(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 432 625**

(51) Int. CI.:

G10L 19/24 (2013.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **07.12.2009 E 09799786 (0)**

(97) Fecha y número de publicación de la concesión europea: **18.09.2013 EP 2382627**

(54) Título: **Cálculo de máscara de escalamiento selectiva basado en detección de picos**

(30) Prioridad:

29.12.2008 US 345141

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
04.12.2013

(73) Titular/es:

MOTOROLA MOBILITY LLC (100.0%)
600 North US Highway 45
Libertyville, IL 60048, US

(72) Inventor/es:

ASHLEY, JAMES P. y
MITTAL, UDAR

(74) Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

DESCRIPCIÓN

Cálculo de máscara de escalamiento selectiva basado en detección de picos.

Referencia a solicitudes relacionadas

La presente invención está relacionada con las siguientes solicitudes de patente europea, que junto con esta solicitud son propiedad de Motorola Mobility, Inc.:

Solicitud EP 2 382 621 A0, titulada "METHOD AND APPARATUS FOR GENERATING AN ENHANCEMENT LAYER WITHIN A MULTIPLE-CHANNEL AUDIO CODING SYSTEM";

Solicitud EP 2 382 622 A0, titulada "METHOD AND APPARATUS FOR GENERATING AN ENHANCEMENT LAYER WITHIN A MULTIPLE-CHANNEL AUDIO CODING SYSTEM"; y

Solicitud EP 2 382 626 A0, titulada "SELECTIVE SCALING MASK COMPUTATION BASED ON PEAK DETECTION".

Campo de la descripción

La presente descripción se refiere en general a sistemas de comunicación y, más en particular, a codificación de voz y señales de audio en dichos sistemas de comunicación.

Antecedentes

La compresión de señales de audio y de voz digital es bien conocida. Generalmente, se requiere compresión para transmitir eficientemente señales sobre un canal de comunicaciones, o para almacenar señales comprimidas en un dispositivo multimedia digital, tal como un dispositivo de memoria de estado sólido o un disco duro de ordenador. Si bien existen muchas técnicas de compresión (o "codificación"), un método que ha seguido siendo muy popular para la codificación digital de la voz se conoce como predicción lineal con excitación por código (CELP, Code Excited Linear Prediction), que es uno de la familia de algoritmos de codificación de "análisis por síntesis". Análisis por síntesis se refiere, en general, a un proceso de codificación mediante el que se utilizan múltiples parámetros de un modelo digital para sintetizar un conjunto de señales candidatas que se comparan con una señal de entrada y cuya distorsión se analiza. A continuación, un conjunto de parámetros que proporciona la mínima distorsión es transmitido o bien almacenado, y eventualmente utilizado para reconstruir una estimación de la señal de entrada original. CELP es un método particular de análisis por síntesis que utiliza uno o varios libros de códigos, cada uno de los cuales comprende esencialmente conjuntos de vectores de código que se recuperan del libro de códigos en respuesta a un índice del libro de códigos.

En los codificadores CELP actuales, existe un problema para mantener la reproducción de voz y audio de alta calidad a velocidades de datos razonablemente bajas. Esto es especialmente cierto para música u otras señales de audio genéricas que no encajan demasiado bien en el modelo de voz CELP. En este caso, el desajuste del modelo puede causar una calidad de audio severamente degradada que puede ser inaceptable para un usuario final del equipo que utiliza dichos métodos. Por lo tanto, sigue existiendo la necesidad de mejorar el comportamiento de los codificadores de voz de tipo CELP a bajas velocidades binarias, especialmente para música y otras entradas no de tipo voz.

Un documento de la técnica anterior en el campo de la codificación de voz/audio es el de Ramprashad S A: "A two stage hybrid embedded speech/ audio coding structure" ACOUSTICS, SPEECH AND SIGNAL PROCESSING, 1998. PROCEEDINGS OF THE 1998 IEEE INTERNATIONAL CONFERENCE ON SEATTLE, WA, USA, 12 a 15 de mayo de 1998, Nueva York, NY, USA, IEEE, US, volumen 1, 12 de mayo de 1998 (12-05-1998), páginas 337 a 340, XP010279163 ISBN: 978-0-7803-4428-0.

Los objetivos anteriores se resuelven mediante las reivindicaciones de la presente invención.

Breve descripción de los dibujos

Las figuras adjuntas, en las que los números de referencia iguales se refieren a elementos idéntica o funcionalmente similares en la totalidad de las diferentes vistas, junto con la siguiente descripción detallada se incorporan a la especificación y forman parte de la misma, y sirven para mostrar diversas realizaciones de conceptos que incluyen la invención reivindicada, y para explicar diversos principios y ventajas de estas realizaciones.

La figura 1 es un diagrama de bloques de un sistema de compresión de voz/audio integrado de la técnica anterior.

La figura 2 es un ejemplo más detallado del codificador de la capa de mejora de la figura 1.

La figura 3 es un ejemplo más detallado del codificador de la capa de mejora de la figura 1.

La figura 4 es un diagrama de bloques de un codificador y un descodificador de la capa de mejora.

La figura 5 es un diagrama de bloques de un sistema de codificación integrado multicapa.

La figura 6 es un diagrama de bloques del codificador y el descodificador de la capa 4.

La figura 7 es un diagrama de flujo que muestra el funcionamiento de los codificadores de la figura 4 y la figura 6.

La figura 8 es un diagrama de bloques de un sistema de compresión de voz/audio integrado de la técnica anterior.

La figura 9 es un ejemplo más detallado del codificador de la capa de mejora de la figura 8.

La figura 10 es un diagrama de bloques de un codificador y un descodificador de la capa de mejora, de acuerdo con diversas realizaciones.

La figura 11 es un diagrama de bloques de un codificador y un descodificador de la capa de mejora, de acuerdo con diversas realizaciones.

La figura 12 es un diagrama de flujo de la codificación de la señal de audio de múltiples canales, de acuerdo con diversas realizaciones.

La figura 13 es un diagrama de flujo de la codificación de la señal de audio de múltiples canales, de acuerdo con diversas realizaciones.

La figura 14 es un diagrama de flujo de la descodificación de una señal de audio de múltiples canales, de acuerdo con diversas realizaciones.

La figura 15 es un gráfico de frecuencias de detección de pico basada en generación de máscaras, de acuerdo con diversas realizaciones.

La figura 16 es un gráfico de frecuencias del escalamiento de la capa central utilizando generación de máscaras de pico, de acuerdo con diversas realizaciones.

Las figuras 17 a 19 son diagramas de flujo que muestran metodología para codificación y descodificación utilizando generación de máscaras basada en detección de picos, de acuerdo con diversas realizaciones.

Los técnicos cualificados apreciarán que los elementos de las figuras se muestran por simplicidad y claridad y no necesariamente han sido dibujados a escala. Por ejemplo, las dimensiones de algunos de los elementos de las figuras pueden estar exageradas con respecto a otros elementos para ayudar a mejorar la comprensión de diversas realizaciones. Además, la descripción y los dibujos no necesariamente requieren el orden mostrado. Se apreciará además que ciertas acciones y/o etapas pueden estar descritas o representadas en un orden de ocurrencia específico, si bien los expertos en la materia comprenderán que dicha especificidad con respecto a la secuencia no es realmente necesaria. Los componentes de aparatos y métodos han sido representados en su caso mediante símbolos convencionales en los dibujos, mostrando solamente aquellos detalles específicos que son pertinentes para la comprensión de las diversas realizaciones, de manera que no se oscurezca la descripción con detalles que resultarán evidentes para los expertos en la materia en beneficio de la descripción del presente documento. Por lo tanto, se apreciará que para mayor simplicidad y claridad de la ilustración, los elementos comunes y bien conocidos que son útiles o necesarios en una realización comercialmente factible pueden no estar representados a efectos de facilitar una visión más clara de estas diversas realizaciones.

Descripción detallada

Para solucionar la necesidad mencionada anteriormente, se describen en la presente memoria un método y un aparato para generar una capa de mejora del sistema de codificación de audio. En funcionamiento, una señal de entrada a codificar es recibida y codificada para producir una señal de audio codificada. A continuación, la señal de audio codificada es escalada con una serie de valores de ganancia para producir una serie de señales de audio codificadas escaladas, que tienen cada una un valor de ganancia asociado y se determinan una serie de valores de error existentes entre la señal de entrada y cada una de dicha serie de señales de audio codificadas escaladas. A continuación, se escoge un valor de ganancia que está asociado con una señal de audio codificada escalada dando como resultado un valor de error bajo, existente entre la señal de entrada y la señal de audio codificada escalada. Finalmente, el valor de error bajo es transmitido junto con el valor de ganancia como parte de una capa de mejora, a la señal de audio codificada.

En la figura 1 se muestra un sistema de compresión de voz/audio integrado de la técnica anterior. El audio de entrada s(n) es procesado en primer lugar mediante un codificador 120 de la capa central, que para esta finalidad puede ser un algoritmo de codificación de voz de tipo CELP. El flujo de bits codificado es transmitido al canal 125, siendo asimismo introducido a un descodificador 115 de la capa central, en el que se genera la señal de audio central $s_c(n)$ reconstruida. A continuación, se utiliza el codificador 120 de la capa de mejora para codificar información adicional en base a cierta comparación de las señales s(n) y $s_c(n)$, y pueden utilizarse opcionalmente parámetros procedentes del descodificador 115 de la capa central. Tal como en el descodificador 115 de la capa central, el descodificador 130 de la capa central transforma parámetros del flujo de bits de la capa central en una

señal de audio de la capa central $\hat{s}_c(n)$. El descodificador 135 de la capa de mejora utiliza a continuación el flujo de bits de la capa de mejora procedente del canal 125 y la señal $\hat{s}_c(n)$ para producir la señal de salida de audio mejorada $\hat{s}(n)$.

La ventaja principal de dicho sistema de codificación integrado es que un canal particular 125 puede no ser capaz de soportar sistemáticamente el requisito de ancho de banda asociado con algoritmos de codificación de audio de alta calidad. Sin embargo, un codificador integrado permite que se reciba un flujo de bits parcial (por ejemplo, solamente el flujo de bits de la capa central) desde el canal 125 para producir, por ejemplo, solamente el audio de salida central cuando el flujo de bits de la capa de mejora se ha perdido o está corrupto. Sin embargo, existen compromisos en la calidad entre codificadores integrados frente a no integrados, y asimismo entre diferentes objetivos de optimización de codificación integrada. Es decir, la codificación de la capa de mejora de alta calidad puede ayudar a conseguir un mejor equilibrio entre las capas central y de mejora, y asimismo a reducir la velocidad de datos global para unas mejores características de transmisión (por ejemplo, congestión reducida), lo que puede tener como resultado menores tasas de error de paquete para las capas de mejora.

En la figura 2 se proporciona un ejemplo más detallado de un codificador 120 de la capa de mejora de la técnica anterior. En este caso, el generador 210 de la señal de error se compone de una señal diferencial ponderada que se transforma al dominio de la transformada de coseno discreta modificada (MDCT, Modified Discrete Cosine Transform) para su procesamiento por el codificador 220 de señal de error. La señal de error E está dada como:

$$\mathbf{E} = \mathrm{MDCT}\{\mathbf{W}(\mathbf{s} - \mathbf{s}_c)\}, \qquad (1)$$

donde W es una matriz de ponderación perceptual basada en coeficientes de filtrado de predicción lineal (LP, Linear Prediction) procedentes del descodificador 115 de la capa central, s es un vector (es decir, una trama) de muestras de la señal de audio de entrada s(n), y s_c es el correspondiente vector de muestras del descodificador 115 de la capa central. En la Recomendación G.729.1 de ITU-T se describe un proceso MDCT a modo de ejemplo. A continuación, la señal de error E es procesada mediante el codificador 220 de la señal de error para producir la palabra de código i_E, que a continuación es transmitida al canal 125. Para este ejemplo, es importante observar que el codificador 120 de la señal de error recibe solamente una señal de error E y emite una palabra de código asociada i_E. El motivo para esto resultará evidente más adelante.

A continuación, el descodificador 135 de la capa de mejora recibe el flujo de bits codificado procedente del canal 125 y desmultiplexa adecuadamente el flujo de bits para producir la palabra de código i_E. El descodificador 230 de la señal de error utiliza la palabra de código i_E para reconstruir la señal de error de la capa de mejora $\hat{E}$, que a continuación se combina como sigue mediante el combinador 240 de señales con la señal de audio de salida de la capa central $\hat{s}_c(n)$, para producir la señal de salida de audio mejorada $\hat{s}(n)$:

$$\hat{\mathbf{s}} = \mathbf{s}_c + \mathbf{W}^{-1}\,\mathrm{MDCT}^{-1}\{\hat{\mathbf{E}}\}, \qquad (2)$$

donde MDCT^{-1} es la MDCT inversa (incluyendo solapamiento y suma), y W^{-1} es la matriz de ponderación perceptual inversa.

En la figura 3 se muestra otro ejemplo de un codificador de la capa de mejora. En este caso, la generación de la señal de error E mediante el generador 315 de la señal de error involucra pre-escalamiento adaptativo, en el que se lleva a cabo alguna modificación sobre la salida de audio de la capa central $s_c(n)$. Este proceso tiene como resultado cierto número de bits a generar, que se muestran en el codificador 120 de la capa de mejora como palabra de código i_s.

Adicionalmente, el codificador 120 de la capa de mejora muestra la señal de audio de entrada s(n) y el audio de salida de la capa central transformado S_c siendo introducidos en el codificador 320 de la señal de error. Estas señales se utilizan para construir un modelo psicoacústico para la codificación mejorada de la señal de error de la capa de mejora E. A continuación, las palabras de código i_s e i_E son multiplexadas mediante el MUX 325, y después enviadas al canal 125 para su posterior descodificación mediante el descodificador 135 de la capa de mejora. El flujo de bits codificado es recibido por el demux 335, que separa el flujo de bits en las componentes i_s e i_E. A continuación, la palabra de código i_E es utilizada por el descodificador 340 de la señal de error para reconstruir la señal de error de la capa de mejora $\hat{E}$. El combinador 345 de señales escala de alguna manera la señal $\hat{s}_c(n)$ utilizando bits de escalamiento i_s, y a continuación combina el resultado con la señal de error de la capa de mejora $\hat{E}$ para producir la señal de salida de audio mejorada $\hat{s}(n)$.

En la figura 4 se proporciona una primera realización de la presente invención. Esta figura muestra el codificador 410 de la capa de mejora que recibe la señal de salida de la capa central $s_c(n)$ mediante la unidad de escalamiento 415. Se utiliza un conjunto predeterminado de ganancias {g} para producir una serie de señales de salida de la capa central escaladas {S}, donde g_j y S_j son los candidatos t-ésimos de los respectivos conjuntos. Dentro de la unidad de escalamiento 415, la primera realización procesa la señal $s_c(n)$ en el dominio (MDCT), como:

$$\mathbf{S}_j = \mathbf{G}_j \times \mathrm{MDCT}\{\mathbf{W}\mathbf{s}_c\}; \quad 0 \le j < M\ , \tag{3}$$

donde W puede ser alguna matriz de ponderación perceptual, s_c es un vector de muestras procedentes del descodificador 115 de la capa central, la MDCT es una operación bien conocida en la técnica, y G_j puede ser una matriz de ganancia formada mediante utilizar un vector de ganancia candidato g_j, y donde M es el número de vectores de ganancia candidatos. En la primera realización, G_j utiliza el vector g_j como la diagonal y ceros en todas las demás posiciones (es decir, una matriz diagonal), aunque existen muchas posibilidades. Por ejemplo, G_j puede ser una matriz banda, o puede ser incluso una simple cantidad escalar multiplicada por la matriz identidad I. Alternativamente, puede haber cierta ventaja dejando la señal S_j en el dominio temporal o pueden existir casos en los que sea ventajoso transformar el audio a un dominio diferente, tal como el dominio de la transformada de Fourier discreta (DFT, Discrete Fourier Transform). En la técnica se conocen muchas transformaciones de este tipo. En estos casos, la unidad de escalamiento puede emitir la S_j adecuada en base al dominio vectorial respectivo.

Pero en cualquier caso, la razón principal para escalar el audio de salida de la capa central es compensar el desajuste del modelo (o alguna otra deficiencia de codificación) que puede provocar diferencias significativas entre la señal de entrada y el códec de la capa central. Por ejemplo, si la señal de audio de entrada es principalmente una señal musical y el códec de la capa central se basa en un modelo de voz, entonces la salida de la capa central puede contener características de señal severamente distorsionadas, en cuyo caso, desde el punto de vista de la calidad del sonido es beneficioso reducir selectivamente la energía de esta componente de la señal antes de aplicar una codificación complementaria de la señal mediante una o varias capas de mejora.

El vector candidato S_j del audio de la capa central escalado en ganancia y el audio de entrada s(n) pueden utilizarse a continuación como entrada para el generador 420 de la señal de error. En una realización a modo de ejemplo, la señal de audio de entrada s(n) es transformada en el vector S, de manera que S y S_j están correspondientemente alineados. Es decir, el vector s que representa la s(n) está alineado en tiempo (fase) con s_c, y pueden aplicarse las operaciones correspondientes de manera que, en esta realización:

$$\mathbf{E}_j = \mathrm{MDCT}\{\mathbf{W}\mathbf{s}\} - \mathbf{S}_j; \quad 0 \le j < M\ . \tag{4}$$

Esta expresión produce una serie de vectores de señal de error E_j que representan la diferencia ponderada entre el audio de entrada y el audio de salida de la capa central escalado en ganancia en el dominio espectral MDCT. En otras realizaciones en las que se consideran dominios diferentes, la expresión anterior puede modificarse en base al dominio de procesamiento respectivo.

A continuación, se utiliza el selector de ganancia 425 para evaluar dicha serie de vectores de señal de error E_j, de acuerdo con la primera realización de la presente invención, a efectos de producir un vector de error óptimo E*, un parámetro de ganancia óptimo g*, y posteriormente, un correspondiente índice de ganancia i_g. El selector de ganancia 425 puede utilizar diversos métodos para determinar los parámetros óptimos, E* y g*, que pueden implicar métodos de bucle cerrado (por ejemplo, minimización de una métrica de distorsión), métodos de bucle abierto (por ejemplo, clasificación heurística, estimación del comportamiento de modelos, etc.), o una combinación de ambos métodos. En la realización a modo de ejemplo, puede utilizarse una métrica de distorsión sesgada, que está dada por la diferencia de energía sesgada entre el vector de la señal de audio original S y el vector de señal reconstruido compuesto:

$$j^* = \underset{0 \le j < M}{\arg\min}\left\{ \beta_j \cdot \left\| \mathbf{S} - \left(\mathbf{S}_j + \hat{\mathbf{E}}_j\right) \right\|^2 \right\}, \tag{5}$$

donde $\hat{E}_j$ puede ser la estimación cuantificada del vector de la señal de error E_j, y βj puede ser un término de sesgo que se utiliza para complementar la decisión de elegir el índice de error de ganancia óptimo perceptualmente j*. Se proporciona un método a modo de ejemplo para la cuantificación vectorial de un vector de señal en la solicitud de patente de EE.UU. de número de serie 11/531122, titulada "APPARATUS AND METHOD FOR LOW COMPLEXITY COMBINATORIAL CODING OF SIGNALS", si bien son posibles muchos otros métodos. Reconociendo que $E_j = S - S_j$, la ecuación (5) puede reescribirse como:

$$j^* = \underset{0 \le j < M}{\arg\min}\left\{ \beta_j \cdot \left\| \mathbf{E}_j - \hat{\mathbf{E}}_j \right\|^2 \right\}. \tag{6}$$

En esta expresión, el término $\varepsilon_j = \| E_j - \hat{E}_j \|^2$ representa la energía de la diferencia entre las señales de error no cuantificada y cuantificada. Para mayor claridad, esta cantidad puede denominarse la "energía residual", y puede utilizarse posteriormente para evaluar un "criterio de selección de ganancia", en el que se selecciona el parámetro de

ganancia óptimo g*. Se proporciona uno de dichos criterios de selección de ganancia en la ecuación (6), aunque son posibles muchos otros.

La necesidad de un término de sesgo β_j puede surgir del caso en que la función de ponderación de error W en las ecuaciones (3) y (4) puede no producir adecuadamente distorsiones igualmente perceptibles a través del vector $\hat{E}_j$. Por ejemplo, aunque puede utilizarse la función de ponderación de error W para intentar "sanear" en cierta medida el espectro de errores, puede haber ciertas ventajas ponderando más las frecuencias bajas, debido a la percepción de la distorsión por el oído humano. Como resultado de una ponderación del error aumentada en las frecuencias bajas, las señales de alta frecuencia pueden estar infra-modeladas mediante la capa de mejora. En estos casos, puede existir un beneficio directo en sesgar la métrica de distorsión hacia valores de g_j que no atenúan los componentes de alta frecuencia de S_j, de manera que la infra-modelación de las altas frecuencias no tenga como resultado artefactos sonoros desagradables o poco naturales en la señal de audio final reconstruida. Un ejemplo de este tipo sería el caso de una señal de voz sorda. En este caso, el audio de entrada se compone generalmente de señales de tipo ruido, de frecuencias media a alta, producidas por el flujo turbulento de aire procedente de la boca humana. Puede ocurrir que el codificador de la capa central no codifique directamente este tipo de forma de onda, pero puede utilizar un modelo de ruido para generar una señal de audio de sonido similar. Esto puede tener como resultado una correlación baja, en términos generales, entre las señales de audio de entrada y de audio salida de la capa central. Sin embargo, en esta realización, el vector de señal de error E_j se basa en la diferencia entre el audio de entrada y las señales de salida de audio de la capa central. Dado que estas señales pueden no estar muy bien correlacionadas, la energía de la señal de error E_j puede no ser necesariamente menor que el audio de entrada y/o el audio de salida de la capa central. En este caso, la minimización del error en la ecuación (6) puede tener como resultado que el escalamiento de la ganancia sea demasiado agresivo, lo que puede tener como resultado potenciales artefactos audibles.

En otro caso, los factores de sesgo β_j pueden basarse en otras características de señal de las señales de audio de entrada y/o de audio de salida de la capa central. Por ejemplo, la relación pico/promedio del espectro de una señal puede proporcionar una indicación del contenido armónico de dicha señal. Las señales tales como la voz y ciertos tipos de música pueden tener un alto contenido armónico y por lo tanto una elevada relación pico/promedio. Sin embargo, una señal de música procesada a través de un códec de voz puede tener como resultado una mala calidad debido al desajuste del modelo de codificación, y como resultado, el espectro de la señal de salida de la capa central puede tener una relación pico/promedio reducida en comparación con el espectro de la señal de entrada. En este caso, puede ser beneficioso reducir la cantidad de sesgo en el proceso de minimización a efectos de permitir que el audio de salida de la capa central sea escalado en ganancia a una energía menor, permitiendo de ese modo que la codificación de la capa de mejora tenga un efecto más pronunciado sobre el audio de salida compuesto. A la inversa, ciertos tipos de señales de entrada de voz o de música pueden presentar relaciones menores pico/promedio, en cuyo caso las señales pueden percibirse siendo más ruidosas, y por lo tanto pueden beneficiarse de un escalamiento menor del audio de salida de la capa central mediante aumentar el sesgo del error. Un ejemplo de una función para generar los factores de sesgo para β_j, está dado por:

$$\beta_j = \begin{cases} 1 + 10^6 \cdot j; \; UVSpeech == \textit{VERDADERO} \; o \; \phi_s < \lambda \phi_{s_t} \\ 10^{(-j \cdot \Delta / 10)}; \; \textit{de lo contrario} \end{cases}, \quad 0 \le j < M \; . \; (7)$$

donde λ Puede ser algún umbral, y la relación pico/promedio para el vector ϕ_y de puede estar dada como:

$$\phi_y = \frac{\max\left\{\left|\mathbf{y}_{k_1 k_2}\right|\right\}}{\frac{1}{k_2 - k_1 + 1}\sum_{k=k_1}^{k_2} |y(k)|}, \qquad (8)$$

y donde y_{k1k2} es un subconjunto vectorial de y(k), de manera que $y_{k1k2} = y(k)$; $k_1 \le k \le k_2$.

Una vez que se ha determinado el índice de ganancia óptimo j* a partir de la ecuación (6), se genera la palabra de código asociada i_g y se envía el vector de error óptimo E* al codificador 430 de la señal de error, donde E* se codifica de forma adecuada para su multiplexación con otras palabras de código (mediante el MUX 440), y se transmite para su utilización por un correspondiente descodificador. En la realización a modo de ejemplo, el codificador 408 de la señal de error utiliza codificación de impulso factorial (FPC, Factorial Pulse Coding). Este método es ventajoso desde el punto de vista de la complejidad del procesamiento, dado que el proceso de enumeración asociado con la codificación del vector E* es independiente del proceso de generación vectorial que se utiliza para generar $\hat{E}_j$.

El descodificador 450 de la capa de mejora invierte estos procesos para producir la salida de audio mejorada ŝ(n). Más específicamente, i_g y i_E son recibidos por el descodificador 450, siendo i_E enviado por el demux 455 al

descodificador 460 de la señal de error, donde el vector de error óptimo E* se obtiene a partir de la palabra de código. El vector de error óptimo E* se pasa al combinador de señal 465, donde el $\hat{s}_c(n)$ recibido se modifica tal como en la ecuación (2) para producir $\hat{s}(n)$.

Una segunda realización de la presente invención involucra un sistema de codificación integrado multicapa, tal como el mostrado en la figura 5. En este caso, se puede ver que existen cinco capas integradas proporcionadas para este ejemplo. Las capas 1 y 2 pueden estar ambas basadas en códec de voz, y las capas 3, 4 y 5 pueden ser capas de mejora MDCT. Por lo tanto, los codificadores 502 y 503 pueden utilizar códecs de voz para producir y emitir la señal de entrada codificada s(n). Los codificadores 510, 610 y 514 comprenden codificadores de la capa de mejora, que emiten cada uno una mejora diferente para la señal codificada. De manera similar a la realización anterior, el vector de la señal de error para la capa 3 (codificador 510) puede estar dado como:

$$\mathbf{E}_3 = \mathbf{S} - \mathbf{S}_2, \qquad (9)$$

donde s = MDCT{Ws} es la señal de entrada transformada ponderada, y S_2 = MDCT{Ws_2} es la señal transformada ponderada generada a partir del descodificador 506 de la capa 1/2. En esta realización, la capa 3 puede ser una capa de cuantificación de baja velocidad y, como tal, puede haber relativamente pocos bits para codificar la correspondiente señal de error cuantificada $\hat{E}_3 = Q\{E_3\}$. Para proporcionar una buena calidad bajo estas limitaciones, solamente puede cuantificarse una fracción de los coeficientes dentro de E_3. Las posiciones de los coeficientes para codificar pueden estar fijas o ser variables, pero si se permite que varíen, puede ser necesario enviar información adicional al descodificador para identificar estas posiciones. Por ejemplo, si el intervalo de posiciones codificadas comienza en k_s y finaliza en k_e, donde $0 \le k_s < k_e < N$, entonces el vector $\hat{E}_3$ de la señal de error cuantificada puede contener valores distintos de cero solamente dentro de dicho intervalo, y ceros para las posiciones exteriores a dicho intervalo. La información de la posición y del intervalo puede asimismo ser implícita, dependiendo del método de codificación utilizado. Por ejemplo, en la codificación de audio se sabe bien que una banda de frecuencias puede considerarse perceptualmente importante, y que la codificación de un vector de señal puede focalizarse en dichas frecuencias. En estas circunstancias, el intervalo codificado puede ser variable, y puede no abarcar un conjunto contiguo de frecuencias. Pero en todo caso, una vez que la señal está cuantificada, el espectro de salida codificado compuesto puede construirse como:

$$\mathbf{S}_3 = \hat{\mathbf{E}}_3 + \mathbf{S}_2, \qquad (10)$$

que a continuación se utiliza como entrada para el codificador 610 de la capa 4.

El codificador 610 de la capa 4 es similar al codificador 410 de la capa de mejora de la realización anterior. Utilizando el candidato a vector de ganancia g_j, el correspondiente vector de error puede escribirse como:

$$\mathbf{E}_4(j) = \mathbf{S} - \mathbf{G}_j \mathbf{S}_3, \qquad (11)$$

donde G_j puede ser una matriz de ganancia con el vector g_j como componente diagonal. Sin embargo, en la realización actual, el vector de ganancia g_j puede estar relacionado del siguiente modo con el vector $\hat{E}_3$ de la señal de error cuantificada. Dado que el vector $\hat{E}_3$ de la señal de error cuantificada puede estar limitado en un intervalo de frecuencias, por ejemplo, comenzando en la posición del vector k_s y finalizando en la posición del vector k_e, se supone que la señal de salida S_3 de la capa 3 se codifica de manera muy precisa dentro de dicho intervalo. Por lo tanto, de acuerdo con la presente invención, el vector de ganancia g_j se regula en base a las posiciones codificadas del vector de señal de error de la capa 3, k_s y k_e. Más específicamente, para conservar la integridad de la señal en estas posiciones, los correspondientes elementos de ganancia individuales pueden fijarse a un valor constante α. Es decir:

$$g_j(k) = \begin{cases} \alpha; & k_s \le k \le k_e \\ \gamma_j(k); & \textit{de lo contrario} \end{cases}, \qquad (12)$$

donde generalmente $0 \le \gamma_j(k) \le 1$, y $g_j(k)$ es la ganancia de la posición de k-ésima del vector candidato j-ésimo. En una realización a modo de ejemplo, el valor de la constante es uno (α = 1), si bien son posibles muchos valores. Además, el intervalo de frecuencias puede abarcar múltiples posiciones de comienzo y terminación. Es decir, la ecuación (12) puede segmentarse en intervalos no contiguos de ganancias variables que se basan en alguna función de la señal de error Ê3, y puede escribirse de manera más general como:

$$g_j(k) = \begin{cases} \alpha; & \hat{E}_3(k) \neq 0 \\ \gamma_j(k); & \textit{de lo contrario} \end{cases}, \qquad (13)$$

Para este ejemplo, se utiliza una ganancia fija α para generar $g_j(k)$ cuando las correspondientes posiciones en la señal de error $\hat{E}_3$ cuantificada previamente son distintas de cero, y se utiliza la función de ganancia $\gamma_j(k)$ cuando las posiciones correspondientes en $\hat{E}_3$ son cero. Una posible función de ganancia puede definirse como:

$$\gamma_j(k) = \begin{cases} \alpha \cdot 10^{(-j \cdot \Delta / 20)}; & k_l \leq k \leq k_h \\ \alpha; & \textit{de lo contrario} \end{cases}, \quad 0 \leq j < M,$$

(14)

donde Δ es un tamaño de paso (por ejemplo, Δ ≈ 2,2 dB), α es una constante, M es el número de candidatos (por ejemplo, M = 4, que puede representarse utilizando solamente 2 bits), y k_l y k_h son límites de baja y alta frecuencia, respectivamente, sobre los cuales puede tener lugar la reducción de ganancia. La introducción de los parámetros k_l y k_h es útil en los sistemas en los que se desea el escalamiento solamente en cierto intervalo de frecuencias. Por ejemplo, en cierta realización, las altas frecuencias pueden no modelarse adecuadamente mediante la capa central, de manera que la energía dentro de la banda de alta frecuencia puede ser inherentemente menor que en la señal de audio de entrada. En tal caso, puede existir poco o ningún beneficio del escalamiento de la salida de la capa 3 en dicha zona de la señal, dado que como resultado puede aumentar la energía de error global.

Resumiendo, la pluralidad de vectores de ganancia candidatos g_j se basa en alguna función de los elementos codificados del vector de señal codificado previamente, en este caso $\hat{E}_3$. Esto puede expresarse en términos generales como:

$$g_j(k) = f\left(k, \hat{\mathbf{E}}_3\right).$$

(15)

Las correspondientes operaciones del descodificador se muestran en el lado derecho de la figura 5. A medida que se reciben las diversas capas de los flujos de bits codificados (i_1 a i_5), las señales de salida de mayor calidad se basan en la jerarquía de las capas de mejora sobre el descodificador de la capa central (capa 1). Es decir, para esta realización específica, dado que las primeras dos capas se componen de codificación del modelo de voz en el dominio de tiempo (por ejemplo, CELP) y las tres capas restantes se componen de codificación en el dominio de la transformada (por ejemplo, MDCT), la salida final para el sistema ŝ(n) se genera según lo siguiente:

$$\hat{s}(n) = \begin{cases} \hat{s}_1(n); \\ \hat{s}_2(n) = \hat{s}_1(n) + \hat{e}_2(n); \\ \hat{s}_3(n) = \mathbf{W}^{-1}\,\mathrm{MDCT}^{-1}\left\{\hat{\mathbf{S}}_2 + \hat{\mathbf{E}}_3\right\}; \\ \hat{s}_4(n) = \mathbf{W}^{-1}\,\mathrm{MDCT}^{-1}\left\{\mathbf{G}_j \cdot \left(\hat{\mathbf{S}}_2 + \hat{\mathbf{E}}_3\right) + \hat{\mathbf{E}}_4\right\}; \\ \hat{s}_5(n) = \mathbf{W}^{-1}\,\mathrm{MDCT}^{-1}\left\{\mathbf{G}_j \cdot \left(\hat{\mathbf{S}}_2 + \hat{\mathbf{E}}_3\right) + \hat{\mathbf{E}}_4 + \hat{\mathbf{E}}_5\right\}; \end{cases}, \qquad (16)$$

donde $\hat{e}_2(n)$ es la señal de la capa de mejora en el dominio temporal de la capa 2, y $\hat{\mathbf{s}}_2 = \mathrm{MDCT}\{\mathrm{Ws}_2\}$ es el vector MDCT ponderado correspondiente a la salida de audio $\hat{s}_2(n)$ de la capa 2. En esta expresión, la señal de salida global ŝ(n) puede determinarse a partir del máximo nivel de capas consecutivas de flujo de bits que se reciben. En esta realización, se supone que las capas de nivel inferior tienen una mayor probabilidad de ser recibidas adecuadamente desde el canal, y por lo tanto, los conjuntos de palabras de código $\{i_1\}$, $\{i_1\ i_2\}$, $\{i_1\ i_2\ i_3\}$, etc., determinan el nivel adecuado de descodificación de la capa de mejora en la ecuación (16).

La figura 6 es un diagrama de bloques que muestra el codificador 610 y el descodificador 650 de la capa 4. El codificador y el descodificador mostrados en la figura 6 son similares a los mostrados en la figura 4, excepto porque el valor de ganancia utilizado por las unidades de escalamiento 615 y 670 se obtiene a través de generadores 630 y 660 de ganancia selectiva en frecuencias, respectivamente. En funcionamiento, la salida de audio S_3 de la capa 3 es emitida desde el codificador de la capa 3 y recibida por la unidad de escalamiento 615. Adicionalmente, el vector de error $\hat{E}_3$ de la capa 3 es emitido desde el codificador 510 de la capa 3 y recibido por el generador 630 de ganancia

selectiva en frecuencias. Tal como se ha descrito, dado que el vector $\hat{E}_3$ de la señal de error cuantificada puede estar limitado en el intervalo de frecuencias, el vector de ganancia g_j se regula, por ejemplo, en base a las posiciones k_s y k_e tal como se muestra en la ecuación 12, o en la expresión más general de la ecuación 13.

El audio escalado S_j es emitido desde la unidad de escalamiento 615 y recibido por el generador 620 de la señal de error. Tal como se ha descrito anteriormente, el generador 620 de la señal de error recibe la señal de audio de entrada S y determina un valor de error E_j para cada vector de escalamiento utilizado por la unidad de escalamiento 615. Estos vectores de error se pasan a los circuitos 635 del selector de ganancia junto con los valores de ganancia utilizados en la determinación de los vectores de error y un error específico E* basado en el valor de ganancia óptimo g*. Una palabra de código (i_g) que representa la ganancia óptima g* se emite desde el selector de ganancia 635, junto con el vector de error óptimo E*, que se pasa al codificador 640 de la señal de error, en el que es determinada y emitida la palabra de código i_E. i_g e i_E son entregadas al multiplexor 645 y transmitidas a través del canal 125 al descodificador 650 de la capa 4.

Durante el funcionamiento del descodificador 650 de la capa 4, i_g y i_E son recibidos desde el canal 125 y desmultiplexados mediante el demux 655. La palabra de código de ganancia i_g y el vector de error $\hat{E}_3$ de la capa 3 se utilizan como entrada para el generador 660 de ganancia selectiva en frecuencias, a efectos de producir el vector de ganancia g*, de acuerdo con el correspondiente método del codificador 610. A continuación, se aplica el vector de ganancia g* al vector de audio reconstruido $\hat{S}3$ de la capa 3 dentro de la unidad de escalamiento 670, cuya salida se combina a continuación en el combinador de señal 675 con el vector de error de la capa de mejora E* de la capa 4, que se obtuvo en el descodificador 655 de la señal de error mediante la descodificación de la palabra de código i_E, para producir la salida de audio reconstruida $\hat{S}_4$ de la capa 4, tal como se muestra.

La figura 7 es un diagrama de flujo 700 que muestra el funcionamiento de un codificador, de acuerdo con la primera y la segunda realizaciones de la presente invención. Tal como se ha descrito anteriormente, ambas realizaciones utilizan una capa de mejora que escala el audio codificado con una serie de valores de escalamiento y a continuación escoge el valor de escalamiento que tiene como resultado un error menor. Sin embargo, en la segunda realización de la presente invención, se utiliza el generador 630 de ganancia selectiva en frecuencias para generar los valores de ganancia.

El flujo lógico comienza en el bloque 710, en el que un codificador de la capa central recibe una señal de entrada a codificar, y codifica la señal de entrada para producir una señal de audio codificada. El codificador 410 de la capa de mejora recibe la señal de audio codificada ($s_c(n)$) y la unidad de escalamiento 415 escala la señal de audio codificada con una serie de valores de ganancia, para producir una serie de señales de audio codificadas escaladas, que tienen cada una un valor de ganancia asociado. (Bloque 720). En el bloque 730, el generador 420 de la señal de error determina una serie de valores de error existentes entre la señal de entrada y cada una de dicha serie de señales de audio codificadas escaladas. A continuación, el selector de ganancia 425 escoge un valor de ganancia entre dicha serie de valores de ganancia (bloque 740). Tal como se ha descrito anteriormente, el valor de ganancia (g*) está asociado con una señal de audio codificada escalada dando como resultado un valor de error bajo (E*) existente entre la señal de entrada y la señal de audio codificada escalada. Finalmente, en el bloque 750, el transmisor 440 transmite a la señal de audio codificada el valor de error bajo (E*) junto con el valor de ganancia (g*), como parte de una capa de mejora. Tal como reconocerá un experto en la materia, E* y g* son codificados adecuadamente antes de la transmisión.

Tal como se ha descrito anteriormente, en el lado del receptor, la señal de audio codificada se recibirá junto con la capa de mejora. La capa de mejora es una mejora a la señal de audio codificada, que comprende el valor de ganancia (g*) y la señal de error (E*) asociada al valor de ganancia.

Escalamiento de la capa central para estéreo

En la descripción anterior, se ha descrito un sistema de codificación integrado en el que cada una de las capas codificaba una señal mono. A continuación, se describe un sistema de codificación integrado para la codificación de estéreo u otras señales de múltiples canales. Para mayor brevedad, se describe la tecnología en el contexto de una señal estéreo que consiste en dos entradas (fuentes) de audio; sin embargo, las realizaciones a modo de ejemplo descritas en la presente memoria pueden extenderse fácilmente a los casos en que la señal estéreo tiene más de dos entradas de audio, tal como en el caso de entradas de audio de múltiples canales. Con propósitos ilustrativos y no limitativos, las dos entradas de audio son señales estéreo que consisten en la señal izquierda (s_L) y la señal derecha (s_R), donde s_L y s_R son vectores columna n-dimensionales que representan una trama de datos de audio. De nuevo para mayor brevedad, se describe en detalle un sistema de codificación integrado que consiste en dos capas, a saber una capa central y una capa de mejora. La idea propuesta puede extenderse fácilmente a sistemas de codificación integrados de múltiples capas. Asimismo, el códec puede no estar integrado por sí mismo, es decir, puede tener solamente una capa, estando parte de los bits de dicho códec dedicados a estéreo y el resto de los bits a la señal mono.

Se conoce un códec estéreo integrado que consiste en una capa central que codifica simplemente una señal mono y capas de mejora que codifican las señales ya sea de alta frecuencia o estéreo. En dicho escenario limitado, la capa central codifica una señal mono (s), obtenida de la combinación de s_L y s_R, para producir una señal mono codificada $\hat{s}$. Sea H una matriz de combinación 2×1 utilizada para generar una señal mono, es decir,

$$\mathbf{s} = (\mathbf{s}_L \quad \mathbf{s}_R)\mathbf{H}$$

(17)

Debe observarse que en la ecuación (17), s_R puede ser una versión retardada de la señal de audio derecha en lugar de ser exactamente la señal del canal derecho. Por ejemplo, el retardo puede calcularse para maximizar la correlación de s_L y la versión retardada de s_R. Si la matriz H es $[0{,}5\ 0{,}5]^T$, entonces la ecuación 17 tiene como resultado una ponderación igual de los respectivos canales derecho e izquierdo, es decir, $s = 0{,}5s_L + 0{,}5s_R$. Las realizaciones presentadas en la presente memoria no se limitan a una capa central que codifica la señal mono y una capa de mejora que codifica la señal estéreo. Tanto la capa central del códec integrado como la capa de mejora pueden codificar señales de audio multicanal. El número de canales en la señal de audio multicanal que son codificados mediante el multicanal de la capa central puede ser menor que el número de canales en la señal de audio multicanal que pueden ser codificados mediante la capa de mejora. Sean (m, n) los números de canales a codificar mediante la capa central y la capa de mejora, respectivamente. Sea s_1, s_2, s_3, ..., s_n una representación de n canales de audio a codificar mediante el sistema integrado. Los m canales a codificar mediante la capa central se derivan de estos, y se obtienen como

$$[\mathbf{s}^1 \quad \mathbf{s}^2 \quad \ldots \quad \mathbf{s}^m] = [\mathbf{s}_1 \quad \mathbf{s}_2 \quad \ldots \quad \mathbf{s}_n]\,\mathbf{H},$$

(17a)

donde H es una matriz n×m.

Tal como se ha mencionado anteriormente, la capa central codifica una señal mono para producir una señal codificada ŝ de la capa central. Para generar estimaciones de los componentes estéreo a partir de ŝ, se calcula un factor de equilibrio. Este factor de equilibrio se calcula como:

$$w_L = \frac{\mathbf{s}_L^T\mathbf{s}}{\mathbf{s}^T\mathbf{s}}, \quad w_R = \frac{\mathbf{s}_R^T\mathbf{s}}{\mathbf{s}^T\mathbf{s}}$$

(18)

Puede demostrarse que si la matriz de combinación H es $[0{,}5\ 0{,}5]^T$, entonces

$$w_L = 2 - w_R$$

(19)

Debe observarse que la relación permite la cuantificación de solamente un parámetro, y del primero puede extraerse fácilmente otro. Las salidas estéreo se calculan a continuación como

$$\hat{\mathbf{s}}_L = w_L\,\hat{\mathbf{s}}, \quad \hat{\mathbf{s}}_R = w_R\,\hat{\mathbf{s}}$$

(20)

En la sección siguiente, trabajaremos en el dominio de frecuencias en lugar de hacerlo en el dominio de tiempo. De este modo, una señal correspondiente en el dominio de frecuencias se representa en letras mayúsculas, es decir, S, Ŝ, S_L, S_R, $\hat{S}_L$, y $\hat{S}_R$ son la representación en el dominio de frecuencias de s, ŝ, s_L, s_R, $\hat{s}_L$, y $\hat{s}_R$, respectivamente. El factor de equilibrio en el dominio de frecuencias se calcula utilizando términos en el dominio de frecuencias y está dado por

$$W_L = \frac{\mathbf{S}_L^T\mathbf{S}}{\mathbf{S}^T\mathbf{S}}, \quad W_R = \frac{\mathbf{S}_R^T\mathbf{S}}{\mathbf{S}^T\mathbf{S}}$$

(21)

y

$$\hat{\mathbf{S}}_L = W_L \hat{\mathbf{S}}, \quad \hat{\mathbf{S}}_R = W_R \hat{\mathbf{S}}$$

(22)

En el dominio de frecuencias, los vectores pueden dividirse adicionalmente en sub-vectores sin solapamiento, es decir, un vector S de dimensión n puede dividirse en t sub-vectores S_1, S, ..., S_t de dimensiones m_1, m_2,...mt, de modo que

$$\sum_{k=1}^{t} m_k = n .$$

(23)

En este caso, puede calcularse un factor de equilibrio diferente para sub-vectores diferentes, es decir,

$$W_{Lk} = \frac{\mathbf{S}_{Lk}^T \mathbf{S}_k}{\mathbf{S}_k^T \mathbf{S}_k}, \quad W_{Rk} = \frac{\mathbf{S}_{Rk}^T \mathbf{S}_k}{\mathbf{S}_k^T \mathbf{S}_k}$$

(24)

En este caso, el factor de equilibrio es independiente de la consideración de la ganancia.

Haciendo referencia a continuación a las figuras 8 y 9, se muestran dibujos de la técnica anterior relevantes para señales estéreo y otras señales de múltiples canales. El sistema 800 de compresión de voz/audio integrado de la técnica anterior de la figura 8 es similar al de la figura 1 pero tiene múltiples señales de entrada de audio, en este ejemplo mostradas como señales de entrada estéreo izquierda y derecha S(n). Estas señales de audio de entrada son alimentadas al combinador 810, que produce audio de entrada s(n), tal como se muestra. Las múltiples señales de entrada se proporcionan asimismo al codificador 820 de la capa de mejora, tal como se muestra. En el lado de el descodificador, el descodificador 830 de la capa de mejora produce señales de audio de salida mejoradas $\hat{s}_L$ $\hat{s}_R$, tal como se muestra.

La figura 9 muestra un codificador de la capa de mejora anterior 900, que puede utilizarse en la figura 8. Las múltiples entradas de audio se proporcionan a un generador del factor de equilibrio, junto con la señal de audio de salida de la capa central, tal como se muestra. El generador 920 del factor de equilibrio del codificador 910 de la capa de mejora recibe las múltiples entradas de audio para producir la señal i_B, que se pasa al MUX 325, tal como se muestra. La señal i_B es una representación del factor de equilibrio. En la realización preferida, i_s es una secuencia de bits que representa los factores de equilibrio. En el lado del descodificador, esta señal i_B es recibida por el descodificador 940 del factor de equilibrio, que produce elementos de factor de equilibrio $W_L(n)$ y $W_R(n)$, tal como se muestra, que son recibidos por el combinador de señal 950, tal como se muestra.

Cálculo del factor de equilibrio con múltiples canales

Tal como se ha mencionado anteriormente, en muchas situaciones el códec utilizado para codificar la señal mono está diseñado para voz de un solo canal y tiene como resultado ruido del modelo de codificación siempre que se utiliza para codificar señales que no están soportadas completamente por el modelo de códec. Las señales musicales y otras señales de tipo no de voz son algunas de las señales que no se modelan adecuadamente mediante un códec de la capa central que está basado en un modelo de voz. La descripción anterior, en relación con las figuras 1 a 7, proponía aplicar una ganancia selectiva en frecuencias a la señal codificada mediante la capa central. El escalamiento se optimizaba para minimizar una distorsión particular (valor de error) entre la entrada de audio y la señal codificada escalada. El enfoque descrito anteriormente funciona bien para señales de un solo canal pero puede no ser óptimo para aplicar el escalamiento de la capa central cuando la capa de mejora está codificando estéreo u otras señales de múltiples canales.

Dado que la componente mono de la señal de múltiples canales, tal como una señal estéreo, se obtiene de la combinación de dichas dos o más entradas de audio estéreo, la señal combinada s puede no conformarse tampoco al modelo de voz de un solo canal; por lo tanto, el códec de la capa central puede producir ruido cuando codifica la señal combinada. De este modo, existe la necesidad de un enfoque que permita el escalamiento de la señal codificada de la capa central en un sistema de codificación integrado, reduciendo de ese modo el ruido generado mediante la capa central. En el enfoque de señal mono descrito anteriormente, una medición de la distorsión específica, sobre la cual se obtenía el escalamiento selectivo en frecuencias, está basada en el error en la señal mono. Este error $E_4(j)$ se muestra en la ecuación (11) anterior. Sin embargo, la distorsión de solamente la señal

mono no es suficiente para mejorar la calidad del sistema de comunicación estéreo. El escalamiento contenido en la ecuación (11) puede ser mediante un factor de escalamiento de la unidad (1) o cualquier otra función identificada.

Para una señal estéreo, una medida de la distorsión debería capturar la distorsión del canal derecho y el izquierdo. Sean E_L y E_R el vector de error para los canales izquierdo y derecho, respectivamente, y están dados por

$$\mathbf{E}_L = \mathbf{S}_L - \hat{\mathbf{S}}_L, \quad \mathbf{E}_R = \mathbf{S}_R - \hat{\mathbf{S}}_R \tag{25}$$

En la técnica anterior, tal como se describe en el estándar AMR-WB+, por ejemplo, estos vectores de error se calculan como

$$\mathbf{E}_L = \mathbf{S}_L - W_L \cdot \hat{\mathbf{S}}, \quad \mathbf{E}_R = \mathbf{S}_R - W_R \cdot \hat{\mathbf{S}}. \tag{26}$$

A continuación, consideramos el caso en que se aplican a S vectores de ganancia selectivos en frecuencia g_j ($0 \leq j < M$). Estos vectores de ganancia selectiva en frecuencias se representan en forma matricial como G_j, donde G_j es una matriz diagonal con elementos diagonales g_j. Para cada vector G_j, los vectores de error se calculan como:

$$\mathbf{E}_L(j) = \mathbf{S}_L - W_L \cdot \mathbf{G}_j \cdot \hat{\mathbf{S}}, \quad \mathbf{E}_R(j) = \mathbf{S}_R - W_R \cdot \mathbf{G}_j \cdot \hat{\mathbf{S}} \tag{27}$$

con las estimaciones de las señales estéreo proporcionadas por los términos $W \cdot G_j \cdot \hat{S}$. Puede verse que la matriz de ganancia G puede ser la matriz unitaria (1) o puede ser cualquier otra matriz diagonal; se reconoce que no todas las estimaciones posibles pueden funcionar para todas las señales escaladas.

La medida ε de la distorsión, que se minimiza para mejorar la calidad del estéreo, es función de los dos vectores de error, es decir,

$$\varepsilon_j = f(\mathbf{E}_L(j), \mathbf{E}_R(j)) \tag{28}$$

Puede verse que el valor de la distorsión puede estar compuesto de múltiples medidas de la distorsión.

El índice j del vector de ganancia selectiva en frecuencias que se selecciona, está dado por:

$$j^* = \arg\min_{0 \leq j < M} \varepsilon_j \tag{29}$$

En la realización a modo de ejemplo, la medida de la distorsión es una distorsión cuadrática media dada por:

$$\varepsilon_j = \|\mathbf{E}_L(j)\|^2 + \|\mathbf{E}_R(j)\|^2 \tag{30}$$

O puede ser una distorsión ponderada o sesgada dada por:

$$\varepsilon_j = B_L \|\mathbf{E}_L(j)\|^2 + B_R \|\mathbf{E}_R(j)\|^2 \tag{31}$$

El sesgo B_L y B_R puede ser función de las energías de los canales izquierdo y derecho.

Tal como se ha mencionado anteriormente, en el dominio de frecuencias, los vectores pueden dividirse adicionalmente en sub-vectores no solapados. Con el fin de extender la técnica propuesta para incluir la división de vectores en el dominio de frecuencias en sub-vectores, se calcula el factor de equilibrio utilizado en (27) para cada

sub-vector. Por lo tanto, los vectores de error E_L y E_R para cada ganancia selectiva en frecuencias están formados mediante una concatenación de los sub-vectores de error, dada por

$$\mathbf{E}_{Lk}(j) = \mathbf{S}_{Lk} - W_{Lk} \cdot \mathbf{G}_{jk} \cdot \hat{\mathbf{S}}_k, \quad \mathbf{E}_{Rk}(j) = \mathbf{S}_{Rk} - W_{Rk} \cdot \mathbf{G}_{jk} \cdot \hat{\mathbf{S}}_k$$

(32)

La medida ε de la distorsión en (28) es en este caso una función de los vectores de error formados mediante la concatenación de los sub-vectores de error anteriores.

Cálculo del factor de equilibrio

El factor de equilibrio generado utilizando la técnica anterior (ecuación 21) es independiente de la salida de la capa central. Sin embargo, para minimizar una medida de la distorsión dada en (30) y en (31), puede ser beneficioso calcular asimismo el factor de equilibrio para minimizar la distorsión correspondiente. En este caso, W_L y W_R del factor de equilibrio pueden calcularse como

$$W_L(j) = \frac{\mathbf{S}_L^T \mathbf{G}_j \hat{\mathbf{S}}}{\left\| \mathbf{G}_j \hat{\mathbf{S}} \right\|^2}, \quad W_R(j) = \frac{\mathbf{S}_R^T \mathbf{G}_j \hat{\mathbf{S}}}{\left\| \mathbf{G}_j \hat{\mathbf{S}} \right\|^2}.$$

(33)

donde puede verse que el factor de equilibrio es independiente de la ganancia, tal como se muestra en el dibujo de la figura 11, por ejemplo. Esta ecuación minimiza las distorsiones en las ecuaciones (30) y (31). El problema de utilizar dicho factor de equilibrio es que, en este caso:

$$W_L(j) \neq 2 - W_R(j),$$

(34)

de manera que pueden requerirse campos de bit independientes para cuantificar W_L y W_R. Esto puede evitarse imponiendo la restricción $W_L(j) = 2 - W_R(j)$ sobre la optimización. Con esta restricción, la solución óptima para la ecuación (30) está dada por:

$$W_L(j) = \frac{2B_R}{B_R + B_L} + \frac{(B_R \mathbf{S}_R - B_L \mathbf{S}_L)^T \mathbf{G}_j \hat{\mathbf{S}}}{\left\| \mathbf{G}_j \hat{\mathbf{S}} \right\|^2}, \quad W_R(j) = 2 - W_L(j).$$

(35)

en la que el factor de equilibrio depende de un término de ganancia, tal como se muestra; la figura 10 de los dibujos muestra un factor de equilibrio dependiente. Si los factores de sesgo B_L y B_R son la unidad, entonces

$$W_L(j) = 1 - \frac{(\mathbf{S}_L - \mathbf{S}_R)^T \mathbf{G}_j \hat{\mathbf{S}}}{\left\| \mathbf{G}_j \hat{\mathbf{S}} \right\|^2}, \quad W_R(j) = 2 - W_L(j)$$

(36)

Los términos $STG_j\hat{S}$ en las ecuaciones (33) y (36) son representativos de valores de correlación entre la señal de audio codificada escalada y por lo menos una de las señales de audio de una señal de audio de múltiples canales.

En la codificación estéreo, la dirección y la posición del origen del sonido pueden ser más importantes que la distorsión cuadrática media. Por lo tanto, la relación entre la energía del canal izquierdo y la energía del canal derecho puede ser un indicador mejor de la dirección (o de la posición del origen del sonido), que minimizar una medida de la distorsión ponderada. En dichos escenarios, el factor de equilibrio calculado en las ecuaciones (35) y (36) puede no ser un buen enfoque para calcular el factor de equilibrio. El requisito es mantener igual la relación de

la energía de los canales izquierdo y derecho antes y después de la codificación. La relación de la energía del canal antes de la codificación y después de la codificación está dada por:

$$\nu = \frac{\|\mathbf{S}_L\|^2}{\|\mathbf{S}_R\|^2}, \qquad \hat{\nu} = \frac{W_L^2(j)\|\hat{\mathbf{S}}\|^2}{W_R^2(j)\|\hat{\mathbf{S}}\|^2}, \tag{37}$$

respectivamente. Igualando estas dos relaciones de energía y utilizando la hipótesis (j) = 2 - W_R(j), obtenemos

$$W_L = \frac{2\sqrt{\mathbf{S}_L^T\mathbf{S}_L}}{\sqrt{\mathbf{S}_L^T\mathbf{S}_L} + \sqrt{\mathbf{S}_R^T\mathbf{S}_R}}, \quad W_R = 2 - W_L \ . \tag{38}$$

que proporcionan los componentes de factor de equilibrio, del factor de equilibrio generado. Debe observarse que el factor de equilibrio calculado en (38) es en este caso independiente de G_j, y por lo tanto ya no es función de j, proporcionando un factor de equilibrio autocorrelacionado que es independiente de la consideración de la ganancia; se muestra adicionalmente un factor de equilibrio dependiente en la figura 10 de los dibujos. Utilizando este resultado con las ecuaciones 29 y 32, podemos extender la selección del índice j de escalamiento de la capa central óptimo, para incluir los segmentos de vector concatenados k, de manera que:

$$j^* = \arg\min_{0 \le j < M} \left\{ \sum_k \left(\left\| \mathbf{S}_{Lk} - W_{Lk} \cdot \mathbf{G}_{jk} \cdot \hat{\mathbf{S}}_k \right\|^2 + \left\| \mathbf{S}_{Rk} - W_{Rk} \cdot \mathbf{G}_{jk} \cdot \hat{\mathbf{S}}_k \right\|^2 \right) \right\} \tag{39}$$

una representación del valor de ganancia óptimo. Este índice del valor de ganancia j* es transmitido como una señal de salida del codificador de la capa de mejora.

Haciendo referencia a continuación a la figura 10, se muestra un diagrama de bloques 1000 de un codificador de la capa de mejora y un descodificador de la capa de mejora, de acuerdo con diversas realizaciones. Las señales de audio de entrada s(n) son recibidas por el generador 1050 del factor de equilibrio del codificador 1010 de la capa de mejora y el generador 1030 de la señal de error (señal de distorsión) del generador 1020 del vector de ganancia. La señal de audio codificada S(n) procedente de la capa central es recibida por la unidad de escalamiento 1025 del generador 1020 del vector de ganancia, tal como se muestra. La unidad de escalamiento 1025 funciona para escalar la señal de audio codificada S (n) con una serie de valores de ganancia, a efectos de generar una serie de señales de audio codificadas candidatas, donde por lo menos una de las señales de audio codificadas candidatas está escalada. Tal como se ha mencionado previamente, puede utilizarse el escalamiento mediante la unidad o mediante cualquier función identidad deseada. La unidad de escalamiento 1025 emite el audio escalado S_j, que es recibido por el generador 1030 del factor de equilibrio. Generar el factor de equilibrio con una serie de componentes de factor de equilibrio, asociados cada uno con una señal de audio de las señales de audio de múltiples canales recibidas mediante el codificador 1010 de la capa de mejora, se ha discutido anteriormente en relación con las ecuaciones (18), (21), (24) y (33). Esto se consigue mediante el generador 1050 del factor de equilibrio, tal como se muestra, para producir componentes de factor de equilibrio $\hat{S}_L(n)$, $\hat{S}_R(n)$, tal como se muestra. Tal como se ha comentado en relación con la ecuación (38) anterior, el generador 1030 del factor de equilibrio muestra un factor de equilibrio independiente de la ganancia.

El generador 1020 del vector de ganancia es responsable de determinar un valor de ganancia a aplicar a la señal de audio codificada a efectos de generar una estimación de la señal de audio de múltiples canales, tal como se ha descrito en las ecuaciones (27), (28) y (29). Esto se consigue mediante la unidad de escalamiento 1025 y el generador 1050 del factor de equilibrio, que funcionan conjuntamente para generar la estimación en base al factor de equilibrio y por lo menos a una señal de audio codificada escalada. El valor de ganancia está basado en el factor de equilibrio y en la señal de audio de múltiples canales, donde el valor de ganancia está configurado para minimizar un valor de distorsión entre la señal de audio de múltiples canales y la estimación de la señal de audio de múltiples canales. La ecuación (30) describe la generación de un valor de distorsión en función de la estimación de la señal de entrada de múltiples canales y la propia señal de entrada real. Por lo tanto, los componentes del factor de equilibrio son recibidos por el generador 1030 de la señal de error, junto con las señales de audio de entrada s(n), para determinar un valor de error E_j para cada vector de escalamiento utilizado por la unidad de escalamiento 1025. Estos vectores de error se pasan a los circuitos 1035 del selector de ganancia junto con los valores de ganancia utilizados en la determinación de los vectores de error y un error específico E* basado en el valor de ganancia óptimo g*. A continuación, el selector de ganancia 1035 está operativo para evaluar el valor de la distorsión en base a la

estimación de la señal de entrada de múltiples canales y a la propia señal real, a efectos de determinar una representación de un valor de ganancia óptimo g* de los valores de ganancia posibles. Una palabra de código (i_g) que representa la ganancia óptima g* es emitida desde el selector de ganancia 1035 y recibida por el multiplexor MUX 1040, tal como se muestra.

i_g e i_B son emitidos ambos al multiplexor 1040 y transmitidos mediante el transmisor 1045 al descodificador 1060 de la capa de mejora a través del canal 125. La representación del valor de ganancia i_g es entregada para su transmisión al canal 125, tal como se muestra, pero puede asimismo ser almacenada si se desea.

En el lado del descodificador, durante el funcionamiento del descodificador 1060 de la capa de mejora, i_g e i_E son recibidos desde el canal 125 y desmultiplexados mediante el demux 1065. De este modo, el descodificador de la capa de mejora recibe una señal de audio codificada $\hat{S}(n)$, un factor de equilibrio codificado i_B y un valor de ganancia codificado i_g. El descodificador 1070 del vector de ganancia comprende un generador 1075 de ganancia selectiva en frecuencias y una unidad de escalamiento 1080, tal como se muestra. El descodificador 1070 del vector de ganancia genera un valor de ganancia descodificador a partir del valor de ganancia codificado. El valor de ganancia codificado i_g es introducido al generador 1075 de ganancia selectiva en frecuencias para producir el vector de ganancia g*, de acuerdo con el correspondiente método del codificador 1010. A continuación, el vector de ganancia g* es aplicado a la unidad de escalamiento 1080, que escala la señal de audio codificada S (n) con el valor de ganancia descodificador g* para generar la señal de audio escalada. El combinador de señal 1095 recibe las señales emitidas del factor de equilibrio codificado desde el descodificador 1090 del factor de equilibrio a la señal de audio escalada $G_j\hat{S}$ (n), a efectos de generar y entregar una señal de audio de múltiples canales descodificada, mostrada como las señales de audio de salida mejoradas.

El diagrama de bloques 1100 muestra un codificador de la capa de mejora y un descodificador de la capa de mejora a modo de ejemplo en los que, tal como se ha descrito en relación con la ecuación (33) anterior, el generador 1030 del factor de equilibrio genera un factor de equilibrio que es dependiente de la ganancia. Esto se muestra mediante el generador de señal de error, que genera la señal G_j 1110.

Haciendo referencia a continuación a las figuras 12 a 14, se muestran flujos que abarcan la metodología de las diversas realizaciones presentadas en este documento. En el flujo 1200 de la figura 12, se presenta un método para codificar una señal de audio de múltiples canales. En el bloque 1210, se recibe una señal de audio de múltiples canales que tiene una serie de señales de audio. En el bloque 1220, la señal de audio de múltiples canales se codifica para generar una señal de audio codificada. La señal de audio codificada puede ser una señal mono o de múltiples canales, tal como una señal estéreo que se muestra a modo de ejemplo en los dibujos. Además, la señal de audio codificada puede comprender una serie de canales. Por lo tanto, puede haber más de un canal en la capa central, y el número de canales en la capa de mejora puede ser mayor que el número de canales en la capa central. A continuación, en el bloque 1230, se genera un factor de equilibrio que tiene componentes de factor de equilibrio asociados cada uno con una señal de audio de la señal de audio de múltiples canales. Las ecuaciones (18), (21), (24) y (33) describen la generación del factor de equilibrio. Cada componente del factor de equilibrio puede depender de otros componentes del factor de equilibrio generados, tal como es el caso en la ecuación (38). La generación del factor de equilibrio puede comprender generar un valor de correlación entre la señal de audio codificada escalada y por lo menos una de las señales de audio de la señal de audio de múltiples canales, tal como en las ecuaciones (33) y (36). Puede generarse una autocorrelación entre por lo menos una de las señales de audio, tal como en la ecuación (38), a partir de la cual puede generarse una raíz cuadrada. En el bloque 1240, se determina un valor de ganancia a aplicar a la señal de audio codificada, a efectos de generar una estimación de la señal de audio de múltiples canales en base al factor de equilibrio y a la señal de audio de múltiples canales. El valor de ganancia está configurado para minimizar un valor de distorsión entre la señal de audio de múltiples canales y la estimación de la señal de audio de múltiples canales. Las ecuaciones (27), (28), (29) y (30) describen la determinación del valor de ganancia. Puede escogerse un valor de ganancia entre una serie de valores de ganancia para escalar la señal de audio codificada y para generar las señales de audio codificadas escaladas. El valor de distorsión puede generarse en base a esta estimación; el valor de ganancia puede basarse en el valor de distorsión. En el bloque 1250, se emite una representación del valor de ganancia para transmisión y/o almacenamiento.

El flujo 1300 de la figura 13 describe otra metodología para codificar la señal de audio de múltiples canales, de acuerdo con varias realizaciones. En el bloque 1310, se recibe una señal de audio de múltiples canales que tiene una serie de señales de audio. En el bloque 1320, la señal de audio de múltiples canales se codifica para generar una señal de audio codificada. Los procesos de los bloques 1310 y 1320 se llevan a cabo mediante un codificador de la capa central, tal como se ha descrito anteriormente. Tal como se ha indicado previamente, la señal de audio codificada puede ser una señal mono o de múltiples canales, tal como una señal estéreo que se muestra a modo de ejemplo en los dibujos. Además, la señal de audio codificada puede comprender una serie de canales. Por lo tanto, puede haber más de un canal en la capa central, y el número de canales en la capa de mejora puede ser mayor que el número de canales en la capa central.

En el bloque 1330, la señal de audio codificada es escalada con una serie de valores de ganancia para generar una serie de señales de audio codificadas candidatas, siendo escalada por lo menos una de las señales de audio codificadas candidatas. El escalamiento se consigue mediante la unidad de escalamiento del generador del vector de ganancia. Tal como se ha descrito, escalar la señal de audio codificada puede incluir el escalamiento con un valor

de ganancia de la unidad. El valor de ganancia de dicha serie de valores de ganancia puede ser una matriz de ganancia con el vector g_j como componente diagonal, tal como se ha descrito anteriormente. La matriz de ganancia puede ser selectiva en frecuencias. Ésta puede ser dependiente de la salida de la capa central, la señal de audio codificada mostrada en los dibujos. Puede escogerse un valor de ganancia entre una serie de valores de ganancia para escalar la señal de audio codificada y para generar las señales de audio codificadas escaladas. En el bloque 1340, se genera un factor de equilibrio que tiene componentes de factor de equilibrio asociados cada uno con una señal de audio de la señal de audio de múltiples canales. La generación del factor de equilibrio se lleva a cabo mediante el generador del factor de equilibrio. Cada componente del factor de equilibrio puede depender de otros componentes del factor de equilibrio generados, tal como es el caso en la ecuación (38). La generación del factor de equilibrio puede comprender generar un valor de correlación entre la señal de audio codificada escalada y por lo menos una de las señales de audio de la señal de audio de múltiples canales, tal como en las ecuaciones (33) y (36). Puede generarse una autocorrelación entre por lo menos una de las señales de audio, tal como en la ecuación (38), a partir de la cual puede generarse una raíz cuadrada.

En el bloque 1350, se genera una estimación de la señal de audio de múltiples canales en base al factor de equilibrio y a por lo menos una señal de audio codificada escalada. La estimación se genera en base a la señal o señales de audio codificadas escaladas y al factor de equilibrio generado. La estimación puede comprender una serie de estimaciones correspondientes a dicha serie de señales de audio codificadas candidatas. Un valor de distorsión es evaluado y/o puede ser generado en base a la estimación de la señal de audio de múltiples canales y a la señal de audio de múltiples canales, para determinar una representación de un valor de ganancia óptimo de los valores de ganancia, en el bloque 1360. El valor de distorsión puede comprender una serie de valores de distorsión correspondientes a dicha serie de estimaciones. La evaluación del valor de distorsión se lleva a cabo mediante los circuitos del selector de ganancia. La presentación del valor de ganancia óptimo está dada por la ecuación (39). En el bloque 1370, puede entregarse una representación del valor de ganancia para transmisión y/o almacenamiento. El transmisor del codificador de la capa de mejora puede transmitir la representación del valor de ganancia, tal como se ha descrito anteriormente.

El proceso realizado en el diagrama de flujo 1400 de la figura 14 muestra la descodificación de una señal de audio de múltiples canales. En el bloque 1410, se reciben una señal de audio codificada, un factor de equilibrio codificado y un valor de ganancia codificado. Se genera un valor de ganancia descodificado a partir del valor de ganancia codificado, en el bloque 1420. El valor de ganancia puede ser una matriz de ganancia, descrita previamente, y la matriz de ganancia puede ser selectiva en frecuencias. La matriz de ganancia puede depender asimismo del audio codificado recibido como una salida de la capa central. Además, la señal de audio codificada puede ser una señal mono o de múltiples canales, tal como una señal estéreo que se muestra a modo de ejemplo en los dibujos. Adicionalmente, la señal de audio codificada puede comprender una serie de canales. Por ejemplo, puede haber más de un canal en la capa central y el número de canales en la capa de mejora puede ser mayor que el número de canales en la capa central.

En el bloque 1430, la señal de audio codificada es escalada con el valor de ganancia descodificado, para generar una señal de audio escalada. El factor de equilibrio codificado es aplicado a la señal de audio escalada para generar una señal de audio de múltiples canales descodificada, en el bloque 1440. La señal de audio de múltiples canales descodificada es entregada en el bloque 1450.

Cálculo de máscara de escalamiento selectiva basado en detección de picos

La matriz de ganancia G_j selectiva en frecuencias, que es una matriz diagonal con elementos que forman un vector de ganancia g_j, puede definirse tal como en (14) más arriba:

$$g_j(k) = \begin{cases} \alpha 10^{(-j\cdot\Delta/20)}; & k_l \le k \le k_h \\ \alpha; & \text{de lo contrario} \end{cases}, \quad 0 \le j < M,$$

(40)

donde Δ es un tamaño del paso (por ejemplo, Δ ≈ 2,0 dB), α es una constante, M es el número de candidatos (por ejemplo, M = 8, que puede representarse utilizando solamente 3 bits), y k_l y k_h son límites de baja y alta frecuencia, respectivamente, sobre los cuales puede tener lugar la reducción de ganancia. En este caso, k representa el k-ésimo MDCT o coeficiente de transformada de Fourier. Debe observarse que g_j es selectivo en frecuencias pero es independiente de la salida de la capa anterior. Los vectores de ganancia g_j pueden basarse en alguna función de los elementos codificados de un vector de señal codificado anteriormente, en este caso S. Esto puede expresarse como:

$$g_j(k) = f\left(k, \hat{\mathbf{S}}\right).$$

(41)

En un sistema de codificación integrado de múltiples capas (con más de 2 capas), la salida $\hat{S}$ que ha de ser escalada mediante el vector de ganancia g_j, se obtiene a partir de la contribución de por lo menos dos capas anteriores. Es decir

$$\hat{\mathbf{S}} = \hat{\mathbf{E}}_2 + \hat{\mathbf{S}}_1,$$

(42)

donde $\hat{S}_l$ es la salida de la primera capa (capa central) y $\hat{E}_2$ es la contribución de la segunda capa o de la primera capa de mejora. En este caso, los vectores de ganancia g_j pueden ser alguna función de los elementos codificados de un vector de señal $\hat{S}$ codificado anteriormente y de la contribución de la primera capa de mejora.

$$g_j(k) = f\left(k, \hat{\mathbf{S}}, \hat{\mathbf{E}}_2\right).$$

(43)

Se ha observado que la mayor parte del ruido audible debido al modelo de codificación de la capa inferior está en los valles y no en los picos. En otras palabras, existe una mejor adaptación entre el original y el espectro codificado en los picos espectrales. Por lo tanto, los picos no deberían ser modificados, es decir el escalamiento debería limitarse a los valles. Para utilizar ventajosamente esta observación, en una de las realizaciones la función de la ecuación (41) se basa en picos y valles de S. Sea $\Psi(\hat{S})$ una máscara de escalamiento basada en las magnitudes de los picos de S detectados. La máscara de escalamiento puede ser una función vectorial con valores distintos de cero en los picos detectados, es decir

$$\psi(\hat{\mathbf{S}}) = \begin{cases} \hat{s}_i & \text{pico presente} \\ 0 & \text{de lo contrario} \end{cases},$$

(44)

donde $\hat{s}_i$ es el i-ésimo elemento de S. La ecuación (41) puede modificarse a continuación como:

$$g_j(k) = f\left(k, \hat{\mathbf{S}}\right) = \begin{cases} \alpha 10^{(-j\cdot\Delta/20)}; & k_l \le k \le k_h, \psi_k(\hat{\mathbf{S}}) = 0 \\ \alpha; & \text{de lo contrario} \end{cases} \quad 0 \le j < M$$

(45)

Pueden utilizarse diversos enfoques para la detección de picos. En la realización preferida, los picos se detectan pasando el espectro absoluto $|\hat{S}|$ a través de dos filtros de promediado ponderados independientes y comparando a continuación las salidas filtradas. Sean A_1 y A_2 la representación matricial de dos filtros de promediado. Sean l_1 y l_2 ($h > l_2$) las longitudes de los dos filtros. La función de detección de picos está dada por:

$$\psi(\hat{\mathbf{S}}) = \begin{cases} \hat{s}_i & \mathbf{A}_2 \,|\hat{\mathbf{S}}| > \beta \cdot \mathbf{A}_1 \,|\hat{\mathbf{S}}| \\ 0 & \text{de lo contrario} \end{cases},$$

(46)

donde β es un valor umbral empírico.

A modo de ejemplo ilustrativo, se hace referencia a la figura 15 y la figura 16. En este caso, el valor absoluto de la señal codificada $|\hat{S}|$ en el dominio MDCT está dado en ambas representaciones como 1510. La señal es representativa de un sonido procedente de un diapasón, que crea una secuencia armónica separada regularmente,

tal como se muestra. La señal es difícil de codificar utilizando un codificador de la capa central en base a un modelo de voz, debido a que la frecuencia fundamental de esta señal está más allá del intervalo de lo que se considera razonable para una señal de voz. Esto tiene como resultado un nivel de ruido bastante alto producido por la capa central, lo que puede observarse comparando la señal codificada 1510 con la versión mono de la señal original |S| (1610).

A partir de la señal codificada (1510), se utiliza un generador de umbral para producir el umbral 1520, que corresponde a la expresión $\beta A_1 |\hat{S}|$ de la ecuación 45. En este caso, A_1 es una matriz de convolución que, en la realización preferida, implementa una convolución de la señal | S | con una ventana coseno de longitud 45. Son posibles muchas formas de ventana y éstas pueden comprender diferentes longitudes. Asimismo, en la realización preferida, A_2 es una matriz identidad. El detector de picos compara a continuación la señal 1510 con el umbral 1520 para producir la máscara de escalamiento $\psi(\hat{S})$, mostrada como 1530.

A continuación, pueden utilizarse los candidatos a vector de escalamiento de la capa central (proporcionados en la ecuación 45) para escalar el ruido entre picos de la señal codificada $|\hat{S}|$ a efectos de reproducir una señal reconstruida escalada 1620. El candidato óptimo puede elegirse de acuerdo con el proceso descrito en la ecuación 39 anterior o de otro modo.

Haciendo referencia a continuación a las figuras 17 a 19, se presentan diagramas de flujo que muestran metodología asociada con el cálculo de la máscara de escalamiento selectiva basado en detección de picos, descrito anteriormente en relación con diversas realizaciones. En el diagrama de flujo 1700 del centro de la figura 17, en el bloque 1710 se detecta un conjunto de picos en un vector de audio reconstruido S de una señal de audio recibida. La señal de audio puede estar integrada en múltiples capas. El vector de audio reconstruido S puede estar en el dominio de frecuencias y el conjunto de picos pueden ser picos en el dominio de frecuencias. Detectar el conjunto de picos se lleva a cabo de acuerdo con una función de detección de picos proporcionada por la ecuación (46), por ejemplo. Debe observarse que el conjunto puede estar vacío, tal como es el caso cuando todo está atenuado y no hay picos. En el bloque 1720, se genera una máscara de escalamiento $\psi(\hat{S})$ en base al conjunto de picos detectado. A continuación, en el bloque 1730, se genera un vector de ganancia g* en base a por lo menos la máscara de escalamiento y un índice de j representativo del vector de ganancia. En el bloque 1740, se escala la señal de audio reconstruida con el vector de ganancia para producir una señal de audio reconstruida escalada. En el bloque 1750, se genera una distorsión en base a la señal de audio y a la señal de audio reconstruida escalada. En el bloque 1760, se entrega el índice del vector de ganancia basado en la distorsión generada.

Haciendo referencia continuación a la figura 18, el diagrama de flujo 1800 muestra una realización alternativa de codificación de una señal de audio, de acuerdo con ciertas realizaciones. En el bloque 1810, se recibe una señal de audio. La señal de audio puede estar integrada en múltiples capas. A continuación, la señal de audio es codificada en el bloque 1820 para generar un vector de audio reconstruido $\hat{S}$. El vector de audio reconstruido S puede estar en el dominio de frecuencias y el conjunto de picos pueden ser picos en el dominio de frecuencias. En el bloque 1830, se detecta un conjunto de picos en el vector de audio reconstruido S de una señal de audio recibida. Detectar el conjunto de picos se lleva a cabo de acuerdo con una función de detección de picos proporcionada por la ecuación (46), por ejemplo. De nuevo, debe observarse que el conjunto puede estar vacío, tal como es el caso cuando todo está atenuado y no hay picos. En el bloque 1840, se genera una máscara de escalamiento $\psi(\hat{S})$ basada en el conjunto de picos detectado. En el bloque 1850, se generan una serie de vectores de ganancia g_j basados en la máscara de escalamiento. La señal de audio reconstruida es escalada con dicha serie de vectores de ganancia para producir una serie de señales de audio reconstruidas escaladas, en el bloque 1860. A continuación, en el bloque 1870 se generan una serie de distorsiones en base a la señal de audio y la serie de señales de audio reconstruidas escaladas. En el bloque 1880 se elige un vector de ganancia entre dicha serie de vectores de ganancia en base a dicha serie de distorsiones. El vector de ganancia puede elegirse para corresponder a una distorsión mínima de la serie de distorsiones. El índice representativo del vector de ganancia es entregado para ser transmitido y/o almacenado, en el bloque 1890.

Los flujos de codificador mostrados en las figuras 17 y 18 anteriores pueden implementarse mediante la estructura de aparato descrita previamente. Haciendo referencia al flujo 1700, en un aparato operativo para codificar una señal de audio, un selector de ganancia, tal como el selector de ganancia 1035 del generador en 1020 de vectores de ganancia del codificador 1010 de la capa de mejora, detecta un conjunto de picos en un vector de audio reconstruido S de una señal de audio recibida y genera una máscara de escalamiento $\psi(\hat{S})$ basada en el conjunto de picos detectado. De nuevo, la señal de audio puede estar integrada en múltiples capas. El vector de audio reconstruido S puede estar en el dominio de frecuencias y el conjunto de picos pueden ser picos en el dominio de frecuencias. Detectar el conjunto de picos se lleva a cabo de acuerdo con una función de detección de picos proporcionada por la ecuación (46), por ejemplo. Debe observarse que el conjunto de picos puede ser nulo si la totalidad de la señal ha sido atenuada. Una unidad de escalamiento, tal como la unidad de escalamiento 1025 del generador 1020 del vector de ganancia genera un vector de ganancia g* basado, por lo menos, en la máscara de escalamiento y un índice j representativo del vector de ganancia, y escala la señal de audio reconstruida con el vector de ganancia para producir una señal de audio reconstruida escalada. El generador 1030 de la señal de error del generador 1025 del vector de ganancia genera una distorsión en base a la señal de audio y a la señal de audio reconstruida escalada. Un transmisor, tal como el transmisor 1045 del descodificador 1010 de la capa de mejora es operativo para entregar el índice del vector de ganancia en base a la distorsión generada.

Haciendo referencia al flujo 1800 de la figura 18, en un aparato operativo para codificar una señal de audio, un codificador recibe la señal de audio y codifica la señal de audio para generar un vector de audio reconstruido $\hat{S}$. Una unidad de escalamiento, tal como la unidad de escalamiento 1025 del generador 1020 del vector de ganancia detecta un conjunto de picos en el vector de audio reconstruido S de una señal de audio recibida, genera una máscara de escalamiento $\psi(\hat{S})$ en base al conjunto de picos detectado, genera una serie de vectores de ganancia g_j en base a la máscara de escalamiento, y escala la señal de audio reconstruida con dicha serie de vectores de ganancia para producir la serie de señales de audio reconstruidas escaladas. El generador 1030 de la señal de error genera una serie de distorsiones en base a la señal de audio y a la serie de señales de audio reconstruidas escaladas. Un selector de ganancia, tal como el selector de ganancia 1035, escoge un vector de ganancia entre dicha serie de vectores de ganancia en base a la serie de distorsiones. El transmisor 1045, por ejemplo, entrega para su posterior transmisión y y/o almacenamiento, el índice representativo del vector de ganancia.

En el diagrama de flujo 1900 de la figura 19, se muestra un método de descodificación de una señal de audio. En el bloque 1910 se recibe un vector de audio reconstruido S y un índice representativo de un vector de ganancia. En el bloque 1920, se detecta un conjunto de picos en el vector de audio reconstruido. Detectar el conjunto de picos se lleva a cabo de acuerdo con una función de detección de picos proporcionada por la ecuación (46), por ejemplo. De nuevo, debe observarse que el conjunto puede estar vacío, tal como es el caso cuando todo está atenuado y no hay picos. En el bloque 1930 se genera una máscara de escalamiento $\psi(\hat{S})$ basada en el conjunto de picos detectado. En el bloque 1940 se genera el vector de ganancia g* basado en por lo menos la máscara de escalamiento y el índice representativo del vector de ganancia. En el bloque 1950, el vector de audio reconstruido es escalado con el vector de ganancia para producir la señal de audio reconstruida escalada. El método puede incluir además generar una mejora para el vector de audio reconstruido y combinar a continuación la señal de audio reconstruida escalada y la mejora para el vector de audio reconstruido, a efectos de generar una señal descodificada mejorada.

El flujo de descodificador mostrado en la figura 19 puede ser implementado por la estructura del aparato descrita anteriormente. En un aparato operativo para descodificar una señal de audio, un descodificador 1070 del vector de ganancia de un descodificador 1060 de la capa de mejora, por ejemplo, recibe un vector de audio reconstruido S y un índice representativo de un vector de ganancia i_g. Tal como se muestra en la figura 10, i_g es recibido por el selector de ganancia 1075 mientras que el vector de audio reconstruido S es recibido por la unidad de escalamiento 1080 del descodificador 1070 del vector de ganancia. Un selector de ganancia, tal como el selector de ganancia 1075 del descodificador 1070 del vector de ganancia, detecta un conjunto de picos en el vector de audio reconstruido, genera una máscara de escalamiento $\psi(\hat{S})$ en base al conjunto de picos detectado, y genera el vector de ganancia g* en base a por lo menos la máscara de escalamiento y el índice representativo del vector de ganancia. De nuevo, el conjunto puede estar vacío si la señal está principalmente atenuada. El selector de ganancia detecta el conjunto de picos de acuerdo con una función de detección de picos, tal como la proporcionada en la ecuación (46), por ejemplo. Una unidad de escalamiento 1080, por ejemplo, escala el vector de audio reconstruido con el vector de ganancia para producir una señal de audio reconstruida escalada.

Además, un descodificador de la señal de error, tal como el descodificador 665 de la señal de error del descodificador de la capa de mejora de la figura 6, puede generar una mejora para el vector de audio reconstruido. Un combinador de señal, tal como el combinador de señal 675 de la figura 6, combina la señal de audio reconstruida escalada y la mejora para el vector de audio reconstruido a efectos de generar una señal descodificada mejorada.

Debe observarse además que los flujos dirigidos por factor de equilibrio de las figuras 12 a 14, y los flujos dirigidos por máscara de escalamiento selectiva con detección de picos de las figuras 17 a 19 pueden ambos llevarse a cabo en diversas combinaciones y éstas están soportadas por el aparato y la estructura descritos en la presente memoria.

Si bien la invención ha sido mostrada y descrita en particular haciendo referencia a una realización específica, los expertos en la materia comprenderán que pueden realizarse en la misma diversos cambios en la forma y los detalles sin apartarse del alcance de la invención. Por ejemplo, si bien las técnicas anteriores se han descrito en términos de transmisión y recepción sobre un canal en un sistema de telecomunicaciones, estas técnicas pueden aplicar igualmente a un sistema que utilice el sistema de compresión de señal con el propósito de reducir los requisitos de almacenamiento en un dispositivo multimedia digital, tal como un dispositivo de memoria de estado sólido o un disco duro de ordenador. El alcance de la invención está definido en las reivindicaciones adjuntas.

REIVINDICACIONES

1. Un aparato operativo para codificar una señal de audio, comprendiendo el aparato:

un selector de ganancia de un generador de vectores de ganancia de un codificador de la capa de mejora que detecta un conjunto de picos en un vector de audio reconstruido S de una señal de audio recibida, genera una máscara de escalamiento ψ(Ŝ) basada en el conjunto de picos detectados;

una unidad de escalamiento del generador del vector de ganancia, que genera un vector de ganancia g* en base a por lo menos la máscara de escalamiento y el índice j representativo del vector de ganancia, escala el vector de audio reconstruido Ŝ con el vector de ganancia para producir una señal de audio reconstruida escalada;

un generador de señal de error del generador de vectores de ganancia que genera una distorsión en base a la señal de audio y a la señal de audio reconstruida escalada; y

un transmisor del codificador de la capa de mejora que emite el índice del vector de ganancia basado en la distorsión generada.

2. El aparato según la reivindicación 1, en el que el selector de ganancia detecta el conjunto de picos además de acuerdo con una función de detección de picos dada por:

$$\psi(\hat{\mathbf{S}}) = \begin{cases} \hat{s}_i & \mathbf{A}_2 \,|\hat{\mathbf{S}}| > \beta \cdot \mathbf{A}_1 \,|\hat{\mathbf{S}}| \\ \mathbf{0} & \text{de lo contrario} \end{cases},$$ donde β es un valor umbral.

3. El aparato según la reivindicación 1, que comprende

un codificador que recibe una señal de audio de múltiples canales que comprende una serie de señales de audio y codifica la señal de audio de múltiples canales para generar una señal de audio codificada;

un generador de factor de equilibrio del codificador de la capa de mejora que recibe una señal de audio codificada y genera un factor de equilibrio que tiene una serie de componentes del factor de equilibrio asociados cada uno con una señal de audio de dicha serie de señales de audio de la señal de audio de múltiples canales;

en el que el generador del vector de ganancia del codificador de la capa de mejora determina un valor de ganancia a aplicar a la señal de audio codificada para generar una estimación de la señal de audio de múltiples canales en base al factor de equilibrio y a la señal de audio de múltiples canales, en el que el valor de ganancia está configurado para minimizar un valor de distorsión entre la señal de audio de múltiples canales y la estimación de la señal de audio de múltiples canales,

en el que el transmisor transmite además una representación del valor de ganancia para por lo menos uno de transmisión y almacenamiento.

4. El aparato según la reivindicación 3, en el que la unidad de escalamiento del codificador de la capa de mejora que escala la señal de audio codificada con una serie de valores de ganancia para generar una serie de señales de audio codificadas candidatas, en el que por lo menos una de las señales de audio codificadas candidatas es escalada;

en el que la unidad de escalamiento y el generador del factor de equilibrio generan la estimación de la señal de audio de múltiples canales en base al factor de equilibrio y a dicha por lo menos una señal de audio codificada escalada de dicha serie de señales de audio codificadas candidatas; y

en el que el selector de ganancia del codificador de la capa de mejora evalúa el valor de distorsión en base a la estimación de la señal de audio de múltiples canales y a la señal de audio de múltiples canales para determinar una representación de un valor de ganancia óptimo de dicha serie de valores de ganancia.

5. Un aparato operativo para codificar una señal de audio, comprendiendo el aparato:

un codificador que recibe la señal de audio y codifica la señal de audio para generar un vector de audio reconstruido Ŝ;

una unidad de escalamiento de un generador de vectores de ganancia de un codificador de la capa de mejora que detecta un conjunto de picos en el vector de audio reconstruido Ŝ de una señal de audio recibida, genera una máscara de escalamiento ψ(Ŝ) en base al conjunto de picos detectado, genera una serie de vectores de ganancia g_j en base a la máscara de escalamiento, y escala el vector de audio reconstruido Ŝ con dicha serie de vectores de ganancia para producir una serie de señales de audio reconstruidas escaladas;

un generador de señal de error del generador de vectores de ganancia, que genera una serie de distorsiones en base a la señal de audio y a la serie de señales de audio reconstruidas escaladas;

un selector de ganancia del generador de vectores de ganancia que elige un vector de ganancia entre la serie de vectores de ganancia en base a la serie de distorsiones; y

un transmisor del codificador de la capa de mejora que entrega, para por lo menos uno de transmisión y almacenamiento, el índice representativo del vector de ganancia.

6. El aparato según la reivindicación 5, en el que se elige el vector de ganancia que corresponde a una distorsión mínima de la serie de distorsiones.

7. El aparato según la reivindicación 5, en el que la unidad de escalamiento detecta el conjunto de picos de acuerdo con una función de detección de picos dada por:

$$\psi(\hat{\mathbf{S}}) = \begin{cases} \hat{s}_i & \mathbf{A}_2 \,|\hat{\mathbf{S}}| > \beta \cdot \mathbf{A}_1 \,|\hat{\mathbf{S}}| \\ \mathbf{0} & \text{de lo contrario} \end{cases},$$ donde β es un valor umbral.

8. El aparato según la reivindicación 1 o la reivindicación 5, en el que la señal de audio está integrada en múltiples capas.

9. El aparato según la reivindicación 1 o la reivindicación 5, en el que el vector de audio reconstruido S está en el dominio de frecuencias y el conjunto de picos son picos en el dominio de frecuencias.

10. Un método para codificar una señal de audio, comprendiendo el método:

detectar un conjunto de picos en un vector de audio reconstruido Ŝ de una señal de audio recibida;

generar una máscara de escalamiento ψ(Ŝ) basada en el conjunto de picos detectado;

generar un vector de ganancia g* basado en por lo menos la máscara de escalamiento y un índice j representativo del vector de ganancia;

escalar el vector de audio reconstruido Ŝ con el vector de ganancia para producir una señal de audio reconstruida escalada;

generar una distorsión en base a la señal de audio y a la señal de audio reconstruida escalada; y

entregar el índice del vector de ganancia en base a la distorsión generada.

11. El método según la reivindicación 10, en el que la detección del conjunto de picos comprende además una función de detección de picos dada por:

$$\psi(\hat{\mathbf{S}}) = \begin{cases} \hat{s}_i & \mathbf{A}_2 \,|\hat{\mathbf{S}}| > \beta \cdot \mathbf{A}_1 \,|\hat{\mathbf{S}}| \\ \mathbf{0} & \text{de lo contrario} \end{cases},$$ donde β es un valor umbral.

12. El método según la reivindicación 10, en el que la señal de audio está integrada en múltiples capas.

13. El método según la reivindicación 10, en el que el vector de audio reconstruido S está en el dominio de frecuencias y el conjunto de picos son picos del dominio de frecuencias.

14. El método según la reivindicación 10, que comprende además:

recibir una señal de audio de múltiples canales que comprende una serie de señales de audio;

codificar la señal de audio de múltiples canales para generar una señal de audio codificada;

generar un factor de equilibrio que tiene una serie de componentes del factor de equilibrio asociados cada uno con una señal de audio de dicha serie de señales de audio de la señal de audio de múltiples canales;

determinar un valor de ganancia a aplicar a la señal de audio codificada para generar una estimación de la señal de audio de múltiples canales en base al factor de equilibrio y a la señal de audio de múltiples canales, en el que el valor de ganancia está configurado para minimizar un valor de distorsión entre la señal de audio de múltiples canales y la estimación de la señal de audio de múltiples canales; y

entregar una representación del valor de ganancia para por lo menos uno de transmisión y almacenamiento.

15. El método según la reivindicación 10, que comprende además:

recibir una señal de audio de múltiples canales que comprende una serie de señales de audio;

codificar la señal de audio de múltiples canales para generar una señal de audio codificada;

escalar la señal de audio codificada con una serie de valores de ganancia para generar una serie de señales de audio codificadas candidatas, en el que por lo menos una de las señales de audio codificadas candidatas es escalada;

generar un factor de equilibrio que tiene una serie de componentes del factor de equilibrio asociados cada uno con una señal de audio de dicha serie de señales de audio de la señal de audio de múltiples canales;

generar una estimación de la señal de audio de múltiples canales en base al factor de equilibrio y a dicha por lo menos una señal de audio codificada escalada de la serie de señales de audio codificadas candidatas;

evaluar un valor de distorsión en base a la estimación de la señal de audio de múltiples canales y a la señal de audio de múltiples canales para determinar una representación de un valor de ganancia óptimo de dicha serie de valores de ganancia;

entregar, para por lo menos uno de transmisión y almacenamiento, la representación del valor de ganancia óptimo.

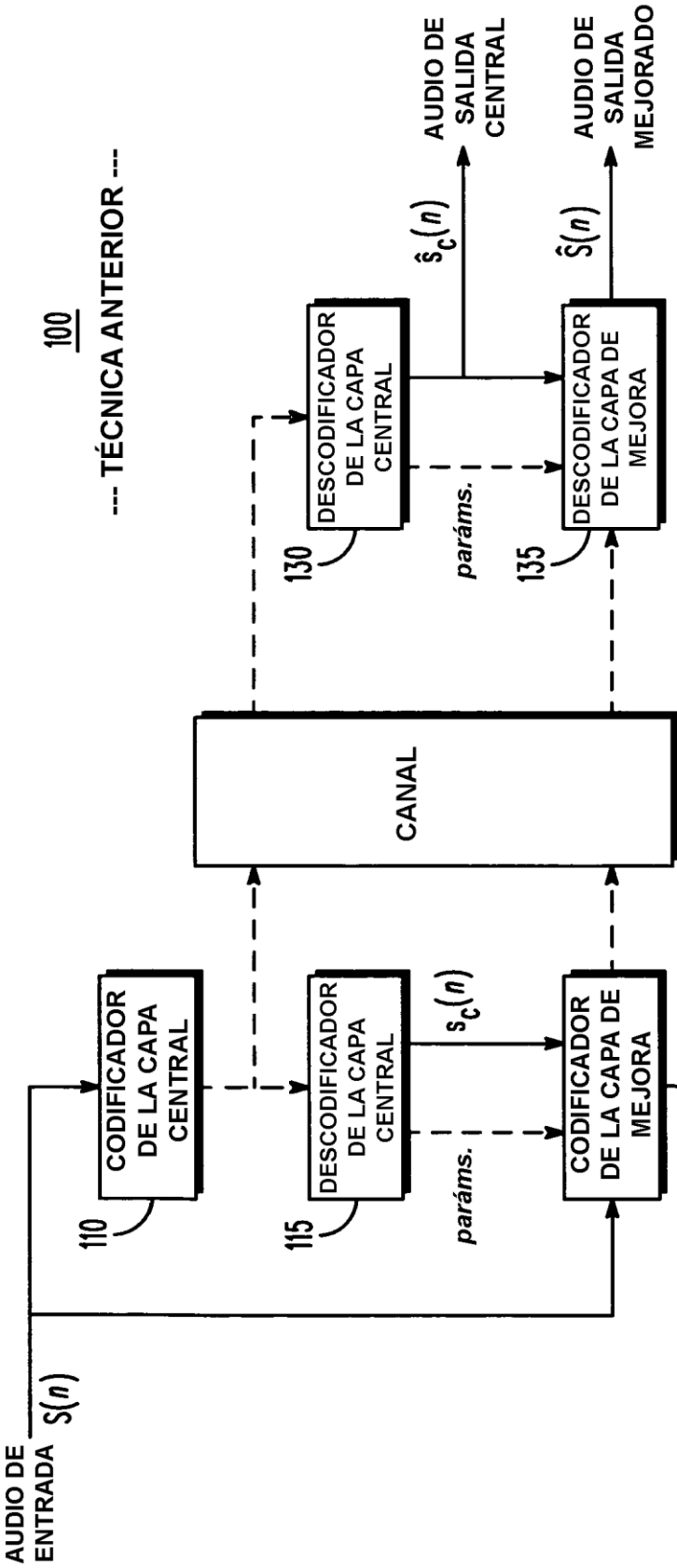

100
--- TÉCNICA ANTERIOR ---
AUDIO DE ENTRADA
S(n)
CODIFICADOR DE LA CAPA CENTRAL
110
DESCODIFICADOR DE LA CAPA CENTRAL
115
paráms.
$s_c(n)$
CODIFICADOR DE LA CAPA DE MEJORA
120
CANAL
125
DESCODIFICADOR DE LA CAPA CENTRAL
130
paráms.
DESCODIFICADOR DE LA CAPA DE MEJORA
135
$\hat{s}_c(n)$
AUDIO DE SALIDA CENTRAL
$\hat{S}(n)$
AUDIO DE SALIDA MEJORADO

FIG. 1

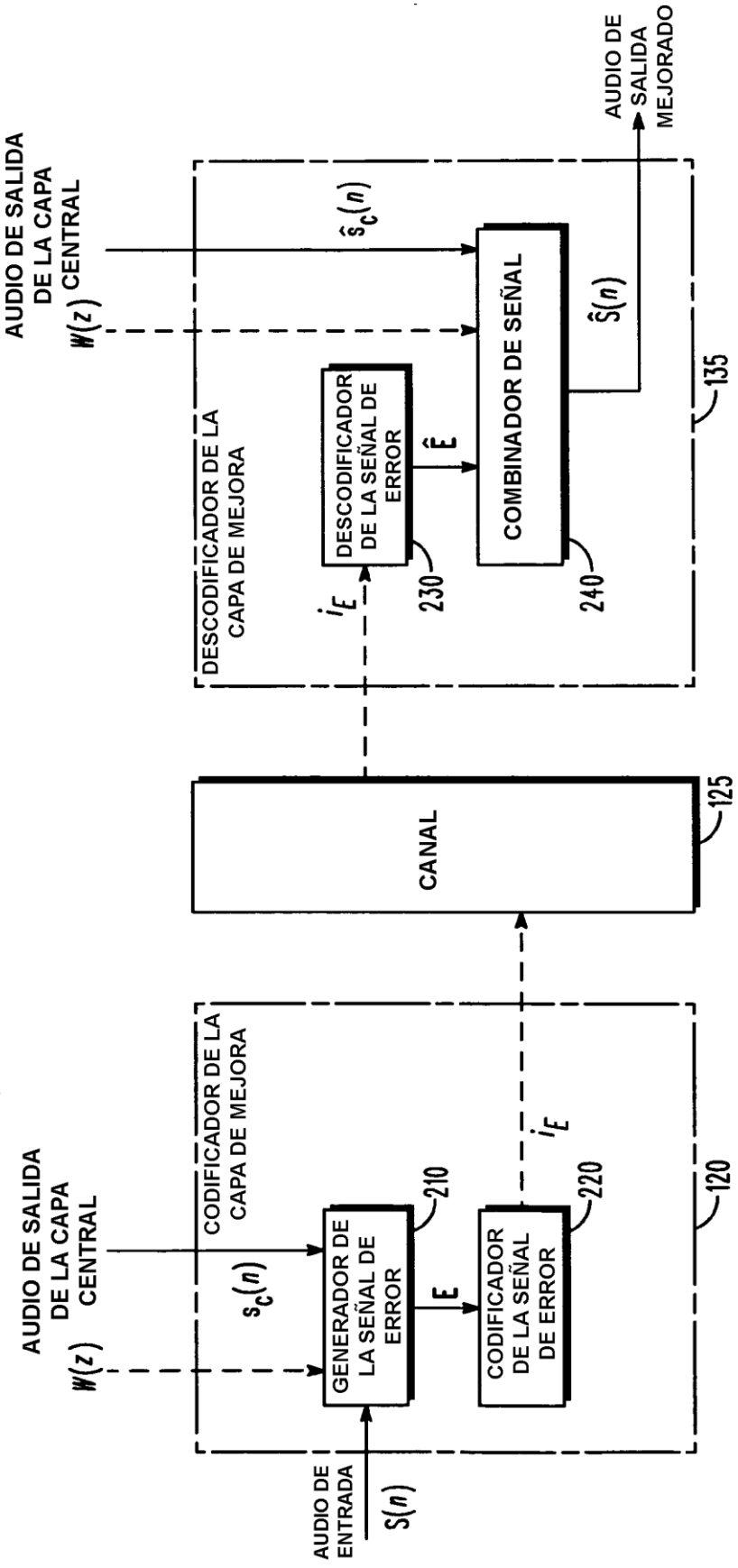
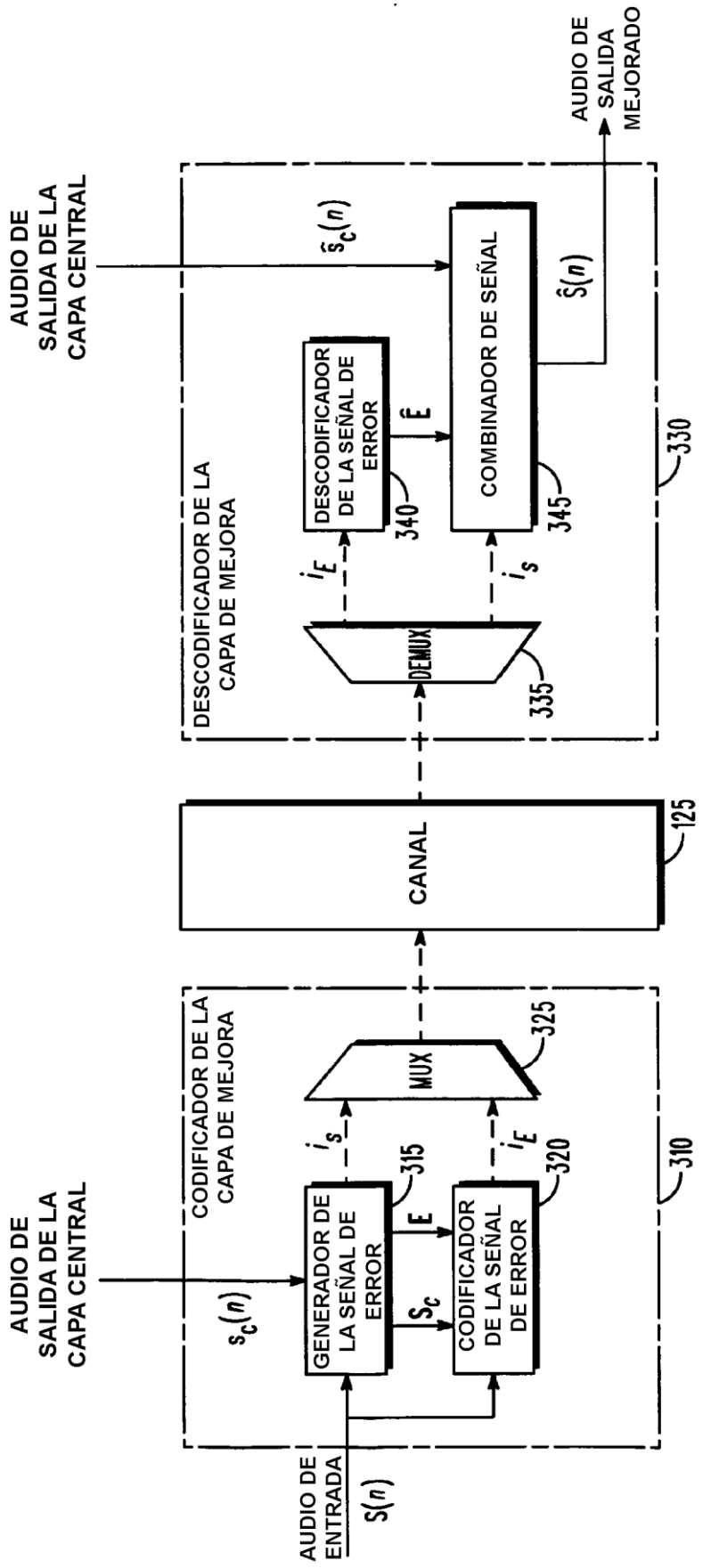

AUDIO DE ENTRADA
S(n)
AUDIO DE SALIDA DE LA CAPA CENTRAL
W(z)
CODIFICADOR DE LA CAPA DE MEJORA
s_c(n)
GENERADOR DE LA SEÑAL DE ERROR
210
E
CODIFICADOR DE LA SEÑAL DE ERROR
220
i_E
120
CANAL
125
200
--- TÉCNICA ANTERIOR ---
AUDIO DE SALIDA DE LA CAPA CENTRAL
W(z)
DESCODIFICADOR DE LA CAPA DE MEJORA
i_E
DESCODIFICADOR DE LA SEÑAL DE ERROR
230
Ê
COMBINADOR DE SEÑAL
240
ŝ_c(n)
Ŝ(n)
135
AUDIO DE SALIDA MEJORADO

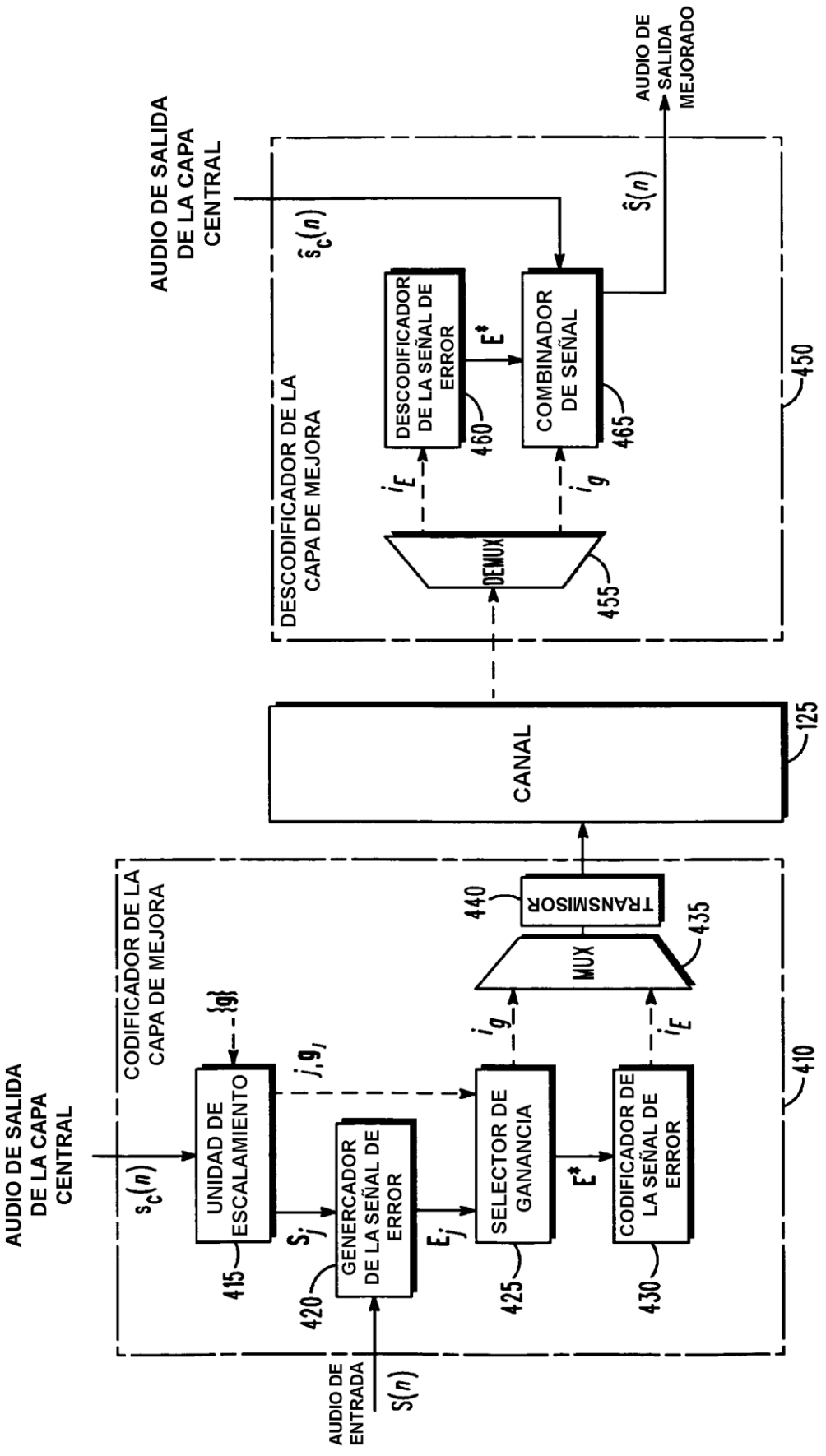
FIG. 2



300

--- TÉCNICA ANTERIOR ---

FIG. 3


AUDIO DE SALIDA DE LA CAPA CENTRAL
CODIFICADOR DE LA CAPA DE MEJORA
UNIDAD DE ESCALAMIENTO
GENERCADOR DE LA SEÑAL DE ERROR
SELECTOR DE GANANCIA
CODIFICADOR DE LA SEÑAL DE ERROR
MUX
TRANSMISOR
CANAL
DESCODIFICADOR DE LA CAPA DE MEJORA
DESCODIFICADOR DE LA SEÑAL DE ERROR
COMBINADOR DE SEÑAL
DEMUX
AUDIO DE SALIDA MEJORADO
AUDIO DE ENTRADA
410
415
420
425
430
435
440
450
455
460
465
125

400

FIG. 4

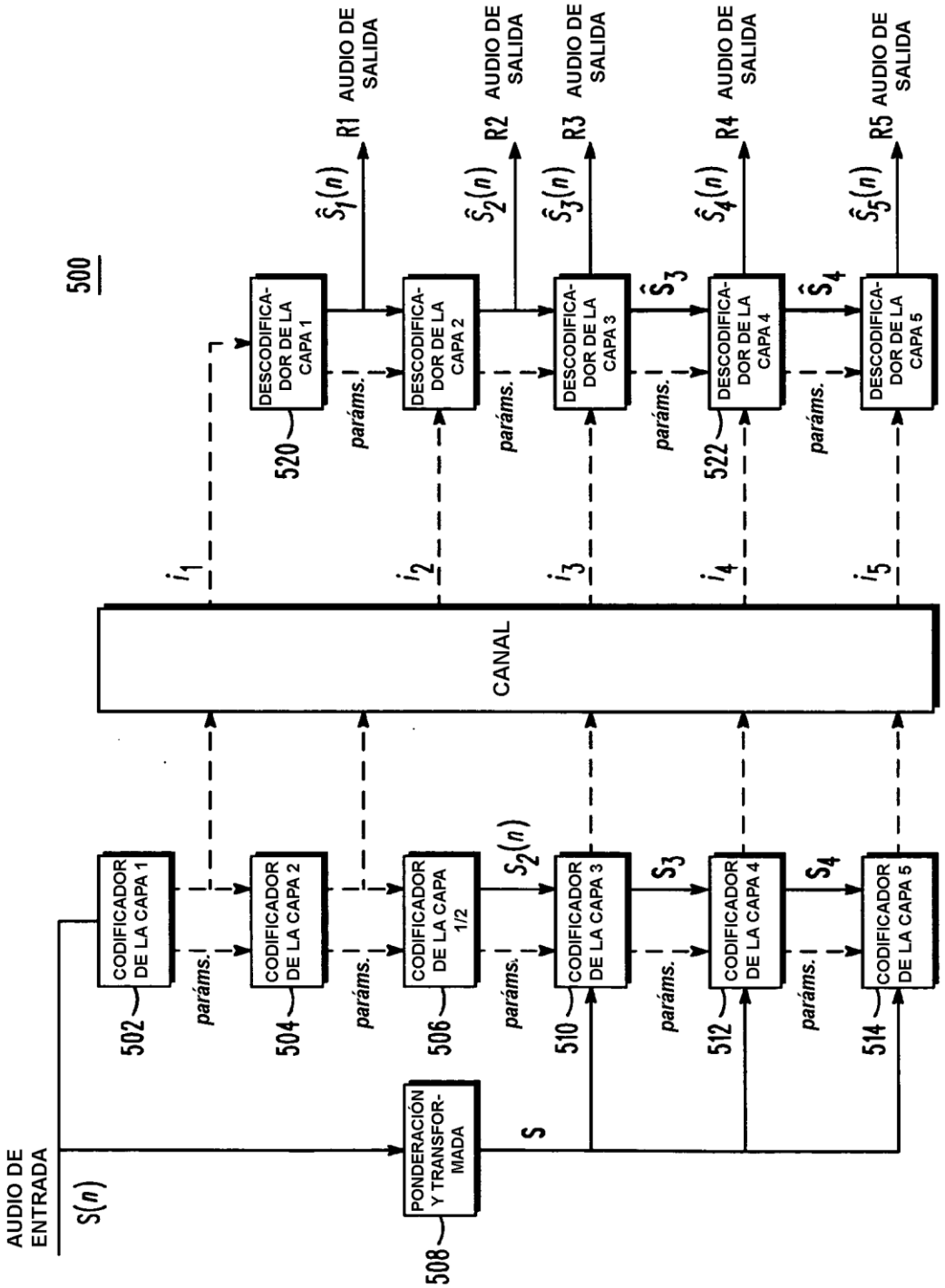

500
AUDIO DE ENTRADA
S(n)
PONDERACIÓN Y TRANSFOR-MADA
508
S
CODIFICADOR DE LA CAPA 1
502
CODIFICADOR DE LA CAPA 2
504
CODIFICADOR DE LA CAPA 1/2
506
CODIFICADOR DE LA CAPA 3
510
CODIFICADOR DE LA CAPA 4
512
CODIFICADOR DE LA CAPA 5
514
paráms.
$S_2(n)$
S_3
S_4
CANAL
i_1
i_2
i_3
i_4
i_5
DESCODIFICA-DOR DE LA CAPA 1
520
DESCODIFICA-DOR DE LA CAPA 2
DESCODIFICA-DOR DE LA CAPA 3
DESCODIFICA-DOR DE LA CAPA 4
522
DESCODIFICA-DOR DE LA CAPA 5
$\hat{S}_1(n)$
$\hat{S}_2(n)$
$\hat{S}_3(n)$
$\hat{S}_4(n)$
$\hat{S}_5(n)$
$\hat{S}_3$
$\hat{S}_4$
R1 AUDIO DE SALIDA
R2 AUDIO DE SALIDA
R3 AUDIO DE SALIDA
R4 AUDIO DE SALIDA
R5 AUDIO DE SALIDA

FIG. 5

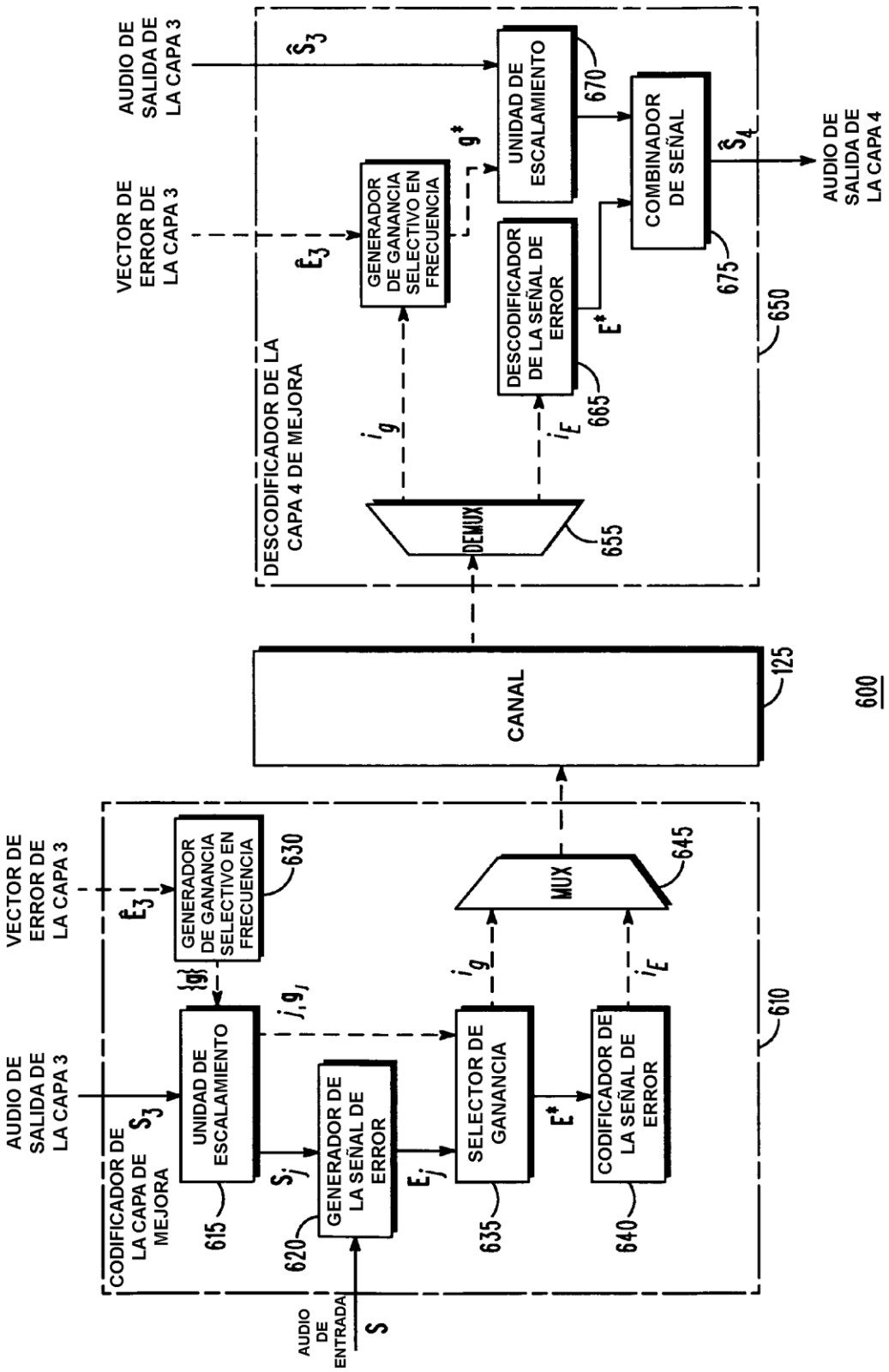

AUDIO DE SALIDA DE LA CAPA 3
VECTOR DE ERROR DE LA CAPA 3
AUDIO DE SALIDA DE LA CAPA 4
DESCODIFICADOR DE LA CAPA 4 DE MEJORA
GENERADOR DE GANANCIA SELECTIVO EN FRECUENCIA
UNIDAD DE ESCALAMIENTO
COMBINADOR DE SEÑAL
DESCODIFICADOR DE LA SEÑAL DE ERROR
DEMUX
650
655
665
670
675
CANAL
125
600
MUX
645
CODIFICADOR DE LA CAPA DE MEJORA
AUDIO DE ENTRADA
GENERADOR DE LA SEÑAL DE ERROR
SELECTOR DE GANANCIA
CODIFICADOR DE LA SEÑAL DE ERROR
610
615
620
630
635
640

FIG. 6

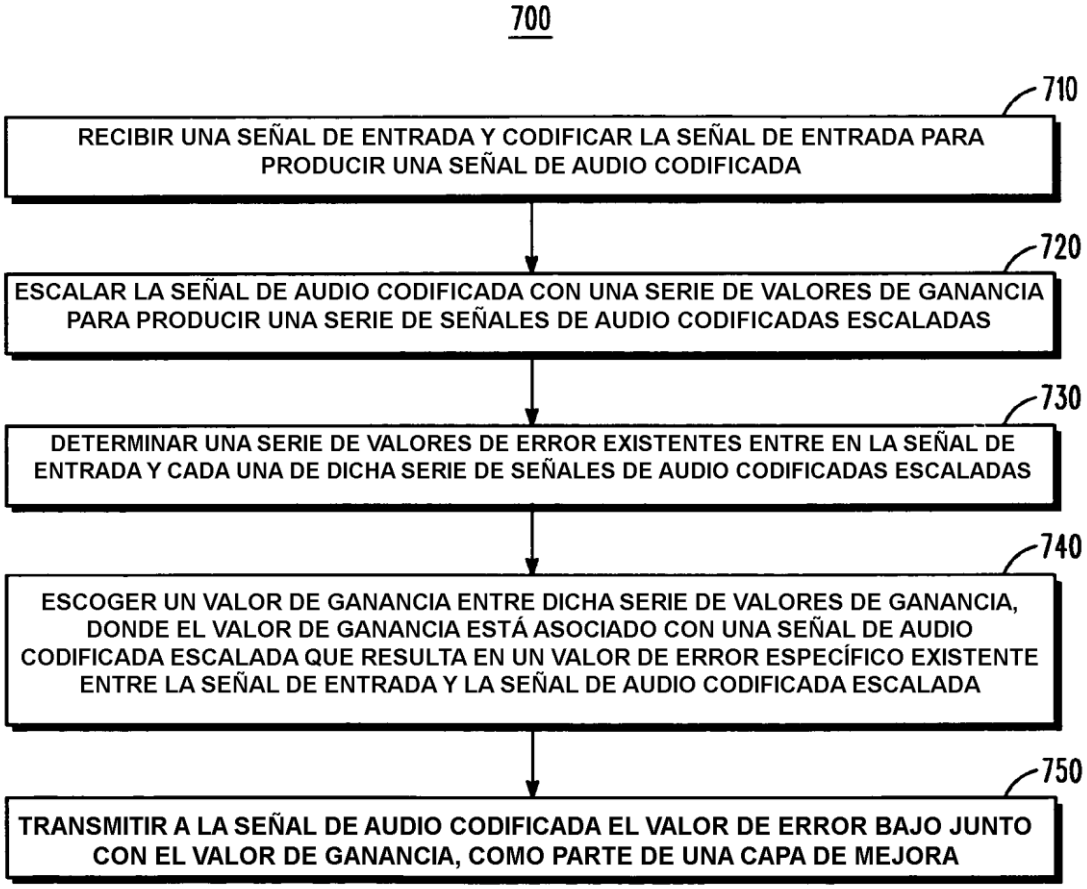

700
710
RECIBIR UNA SEÑAL DE ENTRADA Y CODIFICAR LA SEÑAL DE ENTRADA PARA PRODUCIR UNA SEÑAL DE AUDIO CODIFICADA
720
ESCALAR LA SEÑAL DE AUDIO CODIFICADA CON UNA SERIE DE VALORES DE GANANCIA PARA PRODUCIR UNA SERIE DE SEÑALES DE AUDIO CODIFICADAS ESCALADAS
730
DETERMINAR UNA SERIE DE VALORES DE ERROR EXISTENTES ENTRE EN LA SEÑAL DE ENTRADA Y CADA UNA DE DICHA SERIE DE SEÑALES DE AUDIO CODIFICADAS ESCALADAS
740
ESCOGER UN VALOR DE GANANCIA ENTRE DICHA SERIE DE VALORES DE GANANCIA, DONDE EL VALOR DE GANANCIA ESTÁ ASOCIADO CON UNA SEÑAL DE AUDIO CODIFICADA ESCALADA QUE RESULTA EN UN VALOR DE ERROR ESPECÍFICO EXISTENTE ENTRE LA SEÑAL DE ENTRADA Y LA SEÑAL DE AUDIO CODIFICADA ESCALADA
750
TRANSMITIR A LA SEÑAL DE AUDIO CODIFICADA EL VALOR DE ERROR BAJO JUNTO CON EL VALOR DE GANANCIA, COMO PARTE DE UNA CAPA DE MEJORA

FIG. 7

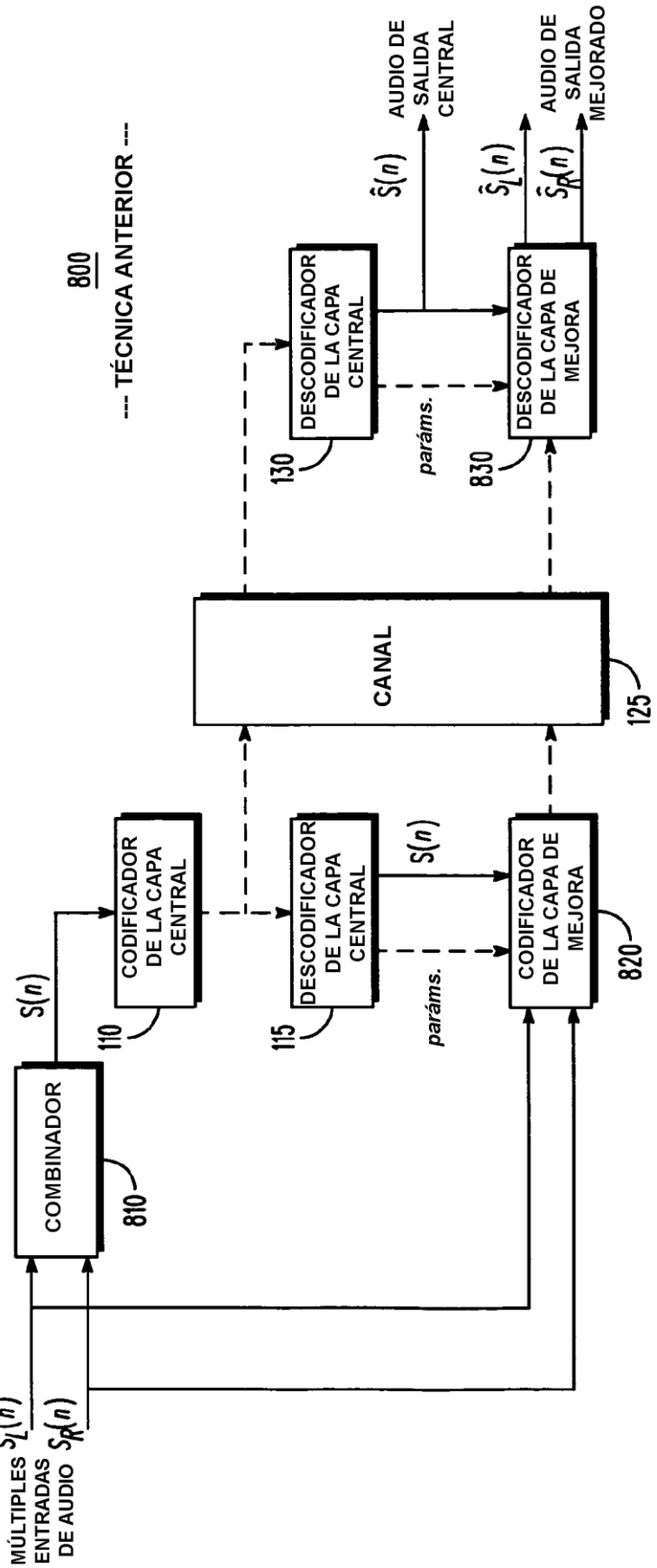

800
--- TÉCNICA ANTERIOR ---
MÚLTIPLES ENTRADAS DE AUDIO
$S_L(n)$
$S_R(n)$
COMBINADOR
810
$S(n)$
CODIFICADOR DE LA CAPA CENTRAL
110
DESCODIFICADOR DE LA CAPA CENTRAL
115
$S(n)$
paráms.
CODIFICADOR DE LA CAPA DE MEJORA
820
CANAL
125
DESCODIFICADOR DE LA CAPA CENTRAL
130
$\hat{S}(n)$
AUDIO DE SALIDA CENTRAL
paráms.
DESCODIFICADOR DE LA CAPA DE MEJORA
830
$\hat{S}_L(n)$
$\hat{S}_R(n)$
AUDIO DE SALIDA MEJORADO

FIG. 8

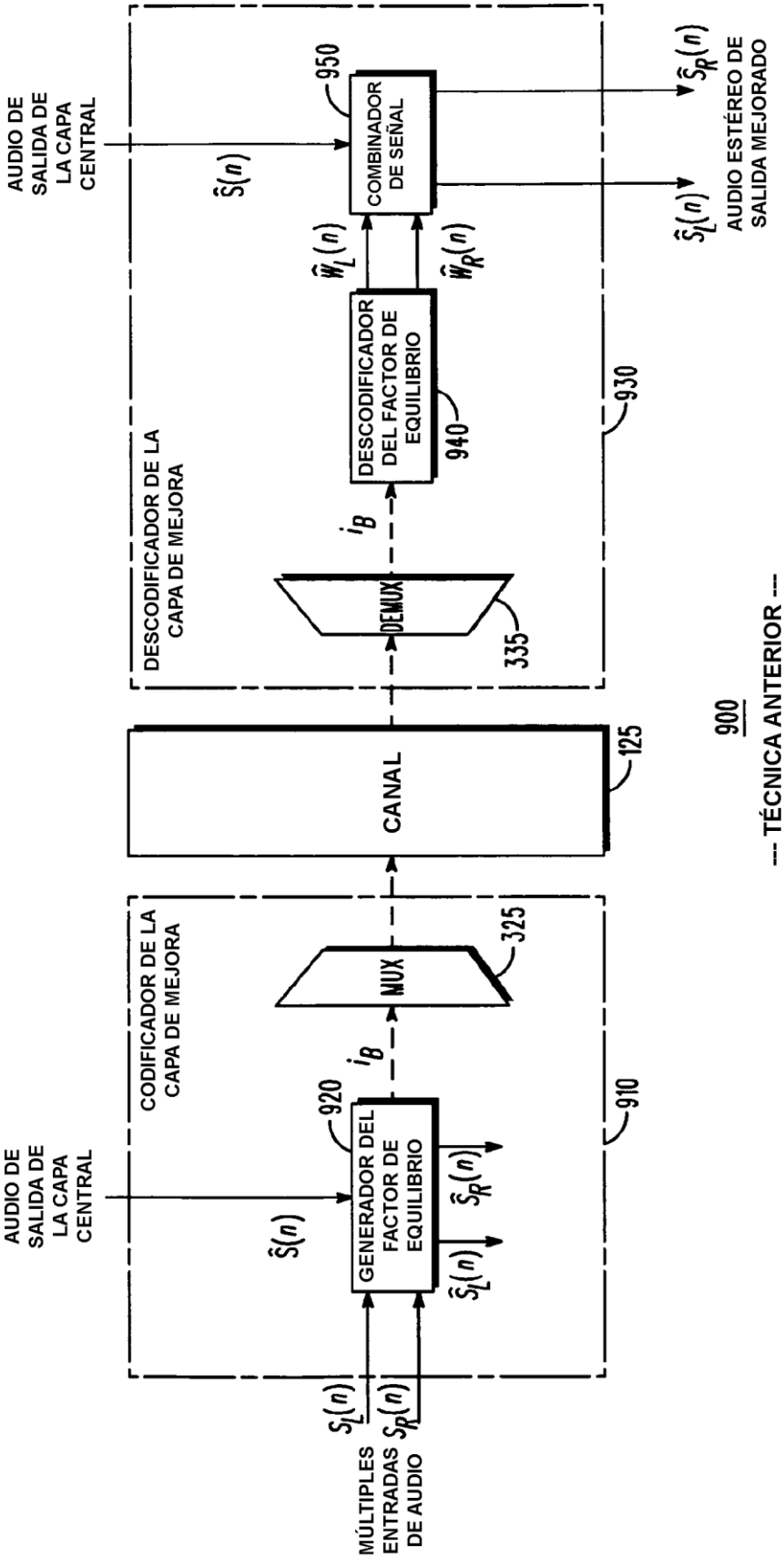

AUDIO DE SALIDA DE LA CAPA CENTRAL
MÚLTIPLES ENTRADAS DE AUDIO
CODIFICADOR DE LA CAPA DE MEJORA
GENERADOR DEL FACTOR DE EQUILIBRIO
920
MUX
325
910
CANAL
125
DESCODIFICADOR DE LA CAPA DE MEJORA
DEMUX
335
DESCODIFICADOR DEL FACTOR DE EQUILIBRIO
940
930
COMBINADOR DE SEÑAL
950
AUDIO ESTÉREO DE SALIDA MEJORADO
900
-- TÉCNICA ANTERIOR --

FIG. 9

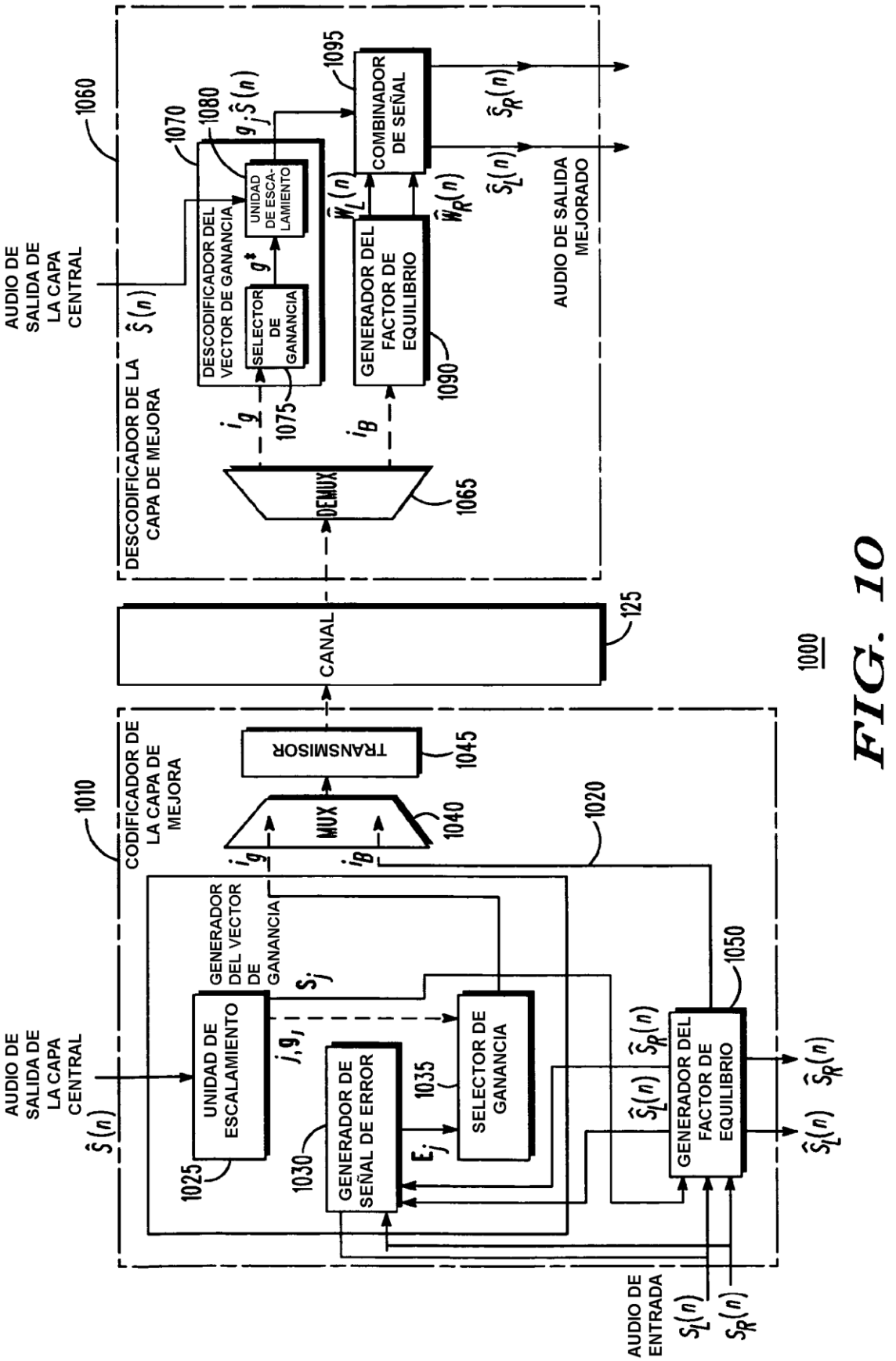

AUDIO DE SALIDA DE LA CAPA CENTRAL
$\hat{S}(n)$
CODIFICADOR DE LA CAPA DE MEJORA
1010
UNIDAD DE ESCALAMIENTO
1025
GENERADOR DEL VECTOR DE GANANCIA
j, g_j
S_j
GENERADOR DE SEÑAL DE ERROR
1030
E_j
1035
SELECTOR DE GANANCIA
$\hat{S}_L(n)$
$\hat{S}_R(n)$
GENERADOR DEL FACTOR DE EQUILIBRIO
1050
1020
MUX
1040
i_g
i_B
TRANSMISOR
1045
CANAL
125
1000
DESCODIFICADOR DE LA CAPA DE MEJORA
1060
DESCODIFICADOR DEL VECTOR DE GANANCIA
1070
SELECTOR DE GANANCIA
g^*
UNIDAD DE ESCALAMIENTO
1080
$g_j \hat{S}(n)$
1075
DEMUX
1065
GENERADOR DEL FACTOR DE EQUILIBRIO
1090
$\tilde{W}_L(n)$
$\tilde{W}_R(n)$
COMBINADOR DE SEÑAL
1095
AUDIO DE SALIDA MEJORADO
AUDIO DE ENTRADA
$S_L(n)$
$S_R(n)$

FIG. 10

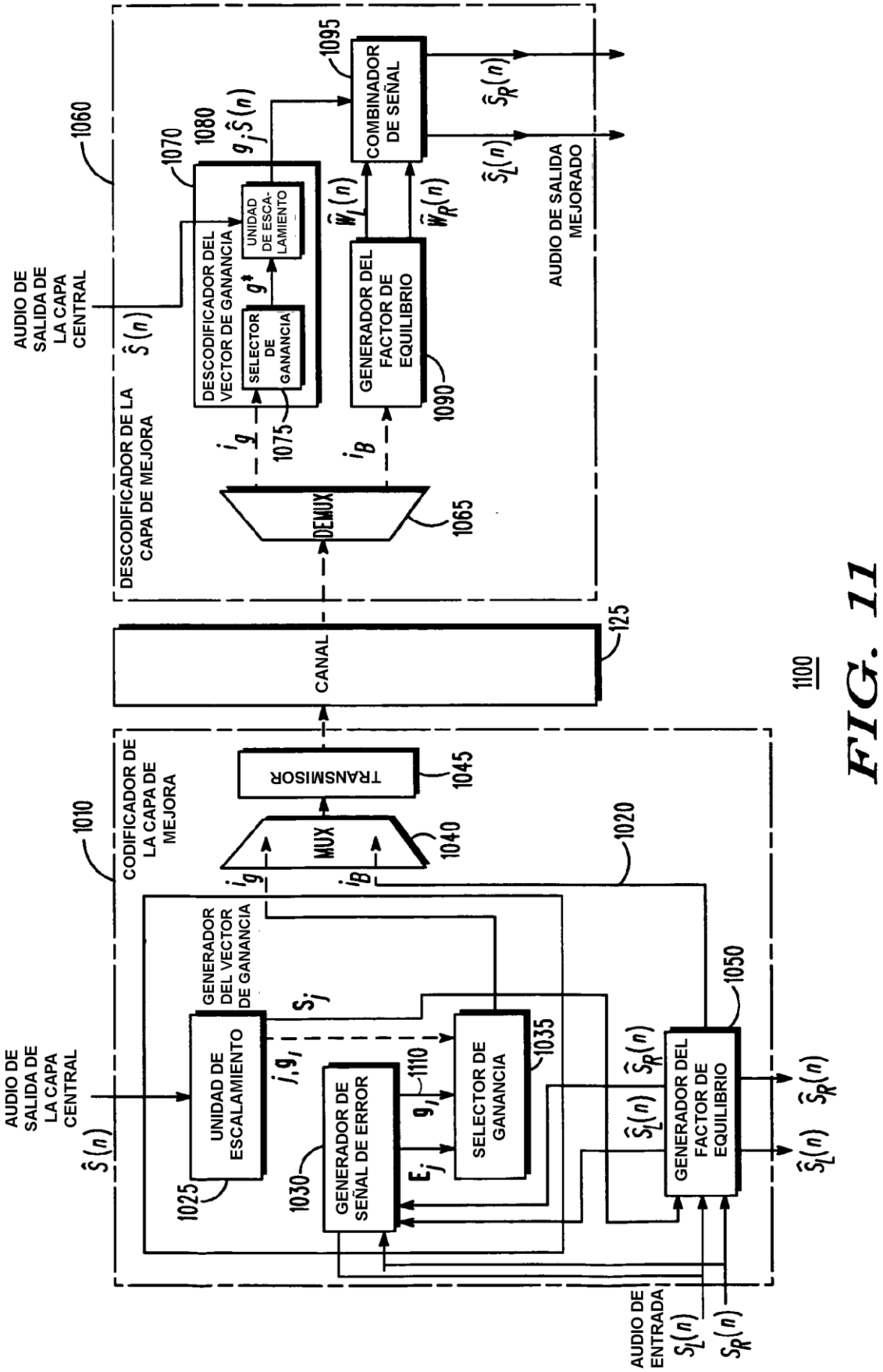

AUDIO DE SALIDA DE LA CAPA CENTRAL
$\hat{S}(n)$
CODIFICADOR DE LA CAPA DE MEJORA
1010
GENERADOR DEL VECTOR DE GANANCIA
UNIDAD DE ESCALAMIENTO
1025
s_j
j, g_j
GENERADOR DE SEÑAL DE ERROR
1030
E_j
1110
g_j
SELECTOR DE GANANCIA
1035
$\hat{S}_L(n)$
$\hat{S}_R(n)$
GENERADOR DEL FACTOR DE EQUILIBRIO
1050
AUDIO DE ENTRADA
$S_L(n)$
$S_R(n)$
MUX
1040
i_g
i_B
1020
TRANSMISOR
1045
CANAL
125
1100
DESCODIFICADOR DE LA CAPA DE MEJORA
1060
DESCODIFICADOR DEL VECTOR DE GANANCIA
1070
UNIDAD DE ESCALAMIENTO
1080
$g_j \hat{S}(n)$
SELECTOR DE GANANCIA
g^*
1075
DEMUX
1065
GENERADOR DEL FACTOR DE EQUILIBRIO
1090
$\tilde{W}_L(n)$
$\tilde{W}_R(n)$
COMBINADOR DE SEÑAL
1095
AUDIO DE SALIDA MEJORADO

FIG. 11


1210
RECIBIR UNA SEÑAL DE AUDIO DE MÚLTIPLES CANALES QUE TIENE UNA SERIE DE SEÑALES DE AUDIO
1220
CODIFICAR LA SEÑAL DE AUDIO DE MÚLTIPLES CANALES PARA GENERAR UNA SEÑAL DE AUDIO CODIFICADA
1230
GENERAR UN FACTOR DE EQUILIBRIO QUE TIENE COMPONENTES DEL FACTOR DE EQUILIBRIO ASOCIADOS CADA UNO CON UNA SEÑAL DE AUDIO DE LA SEÑAL DE AUDIO DE MÚLTIPLES CANALES
1240
DETERMINAR UN VALOR DE GANANCIA A APLICAR A LA SEÑAL DE AUDIO CODIFICADA PARA GENERAR UNA ESTIMACIÓN DE LA SEÑAL DE AUDIO DE MÚLTIPLES CANALES EN BASE AL FACTOR DE EQUILIBRIO Y A LA SEÑAL DE AUDIO DE MÚLTIPLES CANALES, CON EL VALOR DE GANANCIA CONFIGURADO PARA MINIMIZAR UN VALOR DE DISTORSIÓN ENTRE LA SEÑAL DE AUDIO DE MÚLTIPLES CANALES Y LA ESTIMACIÓN DE LA SEÑAL DE AUDIO DE MÑULTIPLES CANALES
1250
ENTREGAR UNA REPRESENTACIÓN DEL VALOR DE GANANCIA PARA TRANSMISIÓN Y/O ALMACENAMIENTO

FIG. 12

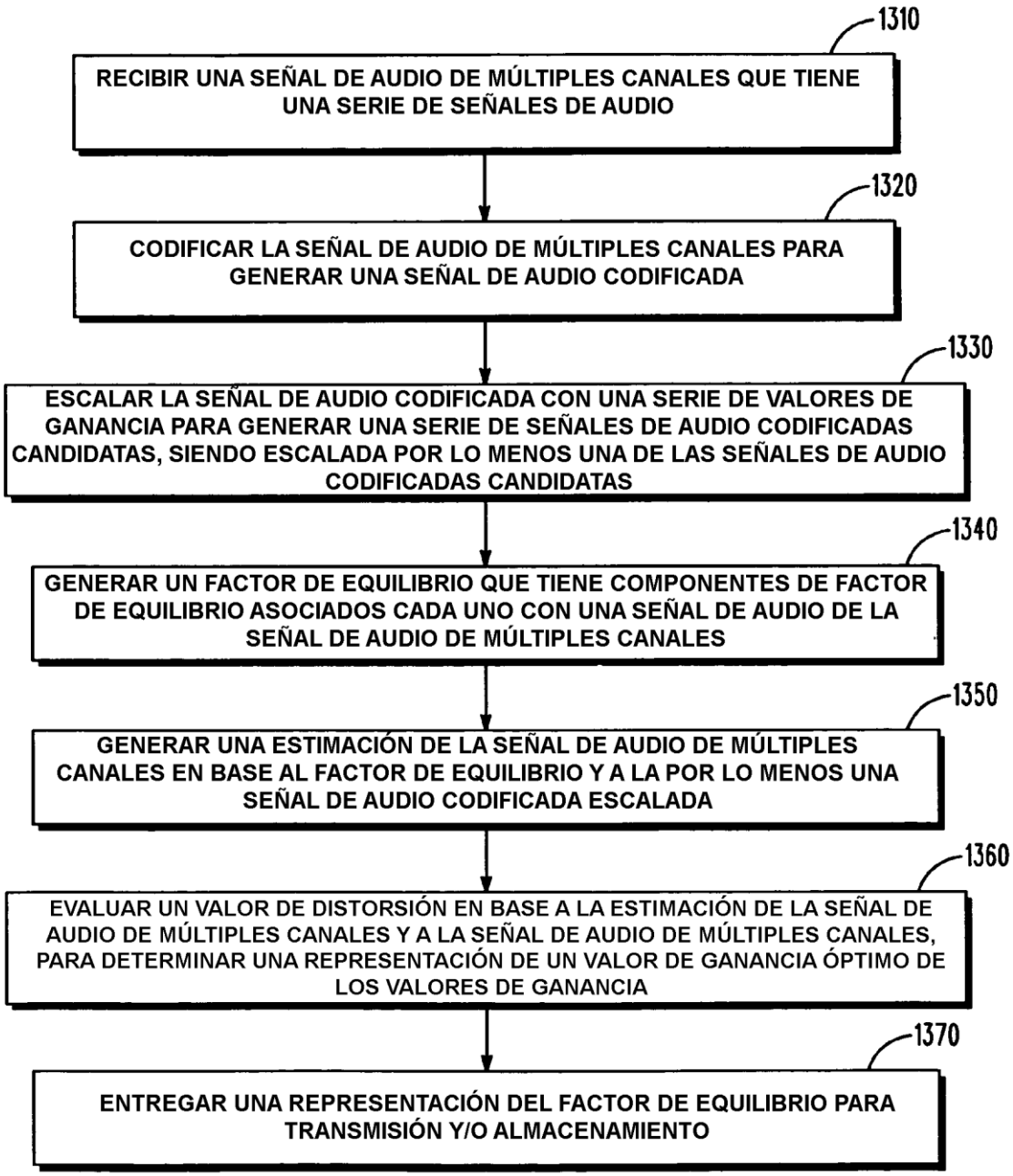

1310
RECIBIR UNA SEÑAL DE AUDIO DE MÚLTIPLES CANALES QUE TIENE UNA SERIE DE SEÑALES DE AUDIO
1320
CODIFICAR LA SEÑAL DE AUDIO DE MÚLTIPLES CANALES PARA GENERAR UNA SEÑAL DE AUDIO CODIFICADA
1330
ESCALAR LA SEÑAL DE AUDIO CODIFICADA CON UNA SERIE DE VALORES DE GANANCIA PARA GENERAR UNA SERIE DE SEÑALES DE AUDIO CODIFICADAS CANDIDATAS, SIENDO ESCALADA POR LO MENOS UNA DE LAS SEÑALES DE AUDIO CODIFICADAS CANDIDATAS
1340
GENERAR UN FACTOR DE EQUILIBRIO QUE TIENE COMPONENTES DE FACTOR DE EQUILIBRIO ASOCIADOS CADA UNO CON UNA SEÑAL DE AUDIO DE LA SEÑAL DE AUDIO DE MÚLTIPLES CANALES
1350
GENERAR UNA ESTIMACIÓN DE LA SEÑAL DE AUDIO DE MÚLTIPLES CANALES EN BASE AL FACTOR DE EQUILIBRIO Y A LA POR LO MENOS UNA SEÑAL DE AUDIO CODIFICADA ESCALADA
1360
EVALUAR UN VALOR DE DISTORSIÓN EN BASE A LA ESTIMACIÓN DE LA SEÑAL DE AUDIO DE MÚLTIPLES CANALES Y A LA SEÑAL DE AUDIO DE MÚLTIPLES CANALES, PARA DETERMINAR UNA REPRESENTACIÓN DE UN VALOR DE GANANCIA ÓPTIMO DE LOS VALORES DE GANANCIA
1370
ENTREGAR UNA REPRESENTACIÓN DEL FACTOR DE EQUILIBRIO PARA TRANSMISIÓN Y/O ALMACENAMIENTO
1300

FIG. 13

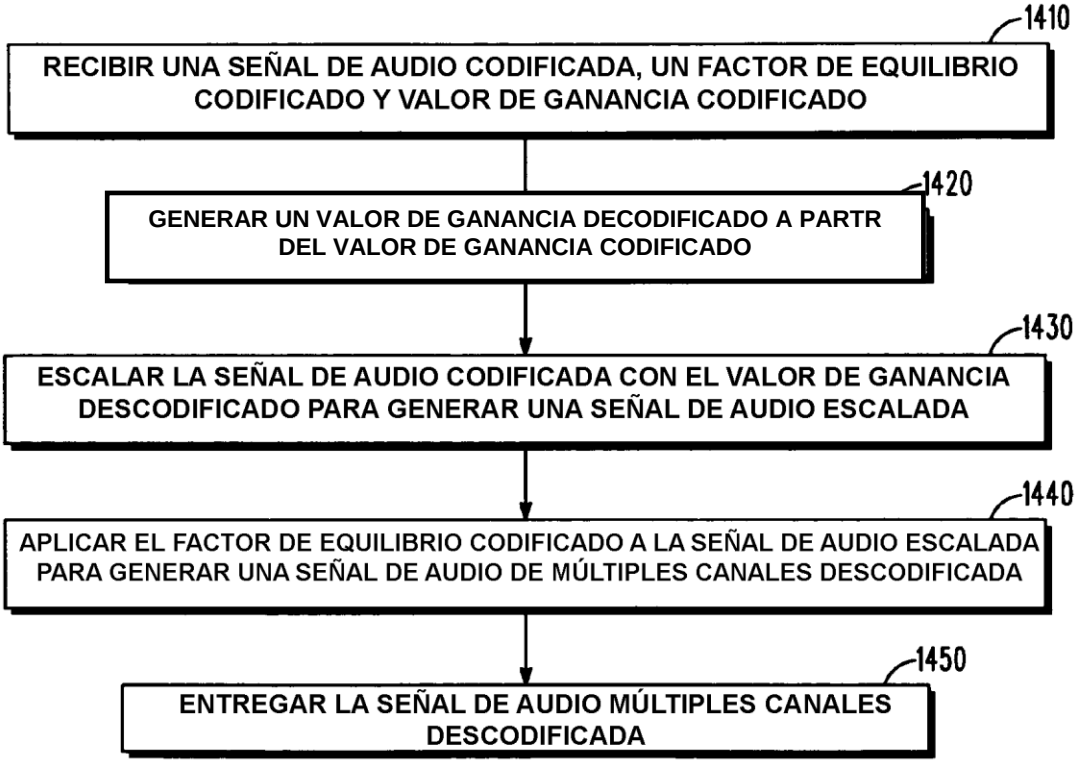

1410
RECIBIR UNA SEÑAL DE AUDIO CODIFICADA, UN FACTOR DE EQUILIBRIO CODIFICADO Y VALOR DE GANANCIA CODIFICADO
1420
GENERAR UN VALOR DE GANANCIA DECODIFICADO A PARTR DEL VALOR DE GANANCIA CODIFICADO
1430
ESCALAR LA SEÑAL DE AUDIO CODIFICADA CON EL VALOR DE GANANCIA DESCODIFICADO PARA GENERAR UNA SEÑAL DE AUDIO ESCALADA
1440
APLICAR EL FACTOR DE EQUILIBRIO CODIFICADO A LA SEÑAL DE AUDIO ESCALADA PARA GENERAR UNA SEÑAL DE AUDIO DE MÚLTIPLES CANALES DESCODIFICADA
1450
ENTREGAR LA SEÑAL DE AUDIO MÚLTIPLES CANALES DESCODIFICADA
1400

FIG. 14

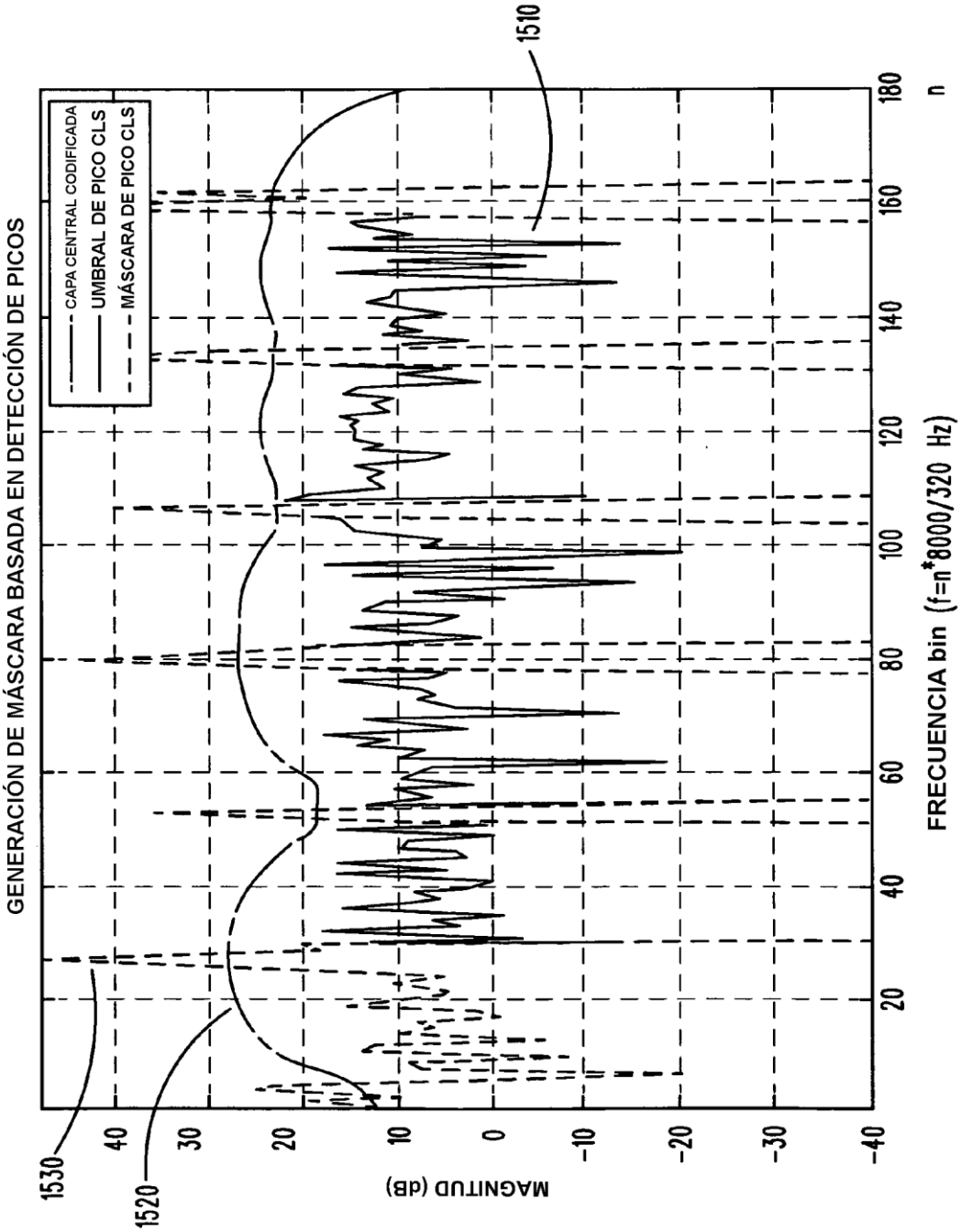

GENERACIÓN DE MÁSCARA BASADA EN DETECCIÓN DE PICOS
CAPA CENTRAL CODIFICADA
UMBRAL DE PICO CLS
MÁSCARA DE PICO CLS
1510
1520
1530
MAGNITUD (dB)
FRECUENCIA bin (f=n*8000/320 Hz)
n

FIG. 15

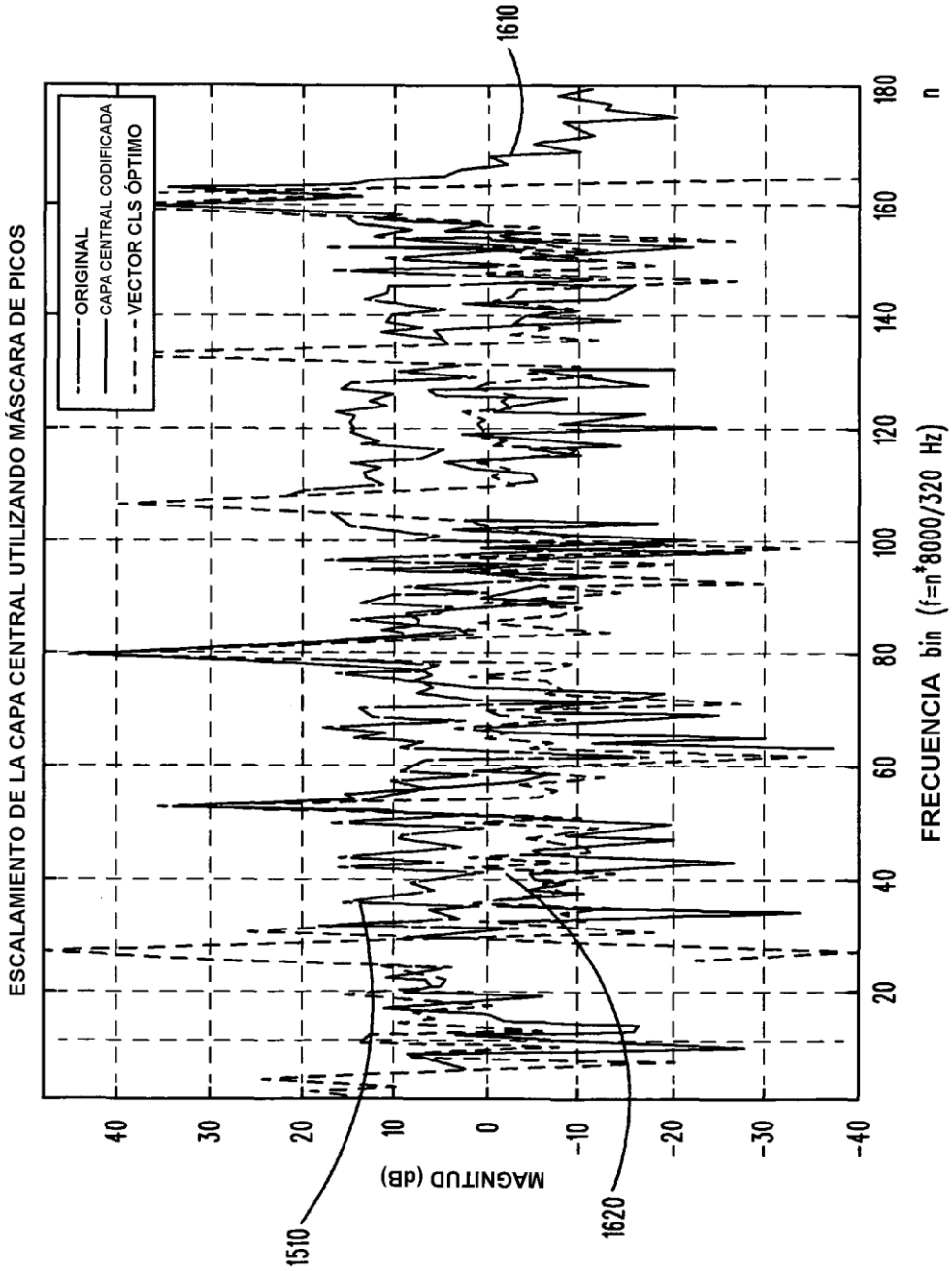

ESCALAMIENTO DE LA CAPA CENTRAL UTILIZANDO MÁSCARA DE PICOS
ORIGINAL
CAPA CENTRAL CODIFICADA
VECTOR CLS ÓPTIMO
MAGNITUD (dB)
FRECUENCIA bin (f=n*8000/320 Hz)
n
1610
1510
1620

FIG. 16

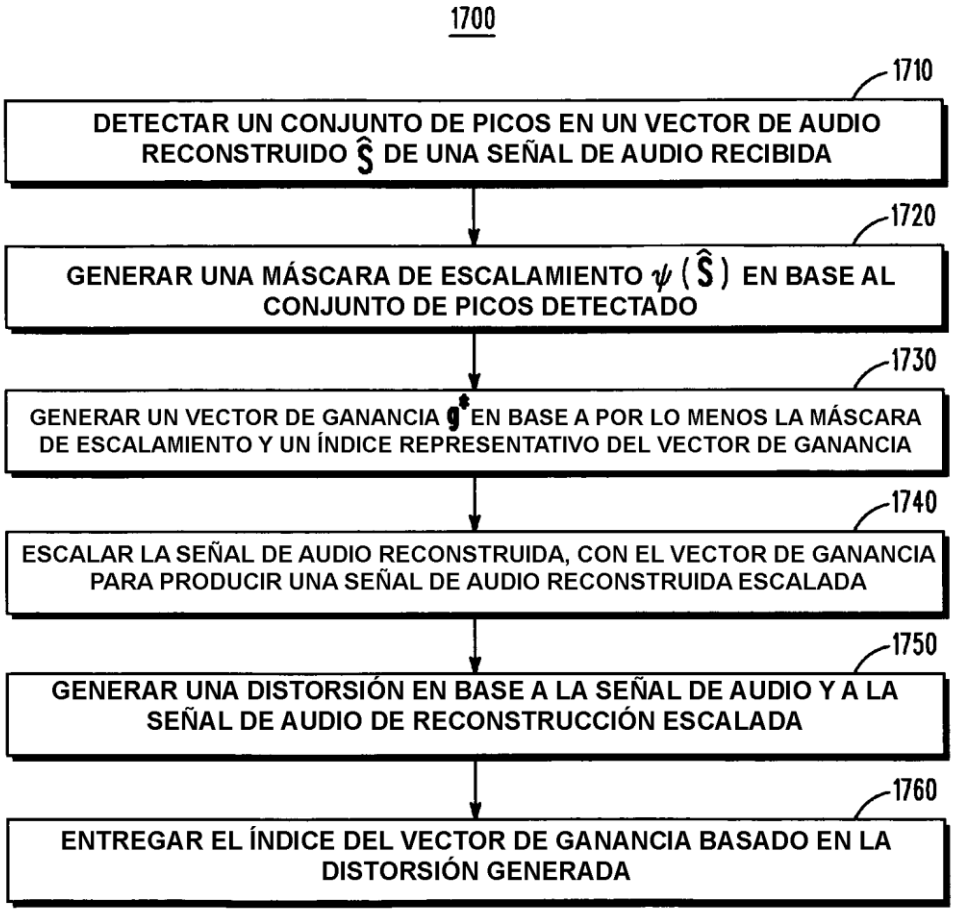

1700
1710
DETECTAR UN CONJUNTO DE PICOS EN UN VECTOR DE AUDIO RECONSTRUIDO $\hat{S}$ DE UNA SEÑAL DE AUDIO RECIBIDA
1720
GENERAR UNA MÁSCARA DE ESCALAMIENTO $\psi(\hat{S})$ EN BASE AL CONJUNTO DE PICOS DETECTADO
1730
GENERAR UN VECTOR DE GANANCIA g_c EN BASE A POR LO MENOS LA MÁSCARA DE ESCALAMIENTO Y UN ÍNDICE REPRESENTATIVO DEL VECTOR DE GANANCIA
1740
ESCALAR LA SEÑAL DE AUDIO RECONSTRUIDA, CON EL VECTOR DE GANANCIA PARA PRODUCIR UNA SEÑAL DE AUDIO RECONSTRUIDA ESCALADA
1750
GENERAR UNA DISTORSIÓN EN BASE A LA SEÑAL DE AUDIO Y A LA SEÑAL DE AUDIO DE RECONSTRUCCIÓN ESCALADA
1760
ENTREGAR EL ÍNDICE DEL VECTOR DE GANANCIA BASADO EN LA DISTORSIÓN GENERADA

FIG. 17

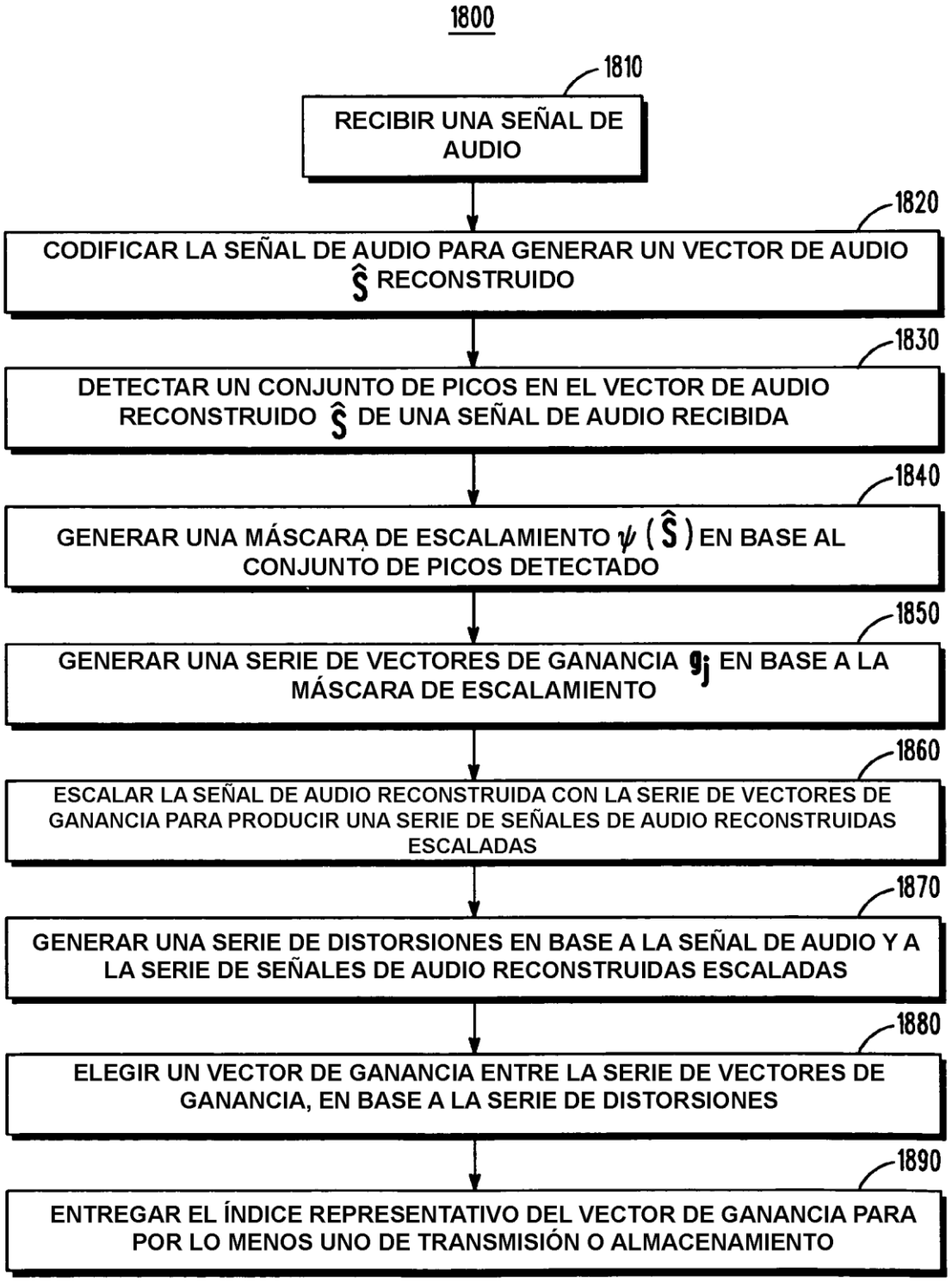

1800
1810
RECIBIR UNA SEÑAL DE AUDIO
1820
CODIFICAR LA SEÑAL DE AUDIO PARA GENERAR UN VECTOR DE AUDIO $\hat{S}$ RECONSTRUIDO
1830
DETECTAR UN CONJUNTO DE PICOS EN EL VECTOR DE AUDIO RECONSTRUIDO $\hat{S}$ DE UNA SEÑAL DE AUDIO RECIBIDA
1840
GENERAR UNA MÁSCARA DE ESCALAMIENTO $\psi(\hat{S})$ EN BASE AL CONJUNTO DE PICOS DETECTADO
1850
GENERAR UNA SERIE DE VECTORES DE GANANCIA g_j EN BASE A LA MÁSCARA DE ESCALAMIENTO
1860
ESCALAR LA SEÑAL DE AUDIO RECONSTRUIDA CON LA SERIE DE VECTORES DE GANANCIA PARA PRODUCIR UNA SERIE DE SEÑALES DE AUDIO RECONSTRUIDAS ESCALADAS
1870
GENERAR UNA SERIE DE DISTORSIONES EN BASE A LA SEÑAL DE AUDIO Y A LA SERIE DE SEÑALES DE AUDIO RECONSTRUIDAS ESCALADAS
1880
ELEGIR UN VECTOR DE GANANCIA ENTRE LA SERIE DE VECTORES DE GANANCIA, EN BASE A LA SERIE DE DISTORSIONES
1890
ENTREGAR EL ÍNDICE REPRESENTATIVO DEL VECTOR DE GANANCIA PARA POR LO MENOS UNO DE TRANSMISIÓN O ALMACENAMIENTO

FIG. 18

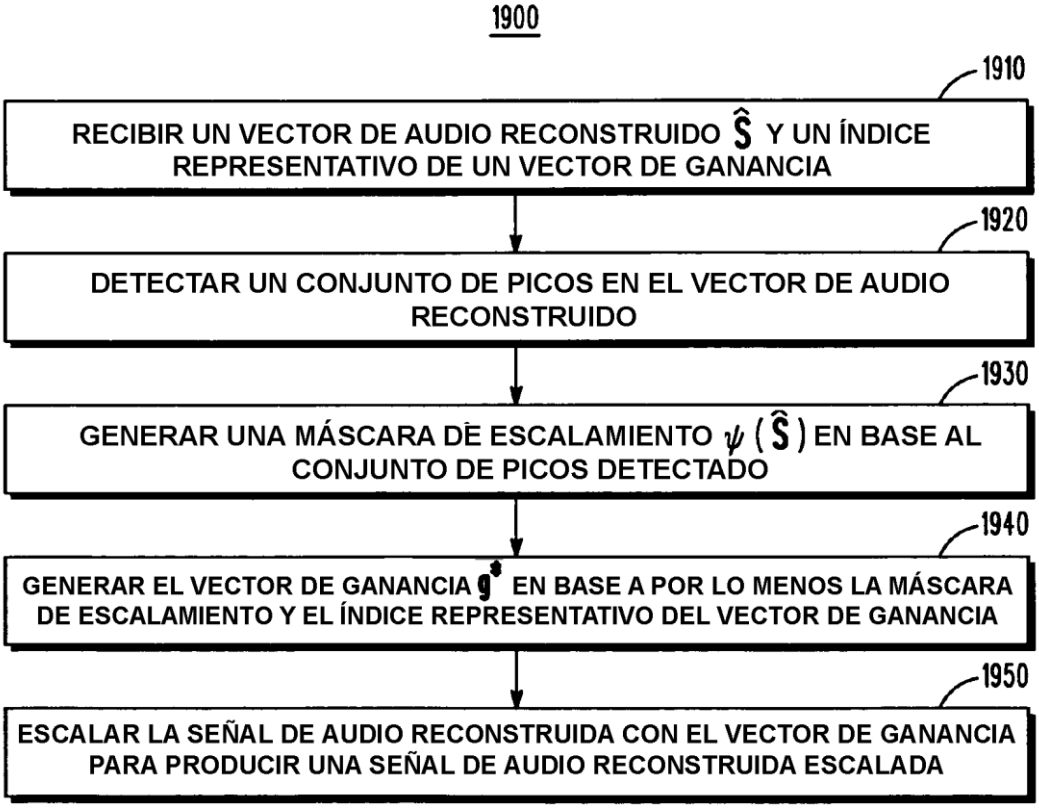

1900
1910
RECIBIR UN VECTOR DE AUDIO RECONSTRUIDO $\hat{s}$ Y UN ÍNDICE REPRESENTATIVO DE UN VECTOR DE GANANCIA
1920
DETECTAR UN CONJUNTO DE PICOS EN EL VECTOR DE AUDIO RECONSTRUIDO
1930
GENERAR UNA MÁSCARA DE ESCALAMIENTO $\psi(\hat{s})$ EN BASE AL CONJUNTO DE PICOS DETECTADO
1940
GENERAR EL VECTOR DE GANANCIA g EN BASE A POR LO MENOS LA MÁSCARA DE ESCALAMIENTO Y EL ÍNDICE REPRESENTATIVO DEL VECTOR DE GANANCIA
1950
ESCALAR LA SEÑAL DE AUDIO RECONSTRUIDA CON EL VECTOR DE GANANCIA PARA PRODUCIR UNA SEÑAL DE AUDIO RECONSTRUIDA ESCALADA

FIG. 19