

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **3 012 833**

51 Int. Cl.:

G10L 19/02 (2013.01)

G10L 19/022 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **10.06.2016** **E 24165637 (0)**

97 Fecha y número de publicación de la concesión europea: **29.01.2025** **EP 4386745**

54 Título: **Decodificación con reducción de escala**

30 Prioridad:

16.06.2015 EP 15172282

12.10.2015 EP 15189398

45 Fecha de publicación y mención en BOPI de la
traducción de la patente:
10.04.2025

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR
FÖRDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (100.00%)**

**Hansastraße 27c
80686 München, DE**

72 Inventor/es:

**SCHNELL, MARKUS;
LUTZKY, MANFRED;
FOTOPOULOU, ELENI;
SCHMIDT, KONSTANTIN;
BENNDORF, CONRAD;
TOMASEK, ADRIAN;
ALBERT, TOBIAS y
SEIDL, TIMON**

74 Agente/Representante:

PONTI & PARTNERS, S.L.P.

ES 3 012 833 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

Decodificación con reducción de escala

5 [0001] La presente solicitud se refiere a un concepto de decodificación con reducción de escala.

[0002] El MPEG-4 AAC Mejorado de Bajo Retardo (AAC-ELD) normalmente opera a tasas de muestreo de hasta 48 kHz, lo que da lugar a un retardo algorítmico de 15 ms. Para algunas aplicaciones, p. ej., transmisión de audio de la sincronización de labios, es deseable un retardo aún más bajo. AAC-ELD ya proporciona una opción tal
10 por medio de la operación a tasas de muestreo más altas, p. ej., 96 kHz, y por lo tanto proporciona modos de funcionamiento en un plazo aún menor, p. ej., 7,5 ms. Sin embargo, este modo de operación viene junto con una alta complejidad innecesaria debido a la alta tasa de muestreo.

[0003] La solución a este problema es aplicar una versión con reducción de escala del banco de filtros y, por lo tanto, renderizar la señal de audio a una tasa de muestreo inferior, p. ej., 48 kHz en lugar de 96 kHz. La operación
15 con reducción de escala ya es parte de AAC-ELD ya que se hereda del códec MPEG-4 AAC-LD, que sirve como base para AAC-ELD.

[0004] Sin embargo, la pregunta que permanece es cómo encontrar la versión con reducción de escala de un banco de filtros específico. Es decir, la única incertidumbre es la forma en que los coeficientes de ventana se derivan
20 al tiempo que permite pruebas de conformidad claras de los modos de operación con reducción de escala del decodificador AAC-ELD.

[0005] En lo sucesivo, se describen los principios del modo de operación con reducción de escala de los códec
25 AAC-(E)LD.

[0006] El modo de operación con reducción de escala o AAC-LD se describe para AAC-LD en la norma ISO/IEC 14496-3:2009 en la sección 4.6.17.2.7 "Adaptation to systems using lower sampling rates" de la siguiente manera:

30 *"En ciertas aplicaciones puede ser necesario integrar el decodificador de bajo retardo en un sistema de audio que se ejecuta a velocidades de muestreo más bajas (p. ej., 16 kHz), mientras que la tasa de muestreo nominal de la carga útil de corriente de bits es mucho más alta (p. ej., 48 kHz, correspondiente a un retardo de códec algorítmico de aprox. 20 ms). En tales casos, es favorable decodificar la salida del códec de bajo retardo directamente en la tasa de muestreo de destino en lugar de utilizar una operación de conversión de tasa de muestreo adicional después de la decodificación."*
35

[0007] *Esto se puede aproximar por medio de la reducción de escala apropiada de ambos, el tamaño de la trama y la tasa de muestreo, por algún factor entero (p. ej., 2, 3), lo que da lugar a la misma resolución de tiempo/frecuencia del códec. Por ejemplo, la salida de códec se puede generar a 16 kHz de tasa de muestreo en lugar de la nominal 48 kHz por medio de la retención de sólo el tercer más bajo (es decir, $480/3 = 160$) de los coeficientes espectrales antes de la síntesis del banco de filtros y la reducción del tamaño de la transformada inversa a un tercio (es decir, tamaño de la ventana $960/3 = 320$).*
40

[0008] *Como consecuencia, la decodificación de una tasa de muestreo más baja reduce tanto la memoria y los requisitos computacionales, pero puede no producir exactamente el mismo resultado que una decodificación de ancho de banda completo, seguido por limitación de banda y la conversión de tasa de muestreo.*
45

[0009] *Se debe tener en cuenta que la decodificación a una menor tasa de muestreo, según se ha descrito con anterioridad, no afecta a la interpretación de los niveles, que se refiere a la tasa de muestreo nominal de la carga útil de corriente de bits de bajo retardo de AAC".*
50

[0010] Se debe tener en cuenta que AAC-LD trabaja con un marco estándar de MDCT y dos formas de ventana, es decir, ventana senoidal y ventana bajo solapamiento. Ambas ventanas están totalmente descritas por las fórmulas y, por lo tanto, se pueden determinar los coeficientes de ventana para cualquier longitud de transformación.
55

[0011] En comparación con AAC-LD, el códec de AAC-ELD muestra dos diferencias principales:

- La ventana de MDCT de bajo retardo (LD-MDCT)
- La posibilidad de utilizar la herramienta de SBR de Bajo Retardo

60 [0012] El algoritmo de IMDCT por el uso de la ventana de MDCT de bajo retardo se describe en 4.6.20.2 en [1], que es muy similar a la versión estándar por el uso de, por ejemplo, IMDCT la ventana senoidal. Los coeficientes de ventana de MDCT de bajo retardo (tamaño de trama de muestras 480 y 512) se presentan en la Tabla 4.A.15 4. A.16 y en [1]. Se debe tener en cuenta que los coeficientes no se pueden determinar por medio de una fórmula, dado
65 que los coeficientes son el resultado de un algoritmo de optimización. La Fig. 9 muestra un gráfico de la forma de la

ventana para el tamaño de trama 512.

[0013] En caso de que la herramienta de SBR de bajo retardo (LD-SBR) se utilice en conjunción con el codificador AAC-ELD, los bancos de filtros del módulo de LD-SBR también tienen reducción de escala. Esto asegura que el módulo de SBR opera con la misma resolución de frecuencia y, por lo tanto, no se requieren más adaptaciones.

[0014] Por lo tanto, la descripción anterior pone de manifiesto que existe una necesidad con reducción de escala de operaciones de decodificación tales como, por ejemplo, reducción de escala de una decodificación en un AAC-ELD. Sería factible encontrar los coeficientes de la función de ventana de síntesis con reducción de escala de nuevo, pero esta es una tarea compleja, que hace necesario el almacenamiento adicional para almacenar la versión con reducción de escala y hacer una verificación de la conformidad entre la decodificación sin reducción de escala y la decodificación con reducción de escala más complicada o, desde otra perspectiva, no se ajusta a la forma con reducción de escala deseada en el AAC-ELD, por ejemplo. Dependiendo de la proporción de escala descendente, es decir, la relación entre la tasa de muestreo original y la tasa de muestreo con reducción de escala, se podría derivar la función de ventana de síntesis con reducción de escala simplemente por medio de submuestreo, es decir, eligiendo cada segundo, tercer, ... coeficiente de ventana de la función de ventana de síntesis original, pero este procedimiento no da como resultado una conformidad suficiente de la decodificación sin reducción de escala y decodificación con reducción de escala, respectivamente. El uso de procedimientos diezmados más sofisticados aplicados a la función de ventana de síntesis, conduce a desviaciones inaceptables de la forma original de la función de ventana de síntesis. Por lo tanto, hay una necesidad en la técnica de un concepto de decodificación con reducción de escala mejorada.

[0015] La publicación científica de JUIN-HWEY CHEN: "A high-fidelity speech and audio codec with low delay and low complexity", Proceedings of 20000 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Estambul, Turquía; 5-9 de junio de 2000, páginas ||1161.||1164, describe la decodificación de una señal de corriente de bits de audio de alta fidelidad directamente en una señal de audio de bajo muestreo sin decodificar primero y a continuación submuestrear a la tasa de muestreo de destino.

[0016] Un objeto de la presente invención es proporcionar un esquema de decodificación de audio que permita una decodificación con reducción de escala mejorada.

[0017] Este objeto se consigue por medio de la materia objetivo de las reivindicaciones independientes.

[0018] La presente invención se basa en el hallazgo de que una versión con reducción de escala de un procedimiento de decodificación de audio se puede lograr de manera más eficaz y/o con un mantenimiento de cumplimiento mejorado si la ventana de síntesis que se utiliza para la decodificación de audio con reducción de escala es una versión submuestreada de una ventana de síntesis de referencia que está implicada en el procedimiento de decodificación de audio sin reducción de escala por medio del submuestreo por el factor de submuestreo por el cual la tasa de muestreo submuestreada y la tasa de muestreo original se desvían, y se submuestrean por el uso de una interpolación segmentaria en los segmentos de 1/4 de la longitud de la trama. Los aspectos ventajosos de la presente solicitud son el objetivo de las reivindicaciones dependientes. Las realizaciones preferidas de la presente solicitud se describen a continuación con respecto a las figuras, entre las cuales:

La Fig. 1 muestra un diagrama esquemático que ilustra las necesidades de reconstrucción perfectas necesarias para ser obedecidas cuando se decodifica con reducción de escala con el fin de preservar una reconstrucción perfecta;

La Fig. 2 muestra un diagrama de bloques de un decodificador de audio para decodificación con reducción de escala según una realización;

La Fig. 3 muestra un diagrama esquemático que ilustra en la mitad superior la forma donde una señal de audio se ha codificado a una tasa de muestreo original, en una corriente de datos y, en la mitad inferior separada de la mitad superior por una línea horizontal de trazos, una operación de decodificación con reducción de escala para la reconstrucción de la señal de audio de la corriente de datos a una tasa de muestreo reducida o con reducción de escala, con el fin de ilustrar el modo de operación del decodificador de audio de la Fig. 2;

La Fig. 4 muestra un diagrama esquemático que ilustra la cooperación del formador de ventanas y cancelador de solapamiento de dominio de tiempo de la Fig. 2;

La Fig. 5 ilustra una posible implementación para lograr la reconstrucción según la Fig. 4 por el uso de un tratamiento especial de las porciones pesadas en cero de las porciones de tiempo moduladas espectrales a tiempo;

La Fig. 6 muestra un diagrama esquemático que ilustra el submuestreo para obtener la ventana de síntesis submuestreada;

La Fig. 7 muestra un diagrama de bloques que ilustra una operación con reducción de escala de AAC-ELD que incluye la herramienta de SBR de bajo retardo;

La Fig. 8 muestra un diagrama de bloques de un decodificador de audio para la decodificación con reducción de escala según una realización donde el modulador, el formador de ventanas y el cancelador se implementan según una implementación de elevación; y

La Fig. 9 muestra un gráfico de los coeficientes de ventana de una ventana de bajo retardo según AAC-ELD de tamaño de trama de la muestra 512 como un ejemplo de una ventana de síntesis de referencia que se va a submuestrear.

[0019] La siguiente descripción comienza con una ilustración de una realización para la decodificación con reducción de escala con respecto al códec AAC-ELD. Es decir, la siguiente descripción comienza con una realización, que podría formar un modo con reducción de escala para AAC-ELD. Esta descripción forma al mismo tiempo una especie de explicación de la motivación subyacente a las realizaciones de la presente solicitud. Más tarde, esta descripción se generaliza, lo que conduce a una descripción de un decodificador de audio y el procedimiento de decodificación de audio según una realización de la presente solicitud.

[0020] Según se ha descrito en la porción introductoria de la memoria descriptiva de la presente solicitud, AAC-ELD utiliza ventanas de MDCT de bajo retardo. Para generar versiones con reducción de escala de la misma, es decir, ventanas de bajo retardo con reducción de escala, la propuesta explicada posteriormente para formar un modo con reducción de escala para AAC-ELD utiliza un algoritmo de interpolación segmentaria de spline que mantiene la propiedad de reconstrucción perfecta (PR, por su sigla en inglés) de la ventana de LD-MDCT con una precisión muy alta. Por lo tanto, el algoritmo permite la generación de coeficientes de ventana en forma directa, según se ha descrito en ISO/IEC 14496-3: 2009, así como también en la forma de elevación, según se ha descrito en [2], de una manera compatible. Esto significa que ambas implementaciones generan una salida conforme de 16 bits.

[0021] La interpolación de la ventana de MDCT de Bajo Retardo se lleva a cabo según lo presentado a continuación.

[0022] En general una interpolación spline se va a utilizar para la generación de los coeficientes de ventana con reducción de escala para mantener la respuesta de frecuencia y en su mayoría la propiedad de reconstrucción perfecta (de aproximadamente 170dB SNR). La interpolación tiene que ser restringida en ciertos segmentos para mantener la propiedad de reconstrucción perfecta. Para los coeficientes de ventana c que cubren el núcleo de la transformación DCT (véase también la Figura 1, $c(1024) \dots c(2048)$), se requiere la siguiente restricción,

$$1 = [(\text{sgn} \cdot c(i) \cdot c(2N - 1 - i) + c(N + i) \cdot c(N - 1 - i))] \text{ para } i = 0 \dots N/2 - 1 \quad (1)$$

donde N denota el tamaño de la trama. Algunas implementaciones pueden utilizar diferentes señales para optimizar la complejidad, en este documento, denotado por sgn . El requisito en (1) se puede ilustrar por medio de la Fig. 1. Se debe recordar que, simplemente, en incluso en caso de $M = 2$, es decir la mitad de la tasa de muestreo, que deja fuera cada segundo coeficiente de ventana de la ventana de síntesis de referencia para obtener la ventana de síntesis con reducción de escala no cumple el requisito.

[0023] Los coeficientes $c(0) \dots c(2N - 1)$ se enumeran a lo largo de la forma de diamante. Los $N/4$ ceros en los coeficientes de ventana, que son responsables de la reducción del retardo del banco de filtros, se marcan por medio de una flecha en negrita. La Fig. 1 muestra las dependencias de los coeficientes provocadas por el plegamiento involucrado en la MDCT y también los puntos donde la interpolación tiene que ser restringida con el fin de evitar cualquier dependencia no deseada.

- Cada $N/2$ coeficientes, la interpolación tiene que parar para mantener (1)
- Además, el algoritmo de interpolación tiene que parar cada $N/4$ coeficientes debido a los ceros insertados. Esto asegura que los ceros se mantengan y que no se transmita el error de interpolación que mantiene la PR.

[0024] La segunda restricción no sólo es necesaria para el segmento que contiene los ceros sino también para los otros segmentos. Sabiendo que algunos coeficientes DCT en el núcleo no se determinaron por medio del algoritmo de optimización, sino que se determinaron por medio de la fórmula (1) para permitir PR, se pueden explicar varias discontinuidades en la forma de la ventana, por ejemplo, de aproximadamente $c(1536+128)$ en la Figura 1. Con el fin de minimizar el error de PR, la interpolación tiene que parar en tales puntos, que aparecen en una rejilla de $N/4$.

[0025] Debido a esta razón, el tamaño de segmento de $N/4$ se elige para la interpolación segmentaria spline para generar los coeficientes de ventana con reducción de escala. Los coeficientes de ventana de origen siempre son dados por los coeficientes utilizados para $N = 512$, también para las operaciones con reducción de escala que da lugar a tamaños de trama de $N = 240$ o $N = 120$. El algoritmo básico se describe muy brevemente en lo sucesivo como el código de MATLAB:

```

FAC = factor de reducción de escala      % p. ej. 0,5
sb = 128;                                % de tamaño de segmento de ventana de origen

w_down = [];                             % de ventana con reducción de escala
nSegments = longitud (W)/(SB);           % del número de segmentos; W = LD
                                           % coeficientes de ventana para N = 512

xn = ((0: (FAC*sb-1))/0,5)/FAC-0,5;      % de spline inicial
para i = 1: nSegments,
    w_down = [w_down, spline ({0: (sb-1)}, W((i-1)*sb+1:(sb)}), xn)];
fin;

```

[0026] Dado que la función de spline puede no ser determinista por completo, el algoritmo completo se especifica exactamente en la sección siguiente, que se puede incluir en la norma ISO/IEC 14496-3: 2009, con el fin de
5 formar un modo con reducción de escala mejorado en AAC-ELD.

[0027] En otras palabras, en la siguiente sección se ofrece una propuesta de cómo la idea descrita con anterioridad se podría aplicar a ER AAC ELD, es decir, en cuanto a cómo un decodificador poco complejo podría decodificar una corriente de bits ER AAC ELD codificada a una primera tasa de datos a una segunda tasa de datos más baja que
10 la primera tasa de datos. Se enfatiza, sin embargo, que la definición de N según lo utilizado en lo sucesivo se adhiere a la norma. En este caso, N corresponde a la longitud del núcleo de DCT, mientras que, en lo anterior, en las reivindicaciones, y las realizaciones generalizadas posteriormente descritas, N corresponde a la longitud de la trama, es decir, la longitud de solapamiento mutuo de los núcleos de DCT, es decir, la media de la longitud del núcleo de DCT. En consecuencia, mientras que N se indicó que era 512 en lo anterior, por ejemplo, se indica para que sea 1024 en lo
15 sucesivo.

[0028] Se proponen los siguientes párrafos para su inclusión a 14496-3: 2009 a través de la enmienda. Las referencias numeradas a continuación se refieren a ese documento [1].

20 **A.0 Adaptación a los sistemas que utilizan tasas de muestreo más bajas**

[0029] Para ciertas aplicaciones, ER AAC LD puede cambiar la tasa de muestreo de reproducción con el fin de evitar medidas de remuestreo adicionales (véase 4.6.17.2.7). ER AAC ELD puede aplicar medidas con reducción de escala similares por el uso de la ventana de MDCT de bajo retardo y la herramienta LD-SBR. En el caso de que AAC-
25 ELD funcione con la herramienta LD-SBR, el factor con reducción de escala se limita a múltiplos de 2. Sin LD-SBR, el tamaño de la trama con reducción de escala tiene que ser un número entero.

A.1 Reducción de escala de la ventana de MDCT de Bajo Retardo

[0030] La ventana LD-MDCT w_{LD} para $N = 1024$ se reduce proporcionalmente un factor F por el uso de una interpolación segmentaria spline. El número de ceros a la izquierda en los coeficientes de ventana, es decir, $N/8$, determina el tamaño del segmento. Los coeficientes de ventana con reducción de escala $w_{LD,d}$ se utilizan para la MDCT inversa según se ha descrito en 4.6.20.2 pero con una longitud de ventana con reducción de escala $N_d = N/F$. Se debe tener en cuenta que el algoritmo también es capaz de generar coeficientes de elevación con reducción de
35 escala de la LD-MDCT.

```

fs_window_size = 2048; /* Número de coeficientes de ventana con fondo de escala. Según la norma ISO/IEC 14496-3: 2009, se
    utilizan 2048. Para implementaciones de elevación, por favor ajuste esta variable en consecuencia */
ds_window_size = N * fs_window_size / (1024 * F); /* Coeficientes de ventana con reducción de escala; N determina la longitud
    transformación según 4.6.20.2 */
fs_segment_size = 128;
num_segments = fs_window_size / fs_segment_size;
ds_segment_size = ds_window_size / num_segments;
tmp[128], y[128]; /* búfers temporales */

/* bucle sobre segmentos */
para (b = 0; b < num_segments; b++) {
    /* copiar segmento actual a tmp */
    copiar (&W_LD [b * fs_segment_size], tmp, fs_segment_size);

    /* aplicar la interpolación spline cúbica para la reducción de escala */

    /* calcular fase de interpolación */
    fase = (fs_window_size - ds_window_size) / (2 * ds_window_size);
    /* calcular los coeficientes c del spline cúbico dado tmp */
    /* matriz de constantes precalculados */
    m = {0,166666672, 0,25, 0,266666681, 0,267857134, 0,267942578, 0,267948717,
    0,267949164};
    n = fs_segment_size; /* por simplicidad */
    /* calcular el vector r necesario para calcular los coeficientes c */
    para (i = n - 3; i >= 0; i--)
        r[i] = 3 * ((tmp[i + 2] - tmp[i + 1]) - (tmp[i + 1] - tmp[i]));
    para (i = 1; i < 7; i++)
        r[i] = m[i - 1] * r[i - 1];
    para (i = 7; i < n - 4; i++)
        r[i] = 0,267949194 * r[i - 1];
    /* Calcular los coeficientes c */
    c[n - 2] = r[n - 3] / 6;
    c[n - 3] = (r[n - 4] - c[n - 2]) * 0,25;
    para (i = n - 4; i > 7; i--)
        c[i] = (r[i - 1] - c[i + 1]) * 0,267949194;
    para (i = 7; i > 1; i--)
        c[i] = (r[i - 1] - c[i + 1]) * m[i - 1];
    c[1] = r[0] * m[0];
    c[0] = 2 * c[1] - c[2];
    c[n - 1] = 2 * c[n - 2] - c[n - 3];
    /* mantener las muestras originales en el búfer y temporal ya que las muestras de tmp se
        reemplazarán con las muestras interpoladas */
    copiar(tmp, y, fs_segment_size);
    /* generar puntos con reducción de escala y hacer la interpolación */
    para (k = 0; k < ds_segment_size; k++) {
        paso = fase + k * fs_segment_size / ds_segment_size;
        idx = suelo(etapa);
        diff = etapa - IDX;
        di = (c[idx + 1] - c[idx]) / 3;
        bi = (y[idx + 1] - y[idx]) - (c[idx + 1] + 2 * c[idx]) / 3;
        /* calcular los valores con reducción de escala y se almacenan en tmp */
        tmp[k] = y[idx] + diff * (bi + diff * (c[idx] + diff * di));
    }
    /* montar ventana con reducción de escala */
    copiar(tmp, &W_LD_d[b * ds_segment_size], ds_segment_size);
}

```

A.2 Reducción de escala de la herramienta de SBR de bajo retardo

[0031] En el caso de que se utilice la herramienta de SBR de bajo retardo en conjunción con ELD, esta herramienta se puede reducir según las tasas de muestreo más bajas, al menos para los factores con reducción de escala de un múltiplo de 2. El factor con reducción de escala F controla el número de bandas utilizadas para el análisis de CLDFB y el banco de filtros de síntesis. Los dos párrafos siguientes describen un análisis de CLDFB con reducción de escala y el banco de filtros de síntesis, véase también 4.6.19.4.

4.6.20.5.2.1 Análisis con reducción de escala de banco de filtros de CLDFB

[0032]

- Definir el número de bandas CLDFB con reducción de escala. $B=32/F$
- Cambiar las muestras en la matriz x por B posiciones. Las B muestras más antiguas se descartan y las nuevas B muestras se almacenan en posiciones de 0 a $B-1$.
- Multiplicar las muestras de matriz x por el coeficiente de ventana ci para conseguir la matriz z . Los coeficientes de ventana se obtienen por interpolación lineal de los coeficientes c , es decir, a través de la ecuación

$$ci(i) = \frac{1}{2} [c(2F \cdot i + 1 + p) + c(2F \cdot i + p)], \quad 0 \leq i < (10B), \quad p = \text{int} \left(\frac{64}{2B} - 0.5 \right).$$

Los coeficientes de ventana de c se pueden encontrar en la Tabla 4.A.90.

- Sumar las muestras para crear la matriz u de $2B$ elementos:

$$u(n) = z(n) + z(n + 2B) + z(n + 4B) + z(n + 6B) + z(n + 8B), \quad 0 \leq n < (2B).$$

- Calcular B nuevas muestras de subbanda por la operación de la matriz Mu , donde

$$M(k, n) = 2 \cdot \exp \left(\frac{j \cdot \pi \cdot (k+0.5) \cdot (2n - (3B-1))}{2B} \right), \quad \begin{cases} 0 \leq k < B \\ 0 \leq n < 2B \end{cases}$$

En la ecuación, $\exp(\)$ denota la función exponencial compleja y j es la unidad imaginaria.

4.6.20.5.2.2 Síntesis con reducción de escala de banco de filtros de CLDFB

[0033]

- Definir número de bandas CLDFB con reducción de escala. $B=64/F$
- Cambiar las muestras en la matriz v por $2B$ posiciones. Las $2B$ muestras más antiguas se descartan.
- Las B nuevas muestras de subbanda con valores complejos se multiplican por la matriz N , donde

$$N(k, n) = \frac{1}{64} \cdot \exp \left(\frac{j \cdot \pi \cdot (k+0.5) \cdot (2n - (B-1))}{2B} \right), \quad \begin{cases} 0 \leq k < B \\ 0 \leq n < 2B \end{cases}$$

En la ecuación, $\exp(\)$ denota la función exponencial compleja y j es la unidad imaginaria. La parte real de la salida de esta operación se almacena en las posiciones 0 a $2B-1$ de la matriz w .

- Extraer muestras de v para crear la matriz g de $10B$ -elemento.

$$\begin{aligned} g(2B \cdot n + k) &= v(4B \cdot n + k) \\ g(2B \cdot n + B + k) &= v(4B \cdot n + 3B + k) \end{aligned} \quad \begin{cases} 0 \leq n \leq 4 \\ 0 \leq k < B \end{cases}$$

- Multiplicar las muestras de matriz g por el coeficiente de ventana ci para producir la matriz w . Los coeficientes de ventana ci se obtienen por medio de la interpolación lineal de los coeficientes c , es decir, a través de la ecuación

$$c(i) = \frac{1}{2} [c(2F \cdot i + 1 + p) + c(2F \cdot i + p)], \quad 0 \leq i < (108), p = \text{int}\left(\frac{64}{12} - 0.5\right).$$

Los coeficientes de ventana de ϵ se pueden encontrar en la Tabla 4.A.90.

- 5 • Calcular B nuevas muestras de salida por medio de la suma de las muestras de la matriz w según la ecuación

$$\text{salida}(n) = \sum_{i=0}^{B-1} w(Bi + n), \quad 0 \leq n < B.$$

- [0034] Se debe tener en cuenta que el ajuste $F=2$ ofrece el banco de filtros de síntesis submuestreado según
10 4.6.19.4.3. Por lo tanto, para procesar una corriente de bits LD-SBR submuestreada con un factor con reducción de escala adicional F , F se tiene que multiplicar por 2.

4.6.20.5.2.3 Banco de filtros de CLDFB de valor real con reducción de escala

- 15 [0035] La reducción de escala de CLDFB se puede aplicar para las versiones de valores reales del modo de bajo consumo de energía de SBR también. A título de ejemplo, por favor también considere 4.6.19.5.

- [0036] Para el banco de filtros de síntesis y análisis de valor real con reducción de escala, siga la descripción en 4.6.20.5.2.1 y 4.6.20.2.2 e intercambie el modulador $\exp()$ en M por un modulador de $\cos()$.

20

A.3 Análisis de MDCT de Bajo Retardo

- [0037] En esta subcláusula se describe el banco de filtros de MDCT de bajo retardo utilizado en el codificador AAC ELD. El algoritmo de núcleo MDCT es prácticamente igual, pero con una ventana más larga, de tal manera que
25 n está en funcionamiento de $-N$ a $N-1$ (en lugar de desde 0 hasta $N-1$)

- [0038] El coeficiente espectral, $X_{i,k}$, se define según lo presentado a continuación:

$$X_{i,k} = -2 \cdot \sum_{n=-N}^{N-1} z_{i,n} \cos\left(\frac{2\pi}{N}(n+n_0)\left(k+\frac{1}{2}\right)\right) \text{ para } 0 \leq k < N/2$$

30

donde:

- z_{in} =secuencia de entrada de ventana
N =índice de muestra
35 K =índice de coeficiente espectral
l =índice de bloque
N =longitud de la ventana
 n_0 = $(-N/2 + 1)/2$

- 40 [0039] La longitud de la ventana N (en base a la ventana senoidal) es 1024 o 960.

- [0040] La longitud de ventana de la ventana de bajo retardo es $2*N$. Las ventanas se extienden al pasado de la siguiente forma:

$$z_{i,n} = w_{LD}(N-1-n) \cdot x'_{i,n}$$

45

para $n = -N, \dots, N-1$, con la ventana de síntesis w utilizada como la ventana de análisis por medio de la inversión del orden.

50 A.4 Síntesis de MDCT de Bajo Retardo

- [0041] El banco de filtros de síntesis se modifica en comparación con el algoritmo IMDCT estándar por el uso de una ventana senoidal con el fin de adoptar un banco de filtros de bajo retardo. El algoritmo de IMDCT de núcleo es prácticamente igual, pero con una ventana más larga, de tal manera que n esté en funcionamiento hasta $2N-1$ (en
55 lugar de hasta $N-1$).

$$x_{i,n} = -\frac{2}{N} \sum_{k=0}^{\frac{N-1}{2}} \text{spec}[i][k] \cos\left(\frac{2\pi}{N}(n+n_0)\left(k+\frac{1}{2}\right)\right) \text{ para } 0 \leq n < 2N$$

donde:

- 5 n =índice de muestra
 i =índice de ventana
 k =índice de coeficiente espectral
 N =longitud de la ventana / el doble de la longitud de la trama
 n₀ =(-N / 2 + 1) / 2

10 con N = 960 o 1024.

[0042] La formación en ventanas y el solapamiento/suma se llevan a cabo de la siguiente manera:

La ventana de longitud N se sustituye por una ventana de longitud 2N con más solapamiento en el pasado, y menos
 15 solapamiento para el futuro (N/8 valores son en realidad cero).

Formación en ventanas para la Ventana de Bajo Retardo:

[0043]

20
$$z_{i,n} = w_{LD}(n) \cdot x_{i,n}$$

[0044] Cuando la ventana ahora tiene una longitud de 2N, por lo tanto, n = 0,..., 2N-1.

[0045] Se superponen y se suma:

25
$$\text{salida}_{i,n} = z_{i,n} + z_{i-1,n+\frac{N}{2}} + z_{i-2,n+N} + z_{i-3,n+N+\frac{N}{2}}$$

para 0 ≤ n < N/2

30 **[0046]** En este caso, los párrafos propuestos para ser incluidos en 14496-3:2009 a través de la enmienda final.

[0047] Naturalmente, la descripción anterior de un posible modo con reducción de escala para AAC-ELD representa meramente una realización de la presente solicitud y varias modificaciones son factibles. En general, las realizaciones de la presente solicitud no se limitan a un decodificador de audio que realiza una versión con reducción
 35 de escala de la decodificación AAC-ELD. En otras palabras, las realizaciones de la presente solicitud se pueden, por ejemplo, derivar por medio de la formación de un decodificador de audio capaz de llevar a cabo el proceso de transformación inversa de una manera con reducción de escala sólo sin apoyo o por el uso de las diversas tareas más específicas AAC-ELD tales como, por ejemplo, la transmisión basada en el factor de escala de la envolvente espectral, el filtrado de TNS (formación por ruido temporal), la replicación de banda espectral (SBR, por su sigla en inglés) o
 40 similares.

[0048] Posteriormente, se describe una realización más general para un decodificador de audio. El ejemplo descrito con anterioridad para un decodificador de audio AAC-ELD que apoya el modo con Reducción de escala descrito podría por lo tanto representar una implementación del decodificador de audio descrito posteriormente. En
 45 particular, el decodificador que se explica posteriormente se muestra en la Fig. 2 mientras que la Fig. 3 ilustra las etapas llevadas a cabo por el decodificador de la Fig. 2.

[0049] El decodificador de audio de la Fig. 2, que por lo general se indica por medio del signo de referencia 10, comprende un receptor 12, un captador 14, un modulador espectral a tiempo 16, un creador de ventanas 18 y un cancelador de solapamiento de dominio de tiempo 20, todos los cuales están conectados en serie el uno al otro en el orden de su mención. La interacción y la funcionalidad de los bloques 12 a 20 del decodificador de audio 10 se describen a continuación con respecto a la Fig. 3. Según se ha descrito al final de la descripción de la presente solicitud, los bloques 12 a 20 se pueden implementar en software, hardware programable o hardware, tales como en forma de un programa informático, una FPGA o un ordenador programado apropiadamente, un microprocesador programado o
 55 un circuito integrado de aplicación específica con los bloques 12 a 20 que representan subrutinas respectivas, trayectorias de circuito o similares.

[0050] De una manera que se describe en más detalles a continuación, el decodificador de audio 10 de la Fig. 2 se ha configurado para, (y los elementos del decodificador de audio 10 se han configurado para cooperar apropiadamente) con el fin de decodificar una señal de audio 22 de una corriente de datos 24 con una mención destacable de que el decodificador de audio 10 decodifica la señal 22 a una tasa de muestreo de $1/F^{\text{va}}$ de la tasa de muestreo donde la señal de audio 22 se ha codificado por transformada en la corriente de datos 24 en el lado de codificación. F puede ser, por ejemplo, cualquier número racional mayor que uno. El decodificador de audio se puede configurar para operar a diferentes o variables factores con reducción de escala F o en uno fijo. Las alternativas se describen con más detalle a continuación.

[0051] La manera en que la señal de audio 22 se codifica por transformada en tasa de la codificación o de muestreo original en la corriente de datos se ilustra en la Fig. 3 en la mitad superior. En el punto 26 de la Fig. 3, se ilustran los coeficientes espectrales por el uso de cajas pequeñas o cuadrados 28 dispuestos en una manera espectro-temporal a lo largo de un eje de tiempo 30 que se extiende horizontalmente en la Fig. 3, y un eje de frecuencia 32 que se extiende verticalmente en la Fig. 3, respectivamente. Los coeficientes espectrales 28 se transmiten dentro de corriente de datos 24. La manera en que se han obtenido los coeficientes espectrales 28, y por lo tanto la forma a través de la cual los coeficientes espectrales 28 representan la señal de audio 22, se ilustra en la Fig. 3 en el punto 34, que ilustra para una porción de eje de tiempo 30 cómo los coeficientes espectrales 28 pertenecientes a, o que representan la porción de tiempo respectiva, se han obtenido a partir de la señal de audio.

[0052] En particular, los coeficientes 28 según lo transmitido dentro de la corriente de datos 24 son coeficientes de una transformada solapada de la señal de audio 22 de tal manera que la señal de audio 22, muestreada en la tasa de muestreo original o de codificación, se dividen en tramas inmediatamente temporalmente consecutivas y que no se solapan de una longitud predeterminada N, donde N coeficientes espectrales se transmiten en corriente de datos 24 para cada trama 36. Es decir, los coeficientes de transformada 28 se obtienen a partir de la señal de audio 22 por el uso de una transformada solapada muestreada críticamente. En la representación de espectrograma espectro-temporal 26, cada columna de la secuencia temporal de columnas de coeficientes espectrales 28 corresponde a una respectiva de las tramas 36 de la secuencia de fotogramas. Los N coeficientes espectrales 28 se obtienen para la trama correspondiente 36 por medio de una transformada de descomposición espectral o modulación espectral a tiempo, las funciones de modulación de las cuales se extienden temporalmente, sin embargo, no sólo a través de la trama 36 a la que pertenecen los coeficientes espectrales resultantes 28, sino también a través de E+1 tramas anteriores, donde E puede ser cualquier número entero o cualquier número entero par mayor que cero. Es decir, los coeficientes espectrales 28 de una columna del espectrograma en el punto 26 que pertenecía a una trama determinada 36 se obtienen por medio de la aplicación de una transformada en una ventana de transformada, que además de la trama respectiva comprende E+1 tramas situadas en el pasado con respecto a la trama actual. La descomposición espectral de las muestras de la señal de audio dentro de esta ventana de transformada 38, que se ilustra en la Fig. 3 para la columna de coeficientes de transformada 28 que pertenecen a la carcasa intermedia 36 de la porción mostrada en el punto 34 se consigue por el uso de una función de ventana de análisis unimodal bajo retardo 40 con el que las muestras espectrales dentro de la ventana de transformada 38 se ponderan antes de someter la misma a una MDCT o MDST u otra transformada de descomposición espectral. Con el fin de reducir el retardo del lado del codificador, la ventana de análisis 40 comprende un intervalo de cero 42 en el extremo delantero temporal del mismo modo que el codificador no tiene que esperar a la porción correspondiente de las nuevas muestras dentro de la trama actual 36 con el fin de calcular los coeficientes espectrales 28 de esta trama actual 36. Es decir, en el intervalo de cero 42 la función de ventana de bajo retardo 40 es cero o tiene cero coeficientes de ventana para que las muestras de audio de ubicación conjunta de la trama actual 36 no lo hacen, debido a la ponderación de ventana 40, contribuyen a los coeficientes de transformada 28 transmitidos para esa trama y una corriente de datos 24. Es decir, resumiendo lo anterior, los coeficientes de transformada 28 que pertenecen a una trama actual 36 se obtienen por medio de la creación de ventanas y la descomposición espectral de muestras de la señal de audio dentro de una ventana de transformada 38 que comprende la trama actual, así como también las tramas temporalmente anteriores y que temporalmente se solapan con las ventanas de transformada correspondientes utilizadas para determinar los coeficientes espectrales 28 pertenecientes a las tramas temporalmente vecinas.

[0053] Antes de reanudar la descripción del decodificador de audio 10, se debe señalar que la descripción de la transmisión de los coeficientes espectrales 28 dentro de la corriente de datos 24 según lo previsto hasta ahora se ha simplificado con respecto a la manera donde los coeficientes espectrales 28 se cuantifican o se codifican en la corriente de datos 24 y/o la manera en que la señal de audio 22 se ha preprocesado antes de someter la señal de audio a la transformada solapada. Por ejemplo, el codificador de audio que tiene la señal de audio codificada por transformada 22 en la corriente de datos 24 se puede controlar por medio de un modelo psicoacústico o puede utilizar un modelo psicoacústico para mantener el ruido de cuantificación y cuantificar los coeficientes espectrales 28 imperceptibles para el oyente y/o por debajo de una función de umbral de enmascaramiento, lo que de ese modo determina los factores de escala para las bandas espectrales por el uso de los coeficientes espectrales cuantificados y transmitidos 28 que se escalan. Los factores de escala también se señalizan en la corriente de datos 24. De forma alternativa, el codificador de audio puede haber sido un tipo de codificador TCX (excitación codificada de transformada). Entonces, la señal de audio se habría sometido a un filtrado de análisis de predicción lineal antes de formar la representación espectro-temporal 26 de coeficientes espectrales 28 por medio de la aplicación de la transformada de solapado sobre

la señal de excitación, es decir, la señal residual de predicción lineal. Por ejemplo, los coeficientes de predicción lineal pueden estar señalados en la corriente de datos 24, de este modo, y se podría aplicar una cuantificación uniforme espectral con el fin de obtener los coeficientes espectrales 28.

5 **[0054]** Además, la descripción presentada hasta ahora también se ha simplificado con respecto a la longitud de la trama de las tramas 36 y/o con respecto a la función de ventana de bajo retardo 40. De hecho, la señal de audio 22 puede haber sido codificada en la corriente de datos 24 de una manera por el uso de diferentes tamaños de trama y/o diferentes ventanas 40. Sin embargo, la descripción presentada a continuación en lo sucesivo se concentra en una ventana 40 y una longitud de la trama, aunque la descripción posterior con facilidad se puede extender a un caso
10 donde el codificador de entropía cambia estos parámetros durante la codificación de la señal de audio en la corriente de datos.

[0055] Volviendo de nuevo al decodificador de audio 10 de la Fig. 2 y su descripción, el receptor 12 recibe la corriente de datos 24 y recibe de este modo, para cada trama 36, N coeficientes espectrales 28, es decir, una columna
15 respectiva de coeficientes 28 mostrados en la Fig. 3. Se debe recordar que la longitud temporal de las tramas 36, medida en muestras de la tasa de muestreo original o de codificación, es N, según lo indicado en la Fig. 3 a 4, pero el decodificador de audio 10 de la Fig. 2 se ha configurado para decodificar la señal de audio 22 a una tasa de muestreo reducida. El decodificador de audio 10 soporta, por ejemplo, simplemente esta funcionalidad de decodificación con reducción de escala que se describe a continuación. De forma alternativa, el decodificador de audio 10 sería capaz de
20 reconstruir la señal de audio a la tasa de muestreo original o de codificación, pero se puede conmutar entre el modo de decodificación con reducción de escala y un modo de decodificación sin reducción de escala con el modo de decodificación con reducción de escala que coincide con el modo de operación del decodificador de audio 10 explicado posteriormente. Por ejemplo, el codificador de audio 10 se podría cambiar a un modo de decodificación con reducción de escala en el caso de un nivel de batería baja, la reducción de las capacidades de reproducción de entorno o
25 similares. Cuando la situación cambia, el decodificador de audio 10 podría, por ejemplo, cambiar de nuevo desde el modo de decodificación con reducción de escala al de sin reducción de escala. En cualquier caso, según el proceso de decodificación con reducción de escala del decodificador 10 según se ha descrito en lo sucesivo, la señal de audio 22 se reconstruye a una tasa de muestreo donde las tramas 36 tienen, en la tasa de muestreo reducida, una longitud inferior medida en muestras de esta tasa de muestreo reducida, es decir, una longitud de las muestras N/F en la tasa
30 de muestreo reducida.

[0056] La salida del receptor 12 es la secuencia de N coeficientes espectrales, es decir, un conjunto de N coeficientes espectrales, es decir, una columna en la Fig. 3, por trama 36. Ya resultó de la breve descripción anterior del proceso de codificación de transformada para la formación de la corriente de datos 24 que el receptor 12 se pueden
35 aplicar varias tareas en la obtención de los N coeficientes espectrales por trama 36. El receptor 12 utiliza la decodificación por entropía para leer los coeficientes espectrales 28 de la corriente de datos 24. El receptor 12 también da forma espectralmente a los coeficientes espectrales leídos de la corriente de datos con los factores de escala dentro de la corriente de datos y/o los factores de escala obtenidos por coeficientes de predicción lineal transportados dentro de la corriente de datos 24. Por ejemplo, el receptor 12 puede obtener los factores de escala de la corriente de datos
40 24, a saber, en una por trama y por base de subbanda, y utilizar estos factores de escala con el fin de ampliar los factores de escala transmitidos dentro de la corriente de datos 24. De forma alternativa, el receptor 12 puede derivar los factores de escala de coeficientes de predicción lineal transmitidos dentro de la corriente de datos 24, para cada trama 36, y utilizar estos factores de escala con el fin de escalar los coeficientes espectrales de transmisión 28. De forma opcional, el receptor 12 puede llevar a cabo el relleno de espacios con el fin de llenar porciones sintéticamente
45 cuantificadas en cero dentro de los conjuntos de N coeficientes espectrales 18 por trama. De forma adicional o alternativa, el receptor 12 puede aplicar un filtro de síntesis de TNS en un coeficiente de filtro TNS transmitido por trama para ayudar a la reconstrucción de los coeficientes espectrales 28 de la corriente de datos con los coeficientes de TNS también transmitidos dentro de la corriente de datos 24. Las posibles funciones simplemente esbozadas del receptor 12 se entenderán como una lista no exhaustiva de posibles medidas y el receptor 12 puede llevar a cabo otras tareas
50 en relación con la lectura de los coeficientes espectrales 28 de la corriente de datos 24.

[0057] El capturador 14 recibe de ese modo desde el receptor 12 el espectrograma 26 de coeficientes espectrales 28 y captura, para cada trama 36, una fracción de baja frecuencia 44 de los N coeficientes espectrales de la trama respectiva 36, es decir, los coeficientes espectrales N/F de menor frecuencia.

55 **[0058]** Es decir, el modulador espectral a tiempo 16 recibe del capturador 14 una corriente o secuencia 46 de N/F coeficientes espectrales 28 por trama 36, que corresponde a una porción de baja frecuencia del espectrograma 26, registrado espectralmente en los coeficientes espectrales de frecuencia más baja que se ilustran por medio del índice "0" en la Fig. 3, y que se extienden hasta los coeficientes espectrales de índice N/F-1.

60 **[0059]** El modulador espectral a tiempo 16 somete, para cada trama 36, la fracción de baja frecuencia correspondiente 44 de coeficientes espectrales 28 a una transformada inversa 48 que tiene funciones de modulación de longitud $(E+2) \cdot N/F$ que temporalmente se extienden sobre la trama respectiva y E+1 tramas anteriores según lo ilustrado en el punto 50 en la Fig. 3, para obtener de ese modo una porción temporal de longitud $(E+2) \cdot N/F$, es decir, un
65 segmento de tiempo 52 aún no colocado en una ventana. Es decir, el modulador espectral a tiempo puede obtener un

segmento de tiempo temporal de $(E+2) \cdot N/F$ muestras de tasa de muestreo reducida por medio de la ponderación y la suma de funciones de modulación de la misma longitud por el uso de, por ejemplo, las primeras fórmulas de la sección de reemplazo A.4 propuesta indicada con anterioridad. Las nuevas N/F muestras del segmento de tiempo 52 pertenecen a la trama actual 36. Las funciones de modulación pueden, según lo indicado, ser funciones coseno en el caso de que la transformada inversa sea una MDCT inversa, o funciones seno en el caso de que la transformada inversa sea una MDCT inversa, por ejemplo.

[0060] Por lo tanto, el formador de ventanas 52 recibe, para cada trama, una porción temporal 52, las N/F muestras en el extremo delantero de la misma temporalmente correspondiente a la trama respectiva, mientras que las otras muestras de la respectiva porción temporal 52 pertenecen a las correspondientes tramas temporalmente anteriores. El formador de ventanas 18 forma una ventana, para cada trama 36, en la porción temporal 52 por el uso de una ventana de síntesis unimodal 54 de longitud $(E+2) \cdot N/F$ que comprende una porción nula 56 de longitud $1/4 \cdot N/F$ en un extremo delantero de la misma, es decir, $1/F \cdot N/F$ coeficientes de ventana valorados en cero, y que tiene un pico 58 dentro de su intervalo temporal subsiguiente, temporalmente, la porción nula 56, es decir, el intervalo temporal de la porción temporal 52 no cubierta por la porción nula 52. El último intervalo temporal se puede denominar la porción no nula de ventana 58 y tiene una longitud de $7/4 \cdot N/F$ medido en muestras de la tasa de muestreo reducida, es decir, $7/4 \cdot N/F$ coeficientes de ventana. El formador de ventanas 18 pesa, por ejemplo, la porción temporal 52 por el uso de la ventana 58. Esta ponderación o multiplicación 58 de cada porción temporal 52 con la ventana 54 da lugar a una porción temporal de ventana 60, una para cada trama 36, y coincide con la porción temporal respectiva en lo que se refiere a la cobertura temporal. En la sección A.4 propuesta con anterioridad, el procesamiento de ventanas que puede ser utilizado por la ventana 18 se describe por las fórmulas relativas $z_{i,n}$ a $x_{i,n}$, donde $x_{i,n}$ corresponde a las porciones temporales 52 mencionadas con anterioridad aún no con una ventana formada y $z_{i,n}$ corresponde a las porciones temporales de ventana 60 con i como la indexación de la secuencia de tramas/ventanas, y n como la indexación, dentro de cada porción temporal 52/60, de las muestras o los valores de las respectivas porciones 52/60 según una tasa de muestreo reducida.

[0061] Por lo tanto, el tiempo de cancelador de solapamiento de dominio de tiempo 20 recibe del formador de ventanas 18 una secuencia de porciones temporales de ventana 60, es decir, una por cada trama 36. El cancelador 20 somete las porciones temporales de ventana 60 de las tramas 36 a un proceso de superposición y suma 62 por medio del registro de cada porción temporal de ventana 60 con sus N/F valores principales para coincidir con la trama correspondiente 36. Por esta medida, una fracción del extremo trasero de longitud $(E+1)/(E+2)$ de la porción temporal de ventana 60 de una trama actual, es decir, el resto que tiene una longitud $(E+1) \cdot N/F$, se superpone con un correspondiente extremo igualmente largo delantero de la porción temporal de la trama inmediatamente anterior. En las fórmulas, el cancelador de solapamiento de dominio de tiempo 20 puede funcionar según lo que se muestra en la última fórmula de la versión anterior propuesta de la sección A.4, donde $salida_{i,n}$ corresponde a las muestras de audio de la señal de audio reconstruida 22 en la tasa de muestreo reducida.

[0062] Los procesos de formación de ventanas 58 y la superposición y suma 62 según lo llevado a cabo por el formador de ventanas 18 y el cancelador de solapamiento de dominio de tiempo 20 se ilustran en más detalle a continuación con respecto a la Fig. 4. La Fig. 4 utiliza tanto la nomenclatura aplicada en la sección A.4 propuesta con anterioridad y los signos de referencia aplicados en las Figs. 3 y 4. $x_{0,0}$ a $x_{0,(E+2) \cdot N/F-1}$ representa la porción temporal 0^{va} 52 obtenida por el modulador espacial a temporal 16 para la trama 0^{va} 36. El primer índice de x indexa las tramas 36 a lo largo del orden temporal, y el segundo índice de x ordena las muestras de la temporal a lo largo del orden temporal, el paso de inter-muestra pertenece a la tasa de muestreo reducida. A continuación, en la Fig. 4, w_0 a $w_{(E+2) \cdot N/F-1}$ indican los coeficientes de ventana de la ventana 54. Al igual que el segundo índice de x , es decir, la porción temporal 52 según la salida del modulador 16, el índice de w es tal que el índice 0 corresponde al índice más antiguo y $(E+2) \cdot N/F-1$ se corresponde con el valor de la muestra más reciente cuando la ventana 54 se aplica a la porción temporal 52 respectiva. El formador de ventanas 18 crea una ventana de la porción temporal 52 por el uso de la ventana 54 para obtener la porción temporal de ventana 60 de tal manera que $z_{0,0}$ a $z_{0,(E+2) \cdot N/F-1}$, que indica la porción temporal de ventana 60 para la trama 0^{va}, se obtenga según $z_{0,0} = x_{0,0} \cdot w_{0,0}, \dots, z_{0,(E+2) \cdot N/F-1} = x_{0,(E+2) \cdot N/F-1} \cdot w_{(E+2) \cdot N/F-1}$. Los índices de z tienen el mismo significado que para x . De esta manera, el modulador 16 y el formador de ventanas 18 actúan para cada trama indexada por el primer índice de x y z . El cancelador 20 resume las $E+2$ porciones temporales de ventana 60 de $E+2$ fotogramas inmediatamente consecutivos con la compensación de las muestras de las porciones temporales de ventana 60 respecto a la otra por una trama, es decir, por el número de muestras por trama 36, es decir, N/F , con el fin de obtener las muestras u de una trama actual, en este caso $u_{-(E+1),0} \dots u_{-(E+1),N/F-1}$. En este caso, de nuevo, el primer índice de u indica el número de trama y el segundo índice ordena las muestras de esta trama a lo largo del orden temporal. El cancelador de este modo une las tramas reconstruidas obtenidas de tal manera que las muestras de la señal de audio reconstruida 22 dentro de las tramas consecutivas 36 se siguen unas a otras según $u_{-(E+1),0} \dots u_{-(E+1),N/F-1}, u_{E,0} \dots u_{E,N/F-1}, u_{(E-1),0}, \dots$ el cancelador 22 calcula cada muestra de la señal de audio 22 dentro de $-(E+1)^{va}$ trama según $u_{-(E+1),0} = z_{0,0} + z_{-1,N/F} + \dots + z_{-(E-1),(E+1) \cdot N/F}, \dots, u_{-(E+1),N/F-1} = z_{0,N/F-1} + z_{-1,2 \cdot N/F-1} + \dots + z_{-(E+1),(E+2) \cdot N/F-1}$ es decir, resumiendo $(E+2)$ sumandos por u muestras de la trama actual.

[0063] La Fig. 5 ilustra una posible explotación del hecho de que, entre las muestras con ventanas recién formadas que contribuyen a las muestras de audio u de trama $-(E+1)$, las que corresponden a, o se han formado en ventanas por el uso de, la porción nula 56 de la ventana 54, a saber, $z_{-(E+1),(E+7/4) \cdot N/F} \dots z_{-(E+1),(E+2) \cdot N/F-1}$ se valoran cero.

Por lo tanto, en lugar de obtener todas las muestras N/F dentro de la $(E+1)^{va}$ trama 36 de la señal de audio u por el uso de $E+2$ sumandos, el cancelador 20 puede calcular el cuarto extremo delantero de la misma, es decir, $u_{(E+1),(E+7/4) \cdot N/F} \dots u_{(E+1),(E+2) \cdot N/F-1}$ por el mero uso de $E+1$ sumandos según $u_{-(E+1),(E+7/4) \cdot N/F} = z_{0,3/4 \cdot N/F} + z_{-1,7/4 \cdot N/F} + \dots + z_{-E,(E+1) \cdot N/F-1}$. De esta manera, el formador de ventanas podría incluso dejar fuera, de manera efectiva, el rendimiento de la ponderación 58 con respecto a la porción nula 56. Las muestras $u_{-(E+1),(E+7/4) \cdot N/F} \dots u_{-(E+1),(E+2) \cdot N/F-1}$ de la $-(E+1)^{va}$ trama actual se podría obtener, por lo tanto, por el uso de sólo $E+1$ sumandos, mientras que $u_{-(E+1),(E+1) \cdot N/F} \dots u_{-(E+1),(E+7/4) \cdot N/F-1}$ que se obtendría por el uso de $E+2$ sumandos.

[0064]

De este modo, en la forma indicada con anterioridad, el decodificador de audio 10 de la Fig. 2 reproduce, en una manera con reducción de escala, la señal de audio codificada en la corriente de datos 24. Para este fin, el decodificador de audio 10 utiliza una función de ventana 54 que es en sí misma una versión submuestreada de una ventana de síntesis de referencia de longitud $(E+2) \cdot N$. Según lo explicado con respecto a la Fig. 6, esta versión submuestreada, es decir, la ventana 54, se obtiene por medio del submuestreo de la ventana de síntesis de referencia por un factor de F , es decir, el factor de submuestreo, por el uso de una interpolación segmentaria, a saber, en segmentos de longitud $1/4 \cdot N$ cuando se miden en el régimen aún sin reducción de escala, en segmentos de longitud $1/4 \cdot N/F$ en el régimen de submuestreo, en los segmentos de cuartos de longitud de la trama de las tramas 36, medidos de forma temporal y expresados de forma independiente de la tasa de muestreo. Por lo tanto, en $4 \cdot (E+2)$ la interpolación se lleva a cabo, lo que de este modo rinde $4 \cdot (E+2)$ veces segmentos de $1/4 \cdot N/F$ de largo, que, concatenados, representan la versión submuestreada de la ventana de síntesis de referencia de longitud $(E+2) \cdot N$. Véase la Fig. 6 para la ilustración. La Fig. 6 muestra la ventana de síntesis 54, que es unimodal y utilizada por el decodificador de audio 10 según un procedimiento de decodificación de audio muestreado descendente de la ventana de síntesis de referencia 70 con su longitud de $(E+2) \cdot N$. Es decir, por medio del procedimiento de submuestreo 72 que conduce desde la ventana de síntesis de referencia 70 a la ventana de síntesis 54 realmente utilizada por el decodificador de audio 10 para la decodificación submuestreada, el número de coeficientes de ventana se reduce en un factor de F . En la Fig. 6, la nomenclatura de las Figs. 5 y 6 se ha adherido a, es decir, w se utiliza con el fin de denotar la ventana de versión submuestreada 54, mientras que w' se ha utilizado para denotar los coeficientes de ventana de la ventana de síntesis de referencia 70.

[0065]

Según lo que se acaba de mencionar, con el fin de llevar a cabo el submuestreo 72, la ventana de síntesis de referencia 70 se procesa en segmentos 74 de igual longitud. En número, hay $(E+2) \cdot 4$ tales segmentos 74. Si se mide en la tasa de muestreo original, es decir, en el número de coeficientes de ventana de la ventana de síntesis de referencia 70, cada segmento 74 tiene $1/4 \cdot N$ coeficientes de ventana w' de largo, y si se mide en la tasa de muestreo reducida o submuestreada, cada segmento 74 es de $1/4 \cdot N/F$ coeficientes de ventana w de largo.

[0066]

Naturalmente, sería posible llevar a cabo el submuestreo 72 para cada coeficiente de ventana submuestreado w_i que coincide accidentalmente con cualquiera de los coeficientes de ventana w'_j de la ventana de síntesis

$$w_i = w'_j$$

de referencia 70 simplemente por medio del ajuste de w_i con el tiempo de la muestra de w_i que coincide con el de w'_j , y/o por medio de la interpolación lineal de cualquier coeficiente de ventana w_i que reside, temporal-

mente, entre dos coeficientes de ventana w'_{j-1} y w'_{j+2} por medio de interpolación lineal, pero este procedimiento daría lugar a una aproximación pobre de la ventana de síntesis de referencia 70, es decir, la ventana de síntesis 54 utilizada por el decodificador de audio 10 para la decodificación submuestreada representaría una aproximación pobre de la ventana de síntesis de referencia 70, con lo que no cumpliría con la solicitud de garantizar las pruebas de conformidad de la decodificación con reducción de escala en relación con la decodificación sin reducción de escala de la señal de audio de la corriente de datos 24. Por lo tanto, el submuestreo 72 implica un procedimiento de interpolación según el cual la mayor parte de los coeficientes de ventana w_i de la ventana submuestreada 54, a saber, los posicionados desplazados de las fronteras de segmentos 74, depende del procedimiento de submuestreo 72 en más de dos coeficientes de ventana w' de la ventana de referencia 70. En particular, aunque la mayoría de los coeficientes de

ventana w_i de la ventana submuestreada 54 dependen de más de dos coeficientes de ventana w'_{j-1}, w'_j, w'_{j+1} de la ventana de referencia 70 con el fin de aumentar la calidad del resultado de la interpolación/submuestreo, es decir, la calidad de aproximación, para cada coeficiente de ventana w_i de la versión submuestreada 54, es válido que el mismo no dependa

de los coeficientes de ventana w'_{j-1}, w'_j, w'_{j+1} que pertenecen a diferentes segmentos 74. Más bien, el procedimiento de submuestreo 72 es un procedimiento de interpolación segmentaria.

[0067]

La ventana de síntesis 54 es una concatenación de funciones de spline de longitud $1/4 \cdot N/F$. Se pueden utilizar funciones de spline cúbicas. Tal ejemplo se ha descrito con anterioridad en la sección A.1 donde el exterior para la próxima secuencia de bucle en bucle sobre los segmentos 74 donde, en cada segmento 74, el submuestreo o

la interpolación 72 implicó una combinación matemática de los coeficientes de ventana w' consecutivos dentro del segmento actual 74 en, por ejemplo, la primera para la siguiente cláusula en la sección "calcular el vector r necesario para calcular los coeficientes c ". Sin embargo, la interpolación aplicada en segmentos, también se puede elegir de manera diferente. Es decir, la interpolación no se limita a los splines o splines cúbicos. Más bien, la interpolación lineal o cualquier otro procedimiento de interpolación se pueden utilizar también. En cualquier caso, la aplicación segmentaria de la interpolación haría que el cálculo de muestras de la ventana de síntesis con reducción de escala, es decir, las muestras más externas de los segmentos de la ventana de síntesis con reducción de escala, que se encuentran próximos a otro segmento, no dependan de los coeficientes de ventana de la ventana de síntesis de referencia que residen en diferentes segmentos.

[0068] Puede ser que el formador de ventanas 18 obtenga la ventana de síntesis submuestreada 54 desde un almacenamiento donde los coeficientes de ventana w_i de esta ventana de síntesis submuestreada 54 se hayan almacenado después de haber sido obtenidos por medio del submuestreo 72. De forma alternativa, según lo ilustrado en la Fig. 2, el decodificador de audio 10 puede comprender un submuestreador segmentario 76 que lleva a cabo el submuestreo 72 de la Fig. 6 sobre la base de la ventana de síntesis de referencia 70.

[0069] Se debe señalar que el decodificador de audio 10 de la Fig. 2 se puede configurar para soportar simplemente un factor de submuestreo F fijo o puede soportar diferentes valores. En ese caso, el decodificador de audio 10 puede ser sensible a un valor de entrada para F según lo ilustrado en la Fig. 2 en el punto 78. El captador 14, por ejemplo, puede ser sensible a este valor F con el fin de capturar, según lo mencionado con anterioridad, los N/F valores espectrales por espectro de trama. De una manera similar, el submuestreador segmentario opcional 76 también puede ser sensible a este valor de F para operar según lo indicado con anterioridad. El modulador S/T 16 puede ser sensible a F o bien con el fin de, por ejemplo, derivar computacionalmente las versiones con reducción de escala/submuestreadas de las funciones de modulación, con reducción de escala/submuestreadas relativas a las utilizadas en el modo de operación sin reducción de escala, donde la reconstrucción conduce a la tasa de muestreo de audio completa.

[0070] Naturalmente, el modulador 16 también sería sensible a la entrada de F 78, dado que el modulador 16 utilizaría versiones submuestreadas en forma adecuada de las funciones de modulación y lo mismo es cierto para el formador de ventanas 18 y el cancelador 20 con respecto a una adaptación de la longitud real de las tramas en la tasa de muestreo reducida o submuestreada.

[0071] Por ejemplo, F puede estar entre 1,5 y 10, ambos inclusive.

[0072] Se debe señalar que el decodificador de la Fig. 2 y 3 o cualquier modificación de la misma indicado en esta invención, se pueden implementar con el fin de llevar a cabo la transición espectral a tiempo por el uso de una aplicación de elevación del MDCT de bajo retardo según lo enseñado en, por ejemplo, el documento de patente EP 2 378 516 B1.

[0073] La Fig. 8 ilustra una implementación del decodificador por el uso del concepto de elevación. El modulador S/T 16 lleva a cabo a modo de ejemplo una DCT-IV inversa y se muestra como seguido de un bloque que representa la concatenación del formador de ventanas 18 y el cancelador de solapamiento de dominio de tiempo 20. En el ejemplo de la Fig. 8 E es 2, es decir $E=2$.

[0074] El modulador 16 comprende un convertidor de frecuencia/tiempo de transformada inversa de coseno discreta de tipo-iv. En lugar de dar salida a las secuencias de largas porciones temporales $(E+2)N/F$ 52, simplemente da salida a porciones temporales 52 de longitud $2 \cdot N/F$, todas derivadas de la secuencia de N/F de espectros largos 46, estas porciones acortadas 52 corresponden al núcleo de DCT, es decir, las $2 \cdot N/F$ muestras más nuevas de las porciones antiguas descritas.

[0075] El formador de ventanas 18 actúa según se ha descrito con anterioridad y genera una porción temporal de ventana 60 para cada porción temporal 52, pero opera únicamente en el núcleo DCT. Con este fin, el formador de ventanas 18 utiliza la función de ventana w_i con $i = 0 \dots 2 \cdot N/F - 1$, que tiene el tamaño del núcleo. La relación entre w_i con $i = 0 \dots (E+2) \cdot N/F - 1$ se describe más adelante, así como también la relación entre los coeficientes de elevación posteriormente mencionados y w_i con $i = 0 \dots (E+2) \cdot N/F - 1$.

[0076] Por el uso de la nomenclatura aplicada con anterioridad, el proceso descrito produce hasta ahora:

$$Z_{k,n} = \omega_n \cdot X_{k,n} \quad \text{para } n = 0, \dots, 2M-1,$$

con la redefinición de $M = N/F$, de tal manera que M corresponde al tamaño de la trama expresada en el dominio con reducción de escala y por el uso de la nomenclatura de las Figs. 2 a 6, donde, sin embargo, $Z_{k,n}$ y $X_{k,n}$ deberá contener simplemente las muestras de la porción temporal de ventana y la porción temporal aún no de ventana dentro del núcleo DCT tiene un tamaño de $2 \cdot M$ y que temporalmente corresponde a las muestras $E \cdot N/F \dots (E+2) \cdot N/F - 1$ en la Fig.

4. Es decir, n es un número entero que indica un índice de muestra y ω_n es un coeficiente de función de ventana de valor real que corresponde al índice de muestra n .

[0077] El proceso de solapamiento/suma del cancelador 20 opera de una manera diferente en comparación con la descripción anterior. Genera porciones temporales intermedias $m_k(0), \dots, m_k(M-1)$ en base a la ecuación o expresión

$$m_{k,n} = z_{k,n} + z_{k-1,n+M} \quad \text{para } n = 0, \dots, M-1$$

10 [0078] En la implementación de la Fig. 8, el aparato comprende además un elevador 80 que se puede interpretar como una parte del modulador 16 y el formador de ventanas 18 dado que el elevador 80 compensa el hecho de que el modulador y el formador de ventanas restringe su procesamiento al núcleo DCT en lugar de procesar la extensión de las funciones de modulación y la ventana de síntesis más allá del núcleo hacia el pasado, dicha extensión se introdujo para compensar la porción nula 56. El elevador 80 produce, por el uso de un marco de los retardadores y los
15 multiplicadores 82 y sumadores 84, las últimas porciones temporales reconstruidas o tramas de longitud M en pares de tramas inmediatamente consecutivas en base a la ecuación o la expresión

$$u_{k,n} = m_{k,n} + l_{n-M/2} \cdot m_{k-1,M-1-n} \quad \text{para } n = M/2, \dots, M-1$$

y
20

$$u_{k,n} = m_{k,n} + l_{M-1-n} \cdot \text{salida}_{k-1,M-1-n} \quad \text{para } n=0, \dots, M/2-1$$

donde l_n con $n = 0 \dots M-1$ son coeficientes de elevación con valores reales relacionados con la ventana de síntesis con reducción de escala de una manera descrita con más detalle a continuación.

25 [0079] En otras palabras, para la superposición prolongada de E tramas en el pasado, sólo se requieren M operaciones de multiplicador-suma adicionales, según lo que se puede observar en el marco del elevador 80. Estas operaciones adicionales a veces también se denominan como "matrices de retardo nulo". A veces, estas operaciones también se conocen como "etapas de elevación". La aplicación eficiente que se muestra en la Fig. 8 puede en algunas
30 circunstancias ser más eficiente como una implementación sencilla. Para ser más precisos, según la aplicación concreta, tal aplicación más eficiente podría dar lugar al ahorro de las operaciones de fusión, como en el caso de una implementación directa de las operaciones de fusión, podría ser aconsejable poner en práctica, ya que la aplicación que se muestra en la Fig. 8, requiere, en principio, las $2M$ operaciones en el marco del módulo 820 y M en operaciones en el marco del elevador 830.

35 [0080] En cuanto a la dependencia de ω_n con $n = 0 \dots 2M-1$ y l_n con $n = 0 \dots M-1$ en la ventana de síntesis w_i con $i = 0 \dots (E+2)M-1$ (se recuerda que en esta invención $E=2$), las siguientes fórmulas describen la relación entre ellos con desplazamiento, sin embargo, los índices de subíndices utilizados hasta ahora en el paréntesis que sigue a la variable respectiva:

40

$$w(i) = l\left(\frac{M}{2} - 1 - n\right) \cdot l(M-1-n) \cdot \omega(M+n)$$

$$w(M/2 + i) = l(n) \cdot l(M/2 + n) \cdot \omega(3M/2 + n)$$

45

$$w(M + i) = l\left(\frac{M}{2} - 1 - n\right) \cdot \omega(M + n)$$

$$w(3M/2 + i) = -l(n) \cdot \omega(3M/2 + n)$$

50

$$w(2M + i) = -\omega(M + n) - l(M-1-n) \cdot \omega(n)$$

$$w(5M/2 + i) = -\omega(3M/2 + n) - l(M/2 + n) \cdot \omega(M/2 + n)$$

$$w(3M + i) = -\omega(n)$$

55

$$w(7M/2 + i) = \omega(M + n)$$

$$\text{para } i, n = 0 \dots \frac{M}{2} - 1$$

[0081] Se debe tener en cuenta que la ventana w_i contiene los valores pico en el lado derecho en esta formulación, es decir, entre los índices $2M$ y $4M-1$. Las fórmulas anteriores refieren coeficientes de l_n con $n = 0 \dots M-1$ y $\omega_n = 0, \dots, 2M-1$ a los coeficientes w_n con $n = 0 \dots (E+2)M-1$ de la ventana de síntesis con reducción de escala. Como puede verse, l_n con $n = 0 \dots M-1$ en realidad sólo depende de $\frac{3}{4}$ de los coeficientes de ventana de síntesis submuestreada, a saber, en w_n con $n = 0 \dots (E+1)M-1$, mientras que ω_n con $n = 0, \dots, 2M-1$ dependen de todo w_n con $n = 0 \dots (E+2)M-1$.

[0082] Según lo indicado con anterioridad, podría ser que el formador de ventanas 18 obtenga la ventana de síntesis submuestreada w_n con $n = 0 \dots (E+2)M-1$ desde un almacenamiento donde los coeficientes de ventana w_i de esta ventana de síntesis submuestreada 54 se han almacenado después de haber sido obtenida por el uso del submuestreo 72, y desde donde los mismos se leen para calcular los coeficientes de l_n con $n = 0 \dots M-1$ y $\omega_n = 0, \dots, 2M-1$ por el uso de la relación anterior, pero, de forma alternativa, el formador de ventanas 18 puede recuperar los coeficientes de l_n con $n = 0 \dots M-1$ y ω_n con $n = 0, \dots, 2M-1$, calculados de ese modo a partir de la ventana de síntesis pre-submuestreada, directamente desde el almacenamiento. De forma alternativa, según lo indicado con anterioridad, el decodificador de audio 10 puede comprender el submuestreador segmentario 76 para llevar a cabo el submuestreo 72 de la Fig. 6 sobre la base de la ventana de síntesis de referencia 70, lo que da lugar a w_n con $n = 0 \dots (E+2)M-1$ sobre la base de que el formador de ventanas 18 calcula los coeficientes de l_n con $n = 0 \dots M-1$ y ω_n con $n = 0, \dots, 2M-1$ por el uso de las relaciones/fórmulas anteriores. Incluso mediante el uso de la implementación de elevación, se puede soportar más de un valor para F .

[0083] Para sintetizar brevemente la implementación de elevación, los mismos resultados en un decodificador de audio 10 configurado para decodificar una señal de audio 22 a una primera tasa de muestreo de una corriente de datos 24 donde la señal de audio está codificada por transformada a una segunda tasa de muestreo, la primera tasa de muestreo es $1/F^{va}$ de la segunda tasa de muestreo, el decodificador de audio 10 comprende el receptor 12 que recibe, por trama de longitud N de la señal de audio, N coeficientes espectrales 28, el capturador 14 que captura para cada trama, una fracción de baja frecuencia de longitud N/F de los N coeficientes espectrales 28, un modulador espectral a tiempo 16 configurado para someter, para cada trama 36, la fracción de baja frecuencia a una transformada inversa que tiene funciones de modulación de longitud $2 \cdot N/F$ que se extiende temporalmente sobre la trama respectiva y una trama anterior con el fin de obtener una porción temporal de longitud $2 \cdot N/F$, y un formador de ventanas 18 que forma una ventana, para cada trama 36, la porción temporal $x_{k,n}$ según $z_{k,n} = \omega_n \cdot x_{k,n}$ para $n = 0, \dots, 2M-1$ con el fin de obtener una porción temporal de ventana $z_{k,n}$ con $n = 0 \dots 2M-1$. El cancelador de solapamiento de dominio de tiempo 20 genera porciones temporales intermedias $m_k(0), \dots, m_k(M-1)$ según $m_{k,n} = z_{k,n} + z_{k-1,n+M}$ para $n = 0, \dots, M-1$. Por último, el elevador 80 calcula tramas $u_{k,n}$ de la señal de audio con $n = 0 \dots M-1$ según $u_{k,n} = m_{k,n} + l_{n-M/2} \cdot m_{k-1,M-1-n}$ para $n = M/2, \dots, M-1$, y $u_{k,n} = m_{k,n} + l_{M-1-n} \cdot \text{salida}_{k-1,M-1-n}$ para $n = 0, \dots, M/2-1$, donde l_n con $n = 0 \dots M-1$ son coeficientes de elevación, donde la transformada inversa es una MDCT inversa o MDST inversa, y donde l_n con $n = 0 \dots M-1$ y ω_n con $n = 0, \dots, 2M-1$ dependen de coeficientes w_n con $n = 0 \dots (E+2)M-1$ de una ventana de síntesis, y la ventana de síntesis es una versión submuestreada de una ventana de síntesis de referencia de longitud $4 \cdot N$, submuestreada por un factor de F por una interpolación segmentaria en segmentos de longitud $1/4 \cdot N$.

[0084] Ya resultó de la discusión anterior de una propuesta de una extensión de AAC-ELD con respecto a un modo de decodificación con reducción de escala que el decodificador de audio de la Fig. 2 pueden ir acompañados de una herramienta de SBR de bajo retardo. A continuación, se describe, por ejemplo, cómo el codificador AAC-ELD extendido para soportar el modo de operación con reducción de escala propuesto con anterioridad, operaría cuando se utiliza la herramienta de SBR de bajo retardo. Según lo mencionado en la porción introductoria de la memoria descriptiva de la presente solicitud, en el caso de que se utilice la herramienta de SBR de bajo retardo en conexión con el codificador AAC-ELD, los bancos de filtros del módulo de SBR de bajo retardo también están con reducción de escala. Esto asegura que el módulo de SBR opera con la misma resolución de frecuencia y por lo tanto no se requieren más adaptaciones. La Fig. 7 describe el recorrido de la señal del decodificador AAC-ELD que opera a 96 kHz, con un tamaño de trama de 480 muestras, en el modo de SBR de muestreo reducido y con un factor con reducción de escala $\frac{1}{2}$ de 2.

[0085] En la Fig. 7, la corriente de bits que llegan como procesada por una secuencia de bloques, a saber, un decodificador AAC, un bloque de LD-MDCT inversa, un bloque de análisis CLDFB, un decodificador de SBR y un bloque de síntesis CLDFB (CLDFB = banco de filtros complejo de bajo retardo). La corriente de bits es igual a la corriente de datos 24 discutida previamente con respecto a las Figs. 3 a 6, pero está, además, junto con los datos de SBR paramétricos que asisten a la conformación espectral de una réplica espectral de una banda de extensión espectral que extiende la frecuencia de los espectros de la señal de audio obtenida por medio de la decodificación de audio con reducción de escala en la salida del bloque de MDCT inversa bajo retardo, la conformación espectral se lleva a cabo por medio del decodificador de SBR. En particular, el decodificador AAC recupera todos los elementos

de sintaxis necesarios por análisis apropiado y decodificación de entropía. El decodificador AAC puede coincidir parcialmente con el receptor 12 del decodificador de audio 10 que, en la Fig. 7, está realizado por el bloque de MDCT de bajo retardo inverso. En la Fig. 7, F es a modo de ejemplo igual a 2. Esto es, el bloque de retardo MDCT inversa bajo de la salida de la Fig. 7, como un ejemplo para la señal de audio reconstruida 22 de la Fig. 2, una señal de tiempo de 48 kHz submuestreada en la mitad de la velocidad a la que la señal de audio se codificó originalmente en la corriente de bits que llega. El bloque de análisis CLDFB subdivide esta señal de tiempo de 48 kHz, es decir, la señal de audio obtenida por la decodificación de audio con reducción de escala, en N bandas, en este caso $N = 16$, y el decodificador de SBR calcula los coeficientes de reconfiguración de estas bandas, da forma de nuevo a las N bandas, que en consecuencia se controlan a través de los datos de SBR en la corriente de bits de entrada que llegan a la entrada del decodificador AAC, y el bloque de síntesis CLDFB se traslada de nuevo desde el dominio espectral al dominio de tiempo con la obtención, de este modo, de una señal de extensión de alta frecuencia que se va a añadir a las señales de audio originales decodificadas emitidas por el bloque de MDCT de bajo retardo inverso.

[0086] Se debe tener en cuenta que la operación estándar de SBR utiliza un CLDFB de 32 bandas. El algoritmo de interpolación para los coeficientes de ventana c_{32} de CLDFB de 32 bandas ya se da en 4.6.19.4.1 en [1],

$$c_{32}(i) = \frac{1}{2} [c_{64}(2i+1) + c_{64}(2i)], \quad 0 \leq i < 320,$$

donde c_{64} son los coeficientes de ventana de la ventana de 64 bandas dada en la Tabla 4.A.90 en [1]. Esta fórmula se puede generalizar aún más para definir los coeficientes de ventana de un menor número de bandas B también

$$c_B(i) = \frac{1}{2} [c_{64}(2F \cdot i + 1 + p) + c_{64}(2F \cdot i + p)], \quad 0 \leq i < (10B), p = \text{int}\left(\frac{64}{2B} - 0.5\right)$$

donde F denota el factor con reducción de escala que es $F = 32/B$. Con esta definición de los coeficientes de ventana, el banco de filtros de análisis y síntesis de CLDFB se puede describir por completo según lo indicado en el ejemplo anterior de la sección A.2.

[0087] Por lo tanto, los ejemplos anteriores proporcionan algunas definiciones que faltan para el códec AAC-ELD con el fin de adaptar el códec para sistemas con tasas de muestreo más bajas. Estas definiciones pueden incluirse en la norma ISO/IEC 14496-3:2009.

REFERENCIAS

[0088]

[1] ISO/IEC 14496-3: 2009

[2] M13958, "Proposal for an Enhanced Low Delay Coding Mode", octubre de 2006, Hangzhou, China

REIVINDICACIONES

1. Procedimiento para decodificar una señal de audio (22) a una primera tasa de muestreo de una corriente de datos (24) donde la señal de audio está codificada por transformada a una segunda tasa de muestreo, siendo la primera tasa de muestreo $1/F^{\text{va}}$ de la segunda tasa de muestreo, comprendiendo el procedimiento:
 - recibir, por trama de longitud N de la señal de audio, N coeficientes espectrales (28);
 - capturar para cada trama, una fracción de baja frecuencia de longitud N/F de los N coeficientes espectrales (28);
 - llevar a cabo una modulación espectral a tiempo sometiendo, para cada trama (36), la fracción de baja frecuencia a una transformada inversa que tiene funciones de modulación de longitud $(E+2) \cdot N/F$ que se extiende temporalmente sobre la trama respectiva y $E+1$ tramas anteriores con el fin de obtener una porción temporal de longitud $(E+2) \cdot N/F$;
 - formar en ventanas, para cada trama (36), la porción temporal utilizando una ventana de síntesis de longitud $(E+2) \cdot N/F$ que comprende una porción nula de longitud $1/4 \cdot N/F$ en un extremo delantero de la misma y que tiene un pico dentro de un intervalo temporal de la ventana de síntesis, sucediendo el intervalo temporal a la porción nula y que tiene una longitud de $7/4 \cdot N/F$ de tal manera que el formador de ventanas obtenga una porción temporal de ventana de longitud $(E+2) \cdot N/F$; y
 - llevar a cabo una cancelación de solapamiento de dominio de tiempo sometiendo la porción temporal formada en ventana de las tramas a un proceso de superposición y suma de tal manera que una fracción del extremo trasero de longitud $(E+1)/(E+2)$ de la porción temporal de ventana de una trama actual se superponga a un extremo delantero de longitud $(E+1)/(E+2)$ de la porción temporal formada en ventana de una trama anterior, donde la transformada inversa es una MDCT inversa o MDST inversa, y
 - donde la ventana de síntesis es una versión submuestreada de una ventana de síntesis de referencia de longitud $(E+2) \cdot N$, submuestreada por un factor de F por una interpolación segmentaria en segmentos de longitud $1/4 \cdot N$,
 - donde la ventana de síntesis es una concatenación de las funciones de spline de longitud $1/4 \cdot N/F$, y
 - donde la decodificación por entropía se utiliza con el fin de leer los coeficientes espectrales a partir de la corriente de datos y dar forma espectralmente a los coeficientes espectrales con factores de escala proporcionados en la corriente de datos o factores de escala derivados por coeficientes de predicción lineales expresados dentro de la corriente de datos (24).
2. Procedimiento según la reivindicación 1, donde el procedimiento comprende soportar diferentes valores para F .
3. Procedimiento según las reivindicaciones 1 o 2, donde F está entre 1,5 y 10, ambos inclusive.
4. Procedimiento según cualquiera de las reivindicaciones anteriores, donde la ventana de síntesis de referencia es unimodal.
5. Procedimiento según cualquiera de las reivindicaciones anteriores, donde la interpolación se lleva a cabo de tal manera que una mayoría de los coeficientes de la ventana de síntesis depende de más de dos coeficientes de la ventana de síntesis de referencia.
6. Procedimiento según cualquiera de las reivindicaciones anteriores, donde se omite la porción nula al ponderar la porción temporal utilizando la ventana de síntesis y se ignora una porción no ponderada correspondiente de la porción temporal formada en ventana en el proceso de superposición-suma, de modo que solamente se suman $E+1$ porciones temporales formadas en ventana para dar como resultado la porción no ponderada correspondiente de un marco correspondiente y se suman $E+2$ porciones formadas en ventana dentro de un recordatorio del marco correspondiente.
7. Programa informático que tiene un código de programa para llevar a cabo, cuando se ejecuta en un ordenador, un procedimiento según la reivindicación 1.

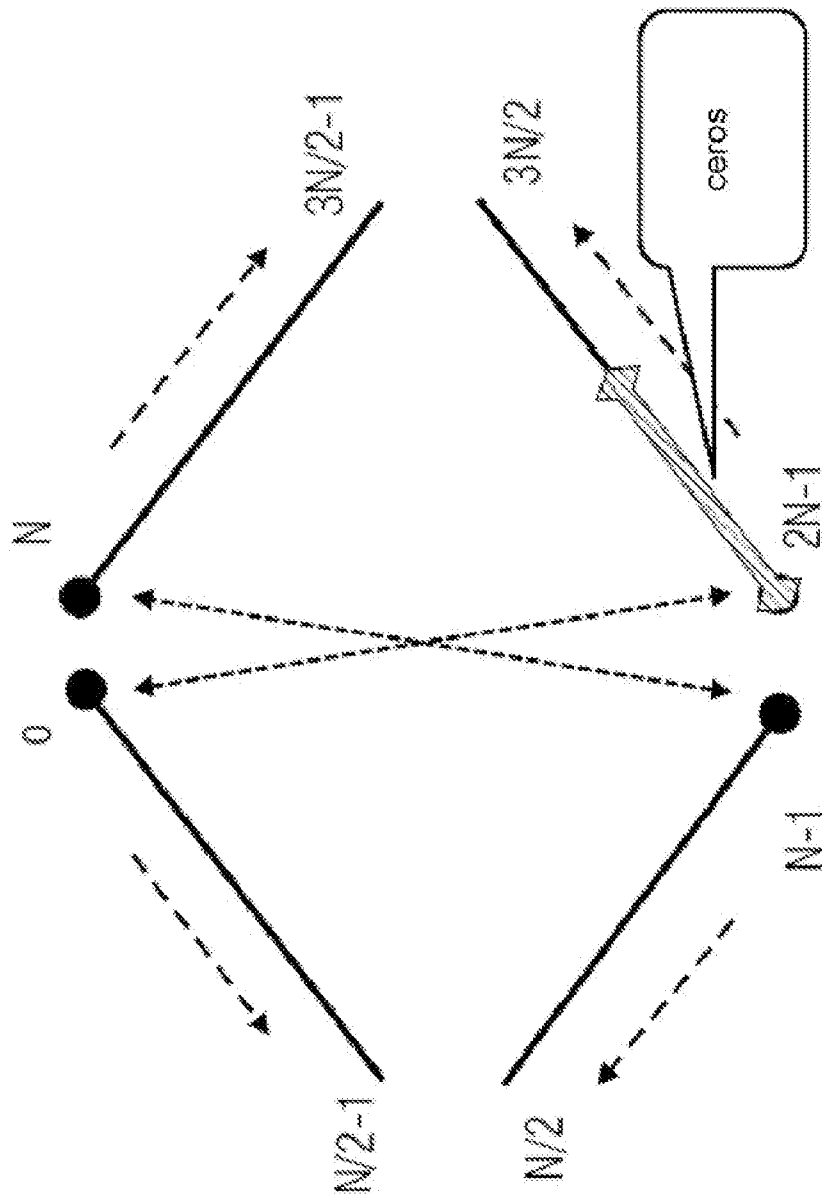
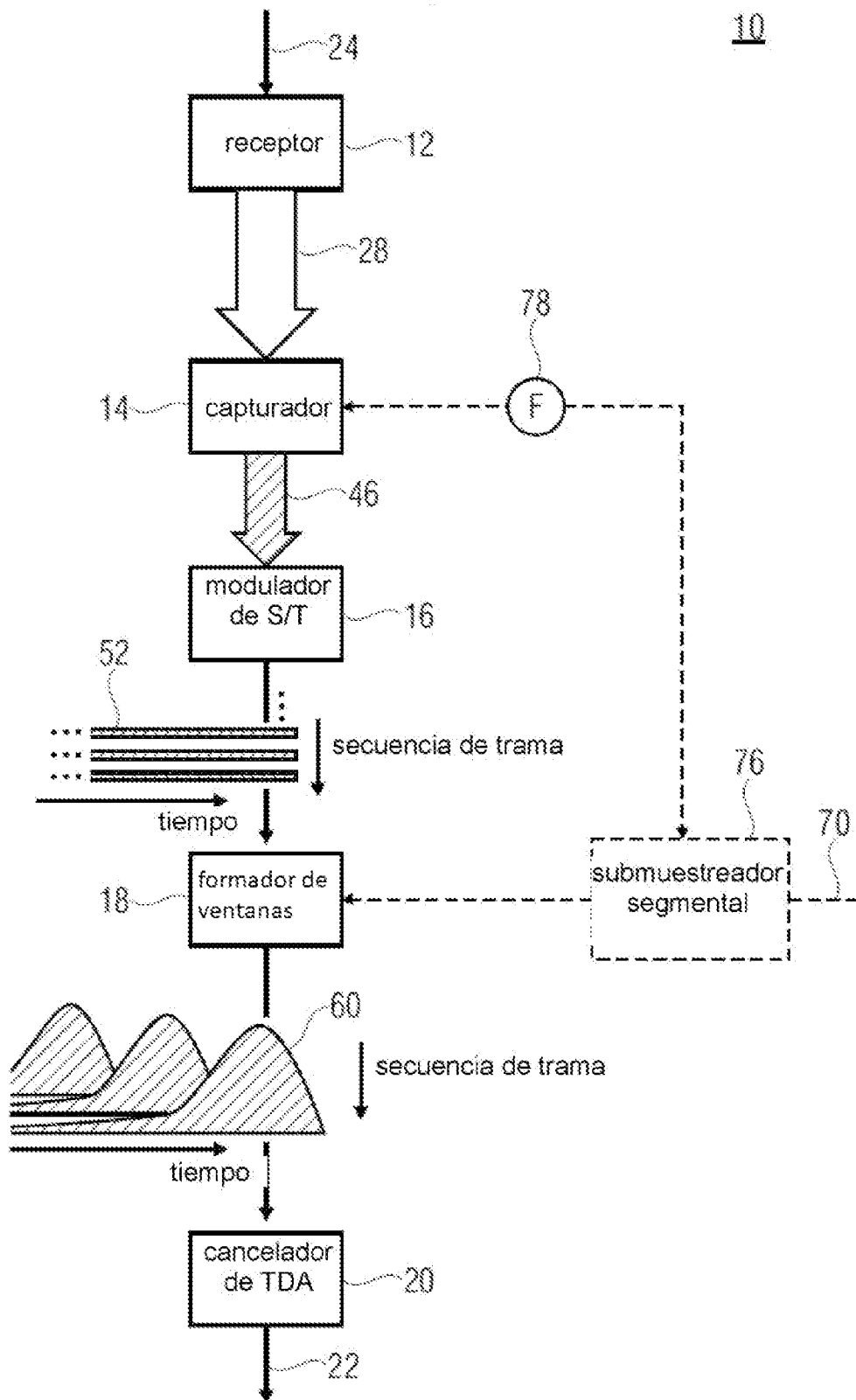


FIG. 1



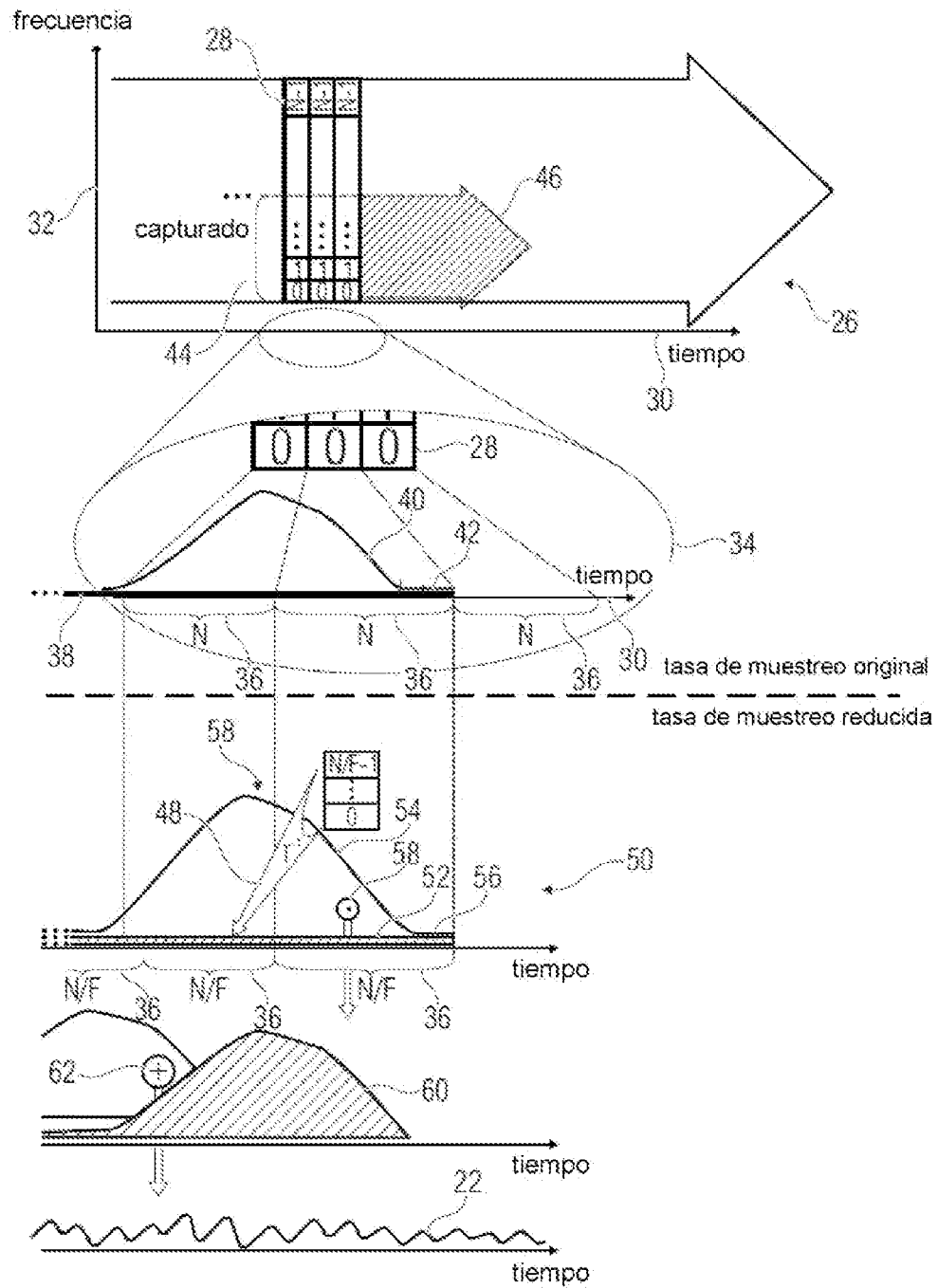


FIG. 3

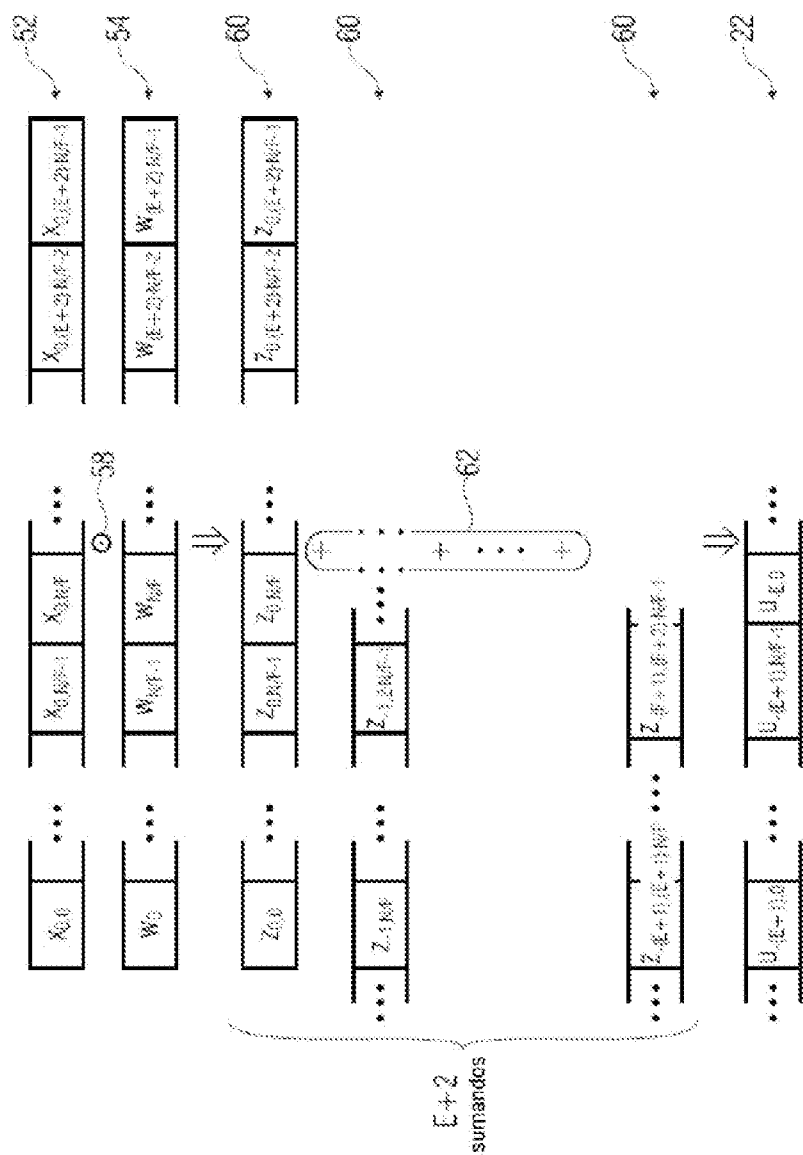


FIG. 4.

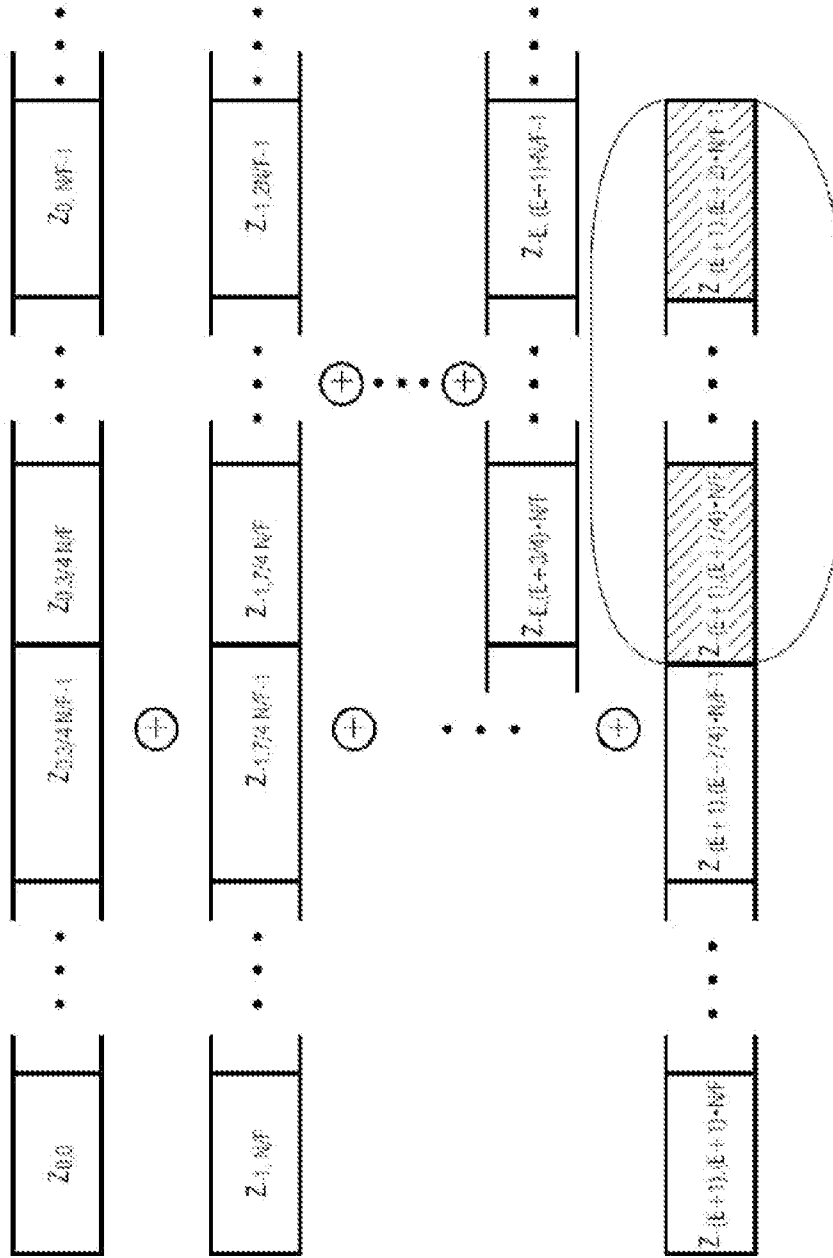


FIG. 5

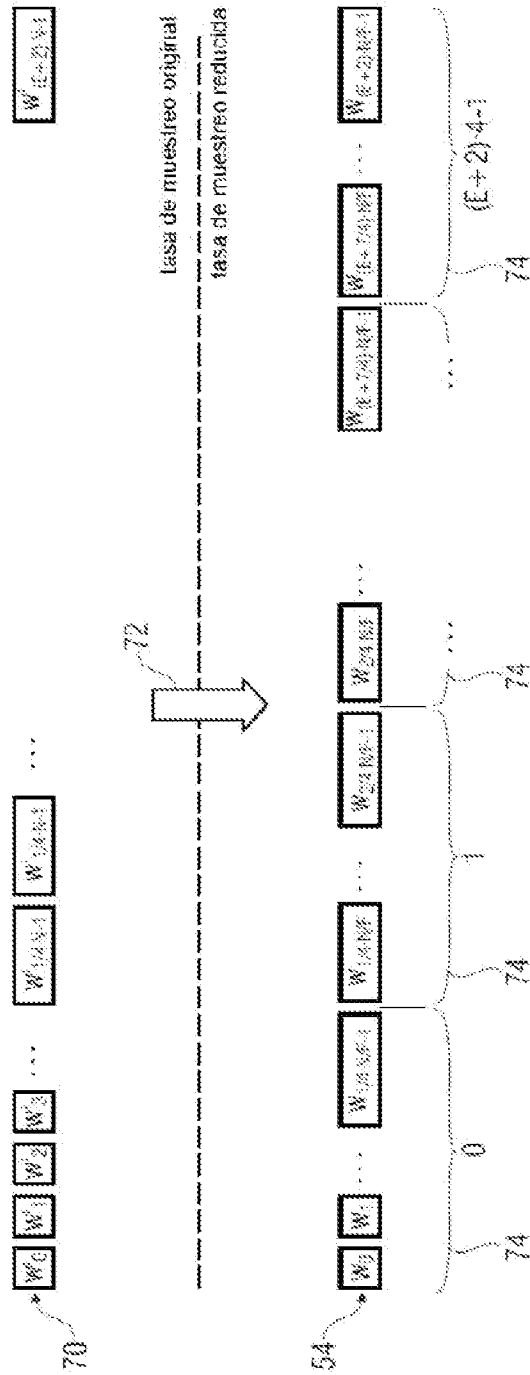


FIG. 6

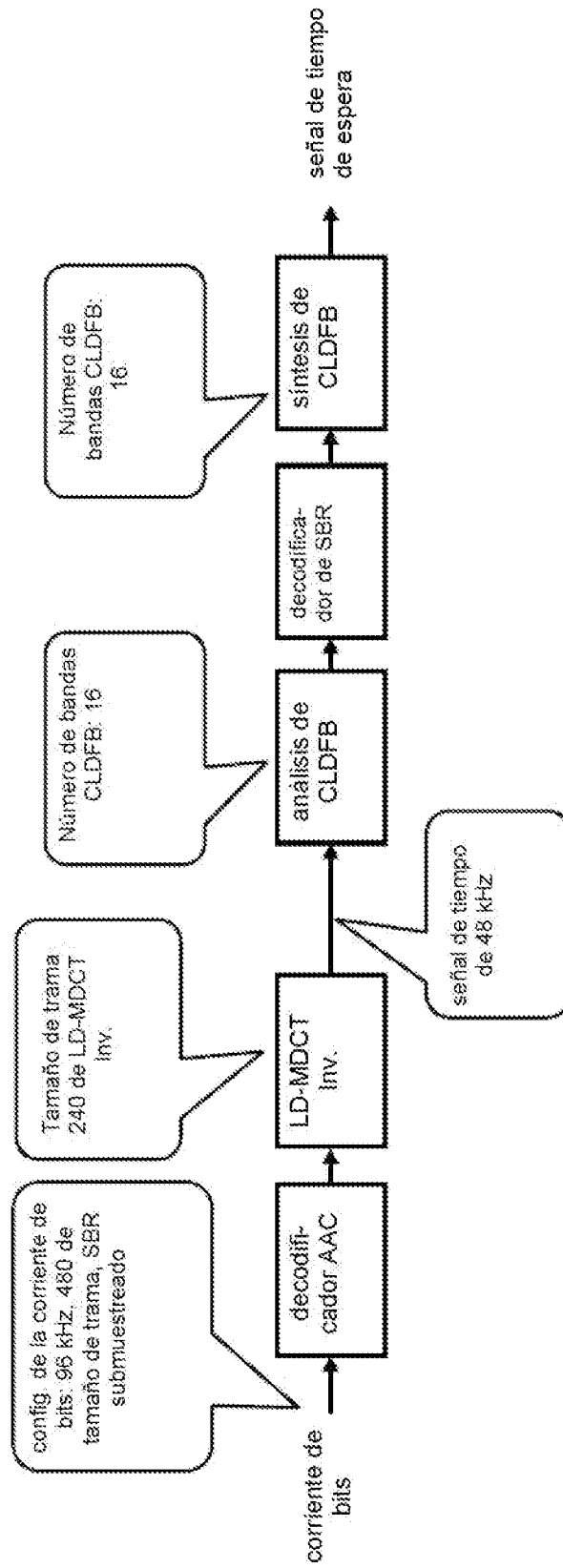
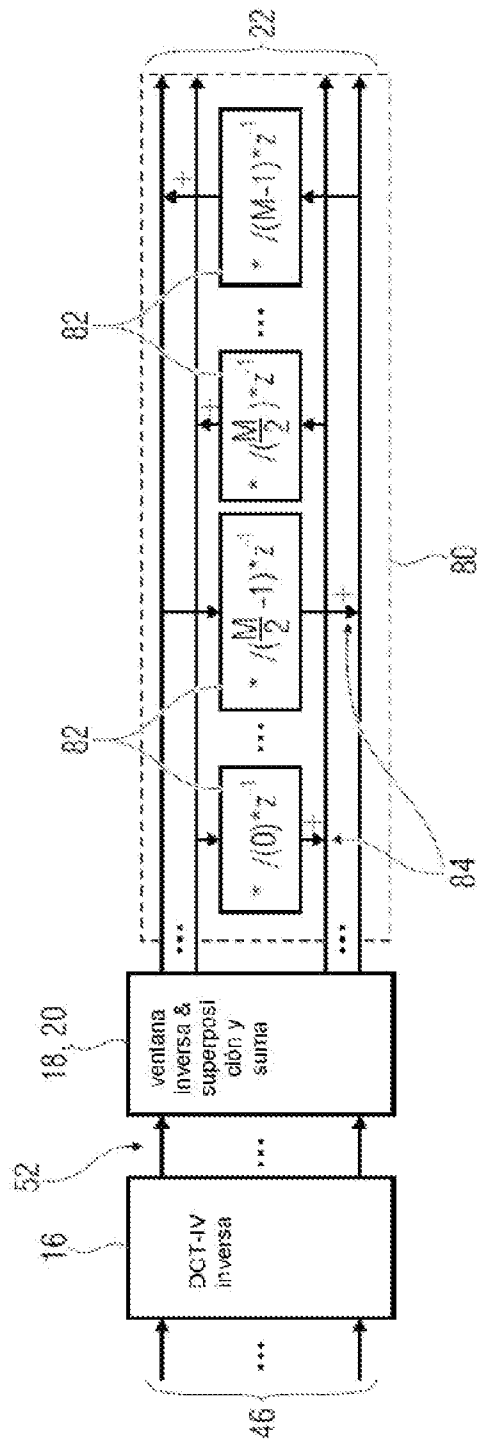


FIG 7



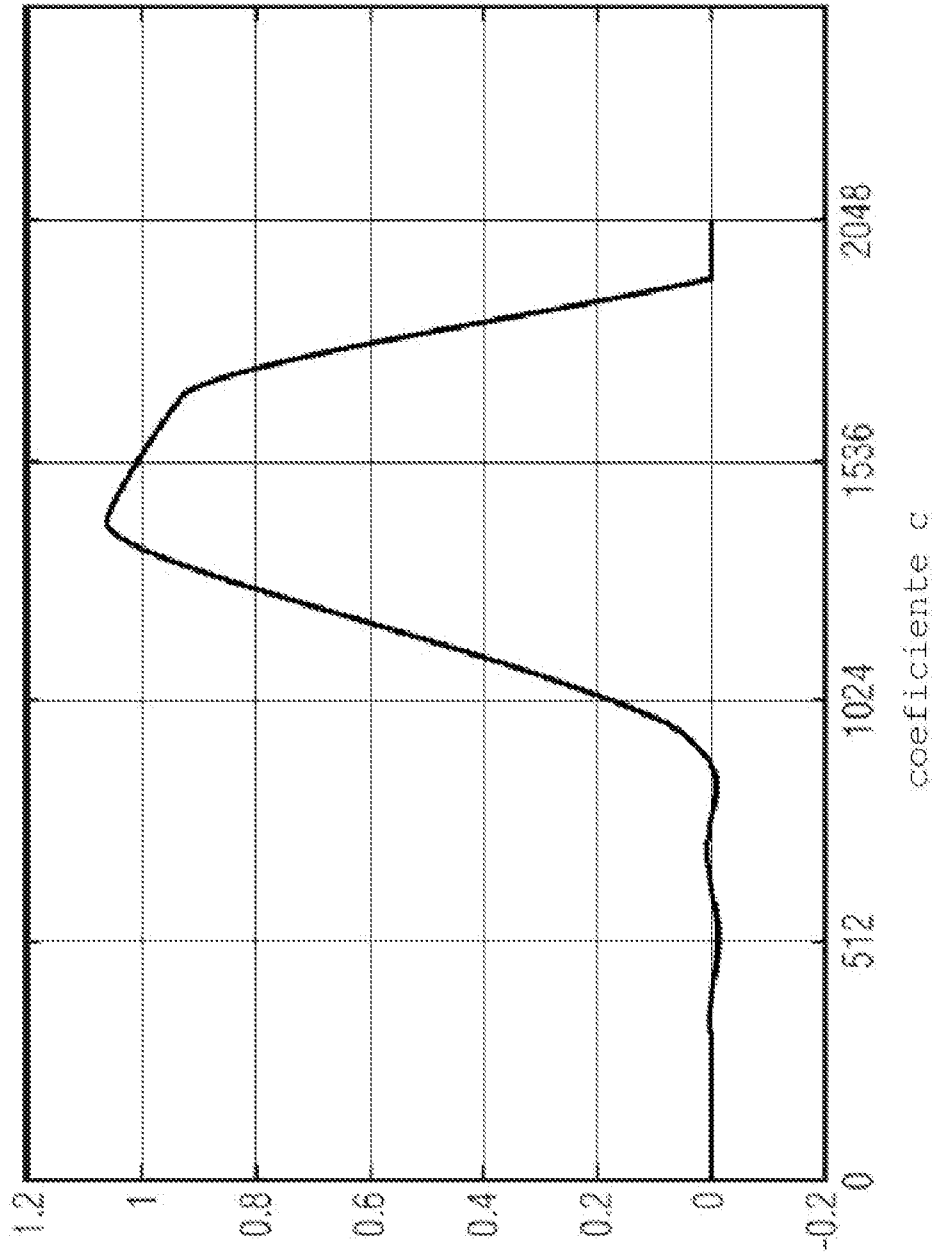


FIG. 9