

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 430 639**

51 Int. Cl.:

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 09799782 (9)**

97 Fecha y número de publicación de la concesión europea: **18.09.2013** **EP 2382621**

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 345165

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

21.11.2013

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

ES 2 430 639 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

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

Referencia a las solicitudes relacionadas

- 5 La presente solicitud está relacionada con las siguientes solicitudes de patente europeas de propiedad común junto con esta solicitud por Motorola, Inc.:
- Solicitud EP 2382 627 A0, titulada "SELECTIVE SCALING MASK COMPUTATION BASED ON PEAK DETECTION";
- Solicitud EP 2382 622 A0, titulada "METHOD AND APPARATUS FOR GENERATING AN ENHANCEMENT LAYER WITHIN A MULTIPLE-CHANNEL AUDIO CODING SYSTEM" (Expediente del Agente N° CS36627AUD); y
- 10 Solicitud EP 2382 626 A0, titulada "SELECTIVE SCALING MASK COMPUTATION BASED ON PEAK DETECTION".

Campo de la descripción

La presente invención se refiere, en general, a sistemas de comunicación y, más particularmente, a codificar señales de habla y audio en tales sistemas de comunicación.

Antecedentes

- 15 La compresión de señales de habla y audio es bien conocida. Se requiere generalmente compresión para transmitir señales eficientemente sobre un canal de comunicaciones, o para almacenar señales comprimidas en un dispositivo de medios digital, tal como un dispositivo de memoria de estado sólido o disco duro de ordenador. Aunque hay muchas técnicas de compresión (o "codificación"), un método que ha permanecido muy popular para codificación de habla digital se conoce como Predicción Lineal Excitada de Código (CELP), que es uno de una familia de algoritmos
- 20 de codificación de "análisis por síntesis". Análisis por síntesis se refiere de manera general a un proceso de codificación por el 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 analizan para distorsión. Un conjunto de parámetros que produce la distorsión más baja entones o bien se transmite o bien se almacena, y eventualmente se usa para reconstruir una estimación de la señal de entrada original. CELP es un método de análisis por síntesis particular que
- 25 usa uno o más libros de códigos cada uno que comprende esencialmente conjuntos de vectores de código que se recuperan del libro de códigos en respuesta a un índice de libro de códigos.

En los codificadores CELP modernos, hay un problema con el mantenimiento de una reproducción de habla y audio de elevada calidad a tasas de datos razonablemente bajas. Esto es verdadero especialmente para música u otras señales de audio genéricas que no cumplen muy bien el modelo de habla CELP. En este caso, el desajuste del

30 modelo puede causar una calidad de audio severamente degradada que puede ser inaceptable para un usuario final del equipo que emplea tales métodos. Por lo tanto, sigue habiendo una necesidad de mejorar el rendimiento de los codificadores de habla tipo CELP a tasas de bit bajas, especialmente para música y otras entradas no de tipo habla.

La EP 1 818 911 (A1) describe un dispositivo de codificación de sonido que tiene una estructura escalable monoaural/estéreo y capaz de codificar eficientemente sonido estéreo incluso cuando es pequeña la correlación

35 entre las señales de canal de una señal estéreo.

Los objetivos antes mencionados se resuelven por las reivindicaciones de la presente invención.

Breve descripción de los dibujos

Las figuras anexas, donde números de referencia iguales se refieren a elementos idénticos o de similar funcionalidad en todas las vistas separadas, las cuales junto con la descripción detallada de más adelante se

40 incorporan en y forman parte de la especificación y sirven para ilustrar además 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 de habla/audio integrado de la técnica anterior.

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

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

45 La FIG. 4 es un diagrama de bloques de un codificador y descodificador de capa de mejora.

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

La FIG. 6 es un diagrama de bloques de un codificador y descodificador de 4 capas.

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 de habla/audio integrado de la técnica anterior.

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

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

5 La FIG. 11 es un diagrama de bloques de un codificador y descodificador de capa de mejora, según diversas realizaciones.

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

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

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

La FIG. 15 es un gráfico de frecuencia de detección de pico en base a generación de máscara, según diversas realizaciones.

15 La FIG. 16 es un gráfico de frecuencia de escalado de capa central que usa generación de máscara pico, según diversas realizaciones.

Las FIG. 17-19 son diagramas de flujo que ilustran la metodología para la codificación y descodificación usando generación de máscara basada en detección de pico, según diversas realizaciones.

20 Los expertos apreciarán que los elementos en las figuras se ilustran para simplicidad y claridad y no han sido dibujados necesariamente a escala. Por ejemplo, las dimensiones de algunos de los elementos en las figuras pueden estar exageradas con respecto a otros elementos para ayudar a mejorar la compresión de diversas realizaciones. Además, la descripción y los dibujos no requieren necesariamente el orden ilustrado. Se apreciará además que se pueden describir o representar ciertas acciones y/o pasos en un orden particular de aparición aunque los expertos en la técnica entenderán que tal especificidad con respecto a la secuencia no se requiere realmente. Los componentes del aparato y método se han representado donde sea adecuado mediante símbolos convencionales en los dibujos, mostrando solamente aquellos detalles específicos que son pertinentes para la comprensión de las diversas realizaciones para no oscurecer la descripción con detalles que serán fácilmente evidentes a aquellos expertos en la técnica que tienen el beneficio de la descripción de la presente memoria. De esta manera, se apreciará que para simplicidad y claridad de ilustración, los elementos comunes y bien conocidos que son útiles o necesarios en una realización comercialmente factible pueden no ser representados a fin de facilitar una vista menos obstruida de estas diversas realizaciones.

Descripción detallada

35 A fin de abordar la necesidad anteriormente mencionada, se describe en la presente memoria un método y aparato para generar una capa de mejora dentro de un sistema de codificación de audio. Durante la operación una señal de entrada a ser codificada se recibe y codifica para producir una señal de audio codificada. La señal de audio codificada entonces se escala con una pluralidad de valores de ganancia para producir una pluralidad de señales de audio codificadas escaladas, cada una que tiene un valor de ganancia asociado y se determina que existen una pluralidad de valores de error entre la señal de entrada y cada una de la pluralidad de señales de audio codificadas escaladas. Se elige entonces un valor de ganancia que está asociado con una señal de audio codificada escalada provocando un valor de error bajo que existe entre la señal de entrada y la señal de audio codificada escalada. Finalmente, el valor de error bajo se transmite junto con el valor de ganancia como parte de una capa de mejora para la señal de audio codificada.

45 Un sistema de compresión de habla/audio integrado de la técnica anterior se muestra en la FIG. 1. El audio de entrada $s(n)$ se procesa en primer lugar por un codificador de capa central 120, que para estos propósitos puede ser un algoritmo de codificación de habla de tipo CELP. El flujo de bits codificado se transmite al canal 125, además de ser introducido a un descodificador de capa central local 115, donde se genera la señal de audio central reconstruida $s_c(n)$. El codificador de capa de mejora 120 se usa entonces para codificar información adicional en base a alguna comparación de señales $s(n)$ y $s_c(n)$, y puede usar opcionalmente parámetros del descodificador de capa central 115. Como en el descodificador de capa central 115, el descodificador de capa central 130 convierte los parámetros de flujo de bits de capa central a una señal de audio de capa central $\hat{s}_c(n)$. El descodificador de capa de mejora 50 entonces usa el flujo de bits de capa de mejora desde el canal 125 y la señal $\hat{s}_c(n)$ para producir la señal de salida de audio mejorada $\hat{s}(n)$.

La ventaja primaria de tal sistema de codificación integrado es que un canal particular 125 puede no ser capaz de soportar consistentemente el requerimiento de ancho de banda asociado con algoritmos de codificación de audio de alta calidad. Un codificador integrado, no obstante, permite que un flujo de bits parcial sea recibido (por ejemplo, solamente el flujo de bits de capa central) desde el canal 125 para producir, por ejemplo, solamente el audio de salida central cuando se pierde o daña el flujo de bits de la capa de mejora. No obstante, hay compromisos de calidad entre codificadores integrados frente a no integrados, y también entre diferentes objetivos de optimización de codificación integrada. Es decir, la codificación de capa de mejora de mayor calidad puede ayudar a lograr un mejor equilibrio entre las capas central y de mejora, y también reducir la tasa de datos total para mejores características de transmisión (por ejemplo, congestión reducida), que puede provocar tasas de error de paquetes inferiores para las capas de mejora.

Un ejemplo más detallado de un codificador de capa de mejora de la técnica anterior 120 se da en la FIG. 2. Aquí, el generador de señal de error 210 está compuesto de una señal de diferencia ponderada que se transforma en el dominio de MDCT (Transformada de Coseno Discreto Modificada) para procesamiento por el codificador de señal de error 220. La señal de error E se da como:

$$E = \text{MDCT}\{W(s - s_c)\}, \quad (1)$$

donde W es una matriz de ponderación perceptiva basada en los coeficientes de filtro de LP (Predicción Lineal) $A(z)$ del descodificador de capa central 115, s es un vector (es decir, una trama) de muestras de la señal de audio de entrada $s(n)$, y s_c es el vector correspondiente de muestras del descodificador de capa central 115. Un ejemplo de proceso MDCT se describe en la Recomendación G.729.1 de la ITU-T. La señal de error E se procesa entonces por el codificador de señal de error 220 para producir la palabra de código i_E , que se transmite posteriormente al canal 125. Para este ejemplo, es importante señalar que el codificador de señal de error 120 se presenta solamente con una señal de error E y saca una palabra de código asociada i_E . La razón para esto llegará a ser evidente más tarde.

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

$$\hat{s} = s_c + W^{-1} \text{MDCT}^{-1}\{\hat{E}\}, \quad (2)$$

donde MDCT^{-1} es la MDCT inversa (que incluye inserción de solape), y W^{-1} es la matriz de ponderación de percepción inversa.

Otro ejemplo de un codificador de capa de mejora se muestra en la FIG. 3. Aquí, la generación de la señal de error E por el generador de señal de error 315 implica escalado previo adaptativo, en el que se realiza alguna modificación a la salida de audio de la capa central $s_c(n)$. Este proceso provoca que algún número de bits sea generado, lo cual se muestra en el codificador de capa de mejora 120 como la palabra de código i_s .

Adicionalmente, el codificador de capa de mejora 120 muestra la señal de audio de entrada $s(n)$ y el audio de salida de la capa central transformado S_c que se introduce al codificador de señal de error 320. Estas señales se usan para construir un modelo psicoacústico para codificación mejorada de la señal de error de capa de mejora E . Las palabras de código i_s e i_E entonces se multiplexan por el MUX 325, y entonces se envían al canal 125 para descodificación posterior por el descodificador de capa de mejora 135. El flujo de bits codificado se recibe por el desmultiplexor 335, que separa el flujo de bits en componentes i_s e i_E . La palabra de código i_E entonces se usa por el descodificador de señal de error 340 para reconstruir la señal de error de capa de mejora \hat{E} . El combinador de señal 345 escala la señal $\hat{s}_c(n)$ de alguna manera usando bits de escalado i_s , y entonces combina el resultado con la señal de error de capa de mejora \hat{E} para producir la señal de salida de audio mejorada $\hat{s}(n)$.

Una primera realización de la presente invención se da en la FIG. 4. Esta figura muestra el codificador de capa de mejora 410 que recibe la señal de salida de capa central $s_c(n)$ mediante la unidad de escalado 415. Se usa un conjunto predeterminado de ganancias $\{g\}$ para producir una pluralidad de señales de salida de capa central escaladas $\{S_j\}$, donde g_j y S_j son las candidatas de orden j de los conjuntos respectivos. Dentro de la unidad de escalado 415, la primera realización procesa la señal $s_c(n)$ en el dominio (MDCT) como:

$$S_j = G_j \times \text{MDCT}\{W s_c\}, \quad 0 \leq j < M, \quad (3)$$

donde W puede ser alguna matriz de ponderación perceptiva, s_c es un vector de muestras del descodificador de capa central 115, la MDCT es una operación bien conocida en la técnica, y G_j puede ser una matriz de ganancia

formada utilizando un candidato de vector de ganancia \mathbf{g}_j , y donde M es el número de candidatos de vector de ganancia. En la primera realización, \mathbf{G}_j usa el vector \mathbf{g}_j como la diagonal y ceros en cualquier otro sitio (es decir, una matriz diagonal), aunque existen muchas posibilidades. Por ejemplo, \mathbf{G}_j puede ser una matriz de banda, o puede ser incluso una simple cantidad escalar multiplicada por la matriz de identidad i . Alternativamente, puede haber alguna ventaja de dejar la señal \mathbf{S}_j en el dominio del tiempo o puede haber casos donde es ventajoso transformar el audio a un dominio diferente, tal como el dominio de Transformada Discreta de Fourier (DFT). Muchas de tales transformaciones son bien conocidas en la técnica. En estos casos, la unidad de escalado puede sacar la \mathbf{S}_j apropiada basada en el dominio de vector respectivo.

Pero en cualquier caso, la razón primaria para escalar el audio de salida de capa central es compensar el desajuste del modelo (o alguna otra deficiencia de codificación) que puede causar diferencias significativas entre la señal de entrada y el códec de la capa central. Por ejemplo, si la señal de audio de entrada es ante todo una señal de música y el códec de la capa central está basado en un modelo de habla, entonces la salida de la capa central puede contener características de señal severamente distorsionadas, en cuyo caso, es beneficioso desde una perspectiva de calidad de sonido reducir selectivamente la energía de esta componente de señal anterior a aplicar codificación suplementaria de la señal por medio de una o más capas de mejora.

El vector candidato de audio de la capa central escalada con ganancia \mathbf{S}_j y el audio de entrada $s(n)$ se pueden usar entonces como entrada al generador de señal de error 420. En una realización ejemplar, la señal de audio de entrada $s(n)$ se convierte al vector \mathbf{S} de manera que \mathbf{S} y \mathbf{S}_j están alineadas en consecuencia. Es decir, el vector \mathbf{s} que representa $s(n)$ está alineado en tiempo (fase) con s_c , y se pueden aplicar de esa manera las operaciones correspondientes en esta realización:

$$\mathbf{E}_j = \text{MDCT}\{\mathbf{W}\mathbf{s}\} - \mathbf{S}_j; \quad 0 \leq j < M. \quad (4)$$

Esta expresión produce una pluralidad de vectores de señal de error \mathbf{E}_j que representan la diferencia ponderada entre el audio de entrada y el audio de salida de la capa central escalada con ganancia en el dominio espectral de MDCT. En otras realizaciones, donde se consideran diferentes dominios, se puede modificar la expresión anterior en base al dominio de procesamiento respectivo.

El selector de ganancia 425 se usa entonces para evaluar la pluralidad de vectores de señal de error \mathbf{E}_j , según la primera realización de la presente invención, para producir un vector de error óptimo \mathbf{E}^* , un parámetro de ganancia óptimo g^* , y posteriormente, un índice de ganancia correspondiente j_g . El selector de ganancia 425 puede usar una variedad de métodos para determinar los parámetros óptimos, \mathbf{E}^* y g^* , que pueden implicar métodos de bucle cerrado (por ejemplo, minimización de una métrica de distorsión), métodos de bucle abierto (por ejemplo, clasificación heurística, estimación de rendimiento del modelo, etc.), o una combinación de ambos métodos. En la realización ejemplar, se puede usar una métrica de distorsión desviada, que se da como la diferencia de energía desviada entre el vector de señal de audio original \mathbf{S} y el vector de señal reconstruido compuesto:

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

donde $\hat{\mathbf{E}}_j$ puede ser la estimación cuantificada del vector de señal de error \mathbf{E}_j , y β_j puede ser un término de desviación que se usa para complementar la decisión de elegir el índice de error de ganancia óptimo perceptivamente j^* . Un método ejemplar para cuantificación de vector de un vector de señal se da en la Solicitud de Patente de EE.UU. N° de Serie 11/531122, titulada APPARATUS AND METHOD FOR LOW COMPLEXITY COMBINATORIAL CODING OF SIGNALS, aunque son posibles muchos otros métodos. Reconociendo que $\mathbf{E}_j = \mathbf{S} - \mathbf{S}_j$, la ecuación (5) se puede reescribir como:

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

En esta expresión, el término $\varepsilon_j = \|\mathbf{E}_j - \hat{\mathbf{E}}_j\|^2$ representa la energía de la diferencia entre las señales de error no cuantificada y cuantificada. Por claridad, esta cantidad se puede conocer como la "energía residual", y se puede usar además para evaluar un "criterio de selección de ganancia", en el que se selecciona el parámetro de ganancia óptimo g^* . Un criterio de selección de ganancia tal se da en la ecuación (6), aunque son posibles muchos.

La necesidad de un término de desviación β_j puede surgir del caso donde la función de ponderación de error W en las ecuaciones (3) y (4) puede no producir adecuadamente distorsiones igualmente perceptibles a través del vector $\hat{\mathbf{E}}_j$. Por ejemplo, aunque la función de ponderación de error \mathbf{W} se puede usar para intentar "blanquear" el espectro de error en algún grado, puede haber ciertas ventajas de colocar más peso en las frecuencias bajas, debido a la percepción de distorsión por el oído humano. Como resultado del aumento de la ponderación de error en las frecuencias bajas, las señales de frecuencia alta pueden ser infra modeladas por la capa de mejora. En estos casos,

puede haber un beneficio directo de desviar la métrica de distorsión hacia valores de \mathbf{g}_j que no atenúan los componentes de frecuencia alta de \mathbf{S}_j , de manera que el infra modelado de las frecuencias altas no provoca artefactos de sonido no naturales u objetables en la señal de audio reconstruida final. Un ejemplo tal sería el caso de una señal de habla sin voz. En este caso, el audio de entrada se compone generalmente de señales tipo ruido de frecuencia media a alta producidas a partir del flujo turbulento de aire desde la boca humana. Puede ser que el codificador de la capa central no codifique este tipo de forma de onda directamente, sino que puede usar un modelo de ruido para generar una señal de audio de sonido similar. Esto puede provocar una correlación generalmente baja entre las señales de audio de entrada y de audio de salida de la capa central. No obstante, en esta realización, el vector de señal de error \mathbf{E}_j está basado en una diferencia entre las señales de audio de entrada y de salida de audio de la capa central. Dado que estas señales no se pueden correlacionar muy bien, la energía de la señal de error \mathbf{E}_j puede no ser necesariamente menor que o bien el audio de entrada o bien el audio de salida de la capa central. En ese caso, la minimización del error en la ecuación (6) puede provocar que el escalado de ganancia sea demasiado agresivo, lo cual puede provocar artefactos audibles potenciales.

En otro caso, los factores de desviación β_j se pueden basar en otras características de señal de las señales de audio de entrada y de audio de salida de la capa central. 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 habla y ciertos tipos de música pueden tener un contenido de armónicos alto y de esta manera una relación pico a media alta. No obstante, una señal de música procesada a través de un códec de habla puede provocar una calidad escasa debido al desajuste del modelo de codificación, y como resultado, el espectro de señal de salida de la capa central puede tener una relación pico a media reducida cuando se compara con el espectro de señal de entrada. En este caso, puede ser beneficioso reducir la cantidad de desviación en el proceso de minimización a fin de permitir que el audio de salida de la capa central sea escalado con ganancia a una energía inferior permitiendo por ello la codificación de la capa de mejora para tener un efecto más pronunciado en el audio de salida compuesto. Al contrario, ciertos tipos de señales de entrada de habla o música pueden exhibir relaciones pico a media inferiores, en cuyo caso, las señales se pueden percibir como que son más ruidosas, y pueden beneficiarse por lo tanto de menos escalado del audio de salida de la capa central aumentando la desviación de error. Un ejemplo de una función para generar los factores de desviación para β_j , se da como:

$$\beta_j = \begin{cases} 1 + 10^6 \cdot j; & UVHabla = VERDADERO \text{ o } \phi_s < \lambda \phi_{s_c}, \quad 0 \leq j < M. \\ 10^{(-j \cdot \Delta / 10)}; & \text{de otro modo} \end{cases} \quad (7)$$

donde λ puede ser algún umbral, y la relación pico a media para el vector ϕ_y se puede dar como:

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

y donde $\mathbf{y}_{k_1 k_2}$ es un subconjunto de vectores de $y(k)$ de manera que $\mathbf{y}_{k_1 k_2} = y(k); k_1 \leq k \leq k_2$.

Una vez que se determina el índice de ganancia óptimo j^* a partir de la ecuación (6), se genera la palabra de código asociada i_g y el vector de error óptimo \mathbf{E}^* se envía al codificador de señal de error 430, donde \mathbf{E}^* se codifica de una forma que es adecuada para multiplexar con otras palabras de código (mediante el MUX 440) y transmite para uso por un descodificador correspondiente. En una realización ejemplar, el codificador de señal de error 408 usa Codificación de Pulso Factorial (FPC). Este método es ventajoso desde un punto de vista de complejidad de procesamiento dado que el proceso de enumeración asociado con la codificación del vector \mathbf{E}^* es independiente del proceso de generación de vector que se usa para generar $\hat{\mathbf{E}}_j$.

El descodificador de la capa de mejora 450 invierte estos procesos para producir la salida de audio de mejora $\hat{S}(n)$. Más específicamente, i_g e i_E se reciben por el descodificador 450, con i_E que se envía por el desmultiplexor 455 al descodificador de señal de error 460 donde se deriva el vector de error óptimo \mathbf{E}^* a partir de la palabra de código. El vector de error óptimo \mathbf{E}^* se pasa al combinador de señal 465 donde la $\hat{S}_c(n)$ recibida se modifica como en la ecuación (2) para producir $\hat{S}(n)$.

Una segunda realización de la presente invención implica un sistema de codificación integrado de múltiples capas como se muestra en la FIG. 5. Aquí, se puede ver que hay cinco capas integradas dadas para este ejemplo. Las capas 1 y 2 pueden estar basadas ambas en códec de habla, y las capas 3, 4, y 5 pueden ser capas de mejora MDCT. De esta manera, los codificadores 502 y 503 pueden utilizar códec de habla para producir y sacar la señal de entrada codificada $s(n)$. Los codificadores 510, 610, y 514 comprenden codificadores de capa de mejora, cada uno

que saca una mejora diferente a la señal codificada. Similar a la realización previa, el vector de señal de error para la capa 3 (codificador 510) se puede dar como:

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

5 donde $\mathbf{S} = \text{MDCT}\{\mathbf{W}s\}$ es la señal de entrada transformada ponderada, y $\mathbf{S}_2 = \text{MDCT}\{\mathbf{W}s_2\}$ es la señal transformada ponderada generada a partir del descodificador de capa 1/2 506. En esta realización, la capa 3 puede ser una capa de cuantificación de tasa baja, y como tal, puede haber relativamente pocos bits para codificar la señal de error cuantificada correspondiente $\hat{\mathbf{E}}_3 = \mathbf{Q}\{\mathbf{E}_3\}$. A fin de proporcionar una buena calidad bajo estas restricciones, solamente se puede cuantificar una fracción de los coeficientes dentro de \mathbf{E}_3 . Las posiciones de los coeficientes a ser codificados pueden ser fijas o pueden ser variables, pero si se permiten variar, se puede requerir enviar información adicional al descodificador para identificar estas posiciones. Si, por ejemplo, el intervalo de posiciones codificadas comienza en k_s y finaliza en k_e , donde $0 \leq k_s < k_e < N$, entonces el vector de señal de error cuantificado $\hat{\mathbf{E}}_3$ puede contener valores no cero solamente dentro de ese intervalo, y ceros para posiciones fuera de ese intervalo. La información de posición e intervalo también puede ser implícita, dependiendo del método de codificación usado. Por ejemplo, es bien conocido en codificación de audio que una banda de frecuencias se puede considerar importante perceptivamente, y que la codificación de un vector de señal puede centrarse en esas frecuencias. En estas circunstancias, el intervalo codificado puede ser variable, y puede no abarcar un conjunto de frecuencias contiguo. Pero a cualquier tasa, una vez que la señal está cuantificada, el espectro de salida codificado compuesto se puede interpretar como:

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

20 que entonces se usa como entrada al codificador de capa 4 610.

El codificador de capa 4 610 es similar al codificador de capa de mejora 410 de la realización previa. Usando el candidato de vector de ganancia \mathbf{g}_j , el vector de error correspondiente se puede describir como:

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

25 donde \mathbf{G}_j puede ser una matriz de ganancia con vector \mathbf{g}_j como el componente diagonal. En la realización actual, no obstante, el vector de ganancia \mathbf{g}_j se puede relacionar con el vector de señal de error cuantificado $\hat{\mathbf{E}}_3$ de la siguiente manera. Dado que el vector de señal de error cuantificado $\hat{\mathbf{E}}_3$ puede estar limitado en intervalo de frecuencia, por ejemplo, iniciando en la posición de vector k_s y terminando en la posición del vector k_e , la señal de salida de capa 3 \mathbf{S}_3 se supone que se codifica con bastante precisión dentro de ese intervalo. Por lo tanto, según la presente invención, el vector de ganancia \mathbf{g}_j se ajusta en base a las posiciones codificadas del vector de señal de error de capa 3, k_s y k_e . Más específicamente, a fin de conservar la integridad de la señal en esas ubicaciones, los elementos de ganancia individuales correspondientes se pueden fijar a un valor constante α . Es decir:

$$g_j(k) = \begin{cases} \alpha; & k_s \leq k \leq k_e \\ \gamma_j(k); & \text{de otro modo} \end{cases} \quad (12)$$

35 donde generalmente $0 \leq \gamma_j(k) \leq 1$ y $g_j(k)$ es la ganancia en la posición de orden k del vector candidato de orden j. En una realización ejemplar, el valor de la constante es uno ($\alpha = 1$), no obstante son posibles muchos valores. Además, el intervalo de frecuencia puede abarcar múltiples posiciones de inicio y finalización. Es decir, la ecuación (12) se puede segmentar en intervalos no continuos de ganancias que varían que están basados en alguna función de la señal de error $\hat{\mathbf{E}}_3$, y se puede escribir de manera general como:

$$g_j(k) = \begin{cases} \alpha; & \hat{E}_3(k) \neq 0 \\ \gamma_j(k); & \text{de otro modo} \end{cases} \quad (13)$$

Para este ejemplo, se usa una ganancia fija α para generar $g_j(k)$ cuando las posiciones correspondientes en la señal de error cuantificada previamente \hat{E}_3 no son cero, y la función de ganancia $\gamma_j(k)$ se usa cuando las posiciones correspondientes en \hat{E}_3 son cero.

5 Una posible función de ganancia se puede definir como:

$$\gamma_j(k) = \begin{cases} \alpha \cdot 10^{(-j \cdot \Delta / 20)}, & k_l \leq k \leq k_h, \\ \alpha; & \text{de otro modo} \end{cases}, \quad 0 \leq j < M, \quad (14)$$

donde Δ es un tamaño de paso (por ejemplo, $\Delta \approx 2,2$ dB), α es una constante, M es el número de candidatos (por ejemplo, M=4, que se pueden representar usando solamente 2 bits), y k_l y k_h son los límites de frecuencia bajo y alto, respectivamente, sobre los cuales puede tener lugar la reducción de ganancia. La introducción de los parámetros k_l y k_h es útil en sistemas donde se desea escalado solamente sobre un cierto intervalo de frecuencia. Por ejemplo, en una realización dada, las frecuencias altas pueden no estar modeladas adecuadamente por la capa central, de esta manera la energía dentro de la banda de frecuencia alta puede ser inherentemente inferior que aquella en la señal de audio de entrada. En ese caso, puede haber poco o ningún beneficio de escalar la salida de capa 3 en esa región de señal dado que la energía de error total puede aumentar como resultado.

15 Resumiendo, la pluralidad de candidatos de vector de ganancia \mathbf{g}_j se basa en alguna función de los elementos codificados de un vector de señal codificado previamente, en este caso \hat{E}_3 . Esto se puede expresar en términos generales como:

$$g_j(k) = f(k, \hat{E}_3). \quad (15)$$

20 Las operaciones de descodificador correspondientes se muestran en el lado derecho de la FIG. 5. Según se reciben las diversas capas de flujos de bits codificados (i_1 a i_5), se construyen las señales de salida de calidad más alta en la jerarquía de capas de mejora sobre el descodificador de la capa central (capa 1). Es decir, para esta realización particular, como las dos primeras capas se componen de codificación del modelo de habla en el dominio del tiempo (por ejemplo, CELP) y las tres capas restantes se componen de codificación en el dominio de transformada (por ejemplo, MDCT), la salida final para el sistema $\hat{s}(n)$ se genera según lo siguiente:

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

25 donde $\hat{e}_2(n)$ es la señal de capa de mejora en el dominio del tiempo de capa 2, y $\hat{\mathbf{S}}_2 = \text{MDCT}\{\mathbf{W}\mathbf{s}_2\}$ es el vector MDCT ponderado que corresponde a la salida de audio de capa 2 $\hat{s}_2(n)$. En esta expresión, la señal de salida total $\hat{s}(n)$ se puede determinar a partir del nivel más alto de capas de flujo de bits consecutivas que se reciben. En esta realización, se supone que las capas de nivel más bajo tienen una probabilidad mayor de ser recibidas

adecuadamente 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 descodificación de capa de mejora en la ecuación (16).

La FIG. 6 es un diagrama de bloques que muestra el codificador de capa 4 610 y el descodificador 650. El codificador y descodificador mostrados en la FIG. 6 son similares a aquellos mostrados en la FIG. 4, excepto que el valor de ganancia usado para unidades de escalado 615 y 670 se deriva a través de generadores de ganancia selectiva en frecuencia 630 y 660, respectivamente. Durante la operación la salida de audio de la capa 3 S_3 se saca del codificador de capa 3 y recibe por la unidad de escalado 615. Adicionalmente, el vector de error de capa 3 \hat{E}_3 se saca del codificador de capa 3 510 y recibe por el generador de ganancia selectivo en frecuencia 630. Como se trató, dado que el vector de señal de error cuantificado \hat{E}_3 puede estar limitado en intervalo de frecuencia, el vector de ganancia g_j se ajusta en base a, por ejemplo, las posiciones k_s y k_e como se muestra en la ecuación 12, o la expresión más general en la ecuación 13.

El audio escalado S_j se saca de la unidad de escalado 614 y se recibe por el generador de señal de error 620. Como se trató anteriormente, el generador de señal de error 620 recibe la señal de audio de entrada S y determina un valor de error E_j para cada vector de escalado utilizado por la unidad de escalado 615. Estos vectores de error se pasan a la circuitería de selector de ganancia 635 junto con los valores de ganancia usados en la determinación de los vectores de error y un error particular E^* basado en el valor de ganancia óptimo g^* . Una palabra de código (i_g) que representa la ganancia óptima g^* se saca desde el selector de ganancia 635, junto con el vector de error óptimo E^* , se pasa al codificador de señal de error 640 donde se determina y saca la palabra de código i_E . Tanto i_g como i_E se sacan al multiplexor 645 y transmiten a través del canal 125 al descodificador de capa 4 650.

Durante la operación del descodificador de capa 4 650, i_g e i_E se reciben desde el canal 125 y desmultiplexan por el desmultiplexor 655. La palabra de código de ganancia i_g y el vector de error de capa 3 \hat{E}_3 se usan como entrada al generador de ganancia selectiva en frecuencia 660 para producir un vector de ganancia g^* según el método correspondiente del codificador 610. El vector de ganancia g^* se aplica entonces al vector de audio reconstruido de capa 3 \hat{S}_3 dentro de la unidad de escalado 670, la salida de la cual se combina entonces en el combinador de señal 675 con el vector de error de la capa de mejora de capa 4 E^* , que fue obtenido a partir del descodificador de señal de error 655 a través de la descodificación de la palabra de código i_E , para producir la salida de audio reconstruida de la capa 4 \hat{S}_4 como se muestra.

La FIG. 7 es un diagrama de flujo 700 que muestra la operación de un codificador según la primera y segunda realizaciones de la presente invención. Como se trató anteriormente, ambas realizaciones utilizan una capa de mejora que escala el audio codificado con una pluralidad de valores de escalado y entonces elige el valor de escalado que provoca un error más bajo. No obstante, en la segunda realización de la presente invención, el generador de ganancia selectiva en frecuencia 630 se utiliza para generar los valores de ganancia.

El flujo lógico comienza en el Bloque 710 donde un codificador de la capa central recibe una señal de entrada a ser codificada y codifica la señal de entrada para producir una señal de audio codificada. El codificador de capa de mejora 410 recibe la señal de audio codificada ($s_c(n)$) y la unidad de escalado 415 escala la señal de audio codificada con una pluralidad de valores de ganancia para producir una pluralidad de señales de audio codificadas escaladas, cada una que tiene un valor de ganancia asociado. (Bloque 720). En el Bloque 730, el generador de señal de error 420 determina una pluralidad de valores de error que existen entre la señal de entrada y cada una de la pluralidad de señales de audio codificadas escaladas. El selector de ganancia 425 entonces elige un valor de ganancia a partir de una pluralidad de valores de ganancia (Bloque 740). Como se trató anteriormente, el valor de ganancia (g^*) está asociado con una señal de audio codificada escalada que provoca un valor de error bajo (E^*) que existe entre la señal de entrada y la señal de audio codificada escalada. Finalmente en el Bloque 750 el transmisor 440 transmite el valor de error bajo (E^*) junto con el valor de ganancia (g^*) como parte de una capa de mejora a la señal de audio codificada. Como reconocerá un experto en la técnica, tanto E^* como g^* se codifican adecuadamente anterior a la transmisión.

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

Escalado de capa central para estéreo

En la descripción anterior, fue descrito un sistema de codificación integrado en el que cada una de las capas estaba codificando una señal mono. Ahora un sistema de codificación integrado para codificación estéreo u otras señales de múltiples canales. Para abreviar, se describe la tecnología en el contexto de una señal estéreo que consta de dos entradas audio (fuentes); no obstante, las realizaciones ejemplares descritas en la presente memoria se pueden extender fácilmente a casos donde la señal estéreo tiene más de dos entradas audio, como es el caso en entradas de audio de múltiples canales. Para propósitos de ilustración y no de limitación, las dos entradas audio son señales estéreo que constan de la señal izquierda (s_L) y la señal derecha (s_R), donde s_L y s_R son vectores de columna n-

dimensionales que representan una trama de datos de audio. De nuevo por abreviar, se tratará en detalle un sistema de codificación integrado que consta de dos capas esto es una capa central y una capa de mejora. La idea propuesta se puede extender fácilmente a un sistema de codificación integrado de múltiples capas. También el códec puede no estar integrado por sí mismo, es decir, puede tener solamente una capa, con algunos de los bits de ese códec que están dedicados para estéreo y el resto de bits para señal mono.

Se conoce un códec estéreo integrado que consta de una capa central que simplemente codifica una señal mono y capas de mejora que codifican o bien la frecuencia más alta o bien las señales estéreo. En ese escenario limitado, la capa central codifica una señal mono (s), obtenida de la combinación de s_L y s_R , para producir una señal mono codificada \hat{s} . Dejemos que H sea una matriz de combinación 2x1 usada para generar una señal mono, es decir,

$$\mathbf{s} = (\mathbf{s}_L \quad \mathbf{s}_R) \mathbf{H} \tag{17}$$

Se señala que en la ecuación (17), s_R puede ser una versión retardada de la señal de audio derecha en lugar de solo la señal de canal derecho. Por ejemplo, se puede calcular el retardo para maximizar la correlación de s_L y la versión retardada de s_R . Si la matriz H es $[0,5 \ 0,5]^T$, entonces la ecuación 17 provoca una ponderación igual de los canales derecho e izquierdo respectivos, es decir, $s = 0,5 s_L + 0,5 s_R$. Las realizaciones presentadas en la presente memoria no están limitadas a la capa central que codifica la señal mono y la capa de mejora que codifica la señal estéreo. Tanto la capa central del códec integrado así 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 se codifican por los múltiples canales de la capa central puede ser menor que el número de canales en la señal de audio de múltiples canales que se puede codificar por la capa de mejora. Dejemos que (m, n) sean los números de canales a ser codificados por la capa central y la capa de mejora, respectivamente. Dejemos que $s_1, s_2, s_3, \dots, s_n$ sean una representación de n canales de audio a ser codificados por el sistema integrado. Los m canales a ser codificados por la capa central se derivan de estos y se obtienen como

$$[s^1 \quad s^2 \quad \dots \quad s^m] = [s_1 \quad s_2 \quad \dots \quad s_n] \mathbf{H}, \tag{17a}$$

donde H es una matriz $n \times m$.

Como se mencionó antes, la capa central codifica una señal mono s para producir una señal codificada de la capa central \hat{s} . A fin de generar estimaciones de los componentes estéreo a partir de \hat{s} , se calcula un factor de equilibrio. Este factor de equilibrio se calcula como:

$$w_L = \frac{\mathbf{s}_L^T \mathbf{s}}{\mathbf{s}^T \mathbf{s}}, \quad w_R = \frac{\mathbf{s}_R^T \mathbf{s}}{\mathbf{s}^T \mathbf{s}} \tag{18}$$

Se puede mostrar que si la matriz de combinación H es $[0,5 \ 0,5]^T$, entonces

$$w_L = 2 - w_R$$

Señalar que la relación permite la cuantificación solamente de un parámetro y el otro se puede extraer fácilmente a partir del primero. Las salidas estéreo ahora se calculan como

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

En la sección posterior, estaremos trabajando en el dominio de la frecuencia en lugar de en el dominio del tiempo.

Así una señal correspondiente en el dominio de la frecuencia se representa en letras mayúsculas, es decir, S, \hat{S} , S_L , S_R , \hat{S}_L , y \hat{S}_R son la representación de s, \hat{s} , s_L , s_R , \hat{s}_L , y \hat{s}_R , respectivamente. El factor de equilibrio en el dominio de la frecuencia se calcula usando términos en el dominio de la frecuencia y se da por

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

(21)

y

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

(22)

5 En el dominio de la frecuencia, los vectores se pueden dividir además en sub vectores de no solapamiento, es decir, un vector \mathbf{S} de dimensión n , se puede dividir en t sub vectores, $\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_t$, de dimensiones m_1, m_2, \dots, m_t , de manera que

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

(23)

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

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

(24)

10 El factor de equilibrio en este caso es independiente de la consideración de ganancia.

Con referencia ahora a las FIG. 8 y 9, se demuestran los dibujos de la técnica anterior relevantes a señales estéreo y otras de múltiples canales. El sistema de compresión de habla/audio integrado 800 de la FIG. 8 es similar a la FIG. 1 pero tiene señales de entrada de audio múltiples, en este ejemplo mostradas como las señales de entrada estéreo izquierda y derecha $\mathbf{S}(n)$. Estas señales de audio de entrada se alimentan al combinador 810 que produce una entrada de audio $s(n)$ como se muestra. Las señales de entradas múltiples se proporcionan también al codificador de capa de mejora 820 como se muestra. En el lado de descodificación, el descodificador de capa de mejora 830 produce señales de audio de salida mejoradas $\hat{\mathbf{S}}_L \hat{\mathbf{S}}_R$ como se muestra.

La FIG. 9 ilustra un codificador de capa de mejora anterior 900 que podría ser usado en la FIG. 8. Las entradas de audio múltiples se proporcionan a un generador de factor de equilibrio, junto con la señal de audio de salida de la capa central como se muestra. El Generador de Factor de Equilibrio 920 del codificador de capa de mejora 910 recibe las entradas de audio múltiples para producir la señal \mathbf{i}_B , que se pasa a lo largo del MUX 325 como se muestra. La señal \mathbf{i}_B es una representación del factor de equilibrio. En la realización preferida \mathbf{i}_B es una secuencia de bits que representa los factores de equilibrio. En el lado del descodificador, la señal \mathbf{i}_B se recibe por el descodificador de factor de equilibrio 940 que produce los elementos de factor de equilibrio $\mathbf{W}_L(\mathbf{n})$ y $\mathbf{W}_R(\mathbf{n})$, como se muestra, que se reciben por el combinador de señal 950 como se muestra.

Cálculo del factor de equilibrio de múltiples canales

Como se mencionó antes, en muchas situaciones el códec usado para la codificación de la señal mono se diseña para habla de canal único y provoca ruido de modelo de codificación siempre que se use para codificación de señales que no están completamente soportadas por el modelo de códec. Las señales de música y otras señales de tipo no de habla son algunas de las señales que no se modelan adecuadamente por un códec de la capa central que se basa en un modelo de habla. La descripción anterior, con respecto a las FIG. 1-7, proponía aplicar una ganancia selectiva en frecuencia a la señal codificada por la capa central. El escalado estaba optimizado para minimizar una distorsión particular (valor de error) entre la señal de entrada audio y la codificada escalada. El planteamiento descrito anteriormente funciona bien para señales de canal único pero puede no ser óptimo para aplicación del escalado de la capa central cuando la capa de mejora está codificando las señales estéreo u otras de múltiples canales.

Dado que el componente mono de la señal de múltiples canales, tal como la señal estéreo, se obtiene a partir de la combinación de dos o más entradas de audio estéreo, la señal combinada s también puede no ser conforme al modelo de habla de canal único; por lo tanto el códec de capa central puede producir ruido cuando se codifica la

señal combinada. De esta manera, hay una necesidad de un planteamiento que permita el escalado de la señal codificada de la capa central en un sistema de codificación integrado, reduciendo por ello el ruido generado por la capa central. En el planteamiento de señal mono descrito anteriormente, una medida de distorsión particular, en la que fue obtenido el escalado selectivo en frecuencia, estaba basada en el error de la señal mono. Este error $E_4(j)$ se muestra en la ecuación (11) anterior. La distorsión de sólo la señal mono, no obstante, no es suficiente para mejorar la calidad del sistema de comunicación estéreo. El escalado contenido en la ecuación (11) puede ser mediante un factor de escalado de la unidad (1) o cualquier otra función identificada.

Para una señal estéreo, una medida de distorsión debería capturar la distorsión tanto del canal derecho como del izquierdo. Dejemos que \mathbf{E}_L y \mathbf{E}_R sean el vector de error para los canales izquierdo y el derecho, respectivamente, y se dan por

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

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

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

Ahora consideremos el caso donde los vectores de ganancia selectiva en frecuencia g_j ($0 \leq j < M$) se aplica a S. Este vector de ganancia selectiva en frecuencia se representa en forma de matriz como \mathbf{G}_j , donde \mathbf{G}_j es una matriz diagonal con elementos diagonales g_j . Para cada vector \mathbf{G}_j , los vectores de error se calculan como:

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

con las estimaciones de las señales estéreo dadas por los términos $\mathbf{W} \cdot \mathbf{G}_j \cdot \hat{\mathbf{S}}$. Se puede ver que la matriz de ganancia \mathbf{G} puede ser una matriz unidad (1) o puede ser cualquier otra matriz diagonal; se reconoce que no toda estimación posible puede ejecutarse para cada señal escalada.

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

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

Se puede ver que el valor de distorsión puede estar compuesto de múltiples medidas de distorsión.

El índice j del vector de ganancia selectiva en frecuencia que se selecciona se da por:

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

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

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

O puede ser una distorsión ponderada o desviada dada por:

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

(31)

La desviación B_L y B_R puede ser una función de las energías del canal izquierdo y derecho.

5 Como se mencionó antes, en el dominio de la frecuencia, los vectores se pueden dividir además en sub vectores de no solapamiento. Para extender la técnica propuesta para incluir la división del vector en el dominio de la frecuencia en sub vectores, se calcula el factor de equilibrio usado en (27) para cada sub vector. De esta manera, los vectores de error \mathbf{E}_L y \mathbf{E}_R para cada una de las ganancias selectivas en frecuencia se forma por concatenación de sub vectores de error dada por

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

(32)

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

Cálculo del factor de equilibrio

15 El factor de equilibrio generado usando la técnica anterior (ecuación 21) es independiente de la salida de la capa central. No obstante, a fin de minimizar una medida de distorsión dada en (30) y (31), puede ser beneficioso también calcular el factor de equilibrio para minimizar la distorsión correspondiente. Ahora el factor de equilibrio W_L y W_R se puede calcular como

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

(33)

20 en la que se puede ver que el factor de equilibrio es independiente de la ganancia, como se muestra en el dibujo de la FIG. 11, por ejemplo. Esta ecuación minimiza las distorsiones en la ecuación (30) y (31). El problema con usar tal factor de equilibrio es que ahora:

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

(34)

por lo tanto los campos de bits separados pueden ser necesarios para cuantificar W_L y W_R . Esto se puede evitar poniendo la restricción $W_L(j)=2-W_R(j)$ en la optimización. Con esta restricción de la solución óptima para la ecuación (30) se da por:

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

(35)

25 en la que el factor de equilibrio es dependiente de un término de ganancia según se muestra; la FIG. 10 de los dibujos ilustran un factor de equilibrio dependiente. Si los factores de desviación B_L y B_R son unidad, entonces

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

(36)

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

En la codificación estéreo, la dirección y ubicación del origen del sonido puede ser más importante que la distorsión cuadrática media. La relación de energía del canal izquierdo y la energía del canal derecho puede ser por lo tanto un mejor indicador de dirección (o ubicación del origen del sonido) más que la minimización de una medida de distorsión ponderada. En tales escenarios, el factor de equilibrio calculado en la ecuación (35) y (36) puede no ser un buen planteamiento para calcular el factor de equilibrio. La necesidad es mantener la relación de la energía del canal izquierdo y derecho antes y después de codificar la misma. La relación de la energía de canal antes de codificar y después de codificar se da por:

$$v = \frac{\|\mathbf{S}_L\|^2}{\|\mathbf{S}_R\|^2}, \quad \hat{v} = \frac{W_L^2(j) \|\hat{\mathbf{S}}\|^2}{W_R^2(j) \|\hat{\mathbf{S}}\|^2},$$

(37)

respectivamente. Igualando estas dos relaciones de energía y usando la suposición $W_L(j) = 2 - W_R(j)$, obtenemos

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

(38)

que da los componentes del factor de equilibrio del factor de equilibrio generado. Señalar que el factor de equilibrio calculado en (38) es ahora independiente de G_j , de esta manera ya no es una función de j , proporcionando un factor de equilibrio auto correlacionado que es independiente de la consideración de ganancia; un factor de equilibrio dependiente se ilustra además en la FIG. 10 de los dibujos. Usando este resultado con las ecuaciones 29 y 32, podemos extender la selección del índice de escalado de la capa central óptimo j para incluir los segmentos de vector concatenados k , de manera que:

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

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

Con referencia ahora a la FIG. 10, se ilustra un diagrama de bloques 1000 de un codificador de capa de mejora y descodificador de capa de mejora según diversas realizaciones. Las señales de audio de entrada $\mathbf{s}(\mathbf{n})$ se reciben por el generador del factor de equilibrio 1050 del codificador de la capa de mejora 1010 y el generador de señal de error (señal de distorsión) 1030 del generador de vector de ganancia 1020. La señal de audio codificada de la capa central $\hat{\mathbf{S}}(\mathbf{n})$ se recibe por la unidad de escalado 1025 del generador del vector de ganancia 1020 como se muestra. La unidad de escalado 1025 opera para escalar la señal de audio codificada $\mathbf{S}(\mathbf{n})$ con una pluralidad de valores de ganancia para generar un número de señales de audio codificadas candidatas, donde se escala al menos una de las señales de audio codificadas candidatas. Como se mencionó previamente, se puede emplear escalado por la unidad o cualquier función de identificación deseada. La unidad de escalado 1025 saca audio escalado \mathbf{S}_j , que se recibe por el generador de factor de equilibrio 1030. La generación del factor de equilibrio que tiene una pluralidad de componentes del factor de equilibrio, cada uno asociado con una señal de audio de las señales de audio de múltiples canales recibidas por el codificador de la capa de mejora 1010, fue discutida anteriormente en conexión con las Ecuaciones (18), (21), (24), y (33). Esto se consuma por el generador de factor de equilibrio 1050 como se muestra, para producir componentes del factor de equilibrio $\hat{S}_L(\mathbf{n})$, $\hat{S}_R(\mathbf{n})$, como se muestra. Como se trató en conexión con la ecuación (38), anterior, el generador de factor de equilibrio 1030 ilustra el factor de equilibrio como independiente de la ganancia.

El generador de vector de ganancia 1020 es responsable de determinar un valor de ganancia a ser aplicado a la señal de audio codificada para generar una estimación de la señal de audio de múltiples canales, como se discutió en las Ecuaciones (27), (28), y (29). Esto se consuma por la unidad de escalado 1025 y el generador de factor de equilibrio 1050, que trabajan juntos para generar la estimación basada en el factor de equilibrio y al menos una señal

de audio codificada escalada. El valor de ganancia se basa en el factor de equilibrio y la señal de audio de múltiples canales, en donde el valor de ganancia está configurado para minimizar un valor de distorsión entre la señal de audio de múltiples canales y la estimación de la señal de audio de múltiples canales. La ecuación (30) trata la generación de un valor de distorsión como una función de la estimación de la señal de entrada de múltiples canales y la señal de entrada real en sí misma. De esta manera, los componentes del factor de equilibrio se reciben por el generador de señal de error 1030, junto con las señales de audio de entrada $\mathbf{s}(n)$, para determinar un valor de error \mathbf{E}_j para cada vector de escalado utilizado por la unidad de escalado 1025. Estos vectores de error se pasan a la circuitería del selector de ganancia 1035 junto con los valores de ganancia usados en la determinación de los vectores de error y un error particular \mathbf{E}^* basado en el valor de ganancia óptimo \mathbf{g}^* . El selector de ganancia 1035, entonces, está operativo para evaluar el valor de distorsión en base a la estimación de la señal de entrada de múltiples canales y la señal real en sí misma a fin de determinar una representación de un valor de ganancia óptimo \mathbf{g}^* de los valores de ganancia posibles. Una palabra de código (i_g) que representa la ganancia óptima \mathbf{g}^* se saca del selector de ganancia 1035 y recibe por el multiplexor MUX 1040 como se muestra.

Tanto i_g como i_B se sacan al multiplexor 1040 y transmiten por el transmisor 1045 al descodificador de capa de mejora 1060 a través del canal 125. La representación del valor de ganancia i_g se saca para transmisión al Canal 125 como se muestra pero también se puede almacenar si se desea.

En el lado del descodificador, durante la operación del descodificador de la capa de mejora 1060, se reciben i_g e i_E desde el canal 125 y desmultiplexan por el desmultiplexor 1065. De esta manera, el descodificador de la capa de mejora recibe una señal de audio codificada $\hat{S}(n)$, un factor de equilibrio codificado i_B y un valor de ganancia codificado i_g . El descodificador del vector de ganancia 1070 comprende un generador de ganancia selectiva en frecuencia 1075 y una unidad de escalado 1080 como se muestra. El descodificador del vector de ganancia 1070 genera un valor de ganancia descodificado a partir del valor de ganancia codificado. El valor de ganancia codificado i_g se introduce al generador de ganancia selectiva en frecuencia 1075 para producir un vector de ganancia \mathbf{g}^* según el método correspondiente del codificador 1010. El vector de ganancia \mathbf{g}^* se aplica entonces a la unidad de escalado 1080, que escala la señal de audio codificada $S(n)$ con el valor de ganancia descodificado \mathbf{g}^* para generar una señal de audio escalada. El combinador de señal 1095 recibe las señales de salida del factor de equilibrio codificado del descodificador de factor de equilibrio 1090 a la señal de audio escalada $\mathbf{G}_j \hat{S}(n)$ para generar y sacar una señal de audio de múltiples canales descodificados, mostrada como las señales de audio de salida mejorada.

El diagrama de bloques 1100 de un codificador de capa de mejora ejemplar y descodificador de capa de mejora en el que, como se discutió en conexión con la ecuación (33), anterior, el generador de factor de equilibrio 1030 genera un factor de equilibrio que es dependiente de la ganancia. Esto se ilustra por el generador de señal de error que genera la señal G_j 1110.

Con referencia ahora a las FIG. 12-14, los flujos se presentan que cubren la metodología de diversas realizaciones presentadas en la presente memoria. 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 una pluralidad de 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 o bien una señal mono o bien de múltiples canales, tal como una señal estéreo que se ilustra a modo de ejemplo en los dibujos. Además, la señal de audio codificada puede comprender una pluralidad de canales. Puede haber más de un canal en la capa central y el número de canales en la capa de mejora puede ser mayor que el número de canales en la capa central. A continuación, en el Bloque 1230, se genera un factor de equilibrio que tiene componentes de factor de equilibrio cada uno asociado con una señal de audio de la señal de audio de múltiples canales. Las Ecuaciones (18), (21), (24), (33) describen una generación del factor de equilibrio. Cada componente del factor de equilibrio puede ser dependiente de otros componentes del factor de equilibrio generados, como es el caso en la Ecuación (38). La generación del factor de equilibrio puede comprender generar un valor de correlación entre la señal de audio codificada escalada y al menos una de las señales de audio de la señal de audio de múltiples canales, tal como en las Ecuaciones (33), (36). Se puede generar una auto correlación entre al menos una de las señales de audio, como en la Ecuación (38), a partir de la cual se puede generar una raíz cuadrada. En el Bloque 1240, se determina un valor de ganancia a ser aplicado a la señal de audio codificada para generar una estimación de la señal de audio de múltiples canales basada en el factor de equilibrio y la señal de audio de múltiples canales. El valor de ganancia se configura 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), (30) describen la determinación del valor de ganancia. Se puede elegir un valor de ganancia a partir de una pluralidad de valores de ganancia para escalar la señal de audio codificada y generar las señales de audio codificadas escaladas. El valor de distorsión se puede generar en base a esta estimación; el valor de ganancia se puede basar en el valor de distorsión. En el Bloque 1250, una representación del valor de ganancia se saca para o bien transmisión y/o bien 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 una pluralidad de señales de audio. En el Bloque 1320, la señal de audio de múltiples canales se codifica para generar una señal de audio codificada. Los procesos de los Bloques 1310 y 1320 se realizan por un codificador de la capa central,

como se describió previamente. Como se expuso previamente, la señal de audio codificada puede ser o bien una señal mono o una de múltiples canales, tal como una señal estéreo que se ilustra a modo de ejemplo en los dibujos. Además, la señal de audio codificada puede comprender una pluralidad de canales. Puede haber más de un canal en la capa central y el número de canales en la capa de mejora puede ser mayor que el número de canales en la capa central.

En el Bloque 1330, la señal de audio codificada se escala con un número de valores de ganancia para generar un número de señales de audio codificadas candidatas, con al menos una de las señales de audio codificadas candidatas que está escalada. El escalado se consume por la unidad de escalado del generador de vector de ganancia. Como se trató, el escalado de la señal de audio codificada puede incluir escalar con un valor de ganancia de la unidad. El valor de ganancia de la pluralidad de valores de ganancia puede ser una matriz de ganancia con el vector \mathbf{g}_j como el componente diagonal como se describió anteriormente. La matriz de ganancia puede ser selectiva en frecuencia. Puede ser dependiente de la salida de la capa central, la señal de audio codificada ilustrada en los dibujos. Se puede elegir un valor de ganancia a partir de una pluralidad de valores de ganancia para escalar la señal de audio codificada y para generar las señales de audio codificadas escaladas. En el Bloque 1340, se genera un factor de equilibrio que tiene componentes de factor de equilibrio cada uno asociado con una señal de audio de la señal de audio de múltiples canales. La generación del factor de equilibrio se realiza por el generador de factor de equilibrio. Cada componente del factor de equilibrio puede ser dependiente de otros componentes del factor de equilibrio generados, como es el caso en la Ecuación (38). La generación del factor de equilibrio puede comprender generar un valor de correlación entre la señal de audio codificada escalada y al menos una de las señales de audio de la señal de audio de múltiples canales, tal como en las Ecuaciones (33), (36). Se puede generar una auto correlación entre al menos una de las señales de audio, como en la Ecuación (38) a partir de la cual se puede generar una raíz cuadrada.

En el Bloque 1350, se genera una estimación de la señal de audio de múltiples canales en base al factor de equilibrio y la al menos una señal de audio codificada escalada. La estimación se genera en base a la(s) señal(es) codificada(s) escalada(s) y el factor de equilibrio generado. La estimación puede comprender un número de estimaciones que corresponden a la pluralidad de señales de audio codificadas candidatas. Un valor de distorsión se evalúa y/o se puede generar en base a 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 de ganancia óptimo de los valores de ganancia en el Bloque 1360. El valor de distorsión puede comprender una pluralidad de valores de distorsión correspondientes a la pluralidad de estimaciones. La evaluación del valor de distorsión se consume por la circuitería del selector de ganancia. La presentación de un valor de ganancia óptima se da por la Ecuación (39). En el Bloque 1370, se puede sacar una representación del valor de ganancia o bien para transmisión y/o bien almacenamiento. El transmisor del codificador de la capa de mejora puede transmitir la representación del valor de ganancia como se describió previamente.

El proceso incorporado en el diagrama de flujo 1400 de la FIG. 14 ilustra la descodificación de una señal de audio de múltiples canales. En el Bloque 1410, se reciben una señal de audio codificada, un factor de equilibrio codificado y un valor de ganancia codificado. Un valor de ganancia descodificado se genera a partir del valor de ganancia codificado en el Bloque 1420. El valor de ganancia puede ser una matriz de ganancia, previamente descrita y la matriz de ganancia puede ser selectiva en frecuencia. La matriz de ganancia también puede ser dependiente del audio codificado recibido como una salida de la capa central. Además, la señal de audio codificada puede ser o bien una señal mono o bien una de múltiples canales, tal como una señal estéreo que se ilustra a modo de ejemplo en los dibujos. Adicionalmente, la señal de audio codificada puede comprender una pluralidad de canales. Por ejemplo, puede haber más de un canal en la capa central y el número de canales en la capa de mejora puede ser mayor que el número de canales en la capa central.

En el Bloque 1430, la señal de audio codificada se escala con el valor de ganancia descodificado para generar una señal de audio escalada. El factor de equilibrio codificado se aplica a la señal de audio escalada para generar una señal de audio de múltiples canales descodificada en el Bloque 1440. La señal de audio de múltiples canales descodificada se saca al Bloque 1450.

Cálculo de máscara de escalado selectivo basado en detección de pico

La matriz de ganancia selectiva en frecuencia \mathbf{G}_j , que es una matriz diagonal con elementos diagonales que forman un vector de ganancia \mathbf{g}_j , se puede definir como en (14) anterior:

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

(40)

donde Δ es un tamaño de paso (por ejemplo, $\Delta \approx 2,0$ dB), α es una constante, M es el número de candidatos (por ejemplo, M=8, que se puede representar usando solamente 3 bits), y k_l y k_h son los límites de frecuencia bajo y alto,

respectivamente, sobre los cuales puede tener lugar la reducción de ganancia. Aquí k representa el coeficiente de Transformada de Fourier o MDCT de orden k . Señalar que \mathbf{g}_j es selectivo en frecuencia pero es independiente de la salida de la capa previa. Los vectores de ganancia \mathbf{g}_j se pueden basar en alguna función de los elementos codificados de un vector de señal codificado previamente, en este caso $\hat{\mathbf{S}}$. Esto se puede expresar como:

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

(41)

5

En un sistema de codificación integrada de múltiples capas (con más de 2 capas), en que la salida $\hat{\mathbf{S}}$ que va a ser escalada por el vector de ganancia \mathbf{g}_j , se obtiene a partir de la contribución de al menos dos capas previas, Es decir

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

(42)

donde $\hat{\mathbf{S}}_1$ es la salida de la primera capa (capa central) y $\hat{\mathbf{E}}_2$ es la contribución de la segunda capa o la primera capa de mejora. En este caso los vectores de ganancia \mathbf{g}_j pueden ser alguna función de los elementos codificados de un vector de señal codificado previamente \mathbf{S} y la contribución de la primera capa de mejora:

10

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

(43)

Se puede observar que la mayoría del ruido audible debido al modelo de codificación de la capa inferior está en los valles y no en los picos. En otras palabras, hay una coincidencia mejor entre el espectro original y el codificado en los picos espectrales. De esta manera no se deberían alterar los picos, es decir, el escalado se debería limitar a los valles. Para usar ventajosamente esta observación, en una de las realizaciones la función en la ecuación (41) se basa en los picos y valles de \mathbf{S} . Dejemos que $\psi(\hat{\mathbf{S}})$ sea una máscara de escalado basada en las magnitudes pico de \mathbf{S} detectadas. La máscara de escalado puede ser una función de valores vectoriales con valores no cero en los picos detectados, es decir

15

$$\psi(\hat{\mathbf{S}}) = \begin{cases} \hat{s}_i & \text{presente pico} \\ 0 & \text{De otro modo} \end{cases}$$

(44)

20

donde \hat{s}_i es el elemento de orden i de \mathbf{S} . La ecuación (41) se puede modificar ahora como:

$$g_j(k) = f(k, \hat{\mathbf{S}}) = \begin{cases} \alpha 10^{(-j \cdot \Delta / 20)}; & k_i \leq k \leq k_h, \psi_k(\hat{\mathbf{S}}) = 0 \\ \alpha; & \text{de otro modo} \end{cases} \quad 0 \leq j < M$$

(45)

Se pueden usar diversos planteamientos para detección de pico. En la realización preferida, los picos se detectan pasando el espectro absoluto $|\hat{\mathbf{S}}|$ a través de dos filtros promedio ponderados separados y entonces comparar las salidas filtradas. Dejemos que \mathbf{A}_1 y \mathbf{A}_2 sean la representación de matriz de dos filtros promedio. Dejemos que l_1 e l_2 ($h > l_2$) sean las longitudes de los dos filtros. La función de detección de pico se da como:

25

$$\psi(\hat{\mathbf{S}}) = \begin{cases} \hat{s}_i & \mathbf{A}_2 |\hat{\mathbf{S}}| > \beta \cdot \mathbf{A}_1 |\hat{\mathbf{S}}| \\ 0 & \text{De otro modo} \end{cases}$$

(46)

donde β es un valor umbral empírico.

Como un ejemplo ilustrativo, se refiere a la FIG. 15 y la FIG. 16. Aquí, el valor absoluto de la señal codificada $|\hat{S}|$ en el dominio de MDCT se da en ambas gráficas como 1510. Esta señal es representativa de un sonido desde un “diapasón”, que crea una secuencia armónica separada regularmente como se muestra. Esta señal es difícil de codificar usando un codificador de capa central basado en un modelo de habla debido a que la frecuencia fundamental de esta señal está más allá del intervalo del que se considera razonable para una señal de habla. Esto provoca un nivel bastante alto de ruido producido por la capa central, que se puede observar comparando la señal codificada 1510 a la versión mono 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 $\beta A_1 |\hat{S}|$ en la ecuación 45. Aquí A_1 es una matriz de convolución que, en la realización preferida, implementa una convolución de la señal $|\hat{S}|$ con una ventana de coseno de longitud 45. Son posibles muchas formas de ventanas y pueden comprender diferentes longitudes. También, en la realización preferida, A_2 es una matriz de identidad. El detector de pico entonces compara la señal 1510 con el umbral 1520 para producir la máscara de escalado $\psi(\hat{S})$, mostrada como 1530.

Los candidatos de vector de escalado de la capa central (dados en la ecuación 45) se pueden usar entonces para escalar el ruido entre medias de picos de la señal codificada $|\hat{S}|$ para producir una señal reconstruida escalada 1620. El candidato óptimo se puede elegir según el proceso descrito en la ecuación 39 anterior o de otro modo.

Con referencia ahora a las FIG 17-19, están presente diagramas de flujo que ilustran la metodología asociada con el cálculo de máscara de escalado selectivo en base a la detección de pico tratada anteriormente según diversas realizaciones. En el diagrama de flujo 1700 de la FIG. 17, en el Bloque 1710 se detecta un conjunto de picos en un vector de audio reconstruido S de una señal de audio recibida. La señal de audio puede estar integrada en múltiples capas. El vector de audio reconstruido S puede estar en el dominio de la frecuencia y el conjunto de picos puede ser picos en el dominio de la frecuencia. La detección del conjunto de picos se realiza según una función de detección de picos dada por la ecuación (46), por ejemplo. Se señala que el conjunto puede estar vacío, como es el caso en que todo está atenuado y no hay picos. En el Bloque 1720, se genera una máscara de escalado $\psi(\hat{S})$ basada en el conjunto de picos detectados. Entonces, en el Bloque 1730, se genera un vector de ganancia g^* basado en al menos la máscara de escalado y un índice j representativo del vector de ganancia. En el Bloque 1740, se escala la señal de audio reconstruida con el vector de ganancia para producir una señal de audio reconstruida escalada. Una distorsión basada en la señal de audio y la señal de audio reconstruida escalada se genera en el Bloque 1750. El índice del vector de ganancia basado en la distorsión generada se saca al Bloque 1760.

Con referencia ahora a la FIG. 18, el diagrama de flujo 1800 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 integrada en múltiples capas. La señal de audio se codifica entonces en el Bloque 1820 para generar un vector de audio reconstruido \hat{S} . El vector de audio reconstruido S puede estar en el dominio de la frecuencia y el conjunto de picos pueden ser picos en el dominio de la frecuencia. En el Bloque 1830, se detectan un conjunto de picos en el vector de audio reconstruido S de la señal de audio recibida. La detección del conjunto de picos se realiza según una función de detección de picos dada por la ecuación (46), por ejemplo. De nuevo, se señala que el conjunto puede estar vacío, como es el caso en el que todo está atenuado y no hay picos. Una máscara de escalado $\psi(\hat{S})$ basada en el conjunto de picos detectado se genera en el Bloque 1840. En el Bloque 1850, se genera una pluralidad de vectores de ganancia g_j basados en la máscara de escalado. La señal de audio reconstruida se escala con la pluralidad de vectores de ganancia para producir una pluralidad de señales de audio reconstruidas escaladas en el Bloque 1860. A continuación, se generan una pluralidad de distorsiones basadas en la señal de audio y la pluralidad de señales de audio reconstruidas escaladas en el Bloque 1870. Se elige un vector de ganancia a partir de la pluralidad de vectores de ganancia basado en la pluralidad de distorsiones en el Bloque 1880. El vector de ganancia se puede elegir que corresponda con una distorsión mínima de la pluralidad de distorsiones. El índice representativo del vector de ganancia se saca para ser transmitido y/o almacenado en el Bloque 1890.

Los flujos del codificador ilustrados en las FIG. 17-18 anteriores se pueden implementar por la estructura de 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 de ganancia 1035 del generador de vector de ganancia 1020 del codificador de capa de mejora 1010, detecta un conjunto de picos en un vector de audio reconstruido \hat{S} de una señal de audio recibida y genera una máscara de escalado $\psi(\hat{S})$ basada en el conjunto de picos detectados. De nuevo, la señal de audio puede estar integrada en múltiples capas. El vector de audio reconstruido S puede estar en el dominio de la frecuencia y el conjunto de picos pueden ser picos en el dominio de la frecuencia. La detección del conjunto de picos se realiza según una función de detección de picos dada por la ecuación (46), por ejemplo. Se señala que el conjunto de picos puede ser nulo si todo en la señal ha sido atenuado. Una unidad de escalado, tal

como la unidad de escalado 1025 del generador de vector de ganancia 1020 genera un vector de ganancia \mathbf{g}^* basado en al menos la máscara de escalado y un índice j representativo del vector de ganancia, escala la señal de audio reconstruida con el vector de ganancia para producir una señal de audio reconstruida escalada. El generador de señal de error 1030 del generador de vector de ganancia 1025 genera una distorsión basada en la señal de audio y la señal de audio reconstruida escalada. Un transmisor, tal como el transmisor 1045 del descodificador de la capa de mejora 1010 es operable para sacar el índice del vector de ganancia en base a la distorsión generada.

Con referencia al flujo 1800 de la FIG. 18, en un aparato operable para codificar una señal de audio, un codificador recibió una señal de audio y codifica la señal de audio para generar un vector de audio reconstruido \hat{S} . Una unidad de escalado tal como la unidad de escalado 1025 del generador de vector de ganancia 1020 detecta un conjunto de picos en el vector de audio reconstruido S de una señal de audio recibida, genera una máscara de escalado $\psi(\hat{S})$ basada en el conjunto de picos detectados, genera una pluralidad de vectores de ganancia \mathbf{g}_j basados en la máscara de escalado, y escala la señal de audio reconstruida con la pluralidad de vectores de ganancia para producir la pluralidad de señales de audio reconstruidas escaladas. El generador de señal de error 1030 genera una pluralidad de distorsiones basadas en la señal de audio y la pluralidad de señales de audio reconstruidas escaladas. Un selector de ganancia tal como el selector de ganancia 1035 elige un vector de ganancia a partir de la pluralidad de vectores de ganancia basados en la pluralidad de distorsiones. El transmisor 1045, por ejemplo, saca para su posterior transmisión y/o almacenamiento, el índice representativo del vector de ganancia.

En el diagrama de flujo 1900 de la FIG. 19, se ilustra un método de descodificación de una señal de audio. Un vector de audio reconstruido S y un índice representativo de un vector de ganancia se reciben en el Bloque 1910. En el Bloque 1920, se detecta el conjunto de picos en el vector de audio reconstruido. La detección del conjunto de picos se realiza según una función de detección de picos dada por la ecuación (46), por ejemplo. De nuevo, se señala que el conjunto puede estar vacío, como es el caso en que todo está atenuado y no hay picos. Se genera en el Bloque 1930 una máscara de escalado $\psi(\hat{S})$ basada en el conjunto de picos detectados. El vector de ganancia \mathbf{g}^* basado en al menos la máscara de escalado y el índice representativo del vector de ganancia se generan en el Bloque 1940. El vector de audio reconstruido se escala con el vector de ganancia para producir una señal de audio reconstruida escalada en el Bloque 1950. El método puede incluir además generar una mejora al vector de audio reconstruido y entonces combinar la señal de audio reconstruida escalada y la mejora al vector de audio reconstruido para generar una señal descodificada mejorada.

El flujo de descodificador ilustrado en la FIG. 19 se puede implementar por la estructura de aparato previamente descrita. En un aparato operable para descodificar una señal de audio, un descodificador de vector de ganancia 1070 de un descodificador de capa de mejora 1060, por ejemplo, recibe un vector de audio reconstruido \hat{S} y un índice representativo de un vector de ganancia i_g . Como se muestra en la FIG. 10, i_g se recibe por el selector de ganancia 1075 mientras que el vector de audio reconstruido S se recibe por la unidad de escalado 1080 del descodificador de vector de ganancia 1070. Un selector de ganancia, tal como el selector de ganancia 1075 del descodificador de vector de ganancia 1070, detecta un conjunto de picos en el vector de audio reconstruido, genera una máscara de escalado $\psi(\hat{S})$ basada en el conjunto de picos detectados, y genera el vector de ganancia \mathbf{g}^* basado en al menos la máscara de escalado y el índice representativo del vector de ganancia. De nuevo, el conjunto puede estar vacío de ficheros si la señal está mayoritariamente atenuada. El selector de ganancia detecta el conjunto de picos según una función de detección de picos tal como aquella dada en la ecuación (46), por ejemplo. Una unidad de escalado 1080, por ejemplo, escala el vector de audio reconstruido con el vector de ganancia para producir una señal de audio reconstruida escalada.

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

Se señala además que los flujos dirigidos de factor de equilibrio de las FIG. 12-14 y la máscara de escalado selectiva con flujos dirigidos de detección de pico de las FIG. 17-19 se pueden realizar ambos en diversa combinación y tal se soporta por el aparato y la estructura descrita en la presente memoria.

Aunque la invención se ha mostrado y descrito particularmente con referencia a una realización particular, se entenderá por los expertos en la técnica que se pueden hacer diversos cambios en la forma y detalles dentro de la misma sin apartarse del espíritu y alcance de la invención. Por ejemplo, aunque las técnicas anteriores se describen en términos de transmisión y recepción sobre un canal en un sistema de telecomunicación, las técnicas se pueden aplicar igualmente a un sistema que usa el sistema de compresión de señal para los propósitos de reducción de los requerimientos de almacenamiento en un dispositivo de medios digital, tal como un dispositivo de memoria de estado sólido o disco duro de ordenador. El alcance de protección se define en las reivindicaciones adjuntas.

REIVINDICACIONES

1. Un aparato operable para codificar una señal de audio de múltiples canales, que comprende:

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

5 un generador de factor de equilibrio de un codificador de capa de mejora que recibe una señal de audio codificada y genera un factor de equilibrio que tiene una pluralidad de componentes de factor de equilibrio cada uno asociado con una señal de audio de la pluralidad de señales de audio de la señal de audio de múltiples canales;

10 un generador de vector de ganancia de un codificador de capa de mejora que determina un valor de ganancia a ser aplicado a la señal de audio codificada para generar una estimación de la señal de audio de múltiples canales basada en el factor de equilibrio y la señal de audio de múltiples canales, en donde el valor de ganancia se configura para minimizar un valor de distorsión entre la señal de audio de múltiples canales y la estimación de la señal de audio de múltiples canales; y

15 un transmisor que transmite una representación del valor de ganancia para al menos uno de transmisión y almacenamiento.

2. El aparato de la reivindicación 1, que además comprende:

una unidad de escalado del codificador de capa de mejora que escala la señal de audio codificada con una pluralidad de valores de ganancia para generar una pluralidad de señales de audio codificadas candidatas, en donde se escala al menos una de las señales de audio codificadas candidatas;

20 en donde la unidad de escalado y el generador de factor de equilibrio generan la estimación de la señal de audio de múltiples canales en base al factor de equilibrio y la al menos una señal de audio codificada escalada de la pluralidad de señales de audio codificadas candidatas; y

25 un selector de ganancia del codificador de capa de mejora que evalúa el valor de distorsión en base a 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 de ganancia óptimo de la pluralidad de valores de ganancia.

3. El aparato de la reivindicación 1, en donde el codificador codifica la señal de audio para generar un vector de audio reconstruido S y el generador del vector de ganancia que además comprende:

30 una unidad de escalado que detecta un conjunto de picos en el vector de audio reconstruido S de una señal de audio recibida, genera una máscara de escalado $\psi(\hat{S})$ en base al conjunto de picos detectados, genera una pluralidad de vectores de ganancia g_j basados en la máscara de escalado, y escala la señal de audio reconstruida con la pluralidad de vectores de ganancia para producir la pluralidad de señales de audio reconstruidas escaladas;

un generador de señal de error que genera una pluralidad de distorsiones en base a la señal de audio y la pluralidad de señales de audio reconstruidas escaladas; y

35 un selector de ganancia que elige un vector de ganancia a partir de la pluralidad de vectores de ganancia en base a la pluralidad de distorsiones, en donde el transmisor saca para al menos uno de transmisión y almacenamiento el índice representativo del vector de ganancia.

4. El aparato de la reivindicación 3, en donde el selector de ganancia detecta el conjunto de picos adicionales según una función de detección de pico dada como:

$$\psi(\hat{S}) = \begin{cases} \hat{s}_i A_2 |\hat{S}| > \beta \cdot A_1 |\hat{S}|, \\ 0 \text{ De otro modo} \end{cases}$$

40 donde β es un valor umbral.

5. Un aparato operable para codificar una señal de audio de múltiples canales, que comprende:

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

una unidad de escalado de un codificador de capa de mejora que escala la señal de audio codificada con una pluralidad de valores de ganancia para generar una pluralidad de señales de audio codificadas candidatas, en donde se escala al menos una de las señales de audio codificadas candidatas;

5 un generador de factor de equilibrio que genera un factor de equilibrio que tiene una pluralidad de componentes de factor de equilibrio cada uno asociado con una señal de audio de la pluralidad de señales de audio de la señal de audio de múltiples canales;

en donde la unidad de escalado y el generador de factor de equilibrio generan una estimación de la señal de audio de múltiples canales en base al factor de equilibrio y la al menos una señal de audio codificada escalada de la pluralidad de señales de audio codificadas candidatas;

10 un selector de ganancia del codificador de capa de mejora que evalúa un valor de distorsión en base a 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 de ganancia óptimo de la pluralidad de valores de ganancia;

un transmisor que transmite la representación del valor de ganancia óptimo para al menos uno de transmisión y almacenamiento.

15 **6.** El aparato de la reivindicación 5, en donde un valor de ganancia de la pluralidad de valores de ganancia es una matriz de ganancia con el vector g_j como la componente diagonal y en donde la matriz de ganancia es selectiva en frecuencia.

7. El aparato de la reivindicación 5, en donde la representación del valor de ganancia óptimo se da por:

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

20 **8.** El aparato de la reivindicación 5, en donde cada componente del factor de equilibrio se da por:

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

9. El aparato de la reivindicación 5, en donde el generador de factor de equilibrio genera un valor de correlación entre la señal de audio codificada escalada y al menos una de las señales de audio de la señal de audio de múltiples canales.

25 **10.** El aparato de la reivindicación 5, en donde el generador de factor de equilibrio genera una auto correlación entre al menos una de las señales de audio de la señal de audio de múltiples canales y genera una raíz cuadrada de la auto correlación.

30 **11.** El aparato de la reivindicación 5, en donde el selector de ganancia genera un valor de distorsión en base a la estimación de la señal de audio de múltiples canales y la señal de audio de múltiples canales y en donde el valor de ganancia está basado en el valor de distorsión.

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

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

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

35 generar un factor de equilibrio que tenga una pluralidad de componentes de factor de equilibrio cada uno asociado con una señal de audio de la pluralidad de señales de audio de la señal de audio de múltiples canales;

determinar un valor de ganancia a ser aplicado a la señal de audio codificada para generar una estimación de la señal de audio de múltiples canales en base al factor de equilibrio y la señal de audio de múltiples canales, en donde el valor de ganancia se configura para minimizar un valor de distorsión entre la señal de audio de múltiples canales y la estimación de la señal de audio de múltiples canales; y

40 sacar una representación del valor de ganancia para al menos uno de transmisión y almacenamiento.

13. El método de la reivindicación 12, que además comprende:

escalar la señal de audio codificada con una pluralidad de valores de ganancia para generar una pluralidad de señales de audio codificadas candidatas, en donde se escala al menos una de las señales de audio codificadas candidatas;

generar la estimación de la señal de audio de múltiples canales en base al factor de equilibrio y la al menos una señal de audio codificada ponderada de la pluralidad de señales de audio codificadas candidatas; y

5 evaluar el valor de distorsión en base a 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 de ganancia óptimo de la pluralidad de valores de ganancia.

14. El método de la reivindicación 12, que además comprende:

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

generar una máscara de escalado $\psi(\hat{S})$ basada en el conjunto de picos detectados;

10 generar un vector de ganancia g^* basado en al menos la máscara de escalado y un índice j representativo del vector de ganancia;

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

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

sacar el índice del vector de ganancia basado en la distorsión generada.

15 15. El método de la reivindicación 12, que además comprende:

recibir una señal de audio;

codificar la señal de audio para generar un vector de audio reconstruido S ;

detectar un conjunto de picos en el vector de audio reconstruido \hat{S} de una señal de audio recibida;

generar una máscara de escalado $\psi(\hat{S})$ basada en el conjunto de picos detectados;

20 generar una pluralidad de vectores de ganancia g_j basados en la máscara de escalado;

escalar la señal de audio reconstruida con la pluralidad de vectores de ganancia para producir la pluralidad de señales de audio reconstruidas escaladas;

generar una pluralidad de distorsiones basadas en la señal de audio y una pluralidad de señales de audio reconstruidas escaladas;

25 elegir un vector de ganancia de la pluralidad de vectores de ganancia en base a la pluralidad de distorsiones; y

sacar para al menos uno de transmisión y almacenamiento el índice representativo del vector de ganancia.

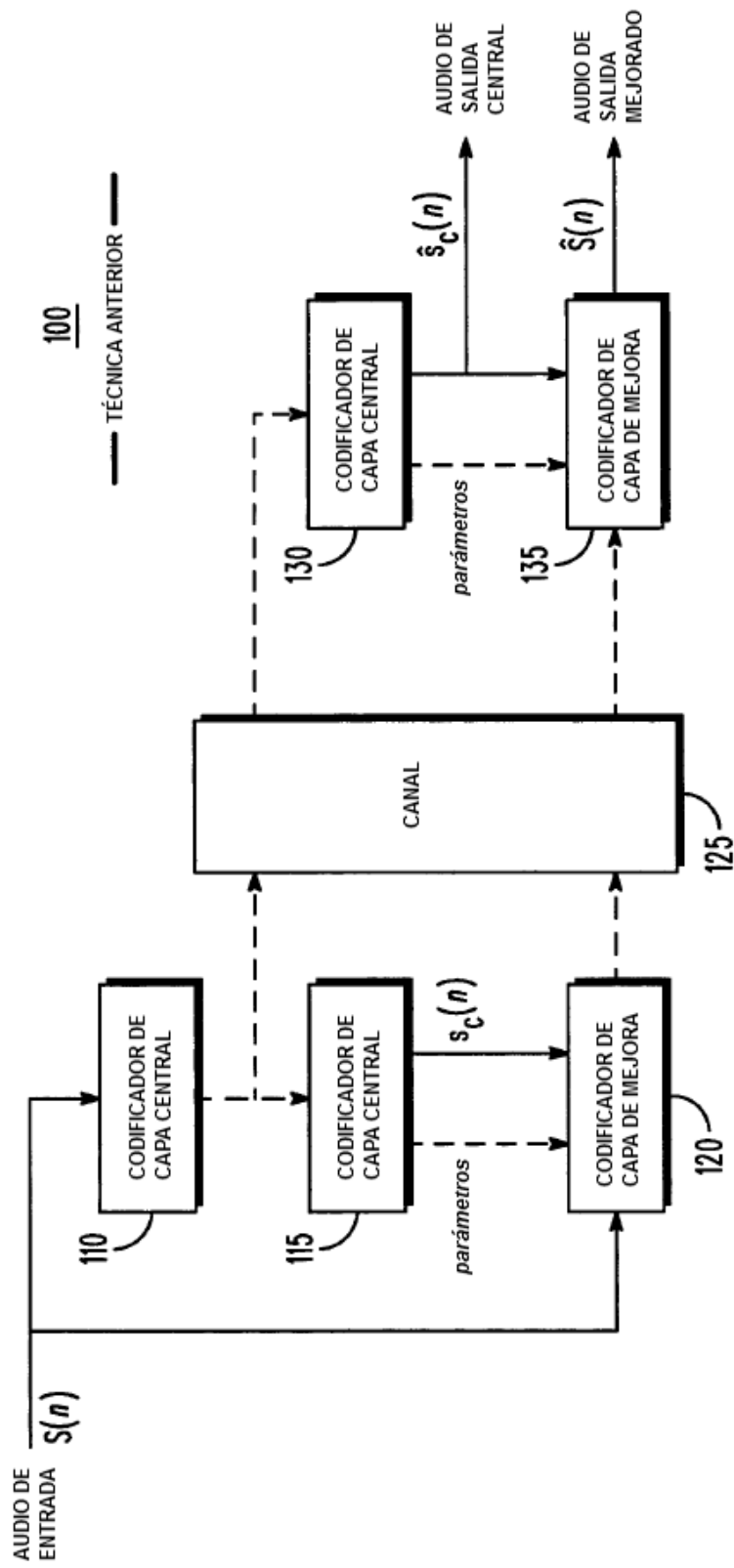
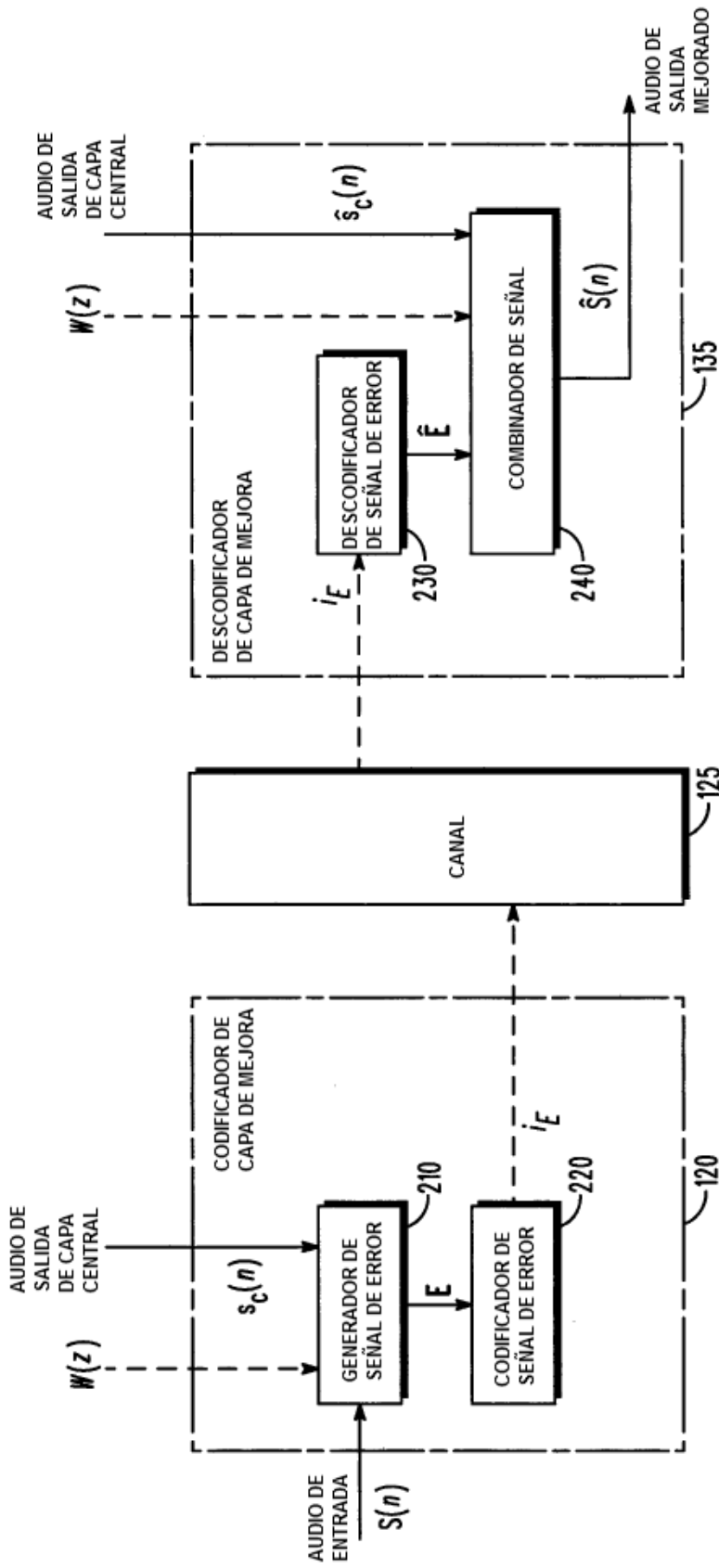
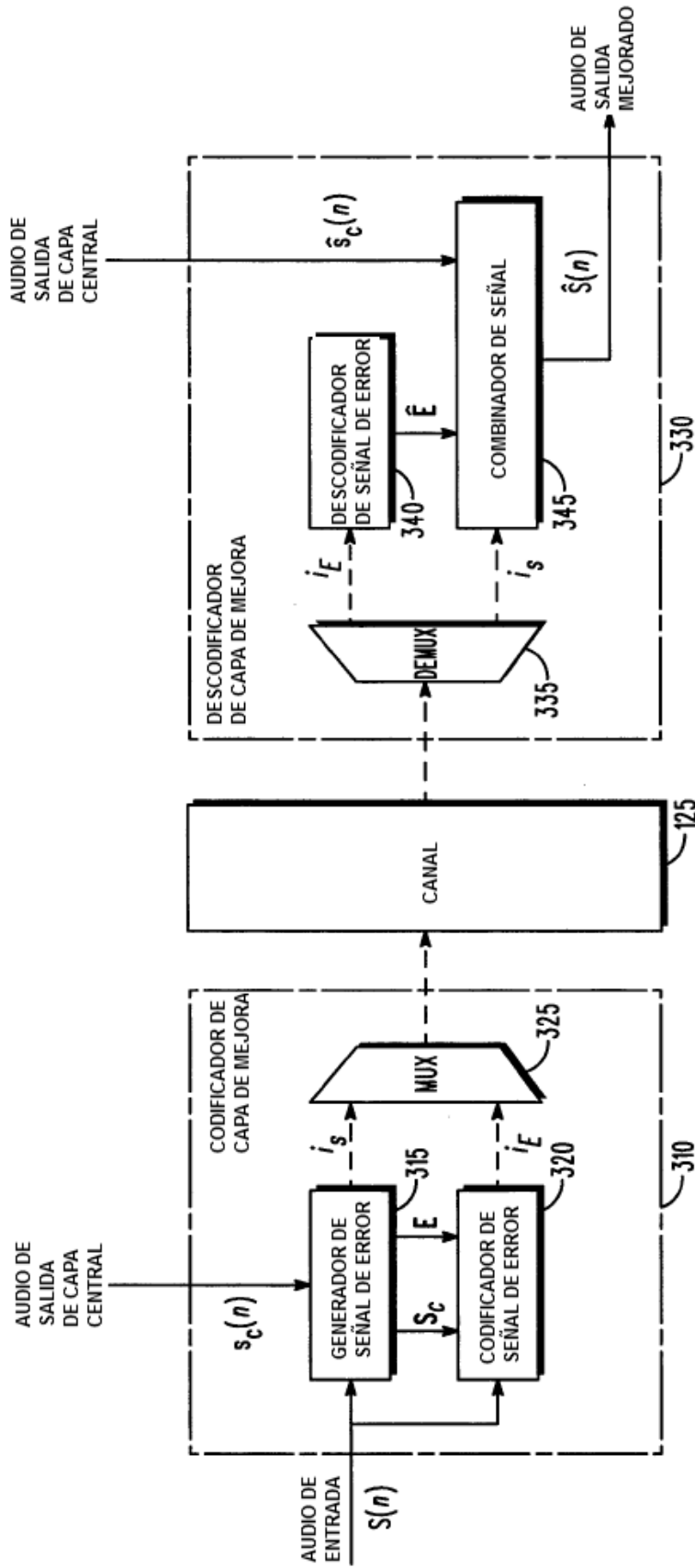


FIG. 1



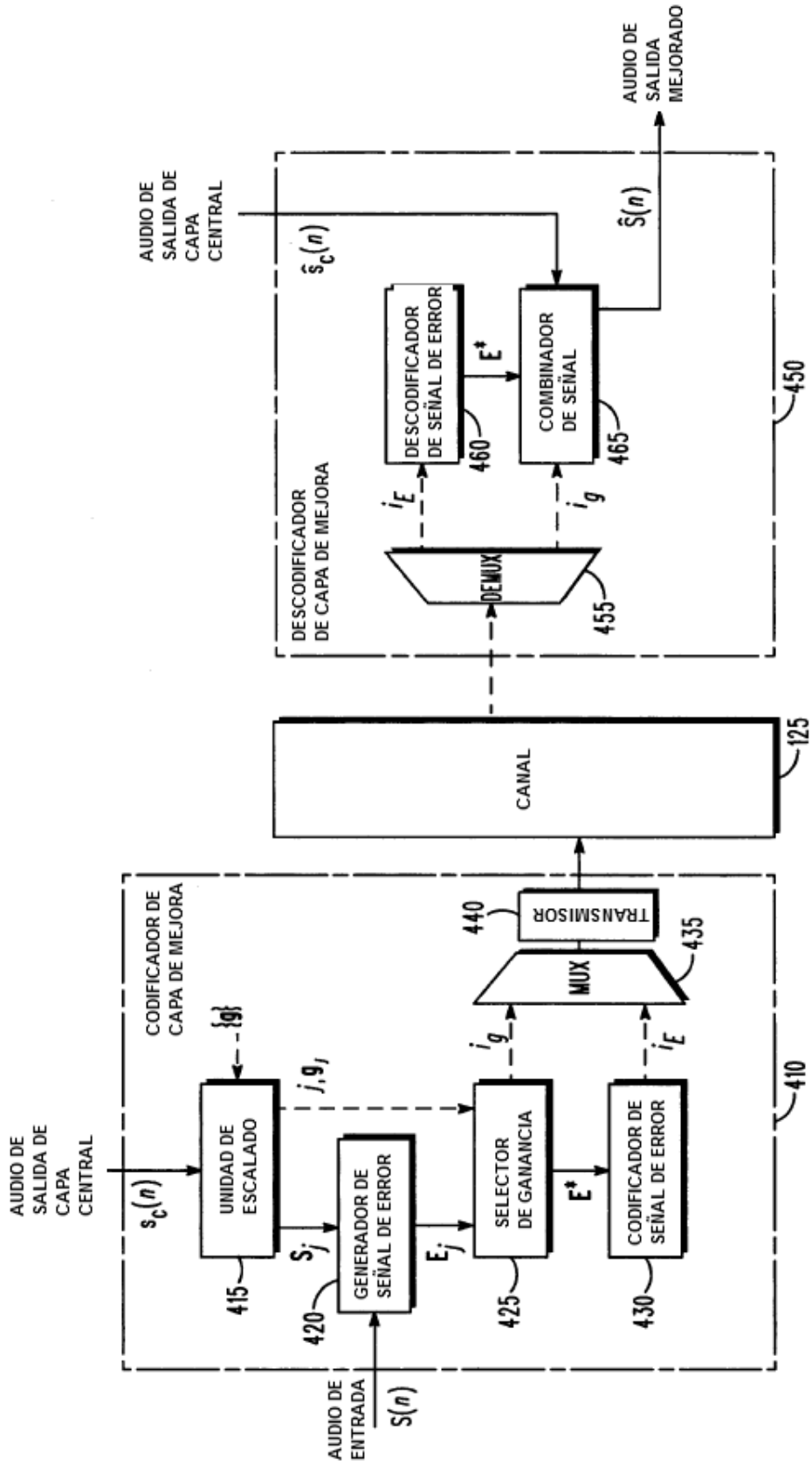
200
 — TÉCNICA ANTERIOR —
FIG. 2



300

— TÉCNICA ANTERIOR —

FIG. 3



400

FIG. 4

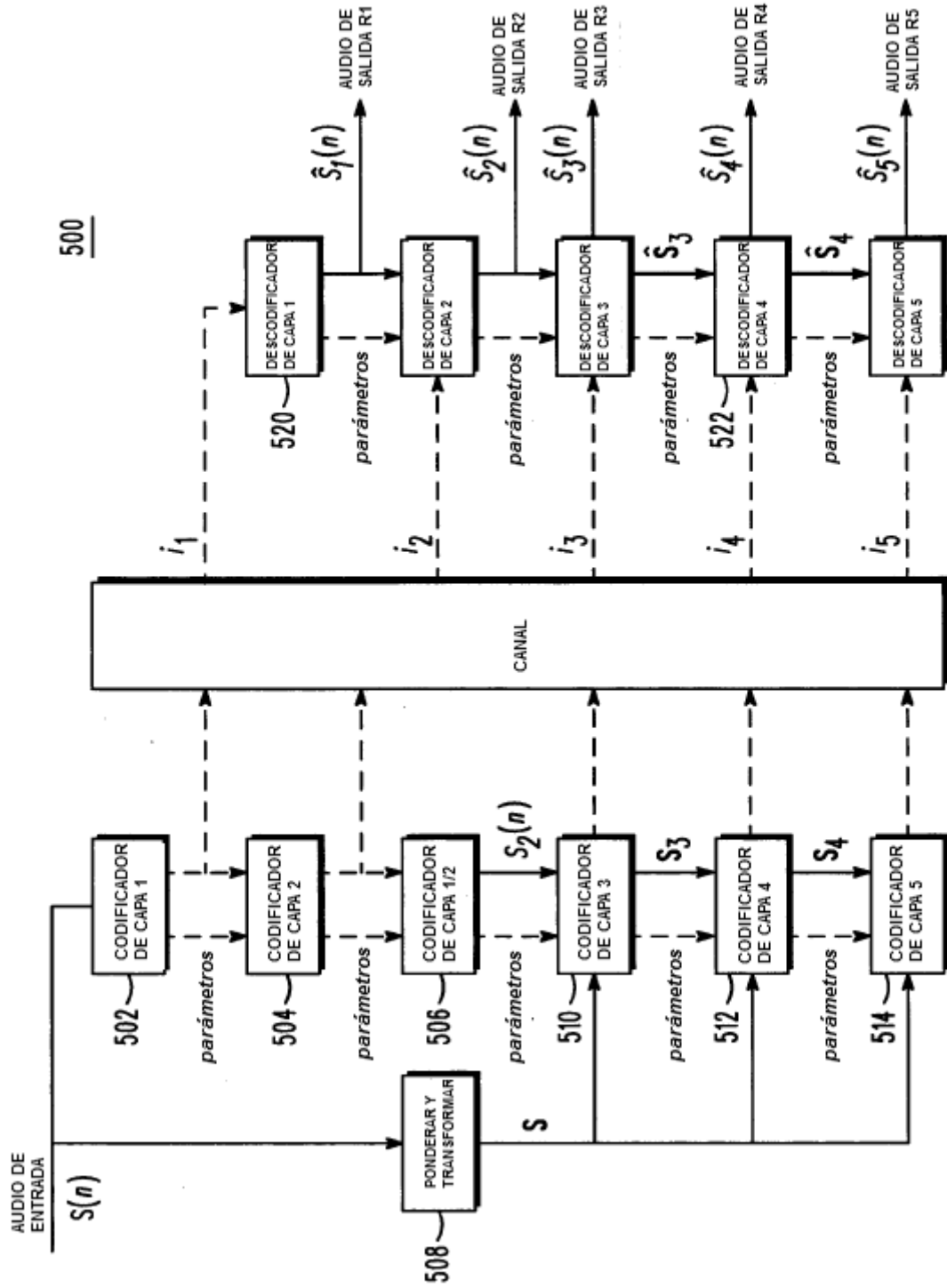
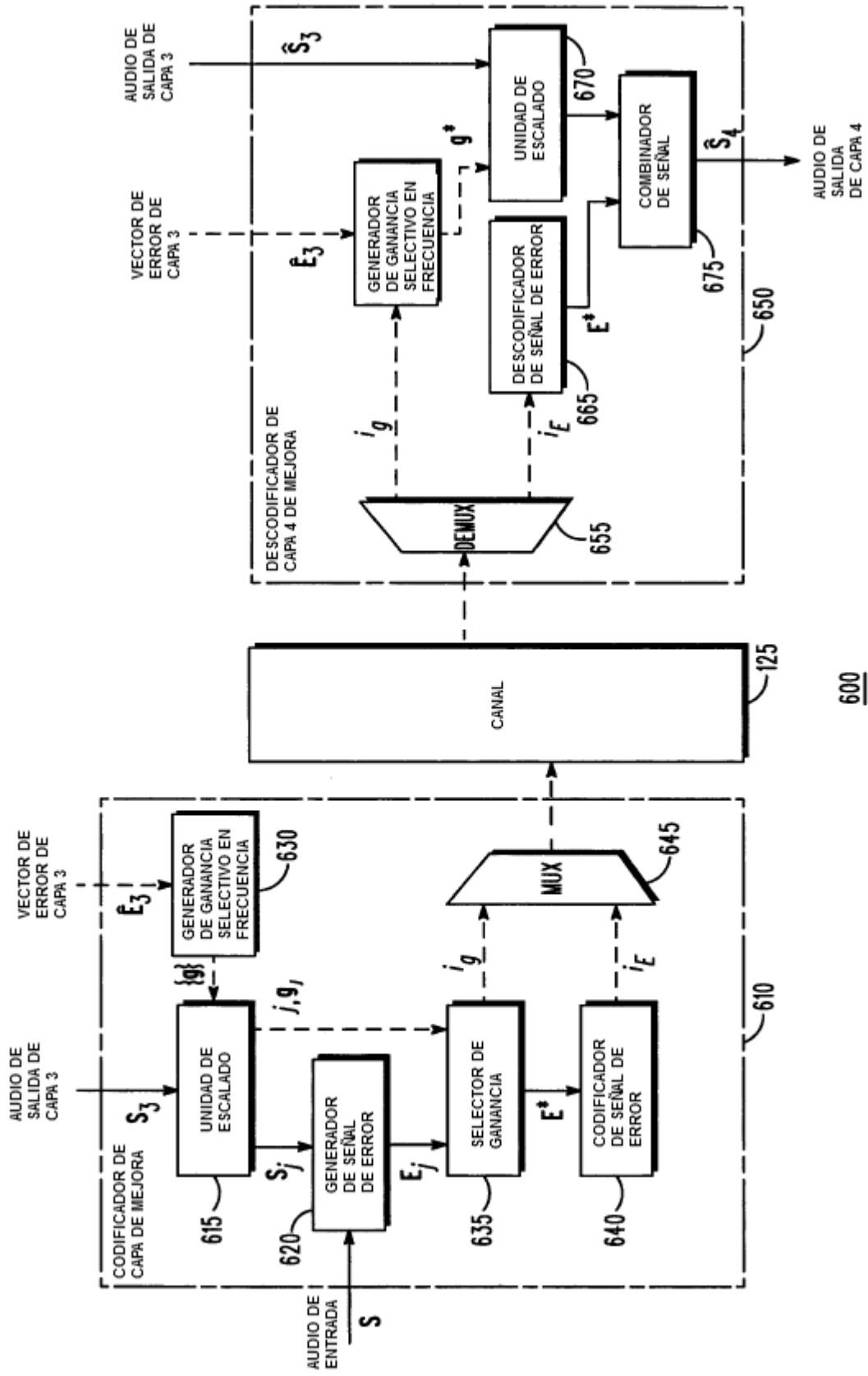


FIG. 5



600

FIG. 6

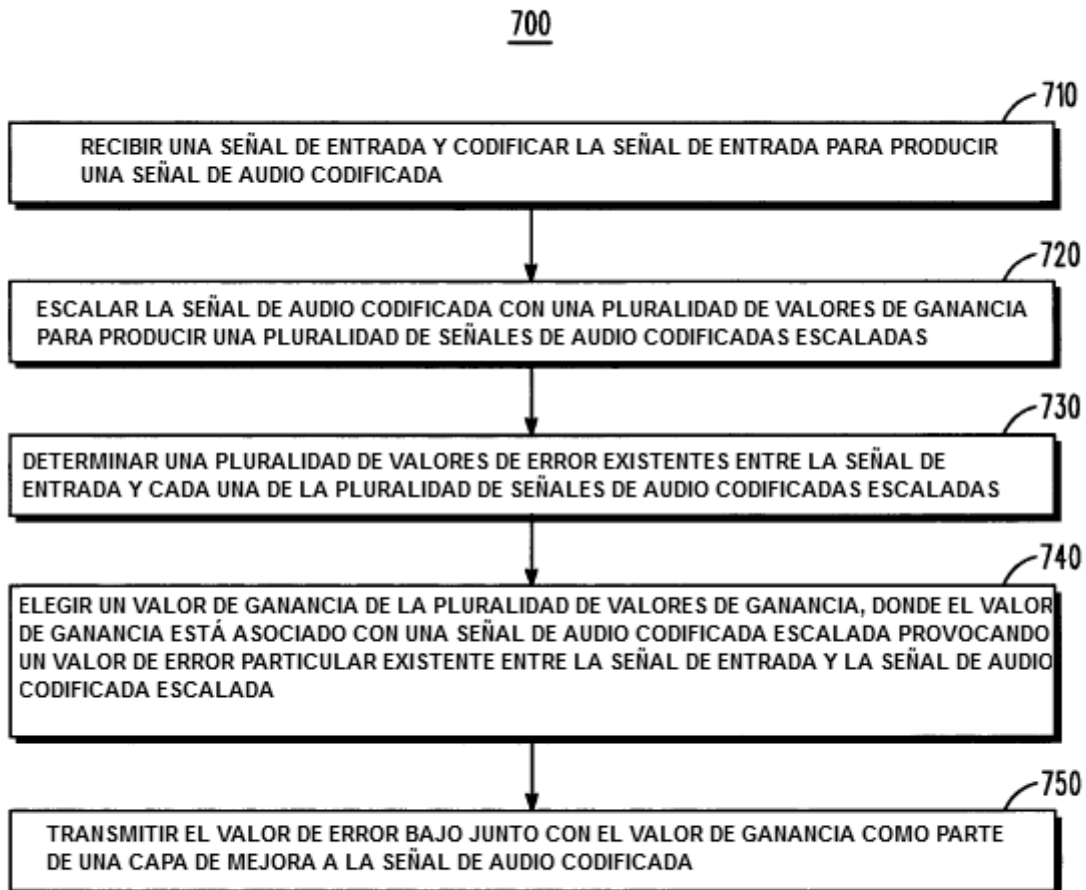


FIG. 7

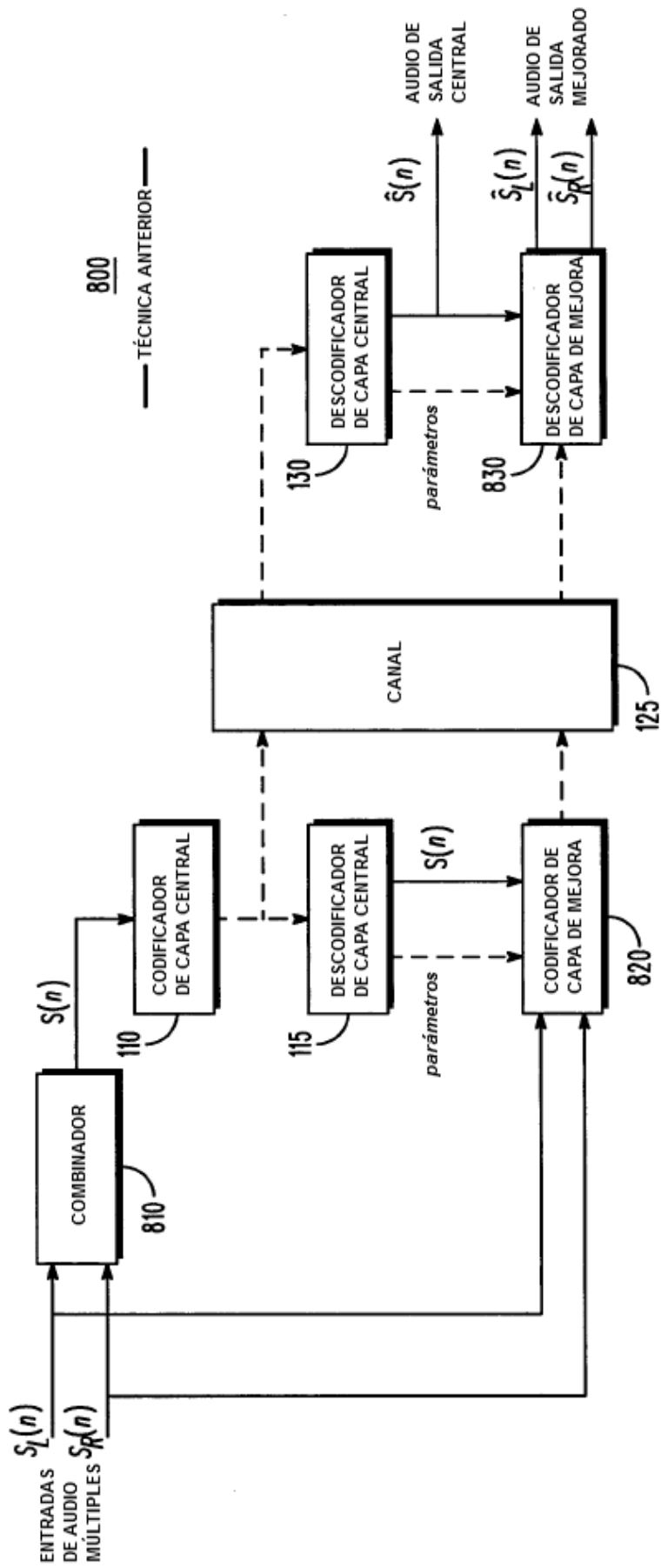


FIG. 8

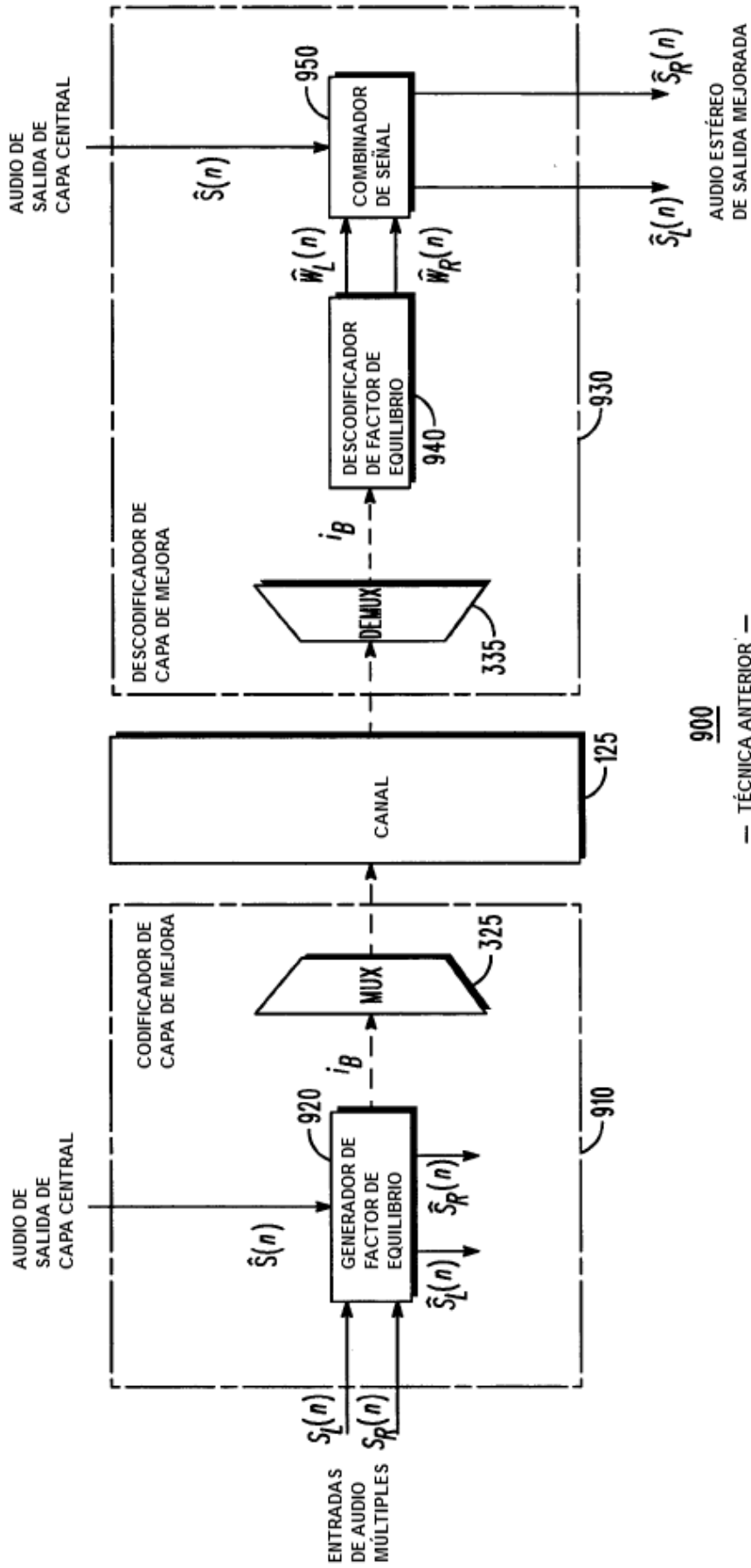
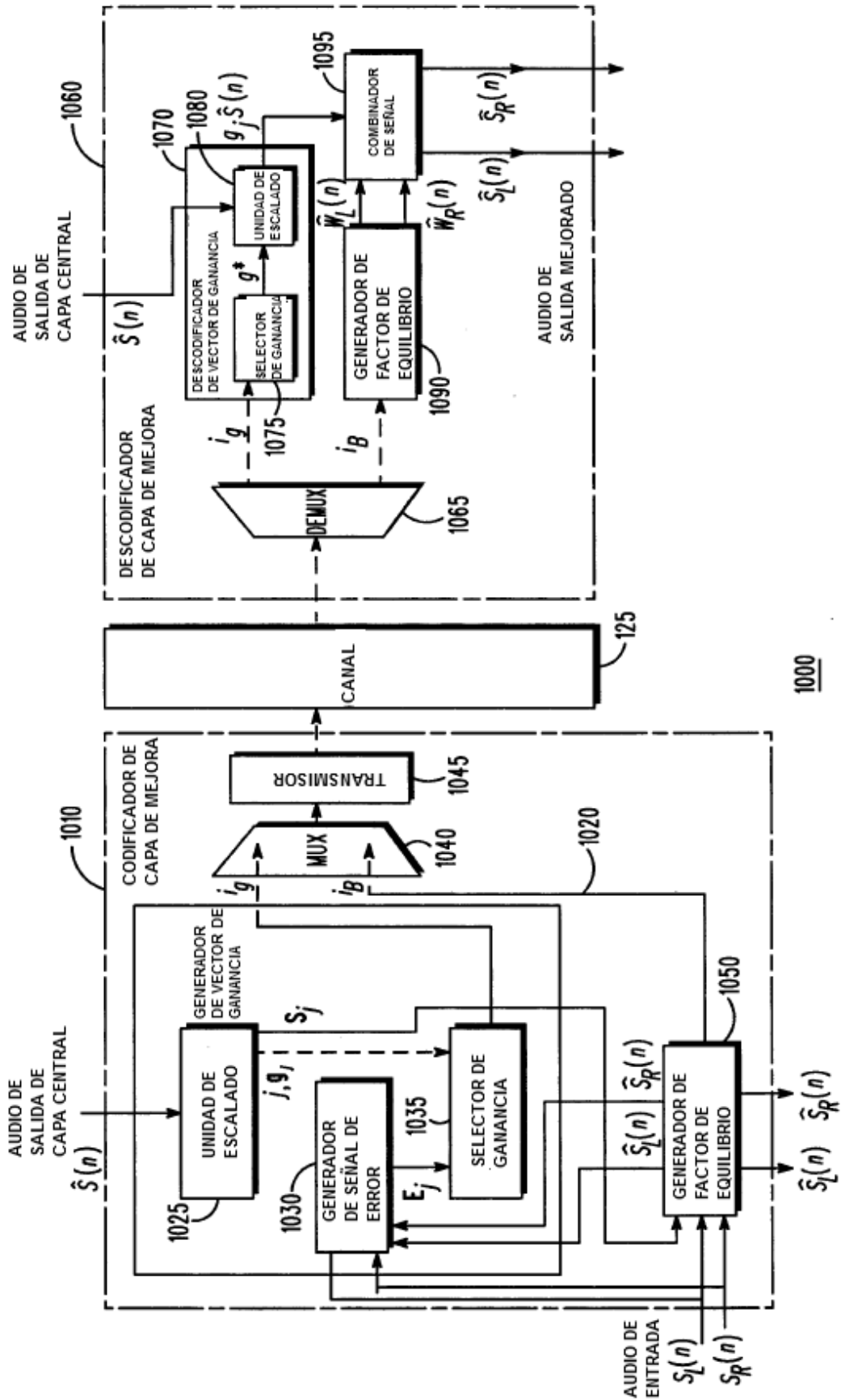


FIG. 9



1000

FIG. 10

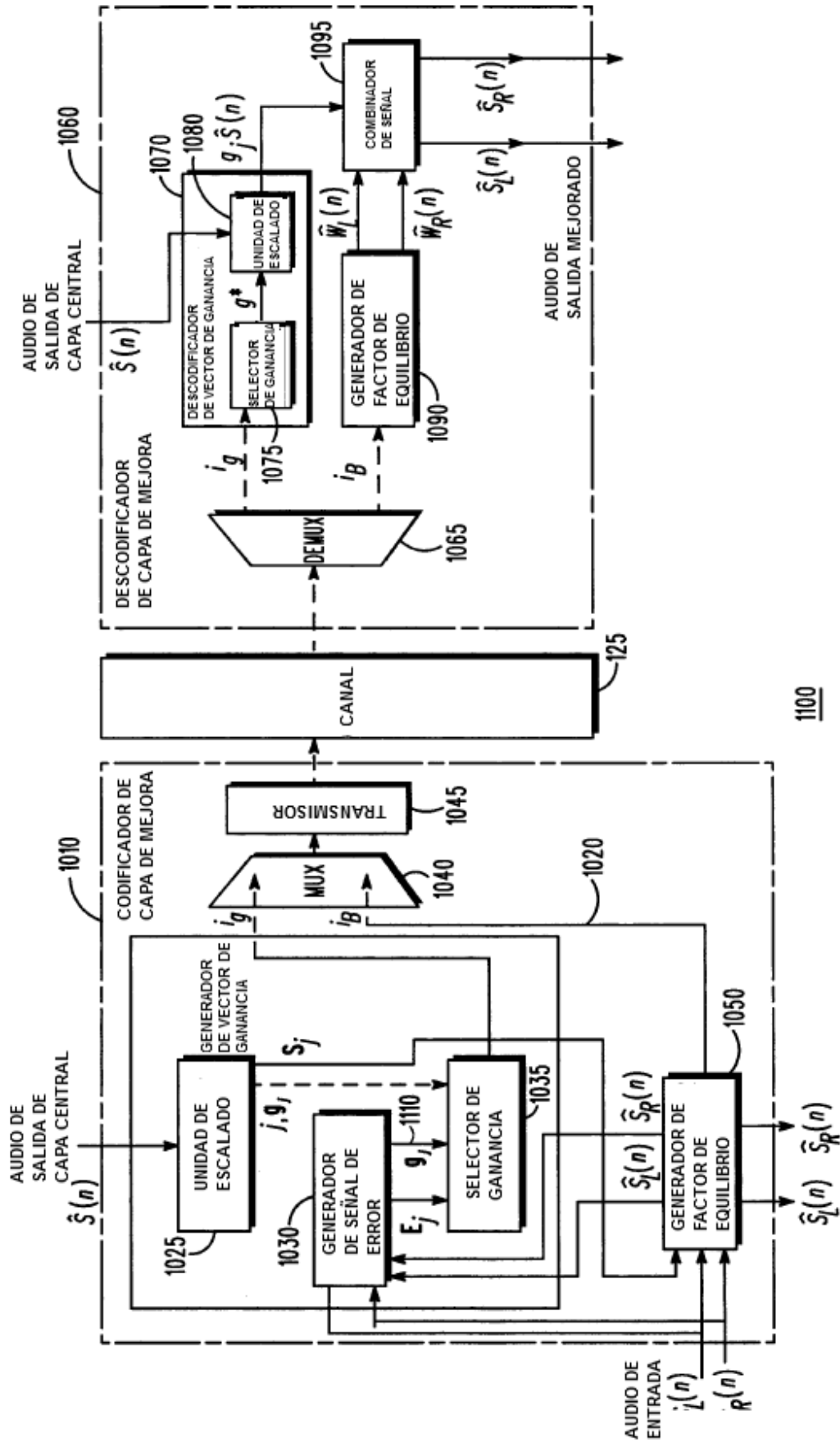


FIG. 11

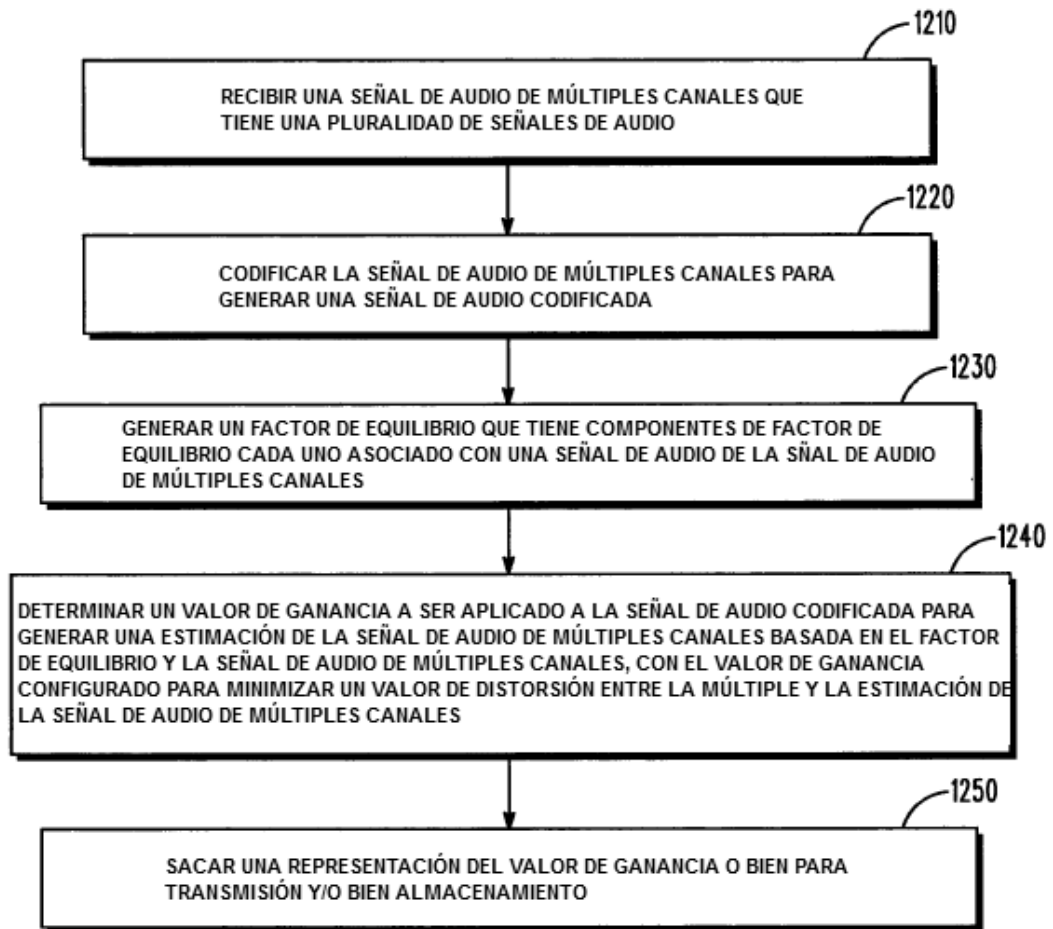
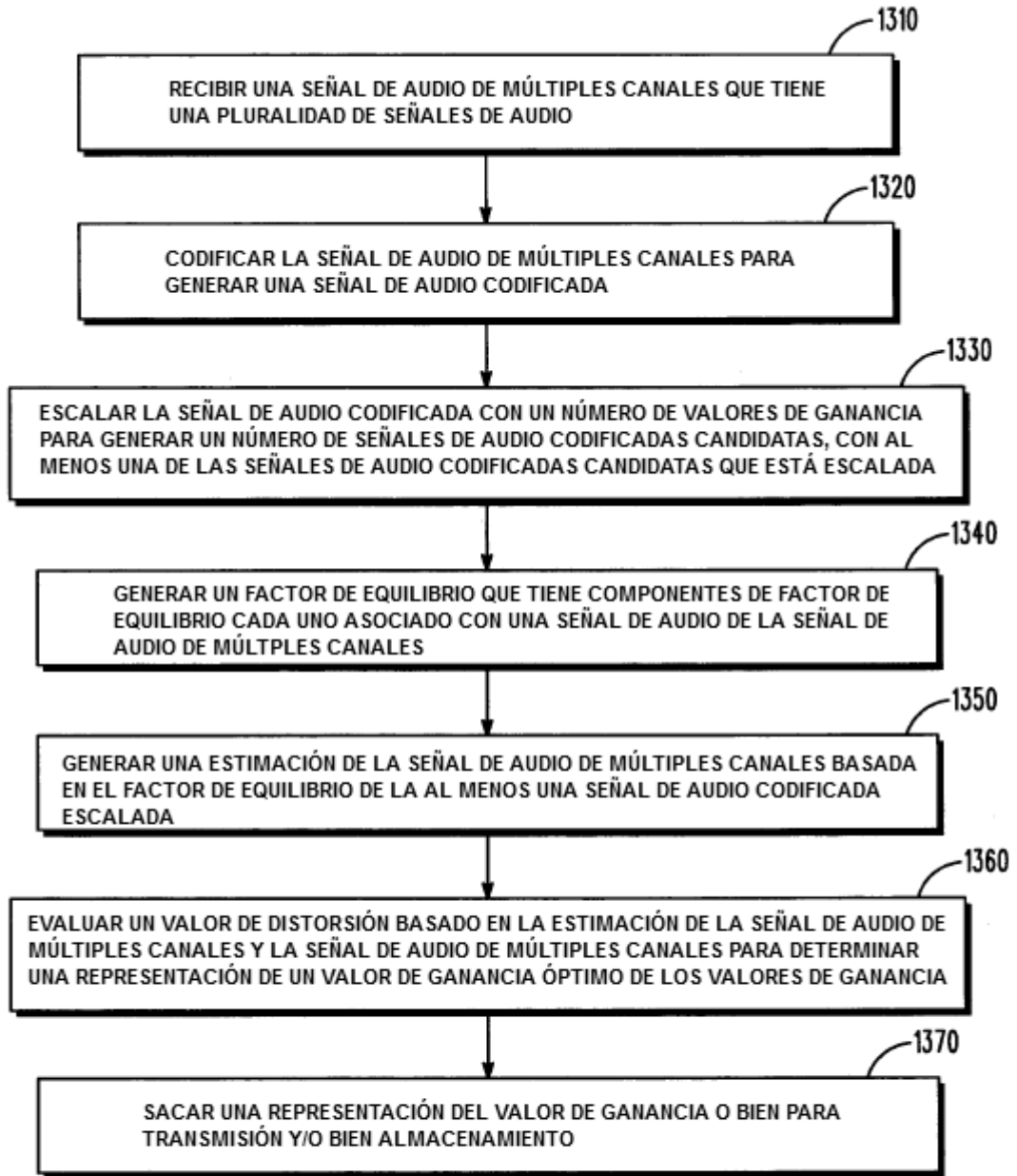


FIG. 12



1300

FIG. 13

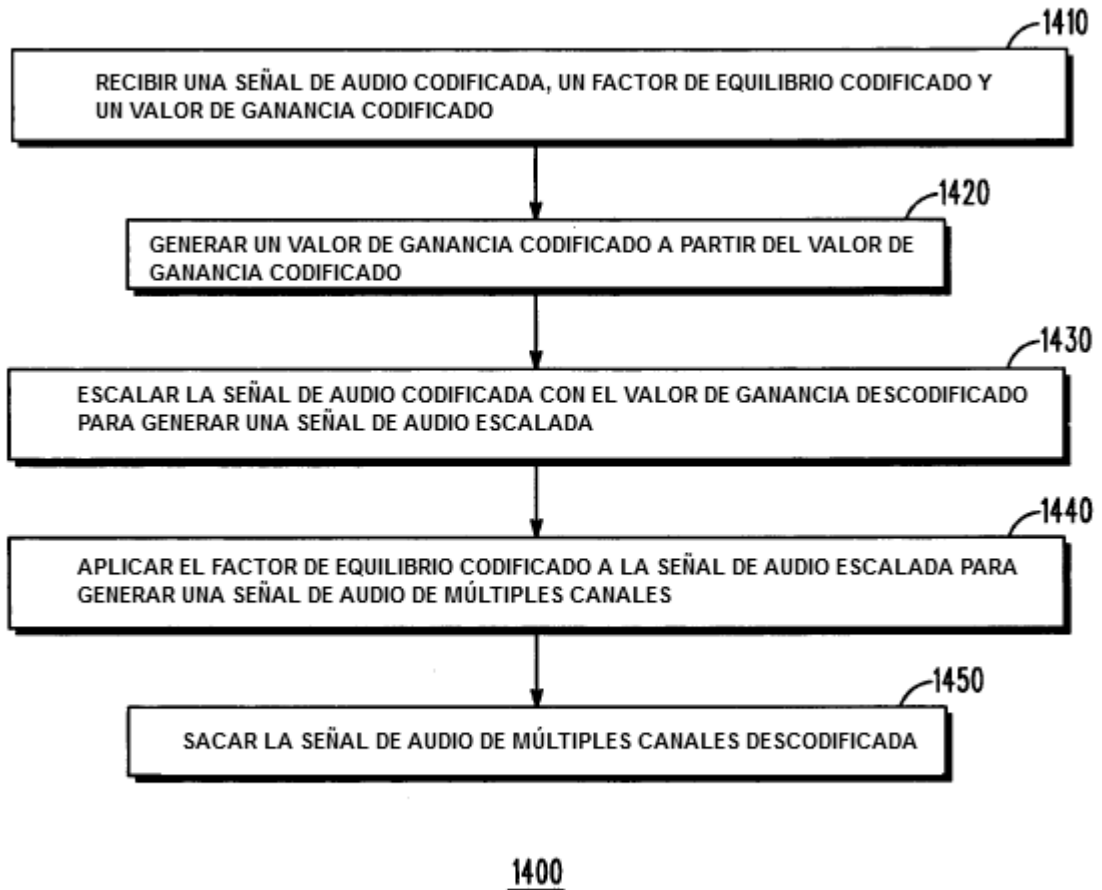


FIG. 14

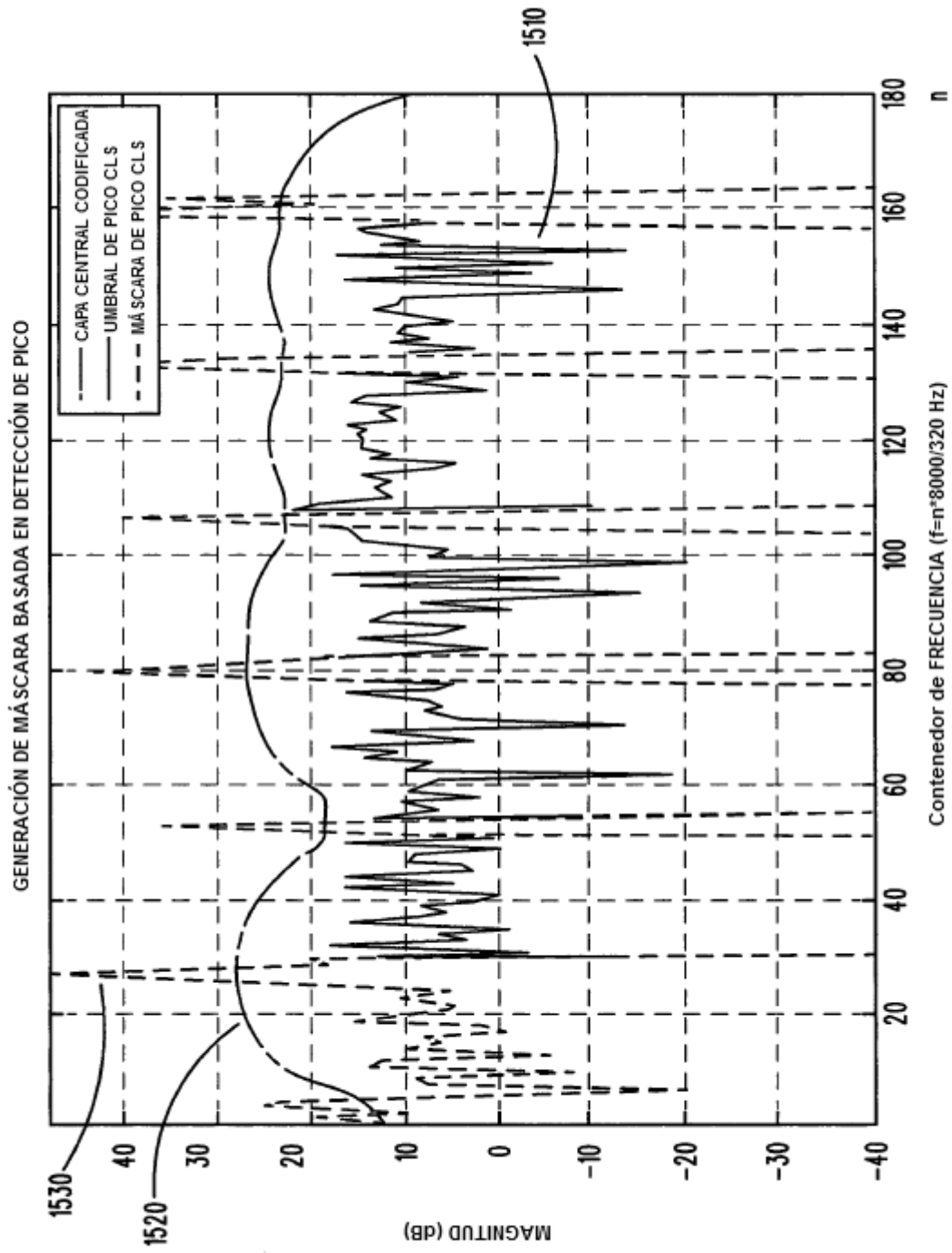


FIG. 15

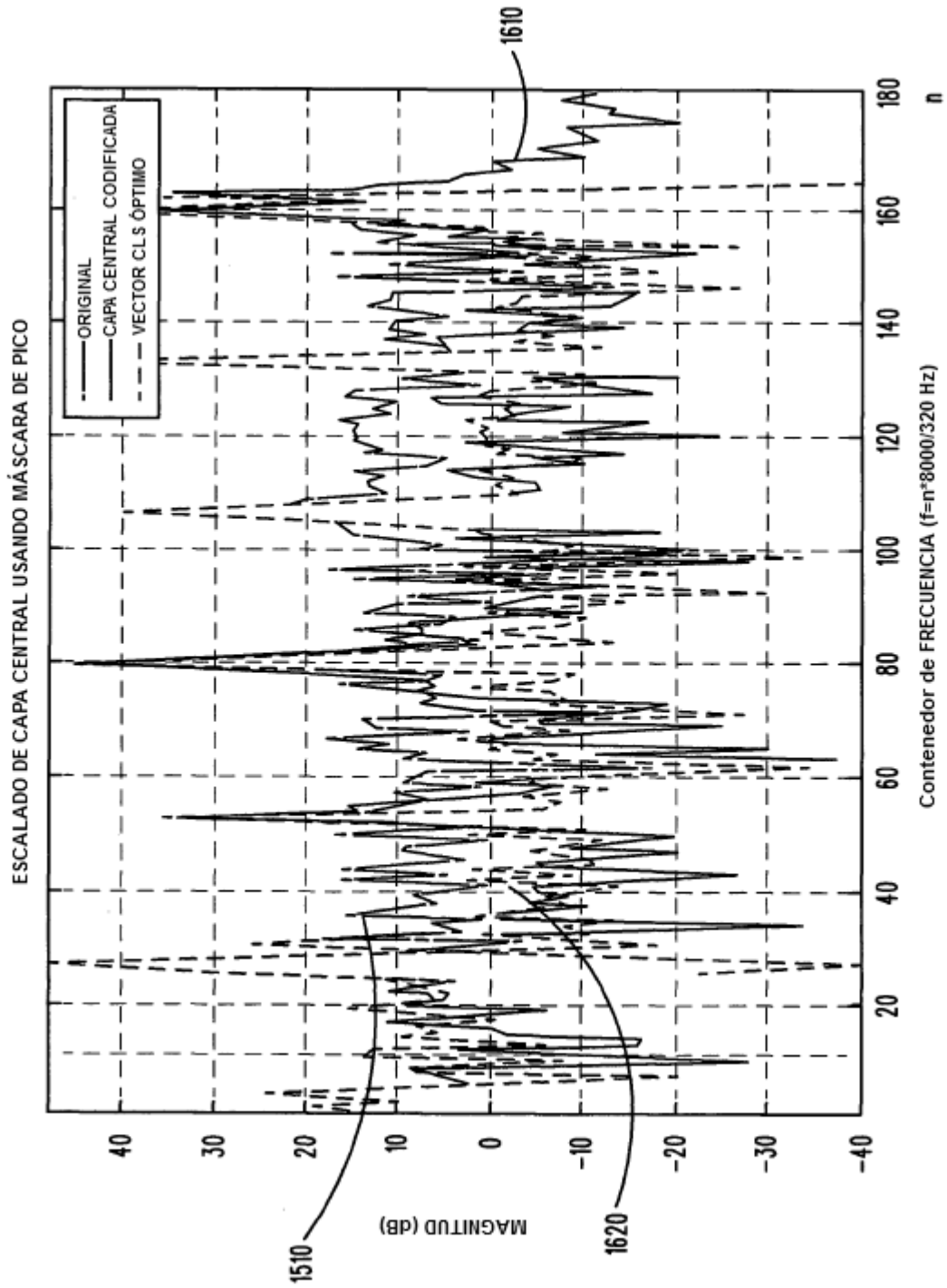


FIG. 16



FIG. 17

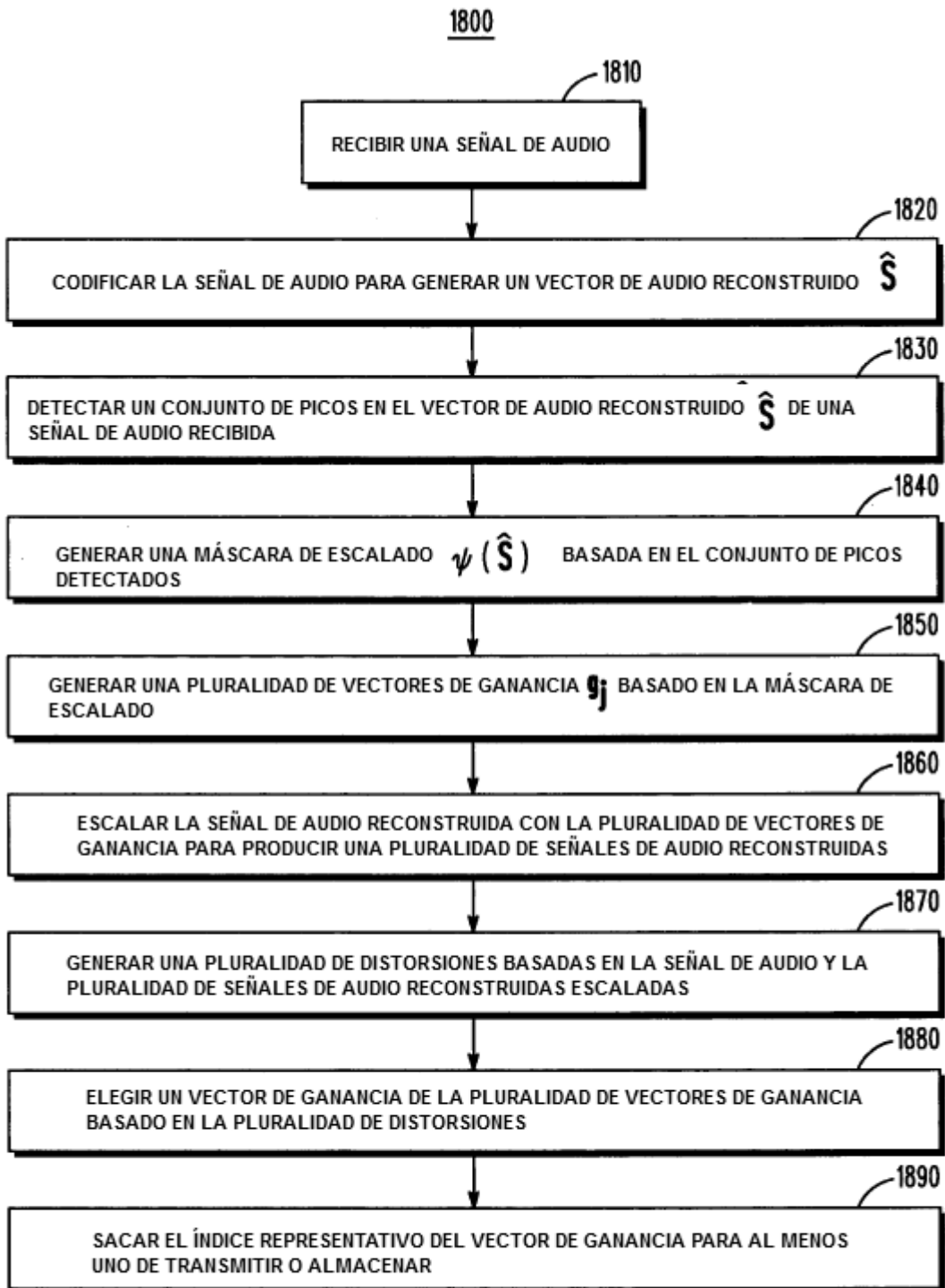


FIG. 18

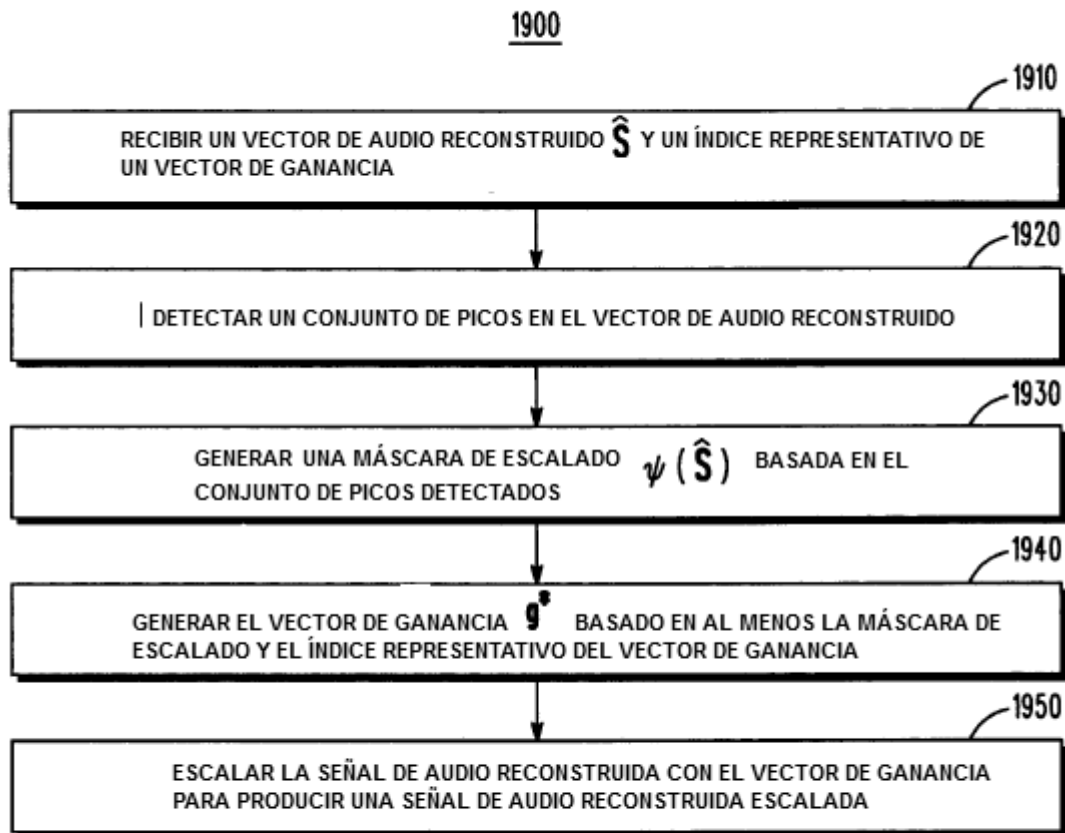


FIG. 19