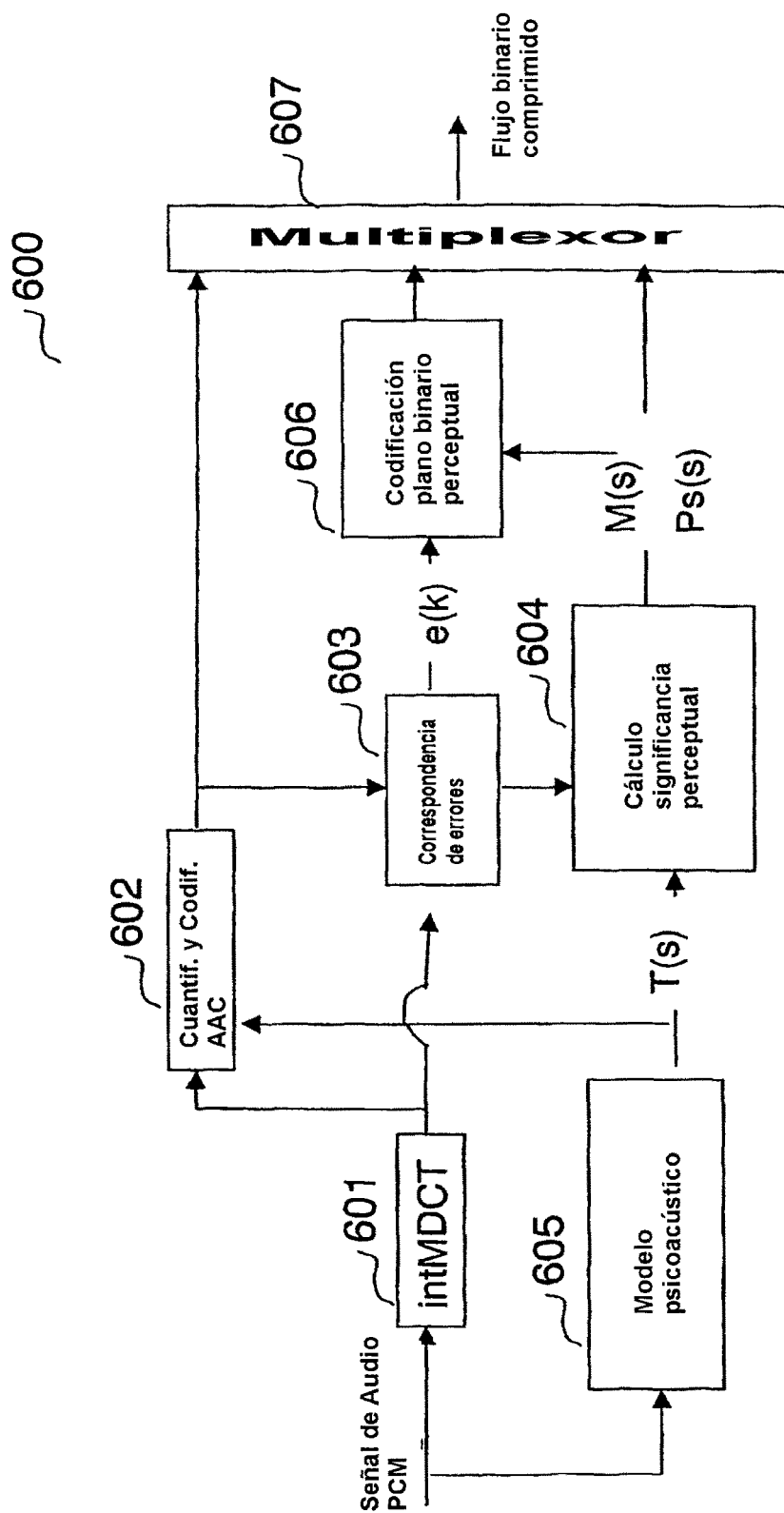


Fig. 6

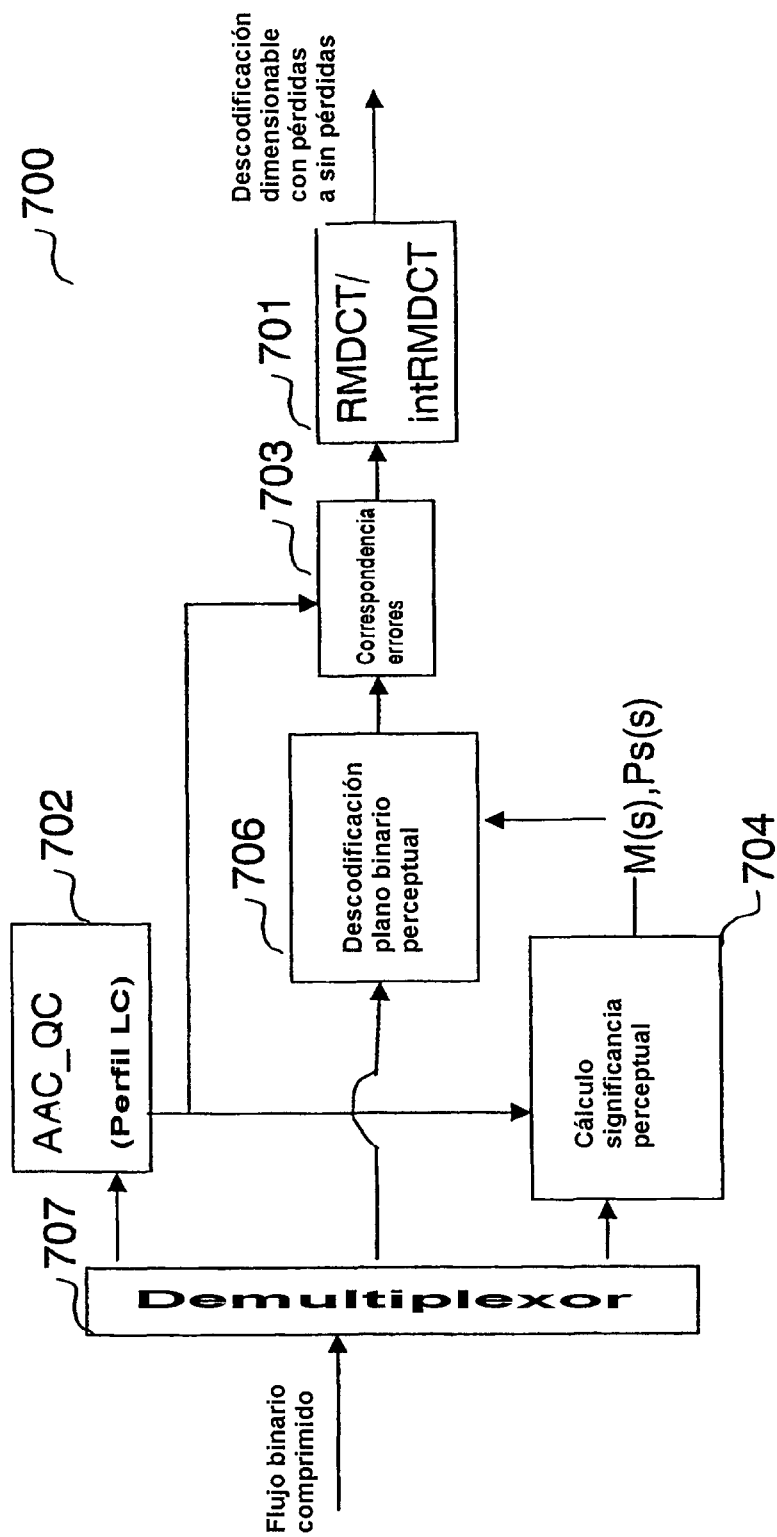


Fig. 7