(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 434 251**

(51) Int. CI.:

G10L 19/008 (2013.01)
G10L 19/24 (2013.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **03.12.2009 E 09799783 (7)**

(97) Fecha y número de publicación de la concesión europea: **25.09.2013 EP 2382622**

(54) Título: **Método y aparato para generar una capa de mejora dentro de un sistema de codificación de audio de múltiples canales**

(30) Prioridad:

29.12.2008 US 345117

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
16.12.2013

(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

Método y aparato para generar una capa de mejora dentro de un sistema de codificación de audio de múltiples canales

Referencia a solicitudes relacionadas

La presente solicitud está relacionada con las siguientes solicitudes de patente europea de propiedad conjunta con esta solicitud 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 2 382 627 A0, titulada “SELECTIVE SCALING MASK COMPUTATION BASED ON PEAK DETECTION”; y

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

Campo de la divulgación

La presente divulgación versa, en general, acerca de sistemas de comunicaciones y, más en particular, acerca de la codificación de señales de voz y audio en tales sistemas de comunicaciones.

Antecedentes

La compresión de señales digitales de voz y audio es bien conocida. Generalmente, se requiere la compresión para transmitir eficientemente señales por un canal de comunicaciones o para almacenar señales comprimidas en un dispositivo de medios digitales, tal como un dispositivo de memoria de estado sólido o un disco duro de ordenador. Aunque hay muchas técnicas de compresión (o “codificación”), un método que ha seguido siendo muy popular para la codificación de voz digital se denomina predicción lineal con excitación por código (CELP), que es uno de una familia de algoritmos de codificación de “análisis por síntesis”. Generalmente, análisis por síntesis se refiere a un procedimiento de codificación por medio del cual se usan múltiples parámetros de un modelo digital para sintetizar un conjunto de señales candidatas que se comparan con una señal de entrada y se analizan en busca de distorsiones. Entonces se transmite o se almacena un conjunto de parámetros que producen la menor distorsión, que acaba usándose para reconstruir una estimación de la señal original de entrada. La CELP es un método particular de análisis por síntesis que usa uno o más 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 modernos existe el problema de mantener una reproducción de voz y audio de alta calidad con velocidades de transferencia de datos razonablemente bajas. Esto se aplica en especial a la música u otras señales genéricas de audio que no encajan muy bien en el modelo de voz de CELP. En este caso, la discordancia del modelo puede causar una calidad de audio muy degradada que puede resultar inaceptable para un usuario final del equipo que emplee tales métodos. Por lo tanto, subsiste la necesidad de mejorar el rendimiento de los codificadores de voz de tipo CELP con velocidades bajas de transferencia de bits, especialmente para música y otras entradas de tipo distinto a la voz.

El documento EP 1 818 911 (A1) da a conocer un dispositivo de codificación del sonido que tiene una estructura monoaural/estereofónica de escala variable y capaz de codificar eficientemente sonido estereofónico cuando la correlación entre las señales de los canales de una señal estereofónica es pequeña.

Los objetivos anteriores son resueltos por las reivindicaciones de la presente invención.

Breve descripción de los dibujos

Las figuras adjuntas, en las que números de referencia semejantes se refieren a elementos idénticos o funcionalmente similares en todas las distintas vistas, que, junto con la descripción detallada que sigue, están incorporadas en la memoria y forman parte de la misma, sirven para ilustrar adicionalmente diversas realizaciones de conceptos que incluyen la invención reivindicada y para explicar diversos principios y ventajas de esas realizaciones.

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

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

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

La FIG. 4 es un diagrama de bloques de un codificador y un decodificador de la capa de mejora.

La FIG. 5 es un diagrama de bloques de un sistema integrado de codificación de múltiples capas.

La FIG. 6 es un diagrama de bloques de un codificador y un decodificador de la capa 4.

La FIG. 7 es un diagrama de flujo que muestra la operación de los codificadores de la FIG. 4 y la FIG. 6.

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

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

La FIG. 10 es un diagrama de bloques de un codificador y un decodificador de la capa de mejora según diversas realizaciones.

La FIG. 11 es un diagrama de bloques de un codificador y un decodificador de la capa de mejora según diversas realizaciones.

La FIG. 12 es un diagrama de flujo de la codificación de una señal de audio de múltiples canales según diversas realizaciones.

La FIG. 13 es un diagrama de flujo de la codificación de una señal de audio de múltiples canales según diversas realizaciones.

La FIG. 14 es un diagrama de flujo de la decodificación de una señal de audio de múltiples canales según diversas realizaciones.

La FIG. 15 es un gráfico de frecuencias de generación de máscara basada en la detección de picos según diversas realizaciones.

La FIG. 16 es un gráfico de frecuencias del cambio de escala de la capa de núcleo usando una generación de máscara de picos según diversas realizaciones.

Las FIGURAS 17-19 son diagramas de flujo que ilustran la metodología para codificar y decodificar usando la generación de máscara basada en la detección de picos según diversas realizaciones.

Los expertos en la técnica apreciarán que algunos elementos de las figuras se ilustran en aras de la simplicidad y la claridad y que no han sido dibujados necesariamente a escala. Por ejemplo, las dimensiones de algunos de los elementos de las figuras pueden estar exageradas con respecto a otros elementos para contribuir a mejorar la comprensión de diversas realizaciones. Además, la descripción y los dibujos no requieren necesariamente el orden ilustrado. Se apreciará, además, que ciertas acciones y/o etapas pueden ser descritas o representadas en un orden particular de incidencia, mientras que los expertos en la técnica entenderán que, en realidad, no se requiere tal especificidad con respecto a la secuencia. Cuando ha sido apropiado, se han representado componentes de aparatos y de métodos mediante símbolos convencionales en los dibujos, mostrando únicamente aquellos detalles específicos que son pertinentes a la comprensión de las diversas realizaciones para no ofuscar la divulgación con detalles que serán inmediatamente evidentes a las personas con un dominio normal de la técnica que cuenten con el beneficio de la descripción del presente documento. Así, se apreciará que, en aras de la simplicidad y la claridad de la ilustración, pueden no estar representados elementos comunes y bien entendidos que son útiles o necesarios en realizaciones comercialmente viables para facilitar una visión menos obstaculizada de estas diversas realizaciones.

Descripción detallada

Para abordar la necesidad mencionada en lo que antecede, en el presente documento se describen un método y un aparato para generar una capa de mejora dentro de un sistema de codificación de audio. Durante la operación, se recibe y se codifica una señal de entrada que ha de ser codificada para producir una señal de audio codificada. A continuación, se cambia la escala de la señal de audio codificada con varios valores de ganancia para producir varias señales de audio codificadas a escala, cada una de las cuales tiene un valor de ganancia asociado, y se determinan varios valores de error que existen entre la señal de entrada y cada una de las varias señales de audio codificadas a escala. Acto seguido, se escoge un valor de ganancia que está asociado con una señal de audio codificada a escala que da como resultado que exista un valor bajo de error entre la señal de entrada y la señal de audio codificada a escala. Por último, se transmite el valor bajo de error junto con el valor de ganancia como parte de una capa de mejora a la señal de audio codificada.

En la FIG. 1 se muestra un sistema de compresión integrado de voz/audio de la técnica anterior. El audio $s(n)$ de entrada es tratado en primer lugar por un codificador 110 de la capa de núcleo que, para estos fines, puede ser un algoritmo de codificación de voz de tipo CELP. La corriente codificada de bits es transmitida al canal 125, además de ser introducida en un decodificador local 115 de la capa de núcleo, en el que se genera la señal reconstruida $s_c(n)$ de audio de núcleo. A continuación, se usa el codificador 120 de la capa de mejora para codificar información adicional en función de cierta comparación de las señales $s(n)$ y $s_c(n)$, y este puede usar opcionalmente parámetros del decodificador 115 de la capa de núcleo. Como en el decodificador 115 de la capa de núcleo, el decodificador 130 de la capa de núcleo convierte los parámetros de la corriente de bits de la capa de núcleo en una señal $\hat{s}_c(n)$ de

audio de la capa de núcleo. Acto seguido, el decodificador 135 de la capa de mejora usa la corriente de bits de la capa de mejora procedente del canal 125 y la señal $\hat{s}_c(n)$ para producir la señal $\hat{s}(n)$ de salida mejorada de audio.

La ventaja fundamental de tal sistema integrado de codificación es que un canal particular 125 puede no ser capaz de soportar coherentemente el requisito de ancho de banda asociado con los algoritmos de codificación de audio de alta calidad. Un codificador integrado, sin embargo, permite que se reciba una corriente parcial de bits (por ejemplo, solo la corriente de bits de la capa de núcleo) desde el canal 125 para producir, por ejemplo, solo el audio de salida de núcleo cuando se pierde o se interrumpe la corriente de bits de la capa de mejora. Sin embargo, hay compromisos en calidad entre los codificadores integrados y los no integrados, y también entre diferentes objetivos de optimización de codificación integrada. Es decir, una codificación de la capa de mejora de mayor calidad puede contribuir a lograr un mejor equilibrio entre las capas de núcleo y de mejora, y también a reducir la velocidad global de transferencia de datos para mejores características de transmisión (por ejemplo, congestión reducida), lo que puede dar como resultado menores tasas de errores de paquete para las capas de mejora.

En la FIG. 2 se da un ejemplo más detallado del codificador 120 de la capa de mejora de la técnica anterior. Aquí, el generador 210 de señales de error comprende una señal ponderada de diferencias que se transforma en el dominio de la transformada discreta del coseno modificada (TDCM) para su tratamiento por el codificador 220 de señales de error. La señal E de error está dada como:

$$\mathbf{E} = \mathrm{TDCM}\left\{\mathbf{W}\left(\mathbf{s} - \mathbf{s}_c\right)\right\}, \tag{1}$$

siendo W una matriz de ponderación perceptual basada en los coeficientes $A(z)$ de filtro de predicción lineal (PL) procedentes del decodificador 115 de la capa de núcleo, siendo s un vector (es decir, una *trama*) de muestras procedentes de la señal $s(n)$ de audio de entrada y siendo s_c el correspondiente vector de muestras procedentes del decodificador 115 de la capa de núcleo. En la Recomendación G.729.1 de ITU-T se describe un procedimiento ejemplar de TDCM. A continuación, el codificador 220 de señales de error procesa la señal E de error para producir la palabra de código i_E, que es transmitida subsiguientemente al canal 125. Para este ejemplo, es importante hacer notar que al codificador 220 de señales de error se le presenta únicamente una señal E de error y que produce una sola palabra asociada de código i_E. La razón de esto se hará evidente más tarde.

El decodificador 135 de la capa de mejora recibe entonces del canal 125 la corriente cifrada de bits y demultiplexa debidamente la corriente de bits para producir la palabra de código i_E. El decodificador 230 de señales de error usa la palabra de código i_E para reconstruir la señal $\hat{\mathbf{E}}$ de errores de la capa de mejora, que luego es combinada como sigue por el combinador 240 de señales con la señal $\hat{s}_c(n)$ de audio de salida de la capa de núcleo para producir la señal $\hat{s}(n)$ de salida mejorada de audio:

$$\hat{\mathbf{s}} = \mathbf{s}_c + \mathbf{W}^{-1}\mathrm{TDCM}^{-1}\left\{\hat{\mathbf{E}}\right\}, \tag{2}$$

siendo TDCM^{-1} la TDCM inversa (incluyendo solapamiento-adición) y siendo W^{-1} la matriz inversa de ponderación perceptual.

En la FIG. 3 se muestra otro ejemplo de un codificador de la capa de mejora. Aquí, la generación de la señal E de error por el generador 315 de señales de error implica un cambio previo adaptativo de escala en el que se lleva a cabo cierta modificación a la salida $s_c(n)$ de audio de la capa de núcleo. Este procedimiento da como resultado que se generen algunos bits que se muestran en el codificador 120 de la capa de mejora como la palabra de código i_s.

Además, el codificador 120 de la capa de mejora muestra la señal $s(n)$ de audio de entrada y el audio transformado S_c de salida de la capa de núcleo que se introduce en el codificador 320 de señales de error. Se usan estas señales para construir un modelo psicoacústico para una codificación mejorada de la señal E de error de la capa de mejora. A continuación, las palabras de código i_s e i_E son multiplexadas por el MUX 325, y luego enviadas al canal 125 para su decodificación subsiguiente por el decodificador 135 de la capa de mejora. La corriente codificada de bits es recibida por el demultiplexor 335, que separa la corriente de bits en componentes i_s e i_E. A continuación, el decodificador 340 de señales de error usa la palabra de código i_E para reconstruir la señal $\hat{\mathbf{E}}$ de errores de la capa de mejora. El combinador 345 de señales cambia la escala de la señal $\hat{s}_c(n)$ de alguna manera usando los bits i_s de escala y luego combina el resultado con la señal $\hat{\mathbf{E}}$ de errores de la capa de mejora para producir la señal $\hat{s}(n)$ de salida mejorada de audio.

La FIG. 4 da una primera realización de la presente invención. Esta figura muestra al codificador 410 de la capa de mejora que recibe la señal $s_c(n)$ de salida de la capa de núcleo por medio de la unidad 415 de escala. Se usa un conjunto predeterminado de ganancias {g} para producir varias señales {S} de salida de la capa de núcleo a escala, siendo g_j y S_j las candidatas j-ésimas de los respectivos conjuntos. Dentro de la unidad 415 de escala, la primera realización procesa la señal $s_c(n)$ en el dominio (TDCM) como:

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

pudiendo ser W alguna matriz de ponderación perceptual, siendo s_c un vector de muestras procedentes del decodificador 115 de la capa de núcleo, siendo la TDCM una operación bien conocida en la técnica y pudiendo ser G_j una matriz de ganancia formada utilizando un vector g_j de ganancia candidato, y siendo M el número de vectores de ganancia candidatos. En la primera realización, G_j usa el vector g_j como la diagonal y ceros en todo lo demás (es decir, una matriz diagonal), aunque existen muchas posibilidades. Por ejemplo, G_j puede ser una matriz de bandas, o puede ser incluso una simple cantidad escalar multiplicada por la matriz de identidad I. Alternativamente, puede haber cierta ventaja en dejar a la señal S_j en el dominio temporal, o puede haber casos en los que sea ventajoso transformar el audio a un dominio diferente, tal como el dominio de una transformada de Fourier discreta (DFT). Son bien conocidas en la técnica muchas transformadas de este tipo. En estos casos, la unidad de escala puede producir la S_j apropiada en función del respectivo dominio vectorial.

Pero, en cualquier caso, la razón fundamental para cambiar la escala del audio de salida de la capa de núcleo es compensar la discordancia del modelo (o alguna otra deficiencia de codificación), que puede causar diferencias significativas entre la señal de entrada y el códec de la capa de núcleo. Por ejemplo, si la señal de audio de entrada es fundamentalmente una señal de música y el códec de la capa de núcleo se basa en un modelo de voz, entonces la salida de la capa de núcleo puede contener características de señal muy distorsionadas, en cuyo caso resulta beneficioso, desde una perspectiva de la calidad del sonido, reducir selectivamente la energía de esta componente de la señal antes de aplicar codificación suplementaria de la señal por medio de una o más capas de mejora.

El vector candidato S_j de audio de la capa de núcleo con ganancia a escala y el audio $s(n)$ de entrada pueden ser usados entonces como entrada al generador 420 de señales de error. En una realización ejemplar, la señal $s(n)$ de audio de entrada se convierte en el vector S, de modo que S y S_j estén correspondientemente alineados. Es decir, el vector s, que representa a $s(n)$, está alineado en el tiempo (fase) con s_c, y pueden aplicarse las operaciones correspondientes, de modo que en esta realización

$$\mathbf{E}_j = \mathrm{TDCM}\{\mathbf{Ws}\} - \mathbf{S}_j; \quad 0 \le j < M, \tag{4}$$

Esta expresión produce varios vectores E_j de señales de error que representan la diferencia ponderada entre el audio de entrada y el audio de salida de la capa de núcleo con ganancia a escala en el dominio espectral de la TDCM. En otras realizaciones en las que se consideran dominios diferentes, la anterior expresión puede ser modificada en función del respectivo dominio de procesamiento.

A continuación, se usa el selector 425 de ganancia para evaluar los varios vectores E_j de señales de error según la primera realización de la presente invención para producir un vector óptimo E* de errores, un parámetro óptimo g* de ganancia y, subsiguientemente, un correspondiente índice i_g de ganancia. El selector 425 de ganancia puede usar varios métodos para determinar los parámetros óptimos, E* y g*, lo que puede implicar métodos de bucle cerrado (por ejemplo, minimización de una métrica de distorsión), métodos de bucle abierto (por ejemplo, clasificación heurística, estimación de rendimiento del modelo, etc.) o una combinación de ambos métodos. En la realización ejemplar puede usarse una métrica sesgada de distorsión, que se da como la diferencia sesgada de energía entre el vector original S de señales de audio y el vector de la señal reconstruida compuesta:

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

en donde $\hat{E}_j$ puede ser la estimación cuantificada del vector E_j de señales de error y β_j puede ser un término de sesgo que se usa para complementar la decisión de escoger el índice j^* de error de ganancia perceptualmente óptimo. La solicitud de patente estadounidense, con nº de serie 11/531122, titulada APPARATUS AND METHOD FOR LOW COMPLEXITY COMBINATORIAL CODING OF SIGNALS, da un método ejemplar para la cuantificación vectorial de un vector de señales, aunque son posibles muchos otros métodos. Reconociendo que $\mathbf{E}_j = \mathbf{S} - \mathbf{S}j$, la Ecuación (5) puede reescribirse como:

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

En esta expresión, el término $\varepsilon_j = \square\, \mathbf{E}_j - \hat{\mathbf{E}}_j \,\square^2$ representa la energía de la diferencia entre las señales de error no cuantificadas y las cuantificadas. En aras de la claridad, esta cantidad puede ser denominada “energía residual” y puede usarse, además, para evaluar un “criterio de selección de la ganancia”, en el que se selecciona el parámetro g* de ganancia óptima. En la Ecuación (6) se da tal criterio de selección de ganancia, aunque muchos son posibles.

La necesidad de un término β_j de sesgo puede surgir del caso en el que la función W de ponderación de errores en las Ecuaciones (3) y (4) pueda no producir de forma adecuada distorsiones igualmente perceptibles en el vector $\hat{E}_j$. Por ejemplo, aunque puede usarse la función W de ponderación de errores para intentar “blanquear” en cierto grado el espectro de errores, puede haber ciertas ventajas en ponderar más las frecuencias bajas, debido a la percepción de distorsión por parte del oído humano. Como consecuencia de la mayor ponderación de los errores en las

frecuencias bajas, las señales de frecuencia elevada pueden ser inframodeladas por la capa de mejora. En estos casos, puede haber un beneficio directo en sesgar la métrica de distorsión hacia valores de g_j que no atenúen los componentes de alta frecuencia de S_j, de modo que el inframodelado de las altas frecuencias no dé como resultado aberraciones de sonido objetable o poco natural en la señal final de audio reconstruida. Un ejemplo de esto sería el caso de una señal de voz sorda. En este caso, el audio de entrada está compuesto generalmente de señales de tipo ruido de frecuencias medias a elevadas producidas por un flujo turbulento de aire procedente de la boca humana. Puede ser que el codificador de la capa de núcleo no codifique directamente este tipo de forma de onda, pero que pueda usar un modelo de ruido para generar una señal de audio de sonido similar. Esto puede dar como resultado una correlación generalmente baja entre el audio de entrada y las señales de audio de salida de la capa de núcleo. Sin embargo, en esta realización, el vector E_j de señales de error está basado en una diferencia entre el audio de entrada y las señales de salida de audio de la capa de núcleo. Dado que estas señales pueden no correlacionarse muy bien, la energía de la señal E_j de error puede no ser necesariamente inferior al audio de entrada o al audio de salida de la capa de núcleo. En ese caso, la minimización del error en la Ecuación (6) puede dar como resultado que el cambio de escala de la ganancia sea demasiado agresivo, lo que puede dar como resultado aberraciones audibles potenciales.

En otro caso, los factores β_j de sesgo pueden basarse en otras características de la señal del audio de entrada y/o de las señales de audio de salida de la capa de núcleo. Por ejemplo, la relación pico a media del espectro de una señal puede dar una indicación del contenido armónico de esa señal. Señales tales como la voz y ciertos tipos de música pueden tener un contenido armónico elevado y, así, una relación pico a media elevada. Sin embargo, una señal de música procesada a través de un códec de voz puede dar como resultado una calidad deficiente debido a la discordancia del modelo de codificación y, en consecuencia, el espectro de la señal de salida de la capa de núcleo puede tener una relación pico a media reducida cuando se lo compara con el espectro de la señal de entrada. En este caso, puede resultar beneficioso reducir la cantidad de sesgo en el procedimiento de minimización para permitir que el audio de salida de la capa de núcleo experimente un cambio de escala en ganancia hasta una energía menor, permitiendo con ello que la codificación de la capa de mejora tenga un efecto más pronunciado en el audio compuesto de salida. En cambio, ciertos tipos de señales de entrada de voz o música pueden presentar menores relaciones pico a media, en cuyo caso puede percibirse que las señales son más ruidosas y, por lo tanto, pueden beneficiarse de un cambio de escala menor del audio de salida de la capa de núcleo aumentando el sesgo de error. Un ejemplo de una función para generar los factores de sesgo para β_j está dado como:

$$\beta_j = \begin{cases} 1+10^6 \cdot j; & VozSorda == VERDADERO \text{ o } \phi_s < \lambda\phi_{s_c} \\ 10^{(-j\cdot\Delta/10)}; & \text{en los demás casos} \end{cases}, \quad 0 \le j < M. \tag{7}$$

pudiendo ser λ cierto umbral, y pudiendo darse la relación pico a media para el vector Φ_y como:

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

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

Una vez se determina el índice j^* de ganancia óptima a partir de la Ecuación (6), se genera la palabra asociada de código i_g y se envía el vector óptimo E* de errores al codificador 430 de señales de error, en el que E* es codificado a una forma que es adecuada para el multiplexado con otras palabras de código (por medio del MUX 440) y transmitido para su uso por un correspondiente decodificador. En una realización ejemplar, el codificador 408 de señales de error usa la codificación factorial de impulsos (FPC). Este método resulta ventajoso desde un punto de vista de la complejidad de procesamiento, dado que el procedimiento de enumeración asociado con la codificación del vector E* es independiente del procedimiento de generación de vectores que se usa para generar $\hat{\mathbf{E}}_j$.

El decodificador 450 de la capa de mejora invierte estos procedimientos para producir la salida mejorada $\hat{s}(n)$ de audio. Más específicamente, el decodificador 450 recibe i_g, siendo enviada i_E por el demultiplexor 455 al decodificador 460 de señales de error, en el que el vector óptimo E* de errores se deriva de la palabra de código. Se pasa el vector óptimo E* de errores al combinador 465 de señales, en el que se modifica la $\hat{s}_c(n)$ recibida como en la Ecuación (2) para producir $\hat{s}(n)$.

Una segunda realización de la presente invención implica un sistema integrado de codificación de múltiples capas como el mostrado en la FIG. 5. Aquí, puede verse que hay cinco capas integradas dadas para este ejemplo. Las capas 1 y 2 pueden estar basadas ambas en códecs de voz, y las capas 3, 4 y 5 pueden ser capas de mejora de TDCM. Así, los codificadores 502 y 503 pueden utilizar códecs de voz para producir y dar salida a la señal codificada $s(n)$ de entrada. Los codificadores 510, 610 y 514 comprenden codificadores de la capa de mejora, produciendo

cada uno una mejora diferente a la señal codificada. De forma similar a la realización anterior, el vector de señales de error para la capa 3 (codificador 510) puede darse como:

$$\mathbf{E}_3 = \mathbf{S} - \mathbf{S}_2, \tag{9}$$

siendo **S** = TDCM{**Ws**} la señal de entrada transformada ponderada, y siendo $\mathbf{S}_2$ = TDCM{$\mathbf{Ws}_2$} la señal transformada ponderada generada a partir del decodificador 506 de las capas 1/2. En esta realización, la capa 3 puede ser una capa de cuantificación de baja velocidad de transferencia y, como tal, puede haber relativamente pocos bits para codificar la correspondiente señal de error cuantificada $\hat{\mathbf{E}}_3 = Q\{\mathbf{E}_3\}$. Para proporcionar una buena calidad con estas limitaciones, puede cuantificarse únicamente una fracción de los coeficientes dentro de E_3. Las posiciones de los coeficientes que han de codificarse pueden ser fijas o pueden ser variables, pero, si se permite que varíen, puede hacer falta enviar información adicional al decodificador para identificar estas posiciones. Por ejemplo, si el intervalo de posiciones codificadas empieza en k_s y termina en k_e, siendo $0 \le k_s < k_e < N$, entonces el vector $\hat{\mathbf{E}}_3$ de señales de error cuantificadas puede contener valores distintos de cero únicamente dentro de ese intervalo, y ceros para las posiciones fuera de ese intervalo. La información de la posición y el intervalo también puede estar implícita, dependiendo del método de codificación usado. Por ejemplo, es bien sabido en la codificación de audio que una banda de frecuencias puede ser considerada perceptualmente importante y que la codificación de un vector de señales puede centrarse en esas frecuencias. En estas circunstancias, el intervalo codificado puede ser variable, y puede no abarcar un conjunto contiguo de frecuencias. Pero, en cualquier caso, una vez que se cuantifica esta señal, el espectro compuesto de salida codificado puede construirse como:

$$\mathbf{S}_3 = \hat{\mathbf{E}}_3 + \mathbf{S}_2, \tag{10}$$

el cual es usado a continuación como entrada al 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. Usando el vector g_j de ganancia candidato, puede describirse el correspondiente vector de errores como:

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

pudiendo ser G_j una correspondiente matriz de ganancia, con el vector g_j como el componente diagonal. En la realización actual, sin embargo, el vector g_j de ganancia puede estar relacionado de la manera siguiente con el vector $\hat{\mathbf{E}}_3$ de señales de error cuantificadas. Dado que el vector $\hat{\mathbf{E}}_3$ de señales de error cuantificadas puede estar limitado en el intervalo de frecuencias, por ejemplo empezando en la posición vectorial k_s y acabando en la posición vectorial k_e, se supone que la señal S_3 de salida de la capa 3 estará codificada con mucha precisión dentro de ese intervalo. Por lo tanto, según la presente invención, se ajusta el vector g_j de ganancia en función de las posiciones codificadas, k_s y k_e, del vector de señales de error de la capa 3. Más específicamente, para mantener la integridad de la señal en esas ubicaciones, pueden fijarse los correspondientes elementos individuales de ganancia a un valor constante α. Es decir:

$$g_j(k) = \begin{cases} \alpha; & k_s \le k \le k_e \\ \gamma_j(k); & \text{en los demás casos} \end{cases}, \tag{12}$$

siendo, generalmente, $0 \le \gamma_j(k) \le 1$ y siendo $g_j(k)$ la ganancia de la posición k-ésima del vector candidato j-ésimo. En una realización ejemplar, el valor de la constante es uno ($\alpha = 1$); sin embargo, son posibles muchos valores. Además, el intervalo de frecuencias puede abarcar múltiples posiciones de comienzo y de finalización. Es decir, la Ecuación (12) puede estar segmentada en intervalos no continuos de ganancias variables que se basan en alguna función de la señal $\hat{\mathbf{E}}_3$ de errores, y puede ser escrita de forma más general como:

$$g_j(k) = \begin{cases} \alpha; & \hat{E}_3(k) \ne 0 \\ \gamma_j(k); & \text{en los demás casos} \end{cases}, \tag{13}$$

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

$$\gamma_j(k) = \begin{cases} \alpha \cdot 10^{(-j \cdot \Delta/20)}; & k_l \le k \le k_h \\ \alpha; & \text{en los demás casos} \end{cases}, \quad 0 \le j < M, \tag{14}$$

siendo Δ un valor de incremento (por ejemplo, $\Delta \approx 2{,}2$ dB), siendo α una constante, siendo M el número de candidatos (por ejemplo, M = 4, lo que puede representarse usando solo 2 bits), y siendo k_l y k_h los cortes bajo y alto de frecuencia, respectivamente, sobre los que puede tener lugar la reducción de ganancia. La introducción de los parámetros k_l y k_h es útil en sistemas en los que se desea un cambio de escala únicamente en cierto intervalo de frecuencias. Por ejemplo, en una realización dada, las frecuencias elevadas puede no estar debidamente modeladas por la capa de núcleo; así, la energía dentro de la banda de alta frecuencia puede ser inherentemente menor que la de la señal de audio de entrada. En ese caso, puede haber poco beneficio, o ninguno, en cambiar la escala de la salida de la capa 3 en la señal de esa región, dado que en consecuencia puede aumentar la energía total de error.

En resumidas cuentas, la pluralidad de vectores g_j de ganancia candidatos se basa en alguna función de los elementos codificados de un vector de señales codificadas previamente, en este caso $\hat{\mathbf{E}}_3$**.** Esto puede expresarse en términos generales como:

$$g_j(k) = f\left(k, \hat{\mathbf{E}}_3\right). \tag{15}$$

A la mano derecha de la FIG. 5 se muestran las operaciones de los correspondientes decodificadores. A medida que se reciben las diversas capas de corrientes de bits codificadas (i_1 a i_5), se construyen las señales de salida de mayor calidad en la jerarquía de las capas de mejora sobre el decodificador de la capa de núcleo (capa 1). Es decir, para esta realización particular, dado que las dos primeras capas comprenden la codificación del modelo de voz en el dominio temporal (por ejemplo, CELP) y las tres capas restantes comprenden la codificación en el dominio de la transformada (por ejemplo, TDCM), la salida final para el sistema $\hat{s}(n)$ se genera según lo siguiente:

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

siendo $\hat{e}_2(n)$ la señal de la capa de mejora del dominio temporal de la capa 2, y siendo $\hat{\mathbf{S}}_2$ = TDCM{$\mathbf{Ws}_2$} el vector ponderado de TDCM correspondiente a la salida $\hat{s}_2(n)$ de audio de la capa 2. En esta expresión, puede determinarse la señal total $\hat{s}(n)$ de salida a partir del nivel mayor de las capas consecutivas de corrientes de bits que se reciban. En esta realización, se supone que las capas de nivel inferior tienen mayor probabilidad de ser recibidas debidamente desde el canal; 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 apropiado de decodificación de la capa de mejora en la Ecuación (16).

La FIG. 6 es un diagrama de bloques que muestra el codificador 610 y el decodificador 650 de la capa 4. El codificador y el decodificador mostrados en la FIG. 6 son similares a los mostrados en la FIG. 4, salvo en que el valor de ganancia usado por las unidades 615 y 670 de escala se deriva, respectivamente, por medio de los generadores 630 y 660 de ganancia selectiva de frecuencia. Durante la operación, la salida S_3 de audio de la capa 3 se produce desde el codificador de la capa 3 y es recibida por la unidad 615 de escala. Además, el vector $\hat{\mathbf{E}}_3$ de errores de la capa 3 se produce desde el codificador 510 de la capa 3 y es recibido por el generador 630 de ganancia selectiva de frecuencia. Según se ha expuesto, dado que el vector $\hat{\mathbf{E}}_3$ de señales de error cuantificadas puede estar limitado en el intervalo de frecuencias, se ajusta el vector g_j de ganancia en función, por ejemplo, de 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 S_j a escala se produce desde la unidad 615 de escala y es recibido por el generador 620 de señales de error. Según se ha expuesto en lo que antecede, el generador 620 de señales de error recibe la señal S de audio de entrada y determina un valor E_j de error para cada vector de escala utilizado por la unidad 615 de escala. Estos vectores de error se pasan a la circuitería selectora 635 de ganancia junto con los valores de ganancia usados en la determinación de los vectores de error y un error E* particular basado en el valor óptimo g* de ganancia. Una palabra de código (i_g) que representa la ganancia óptima g* se produce desde el selector 635 de ganancia, junto con el vector óptimo E* de errores, y es pasada al codificador 640 de señales de error, en el que se determina y se produce la palabra de código i_E. Tanto i_g como i_E son enviadas al multiplexor 645 y transmitidas a través del canal 125 al decodificador 650 de la capa 4.

Durante la operación del decodificador 650 de la capa 4, i_g e i_E se reciben del canal 125 y son demultiplexadas por el demultiplexor 655. La palabra de código i_g de ganancia y el vector $\hat{\mathbf{E}}_3$ de errores de la capa 3 son usados como entrada al generador 660 de ganancia selectiva de frecuencia para producir el vector g* de ganancia según el correspondiente método del codificador 610. A continuación, se aplica el vector g* de ganancia al vector $\hat{S}_3$ de audio reconstruido de la capa 3 dentro de la unidad 670 de escala, cuya salida es combinada entonces en el combinador

675 de señales con el vector E* de errores de la capa de mejora de la capa 4, que se obtuvo del decodificador 655 de señales de error a través de la decodificación de la palabra de código i_E, para producir, según se muestra, la salida $\hat{S}_4$ de audio reconstruido de la capa 4.

La FIG. 7 es un diagrama 700 de flujo que muestra la operación de un codificador según las realizaciones primera y segunda de la presente invención. Según se ha expuesto en lo que antecede, ambas realizaciones utilizan una capa de mejora que cambia la escala del audio codificado con varios valores de escala y luego escoge el valor de escala que dé como resultado el menor error. Sin embargo, en la segunda realización de la presente invención, se utiliza el generador 630 de ganancia selectiva de frecuencia para generar los valores de ganancia.

El flujo lógico comienza en el bloque 710, en el que un codificador de la capa de núcleo recibe una señal de entrada que ha de codificarse 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 415 de escala cambia la escala de la señal de audio codificada con varios valores de ganancia para producir varias señales de audio codificadas a escala, cada una de las cuales tiene un valor de ganancia asociado (bloque 720). En el bloque 730, el generador 420 de señales de error determina varios valores de error que existen entre la señal de entrada y cada una de las varias señales de audio codificadas a escala. A continuación, el selector 425 de ganancia escoge un valor de ganancia entre los varios valores de ganancia (bloque 740). Según se ha expuesto en lo que antecede, el valor de ganancia (g*) se asocia con una señal de audio codificada a escala que dé como resultado que exista un valor de error (E*) bajo entre la señal de entrada y la señal de audio codificada a escala. Por último, en el bloque 750, el transmisor 440 transmite el valor de error (E*) bajo junto con el valor de ganancia (g*) como parte de una capa de mejora a la señal de audio codificada. Tal como reconocerá una persona con un dominio normal de la técnica, tanto E* como g* son debidamente codificados antes de la transmisión.

Según se ha expuesto en lo que antecede, en el lado del receptor, se recibirá la señal de audio codificada 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 con el valor de ganancia.

Cambio de escala de la capa de núcleo para la estereofonía

En la descripción anterior se ha descrito un sistema integrado de codificación en el que cada una de las capas codificaba una señal monoaural. Ahora se describirá un sistema integrado de codificación para codificar señales estereofónicas u otras de múltiples canales. En aras de la brevedad, se describe la tecnología en el contexto de una señal estereofónica que consiste en dos entradas (fuentes) de audio; sin embargo, las realizaciones ejemplares descritas en el presente documento pueden extenderse fácilmente a casos en los que la señal estereofónica tenga más de dos entradas de audio, como ocurre en las entradas de audio de múltiples canales. Con fines de ilustración y no de limitación, las dos entradas de audio son señales estereofónicas que consisten en la señal izquierda (s_I) y la señal derecha (s_D), siendo s_I y s_D vectores de columna de n dimensiones que representan una trama de datos de audio. De nuevo en aras de la brevedad, se expondrá con detalle un sistema integrado de codificación que consiste en dos capas: concretamente, una capa de núcleo y una capa de mejora. La idea propuesta puede extenderse fácilmente a un sistema integrado de codificación de múltiples capas. Además, el códec puede no estar integrado *per se;* es decir, puede tener solo una capa, estando dedicados algunos de los bits de ese códec para una señal estereofónica y el resto de los bits para la señal monoaural.

Se conoce un códec estereofónico integrado consistente en una capa de núcleo que simplemente codifica una señal monoaural y capas de mejora que codifican ya sea las señales de frecuencia mayor o las estereofónicas. En ese escenario limitado, la capa de núcleo codifica una señal monoaural (s), obtenida de la combinación de s_I y s_D, para producir una señal codificada monoaural $\hat{s}$. Sea H una matriz de combinación 2×1 usada para generar una señal monoaural, es decir,

$$\mathbf{s} = \left(\mathbf{s}_I \; \mathbf{s}_D\right)\mathbf{H} \tag{17}$$

Se hace notar que, en la Ecuación (17), s_D puede ser una versión retardada de la señal derecha de audio en vez de simplemente la señal del canal derecho. Por ejemplo, el retardo puede calcularse para maximizar la correlación de S_I y la versión retardada de S_D. Si la matriz H es $[0{,}5 \; 0{,}5]^T$, entonces la Ecuación 17 da como resultado una ponderación igual de los canales derecho e izquierdo respectivos; es decir, $\mathbf{s} = 0{,}5s_I + 0{,}5s_D$. Las realizaciones presentadas en el presente documento no están limitadas a que la capa de núcleo codifique la señal monoaural y la capa de mejora codifique la señal estereofónica. Tanto la capa de núcleo del códec integrado como la capa de mejora pueden codificar señales de audio de múltiples canales. El número de canales en la señal de audio de múltiples canales que son codificados por los múltiples canales de la capa de núcleo puede ser menor que el número de canales de la señal de audio de múltiples canales que pueden ser codificados por la capa de mejora. Sean *(m, n)* los números de canales que han de ser codificados por la capa de núcleo y la capa de mejora, respectivamente. Sea S_1, S_2, S_3, …, S_n una representación de n canales de audio que han de ser codificados por el sistema integrado. Los m canales que han de ser codificados por la capa de núcleo se derivan de estos y se obtienen como

$$\left[\mathbf{s}^1\ \mathbf{s}^2\ \ldots\ \mathbf{s}^m\right] = \left[\mathbf{s}_1\ \mathbf{s}_2\ \ldots\ \mathbf{s}_n\right]\mathbf{H}, \tag{17a}$$

siendo H una matriz de $n \times m$.

Según se ha mencionado antes, la capa de núcleo codifica una señal monoaural s para producir una señal codificada ŝ de la capa de núcleo. Para generar estimaciones de los componentes estereofónicos a partir de ŝ, se calcula un factor de balance. Este factor de balance se calcula como:

$$w_I = \frac{\mathbf{s}_I^T\mathbf{s}}{\mathbf{s}^T\mathbf{s}}, \quad w_D = \frac{\mathbf{s}_D^T\mathbf{s}}{\mathbf{s}^T\mathbf{s}} \tag{18}$$

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

$$w_I = 2 - w_D \tag{19}$$

Obsérvese que la relación permite la cuantificación de únicamente un parámetro y que el otro puede extraerse fácilmente del primero. La salida estereofónica se calcula ahora como

$$\hat{\mathbf{s}}_I = w_I\hat{\mathbf{s}}, \quad \hat{\mathbf{s}}_D = w_D\hat{\mathbf{s}} \tag{20}$$

En la sección subsiguiente, se trabajará en el dominio frecuencial en vez de en el dominio temporal. Por ello, una señal correspondiente en el dominio frecuencial está representada por una letra mayúscula; es decir, S, Ŝ, S_I, S_D, $\hat{S}_I$ y $\hat{S}_D$ son, respectivamente, la representación de s, ŝ, s_I, s_D, $\hat{s}_I$ y $\hat{s}_D$ en el dominio frecuencial. El factor de balance en el dominio frecuencial se calcula usando términos en el dominio frecuencial y está dado por

$$W_I = \frac{\mathbf{S}_I^T\mathbf{S}}{\mathbf{S}^T\mathbf{S}}, \quad W_D = \frac{\mathbf{S}_D^T\mathbf{S}}{\mathbf{S}^T\mathbf{S}} \tag{21}$$

y

$$\hat{\mathbf{S}}_I = W_I\hat{\mathbf{S}}, \quad \hat{\mathbf{S}}_D = W_D\hat{\mathbf{S}} \tag{22}$$

En el dominio frecuencial, los vectores pueden ser divididos adicionalmente en subvectores no solapados; es decir, un vector S de dimensión n puede dividirse en t subvectores, S_1, S_2, …, S_t, de dimensiones m_1, m_2, …m_t, de modo que

$$\sum_{k=1}^{t} m_k = n. \tag{23}$$

En este caso puede calcularse un factor de balance diferente para cada subvector diferente; es decir,

$$W_{Ik} = \frac{\mathbf{S}_{Ik}^T\mathbf{S}_k}{\mathbf{S}_k^T\mathbf{S}_k}, \quad W_{Dk} = \frac{\mathbf{S}_{Dk}^T\mathbf{S}_k}{\mathbf{S}_k^T\mathbf{S}_k} \tag{24}$$

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

Con referencia ahora a las FIGURAS 8 y 9, se muestran dibujos de la técnica anterior relevantes a señales estereofónicas y otras de múltiples canales. El sistema 800 de compresión integrado de voz/audio de la técnica anterior de la FIG. 8 es similar al de la FIG. 1, pero tiene múltiples señales de entrada de audio, en este ejemplo mostradas como señales S(n) de entrada estereofónica izquierda y derecha. Estas señales de audio de entrada son suministradas al combinador 810, que produce, según se muestra, el audio $s(n)$ de entrada. Las múltiples señales de entrada también son proporcionadas, según se muestra, al codificador 820 de la capa de mejora. En el lado de decodificación, el decodificador 830 de la capa de mejora produce, según se muestra, las señales de audio mejorado $\hat{s}_I$, $\hat{s}_D$ de salida.

La FIG. 9 ilustra un codificador 900 anterior de la capa de mejora como podría usarse en la FIG. 8. Según se muestra, se proporcionan las múltiples entradas de audio a un generador del factor de balance, junto con la señal de audio de salida de la capa de núcleo. El generador 920 del factor de balance del codificador 910 de la capa de mejora recibe las múltiples entradas de audio para producir la señal i_B, que pasa, según se muestra, al MUX 325. La señal i_B es una representación del factor de balance. En la realización preferente, i_B es una secuencia de bits que

representa los factores de balance. En el lado del decodificador, esta señal i_B es recibida por el decodificador 940 del factor de balance, que produce, según se muestra, elementos $W_I(n)$ y $W_D(n)$ del factor de balance, que son recibidos, según se muestra, por el combinador 950 de señales.

Cálculo del factor de balance de múltiples canales

Según se ha mencionado antes, en muchas situaciones el códec usado para la codificación de la señal monoaural está diseñado para voz de un solo canal y da como resultado el ruido del modelo de codificación siempre que se use para codificar señales que no estén plenamente soportadas por el modelo del códec. Las señales de música y otras señales de tipo distinto de la voz son algunas de las señales que no son debidamente modeladas por un códec de la capa de núcleo que se base en un modelo de voz. La descripción que antecede, con referencia a las FIGURAS 1-7, ha propuesto aplicar una ganancia selectiva de frecuencia a la señal codificada por la capa de núcleo. El cambio de escala se ha optimizado para minimizar una distorsión particular (valor de error) entre la entrada de audio y la señal codificada a escala. El enfoque descrito en lo que antecede funciona bien para señales de un solo canal, pero puede no ser óptimo para aplicarlo al cambio de escala de la capa de núcleo cuando la capa de mejora codifica las señales estereofónicas u otras de múltiples canales.

Dado que el componente monoaural de la señal de múltiples canales, tal como la señal estereofónica, se obtiene de la combinación de las dos o más entradas estereofónicas de audio, la señal combinada s puede no conformarse al modelo de voz de un solo canal; de aquí que el códec de la capa de núcleo pueda producir ruido cuando codifica la señal combinada. Así, existe la necesidad de un enfoque que permita el cambio de escala de la señal codificada de la capa de núcleo en un sistema integrado de codificación, reduciendo con ello el ruido generado por la capa de núcleo. En el enfoque de la señal monoaural descrito en lo que antecede, una medida particular de la distorsión sobre la que se obtuvo el cambio de escala selectivo de la frecuencia se basaba en el error de la señal monoaural. Este error $E_4(j)$ está mostrado en la anterior Ecuación (11). Sin embargo, la distorsión solo de la señal monoaural no es suficiente para mejorar la calidad del sistema estereofónico de comunicaciones. El cambio de escala contenido en la Ecuación (11) puede ser por un factor de escala de unidad (1) o por cualquier otra función identificada.

Para una señal estereofónica, una medida de distorsión debería capturar la distorsión tanto del canal derecho como del izquierdo. Sean E_I y E_D los vectores de error para los canales izquierdo y derecho, respectivamente, y estén dados por

$$\mathbf{E}_I = \mathbf{S}_I - \hat{\mathbf{S}}_I, \quad \mathbf{E}_D = \mathbf{S}_D - \hat{\mathbf{S}}_D \tag{25}$$

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

$$\mathbf{E}_I = \mathbf{S}_I - W_I \cdot \hat{\mathbf{S}}, \quad \mathbf{E}_D = \mathbf{S}_D - W_D \cdot \hat{\mathbf{S}}. \tag{26}$$

Consideremos ahora el caso en que se aplican, a S, vectores de ganancia g_j ($0 \le j < M$) selectivos de frecuencia. Este vector de ganancia selectivo de frecuencia se representa en forma matricial como G_j, siendo G_j una matriz diagonal con elementos diagonales g_j. Para cada vector G_j, se calculan los vectores de error como:

$$\mathbf{E}_I(j) = \mathbf{S}_I - W_I \cdot \mathbf{G}_j \cdot \hat{\mathbf{S}}, \quad \mathbf{E}_D(j) = \mathbf{S}_D - W_D \cdot \mathbf{G}_j \cdot \hat{\mathbf{S}}, \tag{27}$$

dándose las estimaciones de las señales estereofónicas mediante los términos $W \cdot G_j \cdot S$. Puede verse que la matriz G de ganancia puede ser una matriz unidad (1) o puede ser cualquier otra matriz diagonal; se reconoce que no toda estimación posible puede funcionar para cada señal a escala.

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

$$\varepsilon_j = f\left(\mathbf{E}_I(j), \mathbf{E}_D(j)\right) \tag{28}$$

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

El índice j del vector de ganancia selectivo de frecuencia que se selecciona está dado por:

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

En una realización ejemplar, la medida de distorsión es una distorsión media al cuadrado dada por:

$$\varepsilon_j = \left\|\mathbf{E}_I(j)\right\|^2 + \left\|\mathbf{E}_D(j)\right\|^2 \tag{30}$$

O puede ser una distorsión ponderada o sesgada dada por:

$$\varepsilon_j = B_I \left\|\mathbf{E}_I(j)\right\|^2 + B_D \left\|\mathbf{E}_D(j)\right\|^2 \tag{31}$$

Los sesgos B_I y B_D pueden ser una función de las energías de los canales izquierdo y derecho.

Según se ha mencionado antes, en el dominio frecuencial los vectores pueden ser divididos adicionalmente en subvectores no solapados. Para ampliar la técnica propuesta para que incluya la división del vector del dominio frecuencial en subvectores, se calcula para cada subvector el factor de balance usado en (27). Así, los vectores E_I y E_D de error para cada ganancia selectiva de frecuencia se forman por una concatenación de subvectores de error dada por

$$\mathbf{E}_{Ik}(j) = \mathbf{S}_{Ik} - W_{Ik} \cdot \mathbf{G}_{jk} \cdot \hat{\mathbf{S}}_k, \quad \mathbf{E}_{Dk}(j) = \mathbf{S}_{Dk} - W_{Dk} \cdot \mathbf{G}_{jk} \cdot \hat{\mathbf{S}}_k \tag{32}$$

La medida ε de distorsión de (28) es ahora una función de los vectores de error formados por concatenación de los anteriores subvectores de error.

Cálculo del factor de balance

El factor de balance generado usando la técnica anterior (Ecuación 21) es independiente de la salida de la capa de núcleo. Sin embargo, para minimizar una medida de distorsión dada en (30) y (31), puede ser beneficioso calcular también el factor de balance para minimizar la distorsión correspondiente. Ahora el factor de balance W_I y W_D puede calcularse como

$$W_I(j) = \frac{\mathbf{S}_I^T \mathbf{G}_j \hat{\mathbf{S}}}{\left\|\mathbf{G}_j \hat{\mathbf{S}}\right\|^2}, \quad W_D(j) = \frac{\mathbf{S}_D^T \mathbf{G}_j \hat{\mathbf{S}}}{\left\|\mathbf{G}_j \hat{\mathbf{S}}\right\|^2}, \tag{33}$$

pudiendo verse que el factor de balance es independiente de la ganancia, tal como se muestra, por ejemplo, en el dibujo de la FIG. 11. Esta ecuación minimiza las distorsiones de las Ecuaciones (30) y (31). El problema del uso de tal factor de balance es que ahora

$$W_I(j) \neq 2 - W_D(j), \tag{34}$$

y de ahí que puedan ser necesarios campos separados de bits para cuantificar W_I y W_D. Esto puede evitarse poniendo la limitación $W_I(j) = 2 - W_D(j)$ en la optimización. Con esta limitación, la solución óptima de la Ecuación (30) está dada por:

$$W_I(j) = \frac{2B_D}{B_D + B_I} + \frac{\left(B_D \mathbf{S}_D - B_I \mathbf{S}_I\right)^T \mathbf{G}_j \hat{\mathbf{S}}}{\left\|\mathbf{G}_j \hat{\mathbf{S}}\right\|^2}, \quad W_D(j) = 2 - W_I(j), \tag{35}$$

en donde el factor de balance, según se muestra, depende de un término de ganancia; la FIG. 10 de los dibujos ilustra un factor de balance dependiente. Si los factores B_I y B_D de sesgo son la unidad, entonces

$$W_I(j) = 1 - \frac{\left(\mathbf{S}_I - \mathbf{S}_D\right)^T \mathbf{G}_j \hat{\mathbf{S}}}{\left\|\mathbf{G}_j \hat{\mathbf{S}}\right\|^2}, \quad W_D(j) = 2 - W_I(j). \tag{36}$$

Los términos $\mathbf{S}^T\mathbf{G}_j\hat{\mathbf{S}}$ de las Ecuaciones (33) y (36) son representativos de valores de correlación entre la señal de audio codificada a escala y al menos una de las señales de audio de una señal de audio de múltiples canales.

En la codificación estereofónica, la dirección y la ubicación del origen del sonido pueden ser más importantes que la distorsión media al cuadrado. Por lo tanto, la relación entre la energía del canal izquierdo y la energía del canal derecho puede ser un mejor indicador de la dirección (o de la ubicación del origen del sonido) que la minimización

de una medida ponderada de la distorsión. En tales escenarios, el factor de balance calculado en las Ecuaciones (35) y (36) puede no ser un buen enfoque para el cálculo del factor de balance. Es preciso mantener la relación de la energía entre los canales izquierdo y derecho antes y después de codificarlo. La relación de energía de los canales antes de la codificación y después de la codificación está dada por:

$$v = \frac{\|\mathbf{S}_I\|^2}{\|\mathbf{S}_D\|^2}, \quad \hat{v} = \frac{W_I^2(j)\|\hat{\mathbf{S}}\|^2}{W_D^2(j)\|\hat{\mathbf{S}}\|^2}, \tag{37}$$

respectivamente. Igualando estas dos relaciones de energía y usando la premisa $W_I(j) = 2 - W_D(j)$, obtenemos

$$W_I = \frac{2\sqrt{\mathbf{S}_I^T\mathbf{S}_I}}{\sqrt{\mathbf{S}_I^T\mathbf{S}_I} + \sqrt{\mathbf{S}_D^T\mathbf{S}_D}}, \quad W_D = 2 - W_I, \tag{38}$$

que dan los componentes del factor de balance generado. Obsérvese que el factor de balance calculado en (38) es ahora independiente de G_j; así, ya no es una función de j, proporcionando un factor autocorrelacionado de balance que es independiente de la consideración de ganancia; en la FIG. 10 de los dibujos se ilustra adicionalmente un factor dependiente de balance. Usando este resultado con las Ecuaciones 29 y 32, podemos extender la selección del índice óptimo j de cambio de escala de la capa de núcleo para que incluya los segmentos k de vectores concatenados, de modo que

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

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

Con referencia ahora a la FIG. 10, se ilustra un diagrama 1000 de bloques de un codificador de la capa de mejora y un decodificador de la capa de mejora según diversas realizaciones. Las señales s(n) de audio de entrada son recibidas por el generador 1050 del factor de balance del codificador 1010 de la capa de mejora y el generador 1030 de señales de error (señales de distorsión) del generador 1020 del vector de ganancia. La señal $\hat{S}(n)$ de audio codificada procedente de la capa de núcleo es recibida, según se muestra, por la unidad 1025 de escala del generador 1020 del vector de ganancia. La unidad 1025 de escala opera para cambiar la escala de la señal S(n) de audio codificada con varios valores de ganancia para generar varias señales de audio codificadas candidatas, cambiándose la escala de al menos una de las señales de audio codificadas candidatas. Según se ha mencionado previamente, puede emplearse un cambio de escala por la unidad o por cualquier función deseada de identificación. La unidad 1025 de escala produce el audio S_j a escala, que es recibido por el generador 1050 del factor de balance. En lo que antecede, en conexión con las Ecuaciones (18), (21), (24) y (33), se expuso la generación de un factor de balance que tiene varios componentes del factor de balance, cada uno de los cuales está asociado con una señal de audio de las señales de audio de múltiples canales recibidos por el codificador 1010 de la capa de mejora. Esto se logra, según se muestra, por medio del generador 1050 del factor de balance para producir, según se muestra, componentes $\hat{S}_I(n)$, $\hat{S}_D(n)$ del factor de balance. Según se ha expuesto en lo que antecede en conexión con la Ecuación (38), el generador 1050 del factor de balance ilustra un factor de balance independiente de la ganancia.

El generador 1020 del vector de ganancia es responsable de determinar un valor de ganancia que ha de aplicarse a la señal de audio codificada para generar una estimación de la señal de audio de múltiples canales, según se ha expuesto en las Ecuaciones (27), (28) y (29). Esto se logra por medio de la unidad 1025 de escala y del generador 1050 del factor de balance, que trabajan conjuntamente para generar la estimación en función del factor de balance y de al menos una señal de audio codificada a escala. El valor de ganancia se basa en el factor de balance y en la señal de audio de múltiples canales, configurándose el valor de ganancia 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) expone la generación de un valor de distorsión como una función de la estimación de la señal de entrada de múltiples canales y de la propia señal real de entrada. Así, los componentes del factor de balance son recibidos por el generador 1030 de señales de error, junto con las señales s(n) de audio de entrada, para determinar un valor E_j de error para cada vector de escala utilizado por la unidad 1025 de escala. Estos vectores de error se pasan a la circuitería selectora 1035 de ganancia junto con los valores de ganancia usados en la determinación de los vectores de error y un error E* particular basado en el valor óptimo g* de ganancia. El selector 1035 de ganancia es, entonces, operativo para evaluar el valor de distorsión en función de la estimación de la señal de entrada de múltiples canales y de la propia señal real para determinar una representación de un valor óptimo g* de ganancia de los valores de ganancia posibles. Una palabra de código (i_g) que representa la ganancia óptima g* se produce, según se muestra, desde el selector 1035 de ganancia y es recibida por el multiplexor MUX 1040.

Tanto i_g como i_B son enviados al multiplexor y transmitidos por el transmisor 1045 al decodificador 1060 de la capa de mejora a través del canal 125. Según se muestra, se produce la representación del valor i_g de ganancia para su transmisión al canal 125, pero también este valor puede ser almacenado si se desea.

En el lado del decodificador, durante la operación del decodificador 1060 de la capa de mejora, se reciben i_g e i_E desde el canal 125 y son demultiplexados por el demultiplexor 1065. Así, el decodificador de la capa de mejora recibe una señal S(n) de audio codificada, un factor i_B de balance codificado y un valor i_g de ganancia codificado. El decodificador 1070 del vector de ganancia comprende, según se muestra, un generador 1075 de ganancia selectiva de frecuencia y una unidad 1080 de escala. El decodificador 1070 del vector de ganancia genera un valor de ganancia decodificado a partir del valor de ganancia codificado. Se introduce el valor i_g de ganancia codificado en el generador 1075 de ganancia selectiva de frecuencia para producir el vector g* de ganancia según el correspondiente método del codificador 1010. A continuación, se aplica el vector g* de ganancia a la unidad 1080 de escala, que cambia la escala de la señal S(n) de audio codificada con el valor g* de ganancia decodificado para generar la señal de audio a escala. El combinador 1095 de señales recibe las señales codificadas de salida del factor de balance del decodificador 1090 del factor de balance en la señal $G_j\hat{S}(n)$ de audio a escala para generar y producir una señal de audio de múltiples canales decodificada, mostrada como las señales de audio mejorado de salida.

El diagrama 1100 de bloques ilustra un codificador de la capa de mejora y un decodificador de la capa de mejora ejemplares en los que, según se ha expuesto en lo que antecede en conexión con la Ecuación (33), el generador 1050 del factor de balance genera un factor de balance que depende de la ganancia. Esto se ilustra por medio de un generador de señales de error que genera la señal G_j 1110.

Con referencia ahora a las FIGURAS 12-14, se presentan flujos que cubren la metodología de las diversas realizaciones presentadas en el presente documento. En el flujo 1200 de la FIG. 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 varias señales de audio. En el bloque 1220, se codifica la señal de audio de múltiples canales para generar una señal de audio codificada. La señal de audio codificada puede ser una señal o bien monoaural o de múltiples canales, tal como una señal estereofónica, según se ilustra a título de ejemplo en los dibujos. Además, la señal de audio codificada puede comprender varios canales. Puede haber más de un canal en la capa de núcleo y el número de canales de la capa de mejora puede ser mayor que el número de canales de la capa de núcleo. A continuación, en el bloque 1230, se genera un factor de balance que tiene componentes del factor de balance, cada uno asociado 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 balance. Cada componente del factor de balance puede depender de otros componentes del factor de balance generados, como ocurre en la Ecuación (38). La generación del factor de balance puede comprender generar un valor de correlación entre la señal de audio codificada a escala y al menos una de las señales de audio de la señal de audio de múltiples canales, como en las Ecuaciones (33) y (36). Puede generarse una autocorrelación entre al menos una de las señales de audio, 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 que ha de aplicarse a la señal de audio codificada para generar una estimación de la señal de audio de múltiples canales en función del factor de balance y de la señal de audio de múltiples canales. Se configura el valor de ganancia 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 de los varios valores de ganancia para cambiar la escala de la señal de audio codificada y para generar las señales de audio codificadas a escala. Puede generarse el valor de distorsión en función de esta estimación; el valor de ganancia puede basarse en el valor de distorsión. En el bloque 1250, se produce una representación del valor de ganancia para su transmisión y/o su almacenamiento.

El flujo 1300 de la FIG. 13 describe otra metodología para codificar una señal de audio de múltiples canales según diversas realizaciones. En el bloque 1310 se recibe una señal de audio de múltiples canales que tiene varias señales de audio. En el bloque 1320, se codifica la señal de audio de múltiples canales para generar una señal de audio codificada. Los procedimientos de los bloques 1310 y 1320 los lleva a cabo un codificador de la capa de núcleo, según se ha descrito previamente. Tal como se ha especificado previamente, la señal de audio codificada puede ser una señal o bien monoaural o de múltiples canales, tal como una señal estereofónica, según se ilustra a título de ejemplo en los dibujos. Además, la señal de audio codificada puede comprender varios canales. Puede haber más de un canal en la capa de núcleo y el número de canales de la capa de mejora puede ser mayor que el número de canales de la capa de núcleo.

En el bloque 1330, se cambia la escala de la señal de audio codificada con varios valores de ganancia para generar varias señales de audio codificadas candidatas, cambiándose la escala de al menos una de las señales de audio codificadas candidatas. El cambio de escala lo logra la unidad de escala del generador del vector de ganancia. Según se ha expuesto, el cambio de escala de la señal de audio codificada puede incluir un cambio de escala con un valor de ganancia unidad. El valor de ganancia de los varios valores de ganancia puede ser, según se ha descrito previamente, una matriz de ganancia con el vector g_j como componente diagonal. La matriz de ganancia puede ser selectiva de la frecuencia. La señal de audio codificada ilustrada en los dibujos puede depender de la salida de la capa de núcleo. Puede escogerse un valor de ganancia de los varios valores de ganancia para cambiar la escala de la señal de audio codificada y generar las señales de audio codificadas a escala. En el bloque 1340, se genera un factor de balance que tiene componentes del factor de balance, cada uno asociado con una señal de audio de la

señal de audio de múltiples canales. La generación del factor de balance la lleva a cabo el generador del factor de balance. Cada componente del factor de balance puede depender de otros componentes del factor de balance generados, como ocurre en la Ecuación (38). La generación del factor de balance puede comprender generar un valor de correlación entre la señal de audio codificada a escala y al menos una de las señales de audio de la señal de audio de múltiples canales, como en las Ecuaciones (33) y (36). Puede generarse una autocorrelación entre al menos una de las señales de audio, 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 función del factor de balance y de la al menos una señal de audio codificada a escala. Se genera la estimación en función de la señal (s) de audio codificada a escala y del factor de balance generado. La estimación puede comprender varias estimaciones correspondientes a las varias señales de audio codificadas candidatas. En el bloque 1360 se evalúa y/o puede generarse un valor de distorsión en función de la estimación de la señal de audio de múltiples canales y la señal de audio de múltiples canales para determinar una representación de un valor óptimo de ganancia de los valores de ganancia. El valor de distorsión puede comprender varios valores de distorsión correspondientes a las varias estimaciones. La evaluación del valor de distorsión la logra la circuitería selectora de ganancia. La Ecuación (39) da la presentación de un valor óptimo de ganancia. En el bloque 1370, puede producirse una representación del valor de ganancia para su transmisión y/o su almacenamiento. El transmisor del codificador de la capa de mejora puede transmitir la representación del valor de ganancia según se ha descrito previamente.

El procedimiento implementado en el diagrama 1400 de flujo de la FIG. 14 ilustra la decodificación de una señal de audio de múltiples canales. En el bloque 1410, se reciben una señal de audio codificada, un factor de balance codificado y un valor de ganancia codificado. En el bloque 1420 se genera un valor de ganancia decodificado a partir del valor de ganancia codificado. El valor de ganancia puede ser una matriz de ganancia, descrita previamente, y la matriz de ganancia puede ser selectiva de la frecuencia. La matriz de ganancia también puede depender del audio codificado recibido como una salida de la capa de núcleo. Además, la señal de audio codificada puede ser una señal o bien monoaural o de múltiples canales, tal como una señal estereofónica, según se ilustra a título de ejemplo en los dibujos. Además, la señal de audio codificada puede comprender varios canales. Por ejemplo, puede haber más de un canal en la capa de núcleo y el número de canales de la capa de mejora puede ser mayor que el número de canales de la capa de núcleo.

En el bloque 1430, se cambia la escala de la señal codificada de audio con el valor de ganancia decodificado para generar una señal de audio a escala. En el bloque 1440 se aplica el factor de balance codificado a la señal de audio a escala para generar una señal de audio de múltiples canales decodificada. En el bloque 1450 se produce la señal de audio de múltiples canales decodificada.

Cálculo de la máscara selectiva de escala basado en la detección de picos

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

$$g_j(k) = \begin{cases} \alpha \cdot 10^{(-j\cdot\Delta/20)}; & k_l \le k \le k_h \\ \alpha; & \text{en los demás casos} \end{cases}, \quad 0 \le j < M, \tag{40}$$

siendo Δ un valor de incremento (por ejemplo, $\Delta \approx 2{,}0$ dB), siendo α una constante, siendo M el número de candidatos (por ejemplo, M = 8, lo que puede representarse usando solo 3 bits), y siendo k_l y k_h los cortes bajo y alto de frecuencia, respectivamente, sobre los que puede tener lugar la reducción de ganancia. Aquí k representa el k-ésimo coeficiente de la TDCM o la transformada de Fourier. Obsérvese que g_j es selectiva de la frecuencia, pero es independiente de la salida de la capa anterior. Los vectores g_j de ganancia pueden basarse en alguna función de los elementos codificados de un vector de señales codificadas previamente, en este caso $\hat{\mathbf{S}}$. Esto puede expresarse como:

$$g_j(k) = f\left(k, \hat{\mathbf{S}}\right). \tag{41}$$

En un sistema integrado de codificación de múltiples capas (con más de dos capas), la salida $\hat{\mathbf{S}}$ cuya escala ha de cambiar por el vector g_j de ganancia se obtiene por la aportación de al menos dos capas anteriores. Es decir,

$$\hat{\mathbf{S}} = \hat{\mathbf{E}}_2 + \hat{\mathbf{S}}_1, \tag{42}$$

siendo $\hat{\mathbf{S}}_1$ la salida de la primera capa (capa de núcleo) y siendo $\hat{\mathbf{E}}_2$ la aportación de la segunda capa o la primera capa de mejora. En este caso, los vectores g_j de ganancia pueden ser alguna función de los elementos codificados de un vector $\hat{\mathbf{S}}$ de señales codificadas previamente y la aportación de la primera capa de mejora:

$$g_j(k) = f\left(k, \hat{\mathbf{S}}, \hat{\mathbf{E}}_2\right). \quad (43)$$

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, en los picos espectrales hay mayor coincidencia entre el espectro original y el codificado. Así, los picos no debieran alterarse; es decir, el cambio de escala debería estar limitado a los valles. Para usar ventajosamente esta observación, en una de las realizaciones la función de la Ecuación (41) se basa en picos y valles de Ŝ. Sea $\Psi(\hat{S})$ una máscara de escala basada en las magnitudes de los picos detectados de Ŝ. La máscara de escala puede ser una función de valores vectoriales con valores distintos de cero en los picos detectados, es decir,

$$\psi\left(\hat{\mathbf{S}}\right) = \begin{cases} \hat{s}_i & \text{pico presente} \\ 0 & \text{en los demás casos} \end{cases}, \quad (44)$$

siendo $\hat{s}_i$ el *i*-ésimo elemento de S. La Ecuación (41) puede modificarse ahora como:

$$g_j(k) = f\left(k, \hat{\mathbf{S}}\right) = \begin{cases} \alpha \cdot 10^{(-j\cdot\Delta/20)}; & k_l \le k \le k_h, \psi_k\left(\hat{\mathbf{S}}\right) = 0 \\ \alpha; & \text{en los demás casos} \end{cases}, \quad 0 \le j < M \quad (45)$$

Para la detección de picos pueden usarse diversos enfoques. En la realización preferente, los picos se detectan haciendo pasar el espectro absoluto |Ŝ| a través de dos filtros integradores ponderados y comparando luego las salidas filtradas. Sean A_1 y A_2 la representación matricial de dos filtros integradores. Sean l_1 e l_2 ($l_1 > l_2$) las longitudes de los dos filtros. La función de detección de picos está dada como:

$$\psi\left(\hat{\mathbf{S}}\right) = \begin{cases} \hat{s}_i & \mathbf{A}_2\left|\hat{\mathbf{S}}\right| > \beta \cdot \mathbf{A}_1\left|\hat{\mathbf{S}}\right| \\ 0 & \text{en los demás casos} \end{cases}, \quad (46)$$

siendo β un valor umbral empírico.

Como un ejemplo ilustrativo, remítase el lector a la FIG. 15 y la FIG. 16. Aquí, se da como 1510 el valor absoluto de la señal codificada |**Ŝ**| en el dominio de la TDCM en ambos gráficos. Esta señal es representativa de un sonido de un "diapasón" que crea, según se muestra, una secuencia de armónicos regularmente separados. Esta señal es difícil de codificar usando un codificador de la capa de núcleo basado en un modelo de voz, porque la frecuencia fundamental de esta señal está más allá del intervalo de lo que se considera razonable para una señal de voz. Esto da como resultado un nivel bastante alto de ruido producido por la capa de núcleo, que puede ser observado comparando la señal codificada 1510 con la versión monoaural de la señal original |**S**| (1610).

A partir de la señal codificada (1510), se usa un generador de umbral para producir el umbral 1520, que corresponde a la expresión βA_1 |**Ŝ**| de la Ecuación 45. Aquí, A_1 es una matriz de convolución que, en la realización preferente, implementa una convolución de la señal |Ŝ| con una ventana de coseno de longitud 45. Son posibles muchas formas de ventana y estas pueden comprender longitudes diferentes. También en la realización preferente, A_2 es una matriz de identidad. El detector de picos compara entonces la señal 1510 con el umbral 1520 para producir la máscara ψ(**Ŝ**) de escala, mostrada como 1530.

Los vectores candidatos de cambio de escala de la capa de núcleo (dados en la Ecuación 45) pueden ser usados entonces para cambiar la escala del ruido entre picos de la señal codificada |Ŝ| para producir una señal reconstruida 1620 a escala. Puede escogerse el candidato óptimo según el procedimiento descrito en lo que antecede en la Ecuación 39 o de otro modo.

Con referencia ahora a las FIGURAS 17-19, se presentan diagramas de flujo que ilustran una metodología asociada con el cálculo de la máscara selectiva de escala basado en la detección de picos expuesta en lo que antecede según diversas realizaciones. En el diagrama 1700 de flujo de la FIG. 17, en el bloque 1710 se detecta un conjunto de picos en un vector **Ŝ** de audio reconstruido de una señal de audio recibida. La señal de audio puede estar embebida en múltiples capas. El vector **S** de audio reconstruido puede estar en el dominio frecuencial y el conjunto de picos pueden ser picos en el dominio frecuencial. La detección del conjunto de picos se lleva a cabo según una función de detección de picos dada, por ejemplo, por la Ecuación (46). Se hace notar que el conjunto puede estar vacío, como sucede cuando todo está atenuado y no hay ningún pico. En el bloque 1720 se genera una máscara $\psi(\hat{S})$ de escala en función del conjunto de picos detectado. A continuación, en el bloque 1730, se genera un vector g* de ganancia en función de al menos la máscara de escala y un índice *j* representativo del vector de ganancia. En el bloque 1740, con el vector de ganancia, se cambia la escala de la señal de audio reconstruida para producir una señal de audio reconstruida a escala. En el bloque 1750 se genera una distorsión basada en la señal de audio y la

señal de audio reconstruida a escala. En el bloque 1760 se produce el índice del vector de ganancia en función de la distorsión generada.

Con referencia ahora a la FIG. 18, el diagrama 1800 de flujo ilustra una realización alternativa de codificación de una señal de audio según ciertas realizaciones. En el bloque 1810 se recibe una señal de audio. La señal de audio puede estar embebida en múltiples capas. A continuación, en el bloque 1820 se codifica la señal de audio para generar un vector $\hat{S}$ de audio reconstruido. El vector $\hat{S}$ de audio reconstruido puede estar en el dominio frecuencial y el conjunto de picos pueden ser picos en el dominio frecuencial. En el bloque 1830, se detecta un conjunto de picos en el vector $\hat{S}$ de audio reconstruido de una señal de audio recibida. La detección del conjunto de picos se lleva a cabo según una función de detección de picos dada, por ejemplo, por la Ecuación (46). De nuevo, se hace notar que el conjunto puede estar vacío, como sucede cuando todo está atenuado y no hay ningún pico. En el bloque 1840 se genera una máscara $\psi(\hat{\mathbf{S}})$ de escala basada en el conjunto de picos detectado. En el bloque 1850, se generan varios vectores g_j de ganancia en función de la máscara de escala. En el bloque 1860, con los varios vectores de ganancia se cambia la escala de la señal de audio reconstruida para producir varias señales de audio reconstruidas a escala. A continuación, en el bloque 1870, se generan varias distorsiones en función de la señal de audio y de las varias señales de audio reconstruidas a escala. En el bloque 1880 se escoge un vector de ganancia de los varios vectores de ganancia en función de las varias distorsiones. Puede escogerse el vector de ganancia para que se corresponda con una distorsión mínima de las varias distorsiones. En el bloque 1890 se produce el índice representativo del vector de ganancia para ser transmitido y/o almacenado.

Los flujos de codificador ilustrados en lo que antecede en las FIGURAS 17-18 pueden ser implementados por la estructura del aparato descrita previamente. Con referencia al flujo 1700, en un aparato operable para codificar una señal de audio, un selector de ganancia, tal como el selector 1035 de ganancia del generador 1020 del vector de ganancia del codificador 1010 de la capa de mejora, detecta un conjunto de picos en un vector **S** de audio reconstruido de una señal de audio recibida y genera una máscara $\psi(\hat{S})$ de escala basada en el conjunto de picos detectado. De nuevo, la señal de audio puede estar embebida en múltiples capas. El vector **S** de audio reconstruido puede estar en el dominio frecuencial y el conjunto de picos pueden ser picos en el dominio frecuencial. La detección del conjunto de picos se lleva a cabo según una función de detección de picos dada, por ejemplo, por la Ecuación (46). Se hace notar que el conjunto de picos puede ser nulo si se ha atenuado todo en la señal. Una unidad de escala, tal como la unidad 1025 de escala del generador 1020 del vector de ganancia, genera un vector g* de ganancia en función de al menos la máscara de escala y un índice j representativo del vector de ganancia, cambia, con el vector de ganancia, la escala de la señal de audio reconstruida para producir una señal de audio reconstruida a escala. El generador 1030 de señales de error del generador 1025 del vector de ganancia genera una distorsión basada en la señal de audio y en la señal de audio reconstruida a escala. Un transmisor, tal como el transmisor 1045 del decodificador 1010 de la capa de mejora, es operable para producir el índice del vector de ganancia en función de la distorsión generada.

Con referencia al flujo 1800 de la FIG. 18, en un aparato operable para codificar una señal de audio, un codificador recibe una señal de audio y codifica la señal de audio para generar un vector $\hat{S}$ de audio reconstruido. Una unidad de escala, tal como la unidad 1025 de escala del generador 1020 del vector de ganancia, detecta un conjunto de picos en el vector S de audio reconstruido de una señal de audio recibida, genera una máscara $\psi(\hat{S})$ de escala basada en el conjunto de picos detectado, genera varios vectores g_j de ganancia en función de la máscara de escala, y, con los varios vectores de ganancia, cambia la escala de la señal de audio reconstruida para producir las varias señales de audio reconstruidas a escala. El generador 1030 de señales de error genera varias distorsiones en función de la señal de audio y las varias señales de audio reconstruidas a escala. Un selector de ganancia, tal como el selector 1035 de ganancia, escoge un vector de ganancia de los varios vectores de ganancia en función de las varias distorsiones. El transmisor 1045, por ejemplo, produce, para su transmisión posterior y/o su almacenamiento, el índice representativo del vector de ganancia.

En el diagrama 1900 de flujo de la FIG. 19 se ilustra un método de decodificación de una señal de audio. En el bloque 1910 se reciben un vector $\hat{\mathbf{S}}$ de audio reconstruido 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. La detección del conjunto de picos se lleva a cabo según una función de detección de picos dada, por ejemplo, por la Ecuación (46). De nuevo, se hace notar que el conjunto puede estar vacío, como sucede cuando todo está atenuado y no hay ningún pico. En el bloque 1930 se genera una máscara $\psi(\hat{S})$ de escala basada en el conjunto de picos detectado. En el bloque 1940 se genera el vector g* de ganancia en función de al menos la máscara de escala y del índice representativo del vector de ganancia. En el bloque 1950, con el vector de ganancia se cambia la escala del vector de audio reconstruido para producir una señal de audio reconstruida a escala. El método puede incluir, además, la generación de una mejora al vector de audio reconstruido y luego la combinación de la señal de audio reconstruida a escala y la mejora al vector de audio reconstruido para generar una señal decodificada mejorada.

El flujo de decodificador ilustrado en la FIG. 19 puede ser implementado por la estructura del aparato descrita previamente. En un aparato operable para decodificar una señal de audio, un decodificador 1070 del vector de ganancia de un decodificador 1060 de la capa de mejora, por ejemplo, recibe un vector **S** de audio reconstruido y un índice representativo de un vector i_g de ganancia. Según se muestra en la FIG. 10, el selector 1075 de ganancia recibe i_g mientras la unidad 1080 de escala del decodificador 1070 del vector de ganancia recibe el vector S de audio reconstruido. Un selector de ganancia, tal como el selector 1075 de ganancia del decodificador 1070 del vector de

ganancia, detecta un conjunto de picos en el vector de audio reconstruido, genera una máscara $\psi(\hat{S})$ de escala en función del conjunto de picos detectado, y genera el vector g* de ganancia en función de al menos la máscara de escala y del índice representativo del vector de ganancia. De nuevo, el conjunto puede estar vacío de contenido si la señal está atenuada en su mayor parte. El selector de ganancia detecta el conjunto de picos según una función de detección de picos, tal como la dada, por ejemplo, por la Ecuación (46). Una unidad 1080 de escala, por ejemplo, con el vector de ganancia, cambia la escala del vector de audio reconstruido para producir una señal de audio reconstruida a escala.

Además, un decodificador de señales de error, tal como el decodificador 665 de señales de error del decodificador de la capa de mejora de la FIG. 6, puede generar una mejora al vector de audio reconstruido. Un combinador de señales, como el combinador 675 de señales de la FIG. 6, combina la señal de audio reconstruida a escala y la mejora al vector de audio reconstruido para generar una señal decodificada mejorada.

Se hace notar, además, que los flujos dirigidos del factor de balance de las FIGURAS 12-14 y los flujos dirigidos de máscara de escala selectiva con detección de picos de las FIGURAS 17-19 pueden llevarse a cabo ambos en combinaciones diversas, y ello está soportado por el aparato y la estructura descritos en el presente documento.

Aunque la invención ha sido mostrada y descrita en particular con referencia a una realización particular, los expertos en la técnica entenderán que pueden realizarse en la misma diversos cambios en forma y detalles sin apartarse del alcance de la invención. Por ejemplo, aunque las técnicas anteriores están descritas en términos de transmitir y recibir por un canal en el sistema de telecomunicaciones, las técnicas pueden ser aplicadas igualmente a un sistema que use el sistema de compresión de señales con fines de reducir los requisitos de almacenamiento en un dispositivo de medios digitales, tal como un dispositivo de memoria de estado sólido o un disco duro de ordenador. El alcance de protección se define en las reivindicaciones adjuntas.

REIVINDICACIONES

1. Un aparato que decodifica una señal de audio de múltiples canales, que comprende:

un decodificador que recibe una señal de audio codificada, un factor de balance codificado y un valor de ganancia codificado;

un decodificador de un vector de ganancia de un decodificador de la capa de mejora que genera un valor de ganancia decodificado a partir del valor de ganancia codificado;

una unidad de escala del decodificador de la capa de mejora que cambia la escala de la señal de audio codificada con el valor de ganancia decodificado para generar una señal de audio a escala;

un combinador de señales que aplica el factor de balance codificado a la señal de audio a escala para generar una señal decodificada de audio de múltiples canales y produce la señal decodificada de audio de múltiples canales.

2. El aparato de la reivindicación 1 en el que el valor de ganancia es una matriz de ganancia.

3. El aparato de la reivindicación 1 en el que la señal de audio codificada es una de entre una señal monoaural y una señal de múltiples canales.

4. El aparato de la reivindicación 1 en el que el decodificador del vector de ganancia recibe un vector $\hat{S}$ de audio reconstruido y un índice representativo de un vector de ganancia y en el que el decodificador del vector de ganancia, además, comprende:

un selector de ganancia del decodificador del vector de ganancia que detecta un conjunto de picos en el vector de audio reconstruido, genera una máscara $\psi(\hat{\mathbf{S}})$ de escala en función del conjunto detectado de picos y genera el vector g* de ganancia en función de al menos la máscara de escala y del índice representativo del vector de ganancia; y

una unidad de escala del decodificador del vector de ganancia que cambia la escala del vector de audio reconstruido con el vector de ganancia para producir una señal de audio reconstruida a escala.

5. Un método para decodificar una señal de audio de múltiples canales, que comprende:

recibir una señal de audio codificada, un factor de balance codificado y un valor de ganancia codificado;

generar un valor de ganancia decodificado a partir del valor de ganancia codificado;

cambiar la escala de la señal de audio codificada con el valor de ganancia decodificado para generar una señal de audio a escala;

aplicar el factor de balance codificado a la señal de audio a escala para generar una señal de audio de múltiples canales decodificada; y

producir la señal de audio de múltiples canales decodificada.

6. El método de la reivindicación 5, comprendiendo el método:

recibir un vector $\hat{S}$ de audio reconstruido y un índice representativo de un vector de ganancia;

detectar un conjunto de picos en el vector de audio reconstruido;

generar una máscara $\psi(\hat{\mathbf{S}})$ de escala en función del conjunto de picos detectado;

generar el vector g* de ganancia en función de al menos la máscara de escala y el índice representativo del vector de ganancia; y

cambiar la escala el vector de audio reconstruido con el vector de ganancia para producir una señal de audio reconstruida a escala.

7. Un método para codificar una señal de audio de múltiples canales, que comprende:

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

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

cambiar la escala de la señal de audio codificada con varios valores de ganancia para generar varias señales de audio codificadas candidatas, estando a escala al menos una de las señales de audio codificadas candidatas;

generar un factor de balance que tiene varios componentes del factor de balance, estando asociado cada uno con una señal de audio de las varias señales de audio de la señal de audio de múltiples canales;

generar una estimación de la señal de audio de múltiples canales en función del factor de balance y la al menos una señal de audio codificada a escala de las varias señales de audio codificadas candidatas;

evaluar un valor de distorsión en función de la estimación de la señal de audio de múltiples canales y de la señal de audio de múltiples canales para determinar una representación de un valor óptimo de ganancia de los varios valores de ganancia;

producir para al menos uno de una transmisión y un almacenamiento la representación del valor óptimo de ganancia.

8. El método de la reivindicación 7 en el que un valor de ganancia de los varios valores de ganancia es una matriz de ganancia con el vector g_j como componente diagonal.

9. El método de la reivindicación 7 en el que la representación del valor óptimo de ganancia está dada por:

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

10. El método de la reivindicación 7 en el que cada componente del factor de balance está dado por:

$$W_I = \frac{2\sqrt{\mathbf{S}_I^T \mathbf{S}_I}}{\sqrt{\mathbf{S}_I^T \mathbf{S}_I} + \sqrt{\mathbf{S}_D^T \mathbf{S}_D}}, \quad W_D = 2 - W_I .$$

11. El método de la reivindicación 7 en el que la generación del factor de balance comprende generar un valor de correlación entre la señal de audio codificada a escala y al menos una de las señales de audio de la señal de audio de múltiples canales.

12. El método de la reivindicación 7 en el que la generación del factor de balance comprende:

generar una autocorrelación entre al menos una de las señales de audio de la señal de audio de múltiples canales; y

generar una raíz cuadrada de la autocorrelación.

13. El método de la reivindicación 7 que, además, comprende la generación de un valor de distorsión en función de la estimación de la señal de audio de múltiples canales y de la señal de audio de múltiples canales.

14. El método de la reivindicación 13 en el que el valor de ganancia está basado en el valor de distorsión.

15. El método de la reivindicación 14 en el que el valor de distorsión comprende varios valores de distorsión correspondientes a las varias estimaciones.

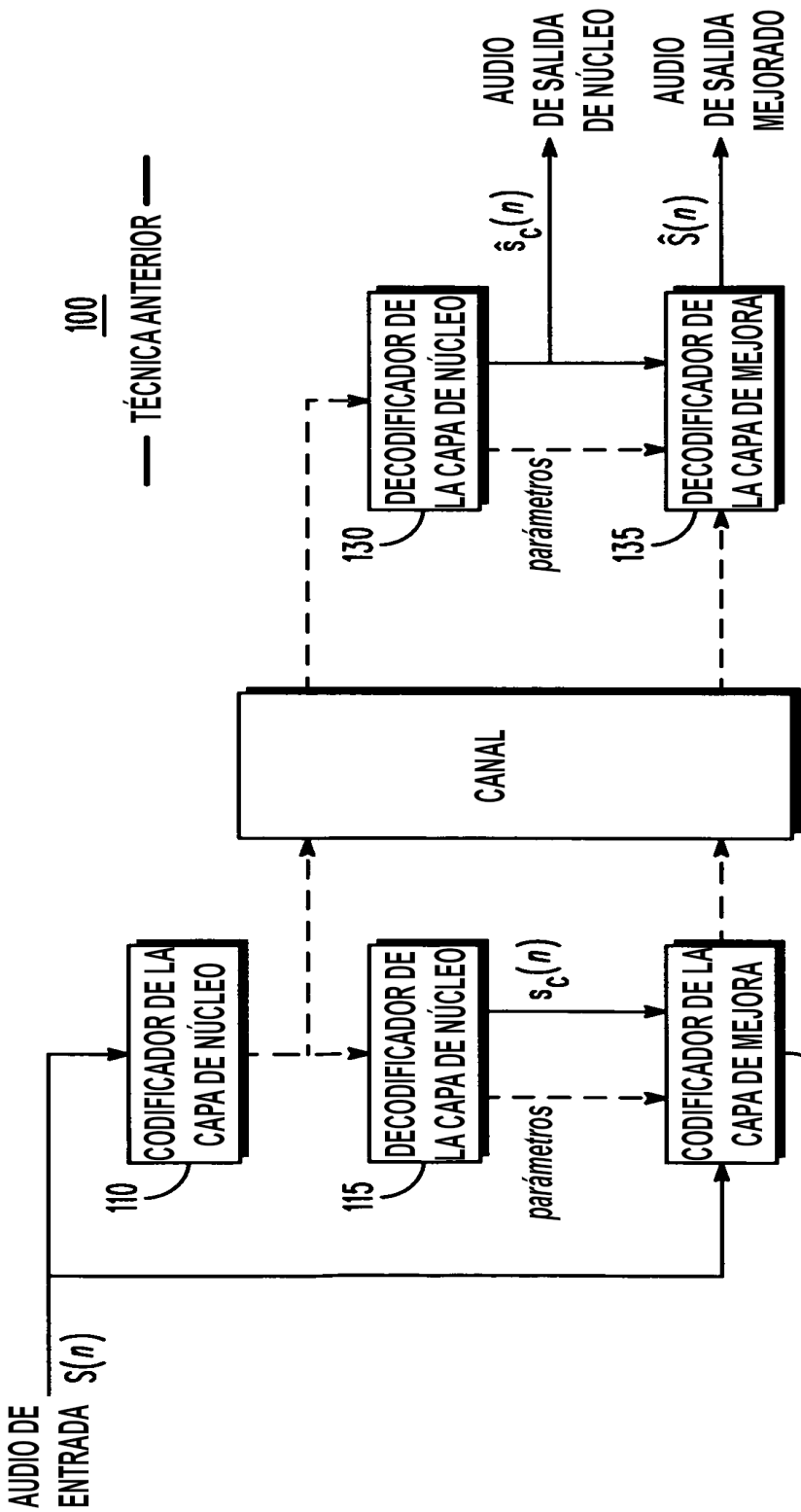
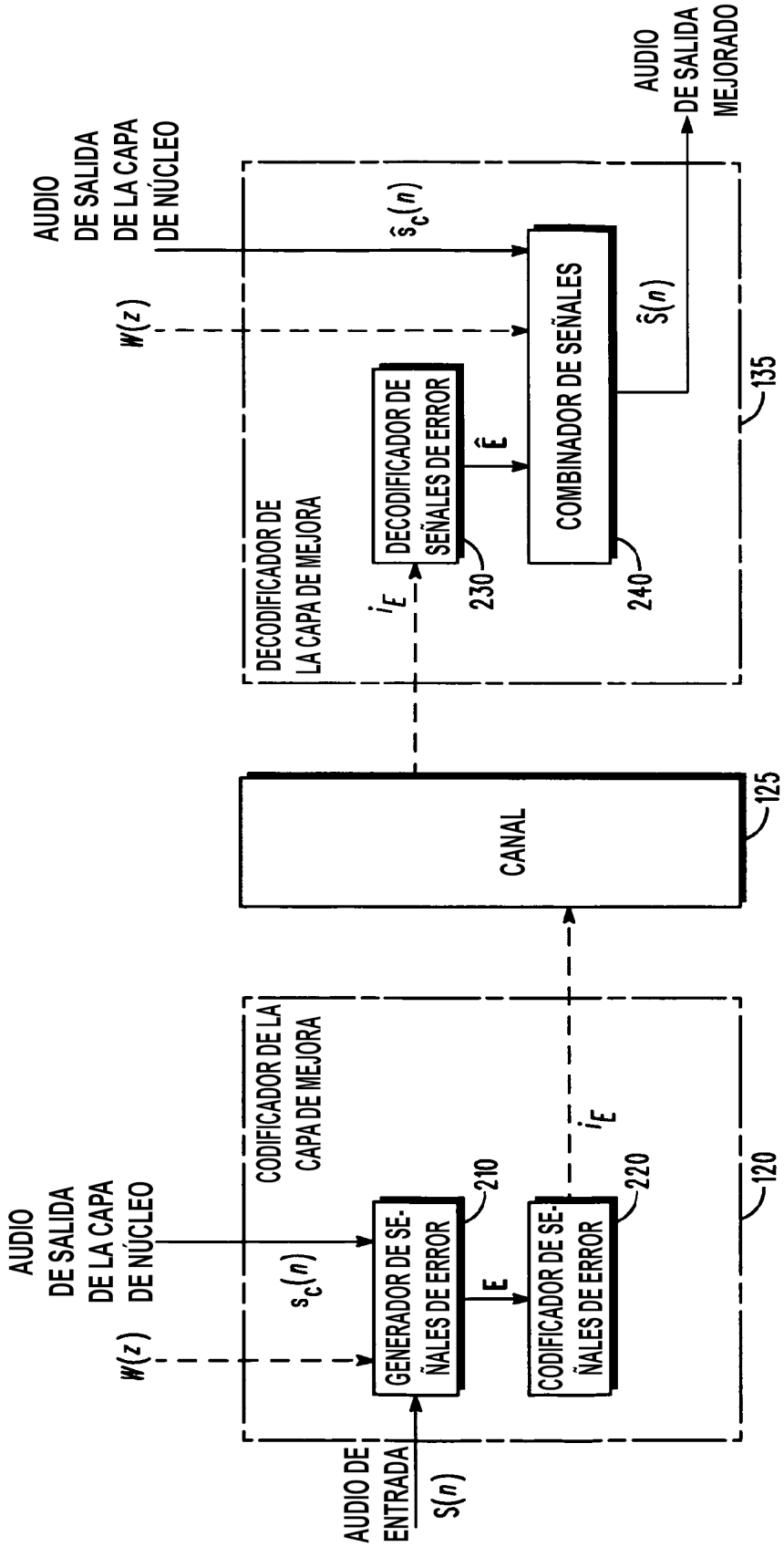

100
TÉCNICA ANTERIOR
AUDIO DE ENTRADA S(n)
CODIFICADOR DE LA CAPA DE NÚCLEO
110
DECODIFICADOR DE LA CAPA DE NÚCLEO
115
$s_c(n)$
parámetros
CODIFICADOR DE LA CAPA DE MEJORA
120
CANAL
125
DECODIFICADOR DE LA CAPA DE NÚCLEO
130
$\hat{s}_c(n)$
AUDIO DE SALIDA DE NÚCLEO
parámetros
DECODIFICADOR DE LA CAPA DE MEJORA
135
$\hat{s}(n)$
AUDIO DE SALIDA MEJORADO

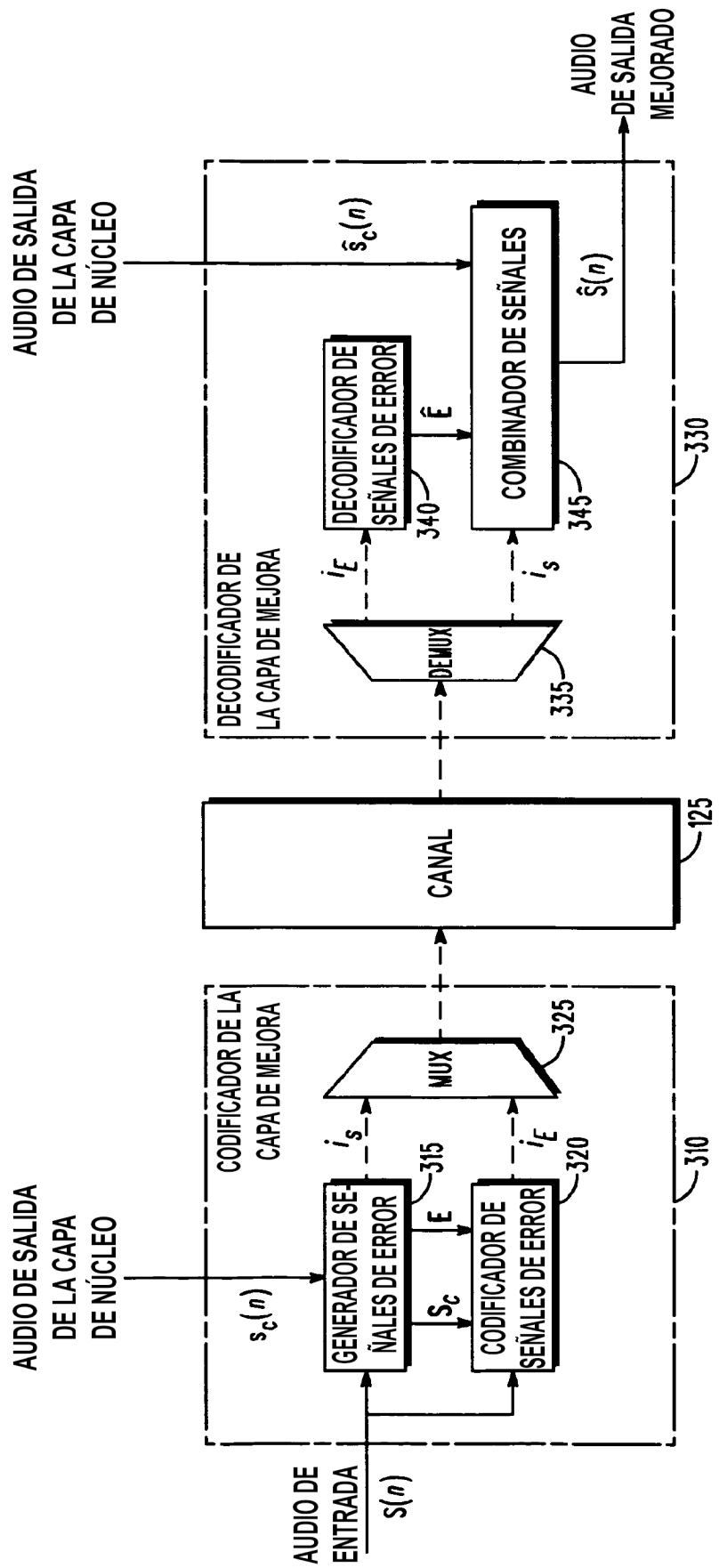
FIG. 1



200

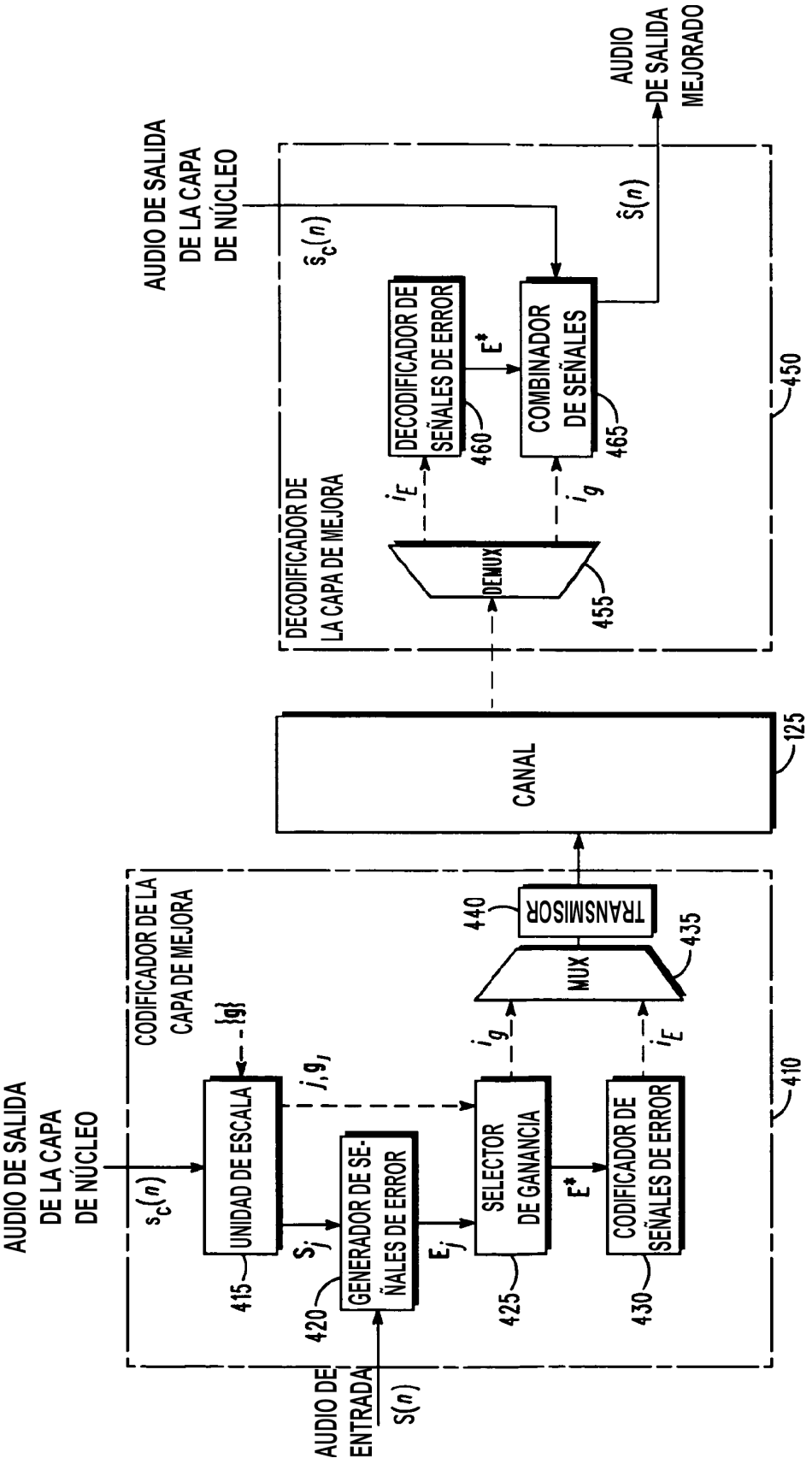
— TÉCNICA ANTERIOR —

FIG. 2



300
—TÉCNICA ANTERIOR—

FIG. 3


AUDIO DE SALIDA DE LA CAPA DE NÚCLEO
$s_c(n)$
CODIFICADOR DE LA CAPA DE MEJORA
{g}
UNIDAD DE ESCALA
415
j, g_j
S_j
GENERADOR DE SEÑALES DE ERROR
420
E_j
SELECTOR DE GANANCIA
425
i_g
E^*
CODIFICADOR DE SEÑALES DE ERROR
430
i_E
MUX
435
TRANSMISOR
440
410
AUDIO DE ENTRADA
$s(n)$
CANAL
125
DECODIFICADOR DE LA CAPA DE MEJORA
DEMUX
455
i_E
i_g
DECODIFICADOR DE SEÑALES DE ERROR
460
E^*
COMBINADOR DE SEÑALES
465
450
$\hat{s}_c(n)$
$\hat{s}(n)$
AUDIO DE SALIDA MEJORADO

400

FIG. 4

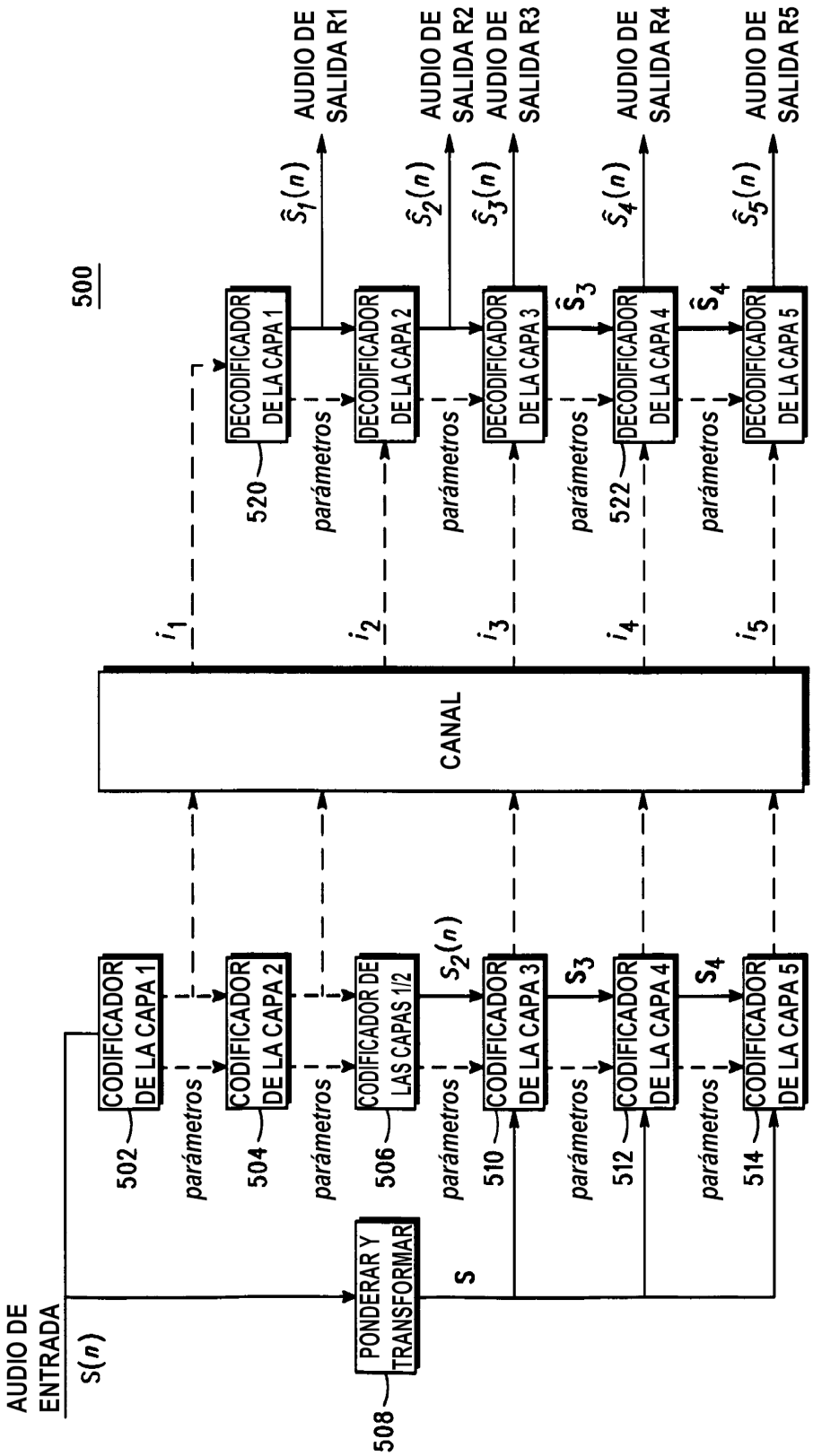

500
AUDIO DE ENTRADA
S(n)
CODIFICADOR DE LA CAPA 1
CODIFICADOR DE LA CAPA 2
CODIFICADOR DE LAS CAPAS 1/2
CODIFICADOR DE LA CAPA 3
CODIFICADOR DE LA CAPA 4
CODIFICADOR DE LA CAPA 5
PONDERAR Y TRANSFORMAR
parámetros
502
504
506
508
510
512
514
S
$s_2(n)$
s_3
s_4
CANAL
i_1
i_2
i_3
i_4
i_5
DECODIFICADOR DE LA CAPA 1
DECODIFICADOR DE LA CAPA 2
DECODIFICADOR DE LA CAPA 3
DECODIFICADOR DE LA CAPA 4
DECODIFICADOR DE LA CAPA 5
520
522
$\hat{s}_1(n)$
$\hat{s}_2(n)$
$\hat{s}_3(n)$
$\hat{s}_4(n)$
$\hat{s}_5(n)$
$\hat{s}_3$
$\hat{s}_4$
AUDIO DE SALIDA R1
AUDIO DE SALIDA R2
AUDIO DE SALIDA R3
AUDIO DE SALIDA R4
AUDIO DE SALIDA R5

FIG. 5

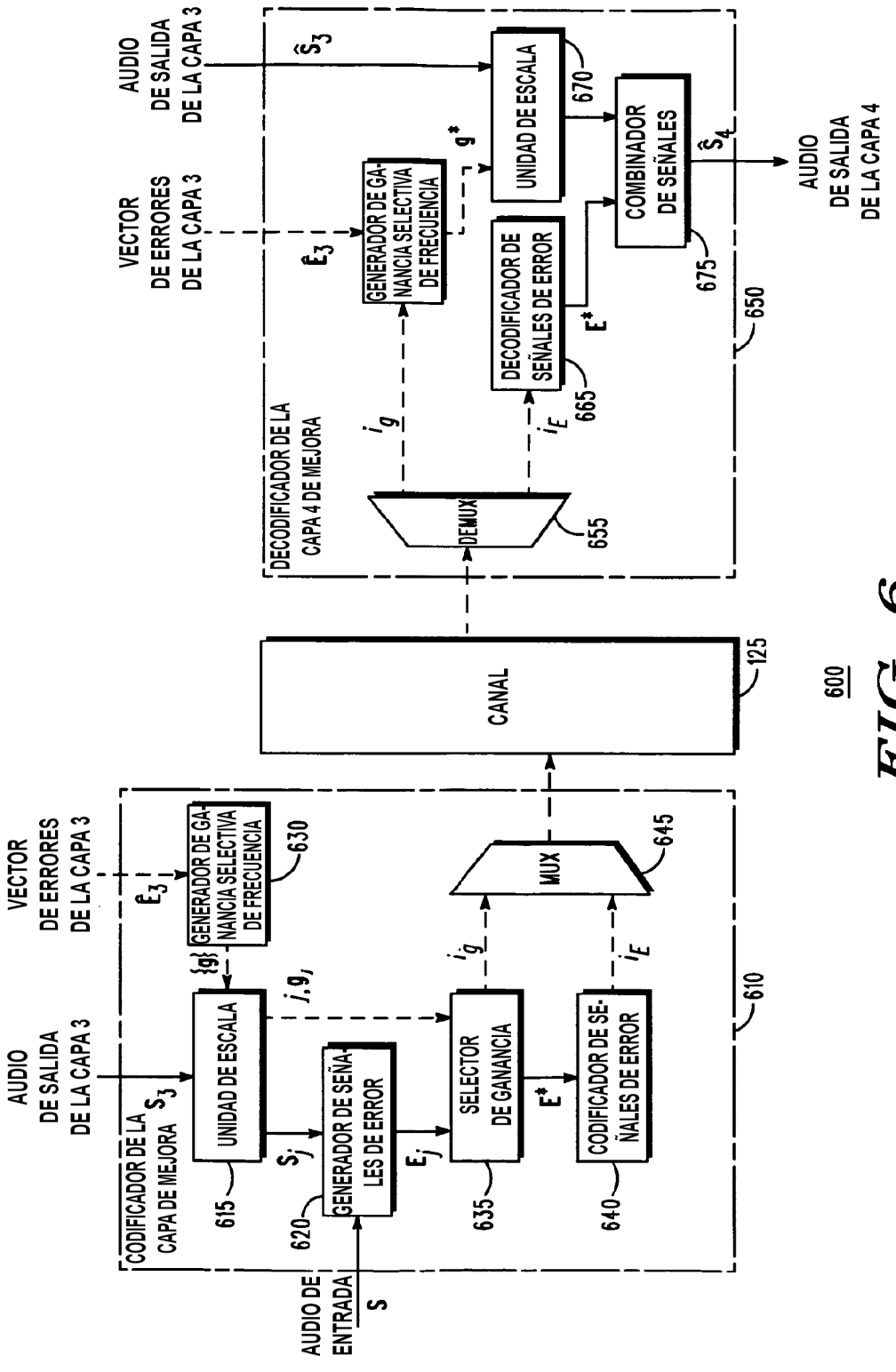

AUDIO DE SALIDA DE LA CAPA 3
VECTOR DE ERRORES DE LA CAPA 3
AUDIO DE ENTRADA
CODIFICADOR DE LA CAPA DE MEJORA
UNIDAD DE ESCALA
GENERADOR DE SEÑALES DE ERROR
GENERADOR DE GANANCIA SELECTIVA DE FRECUENCIA
SELECTOR DE GANANCIA
CODIFICADOR DE SEÑALES DE ERROR
MUX
CANAL
DEMUX
DECODIFICADOR DE LA CAPA 4 DE MEJORA
DECODIFICADOR DE SEÑALES DE ERROR
UNIDAD DE ESCALA
COMBINADOR DE SEÑALES
AUDIO DE SALIDA DE LA CAPA 4
600
610
615
620
625
630
635
640
645
650
655
665
670
675

FIG. 6

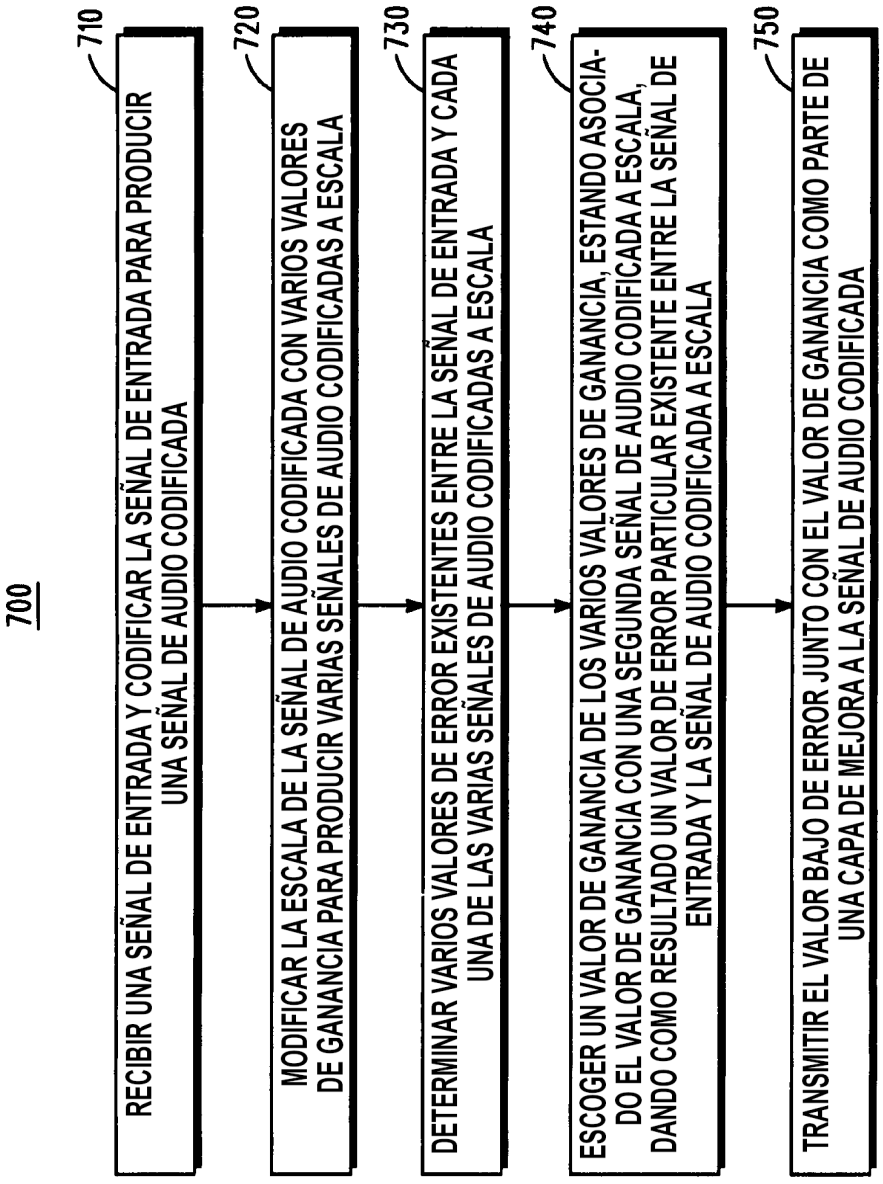

700
710
RECIBIR UNA SEÑAL DE ENTRADA Y CODIFICAR LA SEÑAL DE ENTRADA PARA PRODUCIR UNA SEÑAL DE AUDIO CODIFICADA
720
MODIFICAR LA ESCALA DE LA SEÑAL DE AUDIO CODIFICADA CON VARIOS VALORES DE GANANCIA PARA PRODUCIR VARIAS SEÑALES DE AUDIO CODIFICADAS A ESCALA
730
DETERMINAR VARIOS VALORES DE ERROR EXISTENTES ENTRE LA SEÑAL DE ENTRADA Y CADA UNA DE LAS VARIAS SEÑALES DE AUDIO CODIFICADAS A ESCALA
740
ESCOGER UN VALOR DE GANANCIA DE LOS VARIOS VALORES DE GANANCIA, ESTANDO ASOCIADO EL VALOR DE GANANCIA CON UNA SEGUNDA SEÑAL DE AUDIO CODIFICADA A ESCALA, DANDO COMO RESULTADO UN VALOR DE ERROR PARTICULAR EXISTENTE ENTRE LA SEÑAL DE ENTRADA Y LA SEÑAL DE AUDIO CODIFICADA A ESCALA
750
TRANSMITIR EL VALOR BAJO DE ERROR JUNTO CON EL VALOR DE GANANCIA COMO PARTE DE UNA CAPA DE MEJORA A LA SEÑAL DE AUDIO CODIFICADA

FIG. 7

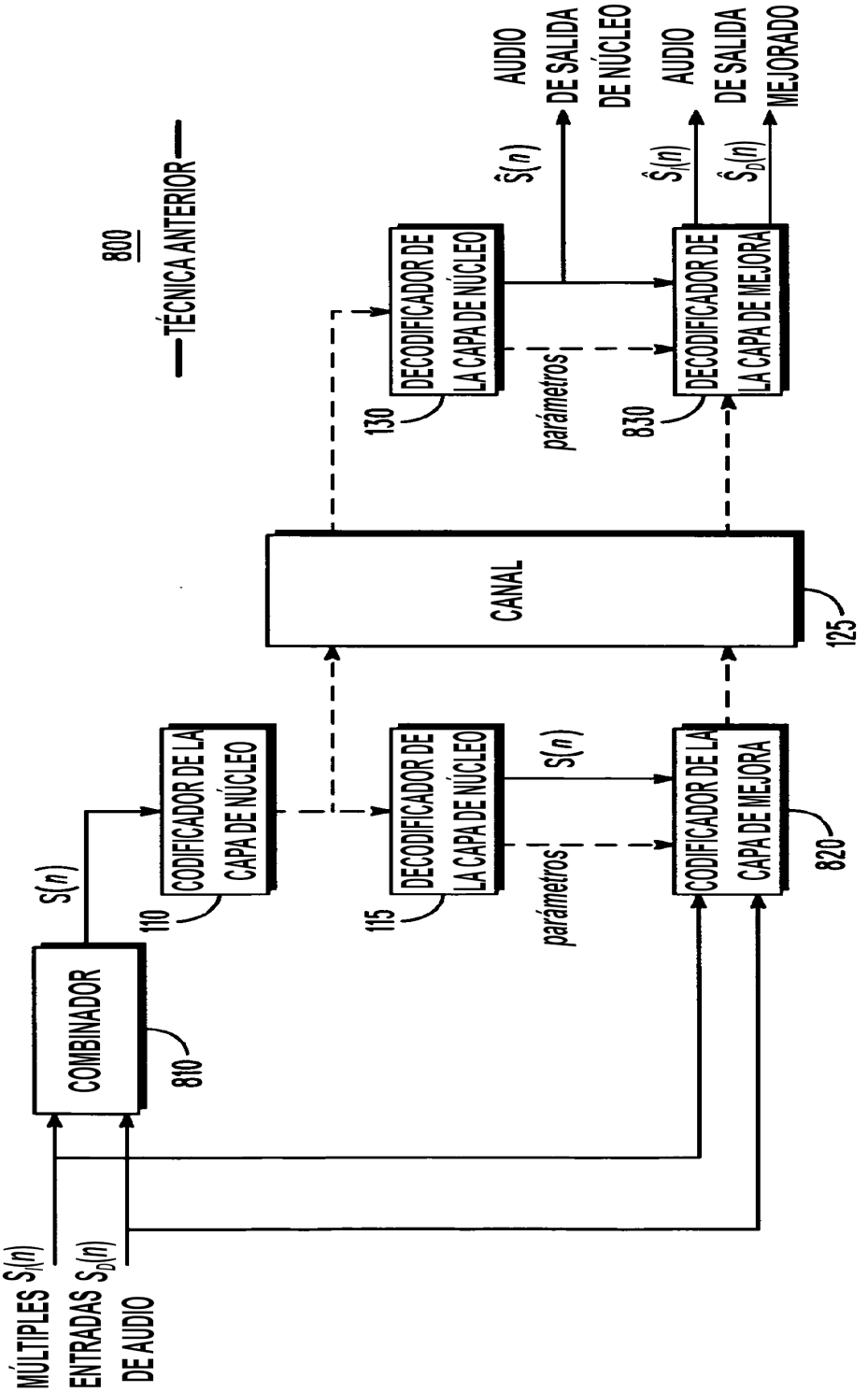
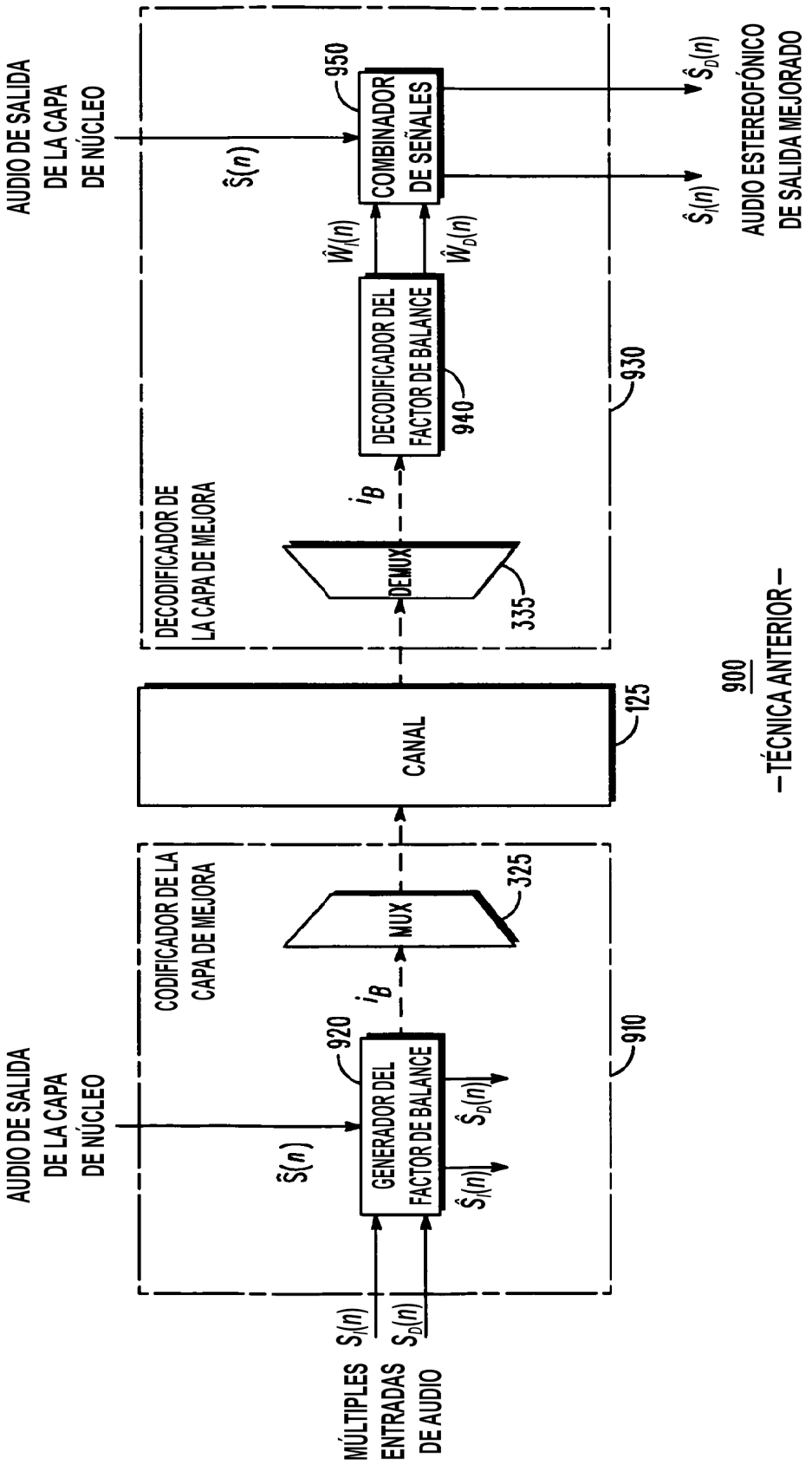

800
TÉCNICA ANTERIOR
MÚLTIPLES ENTRADAS DE AUDIO
$s_1(n)$
$s_D(n)$
COMBINADOR
810
$s(n)$
CODIFICADOR DE LA CAPA DE NÚCLEO
110
DECODIFICADOR DE LA CAPA DE NÚCLEO
115
parámetros
$s(n)$
CODIFICADOR DE LA CAPA DE MEJORA
820
CANAL
125
DECODIFICADOR DE LA CAPA DE NÚCLEO
130
$\hat{s}(n)$
AUDIO DE SALIDA DE NÚCLEO
parámetros
DECODIFICADOR DE LA CAPA DE MEJORA
830
$\hat{s}_1(n)$
$\hat{s}_D(n)$
AUDIO DE SALIDA MEJORADO

FIG. 8



900
–TÉCNICA ANTERIOR–

FIG. 9

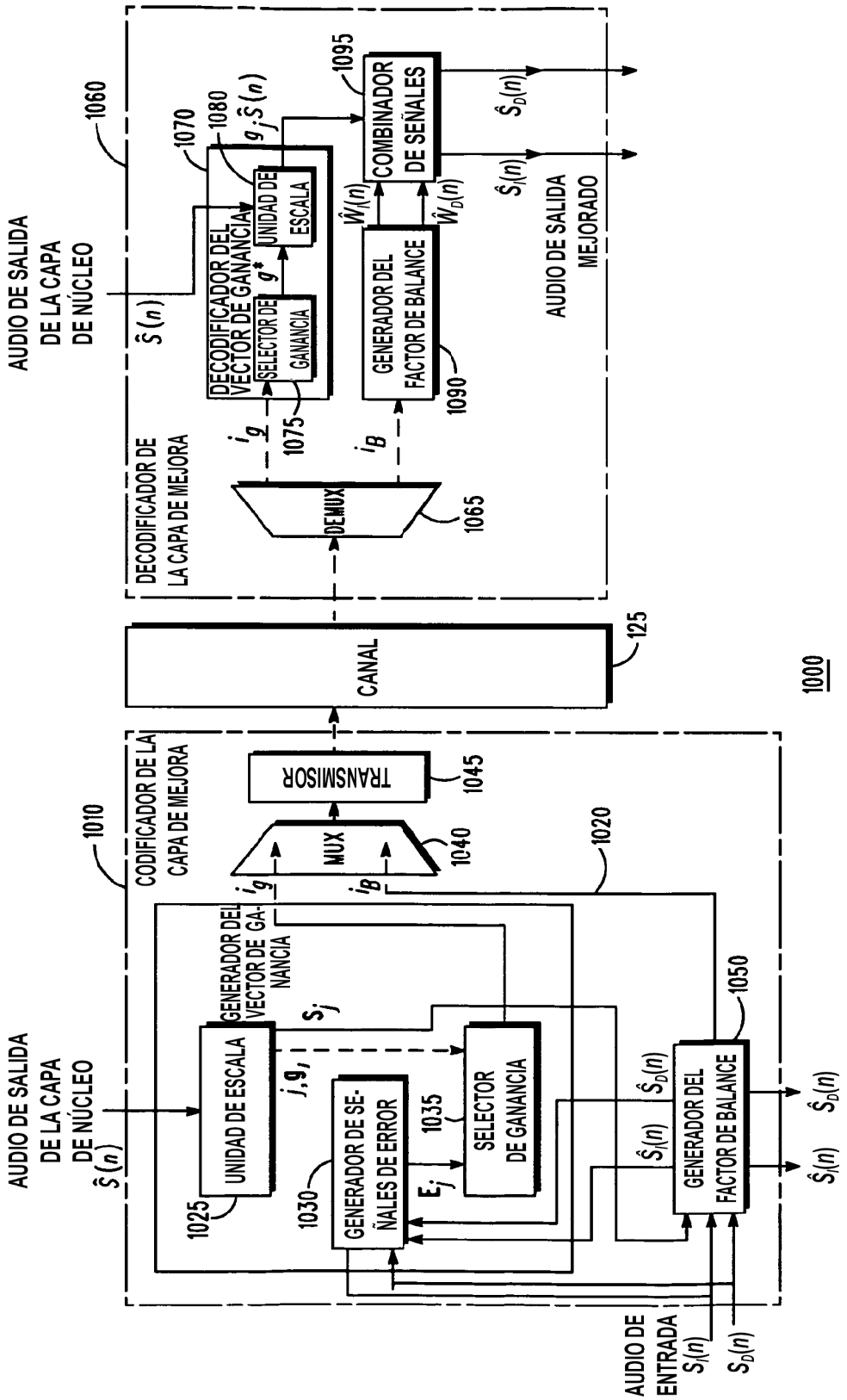

AUDIO DE SALIDA DE LA CAPA DE NÚCLEO
$\hat{S}(n)$
1010
CODIFICADOR DE LA CAPA DE MEJORA
GENERADOR DEL VECTOR DE GANANCIA
UNIDAD DE ESCALA
1025
j, g_j
S_j
GENERADOR DE SEÑALES DE ERROR
1030
E_j
1035
SELECTOR DE GANANCIA
MUX
1040
i_g
i_B
TRANSMISOR
1045
1020
AUDIO DE ENTRADA
$S_I(n)$
$S_D(n)$
GENERADOR DEL FACTOR DE BALANCE
1050
$\hat{S}_I(n)$
$\hat{S}_D(n)$
CANAL
125
1000
AUDIO DE SALIDA DE LA CAPA DE NÚCLEO
$\hat{S}(n)$
1060
DECODIFICADOR DE LA CAPA DE MEJORA
1070
DECODIFICADOR DEL VECTOR DE GANANCIA
SELECTOR DE GANANCIA
UNIDAD DE ESCALA
g^*
1080
$g_j \hat{S}(n)$
1075
i_g
i_B
DEMUX
1065
GENERADOR DEL FACTOR DE BALANCE
1090
$\hat{W}_I(n)$
$\hat{W}_D(n)$
1095
COMBINADOR DE SEÑALES
$\hat{S}_D(n)$
$\hat{S}_I(n)$
AUDIO DE SALIDA MEJORADO

FIG. 10

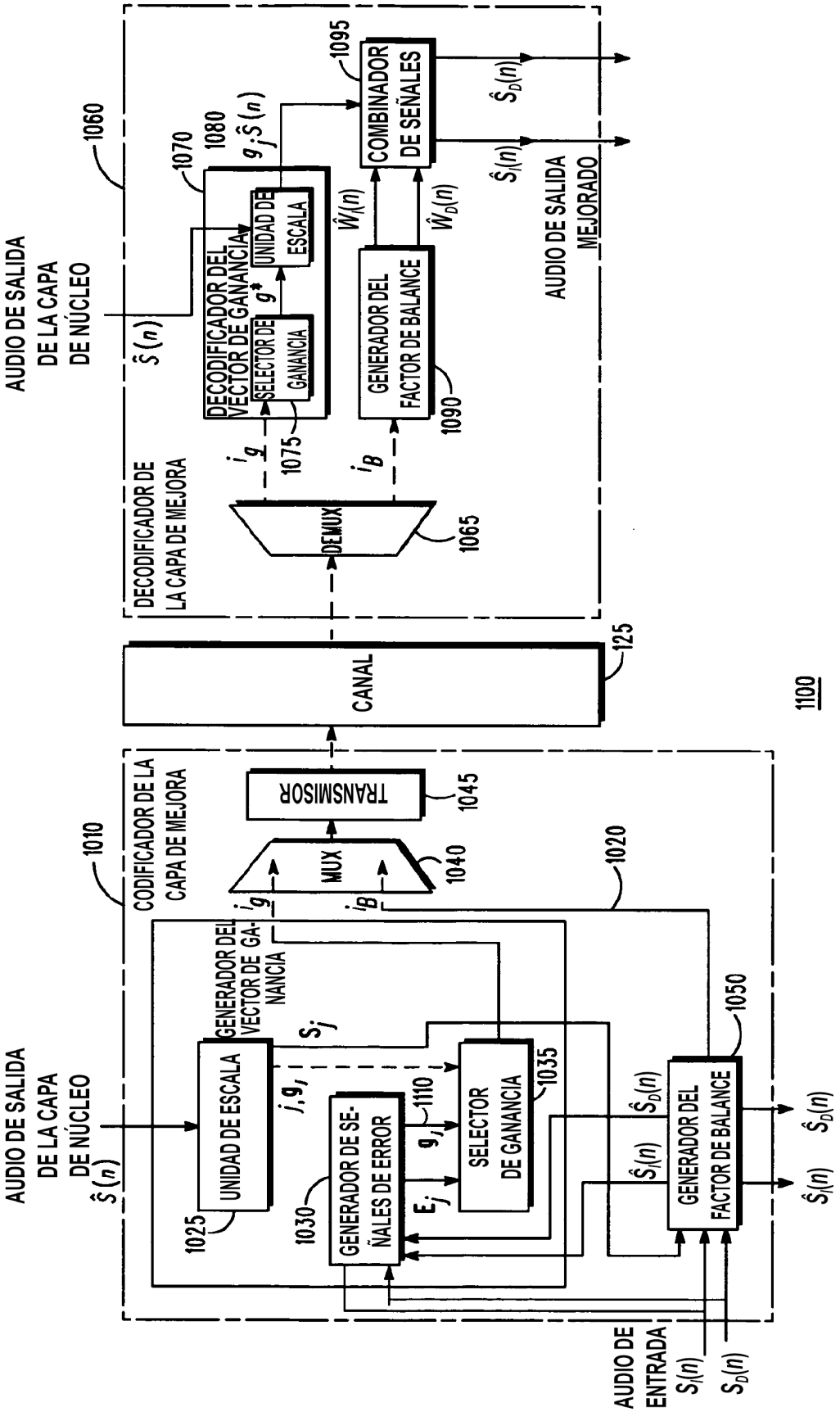

AUDIO DE SALIDA DE LA CAPA DE NÚCLEO
$\hat{s}(n)$
1060
DECODIFICADOR DE LA CAPA DE MEJORA
1070
DECODIFICADOR DEL VECTOR DE GANANCIA
SELECTOR DE GANANCIA
UNIDAD DE ESCALA
1080
$g_j \hat{s}(n)$
1095
COMBINADOR DE SEÑALES
$\hat{S}_D(n)$
$\hat{S}_I(n)$
AUDIO DE SALIDA MEJORADO
$\hat{W}_I(n)$
$\hat{W}_D(n)$
GENERADOR DEL FACTOR DE BALANCE
1090
i_g
1075
i_B
DEMUX
1065
CANAL
125
1100
CODIFICADOR DE LA CAPA DE MEJORA
1010
TRANSMISOR
1045
MUX
1040
1020
GENERADOR DEL VECTOR DE GANANCIA
S_j
j, g_j
UNIDAD DE ESCALA
1025
GENERADOR DE SEÑALES DE ERROR
1030
E_j
1110
g_j
SELECTOR DE GANANCIA
1035
1050
AUDIO DE ENTRADA
$S_I(n)$
$S_D(n)$

FIG. 11

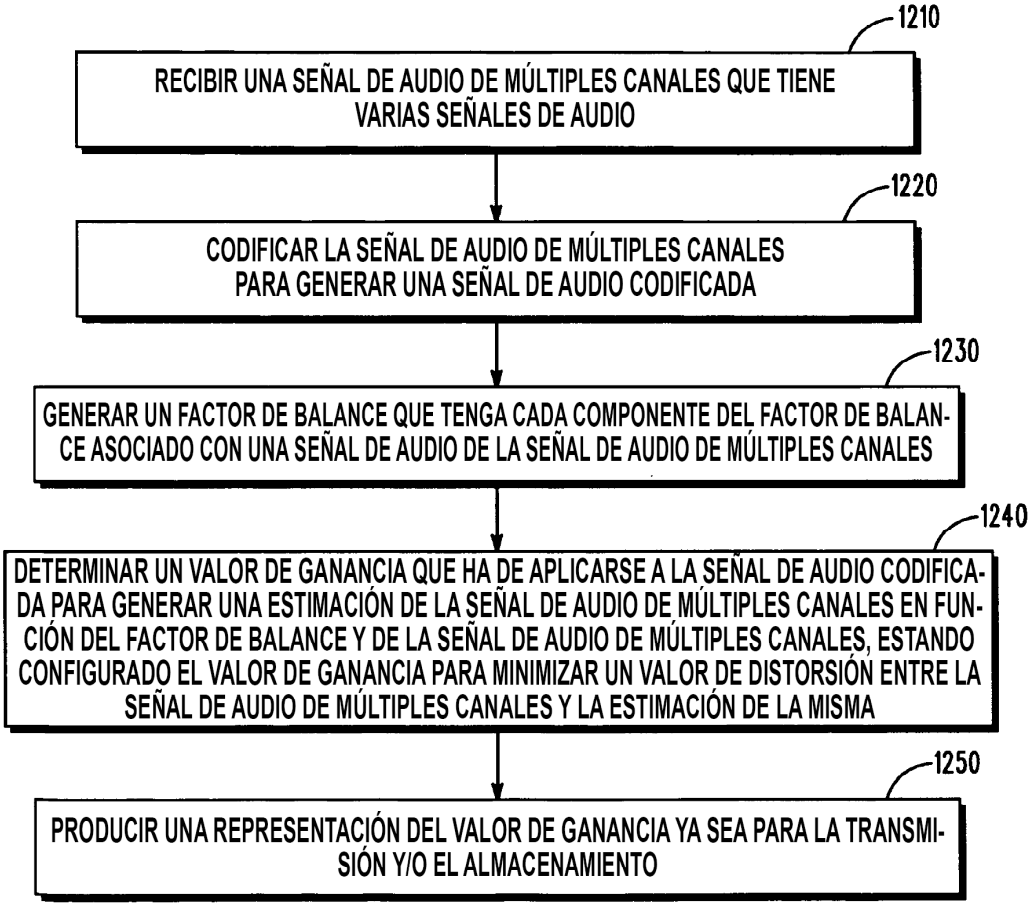

1210
RECIBIR UNA SEÑAL DE AUDIO DE MÚLTIPLES CANALES QUE TIENE VARIAS SEÑALES DE AUDIO
1220
CODIFICAR LA SEÑAL DE AUDIO DE MÚLTIPLES CANALES PARA GENERAR UNA SEÑAL DE AUDIO CODIFICADA
1230
GENERAR UN FACTOR DE BALANCE QUE TENGA CADA COMPONENTE DEL FACTOR DE BALANCE ASOCIADO CON UNA SEÑAL DE AUDIO DE LA SEÑAL DE AUDIO DE MÚLTIPLES CANALES
1240
DETERMINAR UN VALOR DE GANANCIA QUE HA DE APLICARSE A LA SEÑAL DE AUDIO CODIFICADA PARA GENERAR UNA ESTIMACIÓN DE LA SEÑAL DE AUDIO DE MÚLTIPLES CANALES EN FUNCIÓN DEL FACTOR DE BALANCE Y DE LA SEÑAL DE AUDIO DE MÚLTIPLES CANALES, ESTANDO CONFIGURADO EL VALOR DE GANANCIA PARA MINIMIZAR UN VALOR DE DISTORSIÓN ENTRE LA SEÑAL DE AUDIO DE MÚLTIPLES CANALES Y LA ESTIMACIÓN DE LA MISMA
1250
PRODUCIR UNA REPRESENTACIÓN DEL VALOR DE GANANCIA YA SEA PARA LA TRANSMISIÓN Y/O EL ALMACENAMIENTO

FIG. 12

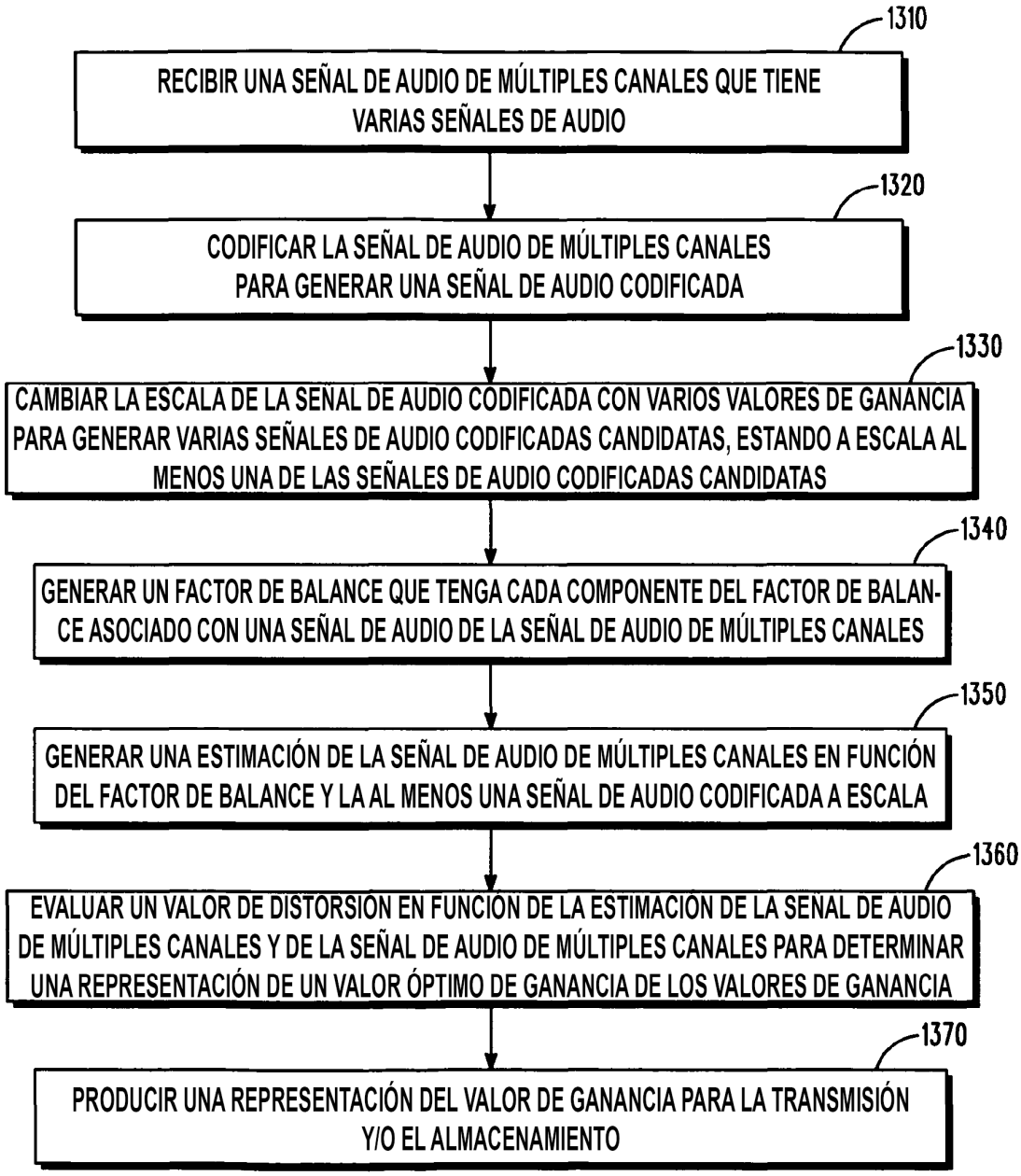
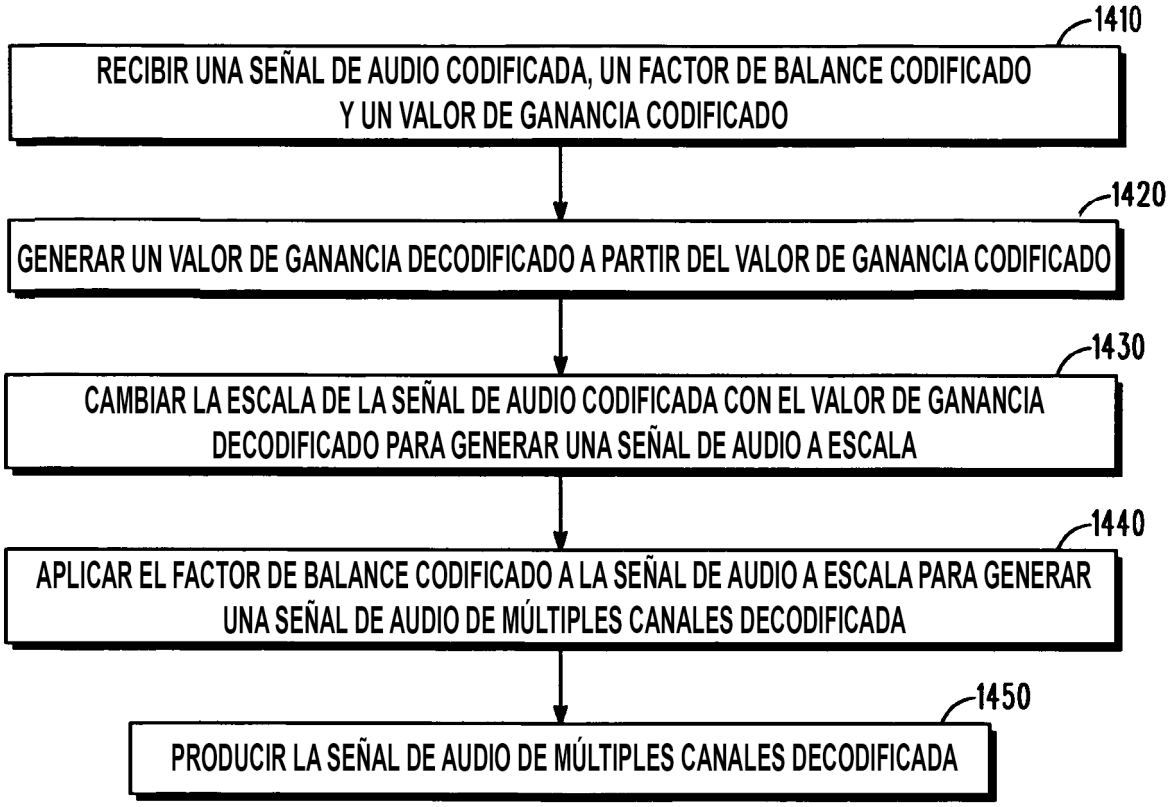

1310
RECIBIR UNA SEÑAL DE AUDIO DE MÚLTIPLES CANALES QUE TIENE VARIAS SEÑALES DE AUDIO
1320
CODIFICAR LA SEÑAL DE AUDIO DE MÚLTIPLES CANALES PARA GENERAR UNA SEÑAL DE AUDIO CODIFICADA
1330
CAMBIAR LA ESCALA DE LA SEÑAL DE AUDIO CODIFICADA CON VARIOS VALORES DE GANANCIA PARA GENERAR VARIAS SEÑALES DE AUDIO CODIFICADAS CANDIDATAS, ESTANDO A ESCALA AL MENOS UNA DE LAS SEÑALES DE AUDIO CODIFICADAS CANDIDATAS
1340
GENERAR UN FACTOR DE BALANCE QUE TENGA CADA COMPONENTE DEL FACTOR DE BALANCE ASOCIADO 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 FUNCIÓN DEL FACTOR DE BALANCE Y LA AL MENOS UNA SEÑAL DE AUDIO CODIFICADA A ESCALA
1360
EVALUAR UN VALOR DE DISTORSIÓN EN FUNCIÓN DE LA ESTIMACIÓN DE LA SEÑAL DE AUDIO DE MÚLTIPLES CANALES Y DE LA SEÑAL DE AUDIO DE MÚLTIPLES CANALES PARA DETERMINAR UNA REPRESENTACIÓN DE UN VALOR ÓPTIMO DE GANANCIA DE LOS VALORES DE GANANCIA
1370
PRODUCIR UNA REPRESENTACIÓN DEL VALOR DE GANANCIA PARA LA TRANSMISIÓN Y/O EL ALMACENAMIENTO
1300

FIG. 13


1410
RECIBIR UNA SEÑAL DE AUDIO CODIFICADA, UN FACTOR DE BALANCE CODIFICADO Y UN VALOR DE GANANCIA CODIFICADO
1420
GENERAR UN VALOR DE GANANCIA DECODIFICADO A PARTIR DEL VALOR DE GANANCIA CODIFICADO
1430
CAMBIAR LA ESCALA DE LA SEÑAL DE AUDIO CODIFICADA CON EL VALOR DE GANANCIA DECODIFICADO PARA GENERAR UNA SEÑAL DE AUDIO A ESCALA
1440
APLICAR EL FACTOR DE BALANCE CODIFICADO A LA SEÑAL DE AUDIO A ESCALA PARA GENERAR UNA SEÑAL DE AUDIO DE MÚLTIPLES CANALES DECODIFICADA
1450
PRODUCIR LA SEÑAL DE AUDIO DE MÚLTIPLES CANALES DECODIFICADA

1400

FIG. 14

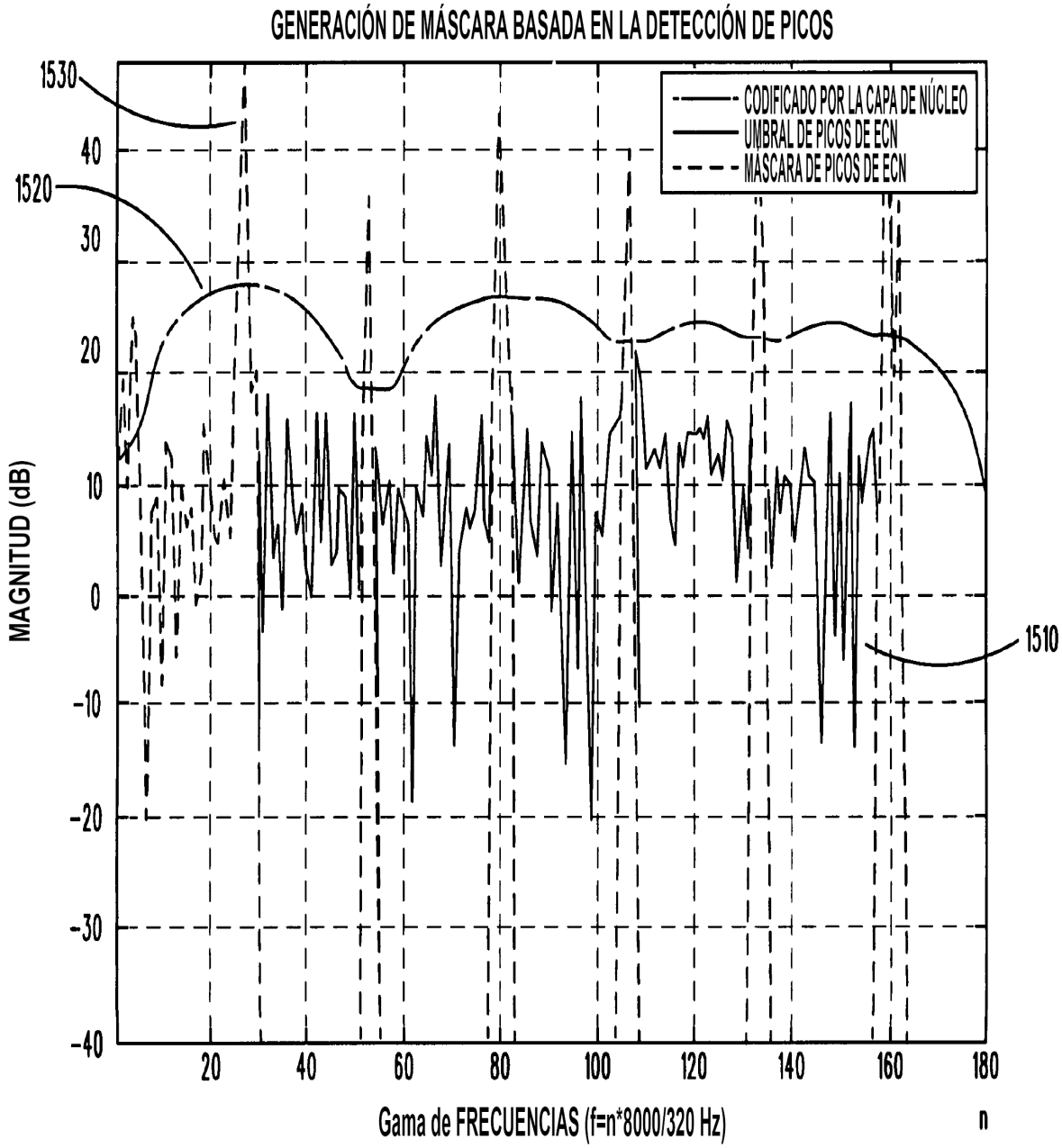

GENERACIÓN DE MÁSCARA BASADA EN LA DETECCIÓN DE PICOS
CODIFICADO POR LA CAPA DE NÚCLEO
UMBRAL DE PICOS DE ECN
MÁSCARA DE PICOS DE ECN
1530
1520
1510
MAGNITUD (dB)
40
30
20
10
0
-10
-20
-30
-40
20
40
60
80
100
120
140
160
180
Gama de FRECUENCIAS (f=n*8000/320 Hz)
n

FIG. 15

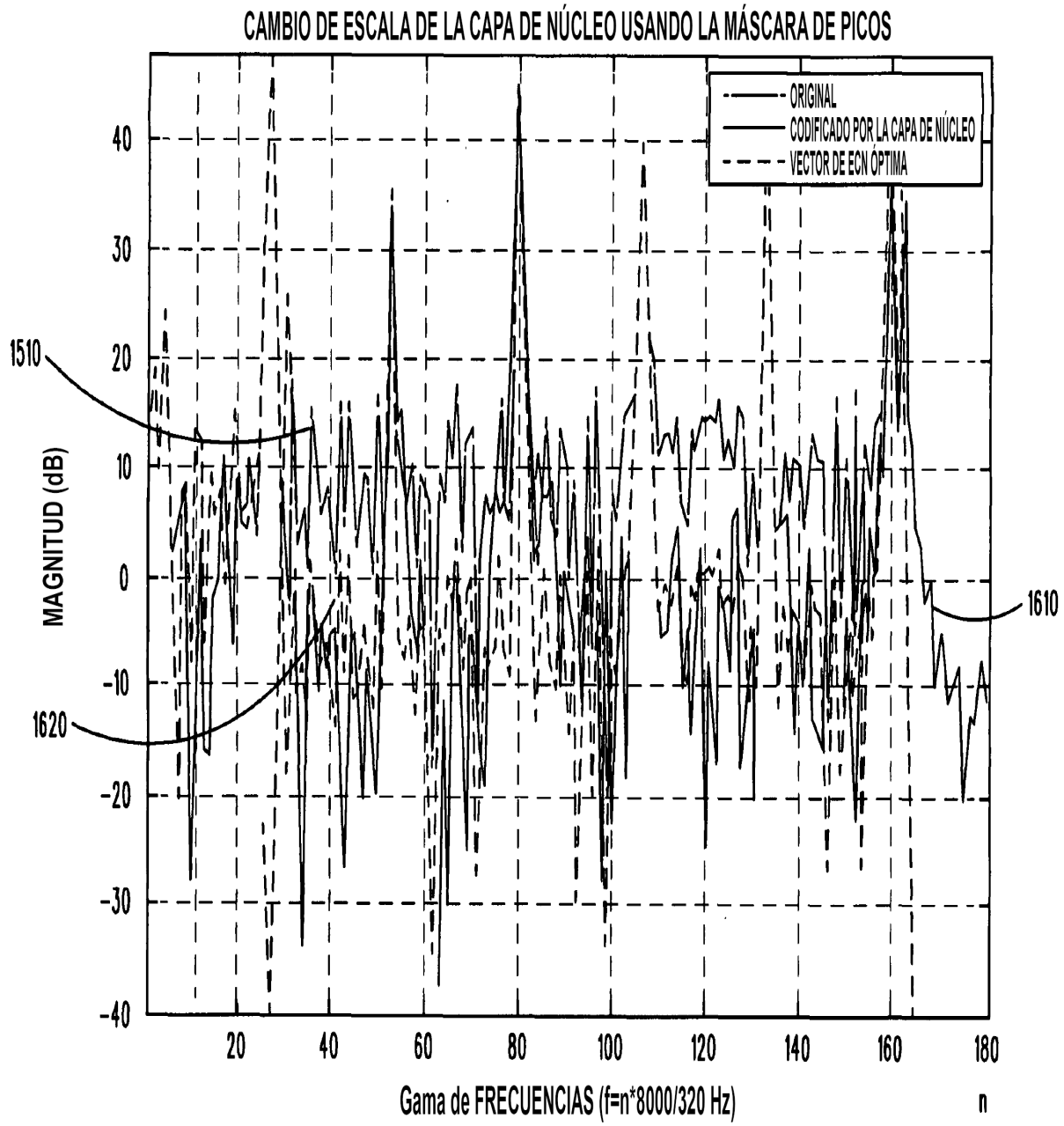

CAMBIO DE ESCALA DE LA CAPA DE NÚCLEO USANDO LA MÁSCARA DE PICOS
ORIGINAL
CODIFICADO POR LA CAPA DE NÚCLEO
VECTOR DE ECN ÓPTIMA
1510
1610
1620
MAGNITUD (dB)
40
30
20
10
0
-10
-20
-30
-40
20
40
60
80
100
120
140
160
180
Gama de FRECUENCIAS (f=n*8000/320 Hz)
n

FIG. 16

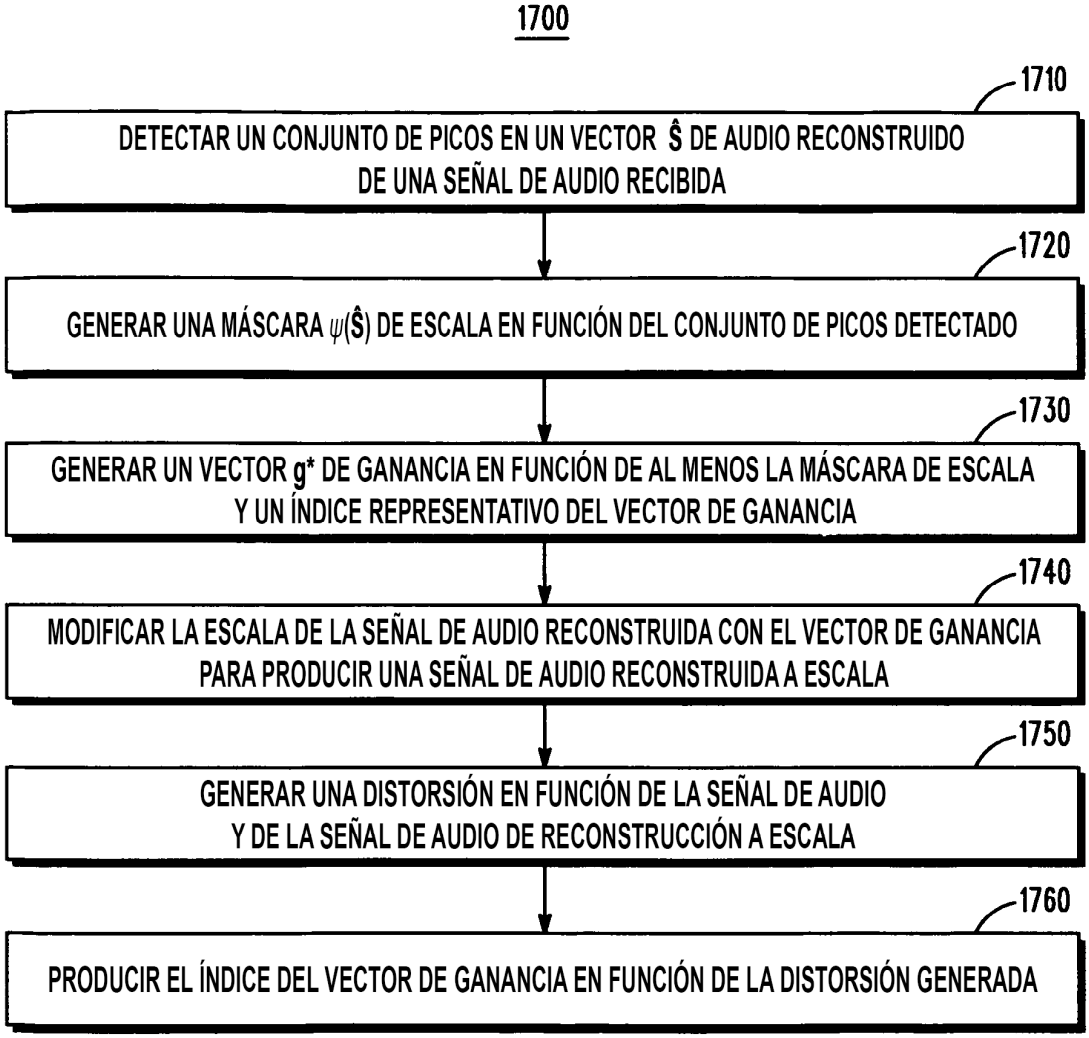

1700
1710
DETECTAR UN CONJUNTO DE PICOS EN UN VECTOR $\hat{S}$ DE AUDIO RECONSTRUIDO DE UNA SEÑAL DE AUDIO RECIBIDA
1720
GENERAR UNA MÁSCARA $\psi(\hat{S})$ DE ESCALA EN FUNCIÓN DEL CONJUNTO DE PICOS DETECTADO
1730
GENERAR UN VECTOR g^* DE GANANCIA EN FUNCIÓN DE AL MENOS LA MÁSCARA DE ESCALA Y UN ÍNDICE REPRESENTATIVO DEL VECTOR DE GANANCIA
1740
MODIFICAR LA ESCALA DE LA SEÑAL DE AUDIO RECONSTRUIDA CON EL VECTOR DE GANANCIA PARA PRODUCIR UNA SEÑAL DE AUDIO RECONSTRUIDA A ESCALA
1750
GENERAR UNA DISTORSIÓN EN FUNCIÓN DE LA SEÑAL DE AUDIO Y DE LA SEÑAL DE AUDIO DE RECONSTRUCCIÓN A ESCALA
1760
PRODUCIR EL ÍNDICE DEL VECTOR DE GANANCIA EN FUNCIÓN DE LA DISTORSIÓN GENERADA

FIG. 17

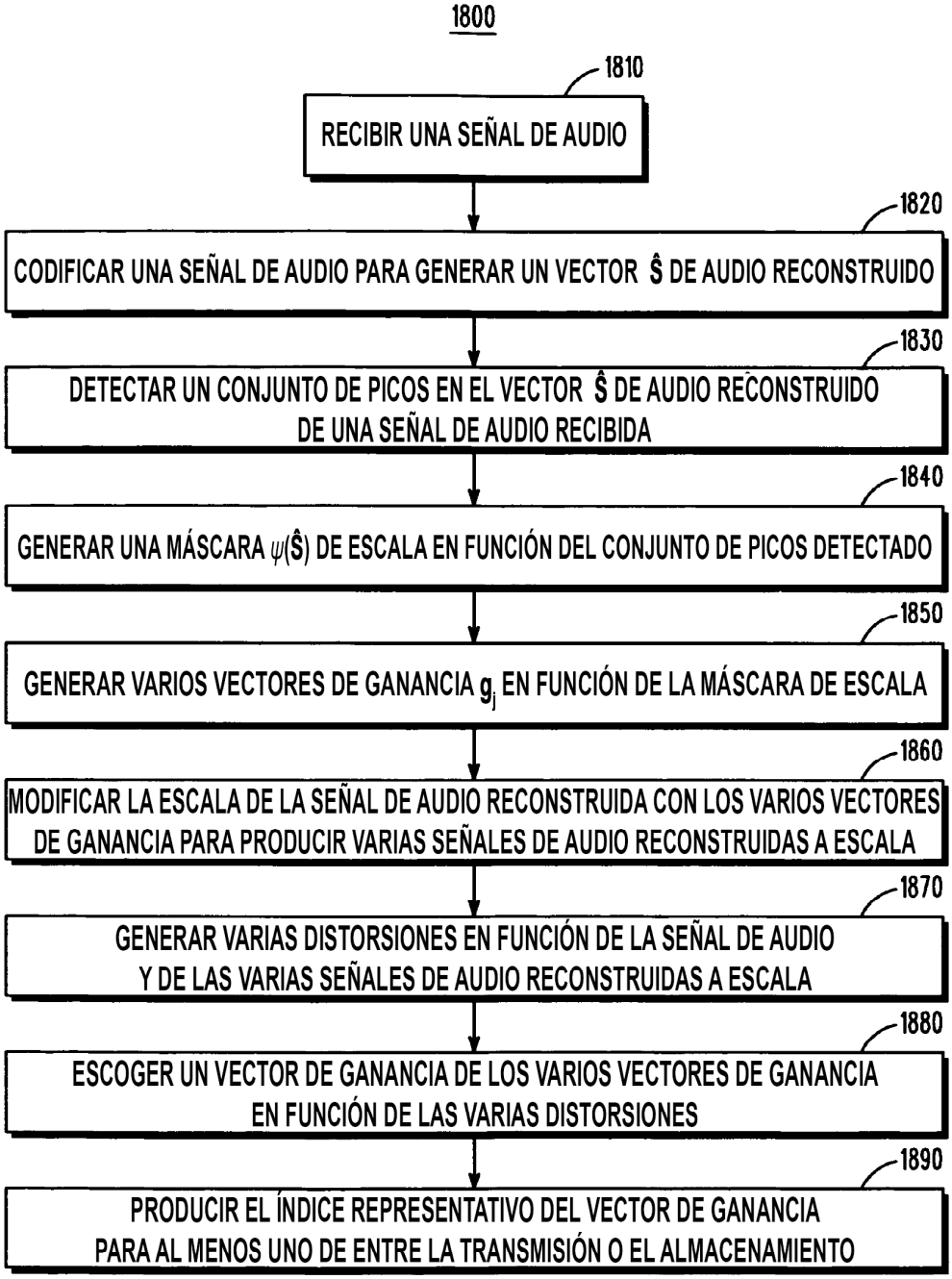

1800
1810
RECIBIR UNA SEÑAL DE AUDIO
1820
CODIFICAR UNA SEÑAL DE AUDIO PARA GENERAR UN VECTOR Ŝ DE AUDIO RECONSTRUIDO
1830
DETECTAR UN CONJUNTO DE PICOS EN EL VECTOR Ŝ DE AUDIO RECONSTRUIDO DE UNA SEÑAL DE AUDIO RECIBIDA
1840
GENERAR UNA MÁSCARA ψ(Ŝ) DE ESCALA EN FUNCIÓN DEL CONJUNTO DE PICOS DETECTADO
1850
GENERAR VARIOS VECTORES DE GANANCIA g_j EN FUNCIÓN DE LA MÁSCARA DE ESCALA
1860
MODIFICAR LA ESCALA DE LA SEÑAL DE AUDIO RECONSTRUIDA CON LOS VARIOS VECTORES DE GANANCIA PARA PRODUCIR VARIAS SEÑALES DE AUDIO RECONSTRUIDAS A ESCALA
1870
GENERAR VARIAS DISTORSIONES EN FUNCIÓN DE LA SEÑAL DE AUDIO Y DE LAS VARIAS SEÑALES DE AUDIO RECONSTRUIDAS A ESCALA
1880
ESCOGER UN VECTOR DE GANANCIA DE LOS VARIOS VECTORES DE GANANCIA EN FUNCIÓN DE LAS VARIAS DISTORSIONES
1890
PRODUCIR EL ÍNDICE REPRESENTATIVO DEL VECTOR DE GANANCIA PARA AL MENOS UNO DE ENTRE LA TRANSMISIÓN O EL ALMACENAMIENTO

FIG. 18

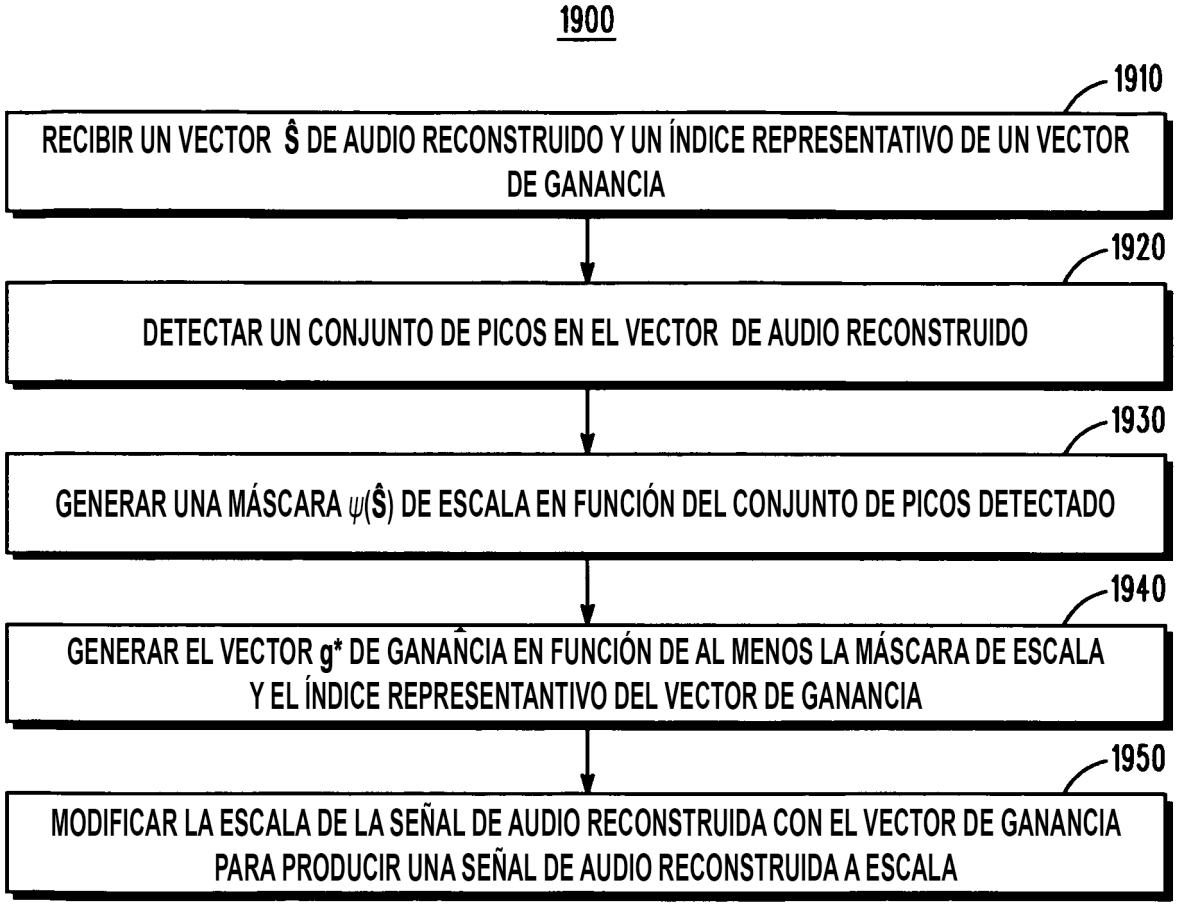

1900
1910
RECIBIR UN VECTOR $\hat{S}$ DE AUDIO RECONSTRUIDO Y UN ÍNDICE REPRESENTATIVO DE UN VECTOR DE GANANCIA
1920
DETECTAR UN CONJUNTO DE PICOS EN EL VECTOR DE AUDIO RECONSTRUIDO
1930
GENERAR UNA MÁSCARA $\psi(\hat{S})$ DE ESCALA EN FUNCIÓN DEL CONJUNTO DE PICOS DETECTADO
1940
GENERAR EL VECTOR g* DE GANANCIA EN FUNCIÓN DE AL MENOS LA MÁSCARA DE ESCALA Y EL ÍNDICE REPRESENTANTIVO DEL VECTOR DE GANANCIA
1950
MODIFICAR LA ESCALA DE LA SEÑAL DE AUDIO RECONSTRUIDA CON EL VECTOR DE GANANCIA PARA PRODUCIR UNA SEÑAL DE AUDIO RECONSTRUIDA A ESCALA

FIG. 19