

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 698 119**

21 Número de solicitud: 201890056

51 Int. Cl.:

**G10L 19/038** (2013.01)

12

SOLICITUD DE PATENTE

A2

22 Fecha de presentación:

**10.03.2017**

30 Prioridad:

**15.03.2016 EP 16160369**

**10.03.2017 EP 16189893**

43 Fecha de publicación de la solicitud:

**31.01.2019**

71 Solicitantes:

**FRAUNHOFER-GESELLSCHAFT ZUR  
FÖRDERUNG DER ANGEWANDTEN  
FORSCHUNG E.V. (100.0%)**

**Hansastraße 27c  
80686 Munich DE**

72 Inventor/es:

**BÄCKSTRÖM, Tom;  
GHIDO, Florin y  
FISCHER, Johannes**

74 Agente/Representante:

**CARPINTERO LÓPEZ, Mario**

54 Título: **Aparato de codificación para el procesamiento de una señal de entrada y aparato de decodificación para el procesamiento de una señal codificada**

57 Resumen:

Aparato de codificación para el procesamiento de una señal de entrada y aparato de decodificación para el procesamiento de una señal codificada.

La invención se refiere a un aparato (1) para el procesamiento de una señal de entrada (3), que comprende un ponderador perceptual (10) y un cuantificador (14). El ponderador perceptual (10) comprende un proveedor de modelo (12) y un aplicador de modelo (13). El proveedor de modelo (12) proporciona un modelo ponderado perceptual (W) basado en la señal de entrada (3). El aplicador de modelo (13) proporciona espectro ponderado perceptualmente (x) aplicando el modelo ponderado perceptual (W) a un espectro (s) basándose en la señal de entrada (3). El cuantificador (14) está configurado para cuantificar el espectro ponderado perceptualmente (x) y para proporcionar una secuencia de bits ( $\hat{u}$ ). El cuantificador (14) comprende un aplicador de matriz aleatoria (16) y un calculador de función de signo (17). El aplicador de matriz aleatoria (16) está configurado para aplicar una matriz aleatoria (P) al espectro ponderado perceptualmente (x) para proporcionar un espectro transformado (u). Un calculador de función de signo (17) está configurado para calcular una función de signo de componentes del espectro transformado (u) para proporcionar la secuencia de bits ( $\hat{u}$ ). La invención adicionalmente se refiere a un aparato (2) para el procesamiento de una señal codificada (4) y a procedimientos correspondientes.

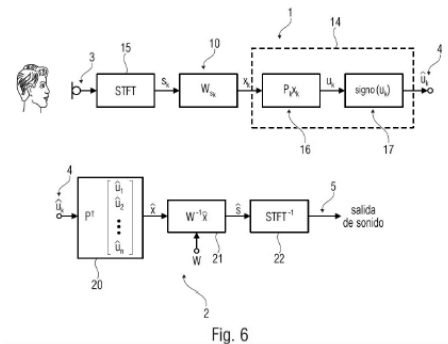


Fig. 6

ES 2 698 119 A2

**DESCRIPCIÓN**

Aparato de codificación para el procesamiento de una señal de entrada y aparato de decodificación para el procesamiento de una señal codificada

5

La invención se refiere a un aparato de codificación para el procesamiento de una señal de entrada y a un aparato de decodificación para el procesamiento de una señal codificada. La invención también se refiere a procedimientos correspondientes y a un programa informático.

10

Una parte central de códecs de voz y audio son sus modelos perceptuales, que describen la importancia perceptual relativa de errores en diferentes elementos de la representación de señales. En la práctica, los modelos perceptuales constan de factores de ponderación dependientes de señales que se usan en cuantificación de cada elemento. Para rendimiento óptimo, sería deseable usar el mismo modelo perceptual en el decodificador. Mientras el modelo perceptual es dependiente de la señal, sin embargo, no se conoce por adelantado en el decodificador, con lo que los códecs de audio generalmente transmiten este modelo explícitamente, con el coste de aumento de consumo de bits.

15

20 La era de Internet de las Cosas (IoT) se acerca, con lo que la siguiente generación de codificadores de voz y audio debería incluirlos. Los objetivos de diseño de los sistema de IoT sin embargo encajan mal con el diseño clásico de codificadores de voz y audio, con lo que se requiere un rediseño más grande de los codificadores.

25 En primer lugar, mientras que codificador de voz y audio del estado de la técnica tal como AMR-WB, EVS, USAC y AAC constan de codificadores inteligentes y complejos y decodificadores relativamente simples [1 - 4], ya que IoT debería soportar nodos de sensor de baja complejidad, se prefiere que los codificadores sean simples.

30 En segundo lugar, ya que los nodos de sensores están codificando la misma señal de fuente, aplicación de la misma cuantificación en cada nodo de sensor representaría sobrecodificación y potencialmente una pérdida grave de eficiencia. Especialmente, ya que el modelo perceptual debería ser más o menos el mismo en cada nodo, transmitir el mismo desde cada modo es casi sobrecodificación pura.

35

Los procedimientos de codificación de voz y audio convencionales constan de tres partes:

1. Un modelo perceptual que especifica el impacto relativo de errores en diferentes parámetros del códec,
2. Un modelo de fuente que describe el intervalo y probabilidad de diferentes entradas y
3. Un codificador por entropía que usa el modelo de fuente para minimizar la distorsión perceptual [5].

Además, el modelo perceptual puede aplicarse en cualquiera de dos formas:

1. Parámetros de señal pueden ponderarse de acuerdo con el modelo perceptual, de tal forma que todos los parámetros pueden entonces cuantificarse con la misma precisión. El modelo perceptual debe transmitirse a continuación al decodificador de tal forma que la ponderación puede deshacerse.
2. El modelo perceptual puede aplicarse como alternativa como un modelo de evaluación, de tal forma que la salida de síntesis de diferentes cuantificaciones se comparan, ponderadas por el modelo perceptual, en una iteración de análisis por síntesis. Aunque en este punto el modelo perceptual no se ha transmitido, este enfoque tiene la desventaja de que las formas de células de cuantificación no tienen una forma regular lo que reduce la eficiencia de codificación. Más importante, sin embargo, para encontrar la cuantificación óptima, tiene que usarse una búsqueda de fuerza bruta computacionalmente compleja de diferentes cuantificaciones.

Ya que el enfoque de análisis por síntesis por lo tanto conduce a un codificador computacionalmente complejo, no es una alternativa viable para IoT. Por lo tanto, el decodificador debe tener acceso al modelo perceptual. Sin embargo, como se ha indicado anteriormente, la transmisión explícita del modelo perceptual (o de forma equivalente, un modelo de envolvente del espectro de señal) no es deseable porque reduce la eficiencia de codificación.

El objetivo de la invención es presentar una forma de recuperar el modelo perceptual en el decodificador desde la señal transmitida sin información secundaria con referencia al modelo perceptual.

El objetivo se consigue mediante un aparato de codificación para el procesamiento de una señal de entrada, así como mediante un aparato de decodificación para el procesamiento de una señal codificada. El objetivo también se consigue mediante correspondientes  
5 procedimientos y un sistema que comprende el aparato de codificación y el aparato de decodificación.

La invención es especialmente útil en redes de sensores distribuidas y la Internet de las Cosas, en las que el coste añadido de consumo de bits de la transmisión de un modelo  
10 perceptual aumenta con el número de sensores.

La invención de acuerdo con una realización también podrían etiquetarse como una recuperación ciega de modelos perceptuales en codificación de voz y audio distribuida.

15 La señal de entrada en este punto es una señal de voz y/o de audio a codificar mediante el aparato de codificación.

El objetivo se consigue mediante un aparato de codificación para el procesamiento de una señal de entrada. El aparato de codificación preferentemente proporciona una señal  
20 codificada.

El aparato de codificación comprende un ponderador perceptual y un cuantificador.

El ponderador perceptual comprende un proveedor de modelo y un aplicador de modelo. El  
25 proveedor de modelo se configura para proporcionar un modelo ponderado perceptual basado en la señal de entrada. El aplicador de modelo se configura para proporcionar un espectro ponderado perceptualmente aplicando el modelo ponderado perceptual a un espectro basado en la señal de entrada.

30 El cuantificador se configura para cuantificar el espectro ponderado perceptualmente y para proporcionar una secuencia de bits. El cuantificador comprende un aplicador de matriz aleatoria y un calculador de función de signo. El aplicador de matriz aleatoria se configura para aplicar una matriz aleatoria al espectro ponderado perceptualmente para proporcionar un espectro transformado. Un calculador de función de signo se configura para calcular una  
35 función de signo (o: función signo) de componentes del vector transformado para proporcionar la secuencia de bits.

La cuantificación comprende al menos dos etapas: en una primera etapa, el espectro ponderado perceptualmente se combina con una matriz aleatoria. Una matriz aleatoria de este tipo tiene la ventaja de que para cada señal de entrada se usa una matriz diferente. Esto es útil cuando una pluralidad de sensores cubre la misma fuente de audio y tiene que evitarse la sobrecodificación. La segunda etapa incluye que la función de signo se calcula de los componentes del vector transformado.

En una realización, el proveedor de modelo se configura para proporcionar el modelo ponderado perceptual basado en una compresión del espectro basado en la señal de entrada.

En el aparato de codificación inventivo la ponderación perceptual de la señal de entrada se basa en una realización en una compresión de la señal de entrada. En una realización, la ponderación perceptual se basa en la compresión de la envolvente del espectro de magnitud de la señal de entrada (una envolvente es una forma continua y normalmente suave que describe una característica de la señal, en este punto del espectro de magnitud). Basado en la compresión se obtiene un modelo ponderado perceptual que se usa finalmente para ponderar perceptualmente el espectro de la señal de entrada.

El aparato de codificación proporciona en una realización una secuencia de bits con información secundaria que cubre algunos aspectos acerca del procedimiento de codificación.

En una realización, se proporciona información acerca de la cuantificación como información secundaria por la señal codificada emitida por el aparato de codificación como resultado del procedimiento de codificación.

En una realización adicional, el ponderador perceptual comprende un calculador de envolvente. El calculador de envolvente se configura para proporcionar una envolvente de un espectro de magnitud basado en la señal de entrada.

En una realización, el proveedor de modelo se configura para calcular una función de compresión que describe la compresión de la envolvente. Además, el proveedor de modelo se configura para calcular el modelo ponderado perceptual basado en la función de compresión. En esta realización, el proveedor de modelo comprime la envolvente y calcula una función que describe esta compresión. Basándose en esta función, se obtiene el modelo

ponderado perceptual. En esta realización, una compresión de una envolvente de un espectro basado en la señal de entrada se realiza, es decir el intervalo de magnitud se reduce y, por lo tanto, más pequeña que antes de la compresión. Comprimiendo la envolvente, el espectro como tal también se comprime, es decir el intervalo de magnitud del espectro se reduce.

En una realización diferente, el modelo ponderado perceptual o la función de compresión se calcula directamente a partir de la señal de entrada o el espectro de magnitud/potencia basado en la señal de entrada.

De acuerdo con una realización, el proveedor de modelo se configura para calcular una función de compresión que describe una compresión del espectro basado en la señal de entrada o que describe una compresión de una envolvente de un espectro de magnitud basado en la señal de entrada. La compresión reduce un intervalo de magnitud del espectro basado en la señal de entrada o reduce el intervalo de magnitud de la envolvente. Además, el proveedor de modelo se configura para calcular el modelo ponderado perceptual basado en la función de compresión.

En una realización, la función de compresión - para comprimir por ejemplo el espectro o la envolvente - cumple con dos criterios:

Primero, la función de compresión está aumentando estrictamente. Esto implica que para cualquier valor escalar positivo y un valor arbitrariamente pequeño el valor de función para el valor escalar positivo es más pequeño que el valor de función para la suma del valor escalar positivo y del valor arbitrariamente pequeño.

Segundo, para un primer valor escalar positivo y un segundo valor escalar positivo mayor que el primer valor escalar positivo, una diferencia entre el valor de función para el segundo valor escalar positivo y el valor de función para el primer valor escalar positivo es más pequeña que una diferencia entre el segundo valor escalar positivo y el primer valor escalar positivo.

El objetivo también se consigue mediante un procedimiento para el procesamiento de una señal de entrada. Esta señal de entrada es preferentemente una señal de audio y/o una señal de voz.

El procedimiento para el procesamiento de una señal de entrada - siendo una señal de audio y/o voz - comprende al menos las siguientes etapas:

5 calcular un modelo ponderado perceptual basado en la señal de entrada, proporcionar un espectro ponderado perceptualmente aplicando el modelo ponderado perceptual a un espectro basado en la señal de entrada, y cuantificar el espectro ponderado perceptualmente para proporcionar una secuencia de bits, en el que cuantificar el espectro ponderado perceptualmente comprende: aplicar una matriz aleatoria al espectro ponderado perceptualmente para proporcionar un espectro transformado, y calcular una función de signo de componentes del espectro transformado para proporcionar la secuencia de bits.

En una realización, se realizan las siguientes etapas:

15 calcular una envolvente de un espectro de magnitud basado en la señal de entrada, calcular el modelo ponderado perceptual basado en una compresión de la envolvente.

Las realizaciones del aparato de codificación también pueden realizarse mediante etapas del procedimiento y correspondientes realizaciones del procedimiento. Por lo tanto, las explicaciones proporcionadas por las realizaciones del aparato también se mantienen para el procedimiento.

20 El objetivo se consigue adicionalmente mediante un procedimiento para el procesamiento de una señal de entrada, que comprende:

proporcionar un modelo de ponderación perceptual basado en la señal de entrada, ponderar un espectro de la señal de entrada aplicando el modelo de ponderación perceptual al espectro de la señal de entrada, y  
25 cuantificar el espectro ponderado calculando una función de signo de proyecciones aleatorias del espectro ponderado.

El procedimiento comprende adicionalmente en una realización:

30 obtener las proyecciones aleatorias del espectro ponderado aplicando una matriz aleatoria al espectro ponderado.

En una realización, proporcionar el modelo de ponderación perceptual comprende comprimir una envolvente de un espectro de magnitud de la señal de entrada.

35 De acuerdo con una realización, el procedimiento comprende adicionalmente: obtener la envolvente usando un banco de filtros y una matriz diagonal que contiene factores de

normalización para cada banda.

El objetivo también se consigue mediante un aparato de decodificación para el procesamiento de una señal codificada.

5

El aparato de decodificación comprende al menos un decuantificador y un deponderador perceptual.

10

El decuantificador se configura para decuantificar una secuencia de bits comprendida por la señal codificada y para proporcionar un espectro ponderado calculado perceptualmente. Además, el decuantificador se configura para decuantificar la secuencia de bits aplicando una pseudo-inversa de una matriz aleatoria a la secuencia de bits. El decuantificador revierte los efectos de una cuantificación que sucedió durante el procedimiento de codificación. El decuantificador se sigue mediante el deponderador perceptual de tal forma que el espectro resultante de la decuantificación se depondera perceptualmente. El espectro ponderado calculado perceptualmente se recibe por el deponderador perceptual para deponderarse perceptualmente. El espectro finalmente obtenido es, por lo tanto, una secuencia de bits decuantificada deponderada perceptualmente comprendida por la señal de entrada.

20

El deponderador perceptual revierte los efectos de una ponderación perceptual que sucedió durante el procedimiento de codificación que conduce a la señal codificada. Esto se hace en una realización sin información secundaria de la señal codificada que comprende el modelo ponderado perceptual. El modelo se reconstruye a partir de la señal de audio codificada como tal.

25

El deponderador perceptual comprende un aproximador de espectro y un aproximador de modelo.

30

La reconstrucción del modelo se hace en una realización de una forma iterativa para la que se requieren valores de inicio o iniciales. Por lo tanto, un proveedor de pronóstico inicial comprendido por el deponderador perceptual se configura para proporcionar datos para un pronóstico inicial de un modelo ponderado perceptual con el que se asocia la señal codificada. Los datos para el pronóstico inicial comprenden en una realización un vector con los elementos diagonales de una matriz que describe el modelo ponderado perceptual.

35



El aproximador de espectro se configura para calcular una aproximación de un espectro basada en el espectro ponderado calculado perceptualmente. Además, el aproximador de modelo se configura para calcular una aproximación de un modelo ponderado perceptual con el que se asocia la señal codificada (es decir que se usó para la codificación de la señal de entrada y, por lo tanto, generación de la señal codificada) basándose en la aproximación de un espectro.

En una realización, el deponderador perceptual comprende un proveedor de pronóstico inicial. El proveedor de pronóstico inicial se configura para proporcionar datos para un pronóstico inicial del modo ponderado perceptual. El aproximador de espectro se configura para calcular la aproximación del espectro basada en la señal codificada y el pronóstico inicial o la aproximación del modelo ponderado perceptual. Además, la aproximación de un espectro se basa en una realización o bien en el pronóstico inicial o bien en una aproximación - especialmente calculada - del modelo ponderado perceptual. Esta elección depende del hecho de si la iteración acaba de comenzar - con el pronóstico inicial - o al menos una iteración con al menos una mejora de la aproximación - conduciendo a la aproximación del modelo ponderado perceptual - ya ha sucedido. La iteración se realiza en una realización hasta que se ha cumplido con un criterio de convergencia.

El decuantificador se configura en una realización para decuantificar la secuencia de bits comprendida por la señal codificada en base una información secundaria acerca de la matriz aleatoria comprendida por la señal codificada. Esta realización se refiere al procedimiento de codificación en el que una cuantificación se realiza usando una matriz aleatoria. La información acerca de la matriz aleatoria usada se comprende por la información secundaria de la señal codificada. La información secundaria comprende en una realización solo la semilla de las columnas de la matriz aleatoria.

En una realización, una pluralidad de señales codificadas se procesan conjuntamente. Cada señal codificada comprende una secuencia de bits con información secundaria con referencia a al menos información acerca de la cuantificación realizada mientras proporciona la respectiva señal codificada. Para este fin, el decuantificador se configura para recibir una pluralidad de señales de entrada y para proporcionar un y en una realización solo un espectro ponderado calculado perceptualmente basado en la pluralidad de señales de entrada. Las señales de entrada preferentemente se refieren a la señal de audio/voz que provienen de la misma fuente de señal.

En otra realización, el aparato de decodificación se configura para aplicar modelización de fuente. Un modelo de fuente describe el intervalo y probabilidad de diferentes entradas.

5 El objetivo también se consigue mediante un procedimiento para el procesamiento de una señal codificada. Esta señal codificada es preferentemente una señal de audio codificada y/o una señal de voz codificada.

El procedimiento para el procesamiento (o decodificación) de una señal codificada comprende al menos las siguientes etapas:

10

decuantificar una secuencia de bits comprendida por la señal codificada y proporcionar un espectro ponderado calculado perceptualmente,

en el que decuantificar la secuencia de bits comprende aplicar una pseudo-inversa de una matriz aleatoria a la secuencia de bits,

15

calcular una aproximación de un espectro basada en el espectro ponderado calculado perceptualmente, y

calcular una aproximación de un modelo ponderado perceptual con el que se asocia la señal codificada (es decir que se usó para generar la señal codificada) basándose en la aproximación de al espectro.

20

En una realización, el procedimiento comprende las etapas:

calcular una aproximación del espectro o bien basada en el espectro ponderado calculado perceptualmente y un pronóstico inicial o bien basada en el espectro ponderado calculado perceptualmente y una aproximación del modelo ponderado perceptual con el que se asocia la señal codificada.

25

Las aproximaciones calculadas del modelo ponderado perceptual se usan preferentemente para el siguiente cálculo de la aproximación de un espectro.

30 El objetivo también se consigue mediante un procedimiento para el procesamiento de una señal codificada,

que comprende:

proporcionar una señal perceptual cuantificada aplicando una pseudo-inversa de una matriz aleatoria a la señal codificada,

35

calcular una estimación de un espectro basado en la señal perceptual cuantificada, y  
calcular una aproximación de un modelo de ponderación perceptual usada para

proporcionar la señal codificada basada en la aproximación del espectro.

De acuerdo con una realización, el procedimiento comprende adicionalmente:

5 proporcionar una cero-ésima aproximación del modelo de ponderación perceptual usando un pronóstico inicial, y

calcular una cero-ésima estimación del espectro basado en la cero-ésima aproximación del modelo de ponderación perceptual.

En una realización adicional, el procedimiento comprende adicionalmente:

10 obtener el pronóstico inicial usando un banco de filtros y una matriz diagonal que contiene factores de normalización para cada banda.

15 Las realizaciones del aparato también pueden realizarse mediante etapas del procedimiento y correspondientes realizaciones del procedimiento. Por lo tanto, las explicaciones proporcionadas por las realizaciones del aparato también se mantienen para el procedimiento.

20 El objetivo también se consigue mediante un sistema que comprende al menos un aparato de codificación y un aparato de decodificación. En una realización, se usa una pluralidad de aparatos de codificación que en una realización se asocian con nodos de sensores, por ejemplo micrófonos.

25 El objetivo también se consigue mediante un programa informático para la realización, cuando se ejecuta en un ordenador o un procesador, del procedimiento de cualquier de las anteriores realizaciones.

La invención se explicará a continuación con respecto a los dibujos adjuntos y las realizaciones representadas en los dibujos adjuntos, en los que:

30 La Figura 1 muestra un diagrama de bloques de un sistema de primera realización que comprende un aparato de codificación y un aparato de decodificación,

La Figura 2 muestra un diagrama de bloques de una realización de un aparato de codificación,

La Figura 3 muestra un diagrama de bloques de un ponderador perceptual como una parte de un aparato de codificación,

35 La Figura 4 muestra señales que pertenecen a un procedimiento de codificación,

La Figura 5 muestra un diagrama de bloques de una segunda realización de un

- sistema que comprende una multitud de aparatos de codificación,
- La Figura 6 muestra un diagrama de bloques más detallado de una tercera realización de un sistema,
- La Figura 7 muestra un diagrama de bloques de una realización de un aparato de decodificación,
- La Figura 8 muestra un diagrama de bloques de una realización diferente de un aparato de decodificación,
- La Figura 9 muestra una realización de un deponderador perceptual como parte de un aparato de decodificación,
- La Figura 10 muestra señales que pertenecen a un procedimiento de decodificación,
- La Figura 11 muestra valores SNR medios con diferentes tasas de bits para datos experimentales y
- La Figura 12 muestra puntuaciones MUSHRA diferenciales para datos experimentales.
- La Figura 1 muestra un sistema para tratar una señal de entrada 3 que es una señal de voz y/o audio. Esta señal de entrada 3 se codifica mediante un aparato de codificación 1. Siendo la señal codificada 4 preferentemente una secuencia de bits se transmite, por ejemplo, a través de la Internet a un aparato de decodificación 2 que decodifica la señal codificada 4 y proporciona una señal de audio extraída 5 a un oyente - no mostrado.
- El aparato de codificación 1 usa un modelo ponderado perceptual para el procesamiento de la señal de entrada 3 pero este modelo no se transmite por la señal codificada 4. El aparato de decodificación 2 extrae el modelo de la señal codificada 4 para deshacer los efectos del modelo.
- El aparato de codificación 1 mostrado en la Figura 2 comprende un transformador 15, un ponderador perceptual 10 y un cuantificador 14.
- El transformador 15 proporciona un espectro  $s$  basado en la señal de entrada 3 siendo una señal de tiempo. Esto se hace, por ejemplo, mediante la transformada de Fourier de tiempo corto (STFT).
- El espectro  $s$  se somete a una ponderación perceptual mediante el ponderador perceptual 10 volviéndose espectro ponderado perceptualmente  $x$ . Este espectro  $x$  se envía al cuantificador 14 que cuantifica el mismo y proporciona una señal cuantificada  $\hat{u}$  siendo una secuencia de bits. Esa señal cuantificada  $\hat{u}$  se combina en esta realización con información

secundaria correspondiente que cubre información acerca de la cuantificación - pero en esta realización no acerca del modelo ponderado perceptual - para proporcionar la señal codificada 4.

5 La Figura 3 muestra una realización del ponderador perceptual 10.

El ponderador perceptual 10 recibe la señal de entrada 3 - o un espectro correspondiente después de una transformación de la señal de entrada 3 al dominio de la frecuencia - y proporciona espectro ponderado perceptualmente  $x$ . Para este fin, el ponderador perceptual 10 comprende un calculador de envolvente 11, un proveedor de modelo 12 y un aplicador de modelo 13.

El calculador de envolvente 11 recibe la señal de entrada 3 o un espectro correspondiente y proporciona una envolvente  $y$  y de un espectro de magnitud  $|x|$  basado en la señal de entrada 3. En la realización mostrada, el calculador de envolvente 11 proporciona la envolvente  $y$  mediante una matriz  $A$  y una matriz diagonal  $\Lambda$ . La matriz  $A$  es un banco de filtros y la matriz diagonal  $\Lambda$  contiene factores de normalización para cada banda de la banda de filtros usada. La envolvente  $y$  se obtiene a continuación basado en el espectro de magnitud  $|x|$  de la señal de entrada 3 mediante la ecuación  $y = A\Lambda A^T|x|$ , mientras que  $A^T$  es la transpuesta de la matriz  $A$ .

Basado en esta envolvente  $y$ , el proveedor de modelo 12 comprime la envolvente  $y$ . El fin de la compresión es obtener una función que se aproxima al rendimiento de percepción del oído. A continuación, se analiza una realización en la que la envolvente se comprime elevando  $y$  a la potencia de  $p$ . Cuando, por ejemplo,  $p$  es igual a 0,3, entonces el intervalo de  $y^p$  será más pequeño entonces que el intervalo de la  $y$  original. La cantidad de compresión, en este ejemplo, por lo tanto depende del valor  $p$ . Por ejemplo, la envolvente se reduce a un intervalo deseado. En una realización, una reducción de intervalo o una función de compresión para comprimir la envolvente  $y$  se proporciona mediante  $y^p$  con  $0 < p < 1$ . Esto significa que la compresión se realiza mediante una función exponencial de la envolvente con un exponente mayor cero y uno menor. Esta compresión se realiza en una realización de muestra.

La función de compresión  $f(y)$  describe un vector  $w$ , es decir  $w = f(y)$  que proporciona elementos diagonales de un modelo ponderado perceptual  $W$ . En el presente documento, la

entrada y se proporciona como un vector y función  $f$  se aplica en todas las muestras del vector  $y$ , resultando en un vector  $w$ . Así si la  $k^{\text{ésima}}$  muestra de  $y$  es  $y_k$ , entonces la  $k^{\text{ésima}}$  muestra de  $w = f(y)$  es  $w_k = f(y_k)$ .

- 5 Por lo tanto, basado en la función de compresión, el modelo ponderado perceptual  $W$  puede obtenerse, en este punto en forma de una matriz.

En otras palabras: la envolvente del espectro de magnitud se comprime y a partir de la función que describe la envolvente comprimida, se calcula el modelo ponderado perceptual que se usa para ponderar perceptualmente el espectro.

10

El aplicador de modelo 13 aplica el modelo ponderado perceptual  $W$  al espectro  $s$  basado en la señal de entrada 3. En la realización mostrada, el aplicador de modelo 13 aplica la matriz del modelo ponderado perceptual  $W$  a un vector basándose en el espectro.

15

La modelización perceptual se explicará ahora una vez más:

Códecs de voz y audio se basan en modelización eficiente de percepción auditiva humana. El objetivo es obtener una ponderación tal de errores de cuantificación que la optimización de la relación señal-ruido en el dominio ponderado proporciona la mejor calidad perceptivamente posible.

20

Códecs de audio operan generalmente en el dominio espectral, en el que el espectro de una de trama entrada  $s$  puede ponderarse perceptualmente con una matriz diagonal  $W$  de tal forma que el espectro ponderado  $x = Ws$  puede cuantificarse  $\hat{x} = [Ws]$ , donde los corchetes  $[ \ ]$  representan cuantificación.

25

En el decodificador, la operación inversa  $\hat{s} = W^{-1}\hat{x}$  puede reconstruirse.

Específicamente, el modelo de ponderación perceptual que consiste en dos partes:

30

- i) Una parte fija que corresponde a los límites de percepción en diferentes bandas de frecuencia. Modelos perceptuales tal como escalas Bark y ERB modelan la densidad de frecuencias de tal forma que el eje combado tiene precisión perceptual uniforme [17]. Sin embargo, ya que un objetivo es medir energía de error en una escala combada, la magnitud de componentes espectrales puede escalarse de forma equivalente de tal forma que la operación de combado computacionalmente compleja puede evitarse [18].
- 35

Esta operación es también similar a la operación de preénfasis aplicada en los códecs de voz [1 - 3]. Ya que esta parte de la ponderación es fija, no necesita transmitirse explícitamente. Puede aplicarse en el codificador y directamente revertida en el decodificador.

5

- ii) La parte adaptativa de señal del modelo perceptual corresponde a las propiedades de enmascaramiento de frecuencia de percepción. En concreto, componentes de alta energía de la señal enmascararán componentes de energía más baja y por lo tanto los convertirá en inaudibles, si los dos están lo suficientemente cerca en frecuencia [5]. La forma de la curva de enmascaramiento de frecuencia es por lo tanto igual a la forma de la señal envolvente, pero con una magnitud más pequeña.

10

Si  $|x|$  es el espectro de magnitud de la señal de entrada, su envolvente espectral y puede obtenerse en una realización mediante  $y=A\Lambda A^T|x|$ , en la que la matriz  $A$  es un banco de filtros tal como en la Figura 4 (a).

15

En diferencia a los bancos de filtros de tipo MFCC [19], se usa una ventana de tipo ventana Hann asimétrica en una realización con una superposición que se extiende desde el  $k^{\text{ésimo}}$  filtro hasta los filtros  $(k - 2)$  y  $(k + 2)$  (véase la Figura 4 (a)).

20

La matriz diagonal  $\Lambda$  contiene factores de normalización para cada banda de tal forma que se obtiene ganancia por unidad.

Dependiendo de la realización, se usa una escala Mel, Bark o ERB con un número de bandas adecuado.

25

En una tasa de muestreo de 12,8 kHz, en este punto se usó un banco de filtros Mel con 20 bandas.

Una alternativa a la matriz de banco de filtros de tipo MFCC es usar ensanchamiento por filtrado, con lo que  $A$  se convierte en una matriz de convolución. Ya que operaciones de filtrado son procedimientos de procesamiento de señales digitales bien conocidos, sus inversos se encuentran fácilmente también.

Los factores de ponderación perceptuales modelan el efecto de enmascaramiento de frecuencia, que a su vez corresponde a ensanchamiento y escalado de energía en

35

frecuencias [20, 5]. La matriz A de modelo envolvente ya consigue el efecto de ensanchamiento, con lo que aún tiene que modelarse el escalado de energía.

5 El escalado de energía corresponde una compresión de la señal, que reduce el intervalo de magnitud de la envolvente (véase la Figura 4 (b)). Por lo tanto, si el espectro s se multiplica con la ponderación perceptual matriz W, se obtiene un espectro  $x = Ws$  que tiene un intervalo reducido (véase la Figura 4 (c)).

10 La ponderación perceptual, por lo tanto, reduce el intervalo o aplanar el espectro, pero no produce un espectro con una envolvente plana en su totalidad. El intervalo de la envolvente se reduce, con lo que una parte de su intervalo se retiene, y es intervalo restante puede usarse para recuperar la señal original que sigue a la envolvente expandida.

15 La reducción de intervalo o función de compresión  $w=f(y)$  para la envolvente y (en la que el vector w proporciona los elementos diagonales de W), puede aplicarse por ejemplo como un exponente de muestras  $f(y)=y^p$  con  $0 < p < 1$ .

20 Dado un espectro s y su k<sup>ésima</sup> muestra  $s_k$ , se aplica la ponderación mediante multiplicación de tal forma que  $x_k = w_k * s_k$ , donde  $x_k$  es la k<sup>ésima</sup> muestra del espectro ponderado x y  $w_k$  es la k<sup>ésima</sup> muestra del vector de ponderación w. La misma operación puede expresarse como una operación de matriz generando una matriz W, que tiene los valores de ponderación en la diagonal  $W_{kk} = w_k$ , y la matriz es cero en todas las demás posiciones. Se deduce que  $x = W*s$ .

25 Mientras es posible usar cualquier función que comprime el intervalo de y, la exponenciación tiene el beneficio de que conduce a una expresión analítica simple en la reconstrucción de envolvente en el decodificador.

Funciones de compresión adecuadas f() cumplen con los siguientes requisitos:

30

1. Las funciones de compresión están aumentando estrictamente, es decir  $f(t) < f(t+\epsilon)$ , donde t es cualquier valor escalar positivo y  $\epsilon$  es un valor arbitrariamente pequeño,
  2. Para cualquier valor escalar positivo (primer y segundo valor escalar,  $t_1, t_2$ )  $t_1 < t_2$ , se mantiene  $f(t_2) - f(t_1) < t_2 - t_1$ . En otras palabras, cualquier función que reduce la distancia
- 35 entre tales dos valores escalares positivos  $t_2 - t_1$  es una función adecuada.

Además de exponenciación  $f(y) = y^p$ , con un exponente pequeño p, en una realización



diferente, la función de compresión es el logaritmo, es decir,  $f(y) = \log(y)$ .

El algoritmo de codificador como un procedimiento de codificación o realizado por el aparato de codificación es entonces en una realización:

- 5 1. Calcular la envolvente del espectro de magnitud.
2. Comprimir la envolvente para obtener el modelo de ponderación perceptual.
3. Aplicar ponderación en espectro  $x = Ws$ .
4. Cuantificar y transmitir espectro ponderado signo(Px).

Este algoritmo se aplica independientemente en cada nodo de sensor.

10 La ponderación perceptual se sigue mediante una cuantificación.

La invención, por lo tanto, que consiste en dos partes:

1. Cuantificación distribuida de la señal de entrada usando proyecciones aleatorias y cuantificación de 1 bit, y
- 15 2. Transmisión implícita del modelo perceptual.

Mediante la cuantificación de proyecciones aleatorias, cada bit transmitido codifica una pieza única de información y se evita la sobre codificación.

20 El modelo perceptual se genera independientemente en cada nodo de sensor (por ejemplo que comprende un micrófono) y se transmite la señal ponderada perceptualmente cuantificada. La ponderación perceptual hace la señal más plana, pero se retiene la forma básica. Por lo tanto, puede deducirse inversamente en el lado de decodificador qué debió haber sido la envolvente original, incluso a partir de la señal ponderada perceptualmente.

25

A continuación una descripción de la cuantificación distribuida:

Aunque la codificación de fuente distribuida es un tema bien estudiado (por ejemplo [7, 8]) y se ha aplicado en otras aplicaciones tal como video [9], únicamente unos pocos han trabajado en codificación de audio distribuida (por ejemplo [10 - 13]), y ninguno de los mismos sin embargo aborda el problema de la sobre codificación con respecto a modelos perceptuales y de envolvente. Incluso el enfoque de codificación escalable en [14] incluye codificación de envolvente con factores de escala. También el enfoque de codificación de descripción múltiple se ha aplicado únicamente la ocultación de pérdida de paquetes [15,

35 16].

A continuación un esquema de cuantificación que se puede implementar fácilmente. Compárese el procedimiento de cuantificación de 1 bit que se ha usado en sistemas de detección compresivos [6].

- 5 El objetivo del cuantificador y el procedimiento de cuantificación es permitir la cuantificación en sensores independientes, de tal forma que se garantiza que cada bit transmitido mejora la calidad, sin comunicación entre nodos de sensores. En el extremo, un sensor podría enviar únicamente un bit y ese único bit podría usarse para mejorar la calidad.
- 10 El esquema de cuantificación propuesto de una realización se basa en proyecciones aleatorias de una representación de valor real del espectro de señal y transmisión del signo de cada dimensión.

15 Sea  $x$  el  $N$  veces  $1$  vector de valor real que contiene el espectro de la señal de entrada, y  $P$  una  $K$  veces  $N$  matriz aleatoria cuyas columnas se normalizan a una longitud de unidad. A continuación,  $x$  se transformará mediante  $u = Px$ . Esto se sigue mediante una cuantificación del signo de cada componente de  $u$ , es decir, la cuantificación es  $\hat{u} = \text{signo}(u)$ , que puede transmitirse sin pérdidas con  $K$  bits.

20 El número de bits de la secuencia de bits, por lo tanto, define una dimensión de la matriz aleatoria.

Las muestras de  $P$  son preferentemente valores pseudo-aleatorios, que significa que parecen valores aleatorios, pero en realidad se generan mediante alguna fórmula matemática compleja o algoritmo. Los generadores pseudo-aleatorios son herramientas matemáticas estándar que tienen cada ordenador o cada librería de software matemático. Relevantes son: que la matriz  $P$  tiene que conocerse tanto en el codificador como en el receptor/decodificador, y que la distribución de los números aleatorios es la misma para todas las muestras en  $P$ .

30 La reconstrucción como una aproximación de  $x$  puede calcularse fácilmente mediante

$$\hat{x} = P^\dagger \hat{u} = P^\dagger \text{sign}(Px) \quad (1)$$

donde  $P^\dagger$  es la pseudo-inversa de la matriz aleatoria  $P$ .

35 Siempre que la semilla para las columnas pseudo-aleatorias de  $P$  se conozca en el decodificador, el decodificador puede decodificar por lo tanto la señal a partir de únicamente

- $\hat{u}$ . Por lo tanto, en una realización, la semilla de las columnas pseudo-aleatorias se proporcionan como información secundaria de la señal codificada. Generadores pseudo-aleatorios normalmente generan secuencias de valores aleatorios de tal forma que dado el valor previo  $x(k)$  en la secuencia, genera la siguiente muestra aleatoria,  $x(k+1) = f(x(k))$ . Es decir, si se conoce el punto de inicio  $x(1)$  - llamado el valor "semilla" de la secuencia pseudo-aleatoria -, entonces es posible generar toda la secuencia. Por lo tanto, en el lado de codificación y decodificación se usa la misma función de generación de las muestras aleatorias.
- 5
- 10 En el caso de múltiples nodos de sensores, la señal de entrada  $x$  se supone que es la misma o versiones ruidosas de la misma señal, pero cada sensor tiene su propia matriz aleatoria  $P_k$ . En el decodificador las matrices aleatorias pueden compilarse a una única gran matriz  $P = [P_1, P_2, \dots]$  con lo que la Ecuación 1 permanece sin cambios.
- 15 Es bien conocido que si  $K \ll N$ , entonces  $P$  es aproximadamente ortonormal,  $P^T P \approx I$  y la cuantificación es casi óptima.

En el presente documento,  $K$  no es necesariamente más pequeño que  $N$ , con lo que la ortonormalidad se vuelve menos precisa. Usando la transpuesta en lugar de la pseudo-inversa disminuye la complejidad algorítmica y eficiencia de codificación, pero no impone una limitación a nuestros experimentos de modelización perceptual, ya que cada bit transmitido aún mejora la precisión de la señal de salida.

20

Puede esperarse que un modelo de fuente se aplicaría entonces en el lado de decodificador y que un modelo de este tipo aumentaría la precisión de la reconstrucción. Sin embargo no es necesario implementar un modelo de fuente, ya que su efecto puede simularse aumentando la precisión enviando más bits.

25

El diagrama de flujos de una realización del sistema (excluyendo el modelo perceptual) se ilustra en la Figura 5. Se muestran  $n$  micrófonos que capturan las señales de audio de una fuente de audio. Los siguientes aparatos codificadores 1 son en la realización mostrada parte del respectivo nodo micrófono o sensor.

30

Las  $n$  señales de entrada 3 se transforman por los transformadores 15 de los  $n$  aparatos de codificación 1 en  $n$  espectros  $s$  que se convierten en  $n$  espectros ponderados perceptualmente  $x_k = W_k s_k$  que se hace en este punto mediante los ponderadores

35

perceptuales no mostrados.

Los  $n$  espectros ponderados perceptualmente  $x_k$  se someten al  $n$  cuantificador 14.

5 Cada cuantificador 14 comprende un aplicador de matriz aleatoria 16 y un calculador de función de signo 17.

10 Los aplicadores de matriz aleatoria 16 aplican una matriz aleatoria  $P_i$  que es una matriz con dimensión  $K$  veces  $N$  a un  $N$  veces  $1$  vector basándose en el respectivo espectro ponderado perceptualmente  $x_i$  con  $i = 1, 2, \dots, n$ .  $N$  es un escalar entero que corresponde al número de muestras en el espectro  $x$ .  $K$  es un escalar entero que corresponde al número de filas en el espectro aleatorizado. Mediante una cuantificación con el operador de signo, cada fila se cuantifica con un bit, con lo que el número de bits transmitidos es  $K$ . Cuando se usan múltiples codificadores - por ejemplo micrófonos -, entonces cada codificador tiene su propia matriz  $P_k$ , que tienen tamaño de  $K_k \times N$ . Es decir, cada codificador envía  $K_k$  bits al

15 decodificador, donde el número de bits puede variar de codificador a codificador.

Los espectros transformados se proporcionan mediante:  $u_i = P_i x_i$ .

20 Un calculador de función de signo 17 calcula el signo o función de signo del respectivo espectro transformado:  $\hat{u}_i = \text{signo}(u_i)$ . Esto se hace a los  $K$  componentes de los espectros transformados que conducen a  $K$  bits transmitiéndose a través de un canal de transmisión.

25 La Figura 6 muestra un aparato de codificación 1 que comprende el ponderador perceptual 10 y el cuantificador 14.

La señal de audio 3 se transforma mediante un transformador 15 en un espectro  $s_k$ . El ponderador perceptual 10 aplica la ponderación perceptual matriz  $W$  al espectro  $s_k$  para proporcionar el espectro ponderado perceptualmente  $x_k$  que se cuantifica mediante el cuantificador 14.

30 El cuantificador 14 comprende el aplicador de matriz aleatorias 16 que recibe el espectro ponderado perceptualmente  $x_k$  y aplica la matriz aleatoria  $P_k$  al mismo a través de  $P_k x_k$ . Los componentes del espectro transformado resultante  $u_k$  se cuantifican mediante un calculador de función de signo 17 a través del cálculo de la función de signo de cada componente. Esto

35 conduce a unas secuencias de bits con número de bits que igualan el número de componentes del espectro transformado. La señal codificada 4 se proporciona, por lo tanto,

mediante la secuencia de bits  $\hat{u}_k$ .

A continuación una descripción de la forma para decodificar las señales codificadas.

5 La Figura 7 muestra una realización de un aparato de decodificación 2 para el procesamiento de una señal codificada 4.

La señal codificada 4 es en la forma de la secuencia de bits  $\hat{u}$  decuantificado mediante un decuantificador 20.

10 La señal codificada 4 se decuantifica usando la pseudo-inversa  $P^\dagger$  de la matriz aleatoria P. La información acerca de la matriz aleatoria se toma de la información secundaria de la señal codificada 4. El espectro ponderado calculado perceptualmente se proporciona mediante:  $\hat{x} = P^\dagger \hat{u}$ .

15 El espectro ponderado calculado perceptualmente  $\hat{x}$  se de-ponderada por el deponderador perceptual 21. El espectro calculado  $\hat{s}$  proporcionado por el deponderador perceptual 21 se transforma mediante un detransformador 22 (por ejemplo a través de la transformada de Fourier de tiempo corto inversa, STFT-1) en una señal de tiempo que es la señal de audio extraída 5.

20

La Figura 8 muestra una realización en la que el decuantificador 20 recibe una pluralidad de señales codificadas 4 desde diferentes nodos de sensores, es decir desde diferentes micrófonos. Las matrices aleatorias individuales  $P_k$  se recopilan es una única gran matriz  $P = [P_1, P_2, \dots]$ .

25

El espectro ponderado calculado perceptualmente se proporciona a continuación mediante:

$$\hat{x} = \begin{bmatrix} P_1 \\ P_2 \\ \dots \\ P_n \end{bmatrix}^\dagger \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{bmatrix}$$

30 En una realización alternativa, las secuencias de bits se invierten con su respectiva matriz aleatoria  $\hat{x}_k = P_k^\dagger u_k$  y después se fusionan los espectros.

El espectro ponderado calculado perceptualmente obtenido  $\hat{x}$  se trata como se describe para la realización mostrada en la Figura 7.

En la Figura 9 se muestra una realización del deponderador perceptual 21 como parte de un aparato de decodificación.

5 El deponderador perceptual 21 recibe desde el decuantificador 20 el espectro ponderado calculado perceptualmente  $\hat{x}$  que es la secuencia de bits decuantificada de la señal codificada 4.

10 El deponderador perceptual 21 recupera el modelo ponderado perceptual  $W$  usado durante la codificación de la señal de entrada 3 usando un algoritmo iterativo. En el presente documento,  $W_k$  es la  $k^{\text{ésima}}$  aproximación o suposición del modelo ponderado perceptual  $W$ .

El algoritmo comienza con un proveedor de pronóstico inicial 23. El proveedor de pronóstico inicial 23 proporciona una cero<sup>ésima</sup> estimación  $W_0$  del modelo ponderado perceptual  $W$ .

15

En la realización mostrada, se usa el vector  $w_k$  que comprende los elementos diagonales de la respectiva matriz  $W_k$ . Un vector de este tipo se usó con la función de compresión de la envolvente del espectro de magnitud durante la ponderación perceptual de la señal de entrada.

20

Por lo tanto, se proporciona una suposición de 0<sup>ésima</sup> del vector  $w_0$  y la 0<sup>ésima</sup> aproximación  $W_0$  del modelo ponderado perceptual  $W$  se establece apropiadamente.

25 El pronóstico inicial del vector de aproximación  $w_0$  se establece en una realización usando la ecuación usada para el cálculo de la envolvente y del espectro de magnitud.

Así, el pronóstico inicial se vuelve:  $w_0 = A\Lambda A^T \hat{x}$  con la matriz  $A$  y una matriz diagonal  $\Lambda$ .

30 La matriz  $A$  es un banco de filtros y la matriz diagonal  $\Lambda$  contiene factores de normalización para cada banda de la banda de filtros usada. La elección de matriz  $A$  es una elección de diseño, que habitualmente se fijaría por adelantado. En una realización, en la que la matriz  $A$  no cambia con el paso del tiempo, está cableada tanto en el codificador como el decodificador. Esto implica que tanto el codificador como el decodificador usan la misma matriz  $A$  conocida. En una realización diferente, la matriz  $A$  usada para la codificación se transmite como parte de la señal codificada. En una realización, la elección con respecto a  
35 la matriz  $A$  se transmite en una etapa de inicialización o configuración, antes de que

comience la comunicación real. Lo anterior también se refiere a la matriz diagonal  $\Lambda$ .

El pronóstico inicial matriz  $W_0$  se proporciona a un aproximador de espectro 24. Esto también inicia la iteración real con un índice de ejecución  $k$  establecido a 0.

5

El aproximador de espectro 24 calcula basándose en los datos  $w_0$  el pronóstico inicial  $W_0$  del modelo ponderado perceptual  $W$  una aproximación del espectro.

10 Esto se hace a través de la aplicación de la inversa del pronóstico inicial matriz  $W_0$  a un vector que contiene el espectro ponderado calculado perceptualmente  $\hat{x}$ .

El espectro se aproxima en esta etapa con  $k = 0$  mediante  $\hat{s}_0 = W_0^{-1}\hat{x}$

15 El espectro aproximado  $\hat{s}_0$  se proporciona a un aproximador de modelo 25 que calcula una nueva aproximación  $W_1$  para el modelo ponderado perceptual  $W$ .

Esto se basa en el conocimiento de que el vector de los elementos diagonales  $w_k$  es una función del espectro  $\hat{s}_k$ .

20 En una realización, la función de compresión  $f()$ , que se usa durante la codificación para generar la matriz  $W$ , se conoce en el lado de decodificación. Como se ha analizado anteriormente con referencia a la matriz  $A$ , la función de compresión  $f()$  o bien se establece en ambos lados, por ejemplo se fija en el respectivo software para codificar y decodificar, o bien se transmite como parte de la señal codificada.

25

Por lo tanto, la siguiente aproximación del vector se calcula a través de:  $w_1 = f(\hat{s}_0)$  y se obtiene la siguiente - en este punto primera - aproximación  $W_1$ .

30 Esta aproximación  $W_1$  se devuelve al aproximador de espectro 24 para calcular el siguiente espectro aproximado  $\hat{s}_k = W_k^{-1}\hat{x}$ . El índice  $k$  se aumenta en consecuencia.

Esta aproximación sirve después para obtener la siguiente aproximación  $W_{k+1}$  del modelo ponderado perceptual  $W$  basada en el vector con elementos diagonales proporcionados mediante la ecuación:  $w_{k+1} = f(\hat{s}_k)$ .

35

Esto se repite hasta que converge. La convergencia normalmente se comprueba comparando el vector de salida actual  $w(k)$  con el vector anterior  $w(k-1)$ . Si la diferencia  $\|w(k) - w(k-1)\|$  está por debajo de algún umbral preestablecido, entonces la iteración ha convergido y las iteraciones pueden detenerse.

5

El espectro aproximado obtenido finalmente  $\hat{s}$  es la salida del deponderador perceptual 21 y es la aproximación del espectro de la señal de entrada original.

La reconstrucción del modelo perceptual se explica una vez más:

10

En el lado de decodificador, una estimada  $\hat{x}$  de la señal perceptual  $x$  (véase la Ecuación 1) puede recuperarse mediante  $\hat{x} = P^+ \text{signo}(Px)$  donde  $\hat{u} = \text{signo}(Px)$  es el espectro ponderado perceptualmente cuantificado  $x$ . Esto se hace mediante el decuantificador 20.

15

La principal tarea es por lo tanto recuperar una estimada  $\hat{s}$  de la señal original  $s$  a partir de la señal perceptual cuantificada  $\hat{x}$ .

El espectro ponderado perceptualmente  $x$  se basa en el espectro de la señal de entrada a través del modelo ponderado perceptual  $W$  mediante la ecuación:  $x = Ws$ .

20

El objetivo es que la estimada es igual al espectro, es decir  $\hat{x} \approx x$ .

Por lo tanto,  $Ws \approx W\hat{s}$ .

25

Adicionalmente, el vector  $w$  que proporciona los elementos diagonales de la matriz  $W$  son una función del espectro  $s$  de la señal de entrada:  $w = f(s)$ .

Por lo tanto, con una estimada de  $w$ ,  $\hat{s}$  puede estimarse, con lo que  $w$  puede estimarse. Esto puede iterarse hasta convergencia.

30

Esto es por lo tanto un algoritmo de tipo Esperanza-Maximización que puede describirse así:

1. Conseguir un pronóstico inicial de  $w_0$  mediante, por ejemplo,  $w_0 = A\Lambda A^T \hat{x}$  y establecer  $W_0$  apropiadamente.

35

2. Repetir desde  $k = 0$  hasta converger:



- (a) Calcular  $\hat{s}_k = W_k^{-1} \hat{x}$ .
- (b) Calcular  $w_{k+1} = f(\hat{s}_k)$  y establecer  $W_{k+1}$  apropiadamente.
- (c) Aumentar k.

Los últimos valores  $\hat{s}_k$  y  $W_k$  son las estimadas finales de  $\hat{s}$  y  $\hat{W}$ .

5

Normalmente, se requieren menos de 20 iteraciones para convergencia

Diferentes realizaciones del aparato de decodificación 2 también se muestran mediante la Figura 5 y la Figura 6. El aparato de decodificación 2 de la Figura 5 se muestra sin la deponderación. En la Figura 6 se proporciona la aproximación del modelo ponderado perceptual  $W$  y usa para la deponderación del espectro decuantificado  $\hat{x}$ . Esto destaca que la decodificación incluye la reconstrucción del modelo  $W$  basado en la señal codificada.

Para evaluar el rendimiento de cada parte del sistema propuesto, se realizaron los siguientes experimentos.

Se compararon tres versiones del audio de entrada:

la señal cuantificada y reconstruida, 1) sin y 2) con modelización perceptual de tal forma que la perceptual se conoce en el decodificador,

así como 3) la señal perceptivamente cuantificada en la que la reconstrucción se realiza con el modelo perceptual de forma ciega estimado de acuerdo con la invención.

Como material de pruebas se usaron muestras de voz aleatorias del conjunto de datos NTT-AT [21] (véase la Figura 10 fila superior). Las señales de entrada se remuestrearon 12,8 kHz, la STFT se implementó con Transformada de Coseno Discreta para obtener un espectro con valor real y se usó un modelo de envolvente con 20 bandas, distribuidas de acuerdo con la escala Mel [20, 5].

30

Como una primera aproximación del modelo perceptual, se usó la reducción de intervalo función de  $f(y)=y^p$  con  $p = 0,5$ . Este modelo perceptual se eligió meramente como una forma para demostrar el rendimiento de la reconstrucción ciega, y no debería considerarse como un producto final afinado. El rendimiento del modelo de envolvente así como el modelo perceptual ya se ilustraron en la Figura 4.

35

Primero, la SNR perceptual para la cuantificación propuesta sin ( $SNR_O$ ) y con la reconstrucción ciega ( $SNR_B$ ) del modelo perceptual, respectivamente, se estimará mediante:

$$SNR_O = \frac{\|x\|}{\|x-\hat{x}\|} \text{ y } SNR_B = \frac{\|x\|}{\|x-W\hat{W}^{-1}\hat{x}\|} \quad (2)$$

5 La Figura 10 ilustra la SNR perceptual para un archivo de voz cuantificado con los diferentes procedimientos ( $K = 3000$ ).

10 Está claro que cuando el modelo perceptual se conoce (enfoque de oráculo), la SNR está cerca de 8,4 dB. La reconstrucción ciega del modelo perceptual disminuye claramente la calidad (Ciega) especialmente para fonemas de voz. Sin embargo, la SNR del sistema sin un modelo perceptual (No perc) es más de dos veces peor que con recuperación ciega.

15 Para cuantificar adicionalmente la ventaja de la reconstrucción ciega en lugar de sin modelización perceptual, se midió la SNR media con diferentes tasas de bits  $K$  (véase la Figura 11).

20 Los enfoques de recuperación ciega y modelo no perceptual son de media 1,1 dB y 5,8 dB peores que el enfoque de oráculo. Claramente SNR mejora con tasa de bits, aunque el caso de modelo no perceptual mejora más despacio que con un modelo perceptual. Además, con SNR en aumento, la recuperación ciega enfoca la calidad del enfoque de oráculo asintóticamente.

25 Finalmente, para evaluar la calidad subjetiva, se realizó una prueba de audición de MUSHRA con ocho oyentes y seis artículos elegidos aleatoriamente del conjunto de datos NTT-AT. La señal se cuantificó con 3 kbits/trama. Esto es un número relativamente bajo dado que no se usó ninguna modelización de fuente para los experimentos, con lo que la SNR de salida también es relativamente baja. Este escenario se eligió para demostrar una condición problemática y rendimiento se espera que mejore significativamente a mayores tasas de bits así como cuando se aplique un modelo de fuente.

30 A partir de las puntuaciones MUSHRA diferenciales en la Figura 12, puede observarse que para todos los artículos, la modelización perceptual mejora la calidad tanto con la estimación de oráculo como ciega, por 29,9 y 22,3 puntos de media, respectivamente. La importancia estadística de las diferencias se confirmó con la prueba t de Student en  $p > 99 \%$ .

35

La cuantificación de 1 bit y esquema de codificación - ya sea combinados en una realización o en realizaciones separadas - tienen varias consecuencias y propiedades interesantes.

5 En primer lugar, para el análisis de propiedades de cuantificación, obsérvese que cada columna de  $P$  es una proyección a un subespacio de 1 dimensión del espacio de  $N$  dimensiones del vector  $x$ .

10 Codificando el signo de una proyección, el espacio de  $N$  dimensiones se divide en dos partes. Codificando repetidamente signos de  $Px$ , el espacio de  $N$  dimensiones se divide en células de cuantificación incluso más pequeñas. Ya que  $P$  es una matriz aleatoria, sus columnas son aproximadamente ortogonales entre sí, con lo que las células de cuantificación permanecen casi óptimas.

15 En un sistema de un solo nodo, puede diseñarse un enfoque de cuantificación que es más eficiente. Sin embargo, en un sistema distribuido se complica más - se necesita un procedimiento simple para evitar que los nodos codifiquen la misma información, es decir, tiene que evitarse la sobre codificación mientras se retiene una complejidad algorítmica baja.

20 La cuantificación inventiva es muy simple y proporciona rendimiento casi óptimo.

En segundo lugar, no se emplearon procedimientos de codificación de fuente.

25 Sin embargo es bien conocido que tal modelización puede usarse para mejorar la eficiencia de codificación significativamente. Modelización de fuente puede aplicarse en el lado de decodificador modelizando la distribución de probabilidad de señales de voz y audio (por ejemplo [22]). La modelización de fuente es posible, ya que la señal cuantificada puede tratarse como una observación ruidosa de la "verdadera" señal, con lo que, aplicando una distribución previa de la fuente, puede aplicarse máxima probabilidad de optimización (o similar) para aproximarse a la "verdadera" señal. Ya que esta optimización se aplica en la  
30 red o en el decodificador, la carga de cálculo se mantiene alejada de los nodos de sensores y los nodos de sensores pueden permanecer con baja potencia.

35 En tercer lugar, desde una perspectiva de privacidad, el procedimiento de proyección aleatoria puede diseñarse para ser un cifrado altamente eficiente.

Si una escucha no autorizada no conoce la semilla para la matriz aleatoria, entonces los

datos parecerán en su totalidad aleatorios y sin sentido. Suponiendo que la semilla aleatoria se comunica de una manera segura, entonces únicamente el codificador y el receptor previsto pueden descifrar un mensaje. Este enfoque es al contrario que enfoques tal como los dados por [12, 13], en los que la comunicación entre nodos se emplea intencionalmente.

5 Mientras tal colaboración entre nodos puede usarse para mejorar la SNR perceptual, la privacidad es más difícil de garantizar. Incluso cuando se supone que los nodos de sensores están operando en una red segura, puede ser suficiente un único nodo comprometido para ganar acceso a todas las comunicaciones. En el enfoque propuesto, en contraste, si una escucha no autorizada gana acceso a un nodo de sensor, compromete los datos de ese  
10 nodo únicamente, ya que los nodos pueden y deberían usar diferentes semillas. Para limitar la potencia de transmisión de nodos de sensores, puede permitirse sin embargo que los nodos retransmitan paquetes, ya que los paquetes permanecen legibles únicamente por el receptor previsto y por lo tanto no se compromete la privacidad.

15 El procedimiento inventivo se basa en una realización en una idea de cuantificación de 1 bit, en la que en el lado de codificador, la señal de entrada perceptualmente ponderada se proyecta a subespacios aleatorios y en la que el signo de cada dimensión se transmite a continuación. El decodificador puede invertir la cuantificación con una pseudo-inversa o similar, para obtener la señal ponderada perceptualmente cuantificada.

20 La parte principal del procedimiento propuesto es entonces una reconstrucción de una estimada de la señal original, cuando se tiene acceso únicamente a la señal ponderada perceptualmente. El enfoque se basa en un algoritmo de estimación-maximización (EM), en el que se alterna iterativamente entre estimación del modelo perceptual y la señal original.

25 El algoritmo de codificación de voz y audio distribuida propuesto es por lo tanto un enfoque viable para aplicaciones para la Internet de las Cosas. Ofrece rendimiento escalable para cualquier número de nodos de sensores y nivel de consumo de potencia. Además, el algoritmo es seguro por diseño, ya que la privacidad del canal de comunicación puede  
30 garantizarse mediante comunicación cifrada de la semilla aleatoria.

La invención presentada incluye al menos las siguientes características, aspectos, así como aplicaciones obvias y extensiones. El listado, por lo tanto, se refiere a realización diferente y permite diferentes combinaciones de las características mencionadas.

35 1. Sistema de codificación de voz y audio distribuida con uno o varios codificadores, que

pueden implementarse en procesadores escalables, flexible, de baja potencia y bajo coste con transmisión limitada y capacidad de cálculo y otros recursos.

- 5 1.1. Los codificadores distribuidos pueden configurarse para codificar información de tal forma que la información de cada nodo de codificador puede cuantificar independientemente de tal forma que la cantidad de información de un conjunto de nodos puede maximizarse, por ejemplo,
- 1.1.1. haciendo que cada nodo de codificador codifique información que es ortogonal o aproximadamente ortogonal a los otros nodos,
  - 10 1.1.2. usando algoritmos de cuantificación aleatorizados, que pueden ser o no ortogonales o aproximadamente ortogonales a los otros nodos,
  - 1.1.3. usando procedimiento de codificación por entropía para comprimir información cuantificada.
- 15 1.2. Los codificadores distribuidos pueden configurarse para codificar información en un dominio perceptual, que se aproxima a la precisión perceptual del sistema auditivo humano
- 20 1.2.1. en el que la transformación al dominio perceptual puede diseñarse para ser reversible, de tal forma que nodos de codificador pueden recuperar (de forma ciega) el modelo perceptual sin transmisión explícita del modelo perceptual,
  - 1.2.2. en el que alguno o todos los parámetros de los modelos perceptuales pueden transmitirse explícitamente desde algunos o todos los nodos de codificador como información secundaria.
  - 25 1.2.3. en el que los modelos perceptuales pueden describirse usando modelos de envolvente, tal como modelos predictivos lineales, bancos de filtros, transformadas de matriz o bandas de factor de escala (constantes en piezas o interpoladas), y pueden implementarse como operaciones de matriz u operaciones de filtrado.
  - 30 1.2.4. en el que los modelos perceptuales pueden tener partes constantes y adaptativas, donde las partes constantes siempre son la misma y partes adaptativas dependen de la señal de entrada y/o la configuración del sistema y/o la cantidad de recursos (hardware, cálculo, transmisión, capacidad de batería, etc.) disponibles.
  - 35
- 1.3. Los codificadores distribuidos pueden configurarse para codificar información de

tal forma que el error de cuantificación se minimiza en un dominio perceptual, por ejemplo, como el bucle de análisis por síntesis en códecs de tipo CELP.

5 1.4. Los codificadores distribuidos pueden configurarse para codificar información usando cifrado para evitar escucha no autorizada mediante por ejemplo

1.4.1. el cifrado de la descripción de señal final o intermedia antes de la transmisión usando procedimientos que son algorítmicos o basándose en tablas de consulta,

10 1.4.2. el uso de cuantificación con aleatorización, tal como transformadas o proyecciones aleatorias, en las que la aleatorización se consigue mediante un algoritmo o tabla de consulta con un valor de semilla conocido (punto de inicio).

15 1.5. Los codificadores distribuidos cuya configuración puede ser flexible y modificarse en línea, por ejemplo,

20 1.5.1. basado en ubicación y movimiento de fuentes de audio (tal como oradores humanos), nodos de sensores, así como disponibilidad de recursos, nodos de sensores pueden decidir independiente o conjuntamente qué nodos están activos o inactivos,

25 1.5.2. basado en ubicación y movimiento de fuentes de audio (tal como oradores humanos), nodos de sensores, así como disponibilidad de recursos, nodos de sensores pueden decidir independiente o conjuntamente ajustar la asignación de recursos, por ejemplo, de tal forma que un nodo de sensor con un micrófono cerca de un orador usa más recursos que los más alejados.

30 1.6. Un caso especial del sistema propuesto es cuando existe únicamente un único codificador, con lo que el sistema se reduce a seguir la arquitectura clásica de códecs de voz y audio. Importante, sin embargo, la cuantificación aleatoria y recuperación ciega de los aspectos de modelo perceptual de la invención actual pueden usarse también en el diseño clásico de codificador y decodificador, por ejemplo, para el beneficio de eficiencia de codificación, para obtener un codificador de baja complejidad, y para el cifrado de la comunicación.

35 2. Sistema de codificación de voz y audio distribuida en el que la entrada desde los codificadores se fusiona en uno o varios nodos de red (usando procedimientos de

procesamiento en red) o en uno o varios nodos de decodificador/receptor.

2.1. Decodificador o unidades de procesamiento pueden configurarse para invertir la cuantificación mediante por ejemplo

5 2.1.1. transformada inversa exacta, pseudo-inversa o una inversa aproximada tal como la traspuesta de una matriz aleatoria,

2.1.2. la estimación de la señal de entrada mediante procedimientos de optimización, tal como detección compresiva,

2.1.3. secuencias de bits  $u_k$  de nodos  $k \in [1, 2, \dots, n]$  pueden fusionarse mediante

10 una inversión conjunta tal como  $\hat{x} = \begin{bmatrix} P_1 \\ P_2 \\ \dots \\ P_n \end{bmatrix}^\dagger \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{bmatrix}$  o las secuencias de bits pueden

invertirse primero  $\hat{x}_k = P_k^\dagger u_k$  y únicamente después de fusionarse. El beneficio de la última es que se puede aplicar a continuación la modelización, tal como modelos perceptual, fuente, especial y sensor separada o conjuntamente en cada nodo.

15

2.2. Decodificador o unidades de procesamiento pueden configurarse para invertir el modelo perceptual usando información transmitida de lado y/o estimando la misma de forma ciega a partir de la señal (perceptual) reconstruida mediante por ejemplo

2.2.1. procedimientos directos, en los que el modelo perceptual se obtiene directamente a partir de la señal cuantificada,

20

2.2.2. procedimientos iterativos, en los que la estimada del modelo perceptual se mejora sucesivamente iteración a iteración.

2.3. Decodificador o unidades de procesamiento pueden configurarse para usar modelos de fuente y de nodo de sensor para mejorar la calidad de la señal decodificada.

25

2.3.1. Modelizando propiedades estadísticas de señales de fuente y nodos de sensores, se puede optimizar la probabilidad de observación (la señal cuantificada) dada la distribución previa de la señal y/o fuente.

30

2.3.1.1. Las características de fuente modelizadas pueden incluir uno o varios modelos de envolvente espectral (modelos predictivos lineales, cuantificación de distribución, factores de escala, etc.), modelos armónicos

(filtros de peine, predictores de larga duración, etc.), envolventes temporales (modelos predictivos lineales, cuantificación de distribución, factores de escala, etc.), envolventes espaciales y modelos de movimiento espaciales.

5 2.3.1.2. Las características pueden modelarse con tablas de consulta tal como libros de códigos o cuantificación de vectores, o con modelos paramétricos tal como físico (producción de voz, etc.), perceptual (modelos de enmascaramiento, bandas críticas etc.) o modelos estadísticos (distribuciones de probabilidad de una o múltiples variantes).

10 2.3.1.3. Las características pueden describir características acústicas y espaciales de la fuentes, pero también características de los nodos de sensores, tal como señal calidad (por ejemplo SNR alta o baja) y características espaciales (por ejemplo estacionaria, móvil, usable, etc.).

15 2.3.2. Determinando perceptualmente distorsiones molestas de la señal de salida, se puede modificar la señal de salida mediante, por ejemplo, procedimientos post-filtrado.

20 2.4. Decodificador o unidades de procesamiento pueden configurarse para usar procedimientos de mejora de señal tal como atenuación de ruido, transformación de haz y procedimientos de separación de fuente para mejorar la señal calidad.

25 2.4.1. Procedimientos de múltiples micrófonos tal como transformación de haz pueden emplearse para tomar ventaja de la configuración espacial de sensores para extraer fuentes de sonido espacialmente separadas y otra información acerca de la acústica de sala. Importante, los procedimientos de múltiples micrófonos pueden incluir habitualmente estimación de retardo y/o estimación de correlación cruzada.

2.4.2. Procedimientos de atenuación de ruido pueden usarse para suprimir fuentes no deseadas tal como ruido de fondo.

30 2.4.3. Procedimientos de separación de fuente pueden usarse para distinguir entre varias fuentes de sonido.

35 3. Sistema de codificación de voz y audio distribuida en el que la información de la señal información puede transmitirse, retransmitirse y procesarse mediante una selección de dispositivos/nodos en el sistema.



3.1. Nodos de sensores (dispositivos con sensores) reciben la señal en bruto y pueden configurarse para o bien codificar (cuantificar y codificar) la señal o transmitir la misma en un formato en bruto.

5

3.2. Cualquier nodo (habitualmente el nodo de sensor) que tiene acceso a una señal en bruto puede codificar la señal y transmitir la misma.

3.3. Cualquier nodo puede configurarse para retransmitir de otros nodos.

3.4. Cualquier nodo (habitualmente un receptor nodo) que tiene acceso a una señal codificada (y la información de codificación-decodificación opcional), puede configurarse para decodificar la señal.

10

3.5. Un nodo intermedio, tal como dispositivo servidor en el lado de decodificador, pueden configurarse para fusionar información de flujos disponibles en uno o varios flujos. El flujo flujos combinados pueden representar por ejemplo la representación acústica original (tal como una interpretación musical) o fuentes de sonido individuales (tal como un orador individual en una teleconferencia). El flujo combinado puede reproducirse adicionalmente mediante altavoces, almacenarse, transmitirse como tal o codificado por la misma u otra herramienta de codificación de voz y audio.

15

3.6. La configuración de red puede configurarse para ser estática o dinámica, de tal forma que optimiza por ejemplo, uno o varios de los siguientes criterios; calidad de sonido, asignación de recursos, seguridad/privacidad.

20

#### 4. Las aplicaciones previstas incluyen al menos

4.1. Aplicaciones telefónicas, en las que se usa una selección de dispositivos de soporte para captar la señal o señales de voz y audio deseadas.

25

4.1.1. Aplicaciones domésticas y empresariales, en las que se usa una selección de dispositivos de soporte para captar la señal o señales de voz y audio deseadas y transmitirse a una o varias ubicaciones remotas.

4.1.2. Aplicaciones de teleconferencias, en las que se usa una selección de dispositivos de soporte para captar la señal o señales de voz y audio deseadas y transmitirse a una o varias ubicaciones remotas.

30

4.1.3. Sistemas telefónicos en coches, en los que se usan micrófonos fijos en el coche y/o dispositivos de soporte dentro del coche para captar la señal o señales de voz y audio deseadas y transmitirse a una o varias ubicaciones remotas.

35

4.2. Aplicaciones de juegos y realidad virtual/aumentada, en las que la escena de sonido de un jugador se transmite a otros jugadores o un servidor.

4.3. Aplicaciones de producción de eventos de concierto, espectáculo, escenario, ópera, representación, deportes y otros eventos, en las que sonidos de los interpretes, músicos, audiencia o escenario de sonido general se graba o transmite con un sistema de codificación de voz y audio distribuida.

5           4.3.1. La aplicación puede diseñarse para que sea de retardo bajo o ultra bajo para habilitar la interacción y/o reproducción y amplificación simultáneas.

4.3.2. La aplicación puede diseñarse para permitir interacción entre interpretes, dentro de la audiencia, o entre todos los participantes.

10         4.4 Aplicaciones de seguridad y protección, en las que se supervisa una escena de sonido para detectar eventos peligrosos (por ejemplo pánico en un estadio), accidentes domésticos (por ejemplo la caída de una persona mayor) etc.

4.5 Aplicaciones multimedia, en las que señales de voz y audio se combinan con video y/o otros medios.

15          Aunque algunos aspectos se han descrito en el contexto de un aparato, está claro que estos aspectos también representan una descripción del correspondiente procedimiento, en el que un bloque o dispositivo corresponde a una etapa de procedimiento o una característica de una etapa de procedimiento. Análogamente, aspectos descritos en el contexto de una etapa de procedimiento también representan una descripción de un correspondiente bloque o artículo o característica de un correspondiente aparato. Alguna o todas las etapas de procedimiento pueden ejecutarse mediante (o usando) un aparato de hardware, como por ejemplo, un microprocesador, un ordenador programable o un circuito electrónico. En algunas realizaciones, alguna o más de las más importantes etapas de procedimiento pueden ejecutarse mediante un aparato de este tipo.

25          La señal inventiva transmitida o codificada puede almacenarse en un medio de almacenamiento digital o puede transmitirse en un medio de transmisión tal como un medio de transmisión inalámbrica o un medio de transmisión cableado tal como internet.

30          Dependiendo de ciertos requisitos de implementación, realizaciones de la invención pueden implementarse en hardware o en software. La implementación puede realizarse usando un medio de almacenamiento digital, por ejemplo un disco flexible, un DVD, un Blu-Ray, un CD, una ROM, una PROM y EPROM, una EEPROM o una Memoria flash, que tiene señales de control legibles electrónicamente almacenadas en los mismos, que cooperan (o son capaces de cooperar) con un sistema informático programable de tal forma que se realiza el respectivo procedimiento. Por lo tanto, el medio de almacenamiento digital puede ser legible

por ordenador.

5 Algunas realizaciones de acuerdo con la invención comprenden una portadora de datos que tiene señales de control legibles electrónicamente, que con capaces de cooperar con un sistema informático programable, de tal forma que uno de los procedimientos descritos en el presente documento se realiza.

10 En general, realizaciones de la presente invención pueden implementarse como un producto de programa informático con un código de programa, estando el código de programa operativo para la realización de uno de los procedimientos cuando el producto de programa informático se ejecuta en un ordenador. El código de programa puede almacenarse, por ejemplo, en un soporte legible por máquina.

15 Otras realizaciones comprenden el programa informático para realizar uno de los procedimientos descritos en el presente documento, almacenado en un soporte legible por máquina.

20 En otras palabras, una realización del procedimiento inventivo es, por lo tanto, un programa informático que tiene un código de programa para realizar uno de los procedimientos descritos en el presente documento, cuando el programa informático se ejecuta en un ordenador.

25 Una realización adicional del procedimiento inventivo es, por lo tanto, una portadora de datos (o un medio de almacenamiento no transitorio tal como un medio de almacenamiento digital, o un medio legible por ordenador) que comprende, grabada en la misma, el programa informático para realizar uno de los procedimientos descritos en el presente documento. La portadora de datos, el medio de almacenamiento digital o el medio grabado son habitualmente tangibles y/o no transitorios.

30 Una realización adicional de la invención procedimiento es, por lo tanto, un flujo de datos o una secuencia de señales que representan el programa informático para realizar uno de los procedimientos descritos en el presente documento. El flujo de datos o la secuencia de señales puede, por ejemplo, configurarse para transferirse a través de una conexión de comunicación de datos, por ejemplo, a través de la Internet.

35 Una realización adicional comprende un medio de procesamiento, por ejemplo, un

ordenador o un dispositivo de lógica programable, configurado para, o adaptado para, realizar uno de los procedimientos descritos en el presente documento.

5 Una realización adicional comprende un ordenador que tiene instalado en el mismo el programa informático para realizar uno de los procedimientos descritos en el presente documento.

10 Una realización adicional de acuerdo con la invención comprende un aparato o un sistema configurado para transferir (por ejemplo, electrónicamente u ópticamente) un programa informático para realizar uno de los procedimientos descritos en el presente documento a un receptor. El receptor puede ser, por ejemplo, un ordenador, un dispositivo móvil, un dispositivo de memoria o similar. El aparato o sistema puede comprender, por ejemplo, un servidor de archivos para transferir el programa informático al receptor.

15 En algunas realizaciones, un dispositivo de lógica programable (por ejemplo, un Campo de Matriz de Puertas Programables) puede usarse para realizar alguna o todas las funcionalidades de los procedimientos descritos en el presente documento. En algunas realizaciones, un campo de matriz de puertas programables puede cooperar con un microprocesador para realizar uno de los procedimientos descritos en el presente  
20 documento. En general, los procedimientos se realizan preferentemente mediante cualquier aparato de hardware.

Las realizaciones descritas anteriormente son meramente ilustrativas de los principios de la presente invención. Se entiende que modificaciones y variaciones de las disposiciones y los  
25 detalles descritos en el presente documento serán evidentes para los expertos en la técnica. Es la intención, por lo tanto, limitarse únicamente por el alcance de las reivindicaciones de patente inminentes y no por los detalles específicos presentados por medio de descripción y explicación de las realizaciones en el presente documento.

## 30 **REFERENCIAS**

[1] TS 26.445, EVS Codec Detailed Algorithmic Description; 3GPP Technical Specification (Release 12), 3GPP, 2014.

[2] TS 26.190, Adaptive Multi-Rate (AMR-WB) speech codec, 3GPP, 2007.

- [3] ISO/IEC 23003-3:2012, "MPEG-D (MPEG audio technologies), part 3: Unified speech and audio coding," 2012.
- [4] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, and M. Dietz, "ISO/IEC MPEG-2 advanced audio coding," *Journal of the Audio engineering society*, vol. 45, n.º 10, pp. 789-814, 1997.
- 5 [5] M. Bosi and R. E. Goldberg, *Introduction to Digital Audio Coding and Standards*. Dordrecht, Países Bajos: Kluwer Academic Publishers, 2003.
- [6] P. T. Boufounos and R. G. Baraniuk, "1-bit compressive sensing," in *Information Sciences and Systems, 2008. CISS 2008. 42<sup>a</sup> Annual Conference on*. IEEE, 2008, pp. 16-21.
- 10 [7] Z. Xiong, A. D. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Process. Mag.*, vol. 21, n.º 5, pp. 80-94, 2004.
- [8] Z. Xiong, A. D. Liveris, and Y. Yang, "Distributed source coding," *Handbook on Array Processing and Sensor Networks*, pp. 609- 643, 2009.
- [9] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. IEEE*, vol. 93, n.º 1, pp. 71-83, 2005.
- 15 [10] A. Majumdar, K. Ramchandran, and L. Kozintsev, "Distributed coding for wireless audio sensors," in *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on*. IEEE, 2003, pp. 209-212.
- [11] H. Dong, J. Lu, and Y. Sun, "Distributed audio coding in wireless sensor networks," in *Computational Intelligence and Security, 2006 International Conference on*, vol. 2. IEEE, 2006, pp. 1695-1699.
- 20 [12] A. Zahedi, J. Østergaard, S. H. Jensen, P. Naylor, and S. Bech, "Coding and enhancement in wireless acoustic sensor networks," in *Data Compression Conference (DCC), 2015*. IEEE, 2015, pp. 293-302.
- [13] A. Zahedi, J. Østergaard, S. H. Jensen, S. Bech, and P. Naylor, "Audio coding in wireless acoustic sensor networks," *Señal Processing*, vol. 107, pp. 141-152, 2015.
- 25 [14] El documento US 7.835.904.

- [15] G. Kubin and W. B. Kleijn, "Multiple-description coding (MDC) of speech with an invertible auditory model," in *Speech Coding*, IEEE Workshop on, 1999, pp. 81-83.
- [16] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Process. Mag.*, vol. 18, n.º 5, pp. 74- 93, 2001.
- 5 [17] J. O. Smith III and J. S. Abel, "Bark and ERB bilinear transforms," *IEEE Trans. Speech Audio Process.*, vol. 7, n.º 6, pp. 697-708, 1999.
- [18] T. Bäckström, "Vandermonde factorization of Toeplitz matrices and applications in filtering and warping," *IEEE Trans. Signal Process.*, vol. 61, n.º 24, pp. 6257-6263, Dic. 2013.
- 10 [19] F. Zheng, G. Zhang, and Z. Song, "Comparison of different implementations of MFCC," *Journal of Computer Science and Technology*, vol. 16, n.º 6, pp. 582-589, 2001.
- [20] H. Fastl and E. Zwicker, *Psychoacoustics: Facts and models*. Springer, 2006, vol. 22.
- [21] NTT-AT, "Super wideband stereo speech database," <http://www.ntt-at.com/product/widebandspeech>, acceso: 09.09.2014. [En línea]. Disponible:  
15 <http://www.ntt-at.com/product/widebandspeech>
- [22] S. Korse, T. Jähnel, and T. Bäckström, "Entropy coding of spectral envelopes for speech and audio coding using distribution quantization," in *Proc. Interspeech*, 2016.

**REIVINDICACIONES**

1. Aparato de codificación (1) para el procesamiento de una señal de entrada (3),  
que comprende un ponderador perceptual (10) y un cuantificador (14),  
5 en el que el ponderador perceptual (10) comprende un proveedor de modelo (12) y un aplicador de modelo (13),  
en el que el proveedor de modelo (12) está configurado para proporcionar un modelo ponderado perceptual (W) basado en la señal de entrada (3),  
en el que el aplicador de modelo (13) está configurado para proporcionar un espectro  
10 ponderado perceptualmente (x) aplicando el modelo ponderado perceptual (W) a un espectro (s) basado en la señal de entrada (3),  
en el que el cuantificador (14) está configurado para cuantificar el espectro ponderado perceptualmente (x) y para proporcionar una secuencia de bits ( $\hat{u}$ ),  
en el que el cuantificador (14) comprende un aplicador de matriz aleatoria (16) y un  
15 calculador de función de signo (17),  
en el que el aplicador de matriz aleatoria (16) está configurado para aplicar una matriz aleatoria (P) al espectro ponderado perceptualmente (x) para proporcionar un espectro transformado (u), y  
en el que un calculador de función de signo (17) está configurado para calcular una  
20 función de signo de componentes del espectro transformado (u) para proporcionar la secuencia de bits ( $\hat{u}$ ).
2. Aparato de codificación (1) de la reivindicación 1,  
en el que el proveedor de modelo (12) está configurado para proporcionar el modelo  
25 ponderado perceptual (W) basado en una compresión del espectro (s) basado en la señal de entrada (3).
3. Aparato de codificación (1) de las reivindicaciones 1 o 2,  
en el que el ponderador perceptual (10) comprende un calculador de envolvente (11),  
30 en el que el calculador de envolvente (11) está configurado para proporcionar una envolvente (y) de un espectro de magnitud ( $|x|$ ) basado en la señal de entrada (3), y  
en el que el proveedor de modelo (12) está configurado para proporcionar el modelo ponderado perceptual (W) basado en una compresión de la envolvente (y).
- 35 4. Aparato de codificación (1) de cualquiera de las reivindicaciones 1 a 3,  
en el que el proveedor de modelo (12) está configurado para calcular una función de

compresión  $f()$  que describe una compresión del espectro  $(s)$  basada en la señal de entrada (3) o de una envolvente  $(y)$  de un espectro de magnitud  $(|x|)$  basada en la señal de entrada (3), y

5 en el que el proveedor de modelo (12) está configurado para calcular el modelo ponderado perceptual  $(W)$  basado en la función de compresión  $(f())$ .

5. Aparato de codificación (1) de la reivindicación 4,

en el que la función de compresión  $(f())$  está aumentando estrictamente,

10 donde para cualquier valor positivo  $(t)$  y un valor arbitrariamente pequeño  $(\epsilon)$  el valor de función  $(f(t))$  para el valor escalar positivo  $(t)$  es más pequeño que el valor de función  $(f(t+\epsilon))$  para la suma  $(t+\epsilon)$  del valor escalar positivo  $(t)$  y del valor arbitrariamente pequeño  $(\epsilon)$ , y

15 en el que para un primer valor escalar positivo  $(t_1)$  y un segundo valor escalar positivo  $(t_2)$  mayor que el primer valor escalar positivo  $(t_1)$ , una diferencia entre el valor de función  $(f(t_2))$  para el segundo valor escalar positivo  $(t_2)$  y el valor de función  $(f(t_1))$  para el primer valor escalar positivo  $(t_1)$  es más pequeño que una diferencia entre el segundo valor escalar positivo  $(t_2)$  y el primer valor escalar positivo  $(t_1)$ .

6. Aparato de decodificación (2) para el procesamiento de una señal codificada (4),

20 que comprende un decuantificador (20) y un deponderador perceptual (21),

en el que el decuantificador (20) está configurado para decuantificar una secuencia de bits  $(\hat{u})$  comprendida por la señal codificada (4) y para proporcionar un espectro ponderado calculado perceptualmente  $(\hat{x})$ ,

25 en el que el decuantificador (20) está configurado para decuantificar la secuencia de bits  $(\hat{u})$  aplicando una pseudo-inversa  $(P^\dagger)$  de una matriz aleatoria  $(P)$  a la secuencia de bits  $(\hat{u})$ ,

en el que el deponderador perceptual (21) comprende un aproximador de espectro (24) y un aproximador de modelo (25),

30 en el que el aproximador de espectro (24) está configurado para calcular una aproximación de un espectro  $(\hat{s}_k)$  basada en el espectro ponderado calculado perceptualmente  $(\hat{x})$ , y

en el que el aproximador de modelo (25) está configurado para calcular una aproximación  $(W_k)$  de un modelo ponderado perceptual  $(W)$  con el que la señal codificada (4) está asociada basándose en la aproximación del espectro  $(\hat{s}_k)$ .

35



7. Aparato de decodificación (2) de la reivindicación 6,  
 en el que el deponderador perceptual (21) comprende un proveedor de pronóstico  
 inicial (23),  
 en el que el proveedor de pronóstico inicial (23) está configurado para proporcionar  
 5 datos ( $w_0$ ) para un pronóstico inicial del modelo ponderado perceptual ( $W$ ), y  
 en el que el aproximador de espectro (24) está configurado para calcular la  
 aproximación del espectro ( $\hat{s}_k$ ) basada en la señal codificada (4) y el pronóstico inicial ( $W_0$ )  
 o la aproximación ( $W_k$ ) del modelo ponderado perceptual ( $W$ ).
- 10 8. Aparato de decodificación (2) de las reivindicaciones 6 o 7,  
 en el que el decuantificador (20) está configurado para decuantificar la secuencia de  
 bits ( $\hat{u}$ ) basándose en una información secundaria acerca de la matriz aleatoria ( $P$ )  
 comprendida por la señal codificada (3).
- 15 9. Aparato de decodificación (2) de cualquiera de las reivindicaciones 6 a 8,  
 en el que el decuantificador (20) está configurado para recibir una pluralidad de flujos  
 de bits ( $\hat{u}_k$ ) y para proporcionar un espectro ponderado calculado perceptualmente ( $\hat{x}$ )  
 basado en la pluralidad de flujos de bits ( $\hat{u}_k$ ).
- 20 10. Aparato de decodificación (2) de cualquiera de las reivindicaciones 6 a 9,  
 en donde el aparato de decodificación (2) está configurado para aplicar modelización  
 de fuente.
11. Procedimiento para el procesamiento de una señal de entrada (3),  
 25 que comprende:  
 proporcionar un modelo de ponderación perceptual ( $W$ ) basado en la señal de entrada  
 (3),  
 ponderar un espectro ( $s$ ) de la señal de entrada (3) aplicando el modelo de  
 ponderación perceptual ( $W$ ) al espectro ( $s$ ) de la señal de entrada (3), y  
 30 cuantificar el espectro ponderado ( $x$ ) calculando una función de signo de proyecciones  
 aleatorias del espectro ponderado ( $x$ ).
12. Procedimiento de la reivindicación 11,  
 comprendiendo además:  
 35 obtener las proyecciones aleatorias del espectro ponderado ( $x$ ) aplicando una matriz  
 aleatoria ( $P$ ) al espectro ponderado ( $x$ ).

13. Procedimiento de las reivindicaciones 11 o 12,  
 en el que proporcionar el modelo de ponderación perceptual ( $W$ ) comprende comprimir  
 una envolvente de un espectro de magnitud ( $|x|$ ) de la señal de entrada (3).
- 5
14. Procedimiento de la reivindicación 13,  
 que comprende además:  
 obtener la envolvente usando un banco de filtros ( $A$ ) y una matriz diagonal ( $\Lambda$ ) que  
 contiene factores de normalización para cada banda.
- 10
15. Procedimiento para el procesamiento de una señal codificada (4),  
 que comprende:  
 proporcionar una señal perceptual cuantificada ( $\hat{x}$ ) aplicando una pseudo-inversa ( $P^\dagger$ )  
 de una matriz aleatoria ( $P$ ) a la señal codificada (4),  
 15 calcular una estimación de un espectro ( $\hat{s}_k$ ) basado en la señal perceptual cuantificada  
 ( $\hat{x}$ ), y  
 calcular una aproximación ( $W_k$ ) de un modelo de ponderación perceptual ( $W$ ) usado  
 para proporcionar la señal codificada (4) basándose en la aproximación del espectro ( $\hat{s}_k$ ).
- 20
16. Procedimiento de la reivindicación 15,  
 que comprende además:  
 proporcionar una cero-ésima aproximación ( $W_0$ ) del modelo de ponderación perceptual  
 ( $W$ ) usando un pronóstico inicial ( $w_0$ ), y  
 calcular una cero-ésima estimación del espectro ( $\hat{s}_k$ ) basándose en la cero-ésima  
 25 aproximación ( $W_0$ ) del modelo de ponderación perceptual ( $W$ ).
- 30
17. Procedimiento de la reivindicación 16,  
 que comprende además:  
 obtener el pronóstico inicial ( $w_0$ ) usando un banco de filtros ( $A$ ) y una matriz diagonal  
 ( $\Lambda$ ) que contiene factores de normalización para cada banda.
18. Programa informático para la realización, cuando se ejecuta en un ordenador o un  
 procesador, del procedimiento de cualquiera de las reivindicaciones 11 a 17.

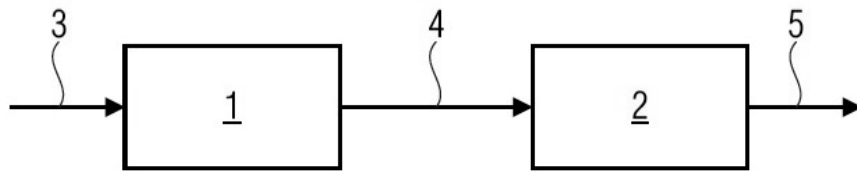


Fig. 1

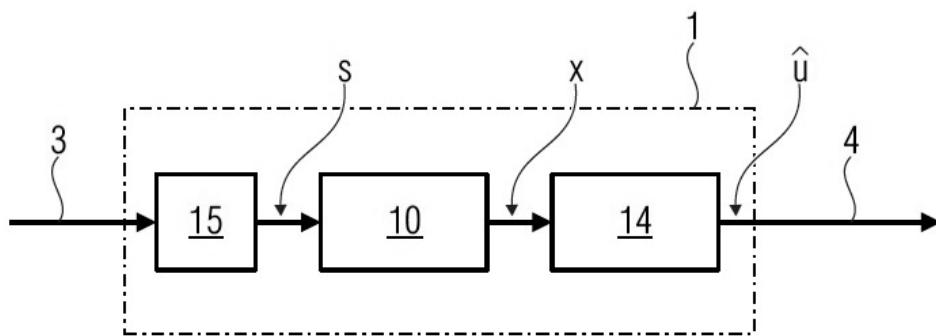


Fig. 2

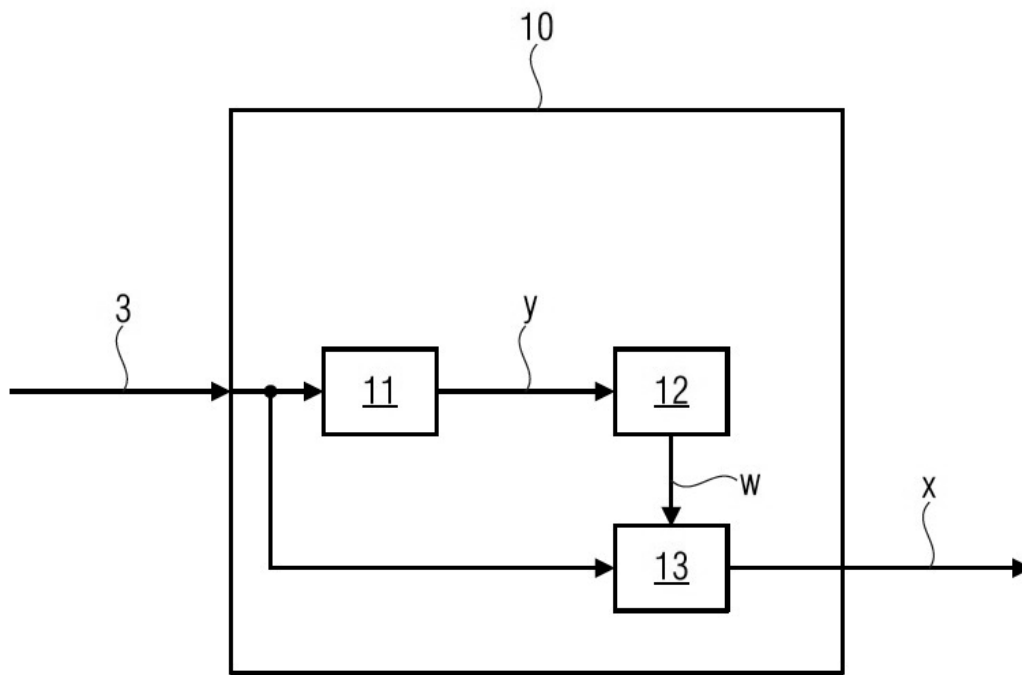


Fig. 3

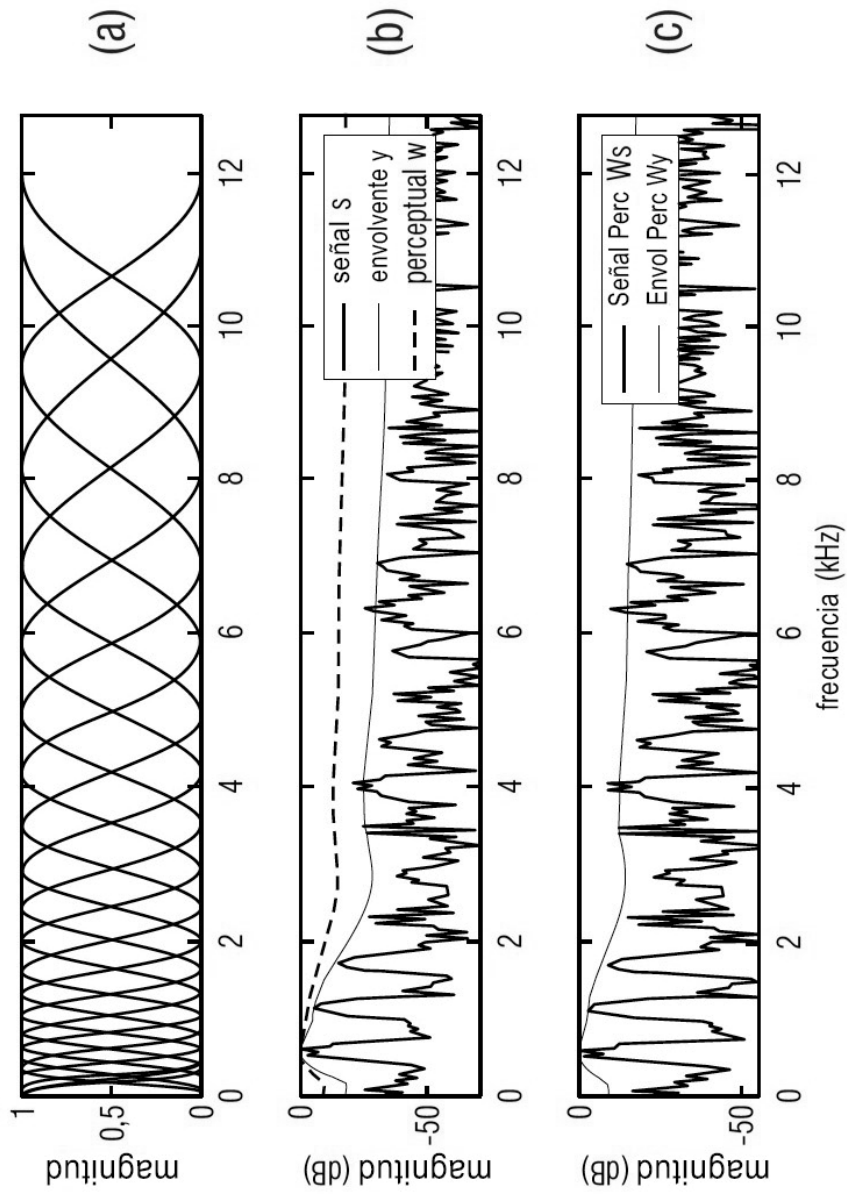


Fig. 4

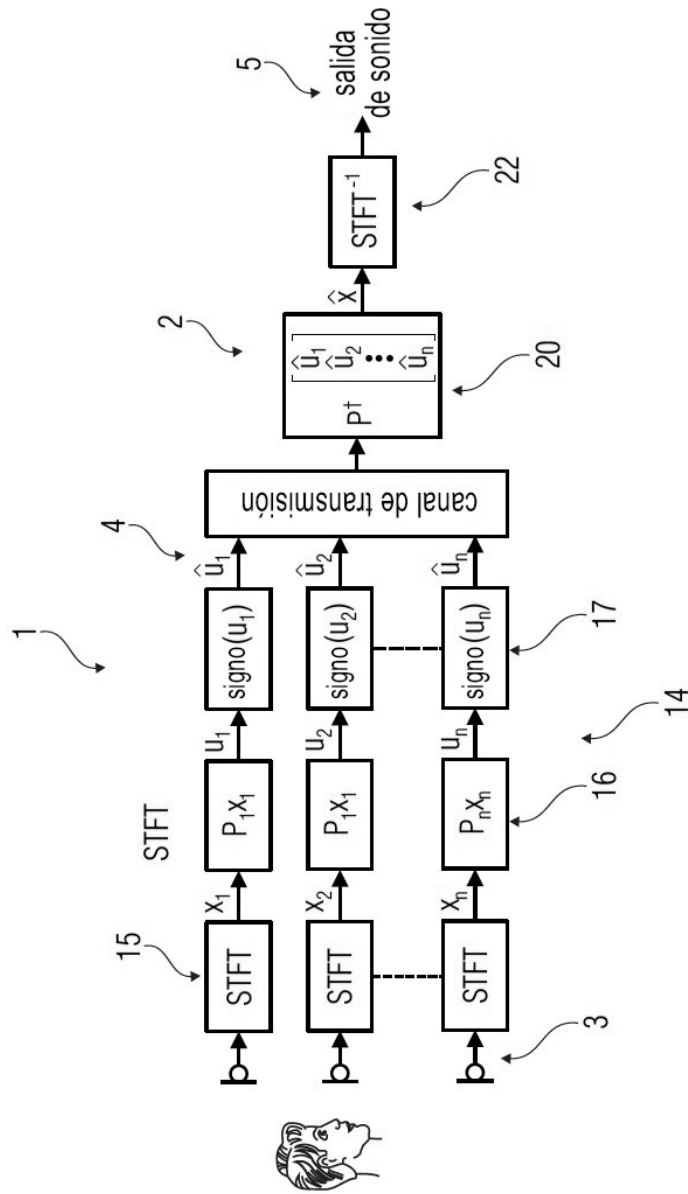


Fig. 5

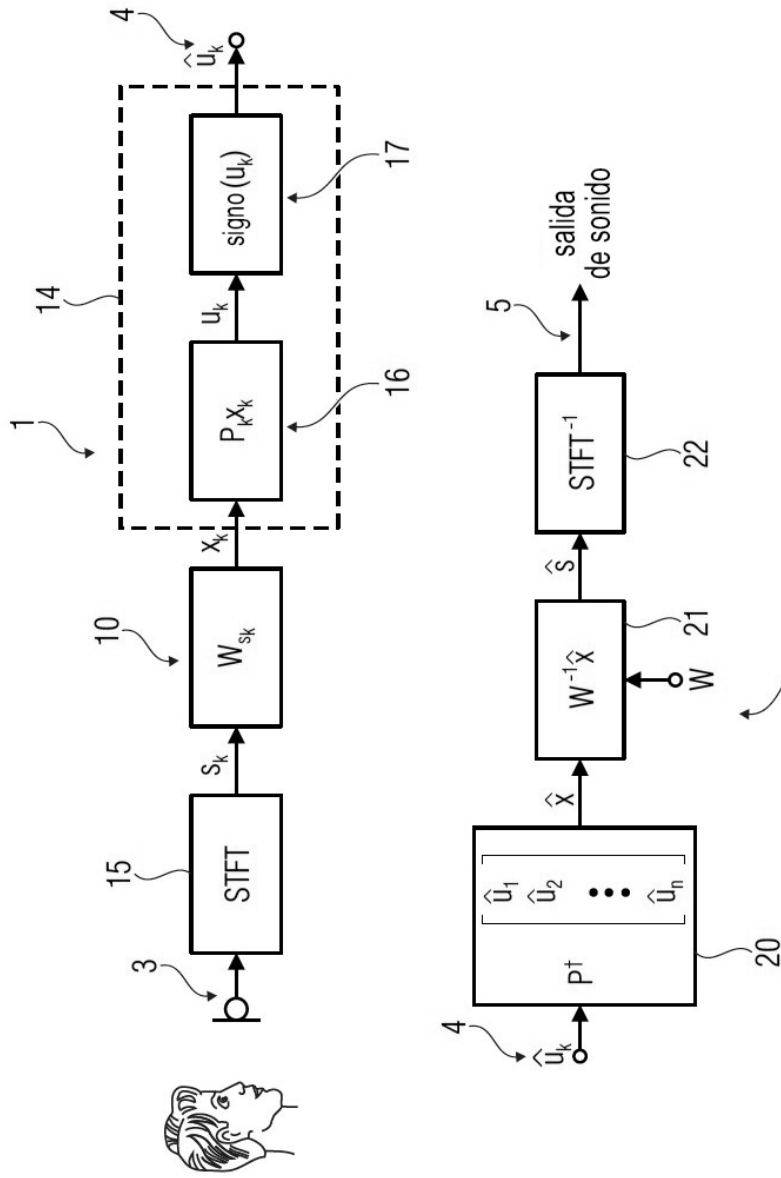


Fig. 6

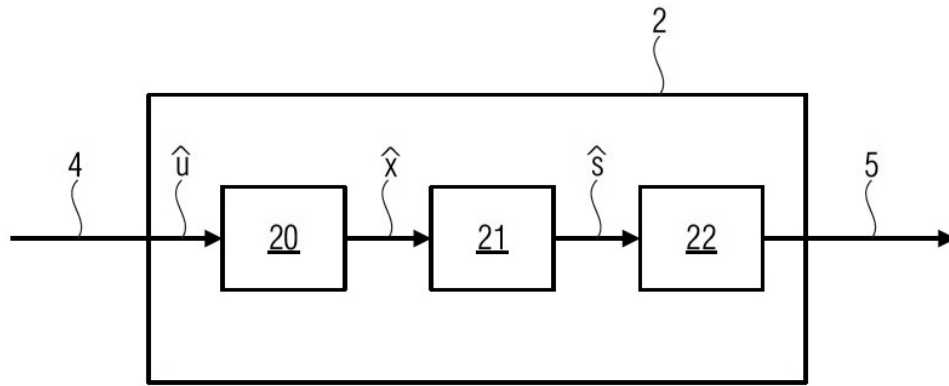


Fig. 7

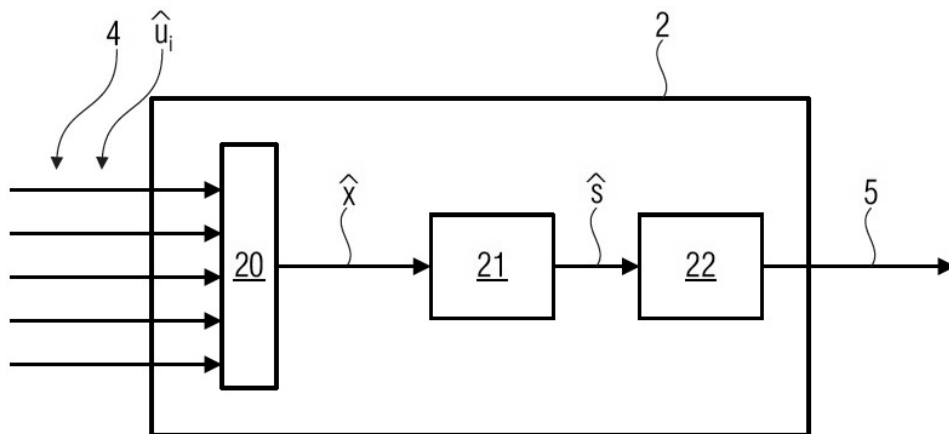


Fig. 8



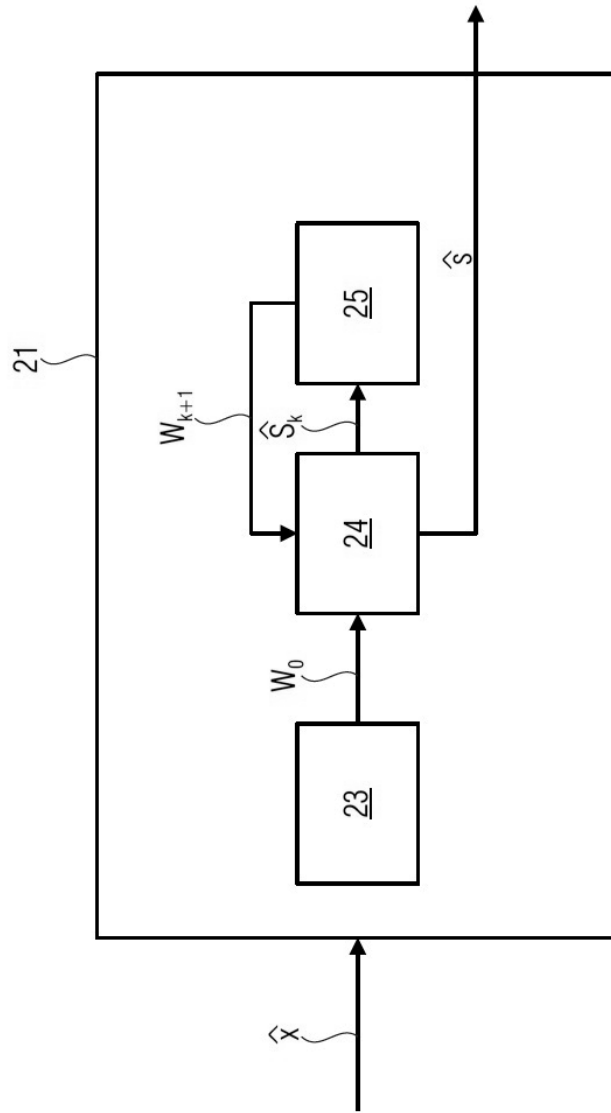


Fig. 9

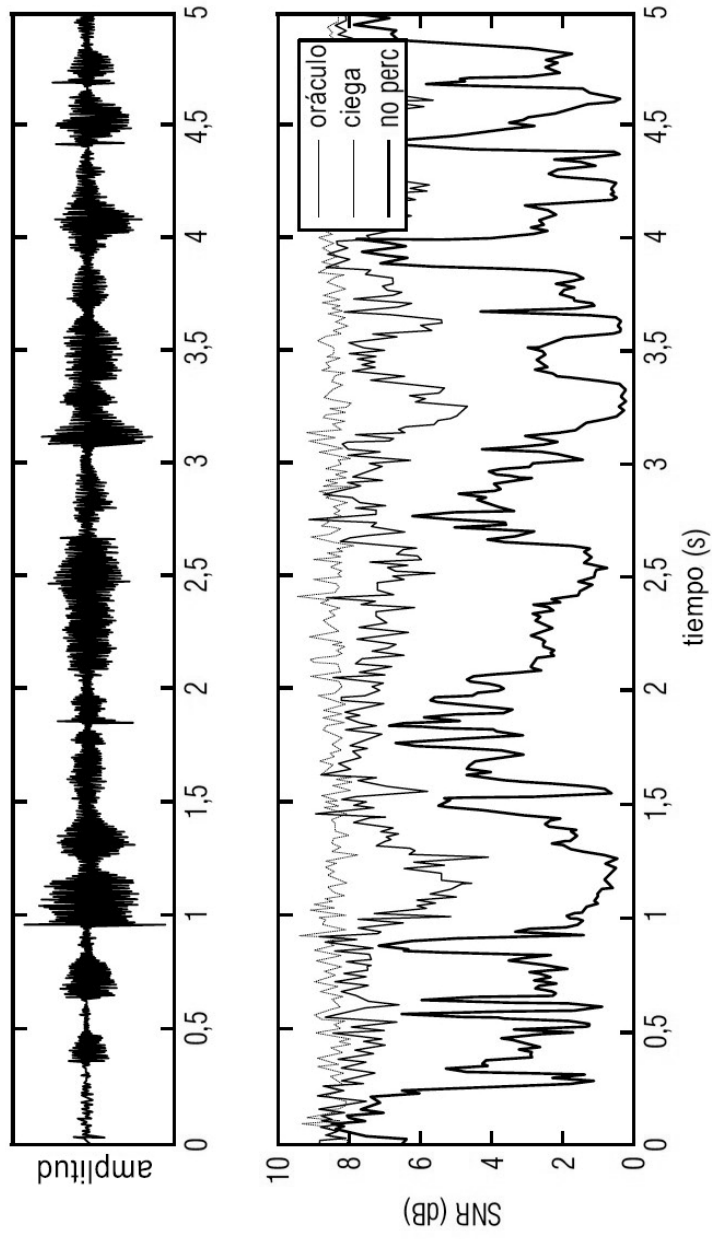


Fig. 10

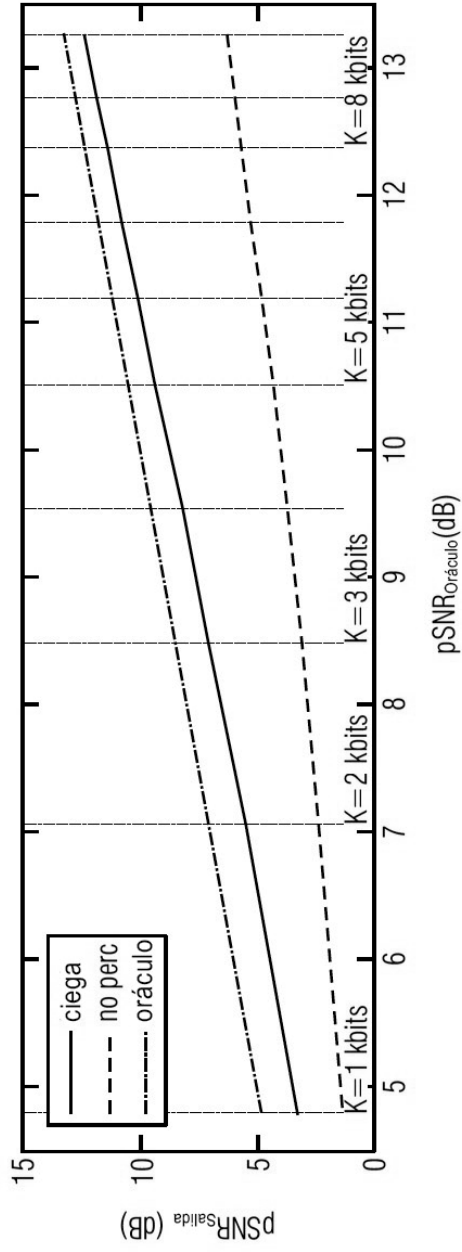


Fig. 11

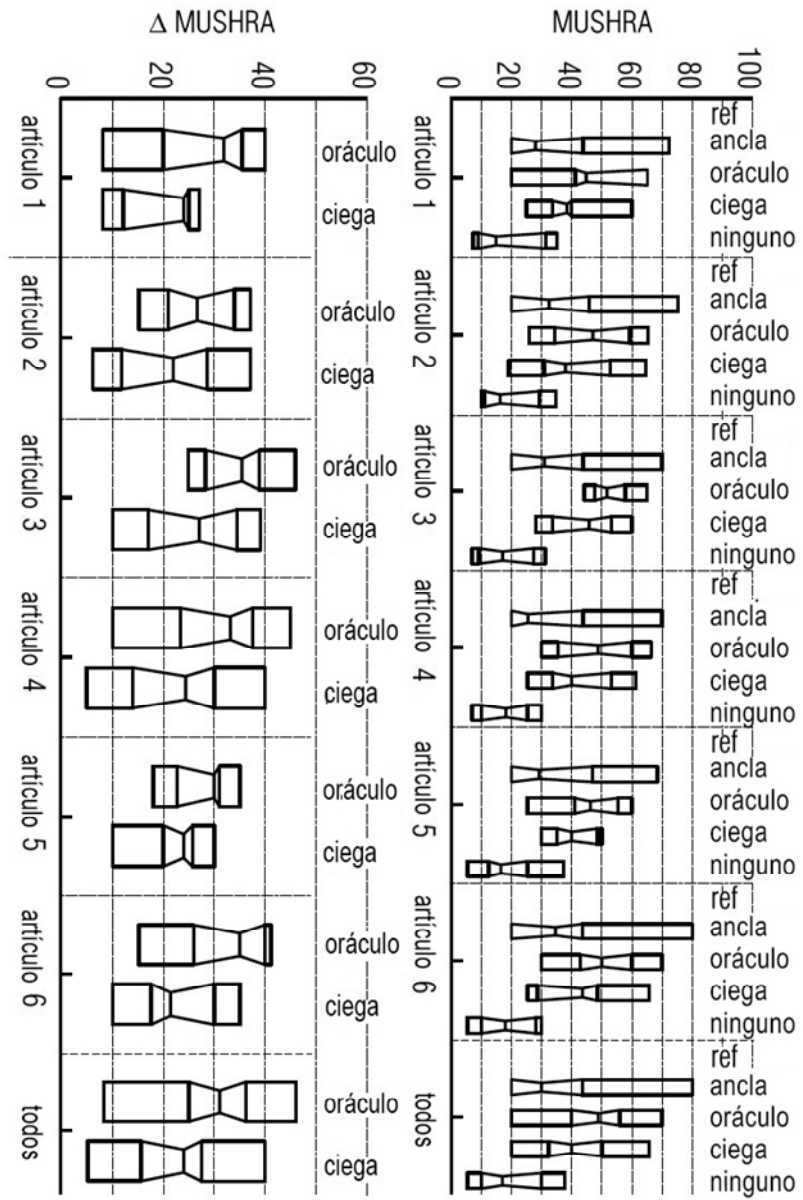


Fig. 12