

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 568 640**

51 Int. Cl.:

G10L 19/02 (2013.01)

G10L 21/0388 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **22.02.2013 E 13705503 (4)**

97 Fecha y número de publicación de la concesión europea: **03.02.2016 EP 2817803**

54 Título: **Procedimientos y sistemas para recuperar de manera eficiente contenido de audio de alta frecuencia**

30 Prioridad:

23.02.2012 EP 12156631
08.08.2012 US 201261680805 P

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

03.05.2016

73 Titular/es:

DOLBY INTERNATIONAL AB (100.0%)
Apollo Building, 3E Herikerbergweg 1-35
1101 CN Amsterdam Zuidoost, NL

72 Inventor/es:

THESING, ROBIN y
SCHUG, MICHAEL

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 568 640 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

Procedimientos y sistemas para recuperar de manera eficiente contenido de audio de alta frecuencia

5 Campo técnico de la invención

El presente documento se refiere al campo técnico de la codificación, la decodificación y el procesamiento de audio. Específicamente, se refiere a procedimientos para recuperar contenido de alta frecuencia de una señal de audio a partir de contenido de baja frecuencia de la misma señal de audio de manera eficiente.

10

Antecedentes de la invención

Codificar y decodificar de manera eficiente señales de audio incluye normalmente reducir la cantidad de datos de audio que van a codificarse, transmitirse y/o descodificarse conforme a principios psicoacústicos. Esto incluye, por ejemplo, descartar el denominado contenido de audio enmascarado que está presente en una señal de audio, pero que un oyente no puede percibir. Como alternativa, o adicionalmente, puede limitarse el ancho de banda de una señal de audio que va a codificarse, mientras que solo se seguiría calculando, de manera respectiva, determinada información acerca de su contenido de frecuencia superior sin tener que codificar tal contenido de frecuencia superior directamente. Después, la señal de banda limitada se codifica y se transmite (o almacena) junto con dicha información de frecuencia superior, donde ésta última requiere menos recursos que si se codificara también el contenido de frecuencia superior directamente.

15

20

La Replicación de Banda Espectral (SBR) en HE-AAC (Codificación de Audio Avanza de Alta Eficiencia) y la Extensión Espectral (SPX) en Dolby Digital Plus son dos ejemplos de sistemas de codificación de audio que aproximan o reconstruyen una componente de alta frecuencia de una señal de audio basándose en una componente de baja frecuencia de la señal de audio y basándose en información complementaria adicional (también denominada información de frecuencia superior). En lo sucesivo se hace referencia al esquema SPX de Dolby Digital Plus. Sin embargo, debe observarse que los procedimientos y sistemas descritos en el presente documento pueden aplicarse, en general, a técnicas de reconstrucción de alta frecuencia, incluyendo SBR en HE-AAC.

25

30

La determinación de la información complementaria en un codificador de audio basado en SPX está sujeta normalmente a una gran complejidad computacional. A modo de ejemplo, la determinación de la información complementaria puede requerir aproximadamente el 50% de los recursos computacionales totales del codificador de audio. El presente documento describe procedimientos y sistemas que permiten reducir la complejidad computacional de los codificadores de audio basados en SPX. En particular, el presente documento describe procedimientos y sistemas que permiten reducir la complejidad computacional para llevar a cabo cálculos de tonalidad en el contexto de codificadores de audio basados en SPX (donde los cálculos de tonalidad pueden suponer aproximadamente el 80% de la complejidad computacional usada para determinar la información complementaria).

35

40

El documento US2010/0094638A1 describe un aparato y un procedimiento para decidir un nivel de ruido adaptativo para la extensión de ancho de banda.

Resumen de la invención

45

Según un aspecto se describe un procedimiento para determinar un primer valor de tonalidad por banda para una primera subbanda de frecuencia de una señal de audio. La señal de audio puede ser la señal de audio de un canal de una señal de audio multicanal (por ejemplo, una señal estéreo, una señal multicanal 5.1 o una señal multicanal 7.1). La señal de audio puede tener un ancho de banda comprendido entre una baja frecuencia de señal y una alta frecuencia de señal. El ancho de banda puede comprender una banda de baja frecuencia y una banda de alta frecuencia. La primera subbanda de frecuencia puede estar dentro de la banda de baja frecuencia o dentro de la banda de alta frecuencia. El primer valor de tonalidad por banda puede indicar una tonalidad de la señal de audio dentro de la primera banda de frecuencia. Puede considerarse que una señal de audio tiene una tonalidad relativamente alta dentro de una subbanda de frecuencia si la subbanda de frecuencia comprende un grado relativamente alto de contenido sinusoidal estable. Por otro lado, puede considerarse que una señal de audio tiene una tonalidad baja dentro de la subbanda de frecuencia si la subbanda de frecuencia comprende un grado de ruido relativamente alto. El primer valor de tonalidad por banda puede depender de la variación de la fase de la señal de audio dentro de la primera subbanda de frecuencia.

50

55

60

El procedimiento para determinar el primer valor de tonalidad por banda puede usarse en el contexto de un codificador de la señal de audio. El codificador puede usar técnicas de reconstrucción de alta frecuencia, tal como la Replicación de Banda Espectral (SBR) (que se usa, por ejemplo, en el contexto de un codificador de audio avanzado de alta eficiencia, HE-AAC) o la Extensión Espectral (SPX) (que se usa, por ejemplo, en el contexto de un codificador Dolby Digital Plus). El primer valor de tonalidad por banda puede usarse para aproximar una componente de alta frecuencia (en la banda de alta frecuencia) de la señal de audio en función de una componente de baja frecuencia (en la banda de baja frecuencia) de la señal de audio. En particular, el primer valor de tonalidad por

65

banda puede usarse para determinar información complementaria que puede usarse por un descodificador de audio correspondiente para reconstruir la componente de alta frecuencia de la señal de audio en función de la componente de baja frecuencia recibida (descodificada) de la señal de audio. Por ejemplo, la información complementaria puede especificar una cantidad de ruido a añadir a las subbandas de frecuencia trasladadas de la componente de baja frecuencia con el fin de aproximar una subbanda de frecuencia de la componente de alta frecuencia.

La invención se describe en las reivindicaciones independientes 1, 7 y 9.

El procedimiento puede comprender determinar un conjunto de coeficientes de transformada en un conjunto correspondiente de celdas de frecuencia (*frequency bins*) en función de un bloque de muestras de la señal de audio. La secuencia de muestras de la señal de audio puede agruparse en una secuencia de tramas, donde cada una comprende un número predeterminado de muestras. Una trama de la secuencia de tramas puede subdividirse en uno o más bloques de muestras. Los bloques adyacentes de una trama pueden solaparse (por ejemplo, hasta en un 50%). Un bloque de muestras puede transformarse del dominio de tiempo al dominio de frecuencia usando una transformada del dominio de tiempo al dominio de frecuencia, tal como una transformada discreta del coseno modificada (MDCT) y/o una transformada discreta del seno modificada (MDST), obteniéndose así el conjunto de coeficientes de transformada. Aplicando una MDST y una MDCT al bloque de muestras, puede proporcionarse un conjunto de coeficientes de transformada complejos. Normalmente, el número N de coeficientes de transformada (y el número N de celdas de frecuencia) corresponde al número N de muestras de un bloque (por ejemplo, $N=128$ o $N=256$). La primera subbanda de frecuencia puede comprender una pluralidad de las N celdas de frecuencia. Dicho de otro modo, las N celdas de frecuencia (que tienen una resolución de frecuencia relativamente alta) pueden agruparse en una o más subbandas de frecuencia (que tienen una resolución de frecuencia relativamente inferior). Como resultado, es posible proporcionar un número reducido de subbandas de frecuencia (lo que es normalmente beneficioso con respecto a velocidades de datos reducidas de la señal de audio codificada), donde las subbandas de frecuencia tienen una selección de frecuencias relativamente altas entre sí (debido a que las subbandas de frecuencia se obtienen mediante la agrupación de una pluralidad de celdas de frecuencia de alta resolución).

El procedimiento puede comprender además determinar un conjunto de valores de tonalidad de celda para el conjunto de celdas de frecuencia usando el conjunto de coeficientes de transformada, respectivamente. Los valores de tonalidad de celda se determinan normalmente para una celda de frecuencia individual (usando el coeficiente de transformada de esta celda de frecuencia individual). De este modo, un valor de tonalidad de celda indica la tonalidad de la señal de audio dentro de una celda de frecuencia individual. A modo de ejemplo, el valor de tonalidad de celda depende de la variación de la fase del coeficiente de transformada dentro de la celda de frecuencia individual correspondiente.

El procedimiento puede comprender además combinar un primer subconjunto de dos o más del conjunto de valores de tonalidad de celda para dos o más celdas de frecuencia adyacentes correspondientes del conjunto de celdas de frecuencia que están dentro de la primera subbanda de frecuencia, obteniéndose así el primer valor de tonalidad por banda para la primera subbanda de frecuencia. Dicho de otro modo, el primer valor de tonalidad por banda puede determinarse combinando dos o más valores de tonalidad de celda para las dos o más celdas de frecuencia que están dentro de la primera subbanda de frecuencia. La combinación del primer subconjunto de dos o más del conjunto de valores de tonalidad de celda puede comprender calcular el promedio de los dos o más valores de tonalidad de celda y/o sumar los dos o más valores de tonalidad de celda. A modo de ejemplo, el primer valor de tonalidad por banda puede determinarse conforme a la suma de los valores de tonalidad de celda de las celdas de frecuencia que están dentro de la primera subbanda de frecuencia.

De este modo, el procedimiento para determinar el primer valor de tonalidad por banda específica la determinación del primer valor de tonalidad por banda dentro de la primera subbanda de frecuencia (que comprende una pluralidad de celdas de frecuencia), en función de los valores de tonalidad de celda de las celdas de frecuencia que están dentro de las primeras subbandas de frecuencia. Dicho de otro modo, se propone determinar en dos etapas el primer valor de tonalidad por banda, donde la primera etapa proporciona un conjunto de valores de tonalidad de celda y donde la segunda etapa combina (al menos algunos de) el conjunto de valores de tonalidad de celda para obtener el primer valor de tonalidad por banda. Como resultado de tal enfoque de dos etapas, es posible determinar diferentes valores de tonalidad por banda (para diferentes estructuras de subbanda) conforme al mismo conjunto de valores de tonalidad de celda, reduciéndose así la complejidad computacional de un codificador de audio que utiliza los diferentes valores de tonalidad por banda.

En una forma de realización, el procedimiento comprende además determinar un segundo valor de tonalidad por banda en una segunda subbanda de frecuencia combinando un segundo subconjunto de dos o más del conjunto de valores de tonalidad de celda para dos o más celdas de frecuencia adyacentes correspondientes del conjunto de celdas de frecuencia que están dentro de la segunda subbanda de frecuencia. La primera y la segunda subbanda de frecuencia pueden comprender al menos una celda de frecuencia común, y el primer y el segundo subconjunto pueden comprender el al menos un valor de tonalidad de celda común correspondiente. Dicho de otro modo, el primer y el segundo valor de tonalidad por banda pueden determinarse en función de al menos un valor de tonalidad de celda común, lo que permite una menor complejidad computacional relacionada con la determinación de los valores de tonalidad por banda. A modo de ejemplo, la primera y la segunda subbanda de frecuencia pueden estar

dentro de la banda de alta frecuencia de la señal de audio. La primera subbanda de frecuencia puede ser más estrecha que la segunda subbanda de frecuencia y puede estar dentro de la segunda subbanda de frecuencia. El primer valor de tonalidad puede usarse en el contexto de la atenuación de varianza elevada de un codificador basado en SPX, y el segundo valor de tonalidad puede usarse en el contexto de la mezcla de ruido del codificador basado en SPX.

Como se ha indicado anteriormente, los procedimientos descritos en el presente documento se usan normalmente en el contexto de un codificador de audio que usa técnicas de reconstrucción de alta frecuencia (HFR). Tales técnicas HFR normalmente trasladan una o más celdas de frecuencia desde la banda de baja frecuencia de la señal de audio hasta una o más celdas de frecuencia de la banda de alta frecuencia, con el fin de aproximar la componente de alta frecuencia de la señal de audio. De este modo, aproximar la componente de alta frecuencia de la señal de audio en función de la componente de baja frecuencia de la señal de audio puede comprender copiar uno o más coeficientes de transformada de baja frecuencia de una o más celdas de frecuencia desde la banda de baja frecuencia, correspondiente a la componente de baja frecuencia, hasta la banda de alta frecuencia, correspondiente a la componente de alta frecuencia de la señal de audio. Este proceso de copia predeterminado puede tenerse en cuenta cuando se determinan valores de tonalidad por banda. En particular, puede tenerse en cuenta que los valores de tonalidad de celda no se ven afectados normalmente por el proceso de copia, permitiendo así que los valores de tonalidad de celda que se han determinado para una celda de frecuencia dentro de la banda de baja frecuencia se usen para celdas de frecuencia copiadas correspondientes dentro de la banda de alta frecuencia.

En una forma de realización, la primera subbanda de frecuencia está dentro de la banda de baja frecuencia, y la segunda subbanda de frecuencia está dentro de la banda de alta frecuencia. El procedimiento puede comprender además determinar el segundo valor de tonalidad por banda en la segunda subbanda de frecuencia combinando un segundo subconjunto de dos o más del conjunto de valores de tonalidad de celda para dos o más celdas de frecuencia correspondientes de las celdas de frecuencia que se han copiado en la segunda subbanda de frecuencia. Dicho de otro modo, el segundo valor de tonalidad por banda (para la segunda subbanda de frecuencia que está dentro de la banda de alta frecuencia) puede determinarse en función de los valores de tonalidad de celda de las celdas de frecuencia que se han copiado en la banda de alta frecuencia. La segunda subbanda de frecuencia puede comprender al menos una celda de frecuencia que se ha copiado desde una celda de frecuencia que está dentro de la primera banda de frecuencia. De este modo, el primer y el segundo subconjunto pueden comprender el al menos un valor de tonalidad de celda común correspondiente, reduciéndose así la complejidad computacional relacionada con la determinación de los valores de tonalidad por banda.

Como se ha indicado anteriormente, la señal de audio está agrupada normalmente en una secuencia de bloques (que comprenden, por ejemplo, N muestras cada uno). El procedimiento puede comprender determinar una secuencia de conjuntos de coeficientes de transformada en función de la secuencia correspondiente de bloques de la señal de audio. Como resultado, para cada celda de frecuencia, puede determinarse una secuencia de coeficientes de transformada. Dicho de otro modo, para una celda de frecuencia particular, la secuencia de conjuntos de coeficientes de transformada puede comprender una secuencia de coeficientes de transformada particulares. La secuencia de coeficientes de transformada particulares puede usarse para determinar una secuencia de valores de tonalidad de celda para la celda de frecuencia particular para la secuencia de bloques de la señal de audio.

Determinar el valor de tonalidad de celda para la celda de frecuencia particular puede comprender determinar una secuencia de fases en función de la secuencia de coeficientes de transformada particulares y determinar una aceleración de fase en función de la secuencia de fases. El valor de tonalidad de celda para la celda de frecuencia particular depende normalmente de la aceleración de fase. A modo de ejemplo, el valor de tonalidad de celda para un bloque actual de la señal de audio puede determinarse en función de una aceleración de fase actual. La aceleración de fase actual puede determinarse en función de la fase actual (determinarse en función del coeficiente de transformada del bloque actual) y en función de dos o más fases anteriores (determinarse en función de dos o más coeficientes de transformada de los dos o más bloques anteriores). Como se ha indicado anteriormente, un valor de tonalidad de celda para una celda de frecuencia particular se determina normalmente solo en función de los coeficientes de transformada de la misma celda de frecuencia particular. Dicho de otro modo, el valor de tonalidad de celda para una celda de frecuencia es normalmente independiente de los valores de tonalidad de celda de otras celdas de frecuencia.

Como se ha descrito anteriormente, el primer valor de tonalidad por banda puede usarse para aproximar una componente de alta frecuencia de la señal de audio en función de una componente de baja frecuencia de la señal de audio usando un esquema de Extensión Espectral (SPX). El primer valor de tonalidad por banda puede usarse para determinar una estrategia de reenvío de coordenadas SPX, un factor de mezcla de ruido y/o una atenuación de varianza elevada.

Según otro aspecto, se describe un procedimiento para determinar un factor de mezcla de ruido. Debe observarse que los diferentes aspectos y procedimientos descritos en el presente documento pueden combinarse entre sí de manera arbitraria. El factor de mezcla de ruido puede usarse para aproximar una componente de alta frecuencia de

la señal de audio en función de una componente de baja frecuencia de la señal de audio. Como se ha descrito anteriormente, la componente de alta frecuencia comprende normalmente componentes de la señal de audio en la banda de alta frecuencia. La banda de alta frecuencia puede subdividirse en una o más subbandas de alta frecuencia (por ejemplo, la primera y/o la segunda subbanda de frecuencia descritas anteriormente). La componente de la señal de audio dentro de una subbanda de alta frecuencia puede denominarse señal de subbanda de alta frecuencia. Asimismo, la componente de baja frecuencia comprende normalmente componentes de la señal de audio de la banda de baja frecuencia, y la banda de baja frecuencia puede subdividirse en una o más subbandas de baja frecuencia (por ejemplo, la primera y/o la segunda subbanda de frecuencia descritas anteriormente). La componente de la señal de audio dentro de una subbanda de baja frecuencia puede denominarse señal de subbanda de baja frecuencia. Dicho de otro modo, la componente de alta frecuencia puede comprender una o más señales de subbanda de alta frecuencia (originales) de la banda de alta frecuencia, y la componente de baja frecuencia puede comprender una o más señales de subbanda de baja frecuencia de la banda de baja frecuencia.

Como se ha descrito anteriormente, aproximar la componente de alta frecuencia puede comprender copiar una o más señales de subbanda de baja frecuencia en la banda de alta frecuencia, obteniéndose así una o más señales de subbanda de alta frecuencia aproximadas. El factor de mezcla de ruido puede usarse para indicar una cantidad de ruido que va a añadirse a la una o más señales de subbanda de alta frecuencia aproximadas con el fin de alinear la tonalidad de las señales de subbanda de alta frecuencia aproximadas con la tonalidad de la señal de subbanda de alta frecuencia original de la señal de audio. Dicho de otro modo, el factor de mezcla de ruido puede indicar una cantidad de ruido a añadir a la una o más señales de subbanda de alta frecuencia aproximadas con el fin de aproximar la componente de alta frecuencia (original) de la señal de audio.

El procedimiento puede comprender determinar un valor de tonalidad por banda objetivo en función de la una o más señales de subbanda de alta frecuencia (originales). Además, el procedimiento puede comprender determinar un valor de tonalidad por banda fuente en función de la una o más señales de subbanda de alta frecuencia aproximadas. Los valores de tonalidad pueden indicar la evolución de la fase de las respectivas señales de subbanda. Además, los valores de tonalidad pueden determinarse como se describe en el presente documento. En particular, los valores de tonalidad por banda pueden determinarse en función del enfoque de dos etapas descrito en el presente documento, es decir, los valores de tonalidad por banda pueden determinarse en función de un conjunto de valores de tonalidad de celda.

El procedimiento puede comprender además determinar el factor de mezcla de ruido en función de los valores de tonalidad por banda objetivo y fuente. En particular, el procedimiento puede comprender determinar el factor de mezcla de ruido en función del valor de tonalidad por banda fuente, si el ancho de banda de la componente de alta frecuencia a aproximar es menor que el ancho de banda de la componente de baja frecuencia que se usa para aproximar la componente de alta frecuencia. Como resultado, la complejidad computacional para determinar el factor de mezcla de ruido puede reducirse en comparación con un procedimiento en el que el factor de mezcla de ruido se determina en función de un valor de tonalidad por banda que se obtiene a partir de la componente de baja frecuencia de la señal de audio.

En una forma de realización, la banda de baja frecuencia comprende una banda de inicio (indicada, por ejemplo, mediante el parámetro *inicio_spx* en el caso de un codificador basado en SPX) que indica la subbanda de baja frecuencia que tiene la frecuencia más baja de las subbandas de baja frecuencia que están disponibles para la copia. Además, la banda de alta frecuencia puede comprender una banda de comienzo (indicada, por ejemplo, mediante el parámetro *comienzo_spx* en el caso de un codificador basado en SPX) que indica la subbanda de alta frecuencia que tiene la frecuencia más baja de las subbandas de alta frecuencia que van a aproximarse. Además, la banda de alta frecuencia puede comprender una banda de finalización (indicada, por ejemplo, mediante el parámetro *finalización_spx* en el caso de un codificador basado en SPX) que indica la subbanda de alta frecuencia que tiene la frecuencia más alta de las subbandas de alta frecuencia que van a aproximarse.

El procedimiento puede comprender determinar un primer ancho de banda entre la banda de inicio (por ejemplo, el parámetro *inicio_spx*) y la banda de comienzo (por ejemplo, el parámetro *comienzo_spx*). Además, el procedimiento puede comprender determinar un segundo ancho de banda entre la banda de comienzo (por ejemplo, el parámetro *comienzo_spx*) y la banda de finalización (por ejemplo, el parámetro *finalización_spx*). El procedimiento puede comprender determinar el factor de mezcla de ruido en función de los valores de tonalidad por banda objetivo y fuente, si el primer ancho de banda es mayor que el segundo ancho de banda. En particular, si el primer ancho de banda es mayor o igual que el segundo ancho de banda, el valor de tonalidad por banda fuente puede determinarse en función de la una o más señales de subbanda de baja frecuencia de la subbanda de baja frecuencia que está entre la banda de inicio y la banda de inicio más el segundo ancho de banda. Normalmente, las últimas señales de subbanda de baja frecuencia son las señales de subbanda de baja frecuencia que se copian en la banda de alta frecuencia. Como resultado, la complejidad computacional puede reducirse en situaciones en las que el primer ancho de banda es mayor o igual que el segundo ancho de banda.

Por otro lado, el procedimiento puede comprender determinar un valor bajo de tonalidad por banda en función de la una o más señales de subbanda de baja frecuencia de la subbanda de baja frecuencia entre la banda de inicio y la banda de comienzo, y determinar el factor de mezcla de ruido en función del valor de tonalidad por banda objetivo y

el valor de tonalidad por banda bajo, si el primer ancho de banda es más pequeño que el segundo ancho de banda. Comparando el primer y el segundo ancho de banda, puede garantizarse que el factor de mezcla de ruido (y los valores de tonalidad por banda) se determinan en un número mínimo de subbandas (independientemente del primer y del segundo ancho de banda), reduciéndose así la complejidad computacional.

5 El factor de mezcla de ruido puede determinarse en función de la varianza del valor de tonalidad por banda objetivo y del valor de tonalidad por banda fuente (o del valor de tonalidad por banda objetivo y el valor de tonalidad por banda bajo). En particular, el factor de mezcla de ruido b puede determinarse como

$$10 \quad b = T_{copia} \cdot (1 - \text{var}\{T_{copia}, T_{alta}\}) + T_{alta} \cdot (\text{var}\{T_{copia}, T_{alta}\}),$$

donde $\text{var}\{T_{copia}, T_{alta}\} = \left(\frac{T_{copia} - T_{alta}}{T_{copia} + T_{alta}} \right)^2$ es la varianza del valor de tonalidad fuente T_{copia} (o del valor de tonalidad bajo) y el valor de tonalidad objetivo T_{alta} .

15 Como se ha indicado anteriormente, los valores de tonalidad por banda (fuente, objetivo o bajo) pueden determinarse usando el enfoque de dos etapas descrito en el presente documento. En particular, un valor de tonalidad por banda de una subbanda de frecuencia puede determinarse determinando un conjunto de coeficientes de transformada en un conjunto correspondiente de celdas de frecuencia en función de un bloque de muestras de la señal de audio. Después puede determinarse un conjunto de valores de tonalidad de celda para el conjunto de celdas de frecuencia usando el conjunto de coeficientes de transformada, respectivamente. El valor de tonalidad por banda de la subbanda de frecuencia puede determinarse después combinando un primer subconjunto de dos o más del conjunto de valores de tonalidad de celda para dos o más celdas de frecuencia adyacentes correspondientes del conjunto de celdas de frecuencia que están dentro de la subbanda de frecuencia.

25 Según un aspecto adicional se describe un procedimiento para determinar un primer valor de tonalidad de celda para una primera celda de frecuencia de una señal de audio. El primer valor de tonalidad de celda puede determinarse según los principios descritos en el presente documento. En particular, el primer valor de tonalidad de celda puede determinarse en función de la variación de la fase del coeficiente de transformada de la primera celda de frecuencia. Además, como también se describe en el presente documento, el primer valor de tonalidad de celda puede usarse para aproximar una componente de alta frecuencia de la señal de audio en función de una componente de baja frecuencia de la señal de audio. De este modo, el procedimiento para determinar un primer valor de tonalidad de celda puede usarse en el contexto de un codificador de audio que usa técnicas HFR.

35 El procedimiento puede comprender proporcionar una secuencia de coeficientes de transformada en la primera celda de frecuencia para una secuencia correspondiente de bloques de muestras de la señal de audio. La secuencia de coeficientes de transformada puede determinarse aplicando a la secuencia de bloques de muestras una transformada del dominio de tiempo al dominio de frecuencia (como se ha descrito anteriormente). Además, el procedimiento puede comprender determinar una secuencia de fases en función de la secuencia de coeficientes de transformada. El coeficiente de transformada puede ser complejo y una fase de un coeficiente de transformada puede determinarse según una función de arcotangente aplicada a la parte real e imaginaria del coeficiente de transformada complejo. Además, el procedimiento puede comprender determinar una aceleración de fase en función de la secuencia de fases. A modo de ejemplo, la aceleración de fase actual para un coeficiente de transformada actual para un bloque de muestras actual puede determinarse en función de la fase actual y en función de dos o más fases anteriores. Además, el procedimiento puede comprender determinar una potencia de celda en función del coeficiente de transformada actual a partir de la secuencia de coeficientes de transformada. La potencia del coeficiente de transformada actual puede basarse en una magnitud elevada al cuadrado del coeficiente de transformada actual.

50 El procedimiento puede comprender además aproximar un factor de ponderación que indica la raíz cuarta de una relación de la potencia de coeficientes de transformada subsiguientes usando una aproximación logarítmica. Después, el procedimiento puede ponderar la aceleración de fase mediante el factor de ponderación aproximado y/o mediante la potencia del coeficiente de transformada actual para obtener el primer valor de tonalidad de celda. Como resultado de aproximar el factor de ponderación usando una aproximación logarítmica, puede conseguirse una aproximación de alta calidad del factor de ponderación correcto, mientras que al mismo tiempo se reduce considerablemente la complejidad computacional en comparación con la determinación del factor de ponderación exacto que implica la determinación de una raíz cuarta de la relación de la potencia de coeficientes de transformada subsiguientes. La aproximación logarítmica puede comprender la aproximación de una función logarítmica mediante una función lineal y/o mediante un polinomio (por ejemplo, de orden 1, 2, 3, 4 o 5).

60 La secuencia de coeficientes de transformada puede comprender un coeficiente de transformada actual (para un bloque de muestras actual) y un coeficiente de transformada directamente anterior (para un bloque de muestras directamente anterior). El factor de ponderación puede indicar la raíz cuarta de la relación de la potencia del

- coeficiente de transformada actual y del coeficiente de transformada directamente anterior. Además, como se ha indicado anteriormente, los coeficientes de transformada pueden ser números complejos que comprenden una parte real y una parte imaginaria. La potencia del coeficiente de transformada actual (anterior) puede determinarse en función de la parte real elevada al cuadrado y la parte imaginaria elevada al cuadrado del coeficiente de transformada actual (anterior). Además, una fase actual (anterior) puede determinarse según una función de arcotangente de la parte real y de la parte imaginaria del coeficiente de transformada actual (anterior). Una aceleración de fase actual puede determinarse en función de la fase del coeficiente de transformada actual y en función de las fases de dos o más coeficientes de transformada directamente anteriores.
- 10 Aproximar el factor de ponderación puede comprender proporcionar la mantisa actual y el exponente actual que representan un coeficiente actual de la secuencia de coeficientes de transformada subsiguientes. Además, aproximar el factor de ponderación puede comprender determinar un valor de índice para una tabla de consulta predeterminada en función de la mantisa actual y el exponente actual. La tabla de consulta proporciona normalmente una relación entre una pluralidad de valores de índice y una pluralidad correspondiente de valores exponenciales de la pluralidad de valores de índice. De este modo, la tabla de consulta puede ser un modo eficaz de aproximar una función exponencial. En una forma de realización, la tabla de consulta comprende 64 entradas (es decir, pares de valores de índice y de valores exponenciales) o un número inferior de entradas. El factor de ponderación aproximado puede determinarse usando el valor de índice y la tabla de consulta.
- 15
- 20 En particular, el procedimiento puede comprender determinar un valor de índice de valor real en función de la mantisa y el exponente. Después puede determinarse un valor de índice (de valor entero) truncando y/o redondeando el valor de índice de valor real. Como resultado de una operación sistemática de truncado o redondeo, puede introducirse una imprecisión sistemática en la aproximación. Tal imprecisión sistemática puede ser beneficiosa con respecto a la calidad percibida de una señal de audio que se codifica usando el procedimiento para determinar el valor de tonalidad de celda descrito en el presente documento.
- 25
- Aproximar el factor de ponderación puede comprender además proporcionar una mantisa anterior y un exponente anterior que representa un coeficiente de transformada anterior al coeficiente de transformada actual. El valor de índice puede determinarse después en función de una o más operaciones de suma y/o resta aplicadas a la mantisa actual, la mantisa anterior, el exponente actual y el exponente anterior. En particular, el valor de índice puede determinarse llevando a cabo una operación de módulo en $(e_y - e_z + 2 \cdot m_y - 2 \cdot m_z)$, donde e_y es la mantisa actual, e_z es la mantisa anterior, m_y es el exponente actual y m_z es el exponente anterior.
- 30
- Como se ha indicado anteriormente, los procedimientos descritos en el presente documento pueden aplicarse a señales de audio multicanal. En particular, los procedimientos pueden aplicarse a un canal de una señal de audio multicanal. Los codificadores de audio para señales de audio multicanal aplican normalmente una técnica de codificación denominada acoplamiento de canal (o, para abreviar, acoplamiento), con el fin de codificar conjuntamente una pluralidad de canales de la señal de audio multicanal. Por este motivo, según un aspecto, se describe un procedimiento para determinar una pluralidad de valores de tonalidad para una pluralidad de canales acoplados de una señal de audio multicanal.
- 35
- 40
- El procedimiento puede comprender determinar una primera secuencia de coeficientes de transformada para una secuencia correspondiente de bloques de muestras de un primer canal de la pluralidad de canales acoplados. Como alternativa, la primera secuencia de coeficientes de transformada puede determinarse en función de una secuencia de bloques de muestras del canal de acoplamiento obtenido a partir de la pluralidad de canales acoplados. Después, el procedimiento puede determinar un primer valor de tonalidad para el primer canal (o para el canal de acoplamiento). Por este motivo, el procedimiento puede comprender determinar una primera secuencia de fases en función de la secuencia de primeros coeficientes de transformada, y determinar una primera aceleración de fase en función de la secuencia de primeras fases. El primer valor de tonalidad para el primer canal (o para el canal de acoplamiento) puede determinarse después en función de la primera aceleración de fase. Además, el valor de tonalidad para un segundo canal de la pluralidad de canales acoplados puede determinarse en función de la primera aceleración de fase. De este modo, los valores de tonalidad para la pluralidad de canales acoplados pueden determinarse en función de la aceleración de fase determinada a partir de un solo canal de los canales acoplados, reduciéndose así la complejidad computacional asociada a la determinación de la tonalidad. Esto es posible debido a que, como resultado del acoplamiento, las fases de la pluralidad de canales acoplados están alineadas.
- 45
- 50
- 55
- Según otro aspecto, se describe un procedimiento para determinar un valor de tonalidad por banda para un primer canal de una señal de audio multicanal en un codificador basado en Extensión Espectral (SPX). El codificador basado en SPX puede configurarse para aproximar una componente de alta frecuencia del primer canal a partir de una componente de baja frecuencia del primer canal. Con este fin, el codificador basado en SPX puede usar el valor de tonalidad por banda. En particular, el codificador basado en SPX puede usar el valor de tonalidad por banda para determinar un factor de mezcla de ruido que indica una cantidad de ruido a añadir a la componente de alta frecuencia aproximada. De este modo, el valor de tonalidad por banda puede indicar la tonalidad de una componente de alta frecuencia aproximada antes de la mezcla de ruido. El primer canal puede acoplarse mediante el codificador basado en SPX a uno o más canales de la señal de audio multicanal.
- 60
- 65

El procedimiento puede comprender proporcionar una pluralidad de coeficientes de transformada en función del primer canal antes del acoplamiento. Además, el procedimiento puede comprender determinar el valor de tonalidad por banda en función de la pluralidad de coeficientes de transformada. De este modo, el factor de mezcla de ruido puede determinarse en función de la pluralidad de coeficientes de transformada del primer canal original y no en función del primer canal acoplado/desacoplado. Esto es beneficioso ya que permite reducir la complejidad computacional asociada a la determinación de tonalidad en un codificador de audio basado en SPX.

Como se ha descrito anteriormente, la pluralidad de coeficientes de transformada que se han determinado en función del primer canal antes del acoplamiento (es decir, en función del primer canal original) puede usarse para determinar los valores de tonalidad de celda y/o los valores de tonalidad por banda que se usan para determinar la estrategia de reenvío de coordenadas SPX y/o para determinar la atenuación de varianza elevada (LVA) de un codificador basado en SPX. Usando el enfoque mencionado anteriormente para determinar el factor de mezcla de ruido del primer canal en función del primer canal original (y no en función del primer canal acoplado/desacoplado), pueden volver a utilizarse los valores de tonalidad de celda que ya se han determinado para la estrategia de reenvío de coordenadas SPX y/o para la atenuación de varianza elevada (LVA), reduciéndose así la complejidad computacional del codificador basado en SPX.

Según otro aspecto se describe un sistema configurado para determinar un primer valor de tonalidad por banda para una primera subbanda de frecuencia de una señal de audio. El primer valor de tonalidad por banda puede usarse para aproximar una componente de alta frecuencia de la señal de audio en función de una componente de baja frecuencia de la señal de audio. El sistema puede estar configurado para determinar un conjunto de coeficientes de transformada en un conjunto correspondiente de celdas de frecuencia en función de un bloque de muestras de la señal de audio. Además, el sistema puede estar configurado para determinar un conjunto de valores de tonalidad de celda para el conjunto de celdas de frecuencia usando el conjunto de coeficientes de transformada, respectivamente. Además, el sistema puede estar configurado para combinar un primer subconjunto de dos o más del conjunto de valores de tonalidad de celda para dos o más celdas de frecuencia adyacentes correspondientes del conjunto de celdas de frecuencia que están dentro de la primera subbanda de frecuencia, obteniéndose así el primer valor de tonalidad por banda para la primera subbanda de frecuencia.

Según otro aspecto, se describe un sistema configurado para determinar un factor de mezcla de ruido. El factor de mezcla de ruido puede usarse para aproximar una componente de alta frecuencia de la señal de audio en función de una componente de baja frecuencia de la señal de audio. La componente de alta frecuencia comprende normalmente una o más señales de subbanda de alta frecuencia de una banda de alta frecuencia, y la componente de baja frecuencia comprende normalmente una o más señales de subbanda de baja frecuencia de una banda de baja frecuencia. Aproximar la componente de alta frecuencia puede comprender copiar una o más señales de subbanda de baja frecuencia en la banda de alta frecuencia, obteniéndose así una o más señales de subbanda de alta frecuencia aproximadas. El sistema puede estar configurado para determinar un valor de tonalidad por banda objetivo en función de la una o más señales de subbanda de alta frecuencia. Además, el sistema puede estar configurado para determinar un valor de tonalidad por banda fuente en función de la una o más señales de subbanda de alta frecuencia aproximadas. Además, el sistema puede estar configurado para determinar el factor de mezcla de ruido en función de los valores de tonalidad por banda objetivo (322) y fuente (323).

Según un aspecto adicional, se describe un sistema configurado para determinar un primer valor de tonalidad de celda para una primera celda de frecuencia de una señal de audio. El primer valor de tonalidad por banda puede usarse para aproximar una componente de alta frecuencia de la señal de audio en función de una componente de baja frecuencia de la señal de audio. El sistema puede estar configurado para proporcionar una secuencia de coeficientes de transformada en la primera celda de frecuencia para una secuencia correspondiente de bloques de muestras de la señal de audio. Además, el sistema puede estar configurado para determinar una secuencia de fases en función de la secuencia de coeficientes de transformada, y para determinar una aceleración de fase en función de la secuencia de fases. Además, el sistema puede estar configurado para aproximar un factor de ponderación que indica la raíz cuarta de una relación de una potencia de coeficientes de transformada subsiguientes usando una aproximación logarítmica, y para ponderar la aceleración de fase mediante el factor de ponderación aproximado para obtener el primer valor de tonalidad de celda.

Según otro aspecto, se describe un codificador de audio (por ejemplo, un codificador de audio basado en HFR, en particular un codificador de audio basado en SPX) configurado para codificar una señal de audio que usa reconstrucción de alta frecuencia. El codificador de audio puede comprender uno cualquiera o más de los sistemas descritos en el presente documento. Además, o como alternativa, el codificador de audio puede estar configurado para llevar a cabo uno cualquiera o más de los procedimientos descritos en el presente documento.

Según un aspecto adicional, se describe un programa de software. El programa de software puede estar adaptado para ejecutarse en un procesador y para llevar a cabo las etapas de procedimiento descritas en el presente documento cuando se ejecuta en el procesador.

Según otro aspecto, se describe un medio de almacenamiento. El medio de almacenamiento puede comprender un programa de software adaptado para ejecutarse en un procesador y para llevar a cabo las etapas de procedimiento descritas en el presente documento cuando se ejecuta en el procesador.

- 5 Según un aspecto adicional, se describe un producto de programa informático. El programa informático puede comprender instrucciones ejecutables para llevar a cabo las etapas de procedimiento descritas en el presente documento cuando se ejecutan en un ordenador.

10 Debe observarse que los procedimientos y los sistemas, incluidas sus realizaciones preferidas, descritos en la presente solicitud de patente pueden usarse de manera independiente o en combinación con otros procedimientos y sistemas dados a conocer en este documento. Además, todos los aspectos de los procedimientos y sistemas descritos en la presente solicitud de patente pueden combinarse de manera arbitraria. En particular, las características de las reivindicaciones pueden combinarse entre sí de manera arbitraria.

15 Breve descripción de las figuras

La invención se describirá a continuación, a modo de ejemplo, con referencia a los dibujos adjuntos, en los que:

- 20 las Fig. 1a, 1b, 1c y 1d ilustran un esquema SPX de ejemplo;
 las Fig. 2a, 2b, 2c y 2d ilustran el uso de la tonalidad en varias fases de un codificador basado en SPX;
 las Fig. 3a, 3b, 3c y 3d ilustran esquemas de ejemplo para reducir el esfuerzo computacional asociado al cálculo de los valores de tonalidad;
 la Fig. 4 ilustra resultados de ejemplo de una prueba de escucha que compara la determinación de tonalidad basada en la señal de audio original y la determinación de tonalidad basada en la señal de audio desacomplada;
 25 la Fig. 5a ilustra resultados de ejemplo de una prueba de escucha que compara varios esquemas para determinar el factor de ponderación usado para el cálculo de valores de tonalidad; y
 la Fig. 5b ilustra grados de aproximación de ejemplo del factor de ponderación usado para el cálculo de valores de tonalidad.

30 Descripción detallada de la invención

Las Fig. 1a, 1b, 1c y 1d ilustran etapas de ejemplo llevadas a cabo por un codificador de audio basado en SPX. La Fig. 1a muestra el espectro de frecuencia 100 de una señal de audio de ejemplo, donde el espectro de frecuencia 100 comprende una banda base 101 (también denominada banda de baja frecuencia 101) y una banda de alta frecuencia 102. En el ejemplo ilustrado, la banda de alta frecuencia 102 comprende una pluralidad de subbandas, es decir, de la banda SE 1 a la banda SE 5 (SE, Extensión Espectral). La banda base 101 comprende las frecuencias inferiores hasta la frecuencia de corte de banda base 103, y la banda de alta frecuencia 102 comprende las altas frecuencias desde la frecuencia de corte de banda base 103 hasta la frecuencia de ancho de banda de audio 104. La banda base 101 corresponde al espectro de una componente de baja frecuencia de la señal de audio y la banda de alta frecuencia 102 corresponde al espectro de una componente de alta frecuencia de la señal de audio. Dicho de otro modo, la componente de baja frecuencia de la señal de audio comprende las frecuencias de la banda base 101, donde la componente de alta frecuencia de la señal de audio comprende las frecuencias de la banda de alta frecuencia 102.

45 Un codificador de audio usa normalmente una transformada del dominio de tiempo al dominio de frecuencia (por ejemplo, una transformada discreta del coseno modificada, MDCT, y/o una transformada discreta del seno modificada, MDST) con el fin de determinar el espectro 100 a partir de la señal de audio de dominio de tiempo. Una señal de audio de dominio de tiempo puede subdividirse en una secuencia de tramas de audio que comprende secuencias respectivas de muestras de la señal de audio. Cada trama de audio puede subdividirse en una pluralidad de bloques (por ejemplo, una pluralidad de hasta seis bloques), donde cada bloque comprende, por ejemplo, N o $2N$ muestras de la señal de audio. Los diversos bloques de una trama pueden solaparse (por ejemplo, con un solapamiento del 50%), es decir, un segundo bloque puede comprender un determinado número de muestras en su parte inicial que son idénticas a las muestras situadas en la parte final de un primer bloque directamente anterior. A modo de ejemplo, un segundo bloque de $2N$ muestras puede comprender una sección central de N muestras, y secciones trasera/delantera de $N/2$ muestras que se solapan con la sección central del primer bloque directamente anterior y de un tercer bloque directamente posterior, respectivamente. La transformada del dominio de tiempo al dominio de frecuencia de un bloque de N (o $2N$) muestras de la señal de audio de dominio de tiempo proporciona normalmente un conjunto de N coeficientes de transformada (TC) para un conjunto correspondiente de celdas de frecuencia (por ejemplo, $N=256$). A modo de ejemplo, la transformada del dominio de tiempo al dominio de frecuencia (por ejemplo, una MDCT o una MDST) de un bloque de $2N$ muestras, que tiene una sección central de N muestras y secciones trasera/delantera solapadas de $N/2$ muestras, puede proporcionar un conjunto de N TC. De este modo, un solapamiento del 50% puede dar como resultado, por término medio, una relación 1:1 de muestras de dominio de tiempo y de TC, obteniéndose así un sistema sumamente muestreado. Las subbandas de la banda de alta frecuencia 102 mostradas en la Fig. 1a pueden obtenerse agrupando M celdas de frecuencia para formar una subbanda (por ejemplo, $M=12$). Dicho de otro modo, una subbanda de la banda de alta frecuencia 102 puede

comprender o abarcar M celdas de frecuencia. La energía espectral de una subbanda puede determinarse en función de los TC de las M celdas de frecuencia que forman la subbanda. A modo de ejemplo, la energía espectral de la subbanda puede determinarse en función de la suma de la magnitud elevada al cuadrado de los TC de las M celdas de frecuencia que forman la subbanda (por ejemplo, en función del promedio de la magnitud elevada al cuadrado de los TC de las M celdas de frecuencia que forman la subbanda). En particular, la suma de la magnitud elevada al cuadrado de los TC de las M bandas de frecuencia que forman la subbanda puede dar como resultado la potencia de subbanda, y la potencia de subbanda dividida por el número M de celdas de frecuencia puede dar como resultado la densidad espectral de potencia (PSD). De este modo, la banda base 101 y/o la banda de alta frecuencia 102 puede comprender una pluralidad de subbandas, donde las subbandas se obtienen a partir de una pluralidad de celdas de frecuencia, respectivamente.

Como se ha indicado anteriormente, un codificador basado en SPX aproxima la banda de alta frecuencia 102 de una señal de audio mediante la banda base 101 de la señal de audio. Con este fin, el codificador basado en SPX determina información complementaria que permite a un descodificador correspondiente reconstruir la banda de alta frecuencia 102 a partir de la banda base codificada y descodificada 101 de la señal de audio. La información complementaria comprende normalmente indicadores de la energía espectral de la una o más subbandas de la banda de alta frecuencia 102 (por ejemplo, una o más relaciones de energía para la una o más subbandas de la banda de alta frecuencia 102, respectivamente). Además, la información complementaria comprende normalmente indicadores de una cantidad de ruido que va a añadirse a la una o más subbandas de la banda de alta frecuencia 102 (lo que se denomina mezcla de ruido). Estos indicadores están normalmente relacionados con la tonalidad de la una o más subbandas de la banda de alta frecuencia 102. Dicho de otro modo, los indicadores de una cantidad de ruido que va a añadirse a la una o más subbandas de la banda de alta frecuencia 102 usan normalmente el cálculo de valores de tonalidad de la una o más subbandas de la banda de alta frecuencia 102.

Las Fig. 1b, 1c y 1d ilustran las etapas de ejemplo para aproximar la banda de alta frecuencia 102 en función de la banda base 101. La Fig. 1b muestra el espectro 110 de la componente de baja frecuencia de la señal de audio que comprende solamente la banda base 101. La Fig. 1c ilustra la traslación espectral de una o más subbandas 121, 122 de la banda base 101 a las frecuencias de la banda de alta frecuencia 102. A partir del espectro 120 puede observarse que las subbandas 121, 122 se copian en las bandas de frecuencia respectivas 123, 124, 125, 126, 127 y 128 de la banda de alta frecuencia 102. En el ejemplo ilustrado, las subbandas 121, 122 se copian tres veces con el fin de llenar la banda de alta frecuencia 102. La Fig. 1d muestra cómo la banda de alta frecuencia original 102 de la señal de audio (véase la Fig. 1a) se aproxima en función de las subbandas copias (o trasladadas) 123, 124, 125, 126, 127 y 128. El codificador de audio basado en SPX puede añadir ruido aleatorio a las subbandas copiadas, de manera que la tonalidad de las subbandas aproximadas 133, 134, 135, 136, 137 y 138 corresponde a la tonalidad de las subbandas originales de la banda de alta frecuencia 102. Esto puede conseguirse determinando indicadores de tonalidad respectivos apropiados. Además, la energía de las subbandas copiadas (y con mezcla de ruido) 123, 124, 125, 126, 127 y 128 puede modificarse de modo que la energía de las subbandas aproximadas 133, 134, 135, 136, 137 y 138 corresponda a la energía de las subbandas originales de la banda de alta frecuencia 102. Esto puede conseguirse determinando indicadores de energía respectivos apropiados. Puede observarse que, como resultado, el espectro 130 aproxima el espectro 100 de la señal de audio original mostrada en la Fig. 1a.

Como se ha indicado anteriormente, la determinación de los indicadores que se usan para la mezcla de ruido (y que normalmente requieren la determinación de la tonalidad de las subbandas) tiene un gran impacto en la complejidad computacional del codificador de audio basado en SPX. En particular, los valores de tonalidad de diferentes segmentos de señal (subbandas de frecuencia) pueden necesitarse para varios fines en diferentes fases del proceso de codificación SPX. Una visión general de las fases que requieren normalmente la determinación de los valores de tonalidad se muestra en las Fig. 2a, 2b, 2c y 2d.

En las Fig. 2a, 2b, 2c y 2d, la frecuencia (en forma de subbandas SPX 0 a 16) se muestra en el eje horizontal con marcadores para la banda de inicio SPX (o frecuencia de inicio SPX) 201 (denominada inicio_spx), la banda de comienzo SPX (o frecuencia de comienzo SPX) 202 (denominada comienzo_spx) y la banda de finalización SPX (o frecuencia de finalización SPX) 203 (denominada finalización_spx). Normalmente, la frecuencia de comienzo SPX 202 corresponde a la frecuencia de corte 103. La frecuencia de finalización SPX 203 puede corresponder al ancho de banda 104 de la señal de audio original o a una frecuencia inferior al ancho de banda de audio 104 (como se ilustra en las Fig. 2a, 2b, 2c y 2d). Tras la codificación, el ancho de banda de la señal de audio codificada/descodificada corresponde normalmente a la frecuencia de finalización SPX 203. En una forma de realización, la frecuencia de inicio SPX 201 corresponde a la celda de frecuencia n.º 25, y la frecuencia de finalización SPX 203 corresponde a la celda de frecuencia n.º 229. Las subbandas de la señal de audio se muestran en tres fases diferentes del proceso de codificación SPX: el espectro 200 (por ejemplo, el espectro MDCT) de la señal de audio original (Fig. 2a, parte superior, y Fig. 2b) y el espectro 210 de la señal de audio tras la codificación/descodificación de la componente de baja frecuencia de la señal de audio (Fig. 2a, parte central, y Fig. 2c). La codificación/descodificación de la componente de baja frecuencia de la señal de audio puede comprender, por ejemplo, la matrización y desmatrización y/o el acoplamiento y desacoplamiento de la componente de baja frecuencia. Además, se muestra el espectro 220 tras la traslación espectral de las subbandas de la banda base 101 a la banda de alta frecuencia 102 (Fig. 2a, parte inferior, y Fig. 2d). El espectro 200 de las partes originales de la señal de audio se muestra en la línea "Original" de la Fig. 2a (es decir, las subbandas de frecuencia 0 a 16); el

espectro 210 de las partes de la señal que se modifican mediante acoplamiento/matrización se muestra en la línea "Banda baja desmatrizada/desacoplada" de la Fig. 2a (es decir, las subbandas de frecuencia 2 a 6 en el ejemplo ilustrado); y el espectro 220 de las partes de la señal modificadas por la traslación espectral se muestra en la línea "banda alta trasladada" de la Fig. 2a (es decir, las subbandas de frecuencia 7 a 14 en el ejemplo ilustrado). Las subbandas 206 que están modificadas por el procesamiento del codificador basado en SPX se ilustran con un sombreado oscuro, mientras que las subbandas 205 que no son modificadas por el codificador basado en SPX se ilustran con un sombreado claro.

Las llaves 231, 232, 233 situadas debajo de las subbandas y/o debajo de los grupos de subbandas SPX indican para qué subbandas o para qué grupos de subbandas se calculan los valores de tonalidad (medidas de tonalidad). Además, se indica la finalidad con que se usan los valores de tonalidad o las medidas de tonalidad. Los valores de tonalidad por banda 231 (es decir, los valores de tonalidad para una subbanda o para un grupo de subbandas) de la señal de entrada original entre la banda de inicio SPX (inicio_spx) 201 y la banda de finalización SPX (finalización_spx) 203 se usan normalmente para ayudar a que el codificador decida si es necesario transmitir o no nuevas coordenadas SPX ("estrategia de reenvío"). Las coordenadas SPX normalmente transportan información acerca de la envolvente espectral de la señal de audio original en forma de factores de ganancia para cada banda SPX. La estrategia de reenvío SPX puede indicar si es necesario transmitir nuevas coordenadas SPX para un nuevo bloque de muestras de la señal de audio o si pueden reutilizarse las coordenadas SPX para un bloque de muestras (directamente) anterior. Además, los valores de tonalidad por banda 231 para las bandas SPX por encima de comienzo_spx 202 pueden usarse como datos de entrada en los cálculos de la atenuación de varianza elevada (LVA), como se ilustra en la Fig. 2a y en la Fig. 2b. La atenuación de varianza elevada es una herramienta de codificación que puede usarse para atenuar posibles errores debidos a la traslación espectral. Las componentes espectrales intensas en la banda de extensión que no tienen una componente correspondiente en la banda base (y viceversa) pueden considerarse errores de extensión. El mecanismo LVA puede usarse para atenuar tales errores de extensión. Como puede observarse mediante las llaves de la Fig. 2b, los valores de tonalidad 231 pueden calcularse para subbandas individuales (por ejemplo, subbandas 0, 1, 2, etc.) y/o para grupos de subbandas (por ejemplo, para el grupo que comprende las subbandas 11 y 12).

Como se ha indicado anteriormente, la tonalidad de señal juega un papel importante para determinar la cantidad de mezcla de ruido aplicada a las subbandas reconstruidas en la banda de alta frecuencia 102. Como se ilustra en la Fig. 2c, los valores de tonalidad 232 se calculan por separado para la banda baja descodificada (por ejemplo, desmatrizada y desacoplada) y para la banda alta original. En este contexto, el término 'descodificación' (por ejemplo, desmatrización y desacoplamiento) significa que las etapas de codificación aplicadas previamente (por ejemplo, las etapas de matrización y de acoplamiento) del codificador se deshacen de la misma manera a como se realizaría en el descodificador. Dicho de otro modo, tal mecanismo de descodificación ya está simulado en el codificador. Por tanto, la banda baja que comprende las subbandas 0 a 6 del espectro 210 es una simulación del espectro que será recreada por el descodificador. La Fig. 2c muestra además que la tonalidad se calcula (solamente) para dos grandes bandas en este caso, a diferencia de la tonalidad de la señal original que se calcula por cada subbanda SPX (que abarca una pluralidad de 12 coeficientes de transformada (TC)) o por cada grupo de subbandas SPX. Como se indica mediante las llaves de la Fig. 2c, los valores de tonalidad 232 se calculan para un grupo de subbandas de la banda base 101 (que comprende, por ejemplo, las subbandas 0 a 6) y para un grupo de subbandas de la banda de alta frecuencia 102 (que comprende, por ejemplo, las subbandas 7 a 14).

Además de lo anterior, los cálculos de la atenuación de varianza elevada (LVA) requieren normalmente otra entrada de tonalidad que se calcula tomando los coeficientes de transformada (TC) trasladados. La tonalidad se mide para la misma región espectral que la de la Fig. 2a, pero tomando datos diferentes, es decir, en relación con las subbandas de banda baja trasladadas y no con las subbandas originales. Esto se ilustra en el espectro 220 mostrado en la Fig. 2d. Puede observarse que los valores de tonalidad 233 se determinan para subbandas y/o grupos de subbandas de la banda de alta frecuencia 102 en función de las subbandas trasladadas.

En términos generales, puede observarse que un codificador típico basado en SPX determina valores de tonalidad 231, 232, 233 en relación con varias subbandas 205, 206 y/o grupos de subbandas de la señal de audio original y/o de señales obtenidas a partir de la señal de audio original durante el proceso de codificación/descodificación. En particular, los valores de tonalidad 231, 232, 233 pueden determinarse para subbandas y/o grupos de subbandas de la señal de audio original, de la componente de baja frecuencia codificada/descodificada de la señal de audio y/o de la componente de alta frecuencia aproximada de la señal de audio. Como se ha indicado anteriormente, la determinación de valores de tonalidad 231, 232, 233 normalmente supone una parte considerable del esfuerzo computacional total de un codificador basado en SPX. A continuación se describen procedimientos y sistemas que permiten reducir considerablemente el esfuerzo computacional asociado a la determinación de los valores de tonalidad 231, 232, 233, reduciéndose así la complejidad computacional del codificador basado en SPX.

El valor de tonalidad de una subbanda 205, 206 puede determinarse analizando la evolución de la velocidad angular $\omega(t)$ de las subbandas 205, 206 a lo largo del tiempo t . La velocidad angular $\omega(t)$ puede ser la variación del ángulo o fase φ en el tiempo. Por consiguiente, la aceleración angular puede determinarse como la variación de la velocidad angular $\omega(t)$ en el tiempo, es decir, la primera derivada de la velocidad angular $\omega(t)$ o la segunda derivada de la fase φ . Si la velocidad angular $\omega(t)$ es constante a lo largo del tiempo, la subbanda 205, 206 es tonal, y si la velocidad

angular $\omega(t)$ varía a lo largo del tiempo, la subbanda 205, 206 es menos tonal. Por tanto, la tasa de cambio de la velocidad angular $\omega(t)$ (es decir, la aceleración angular) es un indicador de la tonalidad. A modo de ejemplo, los valores de tonalidad T_q 231, 232, 233 de una subbanda q de un grupo de subbandas q puede determinarse como

$$T_q = 1 - \left| \frac{\partial \omega(t)}{\partial t} \right| = 1 - |\alpha|, (|\alpha| \leq 1)$$

En el presente documento se propone dividir la determinación de los valores de tonalidad T_q 231, 232, 233 de una subbanda q o de un grupo de subbandas q (también denominados valores de tonalidad por banda) en una determinación de valores de tonalidad T_n para los diferentes coeficientes de transformada TC (es decir, para diferentes celdas de frecuencia n) obtenidos mediante la transformada del dominio de tiempo al dominio de frecuencia (denominados también valores de tonalidad de celda) y para determinar posteriormente los valores de tonalidad por banda T_q 231, 232, 233 en función de los valores de tonalidad de celda T_n . Como se muestra a continuación, esta determinación en dos etapas de los valores de tonalidad por banda T_q 231, 232, 233 permite reducir de manera considerable el esfuerzo computacional asociado al cálculo de los valores de tonalidad por banda T_q 231, 232, 233.

En el dominio de tiempo discreto, el valor de tonalidad de celda $T_{n,k}$ para un coeficiente de transformada TC de una celda de frecuencia n y en un bloque (o instante de tiempo discreto) k puede determinarse, por ejemplo, en función de la fórmula

$$T_{n,k} = w_{n,k} \cdot \left(1 - \frac{|\text{anglenorm}(\varphi_{n,k} - 2 \cdot \varphi_{n,k-1} + \varphi_{n,k-2})|}{\pi} \right) \cdot |TC_{n,k}|^2,$$

donde $\varphi_{n,k}$, $\varphi_{n,k-1}$ y $\varphi_{n,k-2}$ son las fases del coeficiente de transformada TC de la celda de frecuencia n en los instantes de tiempo k , $k-1$ y $k-2$, respectivamente, donde $|TC_{n,k}|^2$ es la magnitud elevada al cuadrado del coeficiente de transformada TC de la celda de frecuencia n en los instantes de tiempo k , y donde $w_{n,k}$ es un factor de ponderación para la celda de frecuencia n en el instante de tiempo k . La función "anglenorm" normaliza su argumento con respecto al intervalo $(-\pi, \pi]$ mediante una adición/sustracción repetida de 2π . La función "anglenorm" viene dada en la Tabla 1.

```
function anglenorm(x)
{
    while (x > pi)
    {
        x = x - 2*pi;
    }

    while (x <= -2*pi)
    {
        x = x + 2*pi;
    }

    return x;
}
```

Tabla 1

El valor de tonalidad $T_{q,k}$ 231, 232, 233 de una subbanda q 205, 206 o de un grupo de subbandas q 205, 206 en un instante de tiempo k (o para un bloque k) puede determinarse en función de los valores de tonalidad $T_{n,k}$ de las celdas de frecuencia n en el instante de tiempo k (o para el bloque k) comprendidas en la subbanda q 205, 206 o en el grupo de subbandas q 205, 206 (por ejemplo, en función de la suma o el promedio de los valores de tonalidad $T_{n,k}$). En el presente documento, el índice de tiempo (o índice de bloque) k y/o el índice de celda n / índice de subbanda q pueden haberse omitido para una mayor claridad.

La fase φ_k (para una celda particular n) puede determinarse a partir de la parte real y de la parte imaginaria de un TC complejo. Los TC complejos pueden determinarse en el lado del codificador, por ejemplo llevando a cabo una transformada MDST y una transformada MDCT de un bloque de N muestras de la señal de audio, obteniéndose así la parte real y la parte imaginaria de los TC complejos, respectivamente. Como alternativa, pueden usarse las transformadas complejas del dominio de tiempo al dominio de frecuencia, obteniéndose así TC complejos. La fase φ_k puede determinarse entonces como

$$\varphi_k = \text{atan2}(\text{Im}\{TC_k\}, \text{Re}\{TC_k\}) - \pi < \varphi_k \leq \pi .$$

La función atan2 está especificada, por ejemplo, en el enlace de Internet <http://de.wikipedia.org/wiki/Atan2#atan2>. En principio, la función atan2 puede describirse como una función de arcotangente de la relación de $y = \text{Im}\{TC_k\}$ y $x = \text{Re}\{TC_k\}$ que tiene en cuenta valores negativos de $y = \text{Im}\{TC_k\}$ y/o $x = \text{Re}\{TC_k\}$. Como se ha descrito en el contexto de las Fig. 2a, 2b, 2c y 2d, puede ser necesario determinar diferentes valores de tonalidad por banda 231, 232, 233 en función de diferentes datos espectrales 200, 210, 220 obtenidos a partir de la señal de audio original. En relación con lo ilustrado en la Fig. 2a, se ha observado que los diferentes cálculos de tonalidad por banda están basados realmente en los mismos datos, en particular en los mismos coeficientes de transformada (TC):

1. La tonalidad de los TC de banda de alta frecuencia original se usa para determinar la estrategia de reenvío de coordenadas SPX y la LVA, así como para calcular el factor de mezcla de ruido b . Dicho de otro modo, los valores de tonalidad de celda T_n de los TC de la banda de alta frecuencia original 102 pueden usarse para determinar los valores de tonalidad por banda 231 y el valor de tonalidad por banda 232 de la banda de alta frecuencia 102.
2. La tonalidad de los TC de banda baja desacoplada/desmatrizada se usa para determinar el factor de mezcla de ruido b y, tras la traslación a la banda alta, se usa en los cálculos de LVA. Dicho de otro modo, los valores de tonalidad de celda T_n que se determinan en función de los TC de la componente de baja frecuencia codificada/descodificada de la señal de audio (espectro 210) se usan para determinar el valor de tonalidad por banda 232 de la banda base 101 y para determinar los valores de tonalidad por banda 233 de la banda de alta frecuencia 102. Esto se debe a que los TC de las subbandas de la banda de alta frecuencia 102 del espectro 220 se obtienen mediante la traslación de una o más subbandas codificadas/descodificadas de la banda base 101 a una o más subbandas de la banda de alta frecuencia 102. Este proceso de traslación no afecta a la tonalidad de los TC copiados, permitiendo así la reutilización de los valores de tonalidad de celda T_n que se determinan en función de los TC de la componente de baja frecuencia codificada/descodificada de la señal de audio (espectro 210).
3. Normalmente, los TC de banda baja desacoplada/desmatrizada solo se diferencian de los TC originales en la región de acoplamiento (suponiendo que la matrización es completamente reversible, es decir, suponiendo que la operación de desmatrización reproduce los coeficientes de transformada originales). Los cálculos de tonalidad para las subbandas (y para los TC) entre la frecuencia de inicio SPX 201 y la frecuencia de comienzo de acoplamiento (*comienzo_cpl*) (suponiendo que está en la subbanda 2 en el ejemplo ilustrado) se basan en los TC originales no modificados y, por tanto, son los mismos para los TC de banda baja desacoplada/desmatrizada y para los TC originales (ilustrados en la Fig. 2a mediante el sombreado claro de las subbandas 0 y 1 del espectro 210).

Las observaciones descritas anteriormente sugieren que algunos de los cálculos de tonalidad no necesitan repetirse o, al menos, no necesitan llevarse a cabo completamente ya que los resultados intermedios calculados anteriormente pueden compartirse, es decir, reutilizarse. En muchos casos, los valores calculados con anterioridad pueden reutilizarse, lo que reduce en gran medida el coste computacional. A continuación se describen varias medidas que permiten reducir el coste computacional relacionado con la determinación de la tonalidad en un codificador basado en SPX.

Como puede observarse a partir de los espectros 200 y 210 de la Fig. 2a, las subbandas 7 a 14 de la banda de alta frecuencia 102 son las mismas en los espectros 200 y 210. De este modo, es posible reutilizar los valores de tonalidad por banda 231 para la banda de alta frecuencia 102 y también para el valor de tonalidad por banda 232. Desafortunadamente, la Fig. 2a muestra que la tonalidad se calcula para una estructura de banda diferente en ambos casos, incluso aunque los TC subyacentes sean idénticos. Por tanto, para poder reutilizar los valores de tonalidad, se propone dividir el cálculo de tonalidad en dos partes, donde la salida de la primera parte puede usarse para calcular los valores de tonalidad por banda 231 y 232.

Como se ha descrito anteriormente, el cálculo de las tonalidades por banda T_q puede dividirse en calcular la tonalidad por celda T_n para cada TC (etapa 1) y en un proceso subsiguiente de alisamiento y agrupación de los valores de tonalidad de celda T_n en bandas (etapa 2), obteniéndose así los respectivos valores de tonalidad por banda T_q 231, 232, 233. Los valores de tonalidad por banda T_q 231, 232, 233 pueden determinarse en función de una suma de los valores de tonalidad de celda T_n de las celdas comprendidas en la banda o subbanda del valor de tonalidad por banda, por ejemplo en función de una suma ponderada de los valores de tonalidad de celda T_n . A modo de ejemplo, un valor de tonalidad por banda T_q puede determinarse en función de la suma de los valores de

tonalidad de celda pertinentes T_n dividida por la suma de los factores de ponderación correspondientes w_n . Además, la determinación de los valores de tonalidad por banda T_q puede comprender una correspondencia y/o correlación de la suma (ponderada) con un intervalo de valores predeterminado (por ejemplo, [0, 1]). A partir del resultado de la etapa 1 pueden obtenerse valores de tonalidad por banda arbitrarios T_q . Debe observarse que la complejidad computacional reside principalmente en la etapa 1 que, de este modo, proporciona la ganancia de eficacia de este enfoque de dos etapas.

El enfoque de dos etapas para determinar los valores de tonalidad por banda T_q se ilustra en la Fig. 3b para las subbandas 7 a 14 de la banda de alta frecuencia 102. Puede observarse que en el ejemplo ilustrado, cada subbanda está constituida por 12 TC en 12 celdas de frecuencia correspondientes. En una primera etapa (etapa 1) se determinan valores de tonalidad de celda T_n 341 para las celdas de frecuencia de las subbandas 7 a 14. En una segunda etapa (etapa 2), los valores de tonalidad de celda T_n 341 se agrupan de diferentes maneras para determinar los valores de tonalidad por banda T_q 312 (que corresponden a los valores de tonalidad por banda T_q 231 de la banda de alta frecuencia 102) y para determinar el valor de tonalidad por banda T_q 322 (que corresponde a los valores de tonalidad por banda T_q 232 de la banda de alta frecuencia 102).

Como resultado, la complejidad computacional para determinar el valor de tonalidad por banda 322 y los valores de tonalidad por banda 312 puede reducirse en casi el 50%, ya que los valores de tonalidad por banda 312, 322 utilizan los mismos valores de tonalidad de celda 341. Esto se ilustra en la Fig. 3a, que muestra que reutilizando también la tonalidad de banda alta de la señal original para la mezcla de ruido y eliminando consecuentemente los cálculos adicionales (número de referencia 302), puede reducirse el número de cálculos de tonalidad. Lo mismo se aplica a los valores de tonalidad de celda 341 para las subbandas 0, 1 por debajo de la frecuencia de comienzo de acoplamiento (*comienzo_cpl*) 303. Estos valores de tonalidad de celda 341 pueden usarse para determinar los valores de tonalidad por banda 311 (que corresponden a los valores de tonalidad por banda T_q 231 de la banda base 101), y pueden reutilizarse para determinar el valor de tonalidad por banda 321 (que corresponde a los valores de tonalidad por banda T_q 232 de la banda base 101).

Debe observarse que el enfoque de dos etapas para determinar los valores de tonalidad por banda es transparente con respecto a la salida del codificador. Dicho de otro modo, los valores de tonalidad por banda 311, 312, 321 y 322 no se ven afectados por el cálculo de dos etapas y, por lo tanto, son idénticos a los valores de tonalidad por banda 231, 232 que se determinan en un cálculo de una etapa.

La reutilización de los valores de tonalidad de celda 341 también puede aplicarse en el contexto de la traslación espectral. Tal escenario de reutilización implica normalmente subbandas desmatrizadas/desacopladas de la banda base 101 del espectro 210. Un valor de tonalidad por banda 321 de estas subbandas se calcula cuando se determina el factor de mezcla de ruido b (véase la Fig. 3a). De nuevo, al menos algunos de los mismos TC que se usan para determinar el valor de tonalidad por banda 321 se usan para calcular valores de tonalidad por banda 233 que controlan la atenuación de varianza elevada (LVA). La diferencia con respecto al primer escenario de reutilización descrito en el contexto de las Fig. 3a y 3b es que los TC se someten a una traslación espectral antes de que se usen para calcular los valores de tonalidad LVA 233. Sin embargo, puede mostrarse que la tonalidad por celda T_n 341 de una celda es independiente de la tonalidad de sus celdas vecinas. En consecuencia, los valores de tonalidad por celda T_n 341 pueden trasladarse en frecuencia de la misma manera que la utilizada para los TC (véase la Fig. 3d). Esto permite reutilizar los valores de tonalidad de celda T_n 341 calculados en la banda base 101 para la mezcla de ruido, en los cálculos de la LVA de la banda de alta frecuencia 102. Esto se ilustra en la Fig. 3c, donde se muestra cómo las subbandas de la banda de alta frecuencia reconstruida 102 se obtienen a partir de las subbandas 0 a 5 de la banda base 101 del espectro 210. Según el proceso de traslación espectral, los valores de tonalidad de celda T_n 341 de las celdas de frecuencia comprendidas en las subbandas 0 a 5 de la banda base 101 pueden reutilizarse para determinar los valores de tonalidad por banda T_q 233. Como resultado, el esfuerzo computacional para determinar los valores de tonalidad por banda T_q 233 se reduce considerablemente, como se ilustra mediante el número de referencia 303. De nuevo, debe observarse que la salida del codificador no se ve afectada por esta manera modificada de obtener la tonalidad de banda de extensión 233.

En términos generales, se ha mostrado que dividiendo la determinación de valores de tonalidad por banda T_q en un enfoque de dos etapas que implica una primera etapa que determina valores de tonalidad por celda T_n y una segunda etapa posterior que determina los valores de tonalidad por banda T_q a partir de los valores de tonalidad por celda T_n , la complejidad computacional global relacionada con el cálculo de los valores de tonalidad por banda T_q puede reducirse. En particular, se ha mostrado que el enfoque de dos etapas permite reutilizar valores de tonalidad por celda T_n para la determinación de una pluralidad de valores de tonalidad por banda T_q (como se ilustra mediante los números de referencia 301, 302, 303 que indican la posibilidad de reutilización), reduciéndose así la complejidad computacional global.

La mejora de rendimiento que se obtiene con el enfoque de dos etapas y con la reutilización de los valores de tonalidad de celda puede cuantificarse comparando el número de celdas para las que se calcula normalmente la tonalidad. El esquema original calcula valores de tonalidad para

$$2 \cdot (\text{finalización_spx} - \text{inicio_spx}) + (\text{finalización_spx} - \text{comienzo_spx}) + 6$$

celdas de frecuencia (donde los 6 valores de tonalidad adicionales se usan para configurar filtros de muesca específicos en el codificador basado en SPX). Reutilizando los valores de tonalidad calculados descritos anteriormente, el número de celdas, para las que se determina un valor de tonalidad, se reduce a

5

$$\begin{aligned} & \text{finalización_spx} - \text{inicio_spx} - \text{comienzo_cpl} + \text{inicio_spx} \\ & + \min(\text{finalización_spx} - \text{comienzo_spx} + 3, \text{comienzo_spx} - \text{inicio_spx}) \\ = & \text{finalización_spx} - \text{comienzo_cpl} + \min(\text{finalización_spx} - \text{comienzo_spx} + 3, \text{comienzo_spx} - \text{inicio_spx}) \end{aligned}$$

10 (donde los 3 valores de tonalidad adicionales se usan para configurar filtros de muesca específicos en el codificador basado en SPX). La proporción de celdas para las que se calcula la tonalidad antes y después de la optimización, da como resultado la mejora del rendimiento (y la reducción de complejidad) para el algoritmo de tonalidad. Debe observarse que el enfoque de dos etapas es, por lo general, algo más complejo que el cálculo directo de valores de tonalidad por banda. La ganancia de rendimiento (es decir, la reducción de complejidad) para el cálculo de tonalidad completo es, por tanto, ligeramente inferior a la proporción de celdas de tonalidad calculada que puede encontrarse en la Tabla 2 para diferentes velocidades binarias.

15

Tabla 2

Velocidad binaria (kbps)	Proporción de celdas de tonalidad antes/después
128	0,50
192	0,52
256	0,45
320	0,41

20 Se observa que puede conseguirse una reducción del 50%, y superior, de la complejidad computacional para calcular los valores de tonalidad.

Como se ha descrito anteriormente, el enfoque de dos etapas no afecta a la salida del codificador. A continuación se describen medidas adicionales para reducir la complejidad computacional de un codificador basado en SPX que pueden afectar a la salida del codificador. Sin embargo, pruebas perceptuales han mostrado que, por término medio, estas medidas adicionales no afectan a la calidad percibida de las señales de audio codificadas. Las medidas descritas a continuación pueden usarse además de, o como alternativa a, las otras medidas descritas en el presente documento.

25

30 Como se muestra, por ejemplo, en el contexto de la Fig. 3c, los valores de tonalidad por banda T_{baja} 321 y T_{alta} 322 son la base para el cálculo del factor de mezcla de ruido b . La tonalidad puede interpretarse como una propiedad que es más o menos inversa a la cantidad de ruido contenido en la señal de audio (es decir, más ruidosa → menos tonal y viceversa). El factor de mezcla de ruido b puede calcularse como

35

$$b = T_{baja} \cdot (1 - \text{var}\{T_{baja}, T_{alta}\}) + T_{alta} \cdot (\text{var}\{T_{baja}, T_{alta}\}),$$

donde T_{baja} 321 es la tonalidad de la banda baja simulada por descodificador, T_{alta} 322 es la tonalidad de la banda

alta original y $\text{var}\{T_{baja}, T_{alta}\} = \left(\frac{T_{baja} - T_{alta}}{T_{baja} + T_{alta}}\right)^2$ es la varianza de los dos valores de tonalidad T_{baja} 321 y T_{alta} 322.

40 El objetivo de la mezcla de ruido es introducir tanto ruido como sea necesario en la banda alta regenerada para hacer que la banda alta regenerada suene como la banda alta original. El valor de tonalidad fuente (que refleja la tonalidad de las subbandas trasladadas en la banda de alta frecuencia 102) y el valor de tonalidad objetivo (que refleja la tonalidad de las subbandas en la banda de alta frecuencia original 102) deben tenerse en cuenta para determinar el nivel de ruido objetivo deseado. Se ha observado que la tonalidad fuente verdadera no se describe correctamente mediante el valor de tonalidad T_{baja} 321 de la banda baja simulada por descodificador, sino mediante un valor de tonalidad T_{copia} 323 de la copia de banda de alta trasladada (véase la Fig. 3c). El valor de tonalidad T_{copia} 323 puede determinarse en función de las subbandas que aproximan las subbandas originales 7 a 14 de la banda de alta frecuencia 102, como se ilustra mediante la llave de la Fig. 3c. Es en la banda alta trasladada donde se lleva a cabo la mezcla de ruido y, por tanto, solamente la tonalidad de los TC de banda baja que se copian realmente en la banda alta debería afectar a la cantidad de ruido que va a añadirse.

50

Como se indica mediante la fórmula anterior, el valor de tonalidad T_{baja} 321 de la banda baja se usa actualmente como una estimación de la tonalidad fuente verdadera. Puede haber dos casos que afecten a la precisión de esta estimación:

1. La banda baja que se usa para aproximar la banda alta es más pequeña que o igual a la banda alta, y el codificador no se encuentra con una envoltura de banda media (es decir, la banda objetivo es más grande que las bandas fuente disponibles al final de la región de copia (es decir, la región entre inicio_spx y comienzo_spx)). El codificador normalmente trata de evitar tales situaciones de envoltura dentro de la banda SPX objetivo. Esto se ilustra en la Fig. 3c, donde la subbanda trasladada 5 va seguida de las subbandas 0 y 1 (con el fin de evitar una situación de envoltura de la subbanda 6 tras la subbanda 0 dentro de la banda SPX objetivo). En este caso, la banda baja se copia, por lo general, completamente, posiblemente múltiples veces, en la banda alta. Puesto que todos los TC van a copiarse, la estimación de tonalidad para la banda baja debería ser muy parecida a la estimación de tonalidad de la banda alta trasladada.
2. La banda baja es más grande que la banda alta. En este caso, solamente la parte inferior de la banda baja se copia en la banda alta. Puesto que el valor de tonalidad T_{baja} 321 se calcula para todos los TC de banda baja, el valor de tonalidad T_{copia} 323 de la banda alta trasladada puede desviarse del valor de tonalidad T_{baja} 321, dependiendo de las propiedades de señal y dependiendo de la relación de tamaño de la banda baja y la banda alta.

De este modo, el uso del valor de tonalidad T_{baja} 321 puede dar lugar a un factor de mezcla de ruido poco preciso b , especialmente en casos en los que no todas las subbandas 0 a 6 que se usan para determinar el valor de tonalidad T_{baja} 321 se trasladan a la banda de alta frecuencia 102 (como es el caso, p. ej., del ejemplo mostrado en la Fig. 3c). Pueden producirse imprecisiones importantes cuando las subbandas que no se copian en la banda de alta frecuencia 102 (por ejemplo, la subbanda 6 de la Fig. 3c) comprenden un contenido tonal significativo. Por lo tanto, se propone determinar el factor de mezcla de ruido b en función del valor de tonalidad por banda T_{copia} 323 de la banda alta trasladada (y no en función del valor de tonalidad por banda T_{baja} 321 de la banda baja simulada por descodificador comprendida entre la frecuencia de inicio SPX 201 y la frecuencia de comienzo SPX 202. En particular, el factor de mezcla de ruido b puede determinarse como

$$b = T_{copia} \cdot (1 - \text{var}\{T_{copia}, T_{alta}\}) + T_{alta} \cdot (\text{var}\{T_{copia}, T_{alta}\}),$$

$$\text{var}\{T_{copia}, T_{alta}\} = \left(\frac{T_{copia} - T_{alta}}{T_{copia} + T_{alta}} \right)^2$$

donde es la varianza de los dos valores de tonalidad T_{copia} 323 y T_{alta} 322.

Además de proporcionar potencialmente una mayor calidad del codificador basado en SPX, el uso del valor de tonalidad por banda T_{copia} 323 de la banda alta trasladada (en lugar del valor de tonalidad por banda T_{baja} 321 de la banda baja simulada por descodificador) puede dar lugar a una menor complejidad computacional del codificador de audio basado en SPX. Esto es particularmente cierto para el caso 2 mencionado anteriormente, donde la banda alta trasladada es más estrecha que la banda baja. Este beneficio aumenta con la disparidad de los tamaños de la banda baja y la banda alta. La cantidad de bandas para las que se calcula la tonalidad fuente puede ser $\min\{\text{comienzo_spx} - \text{inicio_spx}, \text{finalización_spx} - \text{comienzo_spx}\}$, donde el número $(\text{comienzo_spx} - \text{inicio_spx})$ se aplica si el factor de mezcla de ruido b se determina en función del valor de tonalidad por banda T_{baja} 321 de la banda baja simulada por descodificador, y donde el número $(\text{finalización_spx} - \text{comienzo_spx})$ se aplica si el factor de mezcla de ruido b se determina en función del valor de tonalidad por banda T_{copia} 323 de la banda alta trasladada. De este modo, en una forma de realización, el codificador basado en SPX puede configurarse para seleccionar el modo de determinación del factor de mezcla de ruido b (un primer modo basado en el valor de tonalidad por banda T_{baja} 321 y un segundo modo basado en el valor de tonalidad por banda T_{copia} 323), dependiendo del mínimo entre $(\text{comienzo_spx} - \text{inicio_spx})$ y $(\text{finalización_spx} - \text{comienzo_spx})$, reduciéndose así la complejidad computacional (especialmente en los casos en que $(\text{finalización_spx} - \text{comienzo_spx})$ es menor que $(\text{comienzo_spx} - \text{inicio_spx})$).

Debe observarse que el esquema modificado para determinar el factor de mezcla de ruido b puede combinarse con el enfoque de dos etapas para determinar los valores de tonalidad por banda T_{copia} 323 y/o T_{alta} 322. En este caso, el valor de tonalidad por banda T_{copia} 323 se determina en función de los valores de tonalidad de celda T_n 341 de las celdas de frecuencia que se han trasladado a la banda de alta frecuencia 102. Las celdas de frecuencia que contribuyen a la banda de alta frecuencia reconstruida 102 están situadas entre inicio_spx 201 y comienzo_spx 202. En el peor caso con respecto a la complejidad computacional, todas las celdas de frecuencia entre inicio_spx 201 y comienzo_spx 202 contribuyen a la banda de alta frecuencia reconstruida 102. Por otro lado, en otros muchos casos (como se ilustra, por ejemplo, en la Fig. 3c) solamente un subconjunto de las celdas de frecuencia entre inicio_spx 201 y comienzo_spx 202 se copia en la banda de alta frecuencia reconstruida 102. Por este motivo, en una forma de realización, el factor de mezcla de ruido b se determina en función del valor de tonalidad por banda T_{copia} 323 usando los valores de tonalidad de celda T_n 341, es decir, usando el enfoque de dos etapas antes mencionado para determinar el valor de tonalidad por banda T_{copia} 323. Usando el enfoque de dos etapas, se garantiza que incluso en casos en los que $(\text{comienzo_spx} - \text{inicio_spx})$ sea menor que $(\text{finalización_spx} - \text{comienzo_spx})$, la complejidad computacional está limitada por la complejidad computacional requerida para determinar los valores de tonalidad de

celda T_n 341 en la gama de frecuencias entre *inicio_spx* 201 y *comienzo_spx* 202. Dicho de otro modo, el enfoque de dos etapas garantiza que incluso en casos en los que (*comienzo_spx* - *inicio_spx*) sea menor que (*finalización_spx* - *comienzo_spx*), la complejidad computacional para determinar el valor de tonalidad por banda T_{copia} 323 está limitada por el número de TC comprendidos entre (*comienzo_spx* - *inicio_spx*). De este modo, el factor de mezcla de ruido b puede determinarse de manera coherente en función del valor de tonalidad por banda T_{copia} 323. Sin embargo, puede ser beneficioso determinar el mínimo entre (*comienzo_spx* - *inicio_spx*) y (*finalización_spx* - *comienzo_spx*) con el fin de determinar las subbandas en la región de acoplamiento (desde *comienzo_cpl* hasta *comienzo_spx*) para las que deben determinarse los valores de tonalidad. A modo de ejemplo, si (*comienzo_spx* - *inicio_spx*) es mayor que (*finalización_spx* - *comienzo_spx*), no es necesario determinar los valores de tonalidad para al menos algunas de las subbandas de la región de frecuencia (*comienzo_spx* - *inicio_spx*), reduciéndose así la complejidad computacional.

Como puede observarse en la Fig. 3c, el enfoque de dos etapas para determinar los valores de tonalidad por banda a partir de los valores de tonalidad de celda permite una reutilización importante de los valores de tonalidad de celda, reduciéndose así la complejidad computacional. La determinación de los valores de tonalidad de celda se reduce principalmente a la determinación de valores de tonalidad de celda en función del espectro 200 de la señal de audio original. Sin embargo, en caso de acoplamiento, puede ser necesario determinar los valores de tonalidad de celda en función del espectro acoplado/desacoplado 210 para algunas o todas las celdas de frecuencia entre *comienzo_cpl* 303 y *comienzo_spx* 202 (para las celdas de frecuencia de las subbandas sombreadas con color oscuro 2 a 6 de la Fig. 3c). Dicho de otro modo, tras aprovechar los medios antes mencionados para reutilizar la tonalidad por celda previamente calculada, las únicas bandas que pueden requerir un nuevo cálculo de tonalidad son las bandas que están acopadas (véase la Fig. 3c).

El acoplamiento normalmente elimina las diferencias de fase entre los canales de una señal multicanal (por ejemplo, una señal estéreo o una señal multicanal 5.1) que están acoplados. La compartición de frecuencia y la compartición de tiempo de las coordenadas de acoplamiento incrementan adicionalmente la correlación entre los canales acoplados. Como se ha descrito anteriormente, la determinación de los valores de tonalidad se basa en fases y energías del bloque actual de muestras (en el instante de tiempo k) y de uno o más bloques anteriores de muestras (por ejemplo, en los instantes de tiempo $k-1$, $k-2$). Puesto que los ángulos de fase de todos los canales acoplados son idénticos (como resultado del acoplamiento), los valores de tonalidad de esos canales están más correlacionados que los valores de tonalidad de la señal original.

Un decodificador correspondiente a un codificador basado en SPX solo tiene acceso a la señal desacoplada que el decodificador genera a partir del flujo de bits recibido que comprende datos de audio codificados. Las herramientas de codificación, como la mezcla de ruido y la atenuación de varianza elevada (LVA), en el lado del codificador tienen esto en cuenta normalmente cuando calculan relaciones que pretenden reproducir la señal de banda alta original a partir de la señal de banda baja desacoplada transpuesta. Dicho de otro modo, el codificador de audio basado en SPX tiene en cuenta normalmente que el decodificador correspondiente solo tiene acceso a los datos codificados (representativos de la señal de audio desacoplada). Por tanto, la tonalidad fuente para la mezcla de ruido y LVA se calcula normalmente a partir de la señal desacoplada en el codificador basado en SPX actual (como se ilustra, por ejemplo, en el espectro 210 de la Fig. 2a). Sin embargo, incluso aunque conceptualmente tenga sentido calcular la tonalidad en función de la señal desacoplada (es decir, en función del espectro 210), las implicaciones perceptuales de calcular la tonalidad a partir de la señal original no son muy claras. Además, la complejidad computacional puede reducirse adicionalmente si puede evitarse el nuevo cálculo adicional de valores de tonalidad basado en la señal desacoplada.

Con este fin, se ha llevado a cabo un experimento de escucha para evaluar la influencia perceptual de usar la tonalidad de la señal original en lugar de la tonalidad de la señal desacoplada (para determinar los valores de tonalidad por banda 321 y 233). Los resultados del experimento de escucha se ilustran en la Fig. 4. Se han llevado a cabo pruebas MUSHRA (múltiples estímulos con referencia y patrón ocultos) para una pluralidad de diferentes señales de audio. Para cada una de la pluralidad de diferentes señales de audio, las barras (situadas a la izquierda) 401 indican los resultados obtenidos cuando se determinan los valores de tonalidad en función de la señal desacoplada (usando el espectro 210) y las barras (situadas a la derecha) 402 indican los resultados obtenidos cuando se determinan los valores de tonalidad en función de la señal original (usando el espectro 200). Como puede observarse, la calidad de audio obtenida cuando se usa la señal de audio original para la determinación de los valores de tonalidad para la mezcla de ruido y para LVA es la misma, por término medio, que la calidad de audio obtenida cuando se usa la señal de audio desacoplada para la determinación de los valores de tonalidad.

Los resultados del experimento de escucha de la Fig. 4 sugieren que la complejidad computacional para determinar los valores de tonalidad puede reducirse adicionalmente reutilizando los valores de tonalidad de celda 341 de la señal de audio original para determinar el valor de tonalidad por banda 321 y/o el valor de tonalidad por banda 323 (usado para la mezcla de ruido) y los valores de tonalidad por banda 233 (usados para LVA). Por tanto, la complejidad computacional del codificador de audio basado en SPX puede reducirse adicionalmente sin afectar (por lo general) a la calidad de audio percibida de las señales de audio codificadas.

65

Incluso cuando se determinan los valores de tonalidad por banda 321 y 233 en función de la señal de audio desacoplada (es decir, en función de las subbandas sombreadas de color oscuro 2 a 6 del espectro 210 de la Fig. 3c), la alineación de las fases debida al acoplamiento puede usarse para reducir la complejidad computacional asociada a la determinación de tonalidad. Dicho de otro modo, incluso si el nuevo cálculo de la tonalidad para las bandas de acoplamiento no puede evitarse, la señal desacoplada presenta una propiedad especial que puede usarse para simplificar el cálculo de tonalidad habitual. La propiedad especial es que todos los canales acoplados (y posteriormente desacoplados) están en fase. Puesto que todos los canales acoplados comparten la misma fase φ para las bandas de acoplamiento, esta fase φ solo necesita calcularse una vez para un canal y después puede reutilizarse en los cálculos de tonalidad de los otros canales acoplados. En particular, esto significa que la operación "atan2" antes mencionada para determinar la fase φ_k en un instante de tiempo k solo necesita llevarse a cabo una vez para todos los canales de una señal multicanal que están acoplados.

Desde un punto de vista numérico, parece beneficioso usar el propio canal de acoplamiento para el cálculo de fase (en lugar de uno de los canales desacoplados), ya que el canal de acoplamiento representa un promedio de todos los canales acoplados. La reutilización de fase para los canales acoplados se ha implementado en el codificador SPX. No hay ningún cambio en la salida del codificador debido a la reutilización de los valores de fase. La ganancia de rendimiento es del 3% aproximadamente (del esfuerzo computacional del codificador SPX) para la configuración medida a una velocidad binaria de 256 kbps, pero se espera que la ganancia de rendimiento aumente para velocidades binarias más bajas cuando la región de acoplamiento empiece más cerca de la frecuencia de inicio SPX 201, es decir, cuando la frecuencia de comienzo de acoplamiento 303 esté más cerca de la frecuencia de inicio SPX 201.

A continuación se describe un enfoque adicional para reducir la complejidad computacional asociada a la determinación de la tonalidad. Este enfoque puede usarse además de, o como alternativa a, los otros procedimientos descritos en el presente documento. A diferencia de las optimizaciones antes descritas que se centraban en reducir el número de cálculos de tonalidad requeridos, el siguiente enfoque tiene como objetivo acelerar el propio cálculo de tonalidad. En particular, el siguiente enfoque tiene como objetivo reducir la complejidad computacional para determinar el valor de tonalidad de celda $T_{n,k}$ de una celda de frecuencia n para un bloque k (el índice k corresponde, por ejemplo, a un instante de tiempo k).

El valor de tonalidad por celda SPX $T_{n,k}$ de la celda n en el bloque k puede calcularse como

$$T_{n,k} = w_{n,k} \cdot \left(1 - \frac{\text{anglenorm}(\varphi_{n,k} - 2 \cdot \varphi_{n,k-1} + \varphi_{n,k-2})}{\pi} \right) \cdot Y_{n,k}$$

donde $Y_{n,k} = \text{Re}\{TC_{n,k}\}^2 + \text{Im}\{TC_{n,k}\}^2$ es la potencia de celda n y bloque k , $w_{n,k}$ es un factor de ponderación y $\varphi_{n,k} = \text{atan2}(\text{Re}\{TC_{n,k}\}, \text{Im}\{TC_{n,k}\})$ es el ángulo de fase de celda n y bloque k . La fórmula antes mencionada para el valor de tonalidad de celda $T_{n,k}$ indica la aceleración del ángulo de fase (como se ha descrito en el contexto de las fórmulas dadas para el valor de tonalidad de celda $T_{n,k}$ anterior). Debe observarse que pueden usarse otras fórmulas para determinar el valor de tonalidad de celda $T_{n,k}$. La aceleración de los cálculos de tonalidad (es decir, la reducción de la complejidad computacional) está dirigida principalmente a la reducción de la complejidad computacional asociada a la determinación del factor de ponderación w .

El factor de ponderación w puede definirse como

$$w_{n,k} = \begin{cases} \sqrt[4]{\frac{Y_{n,k}}{Y_{n,k-1}}} & \text{para } Y_{n,k} \leq Y_{n,k-1} \\ \sqrt[4]{\frac{Y_{n,k-1}}{Y_{n,k}}} & \text{para } Y_{n,k} > Y_{n,k-1} \end{cases}$$

El factor de ponderación w puede aproximarse sustituyendo la raíz cuarta por una raíz cuadrada y la primera iteración del método babilónico/de Herón, es decir,

$$w_{n,k} \cong \begin{cases} 0 & \text{para } Y_{n,k} = 0 \cup Y_{n,k-1} = 0 \\ \frac{1}{2} + \frac{1}{2} \cdot \begin{cases} \sqrt{\frac{Y_{n,k}}{Y_{n,k-1}}} & \text{para } Y_{n,k} \leq Y_{n,k-1} \\ \sqrt{\frac{Y_{n,k-1}}{Y_{n,k}}} & \text{para } Y_{n,k} > Y_{n,k-1} \end{cases} & \end{cases}$$

5 Aunque la eliminación de una operación de raíz cuadrada ya aumenta la eficacia, queda aún una operación de raíz cuadrada y una división por bloque, por canal y por celda de frecuencia. Una aproximación diferente y más eficaz computacionalmente puede obtenerse en el dominio logarítmico reescribiendo el factor de ponderación w como:

$$w_{n,k} = \begin{cases} 2^{\log_2\left(\sqrt[4]{\frac{Y_{n,k}}{Y_{n,k-1}}}\right)} = 2^{\frac{1}{4}(\log_2(Y_{n,k}) - \log_2(Y_{n,k-1}))} & \text{para } Y_{n,k} \leq Y_{n,k-1} \\ 2^{\log_2\left(\sqrt[4]{\frac{Y_{n,k-1}}{Y_{n,k}}}\right)} = 2^{\frac{1}{4}(\log_2(Y_{n,k-1}) - \log_2(Y_{n,k}))} & \text{para } Y_{n,k} > Y_{n,k-1} \end{cases}$$

10 La distinción de los casos puede omitirse observando que la diferencia en el dominio logarítmico es siempre negativa, independientemente de que $(Y_{n,k} \leq Y_{n,k-1})$ o de que $(Y_{n,k} > Y_{n,k-1})$, obteniéndose así:

$$w_{n,k} = 2^{-\frac{1}{4}|\log_2(Y_{n,k}) - \log_2(Y_{n,k-1})|}$$

15 Para facilitar la notación, los índices se omiten e $Y_{n,k}$ e $Y_{n,k-1}$ se sustituyen por y y z , respectivamente:

$$w = 2^{-\frac{1}{4}|\log_2(y) - \log_2(z)|}$$

20 Las variables y y z puede dividirse ahora en un exponente e_y , e_z y en una mantisa normalizada m_y , m_z , respectivamente, obteniéndose así:

$$w = 2^{-\frac{1}{4}|\log_2(m_y \cdot 2^{e_y}) - \log_2(m_z \cdot 2^{e_z})|} = 2^{-\frac{1}{4}|e_y + \log_2(m_y) - e_z - \log_2(m_z)|}$$

25 Suponiendo que el caso especial de una mantisa de valor cero se trata por separado, las mantisas normalizadas m_y, m_z están dentro del intervalo $[0,5;1]$. La función $\log_2(x)$ en este intervalo puede aproximarse mediante la función lineal $\log_2(x) \approx 2 \cdot x - 2$ con un error máximo de 0,0861 y un error medio de 0,0573. Debe observarse que otras aproximaciones (por ejemplo, una aproximación polinómica) pueden ser posibles, dependiendo de la precisión deseada de la aproximación y/o de la complejidad computacional. Usando la aproximación antes mencionada se obtiene:

30

$$w \cong 2^{-\frac{1}{4}|e_y - e_z + 2 \cdot m_y - 2 - (2 \cdot m_z - 2)|} = 2^{-\frac{1}{4}|e_y - e_z + 2 \cdot m_y - 2 \cdot m_z|}$$

35 La diferencia de las aproximaciones de mantisa sigue teniendo un error absoluto máximo de 0,0861, pero el error medio es cero, de modo que el intervalo del error máximo pasa de $[0;0,0861]$ (sesgado de manera positiva) a $[-0,0861;0,0861]$.

Dividir el resultado de la división por 4 en una parte entera y un resto da como resultado:

40

$$w \cong 2^{-\text{int}\left\{\frac{1}{4}|e_y - e_z + 2 \cdot m_y - 2 \cdot m_z|\right\} - \frac{\text{mod}\{|e_y - e_z + 2 \cdot m_y - 2 \cdot m_z|, 4\}}{4}}$$

donde la operación $\text{int}\{\dots\}$ devuelve la parte entera de su operando mediante truncado, y donde la operación $\text{mod}\{a,b\}$ devuelve el resto de a/b . En la aproximación anterior del factor de ponderación w , la primera expresión

$$2^{-\text{int}\left\{\frac{1}{4}|e_y - e_z + 2 \cdot m_y - 2 \cdot m_z|\right\}}$$

se convierte en una sencilla operación de desplazamiento hacia la derecha mediante

5
$$\text{int}\left\{\frac{1}{4}|e_y - e_z + 2 \cdot m_y - 2 \cdot m_z|\right\}$$

en una arquitectura de coma fija. La segunda expresión

10
$$2^{-\frac{\text{mod}\{|e_y - e_z + 2 \cdot m_y - 2 \cdot m_z|, 4\}}{4}}$$

puede calcularse usando una tabla de consulta predeterminada que comprende potencias de 2. La tabla de consulta puede comprender un número predeterminado de entradas, con el fin de proporcionar un error de aproximación predeterminado.

15 Con el fin de diseñar una tabla de consulta adecuada conviene tener en cuenta el error de aproximación de las mantisas. El error introducido por la cuantificación de la tabla de consulta no necesita ser considerablemente inferior al error de aproximación medio absoluto de las mantisas, que es 0,0573, dividido por 4. Esto da como resultado un error de cuantificación deseado inferior a 0,0143. La cuantificación lineal usando una tabla de consulta de 64 entradas da como resultado un error de cuantificación adecuado de $1/128 = 0,0078$. De este modo, la tabla de consulta predeterminada puede comprender un número total de 64 entradas. En general, el número de entradas en la tabla de consulta predeterminada debe alinearse con la aproximación seleccionada de la función logarítmica. En particular, la precisión de la cuantificación proporcionada por la tabla de consulta debe corresponder a la precisión de la aproximación de la función logarítmica.

20 Una evaluación perceptual del procedimiento de aproximación anterior indica que la calidad global de la señal de audio codificada se mejora cuando el error de estimación de los valores de tonalidad de celda está sesgada de manera positiva, es decir, cuando es más probable que la aproximación sobrestime el factor de ponderación (y los valores de tonalidad resultantes) que subestime el factor de ponderación.

25 Con el fin de conseguir tal sobrestimación, puede añadirse un sesgo a la tabla de consulta, por ejemplo puede añadirse un sesgo de la mitad de una etapa de cuantificación. Un sesgo de la mitad de una etapa de cuantificación puede implementarse truncando el índice en la tabla de consulta de cuantificación en lugar de redondear el índice. Puede ser beneficioso limitar el factor de ponderación a 0,5 para que se corresponda con la aproximación obtenida con el método babilónico/ de Herón.

30 La aproximación 503 del factor de ponderación w que se obtiene con la función de aproximación de dominio logarítmico se muestra en la Fig. 5a, junto con su error medio y máximo asociado. La Fig. 5a ilustra además el factor de ponderación exacto 501 usando la raíz cuarta y el factor de ponderación 502 determinado usando la aproximación babilónica. La calidad perceptiva de la aproximación de dominio logarítmico se ha verificado en una prueba de escucha usando el esquema de prueba MUSHRA. En la Fig. 5b puede observarse que la calidad percibida usando la aproximación logarítmica (barras situadas a la izquierda 511) es similar, por lo general, a la calidad percibida usando la aproximación babilónica (barras centrales 512) y la raíz cuarta (barras situadas a la derecha 513). Por otro lado, usando la aproximación logarítmica, la complejidad computacional del cálculo de tonalidad global puede reducirse en un 28%.

35 En el presente documento se han descrito varios esquemas para reducir la complejidad computacional de un codificador de audio basado en SPX. Los cálculos de tonalidad se han identificado como una contribución principal a la complejidad computacional del codificador basado en SPX. Los procedimientos descritos permiten reutilizar valores de tonalidad ya calculados, reduciéndose así la complejidad computacional global. La reutilización de valores de tonalidad ya calculados no afecta por lo general a la salida del codificador de audio basado en SPX. Además, se han descrito maneras alternativas para determinar el factor de mezcla de ruido b que permiten una reducción adicional de la complejidad computacional. Además, se ha descrito un esquema de aproximación eficiente para el factor de ponderación de tonalidad por celda, que puede usarse para reducir la complejidad del propio cálculo de tonalidad sin deteriorar la calidad de audio percibida. Como resultado de los esquemas descritos en el presente documento puede esperarse una reducción global de la complejidad computacional de un codificador de audio basado en SPX en un 50% y más, dependiendo de la configuración y de la velocidad binaria.

40 Los procedimientos y sistemas descritos en el presente documento pueden implementarse como software, firmware y/o hardware. Determinados componentes pueden implementarse, por ejemplo, como software que se ejecuta en un procesador o microprocesador de señales digitales. Otros componentes pueden implementarse, por ejemplo, como hardware y/o como circuitos integrados de aplicación específica. Las señales encontradas en los procedimientos y sistemas descritos pueden almacenarse en medios tales como una memoria de acceso aleatorio o medios de almacenamiento ópticos. Pueden transferirse a través de redes, tales como redes de radio, redes de satélites, redes

inalámbricas o redes cableadas, por ejemplo Internet. Los dispositivos típicos que usan los procedimientos y sistemas descritos en el presente documento son dispositivos electrónicos portátiles u otros equipos de usuario que se usan para almacenar y/o reproducir señales de audio.

- 5 Un experto en la técnica podrá aplicar fácilmente los diversos conceptos descritos anteriormente para diseñar realizaciones adicionales específicamente adaptadas a los requisitos de codificación de audio actuales.

REIVINDICACIONES

- 1.- Un procedimiento para determinar un primer valor de tonalidad por banda (311, 312) para una primera subbanda de frecuencia (205) de una señal de audio, donde el primer valor de tonalidad por banda (311, 312) se usa para aproximar una componente de alta frecuencia de la señal de audio en función de una componente de baja frecuencia de la señal de audio, comprendiendo el procedimiento:
- determinar un conjunto de coeficientes de transformada en un conjunto correspondiente de celdas de frecuencia en función de un bloque de muestras de la señal de audio;
 - determinar un conjunto de valores de tonalidad de celda (341) para el conjunto de celdas de frecuencia usando el conjunto de coeficientes de transformada, respectivamente; y
 - combinar un primer subconjunto de dos o más del conjunto de valores de tonalidad de celda (341) para dos o más celdas de frecuencia adyacentes correspondientes del conjunto de celdas de frecuencia que están dentro de la primera subbanda de frecuencia, obteniéndose así el primer valor de tonalidad por banda (311, 312) para la primera subbanda de frecuencia;
- en el que
- el procedimiento comprende además determinar una secuencia de conjuntos de coeficientes de transformada en función de una secuencia correspondiente de bloques de la señal de audio;
 - para una celda de frecuencia particular, la secuencia de conjuntos de coeficientes de transformada comprende una secuencia de coeficientes de transformada particulares;
 - determinar el valor de tonalidad de celda (341) para la celda de frecuencia particular comprende:
 - determinar una secuencia de fases en función de la secuencia de coeficientes de transformada particulares; y
 - determinar una aceleración de fase en función de la secuencia de fases; y
 - el valor de tonalidad de celda (341) para la celda de frecuencia particular depende de la aceleración de fase.
- 2.- El procedimiento según la reivindicación 1, que comprende además:
- determinar un segundo valor de tonalidad por banda (321, 322) en una segunda subbanda de frecuencia combinando un segundo subconjunto de dos o más del conjunto de valores de tonalidad de celda (341) para dos o más celdas de frecuencia adyacentes correspondientes del conjunto de celdas de frecuencia que están dentro de la segunda subbanda de frecuencia; donde la primera y la segunda subbanda de frecuencia comprenden al menos una celda de frecuencia común y donde el primer y el segundo subconjunto comprenden el al menos un valor de tonalidad de celda común (341) correspondiente.
- 3.- El procedimiento según la reivindicación 1, en el que:
- aproximar la componente de alta frecuencia de la señal de audio en función de la componente de baja frecuencia de la señal de audio comprende copiar uno o más coeficientes de transformada de baja frecuencia de una o más celdas de frecuencia desde una banda de baja frecuencia (101), correspondiente a la componente de baja frecuencia, a una banda de alta frecuencia (102), correspondiente a la componente de alta frecuencia;
 - la primera subbanda de frecuencia está dentro de la banda de baja frecuencia (101);
 - una segunda subbanda de frecuencia está dentro de la banda de alta frecuencia (102);
 - el procedimiento comprende además determinar un segundo valor de tonalidad por banda (233) en la segunda subbanda de frecuencia combinando un segundo subconjunto de dos o más del conjunto de valores de tonalidad de celda (341) para dos o más celdas de frecuencia correspondientes de las celdas de frecuencia que se han copiado en la segunda subbanda de frecuencia;
 - la segunda subbanda de frecuencia comprende al menos una celda de frecuencia que se ha copiado desde una celda de frecuencia que está dentro de la primera subbanda de frecuencia; y
 - el primer y el segundo subconjunto comprenden el al menos un valor de tonalidad de celda común (341) correspondiente.
- 4.- El procedimiento según cualquier reivindicación anterior, en el que combinar el primer subconjunto de dos o más del conjunto de valores de tonalidad de celda (341) comprende:
- calcular el promedio de los dos o más valores de tonalidad de celda (341); o
 - sumar los dos o más valores de tonalidad de celda (341).
- 5.- El procedimiento según cualquier reivindicación anterior, en el que un valor de tonalidad de celda (341) para una celda de frecuencia se determina solamente en función de los coeficientes de transformada de la misma celda de frecuencia.
- 6.- El procedimiento según cualquier reivindicación anterior, en el que:

- el primer valor de tonalidad por banda (311, 312) se usa para aproximar una componente de alta frecuencia de la señal de audio en función de una componente de baja frecuencia de la señal de audio usando un esquema de extensión espectral, denominado SPX; y
- el primer valor de tonalidad por banda (311, 312) se usa para determinar una estrategia de reenvío de coordenadas SPX, un factor de mezcla de ruido y/o una atenuación de varianza elevada.

5

7.- Un sistema configurado para determinar un primer valor de tonalidad por banda (311, 312) para una primera subbanda de frecuencia (205) de una señal de audio, donde el primer valor de tonalidad por banda (311, 312) se usa para aproximar una componente de alta frecuencia de la señal de audio en función de una componente de baja frecuencia de la señal de audio, donde el sistema está configurado para:

10

- determinar un conjunto de coeficientes de transformada en un conjunto correspondiente de celdas de frecuencia en función de un bloque de muestras de la señal de audio;
- determinar un conjunto de valores de tonalidad de celda (341) para el conjunto de celdas de frecuencia usando el conjunto de coeficientes de transformada, respectivamente; y
- combinar un primer subconjunto de dos o más del conjunto de valores de tonalidad de celda (341) para dos o más celdas de frecuencia adyacentes correspondientes del conjunto de celdas de frecuencia que están dentro de la primera subbanda de frecuencia, obteniéndose así el primer valor de tonalidad por banda (311, 312) para la primera subbanda de frecuencia;

15

20 en el que

- el sistema está configurado además para determinar una secuencia de conjuntos de coeficientes de transformada en función de una secuencia correspondiente de bloques de la señal de audio;
- para una celda de frecuencia particular, la secuencia de conjuntos de coeficientes de transformada comprende una secuencia de coeficientes de transformada particulares;
- determinar el valor de tonalidad de celda (341) para la celda de frecuencia particular comprende:
 - determinar una secuencia de fases en función de la secuencia de coeficientes de transformada particulares; y
 - determinar una aceleración de fase en función de la secuencia de fases; y
- el valor de tonalidad de celda (341) para la celda de frecuencia particular depende de la aceleración de fase.

25

30

8.- Un codificador de audio configurado para codificar una señal de audio usando reconstrucción de alta frecuencia, comprendiendo el codificador de audio el sistema de la reivindicación 7.

35

9.- Un programa de software adaptado para ejecutarse en un procesador y para llevar a cabo las etapas de procedimiento de cualquiera de las reivindicaciones 1 a 6 cuando se ejecuta en el procesador.

10.- Un medio de almacenamiento que comprende un programa de software adaptado para ejecutarse en un procesador y para llevar a cabo las etapas de procedimiento de cualquiera de las reivindicaciones 1 a 6 cuando se ejecuta en un dispositivo informático.

40

11.- Un producto de programa informático que comprende instrucciones ejecutables para llevar a cabo las etapas de procedimiento de cualquiera de las reivindicaciones 1 a 6 cuando se ejecutan en un ordenador.

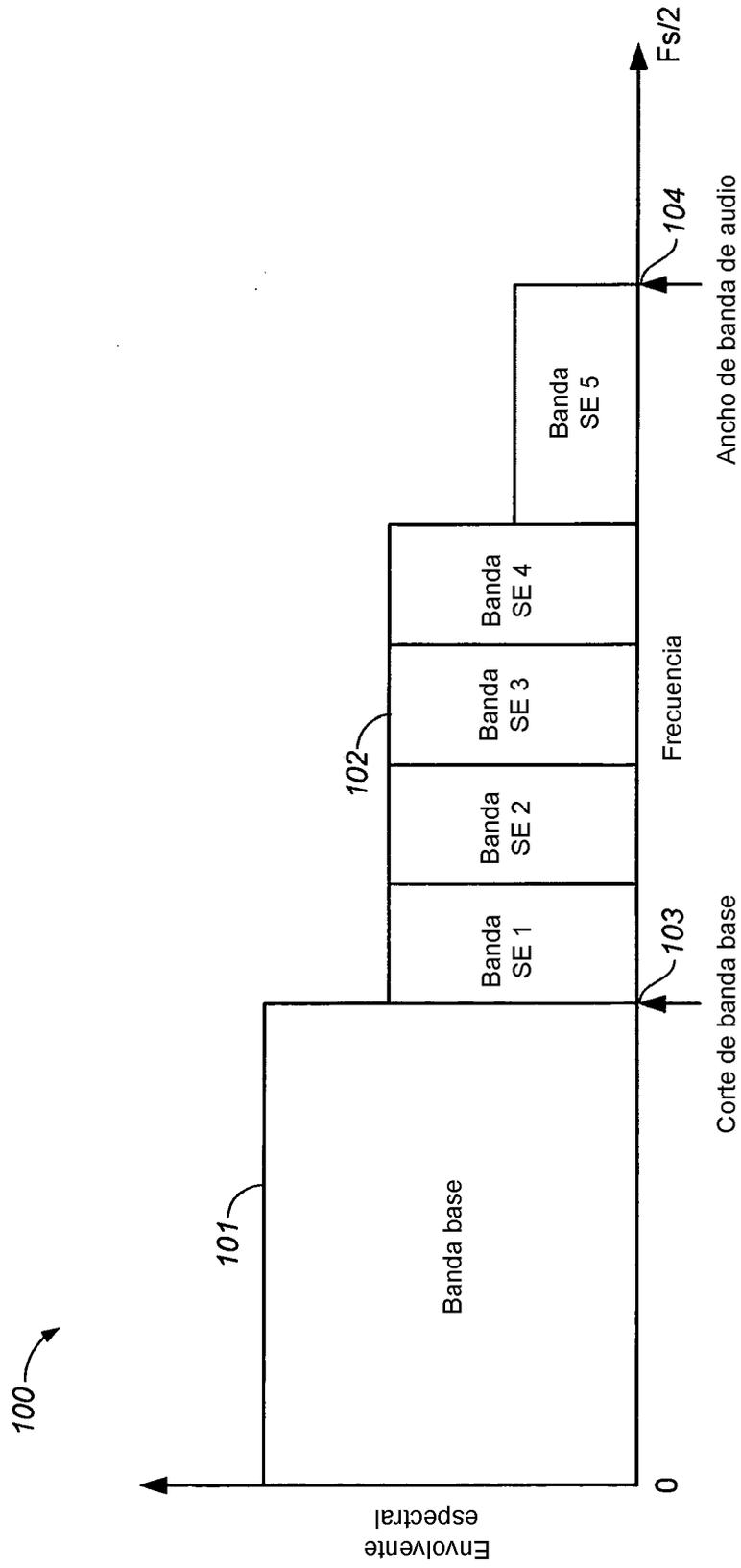


FIG. 1A

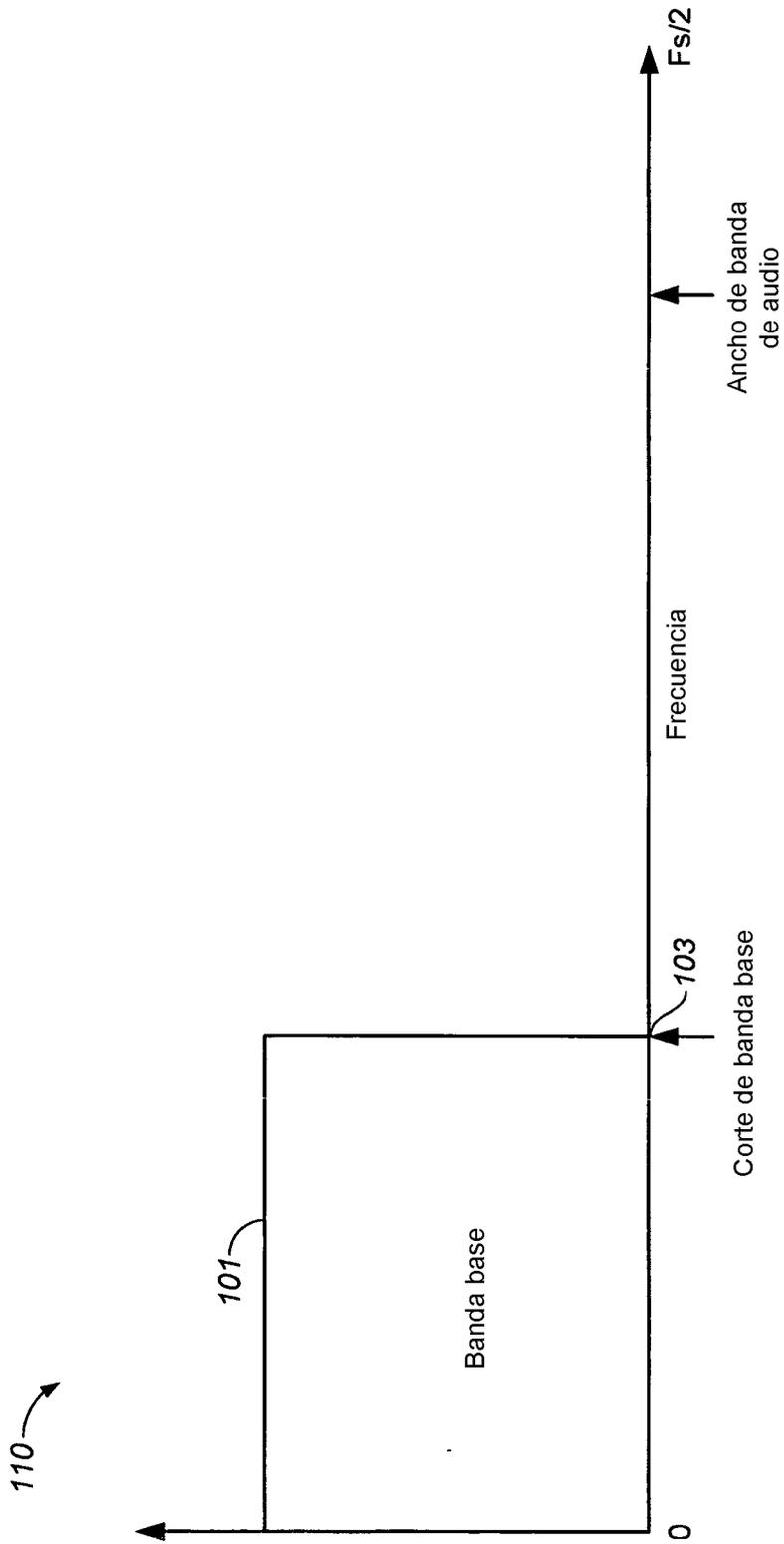


FIG. 1B

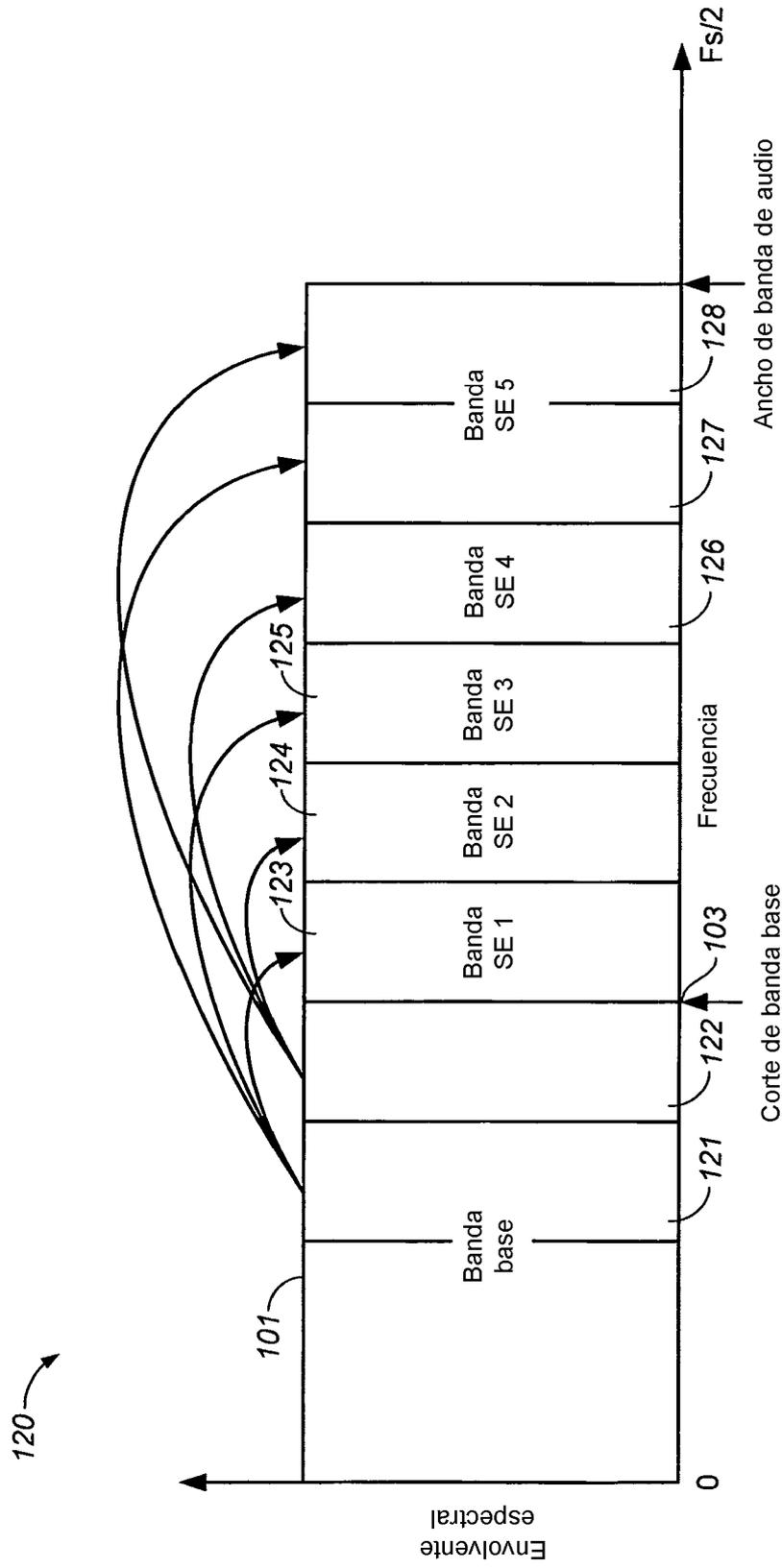


FIG. 1C

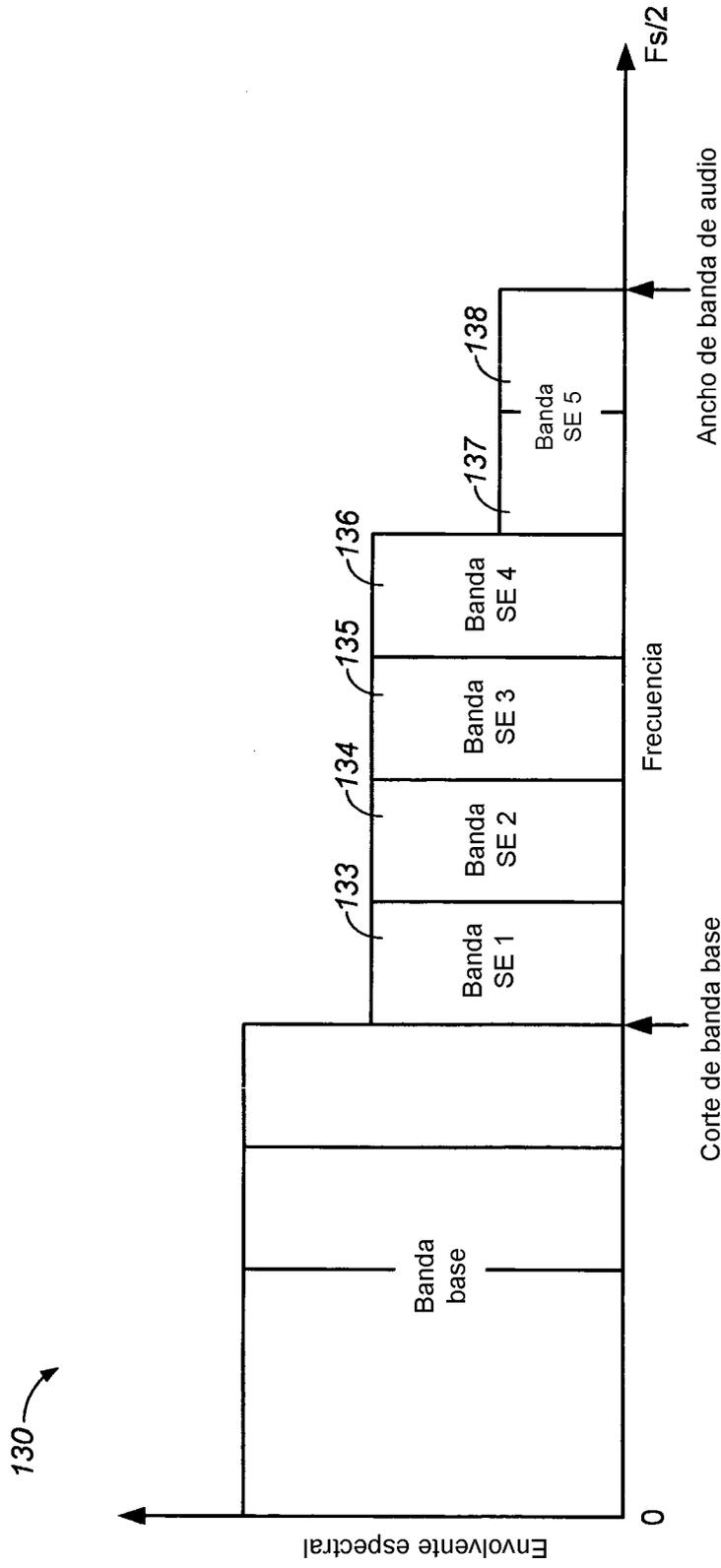


FIG. 1D

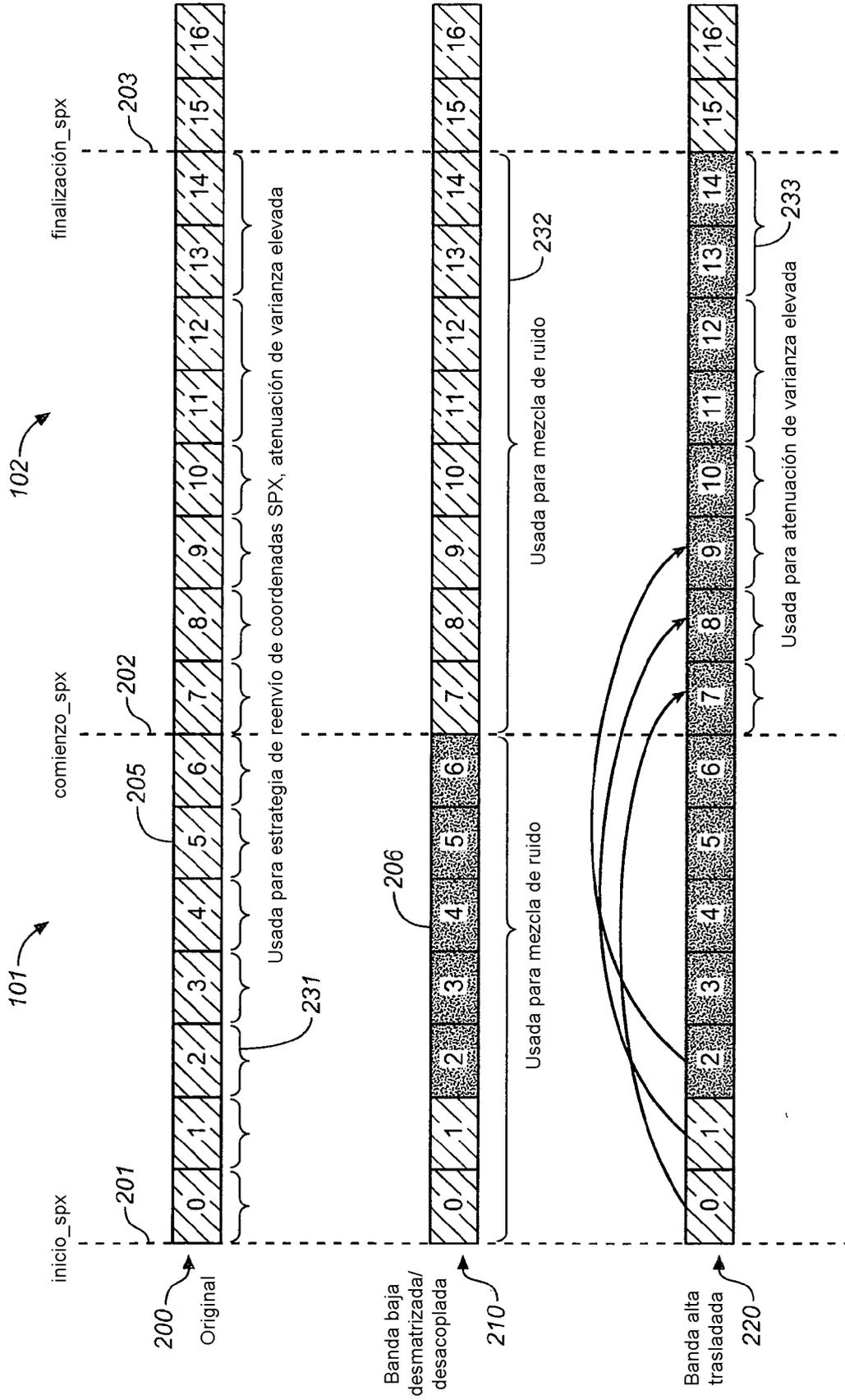


FIG. 2A

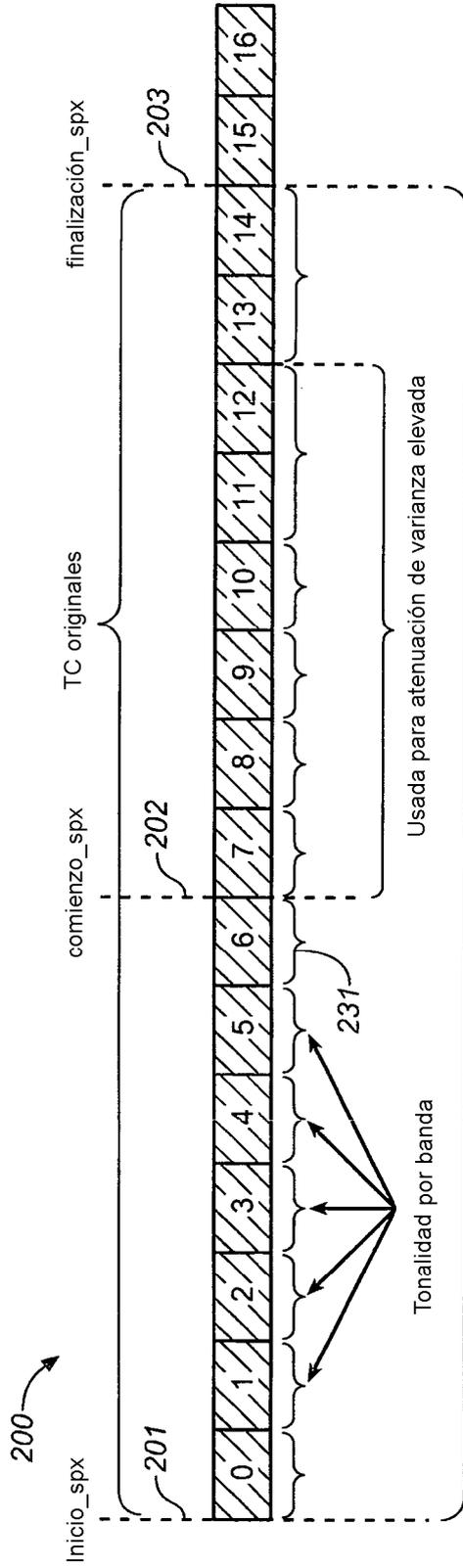


FIG. 2B

Usada para estrategia de reenvío de coordenadas SPX



FIG. 2C

Usada para mezcla de ruido

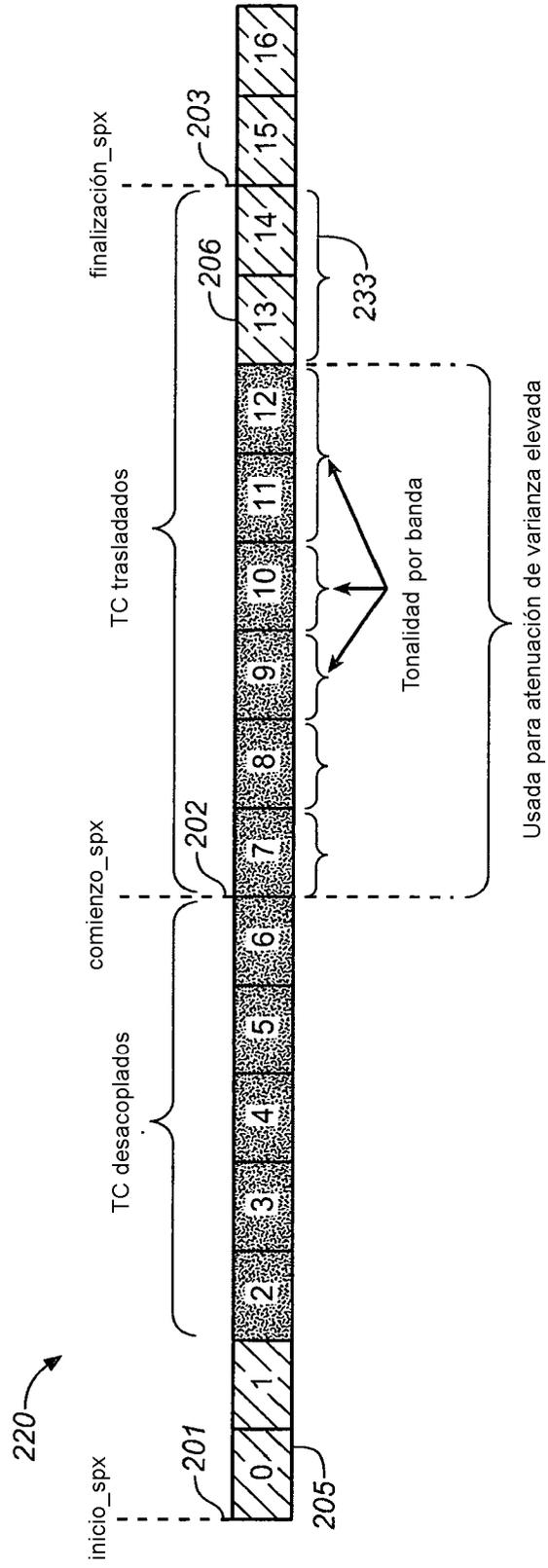


FIG. 2D

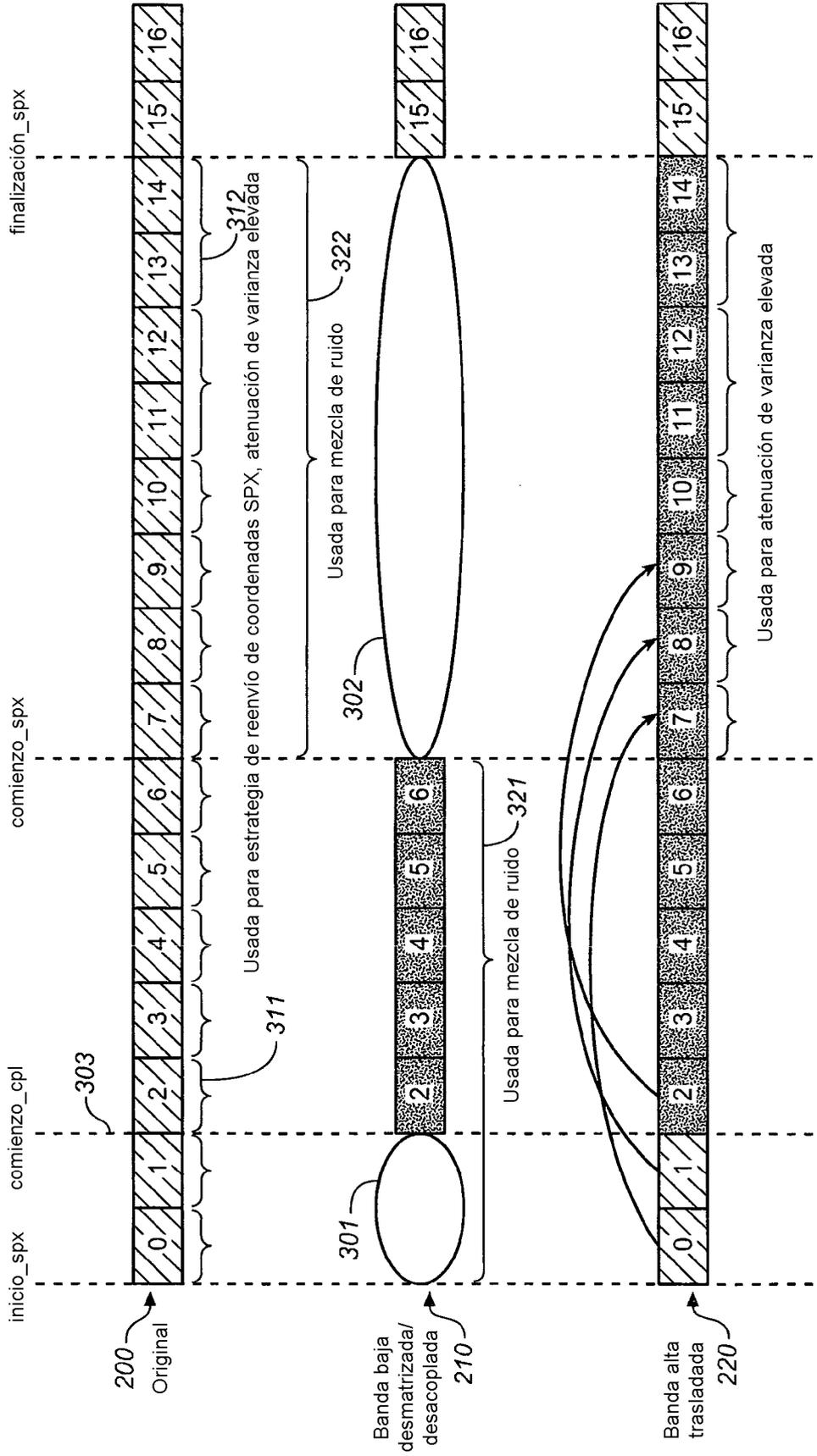


FIG. 3A

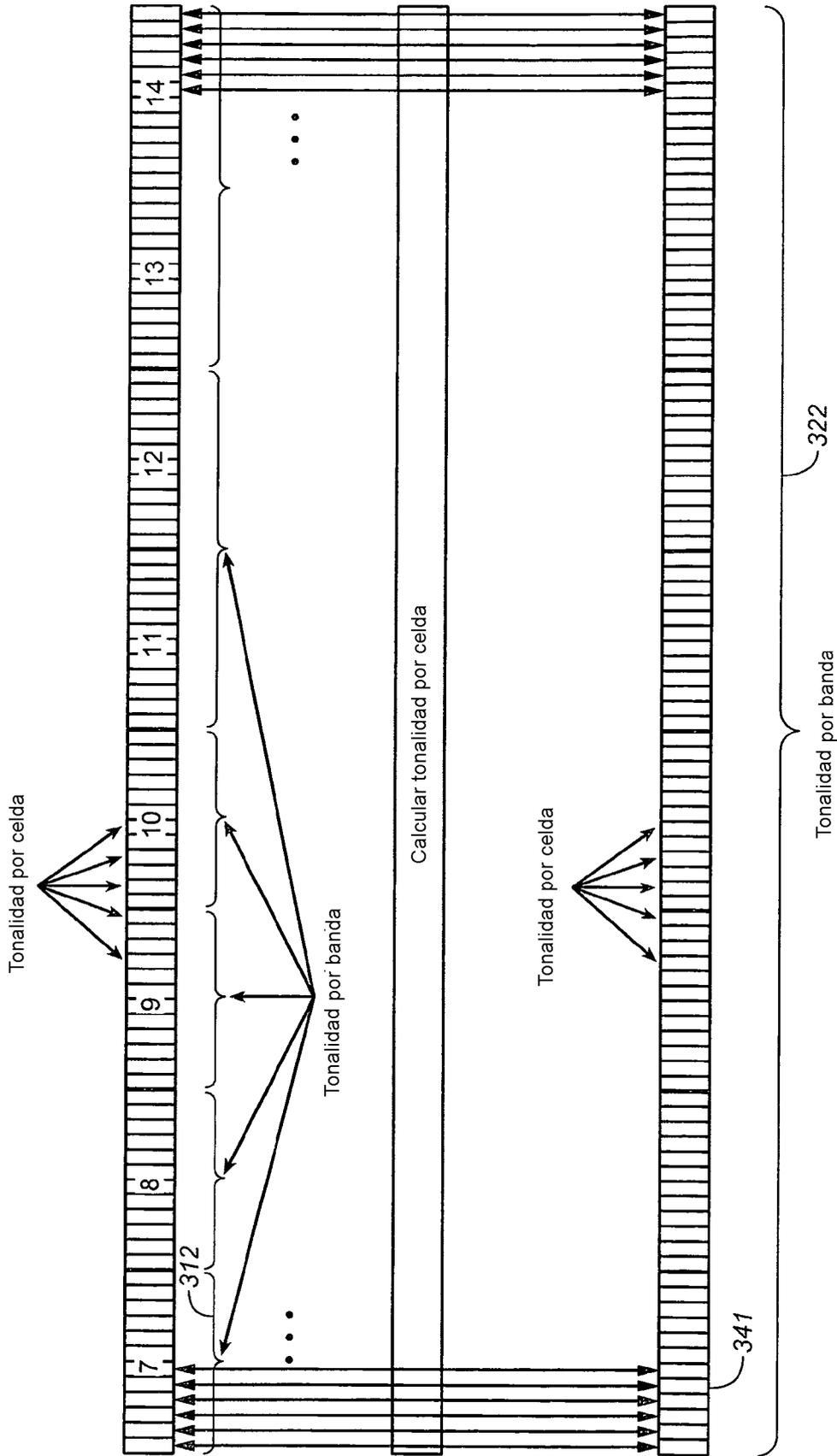


FIG. 3B

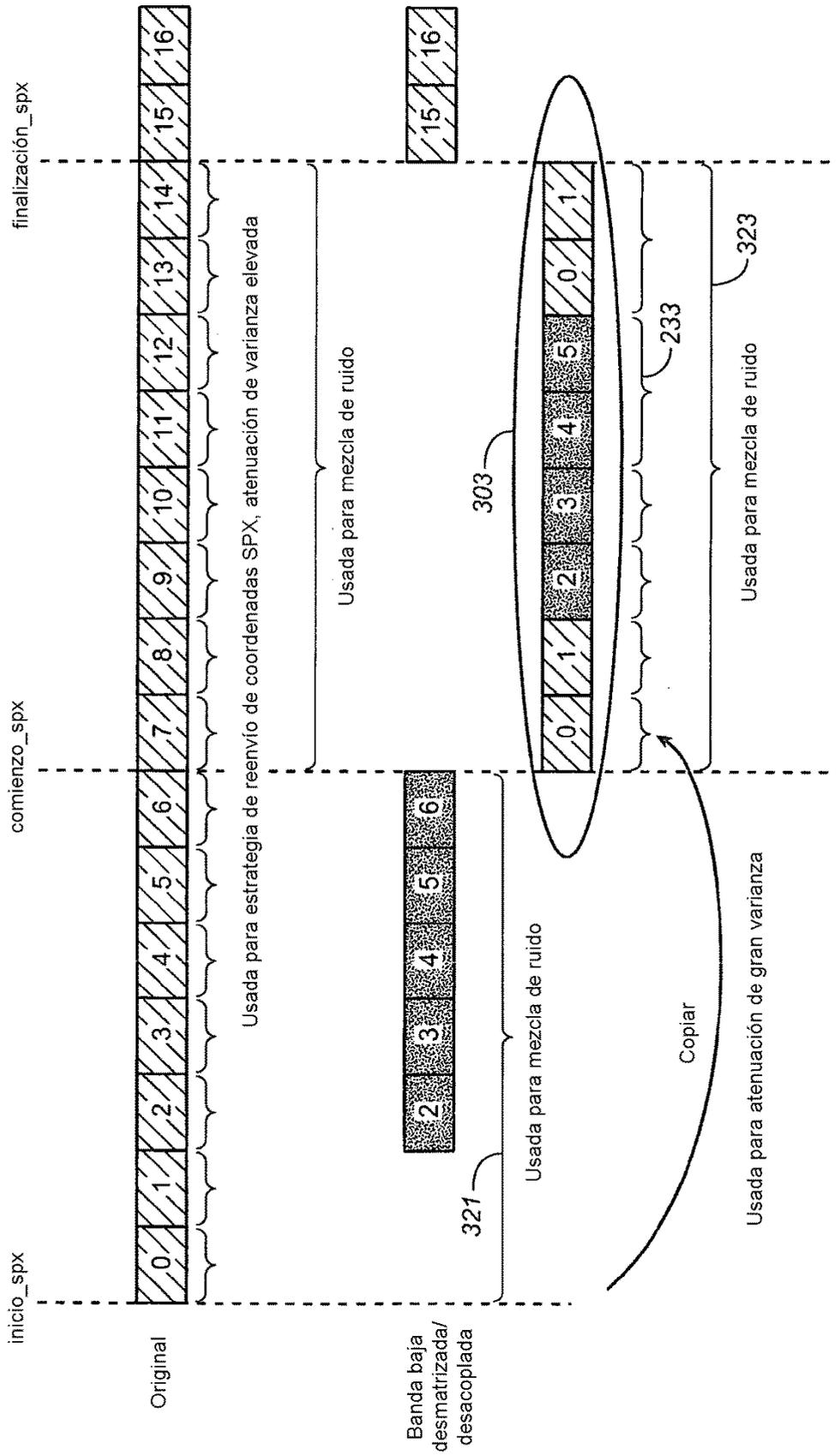


FIG. 3C

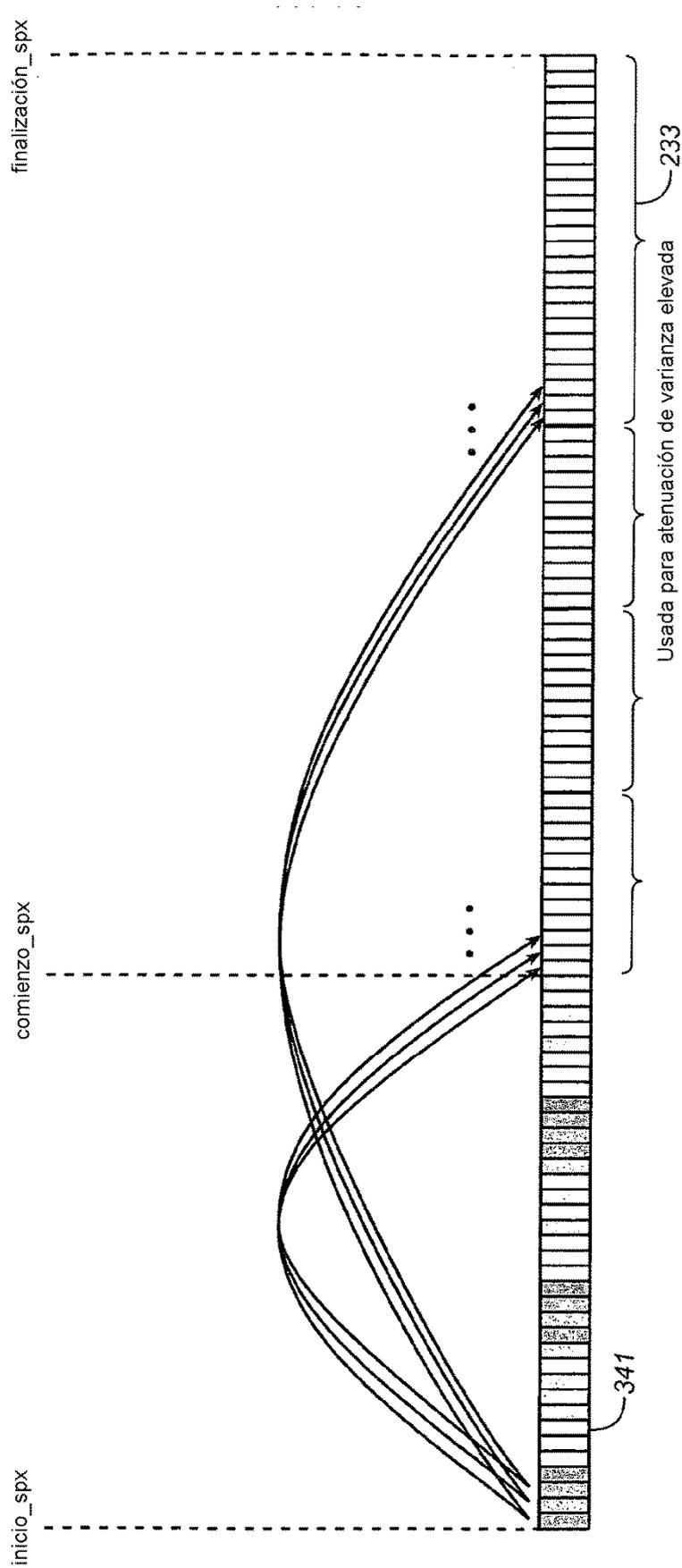


FIG. 3D

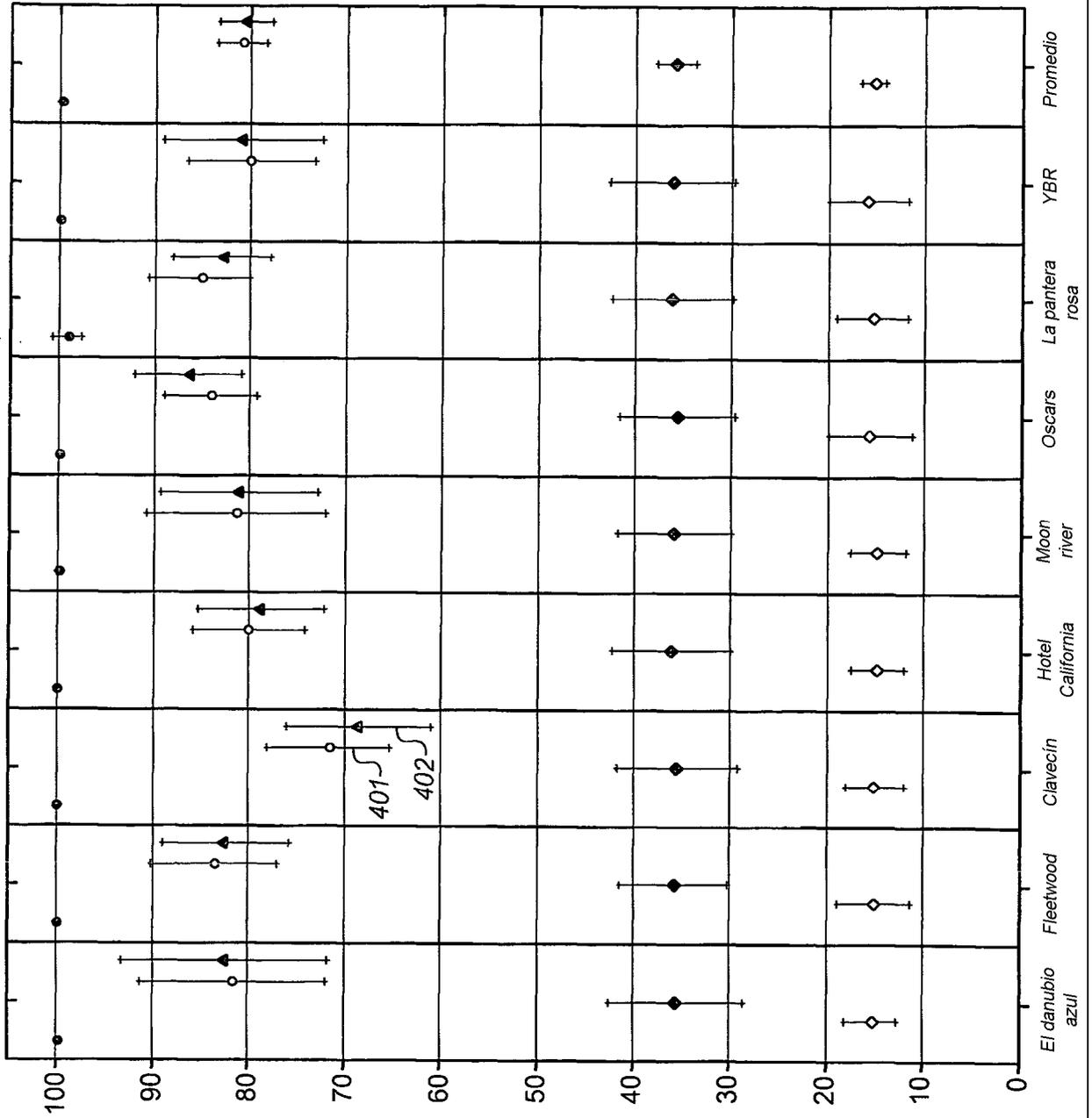


FIG. 4

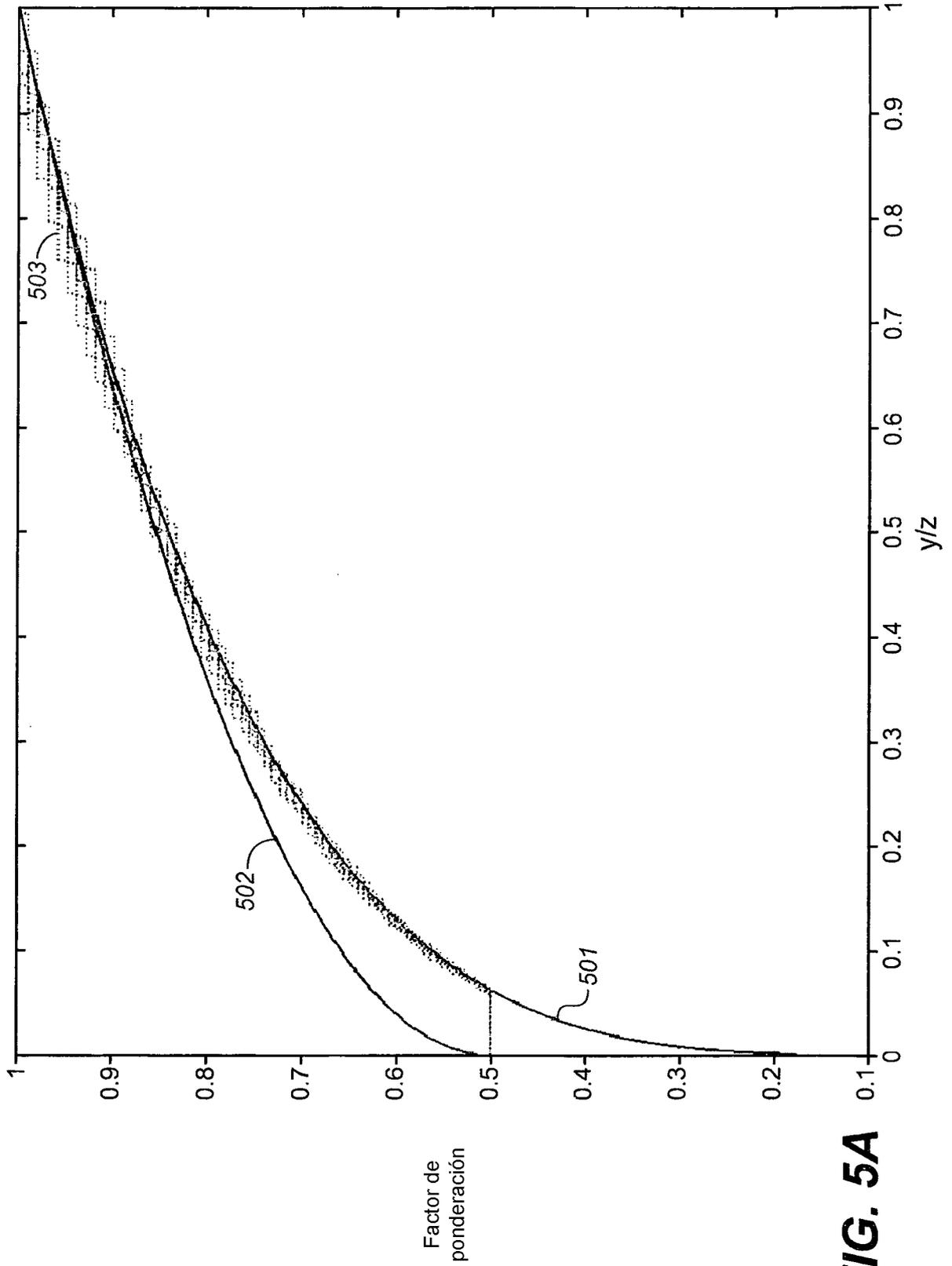


FIG. 5A

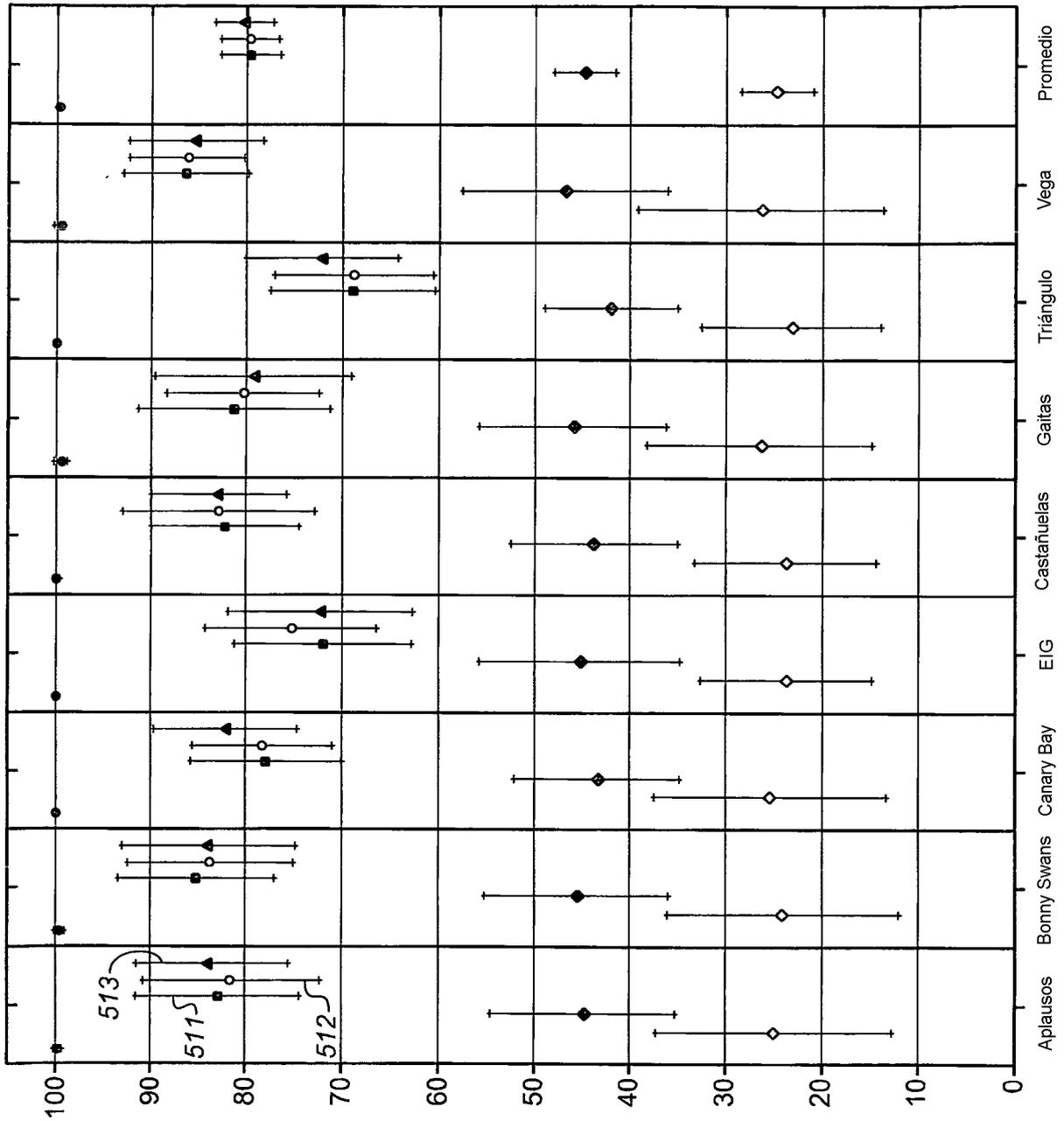


FIG. 5B