

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 763 280**

51 Int. Cl.:

**G10L 25/51** (2013.01)

**G10L 25/18** (2013.01)

**G10L 25/81** (2013.01)

**G10L 19/20** (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **07.05.2015** **E 18172361 (0)**

97 Fecha y número de publicación de la concesión europea: **18.09.2019** **EP 3379535**

54 Título: **Clasificador de señales de audio**

30 Prioridad:

**08.05.2014 US 201461990354 P**

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

**27.05.2020**

73 Titular/es:

**TELEFONAKTIEBOLAGET LM ERICSSON (PUBL)**  
**(100.0%)**

**Unit 32, the Hyde Building The Park,**  
**Carrickmines**  
**164 83 Stockholm, SE**

72 Inventor/es:

**NORVELL, ERIK y**  
**GRANCHAROV, VOLODYA**

74 Agente/Representante:

**ELZABURU, S.L.P**

**ES 2 763 280 T3**

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

## DESCRIPCIÓN

Clasificador de señales de audio

**Campo técnico**

La tecnología propuesta se refiere, en general, a códecs y métodos para la codificación de audio.

**5 Antecedentes**

Los códecs de audio modernos consisten en múltiples esquemas de compresión optimizados para señales con diferentes propiedades. Con prácticamente ninguna excepción, las señales de voz se procesan con códecs en el dominio del tiempo, mientras que las señales de música se procesan con códecs del dominio de la transformación. Esquemas de codificación que se supone que manejan señales de voz y música requieren un mecanismo para reconocer si la señal de entrada comprende voz o música, y cambiar entre los modos de códec apropiados. Tal mecanismo se puede denominar clasificador de voz y música, o discriminador. Una ilustración general de un códec de audio multimodo que utiliza la lógica de decisión de modo basada en la señal de entrada se muestra en la figura 1a.

De manera similar, entre la clase de señales de música, se puede discriminar más ruido como las señales musicales de señales de música armónica, y construir un clasificador y un esquema de codificación óptimo para cada uno de estos grupos. Esta abstracción de crear un clasificador para determinar la clase de una señal, que luego controla la decisión del modo se ilustra en la figura 1b.

Existe una variedad de clasificadores de voz y música en el campo de la codificación de audio. El documento de patente US 2012/015840 A1 proporciona un ejemplo de uno de ellos. Sin embargo, estos clasificadores no pueden discriminar entre diferentes clases en el espacio de las señales de música. De hecho, muchos clasificadores conocidos no proporcionan suficiente resolución para poder discriminar entre clases de música de una manera que es necesaria para la aplicación en un códec multimodo complejo.

**Compendio**

El problema de discriminar entre, por ejemplo, entre segmentos de música armónicos y similares al ruido se trata en el presente documento, mediante la utilización de una nueva métrica, calculada directamente sobre los coeficientes del dominio de la frecuencia. La métrica se basa en la distribución de candidatos a picos espectrales preseleccionados y en la relación media del pico de ruido frente al ruido de fondo.

La solución propuesta permite identificar segmentos de música armónicos y similares al ruido, lo que a su vez permite una codificación óptima de estos tipos de señal. Este concepto de codificación proporciona una calidad superior a los esquemas de codificación convencionales. Las realizaciones descritas en la presente memoria tratan de encontrar un mejor clasificador para la discriminación de señales de música armónicas y similares al ruido.

De acuerdo con un primer aspecto, se proporciona un método de clasificación de señales de audio. El método comprende, para un segmento de una señal de audio, identificar un conjunto de picos espectrales y determinar una distancia media  $S$  entre picos en el conjunto. El método comprende además determinar una relación, PNR, entre una energía de envolvente de picos y una energía de envolvente del ruido de fondo. El método comprende además determinar a qué clase de señales de audio, de entre una pluralidad de clases de señales de audio, pertenece el segmento, en base a, al menos, a la distancia media  $S$  y la relación PNR.

De acuerdo con un segundo aspecto, se proporciona un clasificador de señales de audio. El clasificador está configurado para, para un segmento de una señal de audio, identificar un conjunto de picos espectrales y determinar una distancia media  $S$  entre picos en el conjunto. El clasificador está configurado además para determinar una relación, PNR, entre una energía de envolvente de pico y una energía de envolvente del ruido de fondo, y además para determinar a qué clase de señales de audio, de entre una pluralidad de clases de señales de audio, pertenece el segmento, en base a, al menos, la distancia media  $S$  y la relación PNR.

De acuerdo con un tercer aspecto, se proporciona un dispositivo de comunicación, que comprende un clasificador de señales de audio de acuerdo con el segundo aspecto.

De acuerdo con un cuarto aspecto, se proporciona un programa informático, que comprende instrucciones que, cuando son ejecutadas en al menos un procesador, hacen que el al menos un procesador lleve a cabo el método de acuerdo con el primer aspecto.

De acuerdo con un quinto aspecto, se proporciona una portadora, que contiene el programa informático del cuarto aspecto, en donde la portadora es uno de entre una señal electrónica, señal óptica, señal de radio o medio de almacenamiento legible por ordenador.

**Breve descripción de los dibujos**

Los anteriores y otros objetos, características y ventajas de la tecnología descrita en este documento serán evidentes a partir de la siguiente descripción más particular de las realizaciones, tal como se ilustran en los dibujos adjuntos. Los dibujos no están necesariamente a escala, sino que se destaca en ilustrar los principios de la tecnología descrita en el presente documento.

La figura 1a es una ilustración esquemática de un códec de audio en el que se podrían aplicar realizaciones de la invención. La figura 1b es una ilustración esquemática de un códec de audio que muestra explícitamente un clasificador de señal.

La figura 2 es un diagrama de flujo que ilustra un método de acuerdo con una realización a modo de ejemplo.

La figura 3a es un diagrama que ilustra un algoritmo de selección de pico y valores instantáneos de pico y nivel de ruido de acuerdo con una realización a modo de ejemplo.

La figura 3b es un diagrama que ilustra las distancias de pico  $d_i$ , de acuerdo con una realización a modo de ejemplo.

La figura 4 ilustra un diagrama de Venn de decisiones de acuerdo con una realización a modo de ejemplo.

Las figuras 5a a c ilustran implementaciones de un codificador de acuerdo con realizaciones de ejemplo.

La figura 5d ilustra una implementación de un discriminador de acuerdo con una realización a modo de ejemplo.

La figura 6 ilustra una realización de un codificador.

**Descripción detallada**

La tecnología propuesta se puede aplicar a un codificador y/o descodificador, por ejemplo, de un terminal de usuario o equipo de usuario, que puede ser un dispositivo cableado o inalámbrico. Todos los dispositivos y nodos alternativos descritos en este documento se resumen en el término "dispositivo de comunicación", en el que se podría aplicar la solución descrita en este documento.

Tal como se utiliza en este documento, los términos no limitativos "Equipo de usuario" y "dispositivo inalámbrico" se pueden referir a un teléfono móvil, un teléfono celular, un Asistente Digital Personal, PDA (Personal Digital Assistant, en inglés, equipado con capacidades de comunicación de radio, un teléfono inteligente, un ordenador portátil u ordenador personal, PC (Personal Computer, en inglés, equipado con un módem de banda ancha móvil interno o externo, una PC de tableta con capacidades de comunicación de radio, un dispositivo de destino, un UE de dispositivo a dispositivo, un UE de tipo máquina o un UE capaz de comunicación de máquina a máquina, un iPad, un equipo local del cliente, CPE (Customer Premises Equipment, en inglés, un equipo portátil incorporado, un LEE, un equipo montado en un ordenador portátil, LME (Laptop Mounted Equipment, en inglés, una llave electrónica (dongle, en inglés) de USB, un dispositivo de comunicación de radio electrónico portátil, un dispositivo sensor equipado con capacidades de comunicación de radio o similares. En particular, el término "UE" y el término "dispositivo inalámbrico" se deben interpretar como términos no limitativos que comprenden cualquier tipo de dispositivo inalámbrico que se comunica con un nodo de la red de radio en un sistema de comunicación celular o móvil o cualquier dispositivo equipado con circuitos de radio para comunicación inalámbrica de acuerdo con cualquier estándar relevante para la comunicación dentro de un sistema de comunicación celular o móvil.

Tal como se utiliza en el presente documento, el término "dispositivo cableado" se puede referir a cualquier dispositivo configurado o preparado para conexión por cable a una red. En particular, el dispositivo cableado puede ser al menos alguno de los dispositivos anteriores, con o sin capacidad de comunicación por radio, cuando está configurado para conexión por cable.

La tecnología propuesta también se puede aplicar a un codificador y/o descodificador de un nodo de red de radio. Tal como se utiliza en el presente documento, el término no limitativo "nodo de la red de radio" se puede referir a estaciones base, nodos de control de la red tales como controladores de la red, controladores de la red de radio, controladores de la estación base y similares. En particular, el término "estación base" puede abarcar diferentes tipos de estaciones base de radio, que incluyen estaciones base estandarizadas tales como Nodos B, o Nodos B evolucionados, eNB, y también macro / micro / pico estaciones base de radio, estaciones base domésticas, también conocidas como femto estaciones base, nodos repetidores, repetidores, puntos de acceso de radio, estaciones base transceptoras, BTS (Base Transceiver Station, en inglés, e incluso nodos de control de radio que controlan una o más Unidades de Radio Remotas, RRU (Remote Radio Unit, en inglés), o similares.

Las realizaciones de la solución descrita en este documento son adecuadas para su utilización con un códec de audio. Por lo tanto, las realizaciones se describirán en el contexto de un códec de audio a modo de ejemplo, que opera en bloques cortos, por ejemplo, 20 ms, de la forma de onda de entrada. Se debe observar que la solución descrita en este documento también se puede utilizar con otros códecs de audio que operan en otros tamaños de bloque. Además, las realizaciones presentadas muestran ejemplos de valores numéricos, que son preferidos para la

realización en cuestión. Se debe entender que estos valores numéricos se proporcionan solo como ejemplos y se pueden adaptar al códec de audio en cuestión.

Realizaciones a modo de ejemplo

5 A continuación, se describirán ejemplos de realizaciones relacionadas con un método para codificar una señal de audio haciendo referencia a la figura 2. El método debe ser llevado a cabo por un codificador. El codificador puede ser configurado para cumplir con uno o más estándares para la codificación de audio. El método comprende, para un segmento de la señal de audio: identificar 201 un conjunto de picos espectrales; determinar 202 una distancia media S entre picos en el conjunto; y determinar 203 una relación, PNR, entre una envolvente de picos y una envolvente del ruido de fondo. El método comprende además seleccionar 204 un modo de codificación de una pluralidad de modos de codificación, en base, por lo menos, a la distancia media S y la relación PNR; y aplicar 205 el modo de codificación seleccionado.

15 Los picos espectrales se pueden identificar de diferentes maneras, que también se describirán con más detalle a continuación. Por ejemplo, los coeficientes espectrales cuya magnitud excede un umbral definido podrían ser identificados como pertenecientes a un pico. Cuando se determina la distancia media S entre los picos, cada pico se puede representar por un único coeficiente espectral. Este coeficiente único sería preferiblemente el coeficiente espectral que tiene la amplitud cuadrada máxima de entre los coeficientes espectrales (si hay más de uno) asociados con el pico. Es decir, cuando se identifica que más de un coeficiente espectral está asociado con un pico espectral, entonces se puede seleccionar uno de la pluralidad de coeficientes asociados con el pico para representar el pico cuando se determina la distancia media S. Esto se podría ver en la figura 3b, y se describirá con más detalle a continuación. La distancia media S también se puede consultar, por ejemplo, como la "dispersión de los picos".

20 Para determinar una relación entre una envolvente de picos y una envolvente del ruido de fondo, estas envolventes necesitan ser estimadas. La envolvente del ruido de fondo se puede estimar en base a los valores absolutos de los coeficientes espectrales y a un factor de ponderación que destaca la contribución de los coeficientes de baja energía. De manera correspondiente, la envolvente de picos se puede estimar en base a los valores absolutos de los coeficientes espectrales y a un factor de ponderación que destaca la contribución de los coeficientes de alta energía. Las figuras 3a y 3b muestran ejemplos de envolventes estimadas del ruido de fondo (guiones cortos) y envolventes del pico (guiones largos). Por coeficientes de "baja energía" y "alta energía" se deben entender los coeficientes que tienen una amplitud con una cierta relación con un umbral, donde los coeficientes de baja energía serían típicamente coeficientes que tienen una amplitud por debajo (o posiblemente igual a) de cierto umbral, y los coeficientes de alta energía serían típicamente coeficientes que tienen una amplitud por encima (o posiblemente igual a) de un cierto umbral.

25 De acuerdo con una realización a modo de ejemplo, la forma de onda de entrada, es decir, la señal de audio, se destaca previamente, por ejemplo, con un filtro de paso alto de primer orden  $H(z) = 1 - 0,68z^{-1}$  antes de realizar el análisis espectral. Esto se puede hacer, por ejemplo, con el fin de aumentar la precisión de la modelización para la región de alta frecuencia, pero se debe observar que no es esencial para la invención en cuestión.

30 Se puede utilizar una transformada discontinua de Fourier (DFT – Discrete Fourier Transform, en inglés) para convertir la señal de audio filtrada en el dominio de la transformación o de la frecuencia. En un ejemplo específico, el análisis espectral se realiza una vez por trama utilizando una transformada rápida de Fourier (FFT – Fast Fourier Transform, en inglés) de 256 puntos.

35 Se ejecuta una FFT sobre la señal de entrada enmarcada destacada previamente, es decir, sobre un segmento de la señal de audio, para obtener un conjunto de parámetros espectrales como:

$$X(k) = \sum_{n=0}^{255} x(n)e^{-j2\pi\frac{kn}{256}}$$

40 donde  $k = 0, \dots, 255$ , es un índice de coeficientes de frecuencia o coeficientes espectrales, y  $n$  es un índice de muestras de formas de onda. Se debe observar que se puede utilizar cualquier longitud  $N$  de la transformada. Los coeficientes también se podrían denominar coeficientes de la transformada.

45 Un objetivo de la solución descrita en este documento es lograr un clasificador o discriminador, que no solo pueda discriminar entre voz y música, sino también discriminar entre diferentes tipos de música. A continuación, se describirá con más detalle cómo se puede lograr este objetivo de acuerdo con una realización a modo de ejemplo de un discriminador: El discriminador que se ejemplifica requiere el conocimiento de la ubicación, por ejemplo, en frecuencia, de los picos espectrales de un segmento de la señal de audio de entrada. Los picos espectrales se definen en este documento como coeficientes con un valor absoluto por encima de un umbral adaptativo, que, por ejemplo, se basa en la relación de las envolventes de pico y del ruido de fondo.

50 Se puede utilizar un algoritmo de estimación del ruido de fondo que opera sobre los valores absolutos de los coeficientes de transformación  $|X(k)|$ . Las energías del ruido de fondo instantáneo  $E_{nf}(k)$  se pueden estimar de acuerdo con la repetición:

$$E_{nf}(k) = \alpha E_{nf}(k-1) + (1-\alpha) |X(k)|^2$$

$$\alpha = \begin{cases} 0,9578 & \text{si } |X(k)|^2 > E_{nf}(k-1) \\ 0,6472 & \text{si } |X(k)|^2 \leq E_{nf}(k-1) \end{cases}$$

La forma particular del factor de ponderación  $\alpha$  minimiza el efecto de los coeficientes de transformación de alta energía y destaca la contribución de los coeficientes de baja energía. Finalmente, el nivel del ruido de fondo  $E_{nf}$  se calcula simplemente promediando las energías  $E_{nf}$  instantáneas.

5

$$\bar{E}_{nf} = (\sum_{k=0}^{255} E_{nf}(k)) / 256$$

Una realización del algoritmo de "pico de selección" presentado en la presente memoria requiere el conocimiento de un nivel de energía de la parte entera de ruido y el nivel de energía promedio de los picos espectrales. El algoritmo de estimación de pico de energía utilizado aquí es similar al algoritmo de estimación de la parte entera de ruido anterior, pero en lugar de bajo consumo de energía, rastrea energías de alto espectro como:

$$E_p(k) = \beta E_p(k-1) + (1-\beta) |X(k)|^2$$

$$\beta = \begin{cases} 0,4223 & \text{si } |X(k)|^2 > E_p(k-1) \\ 0,8029 & \text{si } |X(k)|^2 \leq E_p(k-1) \end{cases}$$

10

En este caso, el factor de ponderación  $\beta$  minimiza el efecto los coeficientes de transformación de baja energía y destaca la contribución de los coeficientes de alta energía. La energía de pico  $\bar{E}_p$  global se estima de este modo promediando las energías instantáneas como:

$$\bar{E}_p = (\sum_{k=0}^{255} E_p(k)) / 256$$

15 Cuando se calculan los niveles de pico y el ruido de fondo, se puede formar un nivel umbral  $\tau$  como:

$$\tau = \left( \frac{\bar{E}_p}{\bar{E}_{nf}} \right)^\gamma \bar{E}_{nf}$$

20

estando y establecido en el valor de ejemplo  $\gamma = 0,88579$ . Los coeficientes de transformación de un segmento de la señal de audio de entrada se comparan a continuación con el umbral, y los que tienen una amplitud que excede el umbral forman un vector de candidatos a pico. Es decir, un vector que comprende los coeficientes que se supone que pertenecen a picos espectrales.

Se puede utilizar un valor umbral alternativo,  $\theta(k)$ , para detectar picos, que puede requerir una menor complejidad de cálculo que  $\tau$ . En una realización,  $\theta(k)$  se encuentra como el nivel de la envolvente de picos instantáneo,  $E_p(k)$ , con un factor de escala fijo. En este caso, el factor de escala 0,64 se utiliza como ejemplo, de modo que:

$$\theta(k) = E_p(k) \cdot 0,64$$

25 Cuando se utiliza el umbral alternativo,  $\theta$ , los candidatos a pico se definen como todos los coeficientes con una amplitud cuadrada por encima del nivel de umbral instantáneo, tales como:

$$\begin{cases} |X(k)|^2 > \theta(k), k \in P \\ |X(k)|^2 \leq \theta(k), k \notin P \end{cases}$$

30

donde  $P$  denota el conjunto ordenado por frecuencia de las posiciones de los candidatos a pico. Teniendo en cuenta el espectro de FFT, algunos picos serán amplios y constarán de varios coeficientes de transformación, mientras que otros son estrechos y están representados por un único coeficiente. Para obtener una representación máxima de los coeficientes únicos, es decir, un coeficiente por pico, se supone que los coeficientes candidatos a pico en las posiciones consecutivas forman parte de un pico más amplio. Encontrando la amplitud cuadrada máxima  $|X(k)|^2$  de los coeficientes de transformación en un rango de las posiciones candidatas a pico consecutivas ...  $k-1, k, k+1, \dots$ , se crea un conjunto  $P$  refinado, en el que los picos anchos están representados por la posición máxima en cada rango, es decir, por el coeficiente que tiene el valor más alto de  $|X(k)|^2$  en el rango, que también se podría indicar

35

como el coeficiente que tiene la mayor magnitud espectral en el rango. La figura 3a ilustra la derivación de la envolvente de picos y la envolvente del ruido de fondo, y el algoritmo de selección de pico.

5 Los cálculos anteriores sirven para generar dos características que se utilizan para formar una decisión de clasificación: a saber, una estimación de la dispersión de los picos  $S$  y una relación del pico de ruido frente al ruido de fondo PNR. La dispersión de los picos  $S$  se puede representar o definir utilizando la distancia promedio  $d_i$  entre picos como:

$$S = \begin{cases} \frac{1}{N_d} \sum_{i=1}^N d_i, & N_d > 1 \\ 0, & N_d \leq 1 \end{cases}$$

donde  $N_d$  es el número de picos refinados en el conjunto  $P$ . El PNR se puede calcular como

$$PNR = \frac{\bar{E}_p}{\bar{E}_{nf}}$$

10 La decisión del clasificador se puede formar utilizando estas características en combinación con un umbral de decisión. Podemos nombrar estas decisiones como "issparse" e "isclean", como:

$$issparse = S > S_{THR}$$

$$isclean = PNR > PNR_{THR}$$

15 El resultado de estas decisiones se puede utilizar para formar diferentes clases de señales. Una ilustración de estas clases se muestra en la figura 4. Cuando la clasificación se basa en dos decisiones binarias, el número total de clases puede ser como máximo 4. Como siguiente etapa, la decisión del códec se puede formar utilizando la información de la clase, que se ilustra en la Tabla 1.

Tabla 1: Clases posibles formadas utilizando dos decisiones de características

	isclean	Issparse
Clase A	falso	falso
Clase B	verdadero	falso
Clase C	verdadero	verdadero
Clase D	falso	verdadero

20 En la siguiente etapa del códec de audio, se tomará una decisión acerca de qué etapas de procesamiento aplicar a qué clase. Es decir, se debe seleccionar un modo de codificación en base, por lo menos, a  $S$  y PNR. Esta selección o representación (mapping, en inglés) dependerá de las características y capacidades de los diferentes modos de codificación o etapas de procesamiento disponibles. Como ejemplo, quizás el modo 1 de códec manejaría la clase A y la clase C, mientras que el modo códec 2 manejaría la clase B y la clase D. La decisión del modo de codificación puede ser la salida final del clasificador para guiar el proceso de codificación. La decisión del modo de codificación típicamente se transferiría en el flujo de bits junto con los parámetros del códec del modo de codificación elegido.

25 Se debe entender que las clases anteriores se pueden combinar con otras decisiones de clasificación. La combinación puede dar como resultado un mayor número de clases, o se pueden combinar utilizando un orden de prioridad tal que el clasificador presentado puede ser anulado por otro clasificador, o viceversa, que el clasificador presentado puede anular a otro clasificador.

30 La solución descrita en este documento proporciona un discriminador de tipo de música de alta resolución, que podría, con ventaja, ser aplicado en la codificación de audio. La lógica de decisión del discriminador se basa en estadísticas de la distribución posicional de los coeficientes de frecuencia con energía prominente.

#### Implementaciones

El método y las técnicas descritas anteriormente se pueden implementar en codificadores y/o descodificadores, que pueden formar parte, por ejemplo, de dispositivos de comunicación.

35

Codificador, figuras 5a a 5c.

En la figura 5a se ilustra de una manera general una realización a modo de ejemplo de un codificador. Por codificador se denomina un codificador configurado para la codificación de señales de audio. El codificador se podría configurar posiblemente para codificar otros tipos de señales. El codificador 500 está configurado para realizar al menos una de las realizaciones del método descritas anteriormente, por ejemplo, haciendo referencia a la figura 2. El codificador 500 está asociado con las mismas características técnicas, objetos y ventajas que las realizaciones del método descrito anteriormente. El codificador se puede configurar para cumplir con uno o más estándares para la codificación de audio. El codificador se describirá brevemente para evitar repeticiones innecesarias. El codificador se puede implementar y/o describir de la siguiente manera:

10 El codificador 500 está configurado para la codificación de una señal de audio. El codificador 500 comprende una circuitería de procesamiento, o medios de procesamiento 501 y una interfaz de comunicación 502. La circuitería de procesamiento 501 está configurada para hacer que el codificador 500, para un segmento de la señal de audio, identifique un conjunto de picos espectrales; determinar una distancia media  $S$  entre los picos en el conjunto; y para determinar una relación, PNR, entre una envolvente de picos y una envolvente del ruido de fondo. La circuitería de procesamiento 501 está configurada además para hacer que el codificador seleccione un modo de codificación, de ente una pluralidad de modos de codificación, en base, al menos, a la distancia media  $S$  y la relación PNR; y para aplicar el modo de codificación seleccionado. La interfaz de comunicación 502, que también se puede indicar, por ejemplo, como la interfaz de entrada / salida (I/O – Input / Output, en inglés) incluye una interfaz para enviar y recibir datos de otras entidades o módulos.

20 La circuitería de procesamiento 501 podría, como se ilustra en la figura 5b, comprender medios de procesamiento, tales como un procesador 503, por ejemplo, una CPU y una memoria 504 para almacenar o contener instrucciones. La memoria comprendería, por lo tanto, instrucciones, por ejemplo, en forma de un programa informático 505, el cual, cuando es ejecutado por los medios de procesamiento 503 hace que el codificador 500 realice las acciones descritas anteriormente.

25 Una implementación alternativa de la circuitería de procesamiento 501 se muestra en la figura 5c. La circuitería de procesamiento comprende, en este caso, una unidad de identificación 506, configurada para identificar un conjunto de picos espectrales, para / de un segmento de la señal de audio. La circuitería de procesamiento comprende además una primera unidad de determinación 507, configurada para hacer que el codificador 500 determine una distancia media  $S$  entre los picos en el conjunto. La circuitería de procesamiento comprende además una segunda unidad de determinación 508 configurada para hacer que el codificador determine una relación, PNR, entre una envolvente de picos y una envolvente del ruido de fondo. La circuitería de procesamiento comprende además una unidad de selección 509, configurada para hacer que el codificador seleccione un modo de codificación, de entre una pluralidad de modos de codificación, en base, al menos, a la distancia media  $S$  y la relación PNR. La circuitería de procesamiento comprende además una unidad de codificación 510, configurada para hacer que el codificador aplique el modo de codificación seleccionado. La circuitería de procesamiento 501 podría comprender más unidades, tal como una unidad de filtro configurada para hacer que el codificador filtre la señal de entrada. Esta tarea, cuando se realiza, podría ser realizada alternativamente por una o más de las otras unidades.

Los codificadores, o códecs, descritos anteriormente se podrían configurar para las diferentes realizaciones del método descritas en el presente documento, tales como la utilización de diferentes umbrales para detectar picos. Se puede suponer que el codificador 500 comprende una funcionalidad adicional, para llevar a cabo funciones regulares del codificador.

Ejemplos de la circuitería de procesamiento incluyen, pero no se limitan a, uno o más microprocesadores, uno o más procesadores de señal digital, DSP (Digital Signal Processor, en inglés, una o más unidades de procesamiento central, CPU (Central Processing Unit, en inglés), hardware de aceleración de video y/o cualquier circuitería lógica programable adecuada tal como una o más matrices de puertas programables en campo, FPGA (Field Programmable Gate Array, en inglés, o uno o más controladores lógicos programables, PLC – Programmable Logic Controller, en inglés.

Asimismo, se debe entender que puede ser posible reutilizar las capacidades generales de procesamiento de cualquier dispositivo o unidad convencional en el que se implemente la tecnología propuesta. También es posible reutilizar el software existente, por ejemplo, reprogramando el software existente o agregando nuevos componentes de software.

Discriminador, figura 5d

La figura 5d muestra una implementación a modo de ejemplo de un discriminador, o clasificador, que se podría aplicar en un codificador o decodificador. Tal como se ilustra en la figura 5d, el discriminador descrito en este documento se podría implementar, por ejemplo, mediante uno o más de un procesador y un software adecuado con almacenamiento o memoria adecuados, por lo tanto, para realizar la acción discriminatoria de una señal de entrada, de acuerdo con las realizaciones descritas en este documento. En la realización ilustrada en la figura 5d, una señal entrante es recibida a través de una entrada (ENTRADA), a la que están conectados el procesador y la memoria, y

la representación discriminadora de una señal de audio (parámetros) obtenida a partir del software es enviada a la salida (SALIDA).

5 El discriminador podría discriminar entre diferentes tipos de señales de audio mediante, para un segmento de una señal de audio, identificación de un conjunto de picos espectrales y determinar una distancia media  $S$  entre los picos en el conjunto. Además, el discriminador podría determinar una relación, PNR, entre una envolvente de picos y una envolvente del ruido de fondo y, a continuación, determinar a qué clase de señales de audio, de entre una pluralidad de clases de señales de audio, pertenece el segmento, en base, al menos, a la distancia media  $S$  y la relación PNR. Mediante la realización de este método, el discriminador permite, por ejemplo, una selección adecuada de un método de codificación u otro método relacionado con el procesamiento de señales para la señal de audio.

10 La tecnología descrita anteriormente se puede utilizar, por ejemplo, en un emisor, que puede ser utilizado en un dispositivo móvil (por ejemplo, un teléfono móvil, un ordenador portátil) o un dispositivo estacionario, tal como un ordenador personal, tal como se mencionó previamente.

15 En la figura 6 se puede ver una visión general de un discriminador de señal de audio a modo de ejemplo. La figura 6 muestra un diagrama de bloques esquemático de un codificador con un discriminador de acuerdo con una realización a modo de ejemplo. El discriminador comprende una unidad de entrada configurada para recibir una señal de entrada que representa una señal de audio a manipular, una unidad de organización en tramas, una unidad de pre-énfasis opcional, una unidad de transformación de frecuencia, una unidad de análisis de la envolvente de picos / ruido, una unidad de selección de candidato a pico, una unidad de refinamiento de candidato a pico, una unidad de cálculo de característica, una unidad de decisión de clase, una unidad de decisión de modo de  
20 codificación, una unidad de codificación multimodo, una transmisión / almacenamiento de bits y una unidad de salida para la señal de audio. Todas estas unidades se podrían implementar en hardware. Existen numerosas variantes de elementos de la circuitería que se pueden utilizar y combinar para lograr las funciones de las unidades del codificador. Dichas variantes están abarcadas por las realizaciones. Ejemplos particulares de la implementación en hardware del discriminador son la implementación en hardware del procesador de señal digital (DSP) y tecnología  
25 de circuito integrado, que incluye circuitería electrónica de propósito general y circuitería específica para una aplicación.

30 Un discriminador de acuerdo con una realización descrita en este documento podría ser una parte de un codificador, tal como se describió previamente, y un codificador de acuerdo con una realización descrita en este documento podría ser una parte de un dispositivo o un nodo. Tal como se mencionó anteriormente, la tecnología descrita en este documento se puede utilizar, por ejemplo, en un emisor, que se puede utilizar en un dispositivo móvil, como por ejemplo en un teléfono móvil o un ordenador portátil; o en un dispositivo estacionario, tal como un ordenador personal.

35 Se debe entender que la elección de unidades o módulos interactivos, así como el nombramiento de las unidades son solo para fines de ejemplo, y se pueden configurar en una pluralidad de formas alternativas con el fin de poder ejecutar las acciones de proceso descritas.

Se debe observar asimismo que las unidades o módulos descritos en esta descripción se deben considerar como entidades lógicas, y no necesariamente como entidades físicas separadas. Se apreciará que el alcance de la tecnología descrita en este documento abarca completamente otras realizaciones que pueden llegar a ser obvias para los expertos en la técnica, y que el alcance de esta descripción, por consiguiente, no está limitado.

40 La referencia a un elemento en singular no pretende significar "uno y solo uno" a menos que así se indique explícitamente, sino más bien "uno o más". Todos los equivalentes estructurales y funcionales de los elementos de las formas de realización descritas anteriormente que son conocidas por los expertos en la técnica se incorporan expresamente en este documento como referencia, y se pretende que estén incluidos en este documento. Además, no es necesario que un dispositivo o método aborde todos y cada uno de los problemas que se pretende resolver mediante la tecnología descrita en este documento, para que esté abarcado por este documento.  
45

50 En la descripción anterior, a los fines de explicación y no de limitación, se exponen detalles específicos tales como arquitecturas particulares, interfaces, técnicas, etc. con el fin de proporcionar una comprensión completa de la tecnología descrita. Sin embargo, será evidente para los expertos en la técnica que la tecnología descrita se puede poner en práctica en otras realizaciones y/o en combinaciones de realizaciones que se apartan de estos detalles específicos. Es decir, los expertos en la materia podrán concebir diversas disposiciones que, aunque no se describan explícitamente o se muestren en el presente documento, incorporen los principios de la tecnología descrita. En algunos casos, se omiten las descripciones detalladas de dispositivos, circuitos y métodos bien conocidos para no oscurecer la descripción de la tecnología descrita con detalles innecesarios. Todas las afirmaciones en el presente documento que enumeran principios, aspectos y realizaciones de la tecnología descrita,  
55 así como sus ejemplos específicos, pretenden abarcar sus equivalentes estructurales y funcionales. Además, se pretende que dichos equivalentes incluyan tanto equivalentes actualmente conocidos como equivalentes desarrollados en el futuro, por ejemplo, cualquier elemento desarrollado que realice la misma función, con independencia de la estructura.

5 De este modo, por ejemplo, los expertos en la técnica apreciarán que las figuras de este documento pueden representar vistas conceptuales de circuitería ilustrativa o de otras unidades funcionales que incorporan los principios de la tecnología, y/o diversos procesos que pueden ser representados sustancialmente en un medio legible por ordenador y ejecutados por un ordenador o procesador, incluso aunque dicho ordenador o procesador no se muestre explícitamente en las figuras.

10 Las funciones de los diversos elementos, incluidos los bloques funcionales, se pueden proporcionar mediante la utilización de hardware, tal como hardware de circuito y/o un hardware capaz de ejecutar un software en forma de instrucciones codificadas almacenadas en un medio legible por ordenador. Por lo tanto, dichas funciones y bloques funcionales ilustrados se deben entender como implementados mediante hardware y/o implementados por un ordenador, y por lo tanto implementados en una máquina.

15 Las realizaciones descritas anteriormente se deben entender como ejemplos ilustrativos de la presente invención. Los expertos en la técnica comprenderán que se pueden realizar diversas modificaciones, combinaciones y cambios en las realizaciones sin apartarse del alcance de la presente invención. En particular, las diferentes soluciones de partes en las diferentes realizaciones se pueden combinar en otras configuraciones, cuando sea técnicamente posible.

**Abreviaturas**

- DFT Transformada discontinua de Fourier
- FFT Transformada rápida de Fourier
- MDCT Transformada del coseno discontinua modificada
- 20 PNR Relación del pico de ruido frente al ruido de fondo

**REIVINDICACIONES**

1. Un método para clasificar señales de audio, comprendiendo el método:  
para un segmento de una señal de audio:
  - identificar un conjunto de picos espectrales;
- 5
  - determinar una distancia media  $S$  entre picos en el conjunto;
  - determinar una relación, PNR, entre una energía de envolvente de pico y una energía de envolvente del ruido de fondo;
  - determinar a qué clase de señales de audio, de entre una pluralidad de señales de audio, pertenece el segmento, en base, al menos, a la distancia media  $S$  y la relación PNR.
- 10 2. El método de acuerdo con la reivindicación 1, en el que, cuando se determina  $S$ , cada pico está representado por un coeficiente espectral, estando asociado el coeficiente espectral que tiene la amplitud máxima al cuadrado de los coeficientes espectrales, con el pico.
- 15 3. El método de acuerdo con la reivindicación 1, en el que la energía de envolvente de pico se estima en base a valores absolutos de coeficientes espectrales y a un factor de ponderación que destaca la contribución de los coeficientes de alta energía en comparación con los coeficientes de baja energía.
4. El método de acuerdo con la reivindicación 1, en el que la energía de envolvente del ruido de fondo se estima en base a valores absolutos de coeficientes espectrales y a un factor de ponderación que destaca la contribución de los coeficientes de baja energía en comparación con los coeficientes de alta energía.
5. Un clasificador de señales de audio configurado para:  
para un segmento de una señal de audio:
  - identificar un conjunto de picos espectrales;
  - determinar una distancia media  $S$  entre picos en el conjunto;
  - determinar una relación, PNR, entre una energía de envolvente de pico y una energía de envolvente del ruido de fondo;
- 20
  - determinar a qué clase de señales de audio, de entre una pluralidad de señales de audio, pertenece el segmento, en base, al menos, a la distancia media  $S$  y la relación PNR.
6. El clasificador de señales de audio de acuerdo con la reivindicación 5, en el que, cuando se determina la distancia media  $S$ , cada pico está representado por un coeficiente espectral, que es el coeficiente espectral que tiene la amplitud máxima al cuadrado de los coeficientes espectrales asociados con el pico.
- 30 7. El clasificador de señales de audio de acuerdo con la reivindicación 5, que está configurado para estimar la energía de envolvente de pico en base a valores absolutos de coeficientes espectrales y a un factor de ponderación que destaca la contribución de los coeficientes de alta energía en comparación con los coeficientes de baja energía.
- 35 8. El clasificador de señales de audio de acuerdo la reivindicación 5, que está configurado para estimar la energía de envolvente del ruido de fondo en base a valores absolutos de coeficientes espectrales y a un factor de ponderación que destaca la contribución de los coeficientes de baja energía en comparación con los coeficientes de alta energía.
9. Dispositivo de comunicación que comprende un clasificador de señal de acuerdo con una cualquiera de las reivindicaciones 5 a 8.
10. Programa informático, que comprende instrucciones que, cuando son ejecutadas en al menos un procesador, hacen que el al menos un procesador lleve a cabo el método de acuerdo con una cualquiera de las reivindicaciones 1 a 4.
- 40 11. Un portadora que contiene el programa informático de la reivindicación anterior, en el que el portadora es uno de entre una señal electrónica, señal óptica, señal de radio o medio de almacenamiento legible por ordenador.

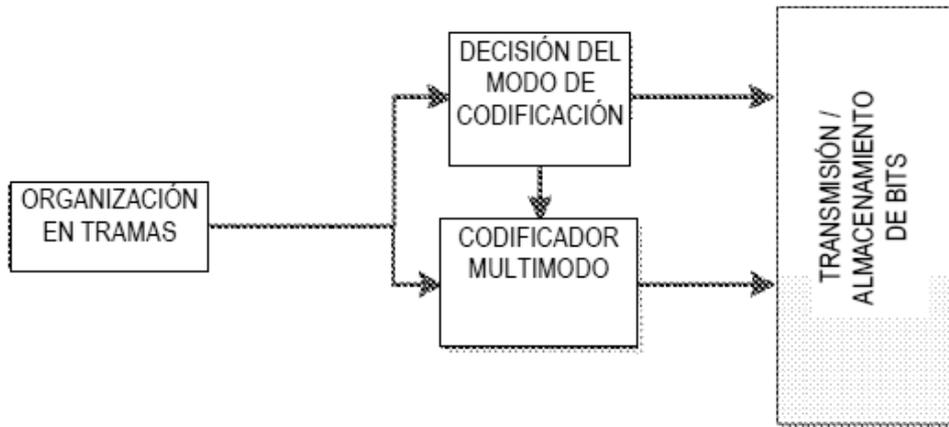


Figura 1a

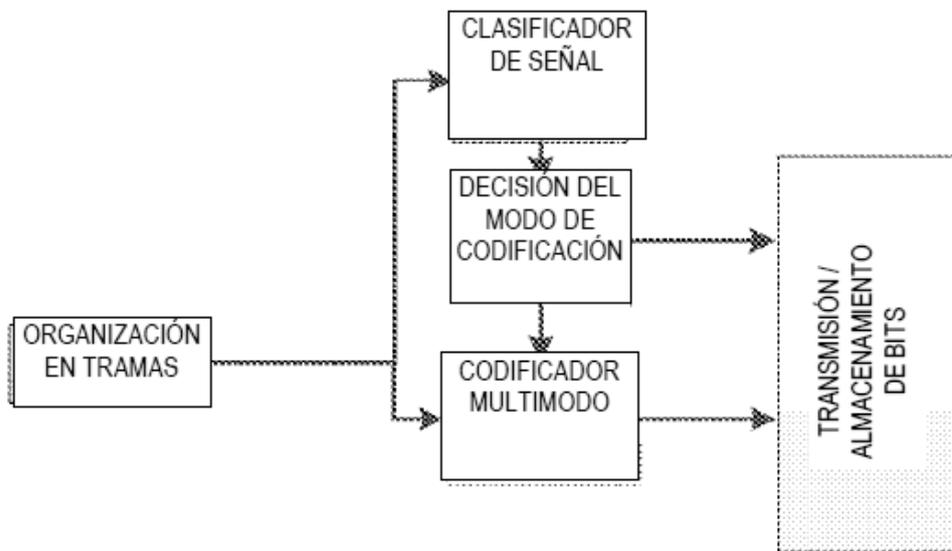


Figura 1b

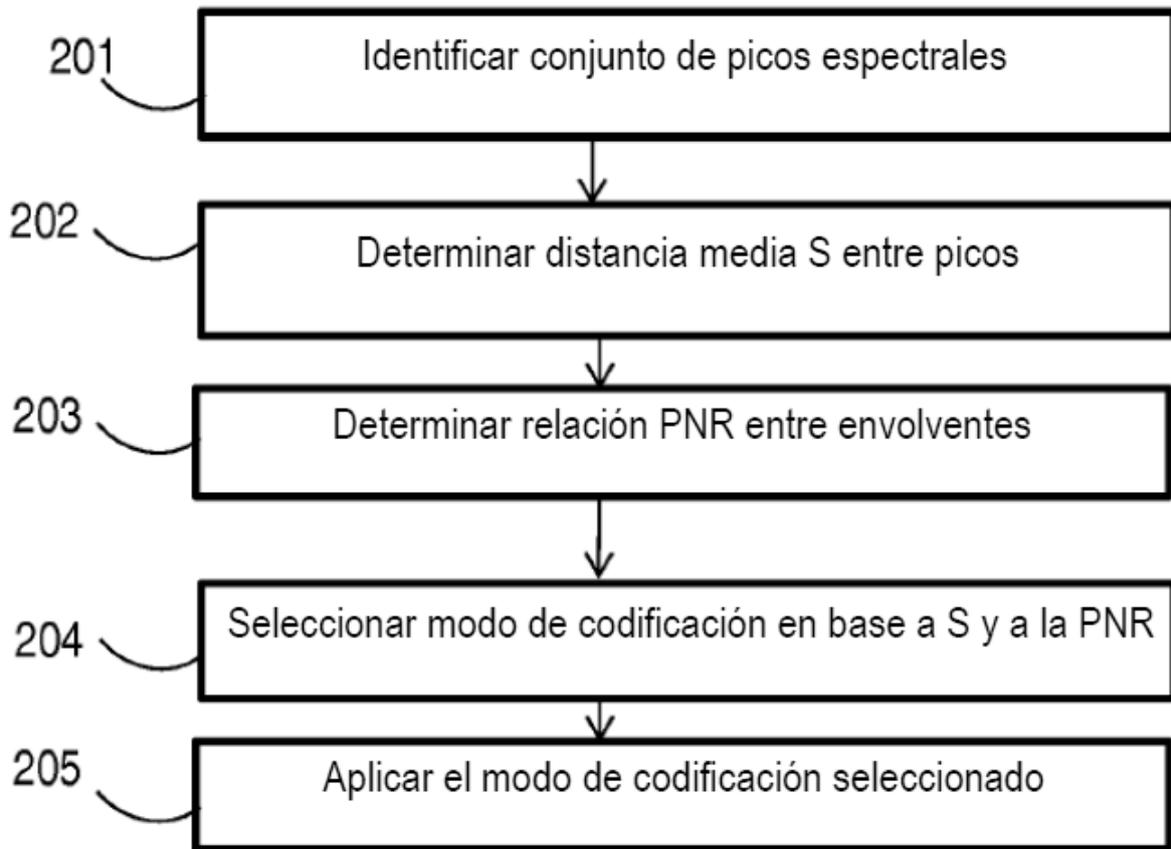


Figura 2

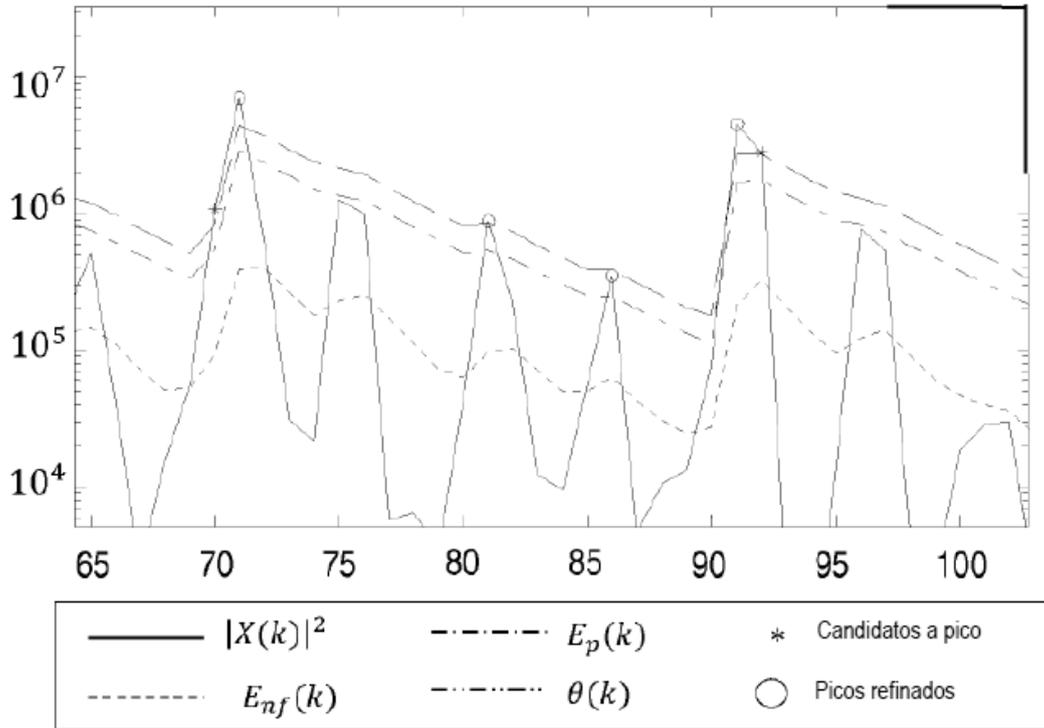


Figura 3a

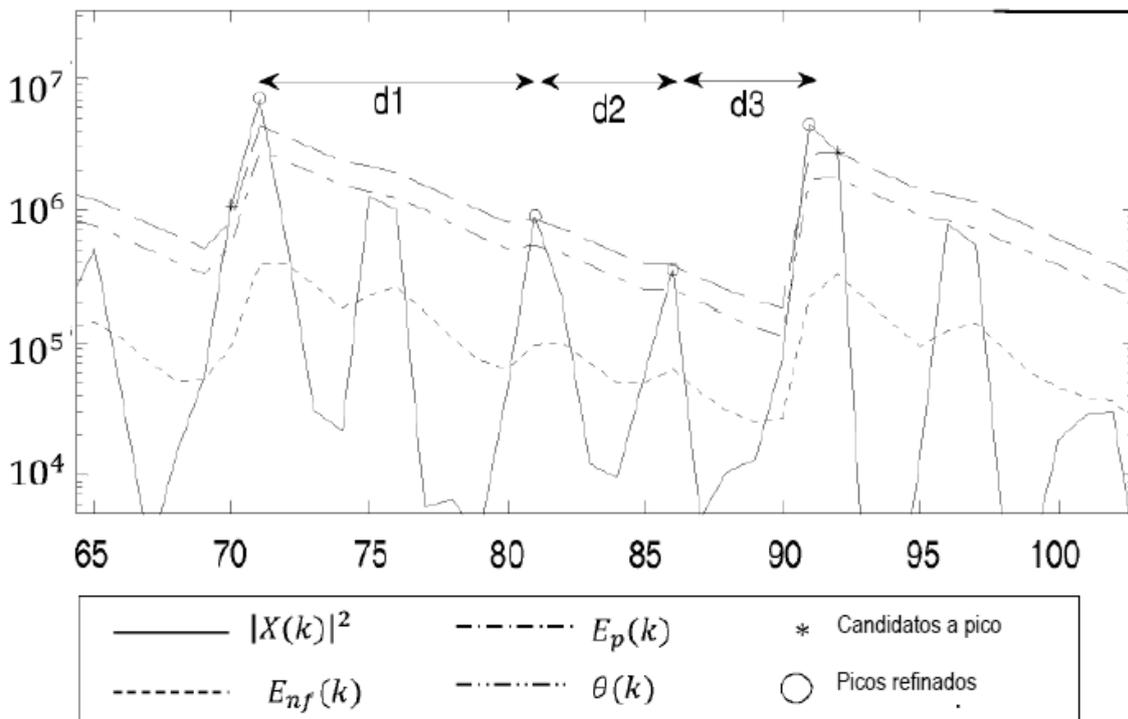


Figura 3b

$$issparse = S > S_{THR}$$
$$isclean = PNR > PNR_{THR}$$

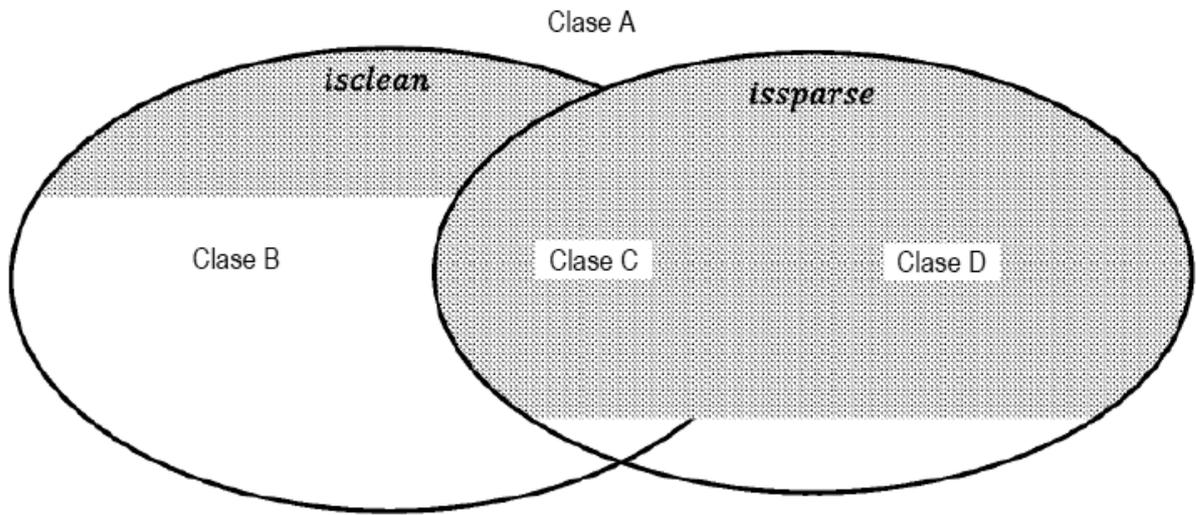


Figura 4

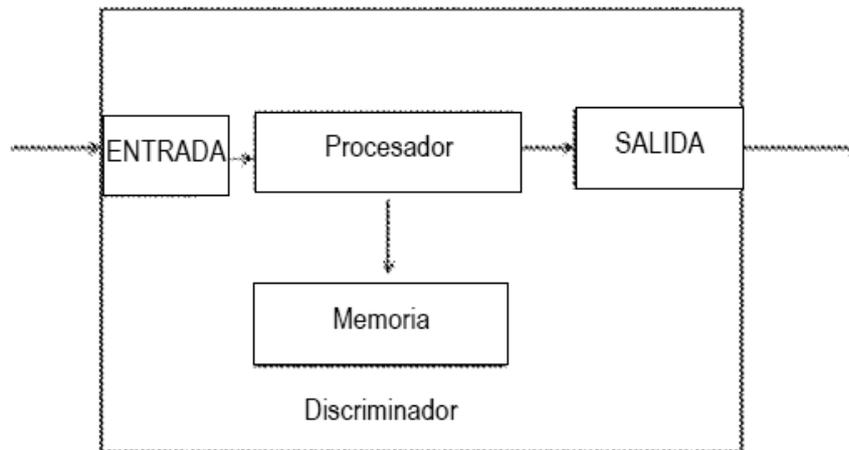


Figura 5d

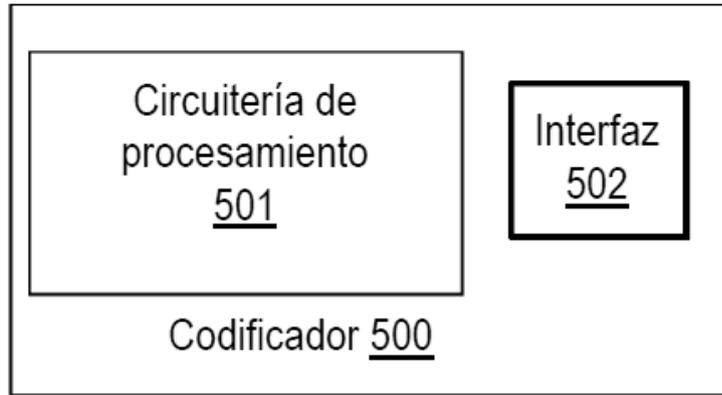


Figura 5a

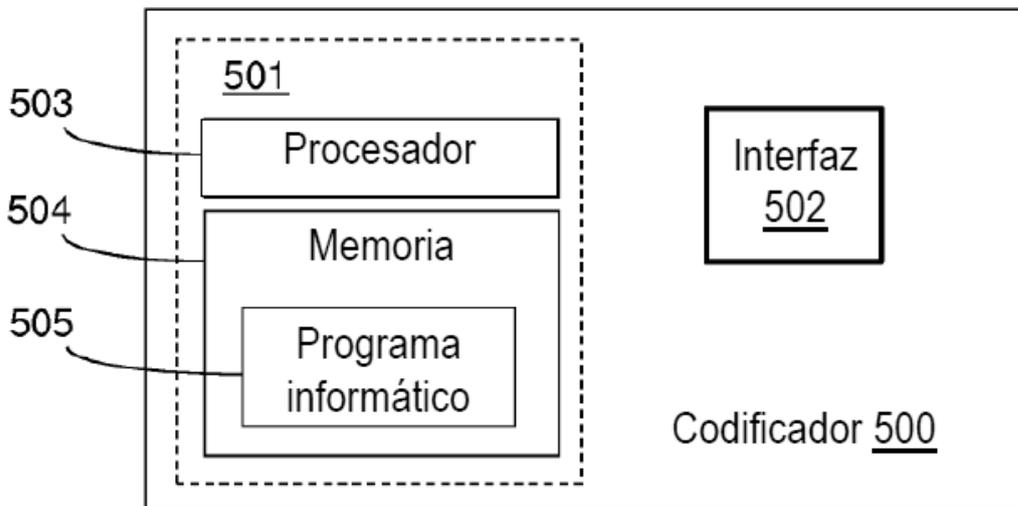


Figura 5b

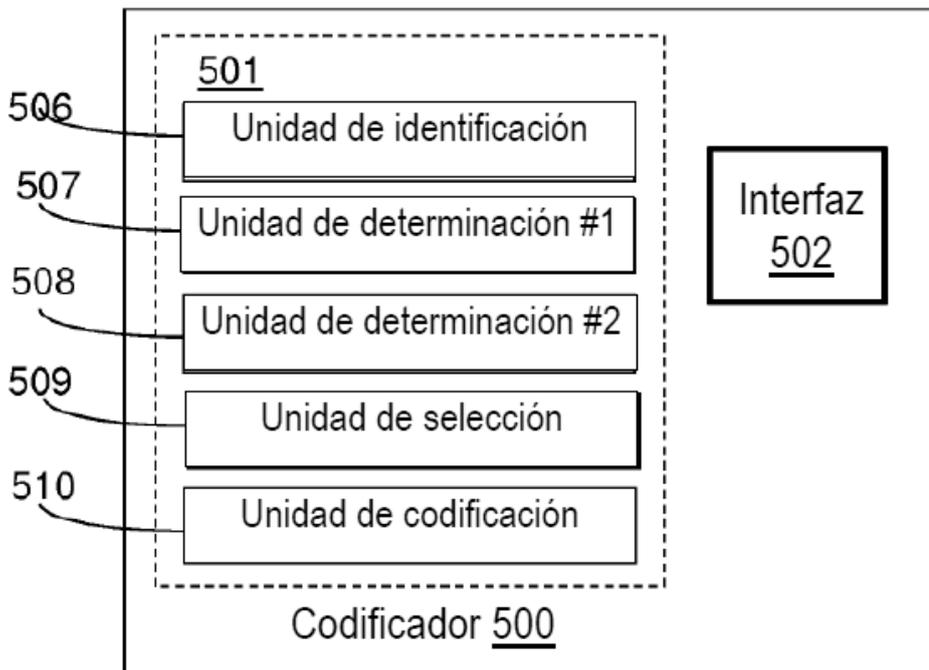


Figura 5c

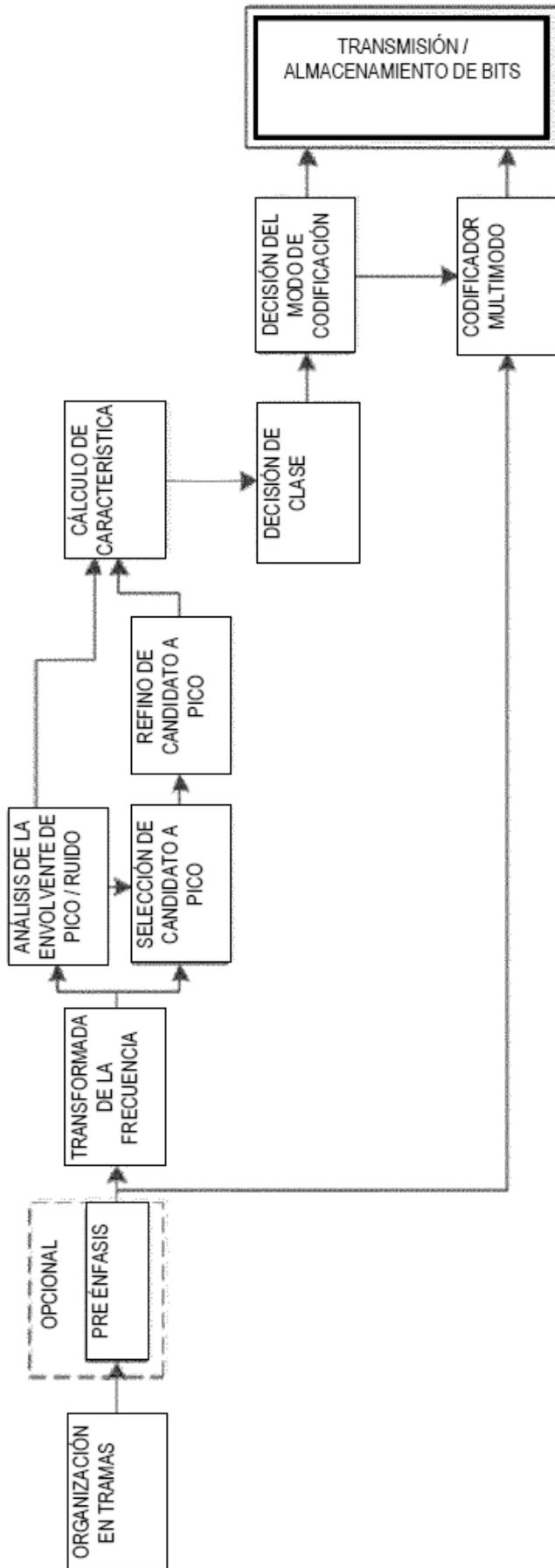


Figura 6