(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 750 783**

(51) Int. CI.:

G10L 19/005 (2013.01)
G10L 19/02 (2013.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **22.01.2014 E 16183917 (0)**

(97) Fecha y número de publicación de la concesión europea: **17.07.2019 EP 3125239**

(54) Título: **Procedimiento y aparato para controlar la ocultación de pérdida de trama de audio**

(30) Prioridad:

05.02.2013 US 201361761051 P
05.02.2013 US 201361760822 P
05.02.2013 US 201361760814 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
27.03.2020

(73) Titular/es:

TELEFONAKTIEBOLAGET LM ERICSSON (PUBL) (100.0%)
164 83 Stockholm, SE

(72) Inventor/es:

BRUHN, STEFAN y
SVEDBERG, JONAS

(74) Agente/Representante:

LINAGE GONZÁLEZ, Rafael

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

Procedimiento y aparato para controlar la ocultación de pérdida de trama de audio

CAMPO TÉCNICO

La solicitud se refiere a procedimientos y aparatos para controlar un procedimiento de ocultación para una trama de audio perdida de una señal de audio recibida.

ANTECEDENTES

Los sistemas de comunicación de audio convencionales transmiten señales de voz y audio en tramas, lo que significa que el lado remitente primero dispone la señal en segmentos cortos o tramas de, por ejemplo, 20-40 ms que posteriormente se codifican y transmiten como una unidad lógica en, por ejemplo, un paquete de transmisión. El receptor descodifica cada una de estas unidades y reconstruye las tramas de señal correspondientes, que a su vez se emiten finalmente como una secuencia continua de muestras de señales reconstruidas. Antes de la codificación, generalmente existe una etapa de conversión de analógico a digital (A/D) que convierte la señal analógica de voz o de audio de un micrófono en una secuencia de muestras de audio. Por el contrario, en el extremo receptor, típicamente existe una etapa final de conversión D/A que convierte la secuencia de muestras de señales digitales reconstruidas en una señal analógica continua en el tiempo para la reproducción en altavoz.

Sin embargo, dicho sistema de transmisión para señales de voz y audio puede sufrir errores de transmisión, lo que podría dar lugar a una situación en la que una o varias de las tramas transmitidas no están disponibles en el receptor para la reconstrucción. En ese caso, el descodificador tiene que generar una señal de sustitución para cada una de las tramas borradas, es decir, no disponibles. Esto se realiza en la llamada unidad de ocultación de pérdida de trama u ocultación de errores del descodificador de señal del lado del receptor. El objetivo de la ocultación de la pérdida de trama es hacer que la pérdida de trama sea lo más inaudible posible y, por lo tanto, mitigar el impacto de la pérdida de trama en la calidad de la señal reconstruida tanto como sea posible.

Los procedimientos de ocultación de pérdida de trama convencionales pueden depender de la estructura o arquitectura del códec, por ejemplo, aplicando una forma de repetición de los parámetros de códec recibidos previamente. Dichas técnicas de repetición de parámetros dependen claramente de los parámetros específicos del códec usado y, por lo tanto, no son fácilmente aplicables para otros códecs con una estructura diferente. Los procedimientos actuales de ocultación de pérdida de trama pueden, por ejemplo, aplicar el concepto de congelación y extrapolación de parámetros de una trama recibida previamente para generar una trama de sustitución para la trama perdida.

Estos procedimientos de ocultación de pérdida de trama del estado de la técnica incorporan algunos esquemas de manejo de pérdida de ráfaga. En general, después de un número de pérdidas de trama seguidas, la señal sintetizada se atenúa hasta que se silencia por completo después de largas ráfagas de errores. Además, los parámetros de codificación que se repiten y extrapolan esencialmente se modifican de modo que se logra la atenuación y que los picos espectrales se aplanan.

Las técnicas de ocultación de pérdida de trama del estado de la técnica actuales típicamente aplican el concepto de congelación y extrapolación de parámetros de una trama recibida previamente para generar una trama de sustitución para la trama perdida. Muchos códecs de voz paramétricos, tales como los códecs predictivos lineales como AMR o AMR-WB, típicamente congelan los parámetros recibidos anteriormente o usan alguna extrapolación de los mismos y usan el descodificador con ellos. En esencia, el principio es tener un modelo dado para codificar/descodificar y aplicar el mismo modelo con parámetros congelados o extrapolados. Las técnicas de ocultación de pérdida de trama de AMR y AMR-WB se pueden considerar representativas. Se especifican en detalle en las especificaciones de las normas correspondientes.

Se aplican muchos códecs de la clase de códecs de audio para codificar técnicas de dominio de frecuencia. Esto significa que después de alguna transformación del dominio de frecuencia, se aplica un modelo de codificación en los parámetros espectrales. El descodificador reconstruye el espectro de señal a partir de los parámetros recibidos y finalmente transforma el espectro de nuevo en una señal de tiempo. Típicamente, la señal de tiempo se reconstruye trama por trama. Dichas tramas se combinan mediante técnicas de solapamiento y suma a la señal reconstruida final. Incluso en el caso de los códecs de audio, la ocultación de errores del estado de la técnica típicamente aplica el mismo modelo de descodificación para tramas perdidos o al menos uno similar. Los parámetros de dominio de frecuencia de una trama recibida previamente se congelan o se extrapolan adecuadamente y a continuación se usan en la conversión de dominio de frecuencia a tiempo. Se proporcionan ejemplos de dichas técnicas con los códecs de audio 3GPP de acuerdo con las normas 3GPP.

El documento US 2004/122680 describe un sistema para la ocultación de errores de trama que enseña a ajustar la magnitud de la trama de sustitución de acuerdo con el número de tramas perdidas consecutivas.

El documento EP 1 722 359 A1 describe un procedimiento de ocultación que incluye detección de transitorios.

SUMARIO

Las soluciones para el ocultación de pérdida de trama del estado de la técnica actuales típicamente sufren deficiencias de calidad. El problema principal es que la técnica de congelación y extrapolación de parámetros y la nueva aplicación del mismo modelo de descodificador, incluso para tramas perdidas, no siempre garantizan una evolución de señal suave y fiel desde las tramas de señal previamente descodificadas hasta la trama perdida. Típicamente, esto da lugar a discontinuidades de señal audibles con el correspondiente impacto en la calidad.

Se describen nuevos esquemas para el ocultación de pérdida de trama para sistemas de transmisión de audio y voz. Los nuevos esquemas mejoran la calidad en caso de pérdida de tramas respecto a la calidad que se puede lograr con las técnicas de ocultación de pérdida de trama de la técnica anterior.

El objetivo de los presentes modos de realización es controlar un esquema de ocultación de pérdida de trama que preferentemente sea del tipo de los nuevos procedimientos relacionados descritos de modo que se logre la mejor calidad de sonido posible de la señal reconstruida. Los modos de realización tienen como objetivo optimizar esta calidad de reconstrucción tanto con respecto a las propiedades de la señal como a la distribución temporal de las pérdidas de trama. Son particularmente problemáticos para la ocultación de pérdida de trama para proporcionar una buena calidad los casos en que la señal de audio tiene propiedades fuertemente variables, tales como inicios o desviaciones de la energía o si es espectralmente muy fluctuante. En ese caso, los procedimientos de ocultación descritos pueden repetir el inicio, la desviación o la fluctuación espectral que da lugar a grandes desviaciones de la señal original y la pérdida de calidad correspondiente.

Otro caso problemático es si se producen ráfagas de pérdida de trama seguidas. Conceptualmente, el esquema para la ocultación de pérdida de trama de acuerdo con los procedimientos descritos puede hacer frente a dichos casos, aunque resulta que todavía se pueden producir artefactos tonales molestos. Es otro objetivo de los presentes modos de realización mitigar dichos artefactos en el grado más alto posible.

De acuerdo con un primer aspecto, se divulga un procedimiento de ocultación de pérdida de trama de acuerdo con la reivindicación 1.

De acuerdo con un segundo aspecto, se describe un aparato para crear un trama de sustitución para una trama de audio perdida de acuerdo con la reivindicación 9.

De acuerdo con un tercer aspecto, se define un programa informático para ocultar una trama de audio perdida, y el programa informático comprende instrucciones que, cuando son ejecutadas por un procesador, provocan que el procesador oculte una trama de audio perdida, de acuerdo con el primer aspecto descrito anteriormente.

De acuerdo con un cuarto aspecto, un producto de programa informático comprende un medio legible por ordenador que almacena un programa informático de acuerdo con el tercer aspecto descrito anteriormente.

Una ventaja con un modo de realización aborda el control de adaptaciones de procedimientos de ocultación de pérdida de trama que permite mitigar el impacto audible de la pérdida de trama en la transmisión de señales de voz y audio codificadas aún más con respecto a la calidad lograda solo con los procedimientos de ocultación descritos. El beneficio general de los modos de realización es proporcionar una evolución suave y fiel de la señal reconstruida incluso para tramas perdidas. El impacto audible de las pérdidas de trama se reduce en gran medida en comparación con el uso de técnicas del estado de la técnica.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

Todas las siguientes referencias a la expresión "modo(s) de realización", si se refieren a combinaciones de características diferentes de las definidas por las reivindicaciones independientes, se refieren a ejemplos que se presentaron originalmente pero que no representan modos de realización de la invención reivindicada actualmente; estos ejemplos todavía se muestran solo con propósitos ilustrativos.

Para una comprensión más completa de los modos de realización de ejemplo en la presente invención, ahora se hace referencia a la siguiente descripción tomada en relación con los dibujos adjuntos, en los que:

la figura 1 muestra una función de ventana rectangular.

La figura 2 muestra una combinación de la ventana de Hamming con la ventana rectangular.

La Figura 3 muestra un ejemplo de un espectro de magnitud de una función de ventana.

La figura 4 ilustra un espectro de línea de una señal sinusoidal ejemplar con la frecuencia f_k.

La figura 5 muestra un espectro de una señal sinusoidal en ventana con la frecuencia f_k.

La figura 6 ilustra las barras correspondientes a la magnitud de los puntos de la cuadrícula de una DFT, basadas en una trama de análisis.

La figura 7 ilustra un ajuste de parábola a través de los puntos de cuadrícula de DFT P1, P2 y P3.

La figura 8 ilustra un ajuste de un lóbulo principal de un espectro de ventana.

La figura 9 ilustra un ajuste de la función de aproximación del lóbulo principal P a través de los puntos de cuadrícula de DFT P1 y P2.

La figura 10 es un diagrama de flujo que ilustra un procedimiento de ejemplo de acuerdo con modos de realización de la invención para controlar un procedimiento de ocultación para una trama de audio perdida de una señal de audio recibida.

La figura 11 es un diagrama de flujo que ilustra otro procedimiento de ejemplo de acuerdo con modos de realización de la invención para controlar un procedimiento de ocultación para una trama de audio perdida de una señal de audio recibida.

La figura 12 ilustra otro modo de realización de ejemplo de la invención.

La figura 13 muestra un ejemplo de un aparato de acuerdo con un modo de realización de la invención.

La figura 14 muestra otro ejemplo de un aparato de acuerdo con un modo de realización de la invención.

La figura 15 muestra otro ejemplo de un aparato de acuerdo con un modo de realización de la invención.

DESCRIPCIÓN DETALLADA

El nuevo esquema de control para las nuevas técnicas de ocultación de pérdida de trama descritas involucra las siguientes etapas como se muestra en la figura 10. Cabe señalar que el procedimiento se puede implementar en un controlador en un descodificador.

1. Detectar condiciones en las propiedades de la señal de audio recibida y reconstruida previamente o en las propiedades estadísticas de las pérdidas de trama observadas para las cuales la sustitución de una trama perdida de acuerdo con los procedimientos descritos proporciona una calidad relativamente reducida, **101.**

2. En caso de que se detecte dicha condición en la etapa 1, modificar el elemento de los procedimientos de acuerdo con el cual se calcula el espectro de la trama de sustitución mediante $Z(m) = Y(m) \cdot e^{j\theta_k}$ ajustando selectivamente las fases o las magnitudes del espectro, **102.**

Análisis sinusoidal

Una primera etapa de la técnica de ocultación de pérdida de trama a la que se puede aplicar la nueva técnica de control implica un análisis sinusoidal de una parte de la señal recibida previamente. El propósito de este análisis sinusoidal es encontrar las frecuencias de las sinusoides principales de esa señal, y la suposición subyacente es que la señal está compuesta de un número limitado de sinusoides individuales, es decir, que es una señal de seno múltiple del tipo siguiente:

$$s(n) = \sum_{k=1}^{K} a_k \cdot \cos(2\pi \frac{f_k}{f_s} \cdot n + \varphi_k) .$$

En esta ecuación, K es el número de sinusoides en los que se supone que consiste la señal. Para cada una de las sinusoides con índice k = 1...K, a_k es la amplitud, f_k es la frecuencia y φ_k es la fase. La frecuencia de muestreo se denomina por f_s y el índice de tiempo de muestras de señal discreta de tiempo $s(n)$ por n.

Es de suma importancia encontrar frecuencias de las sinusoides lo más exactas posibles. Mientras que una señal sinusoidal ideal tendría un espectro de línea con frecuencias de línea f_k, encontrar sus valores verdaderos en principio requeriría un tiempo de medición infinito. Por lo tanto, en la práctica es difícil encontrar estas frecuencias, ya que solo se pueden estimar en base a un período de medición corto, que corresponde al segmento de señal usado para el análisis sinusoidal descrito en el presente documento; este segmento de señal se denomina a continuación en el presente documento trama de análisis. Otra dificultad es que, en la práctica, la

señal puede ser variable en el tiempo, lo que significa que los parámetros de la ecuación anterior varían con el tiempo. Por lo tanto, por un lado, es deseable usar una trama de análisis larga que haga la medición más exacta; por otro lado, se necesitaría un período de medición corto para hacer frente mejor a las posibles variaciones de señal. Una buena compensación es usar una longitud de trama de análisis del orden de, por ejemplo, 20-40 ms.

Una posibilidad preferente para identificar las frecuencias de las sinusoides f_k es hacer un análisis de dominio de frecuencia de la trama de análisis. Para este fin, la trama de análisis se transforma en el dominio de frecuencia, por ejemplo, por medio de DFT o DCT o transformadas de dominio de frecuencia similares. En caso de que se use una DFT de la trama de análisis, el espectro viene dado por:

$$X(m) = DFT(w(n) \cdot x(n)) = \sum_{n=0}^{L-1} e^{-j\frac{2\pi}{L}mn} \cdot w(n) \cdot x(n) .$$

En esta ecuación w(n) indica la función de ventana con la que se extrae y pondera la trama de análisis de longitud *L*. Las funciones de ventana típicas son, por ejemplo, ventanas rectangulares que son iguales a 1 para n $\in$ [0...*L*-1] y de otro modo 0 como se muestra en la figura 1. Aquí se supone que los índices de tiempo de la señal de audio recibida previamente se establecen de modo que la trama de análisis esté referenciada por los índices de tiempo *n*= 0...*L*-1. Otras funciones de ventana que pueden ser más adecuadas para el análisis espectral son, por ejemplo, la ventana de Hamming, la ventana de Hanning, la ventana de Kaiser o la ventana de Blackman. Una función de ventana que se encuentra particularmente útil es una combinación de la ventana de Hamming con la ventana rectangular. Esta ventana tiene una conformación de borde ascendente como la mitad izquierda de una ventana de Hamming de longitud *L*1 y una conformación de borde descendente como la mitad derecha de una ventana de Hamming de longitud *L*1 y entre los bordes ascendente y descendente la ventana es igual a 1 para la longitud de *L*-*L* 1, como se muestra en la figura 2.

Los picos del espectro de magnitud de la trama de análisis en ventana |*X*(*m*)| constituyen una aproximación de las frecuencias sinusoidales requeridas f_k. Sin embargo, la exactitud de esta aproximación está limitada por el espaciado de frecuencia de la DFT. Con la DFT con longitud de bloque L, la exactitud se limita a $\frac{f_s}{2L}$.

Los experimentos muestran que este nivel de exactitud puede ser demasiado bajo en el alcance de los procedimientos descrito en el presente documento. Se puede obtener una exactitud mejorada en base a los resultados de la siguiente consideración:

El espectro de la trama de análisis en ventana viene dado por la convolución del espectro de la función de ventana con el espectro de línea de la señal de modelo sinusoidal *S*(Ω), posteriormente muestreado en los puntos de cuadrícula de la DFT:

$$X(m) = \int_{2\pi} \delta(\Omega - m \cdot \tfrac{2\pi}{L}) \cdot \left(W(\Omega) * S(\Omega)\right) \cdot d\Omega .$$

Usando la expresión espectral de la señal de modelo sinusoidal, esto se puede escribir como

$$X(m) = \tfrac{1}{2} \int_{2\pi} \delta(\Omega - m \cdot \tfrac{2\pi}{L}) \cdot \sum_{k\ 1}^{K} a_k \cdot \left((W(\Omega + 2\pi \frac{f_k}{f_s}) \cdot e^{\ j\varphi_k} + W(\Omega - 2\pi \frac{f_k}{f_s}) \cdot e^{j\varphi_k} \right) \cdot d\Omega .$$

Por lo tanto, el espectro muestreado viene dado por

$$X(m) = \tfrac{1}{2} \sum_{k-1}^{K} a_k \cdot \left((W(2\pi(\frac{m}{L} + \frac{f_k}{f_s})) \cdot e^{-j\varphi_k} + W(2\pi(\frac{m}{L} - \frac{f_k}{f_s})) \cdot e^{j\varphi_k}) \right),$$

con

m=0...L-1.

En base a esta consideración, se supone que los picos observados en el espectro de magnitud de la trama de análisis provienen de una señal sinusoidal en ventana con *K* sinusoides donde las frecuencias sinusoides verdaderas se encuentran en la vecindad de los picos.

Sea m_k el índice de DFT (punto de cuadrícula) del pico $k^{ésimo}$ observado, luego la frecuencia correspondiente es $\hat{f}_k = \frac{m_k}{L} \cdot f_s$ que se puede considerar una aproximación de la frecuencia sinusoidal verdadera f_k. Se puede suponer que la frecuencia sinusoide verdadera f_k se encuentra dentro del intervalo

$$\left[\left(m_k - 1/2\right) \cdot \frac{f_s}{L}, \left(m_k + 1/2\right) \cdot \frac{f_s}{L} \right].$$

Para mayor claridad, cabe señalar que la convolución del espectro de la función de ventana con el espectro del espectro de línea de la señal de modelo sinusoidal se puede entender como una superposición de versiones desplazadas en frecuencia del espectro de función de ventana, con lo que las frecuencias de desplazamiento son las frecuencias de las sinusoides. Esta superposición se muestrea a continuación en los puntos de la cuadrícula de DFT. Estas etapas se ilustran en las siguientes figuras. La figura 3 muestra un ejemplo del espectro de magnitud de una función de ventana. La figura 4 muestra el espectro de magnitud (espectro de línea) de una señal sinusoidal de ejemplo con una sola sinusoide de frecuencia. La figura 5 muestra el espectro de magnitud de la señal sinusoidal en ventana que replica y superpone los espectros de ventana desplazados en frecuencia a las frecuencias de la sinusoide. Las barras en la figura 6 corresponden a la magnitud de los puntos de la cuadrícula de la DFT de la sinusoide en ventana que se obtienen calculando la DFT de la trama de análisis. Cabe señalar que todos los espectros son periódicos con el parámetro de frecuencia normalizada Ω donde Ω = 2 π que corresponde a la frecuencia de muestreo f_s.

El análisis previo y la ilustración de la figura 6 sugieren que una mejor aproximación de las frecuencias sinusoidales verdaderas solo se puede encontrar incrementando la resolución de la búsqueda con respecto a la resolución de frecuencia de la transformada de dominio de frecuencia usada.

Una forma preferente de encontrar mejores aproximaciones de las frecuencias f_k de las sinusoides es aplicar la interpolación parabólica. Un enfoque de este tipo es ajustar las parábolas a través de los puntos de cuadrícula del espectro de magnitud de DFT que rodean los picos y calcular las frecuencias respectivas que pertenecen a la parábola máxima. Una elección adecuada para el orden de las parábolas es 2. En detalle se puede aplicar el siguiente procedimiento:

1. Identificar los picos de la DFT de la trama de análisis en ventana. La búsqueda de picos entregará el número de picos K y los índices de DFT correspondientes de los picos. La búsqueda de picos típicamente se puede hacer en el espectro de magnitud de DFT o en el espectro de magnitud de DFT logarítmico.

2. Para cada pico k (con k = 1...K) con el correspondiente índice de DFT m_k, ajustar una parábola a través de los tres puntos {*P1; P2; P3*} = {(m_k - 1, log(|X(m_k - 1)|); (m_k, log(|X(m_k)|); (m_k +1, log(|X(m_k + 1)|)}. Esto da como resultado coeficientes de parábola $b_k(0)$, $b_k(1)$, $b_k(2)$ de la parábola definida por

$$p_k(q) = \sum_{i=0}^{2} b_k(i) \cdot q^i .$$

Este ajuste de parábola se ilustra en la figura 7.

3. Para cada una de las parábolas K, calcular el índice de frecuencia interpolado $\hat{m_k}$ correspondiente al valor de q para el cual la parábola tiene su máximo. Usar $f_k = \hat{m}_k \cdot f_s / L$ como aproximación para la frecuencia sinusoide f_k

El enfoque descrito proporciona buenos resultados, pero puede tener algunas limitaciones, ya que las parábolas no se aproximan a la conformación del lóbulo principal del espectro de magnitud |W(Ω)| de la función de ventana. Un esquema alternativo que hace esto es una estimación de frecuencia perfeccionada usando una aproximación del lóbulo principal, que se describe como sigue. La idea principal de esta alternativa es ajustar una función *P (q),* que se aproxima al lóbulo principal de $|W(\frac{2\pi}{L} \cdot q)|$, a través de los puntos de cuadrícula del espectro de magnitud de DFT que rodean los picos y calcular las frecuencias respectivas que pertenecen a la función máxima. La función *P (q)* podría ser idéntica al espectro de magnitud desplazado en frecuencia $|W(\frac{2\pi}{L} \cdot (q - \hat{q}))|$ de la función de ventana. Sin embargo, por simplicidad numérica, debería ser más bien, por ejemplo, un polinomio que permita el cálculo sencillo de la función máxima. Se puede aplicar el siguiente procedimiento detallado:

1. Identificar los picos de la DFT de la trama de análisis en ventana. La búsqueda de picos entregará el número de picos *K* y los índices de DFT correspondientes de los picos. La búsqueda de picos típicamente se puede hacer en el espectro de magnitud de DFT o en el espectro de magnitud de DFT logarítmico.

2. Derivar la función *P (q)* que se aproxima al espectro de magnitud $|W(\frac{2\pi}{L}\cdot q)|$ de la función de ventana o del espectro de magnitud logarítmica $\log|W(\frac{2\pi}{L}\cdot q)|$ para un intervalo dado (q_1, q_2). La figura 8 ilustra la elección de la función de aproximación que se aproxima al lóbulo principal del espectro de ventana.

3. Para cada pico *k* (con *k* = 1...*K*) con el índice de DFT correspondiente m_k, ajustar la función desplazada en frecuencia $P(q-\hat{q}_k)$ a través de los dos puntos de cuadrícula de DFT que rodean el pico verdadero esperado del espectro continuo de la señal sinusoidal en ventana. Por lo tanto, si $|X(m_k - 1)|$ es mayor que $|X(m_k + 1)|$ ajustar $P(q - \hat{q}_k)$ a través de los puntos $\{P_1; P_2\} = \{(m_k\text{-}1, \log(|X(m_k\text{-}1)|); (m_k, \log(|X(m_k)|)\}$ y de otro modo a través de los puntos $\{P_1 ; P_2\} = \{(m_k, \log(|X(m_k)|); (m_k +1, \log(|X(m_k+1)|)\}$. *P (q)* se puede elegir, por simplicidad, como un polinomio de orden 2 o 4. Esto hace que la aproximación en la etapa 2 sea un cálculo de regresión lineal simple y el cálculo sencillo de $\hat{q}_k$. El intervalo (q_1,q_2) se puede elegir para que sea fijo e idéntico para todos los picos, por ejemplo, (Q_1,q_2) = (-1,1) o adaptativo.

En el enfoque adaptativo, el intervalo se puede elegir de modo que la función $P(q-\hat{q}_k)$ se ajuste al lóbulo principal del espectro de la función de ventana en el intervalo de los puntos de cuadrícula de DFT relevantes $\{P_1; P_2\}$. El proceso de ajuste se visualiza en la figura 9.

4. Para cada uno de los parámetros de desplazamiento de frecuencia *K* $\hat{q}_k$ para los cuales se espera que el espectro continuo de la señal sinusoidal en ventana tenga su pico de cálculo $\hat{f}_k = \hat{q}_k \cdot f_s/L$ como aproximación para la frecuencia sinusoide f_k.

Existen muchos casos en los que la señal transmitida es armónica, lo que significa que la señal consiste en ondas sinusoidales cuyas frecuencias son múltiplos de enteros de alguna frecuencia fundamental f_0. Este es el caso cuando la señal es muy periódica, como por ejemplo para voz sonora o los tonos sostenidos de algún instrumento musical. Esto significa que las frecuencias del modelo sinusoidal de los modos de realización no son independientes, sino que tienen una relación armónica y provienen de la misma frecuencia fundamental. Tener en cuenta esta propiedad armónica puede, en consecuencia, mejorar sustancialmente el análisis de las frecuencias de componente sinusoidal.

Una posibilidad de perfeccionamiento se explica como sigue:

1. Verificar si la señal es armónica. Esto se puede hacer, por ejemplo, evaluando la periodicidad de la señal antes de la pérdida de trama. Un procedimiento sencillo es realizar un análisis de autocorrelación de la señal. El máximo de dicha función de autocorrelación para algún desfase temporal $\tau > 0$ se puede usar como un indicador. Si el valor de este máximo excede un umbral dado, la señal se puede considerar armónica. El desfase temporal τ correspondiente corresponde al período de la señal que está relacionado con la frecuencia fundamental a través de $f_0 = \frac{f_s}{\tau}$. Muchos procedimientos de codificación de voz predictiva lineal aplican la llamada predicción de tono de bucle abierto o cerrado o codificación CELP usando libros de códigos adaptativos. La ganancia de tono y los parámetros de desfase de tono asociados derivados de dichos procedimientos de codificación también son indicadores útiles si la señal es armónica y, respectivamente, para el desfase temporal. A continuación se describe otro procedimiento para obtener f_0.

2. Para cada índice armónico *j* dentro del intervalo de enteros 1...J_{max} verificar si existe un pico en el espectro de magnitud de DFT (logarítmico) de la trama de análisis en la vecindad de la frecuencia armónica $f_j = j \cdot f_0$. La vecindad de f_j se puede definir como el intervalo delta alrededor de f_j donde delta corresponde a la resolución de frecuencia de la DFT $\frac{f_s}{L}$, es decir, el intervalo $\left[j\cdot f_0 - \frac{f_s}{2\cdot L}, j\cdot f_0 + \frac{f_s}{2\cdot L}\right]$.

En caso de estar presente un pico de este tipo con la frecuencia sinusoidal estimada correspondiente f_k, sustituir f_k por

$$f_k = j \cdot f_0.$$

Para el procedimiento de dos etapas dado anteriormente también existe la posibilidad de verificar si la señal es armónica y la derivación de la frecuencia fundamental implícitamente y posiblemente de manera iterativa sin

necesariamente usar indicadores de algún procedimiento separado. Un ejemplo para una técnica de este tipo se da como sigue: Para cada $f_{0,p}$ de un conjunto de valores candidatos $\{f_{0,1}... f_{0,P}\}$ aplicar la etapa de procedimiento 2, aunque sin sustituir f_k pero contando cuántos picos de DFT están presentes en la vecindad de las frecuencias armónicas, es decir, los múltiplos de enteros de $f_{0,p}$. Identificar la frecuencia fundamental $f_{0,\,p\max}$ para la cual se obtiene el mayor número de picos en o alrededor de las frecuencias armónicas. Si este mayor número de picos excede un umbral dado, se supone que la señal es armónica. En ese caso, se puede suponer que $f_{0,p\max}$ es la frecuencia fundamental con la que se ejecuta a continuación la etapa 2, dando lugar a frecuencias sinusoidales perfeccionadas f_k. Sin embargo, una alternativa más preferente es primero optimizar la frecuencia fundamental f_0 en base a las frecuencias pico f_k que se ha encontrado que coinciden con frecuencias armónicas. Suponer un conjunto de armónicos M, es decir, múltiplos de enteros $\{n_1... n_M\}$ de alguna frecuencia fundamental que se ha encontrado que coincide con algún conjunto de picos espectrales M en frecuencias $f_{k(m)}$, m = 1...M, luego la frecuencia fundamental subyacente (optimizada) $f_{0,opt}$ se puede calcular para minimizar el error entre las frecuencias armónicas y las frecuencias de pico espectral. Si el error que se va a minimizar es el error cuadrático medio $E_2 = \sum_{m\ 1}^{M}(n_m \cdot f_0 - \hat{f}_{k(m)})^2$, la frecuencia fundamental óptima se calcula como

$$f_{0.opt} = \frac{\sum_{m=1}^{M} n_m \cdot \hat{f}_{k(m)}}{\sum_{m\ 1}^{M} n_m^2}.$$

El conjunto inicial de valores candidatos $\{f_{0,1}... f_{0,P}\}$ se puede obtener de las frecuencias de los picos DFT o de las frecuencias sinusoidales estimadas f_k.

Otra posibilidad para mejorar la exactitud de las frecuencias sinusoidales estimadas f_k es considerar su evolución temporal. Con ese fin, las estimaciones de las frecuencias sinusoidales de un múltiplo de tramas de análisis se pueden combinar, por ejemplo, por medio de promediados o predicciones. Antes del promediado o la predicción, se puede aplicar un rastreo de picos que conecte los picos espectrales estimados con los mismos sinusoides subyacentes respectivos.

Aplicando el modelo sinusoidal

La aplicación de un modelo sinusoidal para realizar una operación de ocultación de pérdida de trama descrita en el presente documento se puede describir como sigue.

Se supone que un segmento dado de la señal codificada no se puede reconstruir por el descodificador ya que la información codificada correspondiente no está disponible. Se supone además que una parte de la señal anterior a este segmento está disponible. Sea $y(n)$ con n = 0...N-1 el segmento no disponible para el que se debe generar una trama de sustitución $z(n)$ e $y(n)$ con n<0 la señal descodificada previamente disponible. A continuación, en una primer etapa, una trama prototipo de la señal disponible de longitud L y el índice de inicio n_{-1} se extrae con una función de ventana $w(n)$ y se transforma en dominio de frecuencia, por ejemplo, por medio de DFT:

$$Y_{-1}(m) = \sum_{n=0}^{L-1} y(n - n_{-1}) \cdot w(n) \cdot e^{-j\frac{2\pi}{L}nm}.$$

La función de ventana puede ser una de las funciones de ventana descritas anteriormente en el análisis sinusoidal. Preferentemente, para ahorrar complejidad numérica, la trama transformada en el dominio de la frecuencia debe ser idéntica a la usada durante el análisis sinusoidal.

En la siguiente etapa se aplica el supuesto del modelo sinusoidal. De acuerdo con esto, la DFT de la trama prototipo se puede escribir como sigue:

$$Y_{-1}(m) = \tfrac{1}{2}\sum_{k-1}^{K} a_k \cdot \left((W(2\pi(\frac{m}{L} + \frac{f_k}{f_s})) \cdot e^{-j\varphi_k} + W(2\pi(\frac{m}{L} - \frac{f_k}{f_s})) \cdot e^{j\varphi_k} \right).$$

La siguiente etapa es darse cuenta de que el espectro de la función de ventana usada solo tiene una contribución significativa en un intervalo de frecuencia cercano a cero. Como se ilustra en la figura 3, el espectro de magnitud de la función de ventana es grande para frecuencias cercanas a cero y de otro modo pequeño (dentro del intervalo de frecuencia normalizado desde -π a π, correspondiente a la mitad de la frecuencia de muestreo). Por lo tanto, como una aproximación se supone que el espectro de ventana $W(m)$ no es cero solo para un intervalo $M = [-m_{min}, m_{max}]$, siendo m_{min} y m_{max} números positivos pequeños. En particular, se usa una aproximación del espectro de función de ventana de modo que para cada k las contribuciones de los espectros de ventana desplazados en la expresión anterior son estrictamente no superpuestas. Por lo tanto, en la ecuación anterior para cada índice de frecuencia, siempre existe solo como máximo la contribución de un sumando, es decir, de un espectro de ventana desplazado. Esto significa que la expresión anterior se reduce a la siguiente expresión aproximada:

$$\hat{Y}_{-1}(m) = \frac{a_k}{2} \cdot W\left(2\pi\left(\frac{m}{L} - \frac{f_k}{f_s}\right)\right) \cdot e^{j\varphi_k}$$

para m no negativo $\in M_k$ y para cada k.

Aquí, M_k indica el intervalo de entero $M_k = \left[redondeo\left(\frac{f_k}{f_s} \cdot L\right) - m_{min,k}, redondeo\left(\frac{f_k}{f_s} \cdot L\right) - m_{max,k}\right]$ donde $m_{min,k}$ y $m_{max,\,k}$ cumplen la restricción expuesta anteriormente de modo que los intervalos no se superponen. Una elección adecuada para $m_{min,k}$ y $m_{max,k}$ es establecerlos en un valor de entero pequeño δ, por ejemplo, δ = 3. Sin embargo, si los índices de DFT relacionados con dos frecuencias sinusoidales vecinas f_k y f_{k+1} son menores que 2δ, entonces δ se establecen a suelo $\left(\frac{redondeo\left(\frac{f_{k+1}}{f_s} \cdot L\right) \cdot redondeo\left(\frac{f_k}{f_s} \cdot L\right)}{2}\right)$ de modo que se garantice que los intervalos no se superpongan. El suelo de la función (·) es el entero más cercano al argumento de la función que es más pequeño o igual a él.

La siguiente etapa de acuerdo con el modo de realización es aplicar el modelo sinusoidal de acuerdo con la expresión anterior y evolucionar sus sinusoides K en el tiempo. La suposición de que los índices de tiempo del segmento borrado en comparación con los índices de tiempo de la trama prototipo difieren en n_{-1} muestras significa que las fases de las sinusoides avanzan en

$$\theta_k = 2\pi \cdot \frac{f_k}{f_s} n_{-1} .$$

Por lo tanto, el espectro de DFT del modelo sinusoidal evolucionado viene dado por:

$$Y_0(m) = \frac{1}{2}\sum_{k=1}^{K} a_k \cdot \left((W(2\pi(\frac{m}{L} + \frac{f_k}{f_s})) \cdot e^{-j(\varphi_k + \theta_k)} + W(2\pi(\frac{m}{L} - \frac{f_k}{f_s})) \cdot e^{j(\varphi_k + \theta_k)}) \right) .$$

Aplicando nuevamente la aproximación de acuerdo con la cual los espectros de función de ventana desplazados no se superponen da:

$$\hat{Y}_0(m) = \frac{a_k}{2} \cdot W\left(2\pi\left(\frac{m}{L} - \frac{f_k}{f_s}\right)\right) \cdot e^{j(\varphi_k + \theta_k)}$$

para $m \in M_k$ no negativo y para cada k.

Comparando la DFT de la trama prototipo $Y_{-1}(m)$ con la DFT del modelo sinusoidal evolucionado $Y_0(m)$ usando la aproximación, se encuentra que el espectro de magnitud permanece sin cambios mientras la fase se desplaza por $\theta_k = 2\pi \cdot \frac{f_k}{f_s} n_{-1}$, para cada $m \in M_k$. Por lo tanto, los coeficientes del espectro de frecuencia de la trama prototipo en la vecindad de cada sinusoide se desplazan proporcionalmente a la frecuencia sinusoidal f_k y la diferencia de tiempo entre la trama de audio perdida y la trama prototipo n_{-1}.

Por lo tanto, de acuerdo con el modo de realización, la trama de sustitución se puede calcular mediante la siguiente expresión: $z(n) = IDTf\{Z(m)\}$ con $Z(m) = Y(m) \cdot e^{j\theta_k}$ para $m \in M_k$ no negativo y para cada k.

Un modo de realización específico aborda la aleatorización de fase para índices de DFT que no pertenecen a ningún intervalo M_k. Como se describe anteriormente, los intervalos M_k, $k = 1...K$ se deben establecer de modo que estrictamente no sean superpuestos, lo que se hace usando algún parámetro δ que controla el tamaño de los intervalos. Puede suceder que δ sea pequeño en relación con la distancia de frecuencia de dos sinusoides vecinas. Por lo tanto, en ese caso sucede que existe un hueco entre dos intervalos. En consecuencia, para los índices de DFT correspondientes m no se identifica desplazamiento de fase de acuerdo con la expresión anterior $Z(m) = Y(m) \cdot e^{j\theta_k}$. Una elección adecuada de acuerdo con este modo de realización es aleatorizar la fase para estos índices, proporcionando $Z(m) = Y(m) \cdot e^{j2\pi rand(\cdot)}$, donde la función rand(·) devuelve algún número aleatorio.

Se ha encontrado beneficioso para la calidad de las señales reconstruidas optimizar el tamaño de los intervalos M_k. En particular, los intervalos deberían ser mayores si la señal es muy tonal, es decir, cuando tiene picos

espectrales claros y distintos. Este es el caso, por ejemplo, cuando la señal es armónica con una periodicidad clara. En otros casos donde la señal tiene una estructura espectral menos pronunciada con máximos espectrales más amplios, se ha encontrado que usar intervalos pequeños da lugar a una mejor calidad. Este hallazgo da lugar a otra mejora de acuerdo con la cual el tamaño del intervalo se adapta de acuerdo con las propiedades de la señal. Una realización es usar un detector de tonalidad o de periodicidad. Si este detector identifica la señal como tonal, el parámetro δ que controla el tamaño del intervalo se establece en un valor relativamente grande. De otro modo, el parámetro δ se establece en valores relativamente más pequeños.

En base a lo anterior, los procedimientos de ocultación de pérdida de trama de audio implican las siguientes etapas:

1. Analizar un segmento de la señal disponible previamente sintetizada para obtener las frecuencias sinusoidales constituyentes f_k de un modelo sinusoidal, opcionalmente usando una estimación de frecuencia perfeccionada.

2. Extraer un trama prototipo y_{-1} de la señal disponible sintetizada previamente y calcular la DFT de esa trama.

3. Calcular el desplazamiento de fase θ_k para cada sinusoide k en respuesta a la frecuencia sinusoidal f_k y el avance de tiempo n_{-1} entre la trama prototipo y la trama de sustitución. Opcionalmente, en este etapa, el tamaño del intervalo M se puede haber adaptado en respuesta a la tonalidad de la señal de audio.

4. Para cada sinusoide k, avanzar la fase de la DFT de trama prototipo con θ_k selectivamente para los índices de DFT relacionados con una vecindad alrededor de la frecuencia sinusoide f_k.

5. Calcular la DFT inversa del espectro obtenido en la etapa 4.

Análisis y detección de propiedades de pérdida de trama y señal

Los procedimientos descritos anteriormente se basan en la suposición de que las propiedades de la señal de audio no cambian significativamente durante la corta duración de tiempo desde la trama de señal recibida y reconstruida previamente y una trama perdida. En ese caso, es una muy buena elección retener el espectro de magnitud de la trama previamente reconstruida y evolucionar las fases de los componentes principales sinusoidales detectados en la señal previamente reconstruida. Sin embargo, existen casos en los que esta suposición es incorrecta, por ejemplo, transitorios con cambios repentinos de energía o cambios espectrales repentinos.

En consecuencia, un primer modo de realización de un detector de transitorios de acuerdo con la invención se puede basar en variaciones de energía dentro de la señal previamente reconstruida. Este procedimiento, ilustrado en la figura 11, calcula la energía en una parte izquierda y una parte derecha de alguna trama de análisis 113. La trama de análisis puede ser idéntica a la trama usada para el análisis sinusoidal descrito anteriormente. Una parte (bien izquierda o derecha) de la trama de análisis puede ser la primera o respectivamente la última mitad de la trama de análisis o, por ejemplo, el primer o respectivamente el último cuarto de la trama de análisis, **110.** El cálculo de energía respectivo se realiza sumando los cuadrados de las muestras en estas tramas parciales:

$$E_{izquierda} = \sum_{n=0}^{N_{part}-1} y^2(n - n_{izquierda}), \text{y } E_{derecha} = \sum_{n=0}^{N_{part}-1} y^2(n - n_{derecha}).$$

En el presente documento $y(n)$ indica la trama de análisis, $n_{izquierda}$ y $n_{derecha}$ indican los respectivos índices de inicio de las tramas parciales que son de tamaño N_{part}.

Ahora las energías de trama parcial izquierda y derecha se usan para la detección de una discontinuidad de señal. Esto se hace calculando la proporción

$$R_{l/r} = \frac{E_{izquierda}}{E_{derecha}}.$$

Se puede detectar una discontinuidad con una disminución repentina de energía (desviación) si la proporción $R_{l/r}$ excede algún umbral (por ejemplo, 10), **115.** De forma similar, se puede detectar una discontinuidad con un incremento de energía repentino

(inicio) si la proporción $R_{l/r}$ está por debajo de algún otro umbral (por ejemplo, 0,1), **117.**

En el contexto de los procedimientos de ocultación descritos anteriormente, se ha encontrado que la proporción de energía definida anteriormente puede ser en muchos casos un indicador demasiado insensible. En particular, en las señales reales y especialmente en la música, existen casos en que un tono en alguna frecuencia emerge repentinamente, mientras que otro tono en alguna otra frecuencia se detiene repentinamente. Analizar una trama

de señal de este tipo con la proporción de energía definida anteriormente daría lugar en cualquier caso a un resultado de detección incorrecto para al menos uno de los tonos, ya que este indicador es insensible a diferentes frecuencias.

Una solución a este problema se describe en el siguiente modo de realización. La detección de transitorios ahora se realiza en el plano de frecuencia de tiempo. La trama de análisis se divide nuevamente en una trama parcial izquierda y derecha, **110.** Aunque ahora, estas dos tramas parciales se transforman (después de aplicar adecuadamente una ventana con, por ejemplo, una ventana de Hamming, **111**) en el dominio de la frecuencia, por ejemplo, por medio de una DFT de punto N_{part}, **112.**

$$Y_{izquierda}(m) = DFT\{y(n - n_{izquierda})\}_{Npart}$$

y

$$Y_{derecha}(m) = DFT\{y(n - n_{derecha})\}_{Npart},$$

con $m = 0... N_{part} - 1$.

Ahora la detección de transitorios se puede hacer selectivamente por frecuencia para cada periodo de DFT con índice m. Usando las potencias de los espectros de magnitud de trama parcial izquierda y derecha, para cada índice de DFT m se puede calcular una proporción de energía respectiva **113** como

$$R_{l/r}(m) = \frac{|Y_{izquierda}(m)|^2}{|Y_{derecha}(m)|^2}.$$

Los experimentos muestran que la detección de transitorios selectiva de frecuencia con resolución de periodo de DFT es relativamente imprecisa debido a fluctuaciones estadísticas (errores de estimación). Se descubrió que la calidad de la operación se potencia bastante cuando se realiza la detección de transitorios selectiva de frecuencia sobre la base bandas de frecuencia. Siendo que $I_k = [m_{k-1} + 1,..., m_k]$ especifica el intervalo $k^{ésimo}$, $k = 1...K$, cubriendo los periodos de DFT desde $m_{k-1} + 1$ hasta m_k, luego estos los intervalos definen K bandas de frecuencia. La detección de transitorios selectiva de grupo de frecuencia ahora se puede basar en la proporción de banda entre las energías de banda respectivas de las tramas parciales izquierda y derecha:

$$R_{l/r,banda}(k) = \frac{\sum_{m \in I_k} |Y_{izquierda}(m)|^2}{\sum_{m \in I_k} |Y_{derecha}(m)|^2}.$$

Cabe señalar que el intervalo $I_k = [m_{k-1} + 1,..., m_k]$ corresponde a la banda de frecuencia $B_k = [\frac{m_{k-1}+1}{N_{part}} \cdot f_s, ..., \frac{m_k}{N_{part}} \cdot f_s]$, donde f_s indica la frecuencia de muestreo de audio.

El delimitador más bajo de banda de frecuencia más baja se puede establecer en 0 pero también se puede establecer en un índice de DFT correspondiente a una frecuencia más grande para mitigar los errores de estimación que crecen con frecuencias más bajas. El delimitador más alto de banda de frecuencia más alta m_K se puede establecer en $\frac{Npart}{2}$ pero se elige preferentemente para corresponder a alguna frecuencia más baja en la que un transitorio todavía tiene un efecto audible significativo.

Una elección adecuada para estos tamaños o anchos de banda de frecuencia es hacerlos de igual tamaño con, por ejemplo, un ancho de varios 100 Hz. Otra forma preferente es hacer los anchos de banda de frecuencia siguiendo el tamaño de las bandas críticas auditivas humanas, es decir, relacionarlos con la resolución de frecuencia del sistema auditivo. Esto significa aproximadamente hacer que los anchos de banda de frecuencia sean iguales para frecuencias de hasta 1 kHz e incrementarlos exponencialmente por encima de 1 kHz. El incremento exponencial significa, por ejemplo, duplicar el ancho de banda de frecuencia cuando se incrementa el índice de banda k.

Como se describe en el primer modo de realización del detector de transitorios que se basó en una proporción de energía de dos tramas parciales, cualquiera de las proporciones relacionadas con las energías de banda o las energías de periodo de DFT de dos tramas parciales se comparan con ciertos umbrales. Se usa un umbral superior respectivo para la detección de desviación **115** (selectiva de frecuencia) y un umbral inferior respectivo para la detección de inicio **117** (selectiva de frecuencia).

Otro indicador dependiente de la señal de audio que es adecuado para una adaptación del procedimiento de ocultación de pérdida de trama se puede basar en los parámetros de códec transmitidos al descodificador. Por

ejemplo, el códec puede ser un códec multimodo como ITU-T G.718. Dicho códec puede usar modos de códec particulares para diferentes tipos de señal y un cambio del modo de códec en una trama poco antes de que la pérdida de trama se pueda considerar como un indicador de un transitorio.

Otro indicador útil para la adaptación de la ocultación de pérdida de trama es un parámetro de códec relacionado con una propiedad de emisión sonora y la señal transmitida. La emisión sonora se refiere a la voz altamente periódica que se genera por una excitación glótica periódica del tracto vocal humano.

Otro indicador preferente es si se estima que el contenido de la señal sea música o voz. Dicho indicador se puede obtener de un clasificador de señal que típicamente puede ser parte del códec. En caso de que el códec realice dicha clasificación y ponga a disposición del descodificador una decisión de clasificación correspondiente como parámetro de codificación, este parámetro se usa preferentemente como indicador de contenido de señal para adaptar el procedimiento de ocultación de pérdida de trama.

Otro indicador que se usa preferentemente para la adaptación de los procedimientos de ocultación de pérdida de trama son las ráfagas de las pérdidas de trama. Las ráfagas de pérdidas de trama significa que se producen varias pérdidas de trama seguidas, lo que dificulta que el procedimiento de ocultación de pérdida de trama use partes de señal recientemente descodificadas válidas para su funcionamiento. Un indicador del estado de la técnica es el número n_{burst} de pérdidas de trama seguidas observadas. Este contador se incrementa con uno en cada pérdida de trama y se restablece a cero cuando se recibe una trama válida. Este indicador también se usa en el contexto de los modos de realización de ejemplo de la presente invención.

Adaptación del procedimiento de ocultación de pérdida de trama

En caso de que las etapas llevadas a cabo anteriormente indiquen una condición que sugiera una adaptación de la operación de ocultación de pérdida de trama, se modifica el cálculo del espectro de la trama de sustitución.

Mientras que el cálculo original del espectro de la trama de sustitución se realiza de acuerdo con la expresión $Z(m) = Y(m) \cdot e^{j\theta_k}$, ahora se introduce una adaptación que modifica la magnitud y la fase. La magnitud se modifica por medio del ajuste a escala con dos factores $\alpha(m)$ y $\beta(m)$ y la fase se modifica con un componente de fase aditivo $\vartheta(m)$. Esto da lugar al siguiente cálculo modificado de la trama de sustitución:

$$Z(m) = \alpha(m) \cdot \beta(m) \cdot Y(m) \cdot e^{j(\theta_k - \vartheta(m))}.$$

Cabe señalar que los procedimientos de ocultación de pérdida de trama originales (no adaptados) se usan si $\alpha(m) = 1$, $\beta(m) = 1$ y $\vartheta(m) = 0$. Estos valores respectivos son, por lo tanto, los predeterminados.

El objetivo general con la introducción de adaptaciones de magnitud es evitar artefactos audibles del procedimiento de ocultación de pérdida de trama. Dichos artefactos pueden ser sonidos musicales o tonales o sonidos extraños que surgen de repeticiones de sonidos transitorios. Dichos artefactos a su vez darían lugar a degradaciones de la calidad, cuya evitación es el objetivo de las adaptaciones descritas. Una forma adecuada de dichas adaptaciones es modificar el espectro de magnitud de la trama de sustitución en un grado adecuado.

La figura 12 ilustra un modo de realización de una modificación del procedimiento de ocultación. La adaptación de magnitud, **123**, se realiza preferentemente si el contador de pérdida de ráfaga n_{burst} excede algún umbral thr_{burst}, por ejemplo, $thr_{burst} = 3$, **121**. En ese caso, se usa un valor más pequeño que 1 para el factor de atenuación, por ejemplo, $\alpha(m) = 0{,}1$.

Sin embargo, se ha encontrado que es beneficioso realizar la atenuación con un grado gradualmente creciente. Un modo de realización preferente que logra esto es definir un parámetro logarítmico que especifique un incremento logarítmico en la atenuación por trama, *att_per_frame.* Luego, en caso de que el contador de ráfaga exceda el umbral, el factor de atenuación gradualmente creciente se calcula mediante

$$\alpha(m) = 10^{c \cdot att_per_frame \cdot (n_{ráfaga} - thr_{ráfaga})}.$$

Aquí la constante c es una simple constante de ajuste a escala que permite especificar el parámetro *att_per_frame,* por ejemplo, en decibelios (dB).

Se realiza una adaptación preferente adicional en respuesta al indicador de si se estima que la señal es música o voz. Para el contenido musical en comparación con el contenido de voz, es preferente incrementar el umbral thr_{burst} y disminuir la atenuación por trama. Esto es equivalente a realizar la adaptación del procedimiento de ocultación de pérdida de trama con un grado más bajo. El trasfondo de este tipo de adaptación es que la música es en general menos sensible a ráfagas de pérdida más largas que la voz. Por lo tanto, el procedimiento de ocultación de pérdida de trama original, es decir, sin modificar, sigue siendo preferente para este caso, al menos para un mayor número de pérdidas de trama seguidas.

Otra adaptación del procedimiento de ocultación con respecto al factor de atenuación de magnitud se realiza preferentemente en caso de que se haya detectado un transitorio basado en que el indicador $R_{llr,\ banda}(k)$ o alternativamente $R_{llr}(m)$ o R_{llr} ha pasado un umbral, **122.** En ese caso, una acción de adaptación adecuada, **125**, es modificar el segundo factor de atenuación de magnitud $\beta(m)$ de modo que la atenuación total esté controlada por el producto de los dos factores $\alpha(m) \cdot \beta(m)$.

$\beta(m)$ se establece en respuesta a un transitorio indicado. En caso de que se detecte una desviación, el factor $\beta(m)$ se elige preferentemente para reflejar la disminución de energía de la desviación. Una elección adecuada es establecer $\beta(m)$ en el cambio de ganancia detectado:

$$\beta(m) = \sqrt{R_{llr,banda}(k)},$$

para $m \in I_k$, $k = 1...K$.

En caso de que se detecte un inicio, se considera ventajoso limitar el incremento de energía en la trama de sustitución. En ese caso, el factor se puede establecer en un valor fijo de, por ejemplo, 1, lo que significa que no existe atenuación pero tampoco amplificación.

En lo anterior, cabe señalar que el factor de atenuación de magnitud se aplica preferentemente selectivamente por frecuencia, es decir, con factores calculados individualmente para cada banda de frecuencia. En caso de que no se use el enfoque de banda, los factores de atenuación de magnitud correspondientes todavía se pueden obtener de forma análoga. *A continuación,* $\beta(m)$ se puede establecer individualmente para cada periodo de DFT en caso de que se use la detección de transitorios selectiva de frecuencia en el nivel de periodo de DFT. O, en caso de que no se use ninguna indicación de transitorios selectiva de frecuencia, $\beta(m)$ puede ser globalmente idéntico para todos los *m.*

Otra adaptación preferente del factor de atenuación de magnitud se realiza junto con una modificación de la fase por medio del componente de fase adicional $\vartheta(m)$ **127.** En caso de que para un *m* dado se utiliza una modificación de fase de este tipo, el factor de atenuación *β(m)* se reduce aún más. Preferentemente, incluso el grado de modificación de fase se tiene en cuenta. Si la modificación de fase es solo moderada, $\beta(m)$ solo se reduce a escala ligeramente, mientras que si la modificación de fase es fuerte, $\beta(m)$ se reduce a escala en un grado mayor.

El objetivo general con la introducción de adaptaciones de fase es evitar una tonalidad demasiado fuerte o una periodicidad de señal en las tramas de sustitución generadas, lo que a su vez daría lugar a degradaciones de calidad. Una forma adecuada para dichas adaptaciones es aleatorizar o someter a tramado la fase en un grado adecuado.

Dicho tramado de fase se logra si el componente de fase adicional $\vartheta(m)$ se establece en un valor aleatorio ajustado a escala con algún factor de control: $\vartheta(m) = a(m) \cdot \text{rand}(\cdot)$.

El valor aleatorio obtenido por la función rand(·) se genera, por ejemplo, por algún generador de números pseudoaleatorios. Aquí se supone que proporciona un número aleatorio dentro del intervalo $[0, 2\pi]$.

El factor de ajuste a escala $a(m)$ en la ecuación anterior controla el grado por el cual la fase original θ_k se somete a tramado. Los siguientes modos de realización abordan la adaptación de fase por medio del control de este factor de ajuste a escala. El control del factor de ajuste a escala se realiza de forma análoga como el control de los factores de modificación de magnitud descritos anteriormente.

De acuerdo con un primer modo de realización, el factor de ajuste a escala $a(m)$ se adapta en respuesta al contador de pérdida de ráfaga. Si el contador de pérdida de ráfaga n_{burst} excede un umbral dethr_{burst}, por ejemplo, $thr_{burst} = 3$, se usa un valor mayor que 0, por ejemplo $a(m) = 0{,}2$.

Sin embargo, se ha encontrado que es beneficioso realizar el tramado con un grado gradualmente creciente. Un modo de realización preferente que logra esto es definir un parámetro que especifique un incremento en el tramado por trama, *dith_increase_per_frame.* Luego, en caso de que el contador de ráfaga exceda el umbral, el factor de control de tramado gradualmente creciente se calcula mediante

$$a(m) = dith_increase_per_frame \cdot (n_{ráfaga} - thr_{ráfaga}).$$

Cabe señalar en la fórmula anterior que $a(m)$ se tiene que limitar a un valor máximo de 1 para el cual se logra el tramado de fase completo.

Cabe señalar que el valor umbral de pérdida de ráfaga thr_{burst} usado para iniciar el tramado de fase puede ser el mismo umbral que el usado para la atenuación de magnitud. Sin embargo, se puede obtener una mejor calidad al establecer estos umbrales en valores óptimos individualmente, lo que en general significa que estos umbrales pueden ser diferentes.

Se realiza una adaptación preferente adicional en respuesta al indicador de si se estima que la señal es música o voz. Para el contenido de música en comparación con el contenido de voz, es preferente incrementar el umbral de thr_{burst}, lo que significa que el tramado de fase para la música en comparación con voz se realiza solo en caso de más tramas perdidas seguidas. Esto es equivalente a realizar la adaptación del procedimiento de ocultación de pérdida de trama para música con un grado más bajo. El trasfondo de este tipo de adaptación es que la música es en general menos sensible a ráfagas de pérdida más largas que la voz. Por lo tanto, el procedimiento de ocultación de pérdida de trama original, es decir, sin modificar, sigue siendo preferente para este caso, al menos para un mayor número de pérdidas de trama seguidas.

Otro modo de realización preferente es adaptar el tramado de fase en respuesta a un transitorio detectado. En ese caso, se puede usar un grado más fuerte de tramado de fase para los periodos de DFT m para los que se indica un transitorio, bien para ese periodo, para los periodos de DFT de la banda de frecuencia correspondiente o de la trama entera.

Parte de los esquemas descritos abordan la optimización del procedimiento de ocultación de pérdida de trama para señales armónicas y en particular para voz sonora.

En caso de que los procedimientos que usan una estimación de frecuencia perfeccionada como se describe anteriormente no se realicen, otra posibilidad de adaptación para el procedimiento de ocultación de pérdida de trama que optimiza la calidad de las señales de voz sonora es cambiar a algún otro procedimiento de ocultación de pérdida de trama que esté específicamente diseñado y optimizado para la voz en lugar de para señales de audio generales que contienen música y voz. En ese caso, el indicador de que la señal comprende una señal de voz sonora se usa para seleccionar otro esquema de ocultación de pérdida de trama optimizado para voz en lugar de los esquemas descritos anteriormente.

Las modos de realización se aplican a un controlador en un descodificador, como se ilustra en la figura 13. La figura 13 es un diagrama de bloques de un descodificador de acuerdo con los modos de realización. El descodificador **130** comprende una unidad de entrada **132** configurada para recibir una señal de audio codificada. La figura ilustra la ocultación de pérdida de trama mediante una unidad de ocultación de pérdida de trama lógica **134,** que indica que el descodificador está configurado para implementar una ocultación de una trama de audio perdida, de acuerdo con los modos de realización descritos anteriormente. Además, el descodificador comprende un controlador **136** para implementar los modos de realización descritos anteriormente. El controlador **136** está configurado para detectar condiciones en las propiedades de la señal de audio recibida y reconstruida previamente o en las propiedades estadísticas de las pérdidas de trama observadas para las cuales la sustitución de una trama perdida de acuerdo con los procedimientos descritos proporciona una calidad relativamente reducida. En caso de que se detecte dicha condición, el controlador **136** está configurado para modificar el elemento de los procedimientos de ocultación de acuerdo con los cuales el espectro de la trama de sustitución se calcula mediante $Z(m) = Y(m) \cdot e^{j\theta_k}$ ajustando selectivamente las fases o las magnitudes del espectro. La detección se puede realizar mediante una unidad detectora **146** y la modificación se puede realizar mediante una unidad modificadora **148** como se ilustra en la figura 14.

El descodificador con sus unidades integradoras se podría implementar en hardware. Existen numerosas variantes de elementos de circuitos que se pueden usar y combinar para lograr las funciones de las unidades del descodificador. Dichas variantes están abarcadas por los modos de realización. Ejemplos particulares de implementación de hardware del descodificador es la implementación en hardware de procesador de señal digital (DSP) y tecnología de circuito integrado, incluyendo los circuitos electrónicos de propósito general y los circuitos específicos de la aplicación.

El descodificador **150** descrito en el presente documento se podría implementar de forma alternativa, por ejemplo, como se ilustra en la figura 15, es decir, mediante uno o más de un procesador **154** y un programa informático adecuado **155** con almacenamiento o memoria adecuados **156**, por lo tanto, para reconstruir la señal de audio, que incluye realizar la ocultación de pérdida de trama de audio de acuerdo con los modos de realización descritos en el presente documento, como se muestra en la figura 13. La señal de audio codificada entrante es recibida por una entrada (IN) **152,** a la que están conectados el procesador **154** y la memoria **156**. La señal de audio descodificada y reconstruida obtenida del programa informático se emite desde la salida (OUT) **158.**

La tecnología descrita anteriormente se puede usar, por ejemplo, en un receptor, que se puede usar en un dispositivo móvil (por ejemplo, teléfono móvil, ordenador portátil) o un dispositivo estacionario, como una ordenador personal.

Se debe entender que la elección de unidades o módulos de interacción, así como la denominación de las unidades son solo para un propósito ejemplar, y se pueden configurar en una pluralidad de formas alternativas para poder ejecutar las acciones de proceso divulgadas.

También cabe destacar que las unidades o módulos descritos en esta divulgación se deben considerar como entidades lógicas y no necesariamente como entidades físicas separadas. Se apreciará que el alcance de la tecnología descrita en el presente documento abarca completamente otros modos de realización que pueden resultar obvias para los expertos en la técnica, y que el alcance de esta divulgación no se ha de limitar en consecuencia.

La referencia a un elemento en singular no pretende significar "uno y solo uno" a menos que se indique explícitamente, sino "uno o más".

Además, no es necesario que un dispositivo o procedimiento aborde todos y cada uno de los problemas que se intenta resolver con la tecnología descrita en el presente documento, para que se abarque por el presente documento.

En la descripción anterior, para propósitos de explicación y no de limitación, se exponen detalles específicos tales como arquitecturas, interfaces, técnicas particulares, etc. a fin de proporcionar una comprensión exhaustiva de la tecnología divulgada. Sin embargo, para los expertos en la técnica resultará evidente que la tecnología divulgada se puede llevar a la práctica en otros modos de realización y/o combinaciones de modos de realización que se apartan de estos detalles específicos. Es decir, los expertos en la técnica podrán concebir diversas disposiciones que, aunque no se describan o se muestren explícitamente en el presente documento, realizan los principios de la tecnología divulgada. En algunos casos, se omiten descripciones detalladas de dispositivos, circuitos y procedimientos bien conocidos para no oscurecer la descripción de la tecnología divulgada con detalles innecesarios. Todo lo expuesto en el presente documento relatando los principios, aspectos y modos de realización de la tecnología divulgada, así como los ejemplos específicos de los mismos, está concebido para abarcar equivalentes, tanto estructurales como funcionales, de los mismos. Además, se pretende que dichos equivalentes incluyan tanto equivalentes actualmente conocidos como equivalentes desarrollados en el futuro, es decir, cualquier elemento desarrollado que realice la misma función, independientemente de la estructura.

Por tanto, por ejemplo, los expertos en la técnica apreciarán que las figuras en el presente documento pueden representar vistas conceptuales de circuitos ilustrativos u otras unidades funcionales que realizan los principios de la tecnología, y/o diversos procesos que se pueden representar sustancialmente en un medio legible por ordenador y ejecutado por un ordenador o procesador, aunque dicho ordenador o procesador no se muestre explícitamente en las figuras.

Las funciones de los diversos elementos, incluyendo los bloques funcionales, se pueden proporcionar mediante el uso de hardware tal como hardware de circuito y/o hardware capaz de ejecutar un programa informático en forma de instrucciones codificadas almacenadas en un medio legible por ordenador. Por tanto, dichas funciones y bloques funcionales ilustrados se deben entender como implementados por hardware y/o implementados por ordenador, y por tanto implementados por máquina.

Las modos de realización descritos anteriormente se deben entender como unos pocos ejemplos ilustrativos de la presente invención. Los expertos en la técnica entenderán que se pueden realizar diversas modificaciones, combinaciones y cambios a los modos de realización sin apartarse del alcance de la presente invención. En particular, diferentes soluciones parciales en los diferentes modos de realización se pueden combinar en otras configuraciones, cuando sea técnicamente posible.

REIVINDICACIONES

1. Un procedimiento de ocultación de pérdida de trama, en el que un segmento de una señal de audio recibida o reconstruida previamente se usa como una trama prototipo para crear una trama de sustitución para una trama de audio perdida, comprendiendo el procedimiento:

- transformar la trama prototipo en un dominio de frecuencia;

- analizar una trama de señal reconstruida previamente y estadísticas de pérdida de trama para detectar condiciones predeterminadas (101, 121, 122), en el que dichas condiciones comprenden pérdidas de transitorios y de ráfaga detectadas con varias pérdidas de trama consecutivas;

- si dichas condiciones no se detectan, aplicar un primer procedimiento de ocultación, en el que el primer procedimiento de ocultación comprende:

aplicar un modelo sinusoidal a la trama prototipo para identificar una frecuencia de un componente sinusoidal de la señal de audio, calculando un desplazamiento de fase θ_k para el componente sinusoidal y realizando un desplazamiento de fase del componente sinusoidal por θ_k;

- si se detectan dichas condiciones, aplicar un segundo procedimiento de ocultación, en el que el segundo procedimiento de ocultación comprende:

adaptar el primer procedimiento de ocultación (102) ajustando selectivamente una magnitud del espectro de trama prototipo (125); y

- crear la trama de sustitución realizando una transformación de frecuencia inversa de un espectro de frecuencia de la trama prototipo.

2. El procedimiento de acuerdo con la reivindicación 1, en el que cuando se aplica el primer procedimiento de ocultación, la magnitud del espectro de trama prototipo se mantiene sin cambios.

3. El procedimiento de acuerdo con la reivindicación 1 o 2, en el que dicho transitorio detectado comprende una desviación.

4. El procedimiento de acuerdo con una cualquiera de las reivindicaciones 1 a 3, en el que la detección de transitorios se realiza selectivamente por frecuencia para cada banda de frecuencia.

5. El procedimiento de acuerdo con una cualquiera de las reivindicaciones 1 a 4, en el que el ajuste selectivo de la magnitud del espectro de trama prototipo se realiza selectivamente por banda de frecuencia.

6. El procedimiento de acuerdo con una cualquiera de las reivindicaciones 1 a 5, en el que el segundo procedimiento de ocultación comprende además ajustar el desplazamiento de fase θ_k añadiendo un componente aleatorio.

7. El procedimiento de acuerdo con la reivindicación 6, en el que el desplazamiento de fase θ_k se ajusta si un contador de pérdida de ráfaga excede un umbral determinado.

8. El procedimiento de acuerdo con la reivindicación 7, en el que el umbral es 3.

9. Un aparato (134, 136) para crear una trama de sustitución para una trama de audio perdida, comprendiendo el aparato:

- medios para generar una trama prototipo a partir de un segmento de una señal de audio recibida o reconstruida previamente;

- medios para transformar la trama prototipo en un dominio de frecuencia;

- medios para analizar una trama de señal reconstruida previamente y estadísticas de pérdida de trama para detectar condiciones predeterminadas (146), en el que dichas condiciones comprenden pérdidas de transitorios y de ráfaga detectadas con varias pérdidas de trama consecutivas;

- medios para aplicar un primer procedimiento de ocultación si dichas condiciones no se detectan, en el que el primer procedimiento de ocultación comprende:

aplicar un modelo sinusoidal a la trama prototipo para identificar una frecuencia de un componente sinusoidal de la señal de audio, calculando un desplazamiento de fase θ_k para el componente sinusoidal y realizando un desplazamiento de fase del componente sinusoidal por θ_k;

- medios para aplicar un segundo procedimiento de ocultación, si se detectan dichas condiciones, en el que el segundo procedimiento de ocultación comprende:

adaptar el primer procedimiento de ocultación ajustando selectivamente una magnitud del espectro de trama prototipo (148); y

- medios para crear la trama de sustitución realizando una transformación de frecuencia inversa de un espectro de frecuencia de la trama prototipo.

10. El aparato de acuerdo con la reivindicación 9, en el que el aparato comprende además medios para realizar el procedimiento de acuerdo con al menos una de las reivindicaciones 2 a 8.

11. El aparato de acuerdo con la reivindicación 9 o 10, en el que el aparato está comprendido en un descodificador de audio.

12. Un dispositivo que comprende el descodificador de audio (130) de acuerdo con la reivindicación 11.

13. Un programa informático (155) que comprende instrucciones que, cuando se ejecutan por al menos un procesador, provocan que el al menos un procesador lleve a cabo el procedimiento de acuerdo con una cualquiera de las reivindicaciones 1 a 8.

14. Un producto de programa informático (156) que comprende un medio legible por ordenador que almacena un programa informático (155) de acuerdo con la reivindicación 13.

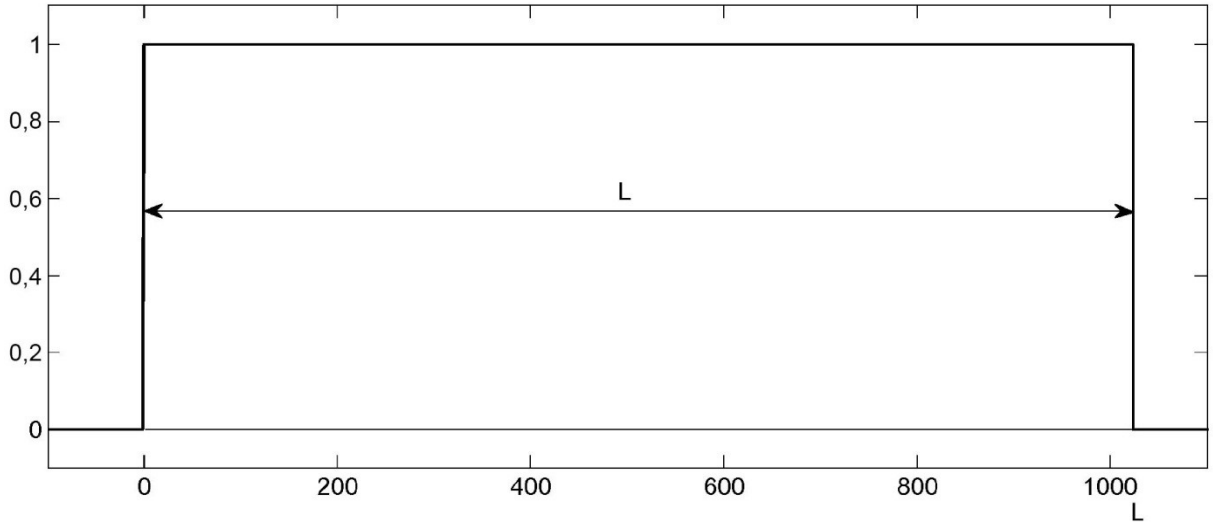
1
0,8
0,6
0,4
0,2
0
L
0
200
400
600
800
1000
L

FIG. 1

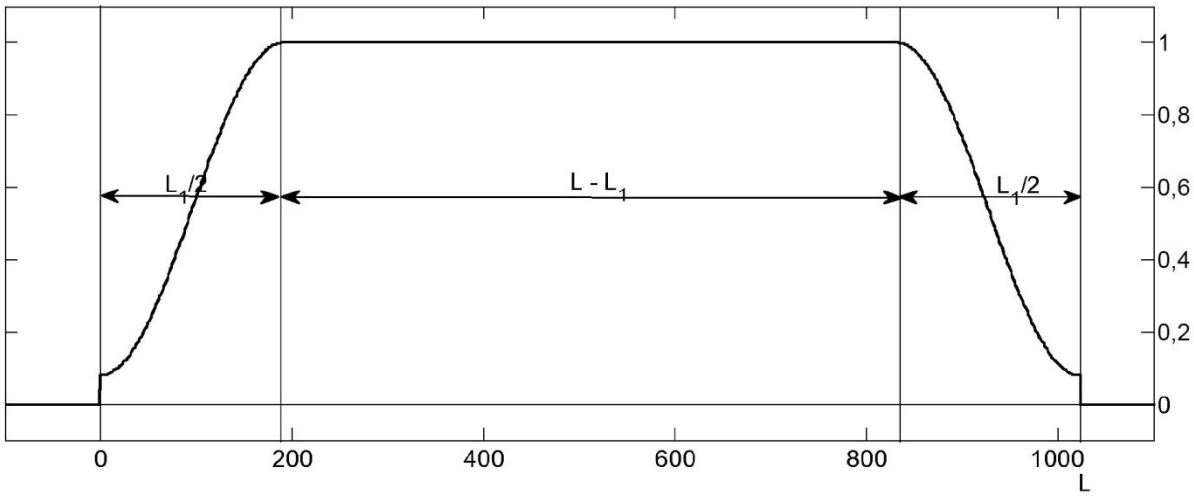
L1/2
L - L1
L1/2
1
0,8
0,6
0,4
0,2
0
0
200
400
600
800
1000
L

FIG. 2

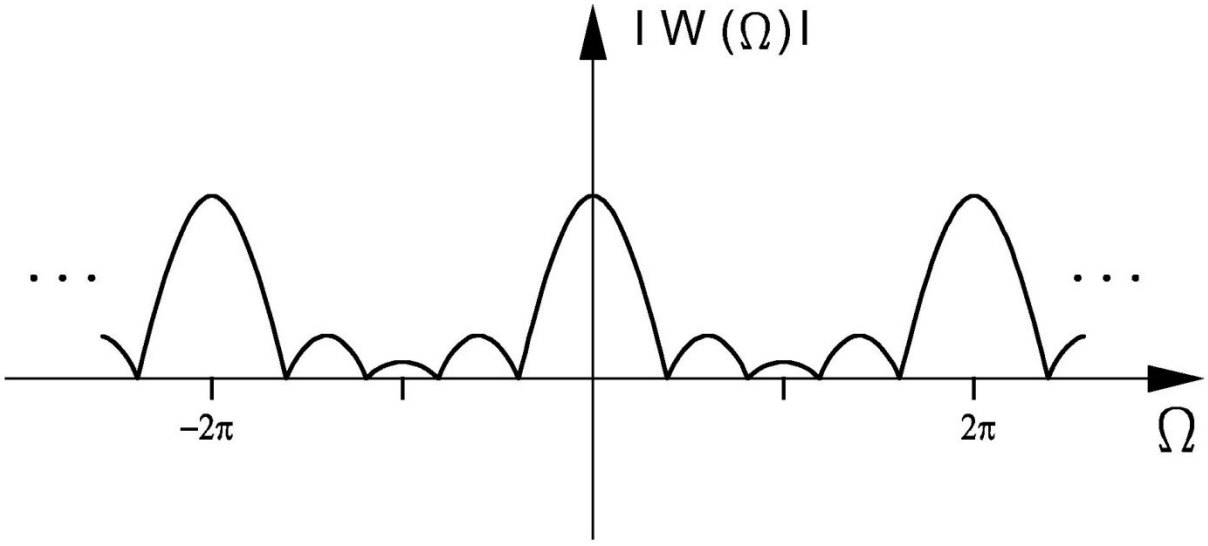

FIG. 3

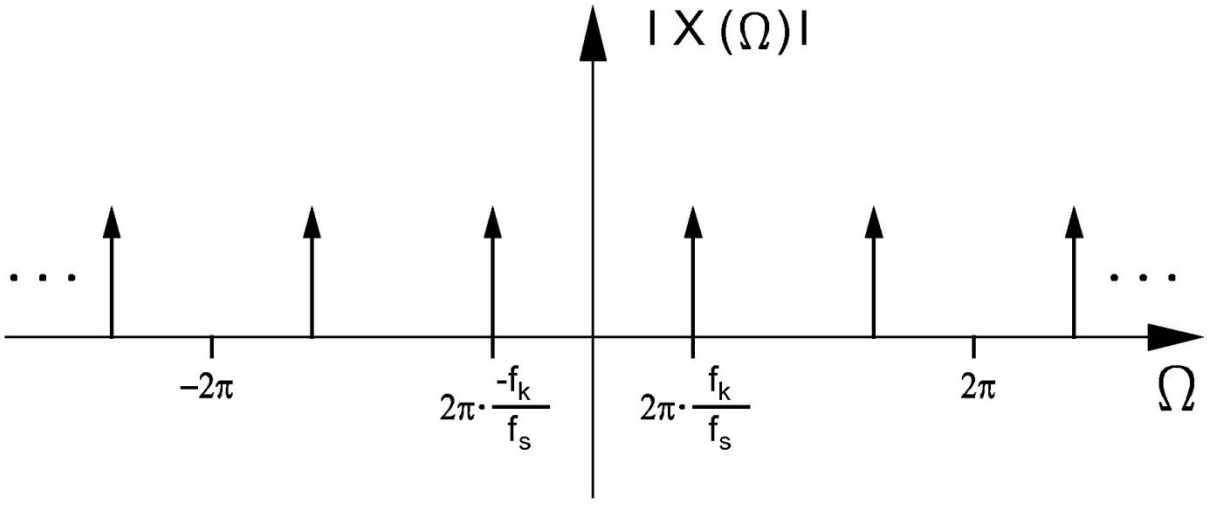

FIG. 4

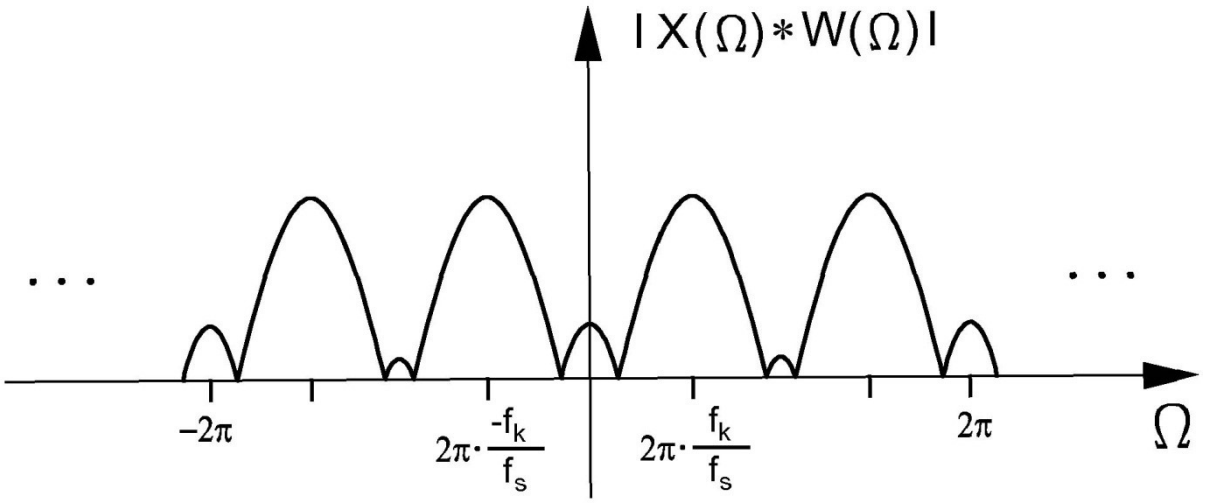

FIG. 5

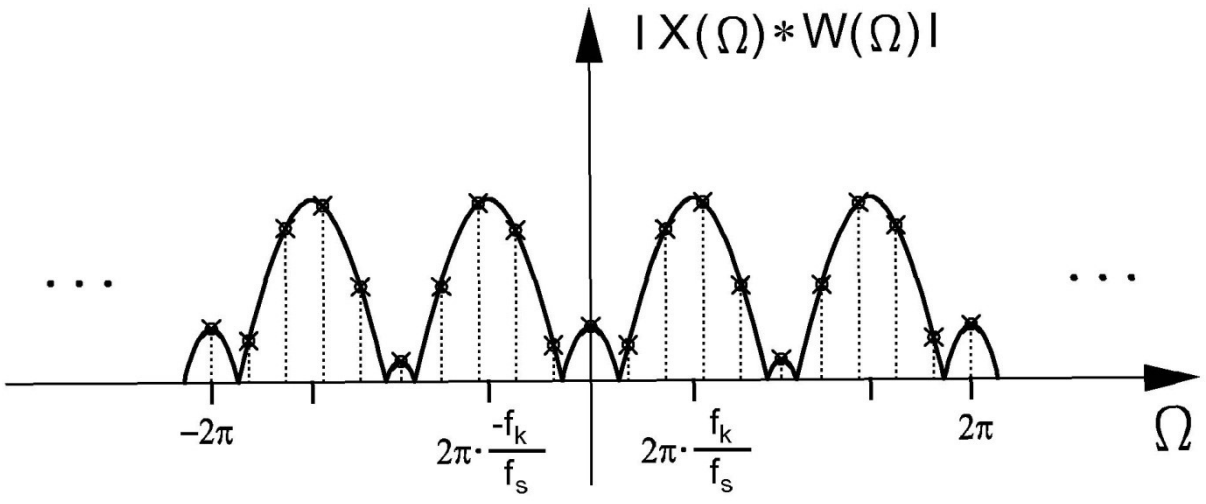

FIG. 6

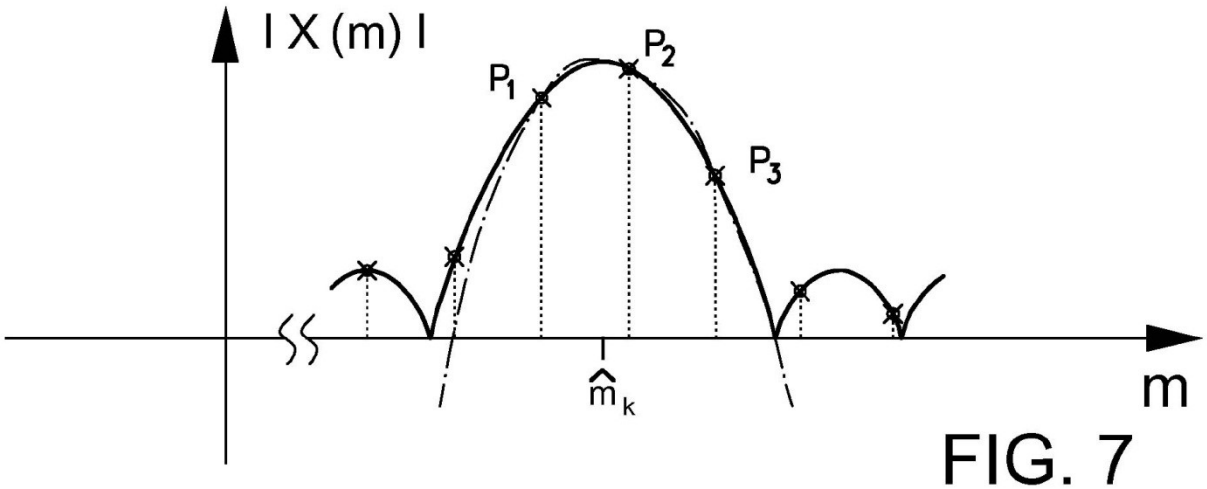

FIG. 7

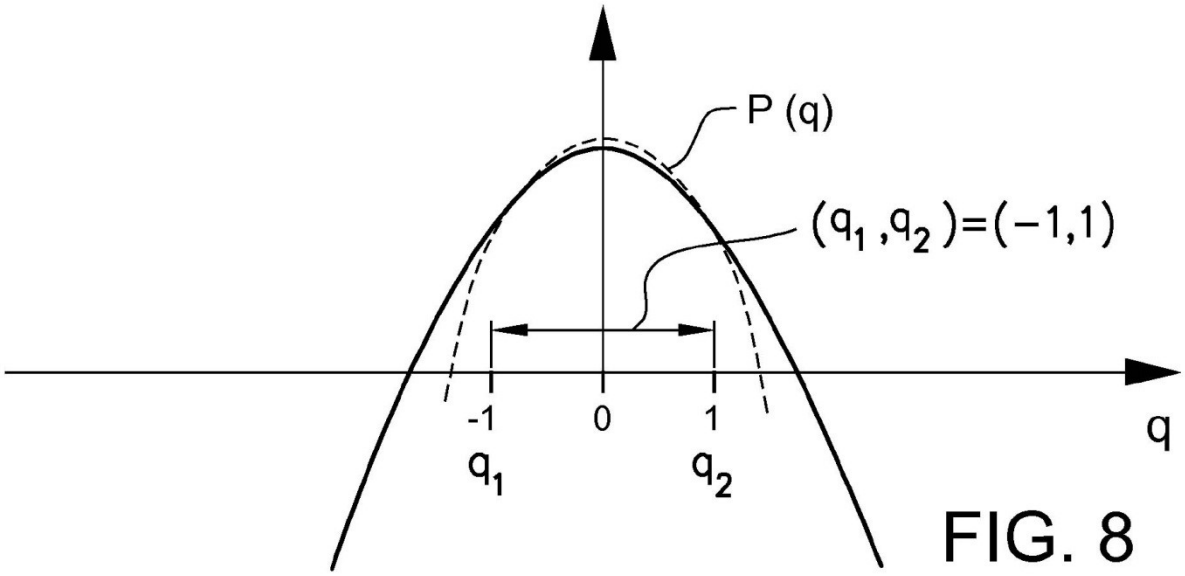

FIG. 8

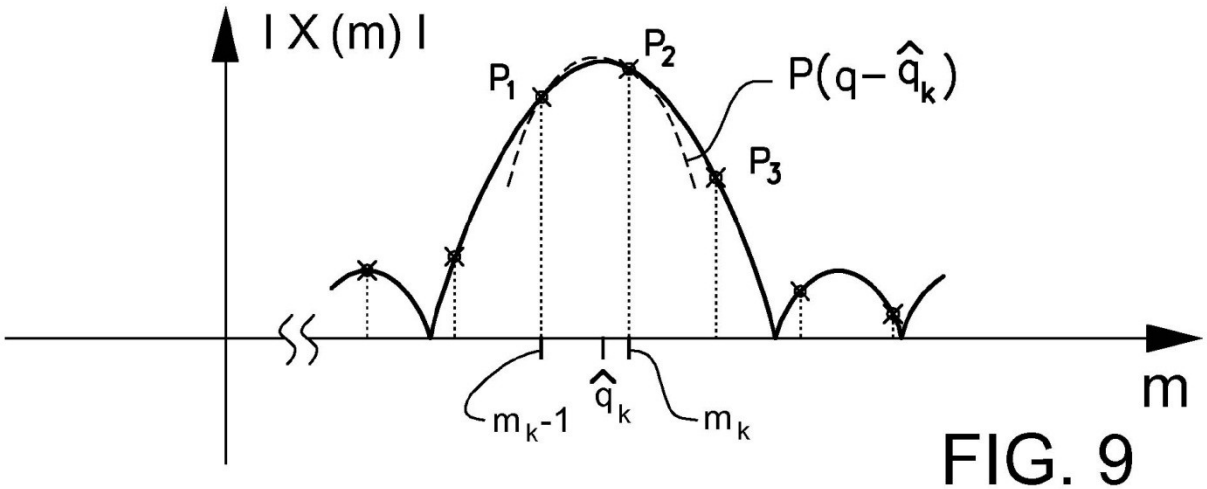

FIG. 9

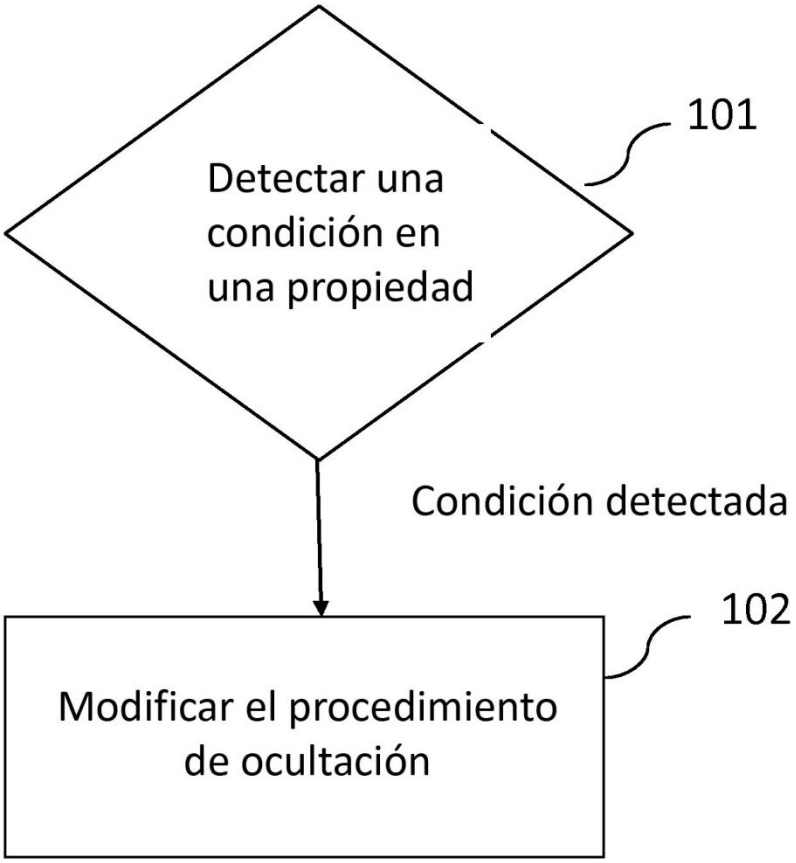
Detectar una condición en una propiedad
101
Condición detectada
Modificar el procedimiento de ocultación
102

FIG. 10

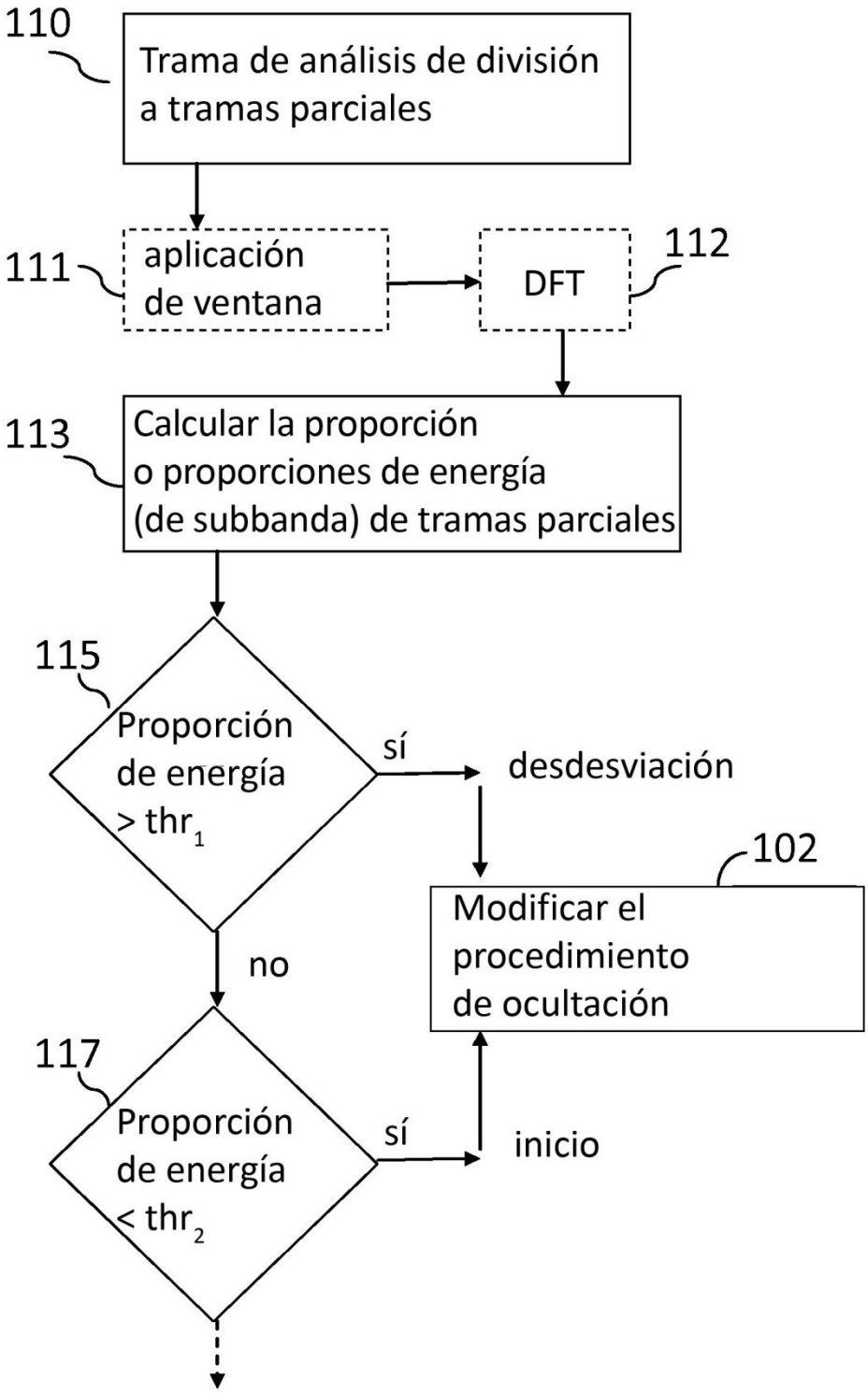
110
Trama de análisis de división a tramas parciales
111
aplicación de ventana
112
DFT
113
Calcular la proporción o proporciones de energía (de subbanda) de tramas parciales
115
Proporción de energía > thr1
sí
desdesviación
102
Modificar el procedimiento de ocultación
no
117
Proporción de energía < thr2
sí
inicio

FIG. 11

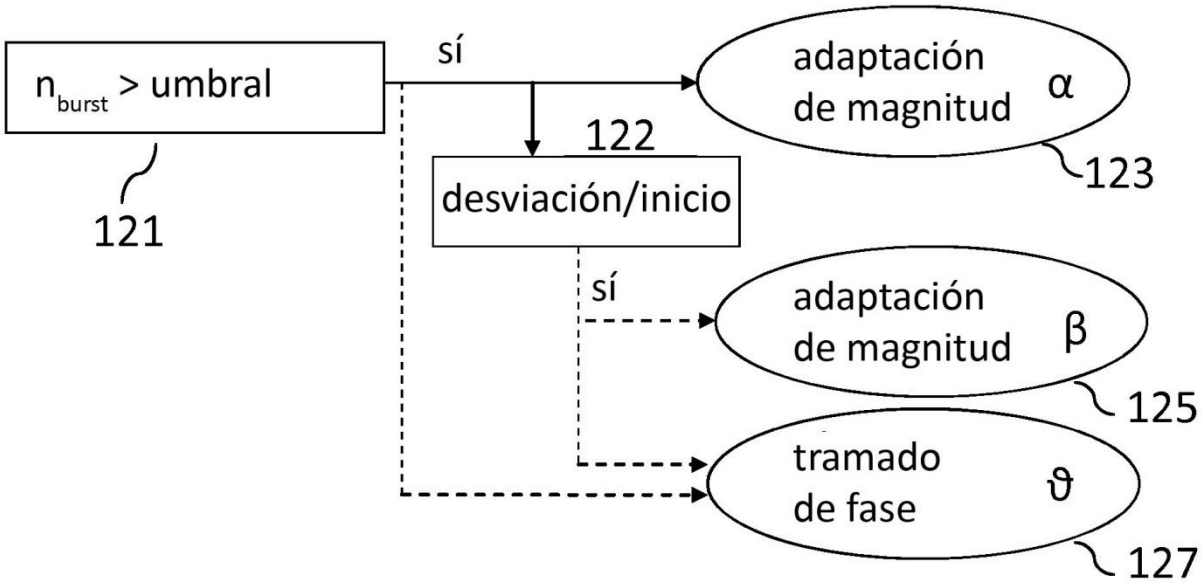
n_{burst} > umbral
sí
adaptación de magnitud α
122
desviación/inicio
123
121
sí
adaptación de magnitud β
125
tramado de fase ϑ
127

FIG. 12

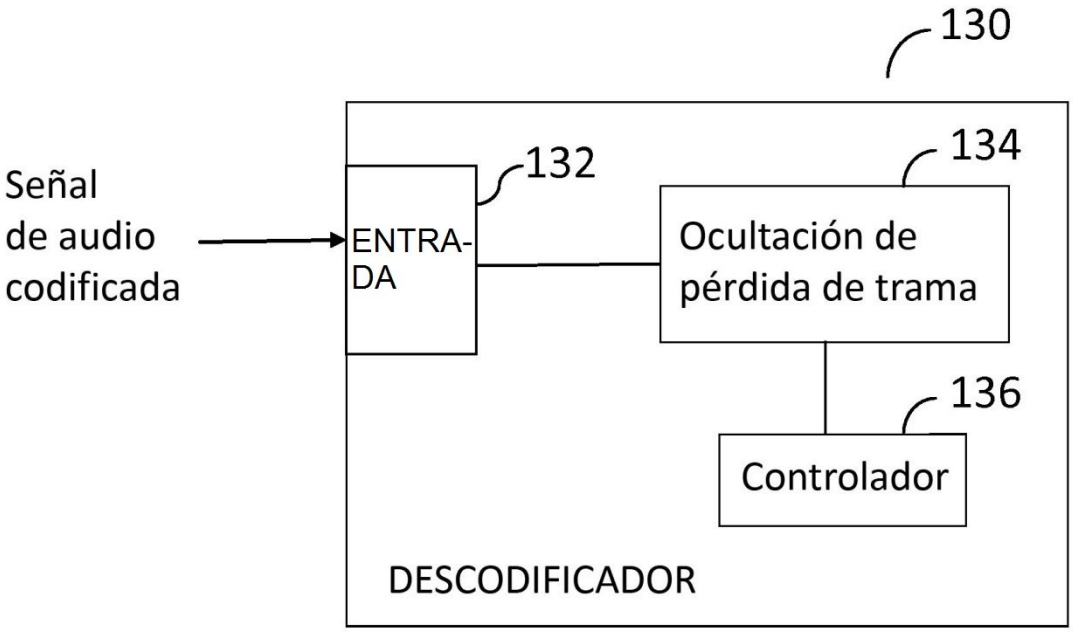
130
Señal de audio codificada
ENTRA-DA
132
134
Ocultación de pérdida de trama
136
Controlador
DESCODIFICADOR

FIG. 13

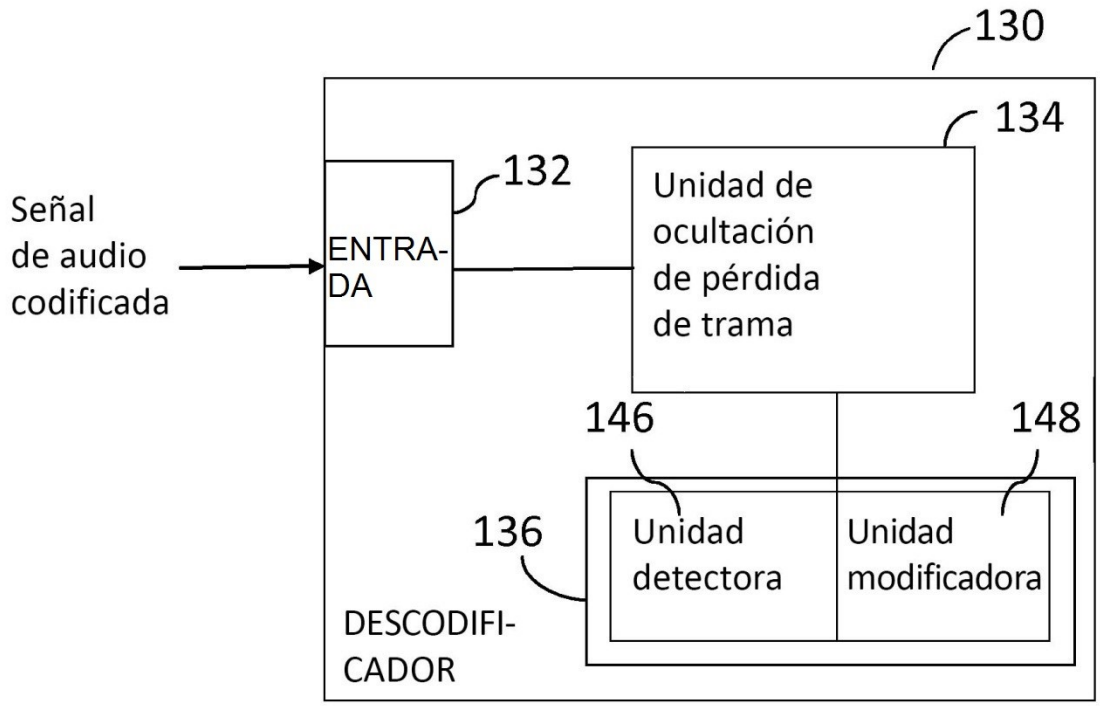
130
134
132
Señal
de audio
codificada
ENTRA-
DA
Unidad de
ocultación
de pérdida
de trama
146
148
136
Unidad
detectora
Unidad
modificadora
DESCODIFI-
CADOR

FIG. 14

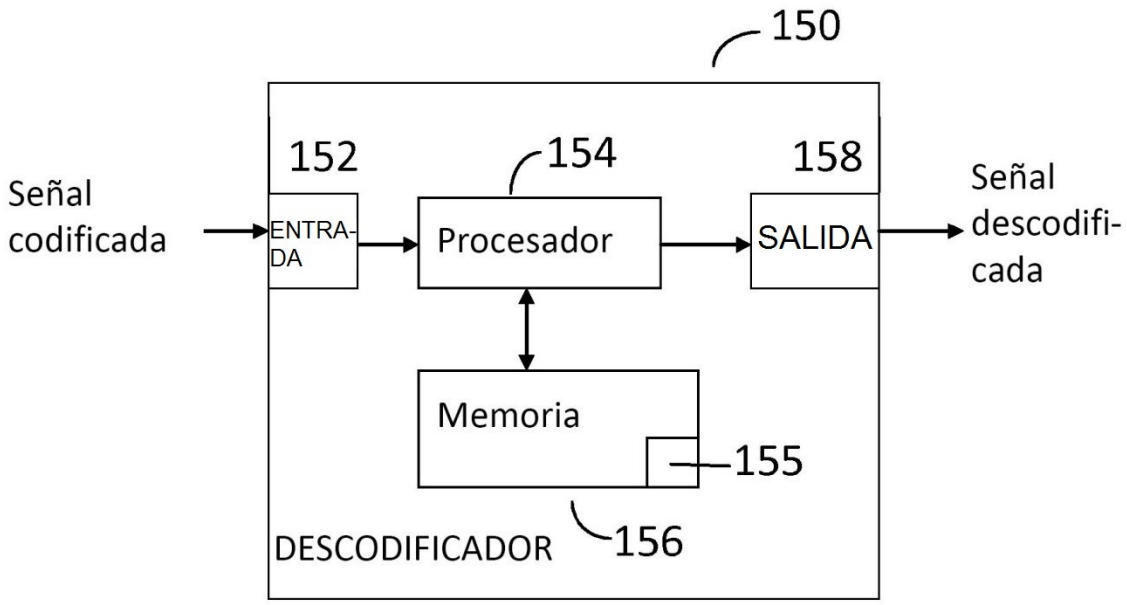
150
152
154
158
Señal
codificada
ENTRA-
DA
Procesador
SALIDA
Señal
descodifi-
cada
Memoria
155
DESCODIFICADOR
156

FIG. 15