



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 338 891**

51 Int. Cl.:
G10K 15/12 (2006.01)
H04S 5/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **05017013 .3**
96 Fecha de presentación : **10.07.2002**
97 Número de publicación de la solicitud: **1603119**
97 Fecha de publicación de la solicitud: **07.12.2005**

54 Título: **Control adaptativo de cola de eco para síntesis de audio pseudoestereofónica.**

30 Prioridad: **10.07.2001 SE 2001102481**
15.03.2002 SE 2002100796
09.07.2002 SE 2002102159

45 Fecha de publicación de la mención BOPI:
13.05.2010

45 Fecha de la publicación del folleto de la patente:
13.05.2010

73 Titular/es: **Dolby Sweden AB.**
Gävlegatan 12A
113 30 Stockholm, SE

72 Inventor/es: **Henn, Frederik;**
Kjörling, Kristofer;
Liljeryd, Lars;
Röden, Jonas y
Engdegard, Jonas

74 Agente: **Arizti Acha, Mónica**

ES 2 338 891 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

Control adaptativo de cola de eco para síntesis de audio pseudoestereofónica.

5 **Campo técnico**

La presente invención se refiere a sistemas de codificación de fuente de audio a baja velocidad de transferencia de bits. Se introducen diversas representaciones paramétricas de propiedades estereofónicas de una señal de entrada y se explica la aplicación de las mismas en el lado del decodificador, abarcando desde codificación pseudoestereofónica a
10 codificación estereofónica completa de envolventes espectrales, siendo la última de éstas especialmente adecuada para códecs basados en HFR (recombinación de alta frecuencia).

Antecedentes de la invención

Las técnicas de codificación de fuente de audio pueden dividirse en dos clases: codificación de audio natural y
15 codificación de voz. A las velocidades de transferencia de bits medias a altas, la codificación de audio natural se utiliza normalmente para señales de música y voz, y es posible la transmisión y reproducción estereofónica. En aplicaciones en las que sólo se dispone de bajas velocidades de transferencia de bits, por ejemplo, en transmisiones de audio en flujo continuo (*streaming*) por Internet dirigidas a usuarios con conexiones telefónicas por módem lentas, o en los sistemas
20 de radiodifusión digital AM emergentes, es inevitable la codificación monofónica del material del programa de audio. Sin embargo, todavía puede desearse una sensación estereofónica, en particular cuando se escucha con auriculares, en cuyo caso se percibe una señal monofónica pura como si proviniese de “dentro de la cabeza”, lo cual puede resultar una experiencia desagradable.

Un enfoque para tratar este problema es sintetizar una señal estereofónica en el lado del decodificador a partir de una señal monofónica pura recibida. A lo largo de los años se han propuesto varios generadores “pseudoestereofónicos” diferentes. Por ejemplo, en la patente estadounidense 5.883.962 se describe la mejora de señales monofónicas por medio de la adición de versiones desfasadas/retardadas de una señal a la señal sin procesar, creando con ello una ilusión estereofónica. Con ello, la señal procesada se añade a la señal original para cada una de las dos salidas a niveles
30 iguales pero con signos opuestos, garantizando que las señales de mejora se cancelen si los dos canales se añaden posteriormente a la trayectoria de la señal. En el documento PCT WO 98/57436 se muestra un sistema similar, aunque sin la compatibilidad monofónica anterior de la señal mejorada. Los métodos de la técnica anterior tienen en común que se aplican como procesos únicamente posteriores. En otras palabras, no se facilita al decodificador información alguna acerca del grado de amplitud estereofónica, dejando a un lado la posición en la etapa de sonido estereofónica. De esta manera, la señal pseudoestereofónica puede asemejarse o no al carácter estereofónico de la señal original. Una situación particular en la que los sistemas de la técnica anterior resultan deficientes es cuando la señal original es una señal monofónica pura, lo cual es a menudo el caso en las grabaciones de voz. Esta señal monofónica se convierte a ciegas
35 en una señal estereofónica sintética en el decodificador, lo cual en el caso de la voz origina artefactos perturbadores y puede reducir la claridad y la inteligibilidad de la voz.

Otros sistemas de la técnica anterior dirigidos a la transmisión estereofónica real a bajas velocidades de transferencia de bits emplean normalmente un esquema de codificación de sumas y restas. De esta manera, las señales originales izquierda (L) y derecha (R) se convierten en una señal de suma, $S = (L+R)/2$, y una señal de resta, $D = (L-R)/2$, y seguidamente se codifican y transmiten. El receptor decodifica las señales S y D, recreándose la señal L/R original a
45 través de las operaciones $L = S + D$, y $R = S - D$. La ventaja de esto es que con gran frecuencia se encuentra en la banda una redundancia entre L y R, siendo menos la información en D que debe codificarse y requiriendo menos bits que en S. Claramente, el caso extremo es una señal monofónica pura, es decir, L y R son idénticas. Un códec L/R convencional codifica esta señal monofónica dos veces, mientras que un códec S/D detecta esta redundancia, y la señal D no requiere (de forma ideal) ningún bit en absoluto. Otro extremo lo representa la situación en la que $R = -L$, correspondiente a señales “fuera de fase”. Ahora, la señal S es cero, mientras que la señal D computa para L. Nuevamente, el esquema S/D tiene una clara ventaja frente a la codificación L/R estándar. Sin embargo, considérese la situación en la que, por ejemplo, $R = 0$ durante una transición, lo cual no era poco frecuente en los primeros
50 tiempos de las grabaciones estereofónicas. Tanto S como D son iguales a $L/2$, y el esquema S/D no ofrece ninguna ventaja. Por el contrario, la codificación L/R trata esto muy bien: la señal R no requiere ningún bit. Por esta razón, los códecs de la técnica anterior emplean conmutación adaptativa entre estos dos esquemas de codificación, dependiendo de qué método es más beneficioso para usarlo en un momento dado. Los ejemplos anteriores son meramente teóricos (excepto en el caso monofónico dual, que es común en los programas de sólo voz). De esta manera, el material de los programas estereofónicos del mundo real contiene importantes cantidades de información estereofónica, e incluso si se lleva a cabo la conmutación anterior, la velocidad de transferencia de bits resultante a menudo es aún demasiado
60 alta para muchas aplicaciones. Además, tal como puede observarse de las relaciones de resintetización anteriores, no es fiable una cuantificación muy poco definida de la señal D en un intento de reducir adicionalmente la velocidad de transferencia de bits dado que los errores de cuantificación se traducen en errores de nivel que no pueden descuidarse en las señales L y R.

Para controlar la cantidad de reverberación añadida por el generador pseudoestereofónico, se conoce según el documento US 5.559.891 controlar el nivel de reverberación añadida a la señal original. Según el documento adicional EP 0 989 543 A2, se conoce adoptar diferentes funciones de desvanecimiento para conformar la parte de cola de la reverberación.

Un objeto de la presente invención es proporcionar una técnica de reverberación mejorada. Este objeto se consigue mediante una unidad de reverberación según la reivindicación 1 y un método según la reivindicación 4.

La memoria descriptiva describe la detección de propiedades estereofónicas de señal antes de la codificación y la transmisión. En la forma más sencilla, un detector mide la cantidad de perspectiva estereofónica que está presente en la señal estereofónica de entrada. Esta cantidad se transmite entonces como un parámetro de amplitud estereofónica, junto con una suma monofónica codificada de la señal original. El receptor decodifica la señal monofónica y aplica la cantidad apropiada de amplitud estereofónica, usando un generador pseudoestereofónico, que se controla mediante dicho parámetro. Como caso especial, una señal de entrada monofónica se señala como de amplitud estereofónica cero y de manera correspondiente no se aplica síntesis estereofónica en el decodificador. Pueden obtenerse medidas útiles de la amplitud estereofónica, por ejemplo, de la señal de diferencia o de la correlación cruzada del canal derecho e izquierdo original. El valor de tales computaciones puede mapearse con un pequeño número de estados, que se transmiten a una velocidad de transferencia fija en el tiempo, o según sea necesario. La memoria descriptiva también enseña cómo filtrar los componentes estereofónicos sintetizados, con el fin de reducir el riesgo de desenmascarar artefactos de codificación que normalmente están asociados con señales codificadas de baja velocidad de transferencia.

De manera alternativa, el equilibrio estereofónico total o localización en el campo estereofónico se detecta en el codificador. Esta información, opcionalmente junto con el parámetro de amplitud anterior, se transmite eficazmente como un parámetro de equilibrio, junto con la señal monofónica codificada. De esta manera, los desplazamientos a cualquier lado de la etapa de sonido pueden recrearse en el decodificador alterando de forma correspondiente las ganancias de los dos canales de salida. Este parámetro de equilibrio estereofónico puede derivarse del cociente de las potencias de señales izquierda y derecha. La transmisión de los dos tipos de parámetros requiere muy pocos bits, en comparación con la codificación estereofónica completa, con lo cual se mantiene reducida la demanda total de velocidad de transferencia de bits. En una versión adicional, que ofrece una descripción estereofónica paramétrica más precisa, se utilizan varios parámetros de equilibrio y amplitud estereofónica, representando cada uno bandas de frecuencia independientes.

El parámetro de equilibrio, generalizado a una operación por banda de frecuencia, junto con una operación por banda correspondiente de un parámetro de nivel, calculado como la suma de las potencias de señal izquierda y derecha, permite una nueva representación, detallada de forma arbitraria, de la densidad espectral de potencia de una señal estereofónica. Un beneficio particular de esta representación, además de los beneficios de la redundancia estereofónica, de la cual también sacan ventaja los sistemas S/D, es que la señal de equilibrio puede cuantificarse con menos precisión que el nivel mencionado, dado que el error de cuantificación, al convertirse nuevamente a una envolvente espectral estereofónica, ocasiona un "error en el espacio", es decir, la localización percibida en el panorama estereofónico, en lugar de un error de nivel. De forma análoga a un sistema L/R y S/D conmutado tradicional, el esquema nivel/equilibrio puede interrumpirse de forma adaptativa en favor de una señal de nivel L/nivel R, que es más eficaz cuando la señal total está intensamente desfasada hacia cualquier canal. El esquema anterior de codificación de envolvente espectral puede utilizarse cada vez que se requiera una codificación eficaz de envolventes espectrales de potencia, y puede incorporarse como una herramienta en los nuevos códecs de fuente estereofónica. Una aplicación particularmente interesante es en sistemas HFR que se guían mediante información acerca de la envolvente de banda alta de la señal original. En un sistema de este tipo, la banda baja se codifica y decodifica por medio de un códec arbitrario, y la banda alta se regenera en el decodificador utilizando la señal de banda baja decodificada y la información de envolvente de banda alta transmitida [PCT WO 98/57436]. Además, se ofrece la posibilidad de construir un códec estereofónico basado en HFR ajustable a escala sincronizando la codificación de envolvente con la operación de nivel/equilibrio. Con ello, los valores de nivel se alimentan al flujo primario de bits que, dependiendo de la implementación, decodifica normalmente a una señal monofónica. Los valores de equilibrio se alimentan al flujo secundario de bits que está disponible, además del flujo primario de bits, para receptores cercanos al transmisor, tomando como ejemplo un sistema de radiodifusión AM digital IBOC (*In-Band On-Channel*). Cuando se combinan los dos flujos de bits, el decodificador produce una señal de salida estereofónica. Además de los valores de nivel, el flujo primario de bits puede contener parámetros estereofónicos, por ejemplo, un parámetro de amplitud. De esta manera, la decodificación de este flujo de bits únicamente ya produce una salida estereofónica que se mejora cuando están disponibles ambos flujos de bits.

La presente descripción se describirá ahora a modo de ejemplos ilustrativos, sin limitar el alcance o el espíritu de la invención, en relación con los dibujos adjuntos, en los que:

la figura 1 ilustra un sistema de codificación de fuente que contiene un codificador mejorado mediante un módulo codificador estereofónico paramétrico, y un decodificador mejorado mediante un módulo decodificador estereofónico paramétrico,

la figura 2a es un bloque esquemático de un módulo decodificador estereofónico paramétrico,

la figura 2b es un bloque esquemático de un generador pseudoestereofónico con entradas de parámetros de control,

la figura 2c es un bloque esquemático de un ajustador de equilibrio con entradas de parámetros de control,

la figura 3 es un bloque esquemático de un módulo decodificador estereofónico paramétrico que utiliza generación pseudoestereofónica multibanda combinada con ajuste de equilibrio multibanda,

la figura 4a es un bloque esquemático del lado del codificador de un códec estereofónico basado en HFR ajustable a escala, que emplea codificación de nivel/equilibrio de la envolvente espectral,

la figura 4b es un bloque esquemático del lado del decodificador correspondiente.

Las realizaciones descritas más adelante son puramente ilustrativas para los principios de la presente invención. Se entiende que las modificaciones y variaciones de las disposiciones y los detalles descritos en la presente memoria resultarán evidentes para otros expertos en la técnica. Por tanto, la intención es limitarse sólo al alcance de las reivindicaciones que se exponen posteriormente, y no el quedar limitada por los detalles específicos presentados a modo de descripción y explicación de las realizaciones de la presente memoria. Para mayor claridad, todos los ejemplos que se muestran a continuación asumirán sistemas de dos canales, pero como es evidente para otros expertos en la técnica, los métodos pueden aplicarse a sistemas multicanal, tales como un sistema 5.1.

La figura 1 muestra cómo puede mejorarse un sistema de codificación de fuente arbitrario que comprende un codificador, 107, y un decodificador, 115, funcionando el codificador y decodificador en el modo monoaural, mediante la codificación estereofónica paramétrica según la invención. L y R indican las señales de entrada analógicas izquierda y derecha, que se alimentan a un convertidor AD, 101. La salida del convertidor AD se convierte a monofónica, 105, y la señal monofónica se codifica, 107. Adicionalmente, la señal estereofónica se dirige a un codificador estereofónico paramétrico, 103, que calcula uno o varios parámetros estereofónicos que se describirán a continuación. Estos parámetros se combinan con la señal monofónica codificada por medio de un multiplexor, 109, formando un flujo de bits, 111. El flujo de bits se almacena o se transmite y posteriormente se extrae en el lado del decodificador por medio de un demultiplexor, 113. La señal monofónica se decodifica, 115, y se convierte en una señal estereofónica mediante un decodificador estereofónico paramétrico, 119, que utiliza el (los) parámetro(s) estereofónicos, 117, como señal(es) de control. Finalmente, la señal estereofónica se dirige al convertidor DA, 121, que alimenta las salidas analógicas, L' y R'. La topología según la figura 1 es común a un conjunto de métodos de codificación estereofónica paramétrica que se describirá detalladamente, comenzando con las versiones menos complejas.

Un método de parametrización de propiedades estereofónicas es determinar la amplitud estereofónica de la señal original en el lado del codificador. Una primera aproximación de la amplitud estereofónica es la señal de resta $D = L - R$, ya que, por así decirlo, un alto grado de similitud entre L y R computa para un valor pequeño de D y viceversa. Un caso especial es el caso monofónico dual en el que $L = R$ y, por tanto, $D = 0$. De esta manera, incluso este sencillo algoritmo es capaz de detectar el tipo de señal de entrada monofónica comúnmente asociada a las emisiones de noticias, en cuyo caso no se desea pseudoestéreo. Sin embargo, una señal monofónica que se alimenta a L y R a diferentes niveles no produce una señal D cero, incluso si la amplitud percibida es cero. Por tanto, en la práctica podrían necesitarse detectores más elaborados que emplean, por ejemplo, métodos de correlación cruzada. Habría que asegurarse de que el valor que describe la diferencia o correlación izquierda-derecha esté normalizado de alguna manera con el nivel total de señal para conseguir un detector independiente del nivel. Un problema con el detector mencionado anteriormente es el caso en el que se mezcla voz monofónica con una señal estereofónica mucho más débil, por ejemplo, un ruido estereofónico o música de fondo durante transiciones voz a música/música a voz. Durante las pausas de la voz, el detector indicará entonces una señal estereofónica amplia. Esto se solventa normalizando el valor de amplitud estereofónica con una señal que contiene información del nivel de energía total anterior, por ejemplo, una señal de disminución de pico de la energía total. Además, para impedir que el detector de amplitud estereofónica se active por un ruido de alta frecuencia o una distorsión de alta frecuencia de diferente canal, las señales del detector deben filtrarse previamente mediante un filtro paso bajo, normalmente con una frecuencia de corte algo por encima de un segundo formante de la voz y opcionalmente también mediante un filtro paso alto para evitar desfases de señal desequilibradas o zumbidos. Sin tener en cuenta el tipo de detector, la amplitud estereofónica calculada se correlaciona con un conjunto finito de valores que cubren el intervalo entero, de monofónico a estereofónico amplio.

La figura 2a proporciona un ejemplo de los contenidos del decodificador estereofónico paramétrico presentado en la figura 1. El bloque designado como "equilibrio", 211, controlado mediante el parámetro B, se describirá más adelante, y debe considerarse que se ha saltado de momento. El bloque denominado "amplitud", 205, toma una señal de entrada monofónica y recrea sintéticamente la sensación de amplitud estereofónica, controlándose la magnitud de la amplitud mediante el parámetro W. Los parámetros opcionales S y D se describirán más adelante. Según esta técnica, a menudo puede conseguirse una calidad de sonido subjetivamente mejor incorporando un filtro de cruce que comprende un filtro paso bajo, 203, y un filtro paso alto, 201, para mantener la gama de baja frecuencia "ajustada" y sin verse afectada. En el presente documento sólo la salida del filtro paso alto se dirige al bloque de amplitud. La salida estereofónica procedente del bloque de amplitud se añade a la salida monofónica del filtro paso bajo mediante 207 y 209, formando la señal de salida estereofónica.

Cualquier generador pseudoestereofónico de la técnica anterior puede utilizarse para el bloque de amplitud, tal como los mencionados en la sección de los antecedentes, o una unidad de simulación de reflexión temprana de tipo Schroeder (retardo *multi*tap) o reverberador. La figura 2b da un ejemplo de un generador pseudoestereofónico, alimentado mediante una señal M monofónica. La cantidad de amplitud estereofónica se determina por la ganancia de 215, y esta ganancia es una función del parámetro de amplitud estereofónica, W. Cuanto más alta sea la ganancia, más amplia es la sensación estereofónica, una ganancia cero corresponde a la reproducción monofónica pura. La salida de 215 se retarda, 221, y se añade, 223 y 225, a las dos instancias de señal directas, empleando signos opuestos. Para no alterar de manera significativa el nivel total de reproducción cuando se cambia la amplitud estereofónica, puede incorporarse

una atenuación de compensación de la señal directa, 213. Por ejemplo, si la ganancia de la señal retardada es G , la ganancia de la señal directa puede seleccionarse como $\sqrt{1 - G^2}$. Según esta técnica, una atenuación progresiva de alta frecuencia puede incorporarse en la trayectoria de la señal de retardo, 217, que ayuda a evitar el enmascaramiento pseudoestereofónico de artefactos de codificación. Opcionalmente, los filtros de cruce, los filtros de atenuación progresiva y los parámetros de retardo pueden enviarse en el flujo de bits, ofreciendo más posibilidades para imitar las propiedades estereofónicas de la señal original, tal como se muestra también en las figuras 2a y 2b como las señales X, S y D. Si se utiliza una unidad de reverberación para generar una señal estereofónica, la disminución de reverberación podría a veces no desearse justo al final de un sonido. Según la invención, estas colas de reverberación no deseadas pueden atenuarse de manera sencilla o eliminarse completamente mediante la simple alteración de la ganancia de la señal de reverberación. Puede utilizarse para este fin un detector diseñado para encontrar terminaciones de sonidos. Si la unidad de reverberación genera artefactos en algunas señales específicas, por ejemplo, perturbaciones transitorias, puede utilizarse también un detector de esas señales para atenuarlas.

Un método alternativo para detectar propiedades estereofónicas según la invención se describe de la siguiente manera. De nuevo, L y R indican las señales de entrada izquierda y derecha. Las potencias de señal correspondientes vienen dadas entonces por $P_L \sim L^2$ y $P_R \sim R^2$. Ahora puede calcularse una medida del equilibrio estereofónico como el cociente entre las dos potencias de señal, o más específicamente como $B = (P_L + e)/(P_R + e)$, donde e es un número arbitrario muy pequeño que elimina la división por cero. El parámetro de equilibrio, B , puede expresarse en dB dado mediante la relación $B_{dB} = 10\log_{10}(B)$. Como ejemplo, los tres casos $P_L = 10P_R$, $P_L = P_R$ y $P_L = 0,1 P_R$ corresponden a valores de equilibrio de +10 dB, 0 dB, y -10 dB respectivamente. Claramente, estos valores representan las localizaciones “izquierda”, “centro” y “derecha”. Los experimentos han demostrado que el intervalo del parámetro de equilibrio puede limitarse, por ejemplo, a +/-40 dB, dado que esos valores extremos ya se perciben como si el sonido se originara completamente desde uno de los dos altavoces o controladores de auriculares. Esta limitación reduce el espacio de la señal que ha de cubrirse en la transmisión, ofreciendo así reducción de la velocidad de transferencia de bits. Además, puede emplearse un esquema de cuantificación progresiva por el que se utilizan pasos de cuantificación más pequeños alrededor de cero y pasos más grandes hacia los límites exteriores, lo que reduce adicionalmente la velocidad de transferencia de bits. A menudo el equilibrio es constante en el tiempo para transiciones extendidas. Por tanto, puede llevarse a cabo una última etapa para reducir de manera significativa el número de bits promedio necesarios: después de la transmisión de un valor de equilibrio inicial, sólo se transmiten las diferencias entre valores de equilibrio consecutivos, empleándose la codificación de entropía. Con mucha frecuencia esta diferencia es cero, lo cual se señala, por tanto, mediante la palabra de código más corta posible. Claramente, en aplicaciones en las que son posibles errores de bits, esta codificación delta debe reajustarse a un intervalo de tiempo adecuado para eliminar la propagación incontrolada de errores.

El uso más rudimentario por el decodificador del parámetro de equilibrio es simplemente desfazar la señal monofónica hacia uno de los dos canales de reproducción, alimentando la señal monofónica a las dos salidas y ajustando las ganancias de manera correspondiente, tal como se ilustra en la figura 2c, bloques 227 y 229, con la señal de control B. Esto es análogo a girar el botón “panorama” en una mesa de mezclas, “moviendo” sintéticamente una señal monofónica entre los dos altavoces estereofónicos.

El parámetro de equilibrio puede enviarse adicionalmente al parámetro de amplitud descrito anteriormente, ofreciendo la posibilidad de colocar y extender la imagen del sonido en la etapa de sonido de una manera controlada, ofreciendo flexibilidad al simular la sensación estereofónica original. Un problema con la combinación de la generación pseudoestereofónica, tal como se mencionó en una parte anterior, y el equilibrio controlado por parámetros es la aportación no deseada de señales desde el generador pseudoestereofónico en posiciones de equilibrio alejadas de la posición central. Esto se soluciona aplicando una función que favorece el carácter monofónico al valor de la amplitud estereofónica, dando como resultado una atenuación mayor del valor de amplitud estereofónica en posiciones de equilibrio en la posición lateral extrema y menor atenuación o ninguna atenuación en las posiciones de equilibrio cercanas a la posición central.

Los métodos descritos hasta ahora están concebidos para aplicaciones con una velocidad muy baja de transferencia de bits. En aplicaciones en las que se dispone de velocidades de transferencia de bits más altas es posible utilizar versiones más elaboradas de los métodos anteriores de amplitud y equilibrio. La detección de la amplitud estereofónica puede hacerse en varias bandas de frecuencias, dando como resultado valores de amplitud estereofónica individuales para cada banda de frecuencias. De manera similar, el cálculo de equilibrio puede funcionar de una manera multibanda, que es equivalente a aplicar diferentes curvas de filtro a dos canales que se alimentan por una señal monofónica. La figura 3 muestra un ejemplo de un decodificador estereofónico paramétrico que utiliza un conjunto de N generadores pseudoestereofónicos según la figura 2b, representados mediante los bloques 307, 317 y 327, combinados con un ajuste de equilibrio multibanda, representado mediante los bloques 309, 319 y 329, tal como se describe en la figura 2c. Las bandas de paso individuales se obtienen alimentando la señal de entrada monofónica, M , a un conjunto de filtros paso banda, 305, 315 y 325. Se añaden las salidas estereofónicas de la banda de paso procedentes de los ajustadores de equilibrio, 311, 321, 313, 323, formando la señal de salida estereofónica, L y R. Los parámetros de equilibrio y amplitud anteriormente escalares se reemplazan ahora por las disposiciones $W(k)$ y $B(k)$. En la figura 3, cada generador pseudoestereofónico y ajustador de equilibrio tiene parámetros estereofónicos únicos. Sin embargo, para reducir la cantidad total de datos que han de transmitirse o almacenarse, puede calcularse la media de parámetros de varias bandas de frecuencias en grupos en el codificador, y correlacionarse este número más pequeño de parámetros con los grupos correspondientes de bloques de amplitud y equilibrio en el decodificador. Claramente, pueden utilizarse diferentes esquemas y longitudes de agrupación para las disposiciones $W(k)$ y $B(k)$. $S(k)$ representa las ganancias de

las trayectorias de señal de retardo en los bloques de amplitud, y $D(k)$ representa los parámetros de retardo. De nuevo, $S(k)$ y $D(k)$ son opcionales en el flujo de bits.

El método de codificación de equilibrio paramétrico, especialmente para bandas de frecuencias más bajas, puede dar un comportamiento algo inestable debido a la falta de resolución de frecuencia o debido a demasiados eventos de sonido que suceden al mismo tiempo en una banda de frecuencia pero en diferentes posiciones de equilibrio. Estos problemas de equilibrio se caracterizan normalmente por un valor desviado de equilibrio durante simplemente un corto periodo de tiempo, normalmente uno o unos pocos valores consecutivos calculados, dependiendo de la velocidad de actualización. Para evitar problemas de equilibrio perturbadores puede aplicarse un proceso de estabilización en los datos de equilibrio. Este proceso puede utilizar un número de valores de equilibrio antes y después de la posición de tiempo actual, para calcular el valor medio de éstos. En consecuencia, el valor medio puede utilizarse como un valor limitador para el valor de equilibrio actual, es decir, no debería permitirse que el valor de equilibrio actual fuese más allá del valor medio. El valor actual queda limitado entonces por el intervalo entre el último valor y el valor medio. De manera opcional, puede permitirse que el valor de equilibrio actual traspase los valores limitados un determinado factor de exceso. Además, el factor de exceso, así como el número de valores de equilibrio utilizados para calcular la media deberían verse como propiedades dependientes de la frecuencia y, por tanto, ser individuales para cada banda de frecuencias.

A bajas relaciones de actualización de la información de equilibrio, la falta de resolución temporal puede provocar fallos en la sincronización entre los movimientos de la imagen estereofónica y los eventos de sonido actuales. Para mejorar este comportamiento en cuanto a la sincronización puede emplearse un esquema de interpolación basado en identificar eventos de sonido. Aquí la interpolación se refiere a interpolaciones entre dos valores de equilibrio consecutivos en el tiempo. Estudiando la señal monofónica en el lado del receptor, puede obtenerse información sobre los inicios y los finales de diferentes eventos de sonidos. Una manera es detectar un incremento o disminución repentina de la energía de la señal en una banda de frecuencias concreta. La interpolación, tras el guiado a partir de esa envolvente de energía en el tiempo, debería asegurar que los cambios en la posición de equilibrio deberían realizarse preferiblemente durante segmentos de tiempo que contienen poca energía de señal. Dado que el oído humano es más sensible a las partes de entrada que a las de salida de un sonido, el esquema de interpolación se beneficia de encontrar el comienzo de un sonido, aplicando, por ejemplo, retención de pico a la energía y dejando entonces que los aumentos de valor de equilibrio sean una función de la energía de retención de pico, donde un valor de energía pequeño da un gran incremento y viceversa. Para segmentos de tiempo que contienen energía distribuida uniformemente en el tiempo, por ejemplo, como para algunas señales estacionarias, este método de interpolación iguala la interpolación lineal entre los dos valores de equilibrio. Si los valores de equilibrio son cocientes de energías derecha e izquierda, se prefieren los valores de equilibrio logarítmicos, por razones de simetría izquierda-derecha. Otra ventaja de aplicar el algoritmo de interpolación completo en el dominio logarítmico es la tendencia del oído humano a relacionar niveles a una escala logarítmica.

Asimismo, a bajas relaciones de actualización de los valores de ganancia de amplitud estereofónica, puede aplicarse interpolación a los mismos. Una manera sencilla es interpolar linealmente entre dos valores de amplitud estereofónica consecutivos en el tiempo. Un comportamiento más estable de la amplitud estereofónica puede conseguirse suavizando los valores de ganancia de amplitud estereofónica en un segmento de tiempo más largo que contiene varios parámetros de amplitud estereofónica. Utilizando suavizado con diferentes constantes de tiempo de ataque y de liberación se consigue un sistema muy apropiado para material de programa que contiene voz y música mezclados o intercalados. Se produce un diseño apropiado de este tipo de filtro de suavizado utilizando una constante de tiempo de ataque corta para conseguir un breve tiempo de subida y, por lo tanto, una respuesta inmediata a entradas de música en estéreo, y un largo tiempo de liberación para conseguir un largo tiempo de caída. Para poder conmutar rápidamente de un modo estereofónico amplio a un modo monofónico, lo que puede ser deseable para entradas de voz repentinas, hay una posibilidad de saltar o reajustar el filtro de suavizado señalizando este evento. Además, las constantes de tiempo de ataque, las constantes de tiempo de liberación y otras características de filtro de suavizado también pueden señalizarse mediante un codificador.

Para señales que contienen distorsión enmascarada procedente de un códec psicoacústico, un problema común al introducir información estereofónica basada en la señal monofónica codificada es un efecto de desenmascaramiento de la distorsión. Ese fenómeno denominado comúnmente “desenmascaramiento estereofónico” es el resultado de sonidos no centrados que no cumplen el criterio de enmascaramiento. El problema con el desenmascaramiento estereofónico puede resolverse o resolverse parcialmente introduciendo, por el lado del decodificador, un detector destinado a estas situaciones. Las tecnologías conocidas para medir las relaciones señal a máscara pueden utilizarse para detectar un posible desenmascaramiento estereofónico. Una vez detectado puede señalarse explícitamente o los parámetros estereofónicos simplemente pueden disminuirse.

En el lado del codificador, una opción, enseñada por la invención, es emplear un transformador Hilbert en la señal de entrada, por ejemplo, se introduce un desfase de 90 grados entre los dos canales. Al formarse posteriormente la señal monofónica sumando las dos señales, se consigue un mejor equilibrio entre una señal monofónica centrada y señales estereofónicas “verdaderas” dado que la transformación de Hilbert introduce una atenuación de 3 dB para la información de centro. En la práctica esto mejora la codificación monofónica de, por ejemplo, música pop contemporánea, en la que, por ejemplo, los cantantes solistas y el bajo se graban normalmente utilizando una única fuente monofónica.

El método de parámetro de equilibrio multibanda no se limita al tipo de aplicación descrito en la figura 1. Puede utilizarse ventajosamente siempre que el objetivo sea codificar de manera eficaz la envolvente espectral de potencia de una señal estereofónica. Por tanto, puede emplearse como herramienta en códecs estereofónicos en los que, además de la envolvente espectral estereofónica, se codifica un residuo estereofónico correspondiente. La potencia P total se define por $P = P_L + P_R$, donde P_L y P_R son potencias de señal, tal como se ha descrito anteriormente. Obsérvese que esta definición no tiene en cuenta las relaciones de fase izquierda a derecha. (Por ejemplo, señales idénticas derecha e izquierda, pero con signo opuesto, no producen una potencia total cero). Análogamente a B , P puede expresarse en dB como $P_{dB} = 10 \log_{10} (PIP_{ref})$ donde P_{ref} es una potencia de referencia arbitraria y los valores delta pueden codificarse por entropía. En contraposición al caso del equilibrio, se emplea cuantificación no progresiva para P . Para representar la envolvente espectral de una señal estereofónica, P y B se calculan para un conjunto de bandas de frecuencia, normalmente, pero no necesariamente, con anchos de banda que están relacionados con las bandas críticas del oído humano. Por ejemplo, dichas bandas pueden formarse mediante agrupación de canales en un banco de filtros de ancho de banda constante, calculándose P_L y P_R como los promedios en tiempo y frecuencia de los cuadrados de las muestras de subbanda que corresponden a la banda y período temporales respectivos. Los conjuntos $P_0, P_1, P_2, \dots, P_{N-1}$ y $B_0, B_1, B_2, \dots, B_{N-1}$, en los que los subíndices indican la banda de frecuencia en una representación de N bandas, se codifican a modo delta y Huffman, se transmiten o se almacenan, y finalmente se decodifican en los valores cuantificados que se calcularon en el codificador. La última etapa es convertir P y B de nuevo en P_L y P_R . Tal como puede observarse fácilmente a partir de las definiciones de P y B , las relaciones inversas son (al ignorar e en la definición de B) $P_L = BP/(B+1)$ y $P_R = P/(B+1)$.

Una aplicación especialmente interesante del método anterior de codificación de envolvente es codificar envolventes espectrales de banda alta para códecs basados en HFR. En este caso no se transmite ninguna señal residual de banda alta. En su lugar, este residuo se deriva de la banda baja. Por tanto, no hay una relación estricta entre representación residual y representación de envolvente, y la cuantificación de envolvente es más decisiva. Para estudiar los efectos de la cuantificación, P_q y B_q indican los valores cuantificados de P y B respectivamente. P_q y B_q se insertan entonces en las relaciones anteriores y se forma la suma: $P_L q + P_R q = B_q P_q / (B_q + 1) + P_q / (B_q + 1) = P_q (B_q + 1) / (B_q + 1) = P_q$. La característica interesante aquí es que se elimina B_q y el error en la potencia total se determina únicamente mediante el error de cuantificación en P . Esto implica que incluso aunque B se cuantifique intensamente, el nivel percibido es correcto suponiendo que se utiliza suficiente precisión en la cuantificación de P . En otras palabras, la distorsión en B se correlaciona con una distorsión en el espacio, más que en nivel. Mientras que las fuentes de sonido sean estacionarias en el espacio a lo largo del tiempo, esta distorsión en la perspectiva estereofónica es también estacionaria y difícil de notar. Como ya se expuso, la cuantificación del equilibrio estereofónico también puede ser menos precisa hacia los extremos exteriores dado que un error dado en dB corresponde a un error menor en el ángulo percibido cuando el ángulo respecto a la línea central es grande, debido a las propiedades del oído humano.

Al cuantificar datos dependientes de la frecuencia, por ejemplo, valores de ganancia de amplitud estereofónica multibanda o valores de equilibrio multibanda, puede seleccionarse de manera ventajosa la resolución y el rango del método de cuantificación para ajustarse a las propiedades de una escala de percepción. Si tal escala se hace en función de la frecuencia, pueden elegirse diferentes métodos de cuantificación, o las llamadas clases de cuantificación, para las diferentes bandas de frecuencia. Los valores de parámetros codificados que representan las diferentes bandas de frecuencia deberían interpretarse entonces en algunos casos, incluso si tienen valores idénticos, de diferentes maneras, es decir, decodificarse en valores diferentes.

De manera análoga a un esquema de codificación conmutado L/R a S/D, las señales P y B pueden sustituirse de manera adaptativa por las señales P_L y P_R , para hacer frente mejor a las señales extremas. Como se enseñó mediante el documento PCT/SE00/00158, la codificación delta de muestras de envolvente puede conmutarse de delta en tiempo a delta en frecuencia dependiendo de qué dirección es más eficaz con respecto al número de bits en un momento particular. El parámetro de equilibrio puede beneficiarse también de este esquema: considérese, por ejemplo, una fuente que se mueve en el tiempo por el campo estereofónico. Claramente, esto corresponde a un cambio sucesivo de valores de equilibrio a lo largo del tiempo que, dependiendo de la velocidad de la fuente frente a la velocidad de actualización de los parámetros, puede corresponder a valores grandes de delta en tiempo, correspondiendo a grandes palabras de código cuando se emplea la codificación por entropía. Sin embargo, asumiendo que la fuente tiene radiación de sonido uniforme frente a la frecuencia, los valores de delta en frecuencia del parámetro de equilibrio son cero en cualquier punto en el tiempo, correspondiendo de nuevo a palabras de código pequeñas. Por tanto, en este caso se consigue una velocidad de transferencia de bits más baja al utilizar la dirección de codificación de frecuencia delta. Otro ejemplo es una fuente que es estacionaria en el espacio, pero tiene una radiación no uniforme. Ahora, los valores delta en frecuencia son grandes y la elección preferida es delta en tiempo.

El esquema de codificación P/B ofrece la posibilidad de construir un códec HFR ajustable a escala, véase la figura 4. Un códec ajustable a escala se caracteriza porque el flujo de bits se divide en dos o más partes, siendo opcional la recepción y la decodificación de partes de mayor orden. El ejemplo supone dos partes de flujo de bits, denominadas en lo sucesivo primaria, 419, y secundaria, 417, aunque también es claramente posible la extensión a un número mayor de partes. El lado del codificador, figura 4a, comprende un codificador estereofónico de banda baja arbitrario, 403, que opera en la señal de entrada estereofónica, IN (no se muestran en la figura las etapas triviales de conversión AD o respectivamente DA), un codificador estereofónico paramétrico que estima la envolvente espectral de banda alta y, opcionalmente, parámetros estereofónicos adicionales, 401, que también opera en la señal de entrada estereofónica, y dos multiplexores, 415 y 413, para los flujos de bits primario y secundario respectivamente. En esta aplicación, la codificación de envolvente de banda alta se sincroniza con la operación P/B, y la señal P , 407, se envía

al flujo primario de bits mediante 415, mientras que la señal B, 405, se envía al flujo secundario de bits, mediante 413.

Para el códec de banda baja existen diferentes posibilidades: puede operar de manera constante en el modo S/D, y las señales S y D pueden enviarse a los flujos de bits primario y secundario respectivamente. En este caso, una decodificación del flujo primario de bits da como resultado una señal monofónica de banda completa. Por supuesto, esta señal monofónica puede mejorarse mediante métodos estereofónicos paramétricos según la invención, en cuyo caso el (los) parámetro(s) estereofónicos también deben situarse en el flujo primario de bits. Otra posibilidad es alimentar una señal estereofónica de banda baja codificada al flujo primario de bits, opcionalmente junto con parámetros de equilibrio y de amplitud de banda alta. Ahora, la decodificación del flujo primario de bits da como resultado estéreo verdadero para la banda baja, y un pseudo estéreo muy realista para la banda alta dado que las propiedades estereofónicas de la banda baja se reflejan en la reconstrucción de alta frecuencia. Dicho de otra manera: incluso aunque la representación de envolvente de banda alta disponible o la estructura espectral poco definida estén en modo monofónico, el residuo de banda alta sintetizada o la estructura espectral fina, no lo están. En este tipo de implementación, el flujo secundario de bits puede contener más información de banda baja que, combinada con la del flujo primario de bits, produce una reproducción de banda baja de mayor calidad. La topología de la figura 4 ilustra ambos casos dado que las señales primaria y secundaria de salida del codificador de banda baja, 411 y 409, conectadas a 415 y 417 respectivamente, pueden contener cualquiera de los tipos de señal descritos anteriormente.

Los flujos de bits se transmiten o se almacenan y se alimenta o bien sólo 419 o bien tanto 419 como 417, al decodificador, figura 4b. El flujo primario de bits se demultiplexa mediante 423 en la señal 429 primaria del decodificador central de baja banda y la señal P, 431. De manera similar, el flujo secundario de bits se demultiplexa mediante 421 en la señal 427 secundaria del decodificador central de baja banda y la señal B, 425. La(s) señal(es) de baja banda se dirigen al decodificador 433 de banda baja, que produce una salida 435, que de nuevo, en caso de sólo decodificar el flujo primario de bits, puede ser de cualquiera de los tipos descritos anteriormente (monofónica o estereofónica). La señal 435 alimenta la unidad HFR, 437, generándose una banda alta sintética y se ajusta según P, que también se conecta a la unidad HFR. La banda baja decodificada se combina con la banda alta en la unidad HFR, y la banda baja y/o la banda alta se mejoran opcionalmente mediante un generador pseudoestereofónico (también situado en la unidad HFR) antes de alimentarse finalmente a las salidas del sistema, formando la señal de salida, OUT. Cuando el flujo 417 secundario de bits está presente, la unidad HFR también obtiene la señal B como una señal 425 de entrada, y 435 está en modo estereofónico, produciendo el sistema una señal de salida estereofónica completa y los generadores pseudoestereofónicos, si hay alguno, se saltan.

Dicho en otras palabras, un método para codificar propiedades estereofónicas de una señal de entrada incluye, en un codificador, la etapa de calcular un parámetro de amplitud que indica una amplitud estereofónica de dicha señal de entrada, y en un decodificador, una etapa de generar una señal de salida estereofónica, usando dicho parámetro de amplitud para controlar una amplitud estereofónica de dicha señal de salida. El método comprende además en dicho codificador, formar una señal monofónica a partir de dicha señal de entrada, en el que, en dicho decodificador, dicha generación implica un método pseudoestereofónico que opera en dicha señal monofónica. El método implica además dividir dicha señal monofónica en dos señales así como añadir una versión (versiones) retardada(s) de dicha señal monofónica a dichas dos señales, a un nivel(es) controlados(s) por dicho parámetro de amplitud. El método incluye además que dicha(s) versión (versiones) retardada(s) se filtren paso alto y se atenúen de manera progresiva a frecuencias más altas antes de añadirse a dichas dos señales. El método incluye además que dicho parámetro de amplitud es un vector y los elementos de dicho vector corresponden a bandas de frecuencia independientes. El método incluye además que si dicha señal de entrada es de tipo monofónico dual, dicha señal de salida es también de tipo monofónico dual.

Un método para codificar propiedades estereofónicas de una señal de entrada incluye, en un codificador, calcular un parámetro de equilibrio que indica un equilibrio estereofónico de dicha señal de entrada y, en un decodificador, generar una señal de salida estereofónica usando dicho parámetro de equilibrio para controlar un equilibrio estereofónico de dicha señal de salida.

En este método, en dicho codificador, se forma una señal monofónica a partir de dicha señal de entrada y, en dicho decodificador, dicha generación implica dividir dicha señal monofónica en dos señales, y dicho control implica el ajuste de niveles de dichas dos señales. El método incluye además que se calcula una potencia para cada canal de dicha señal de entrada y dicho parámetro de equilibrio se calcula a partir de un cociente entre dichas potencias. El método incluye además que dichas potencias y dicho parámetro de equilibrio son vectores, correspondiendo cada elemento a una banda de frecuencia específica. El método incluye además que en dicho decodificador se interpola entre dos valores consecutivos en el tiempo de dichos parámetros de equilibrio de manera que el valor momentáneo de la potencia correspondiente de dicha señal monofónica controla qué inclinación debería tener la interpolación momentánea. El método incluye además que dicho método de interpolación se realiza sobre valores de equilibrio representados como valores logarítmicos. El método incluye además que dichos valores de parámetros de equilibrio están limitados a un intervalo entre un valor de equilibrio previo y un valor de equilibrio extraído de otros valores de equilibrio mediante un filtro de media u otro proceso de filtro, en el que dicho intervalo puede extenderse adicionalmente moviendo los bordes de dicho intervalo un determinado factor. El método incluye además que dicho método de extraer bordes de límite para valores de equilibrio es, para un sistema multibanda, en función de la frecuencia. El método incluye además que se calcula un parámetro de nivel adicional como una suma de vectores de dichas potencias y se envía a dicho decodificador, proporcionando así a dicho decodificador una representación de una envolvente espectral de dicha

ES 2 338 891 T3

señal de entrada. El método incluye además que dicho parámetro de nivel y dicho parámetro de equilibrio se sustituyen de manera adaptativa por dichas potencias. El método incluye además que dicha envolvente espectral se usa para controlar un proceso HFR en un decodificador. El método incluye además que dicho parámetro de nivel se alimenta a un flujo primario de bits de un códec estereofónico basado en HFR ajustable a escala y dicho parámetro de equilibrio se alimenta a un flujo secundario de bits de dicho códec. Dicha señal monofónica y dicho parámetro de amplitud se alimentan a dicho flujo primario de bits. Además, dichos parámetros de amplitud se procesan por una función que da valores más pequeños para un valor de equilibrio que corresponde a una posición de equilibrio más alejada de la posición central. El método incluye además que una cuantificación de dicho parámetro de equilibrio emplea pasos de cuantificación más pequeños alrededor de una posición central y pasos más grandes hacia posiciones exteriores. El método incluye además que dichos parámetros de amplitud y dichos parámetros de equilibrio se cuantifican usando un método de cuantificación en términos de resolución e intervalo que, para un sistema multibanda, es en función de la frecuencia. El método incluye además que dicho parámetro de equilibrio se codifica de modo delta de manera adaptativa en tiempo o en frecuencia. El método incluye además que dicha señal de entrada se pasa a través de un transformador Hilbert antes de formar dicha señal monofónica.

Un aparato para la codificación estereofónica paramétrica incluye, en un codificador, medios para calcular un parámetro de amplitud que indica una amplitud estereofónica de una señal de entrada, y medios para formar una señal monofónica a partir de dicha señal de entrada y, en un decodificador, medios para generar una señal de salida estereofónica a partir de dicha señal monofónica usando dicho parámetro de amplitud para controlar una amplitud estereofónica de dicha señal de salida.

REIVINDICACIONES

1. Unidad de reverberación para generar un primer canal y un segundo canal de una señal estereofónica o una señal multicanal, que comprende:

un detector para detectar terminaciones de sonido o señales específicas, para los que la unidad de reverberación generaría artefactos; y

un atenuador para atenuar o eliminar completamente cualquier cola de reverberación alterando la ganancia de una señal de reverberación.

2. Unidad de reverberación según la reivindicación 1, en la que la señal específica incluye transitorios y en la que el detector es un detector de transitorios.

3. Unidad de reverberación según la reivindicación 1 o la reivindicación 2, que comprende además:

medios de generación de una señal de reverberación usando una ganancia alterable; y

en la que el atenuador es operativo para alterar la ganancia de los medios de generación.

4. Método de reverberación para generar un primer canal y un segundo canal de una señal estereofónica o una señal multicanal, que comprende:

detectar terminaciones de sonido o señales específicas, para los que la unidad de reverberación generaría artefactos; y

atenuar o eliminar completamente cualquier cola de reverberación.

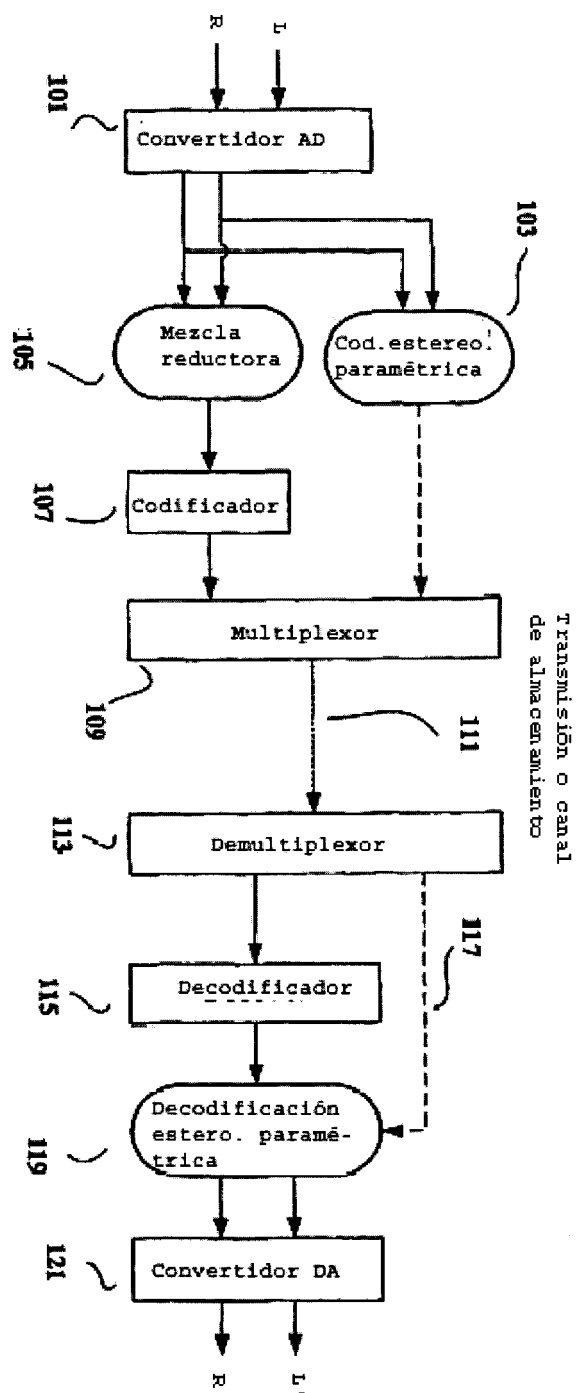


Fig. 1

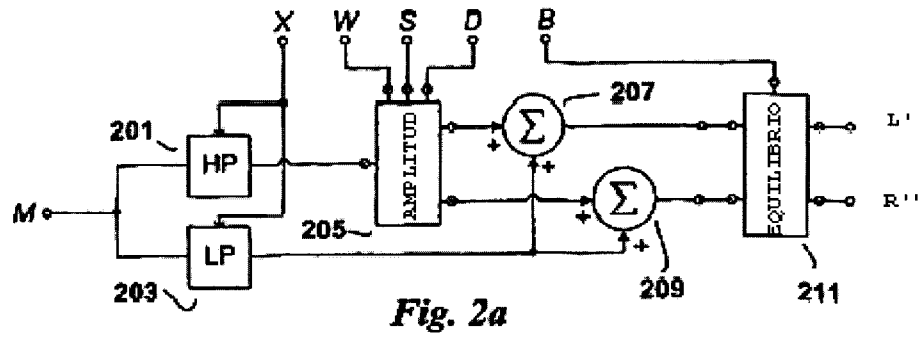


Fig. 2a

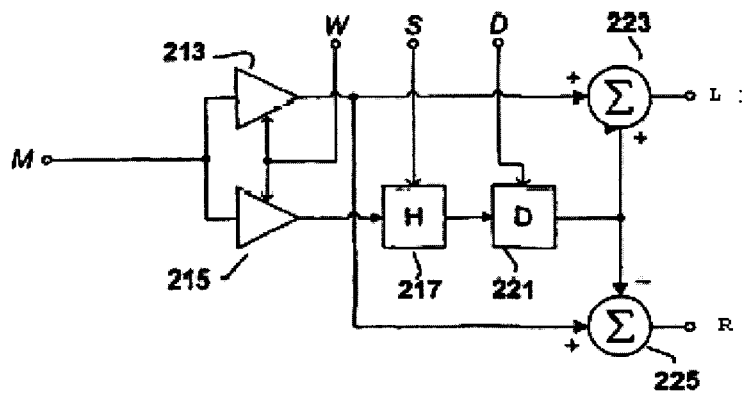


Fig. 2b

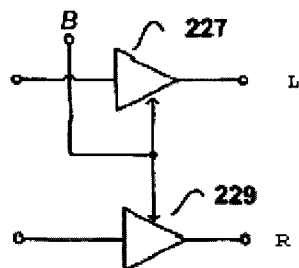


Fig. 2c

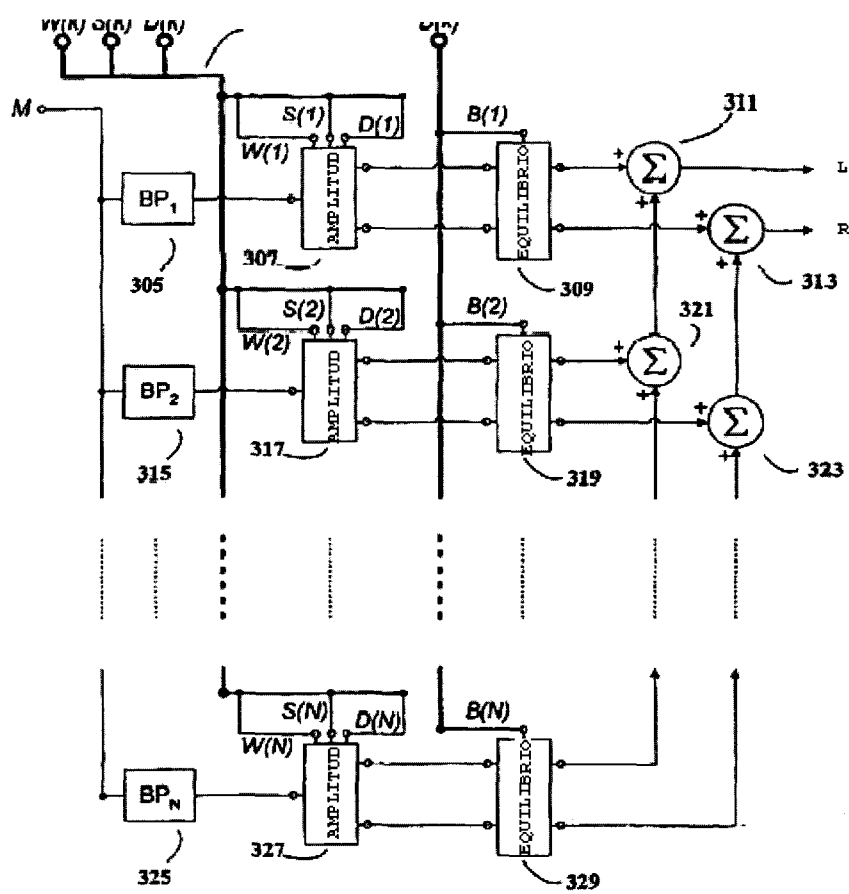


Fig. 3

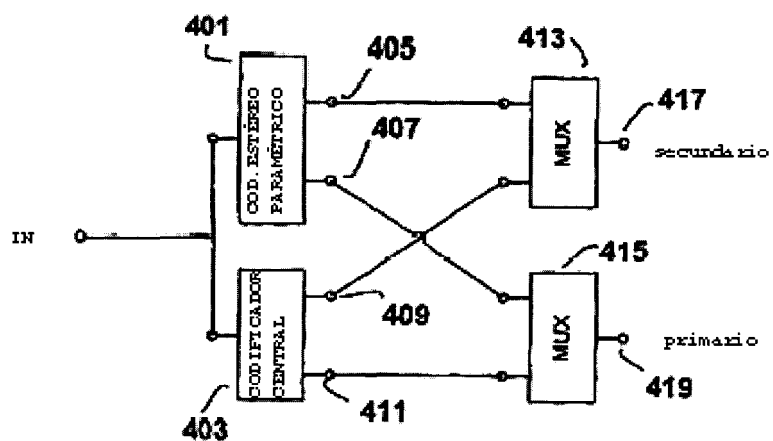


Fig. 4a

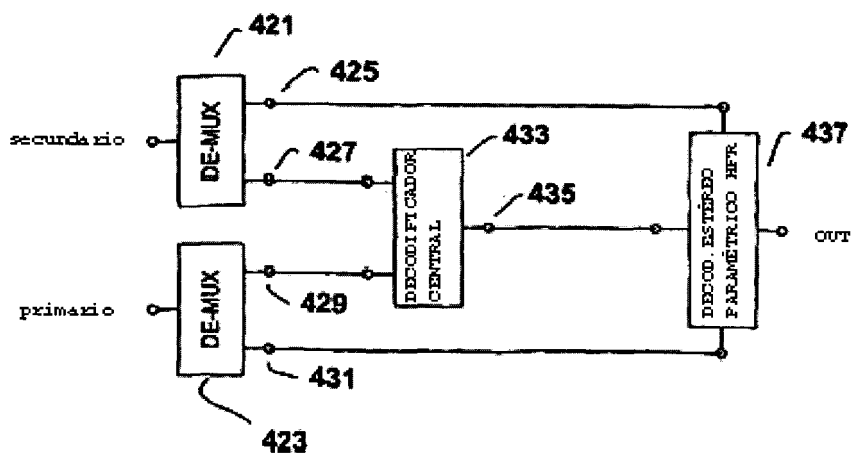


Fig. 4b