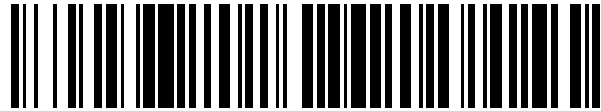


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 576 232**

51 Int. Cl.:

G10L 19/02 (2013.01)

G10L 25/81 (2013.01)

G10L 19/12 (2013.01)

G10L 19/22 (2013.01)

G10L 19/20 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **21.12.2012 E 12810018 (7)**

97 Fecha y número de publicación de la concesión europea: **13.04.2016 EP 2803068**

54 Título: **Clasificación de señal de múltiples modos de codificación**

30 Prioridad:

13.01.2012 US 201261586374 P

20.12.2012 US 201213722669

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

06.07.2016

73 Titular/es:

QUALCOMM INCORPORATED (100.0%)
5775 Morehouse Drive
San Diego, CA 92121, US

72 Inventor/es:

ATTI, VENKATRAMAN SRINIVASA y
DUNI, ETHAN ROBERT

74 Agente/Representante:

FORTEA LAGUNA, Juan José

ES 2 576 232 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

Clasificación de señal de múltiples modos de codificación

5 ANTECEDENTES

10 La transmisión de voz (también denominada como vocal) y música mediante técnicas digitales se ha generalizado e incorporado en una amplia gama de dispositivos, incluyendo, dispositivos de comunicación inalámbrica, asistentes digitales personales (PDA), ordenadores portátiles, ordenadores de mesa, teléfonos móviles y/o radiotelefonos satelitales, y similares. Un campo ejemplar son las comunicaciones inalámbricas. El campo de las comunicaciones inalámbricas tiene muchas aplicaciones, incluyendo, por ejemplo, teléfonos sin cables, paginación, bucles locales inalámbricos, telefonía inalámbrica, tal como sistemas celulares y teléfonos PCS, telefonía móvil por protocolo de Internet (IP), y sistemas de comunicación por satélite.

15 En redes de telecomunicaciones, la información se transfiere de forma codificada entre un dispositivo de comunicación de transmisión y un dispositivo de comunicación de recepción. El dispositivo de comunicación de transmisión codifica información original, tal como señales de voz y/o señales de música, en información codificada y la envía al dispositivo de comunicación de recepción. El dispositivo de comunicación de recepción descodifica la información codificada recibida para recrear la información original. La codificación y la descodificación se realiza usando códecs. La codificación de señales de voz y/o señales de música se realiza en un códec localizado en el dispositivo de comunicación de transmisión, y la descodificación se realiza en un códec situado en el dispositivo de comunicación de recepción.

20 En los códecs modernos, se incluyen múltiples modos de codificación para manejar diferentes tipos de fuentes de entrada, tales como voz, música y contenido mixto. Para un rendimiento óptico, debe seleccionarse y usarse el modo de codificación óptimo para cada trama de la señal de entrada. Es necesaria una clasificación precisa para selección los esquemas de codificación más eficientes y conseguir la menor tasa de datos.

25 Esta clasificación puede realizarse en bucle abierto para evitar complejidad. En este caso, el clasificador de modo óptimo debe tener en cuenta características principales de los diversos modos de codificación. Algunos modos (tal como modos de codificación de voz como predicción lineal excitada por código algebraico (ACELP)) contienen un libro de códigos adaptativo (ACB) que aprovecha la correlación entre las tramas pasadas y actuales. Algunos otros modos (tales como modos de codificación de transformada discreta del coseno modificada (MDCT) par música/audio) pueden no contener tal característica. Por lo tanto, es importante asegurar que las tramas de entrada que tienen una alta correlación con la trama anterior se clasifican en el modo que tiene ACB o que incluye otras técnicas de modelación de correlación entre tramas.

30 Las soluciones anteriores han usado decisiones de modo de bucle cerrado (por ejemplo, AMR- WB+, USAC) o diversos tipos de decisiones de bucle abierto (por ejemplo, AMR-WB+, EVRC-WB), pero estas soluciones son complejas o sus rendimientos son propensos a errores.

RESUMEN

35 Se proporciona una clasificación de audio mejorada para codificar aplicaciones. Se realiza una clasificación inicial, seguida de una clasificación más precisa, para producir clasificaciones de voz y clasificaciones de música con mayor precisión y menos complejidad que las disponibles previamente.

40 El audio se clasifica como voz o música en una base porción por porción (por ejemplo, trama por trama). Si la trama se clasifica como música por la clasificación inicial, esa trama se somete a una segunda clasificación más precisa para confirmar que la trama es música y no voz (por ejemplo, voz que es tonal y/o está estructurada que puede no haberse clasificado como voz por la clasificación inicial).

45 Dependiendo de la implementación, pueden usarse uno o más parámetros en la clasificación más precisa. Los parámetros ejemplares incluyen sonorización, correlación modificada, actividad significativa y ganancia de paso a largo plazo. Estos parámetros son únicamente ejemplos, y no pretenden ser limitantes.

50 Este resumen se proporciona para introducir una selección de conceptos de forma simplificada que se describen adicionalmente a continuación en la descripción detallada. Este resumen no pretende identificar características clave o características esenciales de la materia objeto reivindicada, ni pretende usarse para limitar el alcance de la materia objeto reivindicada.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

55 El resumen anterior, así como la siguiente descripción detallada de las realizaciones ilustrativas, se entiende mejor cuando se lee junto con los dibujos adjuntos. Para el fin de ilustrar las realizaciones, se muestran en los dibujos construcciones ejemplares de las realizaciones; sin embargo, las realizaciones no se limitan a los procedimientos e

instrumentos específicos divulgados. En los dibujos:

La figura 1A es un diagrama de bloques que ilustra un sistema ejemplar en el que un dispositivo origen transmite un flujo de bits codificado a un dispositivo de recepción;

la figura 1B es un diagrama de bloques de dos dispositivos que pueden usarse como se describe en el presente documento;

la figura 2 es un diagrama de bloques de una implementación de un sistema de clasificación de señal de múltiples modos de codificación y codificador;

la figura 3 es un flujo operativo de una implementación de un procedimiento para clasificar audio;

la figura 4 es un diagrama de una estación móvil ejemplar; y

la figura 5 muestra un entorno informático ejemplar.

DESCRIPCIÓN DETALLADA

Las realizaciones divulgadas presentan técnicas de clasificación para una diversidad de modos de codificación en entornos con diversos tipos de audio, tal como voz y música. Los tipos de tramas de audio pueden identificarse de forma fiable y precisa para su codificación de la manera más eficiente. Aunque los ejemplos y la descripción en el presente documento se refieren a tramas de audio, se contemplan más generalmente porciones de señales de audio y pueden usarse de acuerdo con las implementaciones descritas en el presente documento.

La figura 1A es un diagrama de bloques que ilustra un sistema ejemplar 10 en el que un dispositivo origen 12a transmite un flujo de bits codificado a través de un enlace de comunicación 15 a un dispositivo de recepción 14a. El flujo de bits puede representarse como uno o más paquetes. El dispositivo origen 12a y el dispositivo de recepción 14a pueden ser ambos dispositivos digitales. En particular, el dispositivo origen 12a puede codificar datos acordes con el estándar 3GPP2 EVRC-B, o estándares similares que aprovechan la codificación de datos en paquetes para la compresión de voz. Uno o ambos dispositivos 12a, 14a del sistema 10 pueden implementar selecciones de modos de codificación (basados en diferentes modelos de codificación) y tasas de codificación para compresión de audio (por ejemplo, voz y/o música), como se describe en más detalle a continuación, para mejorar el proceso de codificación de audio. Una estación móvil ejemplar, que puede comprender un dispositivo origen o un dispositivo de recepción, se describe con respecto a la figura 4.

El enlace de comunicación 15 puede comprender un enlace inalámbrico, una línea de transmisión física, fibra óptica, una red basada en paquetes tal como una red de área local, una red de área extensa, o una red global tal como Internet, una red telefónica conmutada pública (PSTN), o cualquier otro enlace de comunicación capaz de transferir datos. El enlace de comunicación 15 puede acoplarse a un medio de almacenamiento. Por lo tanto, el enlace de comunicación 15 representa cualquier medio de comunicación adecuado, o posiblemente una colección de diferentes redes y enlaces, para transmitir datos de voz comprimidos desde el dispositivo origen 12a al dispositivo de recepción 14a.

El dispositivo origen 12a puede incluir uno o más micrófonos 16 que capturan el sonido. El sonido continuo se envía al digitalizador 18. El digitalizador 18 muestrea el sonido a intervalos discretos y cuantifica (digitaliza) la voz. La voz digitalizada puede almacenarse en una memoria 20 y/o puede enviarse a un codificador 22, donde las muestras de voz digitalizadas pueden codificarse, a menudo durante una trama de 20 ms.

Más particularmente, el codificador divide la señal de voz entrante en bloques de tiempo, o tramas o porciones de análisis. La duración de cada segmento en tiempo (o trama) se selecciona típicamente para que sea suficientemente corta como para que pueda esperarse que la envolvente espectral de la señal permanezca relativamente estacionaria. Por ejemplo, una longitud de trama típica es de veinte milisegundos (20 ms), que corresponde a 160 muestras a una frecuencia de muestreo típica de ocho kilohertzios (8 kHz), aunque puede usarse cualquier longitud de trama o frecuencia de muestreo que se considere adecuada para la aplicación particular.

El proceso de codificación realizado en el codificador 22 produce uno o más paquetes, para enviar al transmisor 24, que pueden transmitirse por un enlace de comunicación 15 al dispositivo de recepción 14a. Por ejemplo, el codificador analiza la trama entrante para extraer ciertos parámetros relevantes, y después cuantifica los parámetros en representación binaria, es decir, a un conjunto de bits o un paquete de datos binario. Los paquetes de datos se transmiten por el canal de comunicación (es decir, una conexión de red alámbrica y/o inalámbrica) a un receptor y un descodificador. El descodificador procesa los paquetes de datos, los descuantifica para producir los parámetros, y sintetiza de nuevo las tramas de audio usando los parámetros no cuantificados.

El codificador 22 puede incluir, por ejemplo, diverso hardware, software o firmware, o uno o más procesadores de señales digitales (DSP) que ejecutan módulos de software programables para controlar las técnicas de codificación, como se describen en el presente documento. Los circuitos de memoria y lógica asociados pueden proporcionarse

para soportar el DSP en el control de las técnicas de codificación. Como se describirá, el codificador 22 puede comportarse más enérgicamente si los modos de codificación y las tasas pueden cambiarse antes y/o durante la codificación dependiendo de si se ha determinado y se codifica una trama de voz o una trama de música.

5 El dispositivo de recepción 14a puede tomar la forma de cualquier dispositivo de audio digital capaz de recibir y descodificar datos de audio. Por ejemplo, el dispositivo de recepción 14a puede incluir un receptor 26 para recibir paquetes del transmisor 24, por ejemplo, a través de enlaces intermedios, enrutadores, otro equipo de red, y similares. El dispositivo de recepción 14a también puede incluir un descodificador 28 para descodificar el uno o más paquetes, y uno o más altavoces 30 para permitir que un usuario oiga el audio reconstruido después de la
10 descodificación de los paquetes por el decodificador de voz 28.

En algunos casos, un dispositivo origen 12b y un dispositivo de recepción 14b pueden incluir cada uno un codificador/descodificador de voz (código) 32 como se muestra en la figura 1B, para codificar y descodificar datos de audio digital. En particular, tanto el dispositivo origen 12b como el dispositivo de recepción 14b pueden incluir
15 transmisores y receptores, así como una memoria y altavoces. Muchas de las técnicas de codificación contempladas en el presente documento se describen en el contexto de un dispositivo de audio digital que incluye un codificador para comprimir voz y/o música.

Sin embargo, se entenderá que el codificador puede formar parte de un códec 32. En este caso, el códec puede implementarse en hardware, software, firmware, un DSP, un microprocesador, un procesador de propósito general, un circuito integrado de aplicación específica (ASIC), una matriz de puertas programables en campo (FPGA), componentes de hardware discretos o diversas combinaciones de los mismos. Además, se entiende por los expertos en la técnica que los codificadores pueden implementarse con un DSP, un ASIC, una lógica de compuertas discretas, firmware, o cualquier módulo de software programable convencional y un microprocesador. El módulo de
20 software puede residir en una memoria RAM, una memoria flash, registros, o cualquier otra forma de medio de almacenamiento de escritura conocido en la técnica. Como alternativa, cualquier procesador, controlador o máquina de estados convencional puede estar sustituido por el microprocesador. Se describe un dispositivo informático ejemplar con respecto a la figura 5.

30 La figura 2 es un diagrama de bloques de una implementación de un sistema de clasificación de señal de múltiples modos de codificación y codificador 200. En una implementación, el sistema 200 puede usarse con un dispositivo, tal como un dispositivo origen o un dispositivo de recepción descritos con respecto a las figuras 1A y 1B. Por ejemplo, el sistema 200 puede operar junto con el codificador 22 del dispositivo origen 12a.

35 El sistema de clasificación de señal de múltiples modos de codificación y codificador 200 comprende un clasificador inicial 210 (también denominado como un primer clasificador) y un clasificador de precisión 220 (también denominado como un segundo clasificador). El sistema 200 también comprende un conmutador de selección de clasificador de precisión 230 que puede seleccionarse (por ejemplo, por un usuario) para habilitar o deshabilitar el clasificador de precisión 220 y su funcionalidad de clasificación más precisa asociada.

40 Se incluyen diversos tipos de codificadores en el sistema 200, tales como codificadores de voz y un codificador de música. En una implementación, un primer modo de codificación, denominado como "modo de codificación 1" 240 (tal como un codificador de tipo predictivo lineal excitado por código (CELP), o un modo de codificación de voz, por ejemplo) puede proporcionarse y usarse en respuesta a la clasificación por el clasificador inicial 210. Un segundo modo de codificación, denominado como "modo de codificación 2" 260 (tal como un codificador híbrido CELP/transformada, o un segundo modo de codificación de voz, por ejemplo) puede proporcionarse y usarse en
45 respuesta a la clasificación por el clasificador de precisión 220.

Un tercer modo de codificación, denominado como "modo de codificación 3" 250 (tal como un codificador por transformada, o un modo de codificación de música, por ejemplo) puede proporcionarse y usarse en respuesta a la clasificación por el clasificador inicial 210 y/o el clasificador de precisión 220. Estos tipos de modos de codificación y codificadores se conocen bien, y se omiten descripciones adicionales para mayor brevedad. Los modos de codificación y codificadores ejemplares descritos para los modos de codificación 1, 2 y 3 son únicamente ejemplos y no pretenden ser limitantes. Puede usarse cualquier modo o modos de codificación y/o codificador o codificadores de voz y cualquier modo o modos de codificación y/o codificador o codificadores de música apropiados.
50

La figura 3 es un flujo operativo de una implementación de un procedimiento 300 para clasificar audio. En 310, el clasificador inicial 210 recibe una trama de audio de entrada (u otra porción de una señal de audio para clasificar la porción de la señal de audio como una señal de audio tipo voz o una señal de audio tipo música) y la clasifica como voz o música en 320. El clasificador inicial 210 puede ser cualquiera clasificador que clasifica una trama o porción de audio como voz o música.
60

En algunas implementaciones, el clasificador inicial 210 puede comprender más de un clasificador (mostrado en 320 como "clasificador 1" y "clasificador 2", aunque puede usarse cualquier número de clasificadores dependiendo de la implementación). Por ejemplo, el clasificador inicial puede comprender un clasificador que se sesga completamente hacia voz, y otro clasificador diferente, tal como un clasificador que está más sesgado hacia música. Estos dos
65

clasificadores pueden operar en la trama de entrada secuencialmente o, a veces, simultáneamente (dependiendo de la implementación) en 320, estando sus resultados combinados para formar un resultado que se envía a 330 o 340.

Hay una pequeña probabilidad de que la voz se detecte como música por el clasificador inicial 210. Como tal, algunas tramas de voz pueden clasificarse inicialmente como música. Por ejemplo, la voz en presencia de música de fondo de muy bajo nivel o una voz cantante, que son representativos de la voz, pueden no clasificarse como voz por el clasificador inicial 210. En su lugar, el clasificador inicial 210 puede clasificar dichas señales como música. La presencia de otro ruido de fondo, tal como el claxon de un vehículo suena en el ruido de la calle o el teléfono suena en una oficina típica, por ejemplo, también puede contribuir a aumentar una clasificación errónea de la voz como música.

Si la trama se determina en 320 como una trama de voz por el clasificador inicial 210, entonces la trama se proporciona al modo de codificación 1 240 (por ejemplo, un codificador de tipo CELP) para su codificación. En algunas implementaciones, puede usarse cualquier codificador de tipo CELP conocido.

Por otro lado, si la trama se determina en 320 como una trama de música por el clasificador inicial 210, entonces se determina en 340 si está habilitada una clasificación más precisa (por ejemplo, por el usuario que tiene habilitada previamente la característica, a través de un conmutador de selección "on" y "off" en el dispositivo correspondiente a "habilitado" y "no habilitado", respectivamente). Esta clasificación más precisa es una segunda ronda de clasificación que refuerza la decisión de la primera clasificación. En una implementación, la clasificación más precisa para procesar datos de audio puede habilitarse selectivamente por un usuario.

Si la clasificación más precisa no está habilitada como se determina en 340, entonces la trama se proporciona al modo de codificación 3 250 (por ejemplo, un codificador por transformada) para su codificación como una trama de música en 350. Sin embargo, si la clasificación más precisa se habilita como se determina en 340, entonces la trama se proporciona al clasificador de precisión 220 en 360 para una clasificación más precisa adicional. La clasificación más precisa se usa para distinguir adicionalmente una trama de voz de una trama de música.

En una implementación, la clasificación más precisa se usa para confirmar que la trama es tipo ruido de banda ancha, que es una característica de ciertos tipos de música, en lugar de características tonales y/o casi estacionarias de voz. Si la clasificación más precisa en 360 da como resultado la trama que se identifica como una trama de música, entonces la trama se envía al modo de codificación 3 para su codificación como una trama de música en 350.

Si la clasificación más precisa en 360 da como resultado la trama que se identifica como una trama de voz, entonces la trama se envía al modo de codificación 2 260 para su codificación como una trama de voz en 370. Como se ha señalado anteriormente, en una implementación, el modo de codificación 2 260 puede ser un codificador híbrido CELP/transformada, que puede usarse para codificar tramas de voz tonales y/o estructuradas. En una implementación alternativa, el modo de codificación 2 260 en 370 puede ser un codificador de tipo CELP tal como el modo de codificación 1 usado en 330.

En una implementación, la clasificación más precisa realizada en 360 (por ejemplo, el clasificador de precisión 220) puede comparar diversas características o funciones de la trama con uno o más umbrales para determinar si la trama es una trama de voz o una trama de música.

En algunas implementaciones, la sonorización de la trama puede compararse con un primer umbral THR1. Si la sonorización de la trama es mayor que THR1, entonces se determina que la trama es una trama de voz. Un valor ejemplar para THR1 es 0,99, aunque puede usarse cualquier valor dependiendo de la implementación. La sonorización varía de 0 (correspondiente a correlación nula con una trama de voz) a 1 (correspondiente a una alta correlación con una trama de voz).

En algunas implementaciones, la correlación de señal ponderada puede compararse con un segundo umbral THR2. Si la correlación de señal ponderada es mayor que THR2, entonces se determina que la trama es una trama de voz. Un valor ejemplar para THR2 es 0,87, aunque puede usarse cualquier valor dependiendo de la implementación. La correlación de señal varía de 0 (correspondiente a un ruido aleatorio) a 1 (correspondiente a un sonido altamente estructurado).

En algunas implementaciones, la ganancia de paso a largo plazo puede compararse con un tercer umbral THR3. Si la ganancia de paso a largo plazo es mayor que THR3, entonces se determina que la trama es una trama de voz. Un valor ejemplar para THR3 es 0,5, aunque puede usarse cualquier valor dependiendo de la implementación. La ganancia de paso a largo plazo es la correlación cruzada normalizada entre la excitación pasada y el residuo de predicción actual. La ganancia de paso a largo plazo varía de 0 (que indica que el error en la trama pasada no es adecuado en la representación de la trama actual) a 1 (que indica que el uso del error residual en la trama pasada puede representar completamente la trama actual).

En algunas implementaciones, la tonalidad de la trama actual puede determinarse y compararse con un umbral

THR4. La tonalidad de una señal puede medirse usando una medición de planitud espectral o un pico espectral con respecto una medición de relación media. Si el espectro no contiene ningún pico localizado prominente, entonces la medición de planitud espectral tenderá a estar cerca de 1. Por otro lado, si el espectro muestra una fuerte inclinación con picos localizados, entonces la medición de planitud espectral estará cerca de 0. Si la tonalidad es mayor que THR4, entonces se determina que la trama es una trama de voz. Un valor ejemplar para THR4 es 0,75, aunque puede usarse cualquier valor dependiendo de la implementación.

Adicionalmente, o como alternativa, en algunas implementaciones, puede determinarse si hay alguna actividad de señal. Si no hay ninguna actividad de señal (es decir, la trama es silenciosa), entonces se determina que no hay ninguna señal útil que codificar, y puede codificarse como una trama de voz.

En algunas implementaciones, si no se cumple ninguna de las condiciones para determinar en 360 que la trama es una trama de voz, entonces se determina que la trama es una trama de música.

Las comparaciones y umbrales que se describen en el presente documento no pretenden ser limitantes, ya que puede usarse uno cualquiera o más comparaciones y/o umbrales dependiendo de la implementación. También pueden usarse comparaciones y umbrales adicionales y/o alternativos, dependiendo de la implementación.

Por lo tanto, en una implementación, si una trama se clasifica inicialmente (por el clasificador inicial 210) como voz, se pasa a un codificador CELP. Si la trama se clasifica como música, sin embargo, entonces se verifica si se permite o no una clasificación más precisa. La clasificación más precisa puede habilitarse usando un control de usuario externo. Si la clasificación más precisa no está habilitada, entonces la trama que inicialmente se clasifica como música se enruta a un codificador por transformada para su codificador. Si la clasificación más precisa está habilitada, entonces se usa una combinación lógica de ciertos criterios (por ejemplo, voz, correlación modificada, actividad de señal, ganancia de paso a largo plazo, etc.) para seleccionar entre un codificador por transformada y un híbrido de codificador CELP/transformada. El THR1, THR2, THR3 y el THR4 son valores de umbral que pueden determinarse experimentalmente y depender de las tasas de muestreo y tipos de señal, por ejemplo.

En una implementación, se impide que las señales fuertemente tonales se codifiquen en modo MDCT (que carece de libro de códigos adaptativo) y, en su lugar, se proporcionan a modos de codificación predictiva lineal (LPC) que utilizan un libro de códigos adaptativo.

Los componentes de los codificadores y los clasificadores que se describen en el presente documento pueden implementarse como hardware electrónico, como software informático, o combinaciones de ambos. Estos componentes se describen en el presente documento en cuanto a su funcionalidad. Si la funcionalidad se implementa como hardware o software dependerá de la aplicación particular y de las limitaciones de diseño impuestas sobre todo el sistema. Los expertos en la técnica reconocerán la intercambiabilidad del hardware y el software en estas circunstancias, y cuál es la mejor forma de implementar la funcionalidad descrita para cada aplicación particular.

Como se usa en el presente documento, el término "determinar" (y variantes gramaticales del mismo) se usa en un sentido extremadamente amplio. El término "determinación" incluye una amplia variedad de acciones y, por lo tanto, "determinación" puede incluir el cálculo, la computación, el procesamiento, la derivación, la investigación, la consulta (por ejemplo, la consulta en una tabla, la consulta en una base de datos o en otra estructura de datos), la verificación y similares. Además, "determinación" puede incluir la recepción (por ejemplo, la recepción de información), el acceso, (por ejemplo, el acceso a datos de una memoria) y similares. Así mismo, "determinación" puede incluir la resolución, la selección, la elección, el establecimiento y similares.

La expresión "procesamiento de señal" (y variantes gramaticales de la misma) puede referirse al procesamiento e interpretación de señales. Las señales de interés pueden incluir sonido, imágenes, y muchos otros. El procesamiento de dichas señales puede incluir almacenamiento y reconstrucción, separación de información del ruido, compresión, y extracción de características. El término "procesamiento digital de señales" puede referirse al estudio de señales en una representación digital y a los métodos de procesamiento de estas señales. El procesamiento digital de señales es un elemento de muchas tecnologías de comunicaciones tales como estaciones móviles, estaciones no móviles e Internet. Los algoritmos que se utilizan para el procesamiento digital de señales pueden realizarse usando ordenadores especializados, que pueden hacer uso de microprocesadores especializados denominados procesadores de señales digitales (a menudo abreviados como DSP).

A menos que se indique otra cosa, cualquier divulgación de una operación de un aparato que tiene una característica particular también pretende expresamente divulgar un procedimiento que tenga una característica análoga (y viceversa), y cualquier divulgación de una operación de un aparato de acuerdo con una configuración particular también pretende expresamente divulgar un procedimiento de acuerdo con una configuración análoga (y viceversa).

La figura 4 muestra un diagrama de bloques de un diseño de una estación móvil ejemplar 400 en un sistema de comunicación inalámbrica. La estación móvil 400 puede ser un teléfono móvil, un terminal, un dispositivo de mano,

un PDA, un módem inalámbrico, un teléfono sin cables, etc. El sistema de comunicación inalámbrica puede ser un sistema CDMA, un sistema GSM, etc.

5 La estación móvil 400 es capaz de proporcionar una comunicación bidireccional a través de una ruta de recepción y una ruta de transmisión. En la ruta de recepción, las señales transmitidas por las estaciones base se reciben por una antena 412 y se proporcionan a un receptor (RCVR) 414. El receptor 414 acondiciona y digitaliza la señal recibida y proporciona muestras a una sección digital 420 para su procesamiento adicional. En la ruta de transmisión, un transmisor (TMTR) 416 recibe datos que se van a transmitir desde la sección digital 420, procesa y acondiciona los datos, y genera una señal modulada, que se transmite a través de la antena 412 a las estaciones base. El receptor 10 414 y el transmisor 416 pueden ser parte de un transceptor que puede soportar CDMA, GSM, etc.

La sección digital 420 incluye diversas unidades de procesamiento, interfaz y memoria, tales como, por ejemplo, un procesador de módem 422, un ordenador con conjunto reducido de instrucciones/procesador de señales digitales (RISC/DSP) 424, un controlador/procesador 426, una memoria interna 428, un codificador de audio generalizado 15 432, un descodificador de audio generalizado 434, un procesador de gráficos/pantalla 436, y una interfaz de externo (EBI) 438. El procesador de módem 422 puede realizar un procesamiento para la transmisión y recepción de datos, por ejemplo, codificación, modulación, desmodulación y descodificación. El RISC/DSP 424 puede realizar un procesamiento general y especializado para el dispositivo inalámbrico 400. El controlador/procesador 426 puede dirigir la operación y diversas unidades de procesamiento e interfaz dentro de la sección digital 420. La memoria 20 interna 428 puede almacenar datos y/o instrucciones para diversas unidades dentro de la sección digital 420.

El codificador de audio generalizado 432 puede realizar una codificación de señales de entrada de una fuente de audio 442, un micrófono 443, etc. El descodificador de audio generalizado 434 puede realizar una descodificación de datos de audio codificados y puede proporcionar señales de salida a un altavoz/auriculares 444. El procesador de 25 gráficos/pantalla 436 puede realizar un procesamiento de gráficos, vídeos, imágenes y textos, que pueden presentarse a una unidad de pantalla 446. La EBI 438 puede facilitar la transferencia de datos entre la sección digital 420 y una memoria principal 448.

La sección digital 420 puede implementarse con uno o más procesadores, DSP, microprocesadores, RISC, etc. La sección digital 420 también puede fabricarse en uno o más circuitos integrados de aplicaciones específicas (ASIC) 30 y/o alguno otro tipo de circuitos integrados (IC).

La figura 5 muestra un entorno informático ejemplar en el que pueden implementarse implementaciones y aspectos ejemplares. El entorno del sistema informático es únicamente un ejemplo de un entorno informático adecuado y no pretende sugerir ninguna limitación en cuanto al alcance del uso o funcionalidad. 35

Pueden usarse instrucciones ejecutables por ordenador, tales como módulos de programa, que se ejecutan por un ordenador. Generalmente, los módulos de programa incluyen rutinas, programas, objetos, componentes, estructuras de datos, etc. que realizan tareas particulares o implementan tipos de datos abstractos particulares. Pueden usarse 40 entornos informáticos distribuidos donde las tareas se realizan mediante dispositivos de procesamiento remotos que están vinculados a través de una red de comunicaciones u otro medio de transmisión de datos. En un entorno informático distribuido, los módulos de programa y otros datos pueden localizarse tanto en medios de almacenamiento informáticos locales como remotos, incluyendo dispositivos de almacenamiento de memoria.

45 Con referencia a la figura 5, un sistema ejemplar para implementar los aspectos descritos en el presente documento incluye un dispositivo informático, tal como el dispositivo informático 500. En su configuración más básica, el dispositivo informático 500 incluye típicamente al menos una unidad de procesamiento 502 y una memoria 504. Dependiendo de la configuración exacta y el tipo de dispositivo informático, la memoria 504 puede ser volátil (tal como una memoria de acceso aleatorio (RAM)), no volátil (tal como una memoria de sólo lectura (ROM)), una memoria flash, etc.), o alguna combinación de las dos. Esta configuración más básica se ilustra en la figura 5 por la 50 línea discontinua 506.

El dispositivo informático 500 puede tener características y/o una funcionalidad adicionales. Por ejemplo, el dispositivo informático 500 puede incluir un almacenamiento adicional (extraíble y/o no extraíble), incluyendo, pero 55 sin limitación, discos magnéticos y ópticos o una cinta. Dicho almacenamiento adicional se ilustra en la figura 5 por el almacenamiento extraíble 508 y el almacenamiento no extraíble 510.

El dispositivo informático 500 incluye típicamente una diversidad de medios legibles por ordenador. Los medios legibles por ordenador pueden ser cualquier medio disponible al que se pueda acceder mediante el dispositivo 500 y que incluya tanto medios volátiles como no volátiles, y tanto medios extraíbles como no extraíbles. Los medios de 60 almacenamiento informáticos incluyen medios volátiles y no volátiles, y extraíbles y no extraíbles implementados en cualquier procedimiento o tecnología para el almacenamiento de información, tal como instrucciones legibles por ordenador, estructuras de datos, módulos de programa, u otros datos. La memoria 504, el almacenamiento extraíble 508, y el almacenamiento no extraíble 510 son todos ejemplos de medios de almacenamiento informático. Los 65 medios de almacenamiento de ordenador incluyen, pero sin limitación, memoria RAM, ROM, memoria de sólo lectura eléctricamente programable borrable (EEPROM), memoria flash u otras tecnologías de memoria, CDROM,

discos versátiles digitales (DVD) u otro almacenamiento óptico, cassetes magnéticas, cinta magnética, almacenamiento en disco magnético u otros dispositivos magnéticos de almacenamiento, o cualquier otro medio que pueda usarse para almacenar la información deseada y a la que se pueda acceder por el dispositivo informático 500. Cualquier medio de almacenamiento informático puede ser parte del dispositivo informático 500.

El dispositivo informático 500 puede contener una conexión o conexiones de comunicación 512 que permiten que el dispositivo se comunique con otros dispositivos. El dispositivo informático 500 también puede tener un dispositivo o dispositivos de entrada 514, tales como un teclado, ratón, *pen*, dispositivo de entrada de voz, dispositivo de entrada táctil, etc. También puede incluir un dispositivo o dispositivos de salida 516, tales como una pantalla, altavoces, impresora, etc. Todos estos dispositivos se conocen bien en la técnica y no necesitan analizarse ampliamente aquí.

En general, cualquier dispositivo descrito en el presente documento puede representar diversos tipos de dispositivos, tales como un teléfono inalámbrico o alámbrico, un teléfono móvil, un ordenador portátil, un dispositivo multimedia inalámbrico, una tarjeta para PC de comunicación inalámbrica, un PDA, un módem externo o interno, un dispositivo que comunica a través de un canal inalámbrico o alámbrico, etc. Un dispositivo puede tener diversos nombres, tal como terminal de acceso (AT), unidad de acceso, unidad de abonado, estación móvil, dispositivo móvil, unidad móvil, teléfono móvil, móvil, estación remota, terminal remoto, unidad remota, dispositivo de usuario, equipo de usuario, dispositivo de mano, estación no móvil, dispositivo no móvil, punto terminal, etc. Cualquier dispositivo descrito en el presente documento puede tener una memoria para almacenar instrucciones y datos, así como hardware, software, firmware, o combinaciones de los mismos.

Las técnicas descritas en el presente documento pueden implementarse por diversos medios. Por ejemplo, estas técnicas pueden implementarse en hardware, software o una combinación de ambos. Los expertos en la técnica apreciarán además que los diversos bloques lógicos, módulos, circuitos y etapas de algoritmo ilustrativos descritos en relación con la divulgación del presente documento pueden implementarse como hardware electrónico, software informático o combinaciones de ambos. Para ilustrar claramente esta intercambiabilidad de hardware y software, anteriormente se han descrito diversos componentes, bloques, módulos, circuitos y etapas ilustrativos, generalmente, en lo que respecta a su funcionalidad. Si tal funcionalidad se implementa como hardware o software, dependerá de la aplicación particular y de las limitaciones de diseño impuestas sobre todo el sistema. Los expertos en la técnica pueden implementar la funcionalidad descrita de diferentes maneras para cada aplicación particular, pero no debe interpretarse que tales decisiones de implementación suponen un apartamiento del alcance de la presente divulgación.

Para una implementación de hardware, las unidades de procesamiento usadas para realizar las técnicas pueden implementarse en uno o más ASIC, DSP, dispositivos de procesamiento digital de señales (DSPD), dispositivos lógicos programables (PLD), FPGA, procesadores, controladores, microcontroladores, microprocesadores, dispositivos electrónicos, otras unidades electrónicas diseñadas para realizar las funciones descritas en el presente documento, un ordenador, o una combinación de los mismos.

Por lo tanto, los diversos bloques lógicos, módulos y circuitos ilustrativos descritos en relación con la divulgación del presente documento pueden implementarse o realizarse con un procesador de propósito general, un DSP, un ASIC, una FPGA o con otro dispositivo de lógica programable, lógica de transistor o de puertas discretas, componentes de hardware discretos, o con cualquier combinación de los mismos diseñada para realizar las funciones descritas en el presente documento. Un procesador de propósito general puede ser un microprocesador pero, como alternativa, el procesador puede ser cualquier procesador, controlador, microcontrolador o máquina de estados convencional. Un procesador también puede implementarse como una combinación de dispositivos informáticos, por ejemplo una combinación de un DSP y un microprocesador, una pluralidad de microprocesadores, uno o más microprocesadores junto con un núcleo de DSP o cualquier otra configuración de este tipo.

Para una implementación de firmware y/o software, las técnicas pueden realizarse como instrucciones en un medio legible por ordenador, tal como RAM, ROM, RAM no volátil, ROM programable, EEPROM, memoria flash, disco compacto (CD), dispositivo de almacenamiento de datos magnético u óptico, o similares. Las instrucciones pueden ser ejecutables por uno o más procesadores, y pueden hacer que el procesador o los procesadores realicen ciertos aspectos de la funcionalidad que se describe en el presente documento.

Si se implementan en software, las funciones pueden almacenarse en o transmitirse como una o más instrucciones o código en un medio legible por ordenador. Los medios legibles por ordenador incluyen tanto medios de almacenamiento informáticos como medios de comunicación, incluyendo cualquier medio que facilite la transferencia de un programa informático de un lugar a otro. Los medios de almacenamiento pueden ser cualquier medio disponible al que pueda accederse mediante un ordenador de fin general o de fin especial. A modo de ejemplo, y no de manera limitativa, tales medios legibles por ordenador pueden comprender RAM, ROM, EEPROM, CD-ROM u otro almacenamiento de disco óptico, almacenamiento de disco magnético u otros dispositivos de almacenamiento magnético, o cualquier otro medio que pueda usarse para transportar o almacenar medios de código de programa deseado en forma de instrucciones o estructuras de datos y al que pueda accederse mediante un ordenador de propósito general o de propósito especial, o mediante un procesador de propósito general o de propósito especial. Además, cualquier conexión puede denominarse de manera apropiada medio legible por ordenador. Por ejemplo, si

5 el software se transmite desde un sitio web, un servidor u otra fuente remota usando un cable coaxial, un cable de fibra óptica, un par trenzado, una línea de abonado digital (DSL) o tecnologías inalámbricas tales como infrarrojos, radio y microondas, entonces el cable coaxial, el cable de fibra óptica, el par trenzado, la DSL o las tecnologías inalámbricas tales como infrarrojos, radio y microondas se incluyen en la definición de medio. Los discos, como se usan en el presente documento, incluyen CD, discos de láser, discos ópticos, discos versátiles digitales (DVD), discos flexibles y discos blu-ray, donde los discos normalmente reproducen datos de manera magnética así como de manera óptica con láser. Las combinaciones de lo anterior también deben incluirse dentro del alcance de los medios legibles por ordenador.

10 Un módulo de software puede residir en memoria RAM, memoria flash, memoria ROM, memoria EPROM, memoria EEPROM, registros, un disco duro, un disco extraíble, un CD-ROM o en cualquier otra forma de medio de almacenamiento conocida en la técnica. Un medio de almacenamiento a modo de ejemplo está acoplado al procesador de manera que el procesador pueda leer información de, y escribir información en, el medio de almacenamiento. Como alternativa, el medio de almacenamiento puede ser una parte integrante del procesador. El
15 procesador y el medio de almacenamiento pueden residir en un ASIC. El ASIC puede residir en un terminal de usuario. Como alternativa, el procesador y el medio de almacenamiento pueden residir como componentes discretos en un terminal de usuario.

20 La anterior descripción de la divulgación se proporciona para permitir que cualquier experto en la técnica realice o use la divulgación. Por tanto, la divulgación no pretende limitarse a los ejemplos descritos en el presente documento, sino que se le concede el alcance más amplio compatible con los principios y características novedosas dados a conocer en el presente documento.

25 Aunque las implementaciones ejemplares pueden referirse a utilizar aspectos de la presente materia objeto divulgada en el contexto de uno o más sistemas informáticos autónomos, la materia objeto no está limitada, sino que en su lugar puede implementarse en relación con cualquier entorno informático, tal como una red o un entorno informático distribuido. Aún adicionalmente, pueden implementarse aspectos de la presente materia objeto divulgada en o a través de una pluralidad de chips o dispositivos de procesamiento, y el almacenamiento puede realizarse de forma similar en una pluralidad de dispositivos. Dichos dispositivos pueden incluir PC, servidores de red y
30 dispositivos de mano, por ejemplo.

Aunque la materia objeto se ha descrito en un lenguaje específico con respecto a las características estructurales y/o acciones metodológicas, se entenderá que la materia objeto definida en las reivindicaciones adjuntas no se limita necesariamente a las características específicas o acciones que se han descrito anteriormente. En su lugar, las
35 características y acciones específicas que se han descrito anteriormente se divulgan como formas ejemplares de implementar las reivindicaciones.

REIVINDICACIONES

1. Un procedimiento, que comprende:
 - 5 recibir una porción de una señal de audio en un primer clasificador;
 - clasificar la porción de la señal de audio en el primer clasificador como voz o como música; y
 - 10 procesar la porción de la señal de audio, en el que el procesamiento de la porción de la señal de audio comprende:
 - si la porción se clasifica por el primer clasificador como voz, entonces codificar la voz usando un primer modo de codificación; o
 - 15 si la porción se clasifica por el primer clasificador como música, entonces:
 - proporcionar la porción a un segundo clasificador;
 - clasificar la porción en el segundo clasificador como voz o como música; y
 - 20 codificar la porción de la señal de audio, en el que codificar la porción de la señal de audio comprende:
 - 25 si la porción se clasifica en el segundo clasificador como voz, entonces codificar la porción usando un segundo modo de codificación; o
 - si la porción se clasifica en el segundo clasificador como música, entonces codificar la porción usando un tercer modo de codificación.
- 30 2. El procedimiento de la reivindicación 1, en el que la porción de la señal de audio es una trama.
3. El procedimiento de la reivindicación 1, en el que el primer modo de codificación comprende un primer codificador de voz, el segundo modo de codificación comprende un segundo codificador de voz, y el tercer modo de codificación comprende un codificador de música.
- 35 4. El procedimiento de la reivindicación 3, en el que el primer codificador de voz es un codificador de tipo predictivo lineal excitado por código (CELP), el segundo codificador de voz es un codificador híbrido CELP/transformada, y el codificador de música es un codificador por transformada.
- 40 5. El procedimiento de la reivindicación 1, que comprende adicionalmente determinar si el segundo clasificador está habilitado antes de proporcionar la porción al segundo clasificador, y si el segundo clasificador no está habilitado, entonces codificar la porción con el tercer modo de codificación.
- 45 6. El procedimiento de la reivindicación 1, en el que clasificar la porción en el segundo clasificador como voz o como música comprende comparar una pluralidad de características de la porción con uno o más umbrales para clasificar si la porción tiene características de música o características de voz.
- 50 7. El procedimiento de la reivindicación 6, en el que las características de música comprenden características tipo ruido de banda ancha de música, y las características de voz comprenden al menos una de las características tonales de voz o características casi estacionarias de voz.
8. El procedimiento de la reivindicación 1, en el que la clasificación de la porción en el segundo clasificador como voz o como música comprende al menos uno de comparar la sonorización de la porción con un primer umbral, comparar la correlación modificada con un segundo umbral, o comparar la ganancia de paso a largo plazo con un tercer umbral.
- 55 9. El procedimiento de la reivindicación 8, en el que la sonorización varía de 0, correspondiente a una correlación nula con voz, a 1, correspondiente a una alta correlación con voz; en el que la correlación modificada varía de 0, correspondiente a un ruido aleatorio, a 1, correspondiente a un sonido altamente estructurado; en el que la ganancia de paso a largo plazo es la correlación cruzada normalizada entre la excitación pasada y el residuo de predicción actual; y en el que la ganancia de paso a largo plazo varía de 0, que indica que el error en la porción pasada no es adecuado en la representación de la porción actual, a 1, que indica que el uso del error residual en la porción pasada puede representar completamente la porción actual.
- 60 10. El procedimiento de la reivindicación 1, en el que la clasificación de la porción en el segundo clasificador
- 65

como voz o como música comprende determinar si hay alguna actividad de señal en la porción, y si no hay ninguna actividad de señal, determinar entonces que no hay ninguna señal útil que codificar, y codificar la porción como voz.

- 5 11. Un aparato que comprende:
- medios para recibir una porción de una señal de audio en un primer clasificador;
medios para clasificar la porción de la señal de audio en el primer clasificador como voz o como música;
- 10 medios para codificar la voz usando un primer modo de codificación si la porción se clasifica por el primer clasificador como voz, o clasificar la porción en el segundo clasificador como voz o como música cuando la porción se clasifica por el primer clasificador como música; y
- 15 medios para codificar la porción usando un segundo modo de codificación cuando la porción se clasifica en el segundo clasificador como voz, o codificar la porción usando un tercer modo de codificación cuando la porción se clasifica en el segundo clasificador como música.
12. El aparato de la reivindicación 11, que comprende adicionalmente medios para determinar si el segundo clasificador está habilitado antes de proporcionar la porción al segundo clasificador, y si el segundo clasificador no está habilitado, codificar entonces la porción con el tercer modo de codificación.
- 20 13. El aparato de la reivindicación 11, en el que el medio para clasificar la porción en el segundo clasificador como voz o como música comprende medios para comparar una pluralidad de características de la porción con uno o más umbrales para clasificar si la porción tiene características de música o características de voz.
- 25 14. Un medio legible por ordenador no transitorio que comprende instrucciones legibles por ordenador para hacer que un procesador realice el procedimiento de cualquiera de las reivindicaciones 1-10.
- 30 15. Un sistema que comprende:
- un primer clasificador que recibe una porción de señal de audio, clasifica la porción de la señal de audio como voz o como música, y procesa la porción de la señal de audio, en el que el procesamiento de la porción de la señal de audio comprende:
- 35 si la porción se clasifica como voz, entonces codifica la voz usando un primer modo de codificación, o
- si la porción se clasifica como música, entonces proporciona la porción a un segundo clasificador; y
- 40 el segundo clasificador, en el que si la porción se clasifica por el primer clasificador como música, clasifica la porción como voz o como música y codifica la porción de la señal de audio, en el que la codificación de la porción de la señal de audio comprende:
- si la porción se clasifica en el segundo clasificador como voz, codificar la porción usando un segundo modo de codificación; o
- 45 si la porción se clasifica en el segundo clasificador como música, codificar la porción usando un tercer modo de codificación.

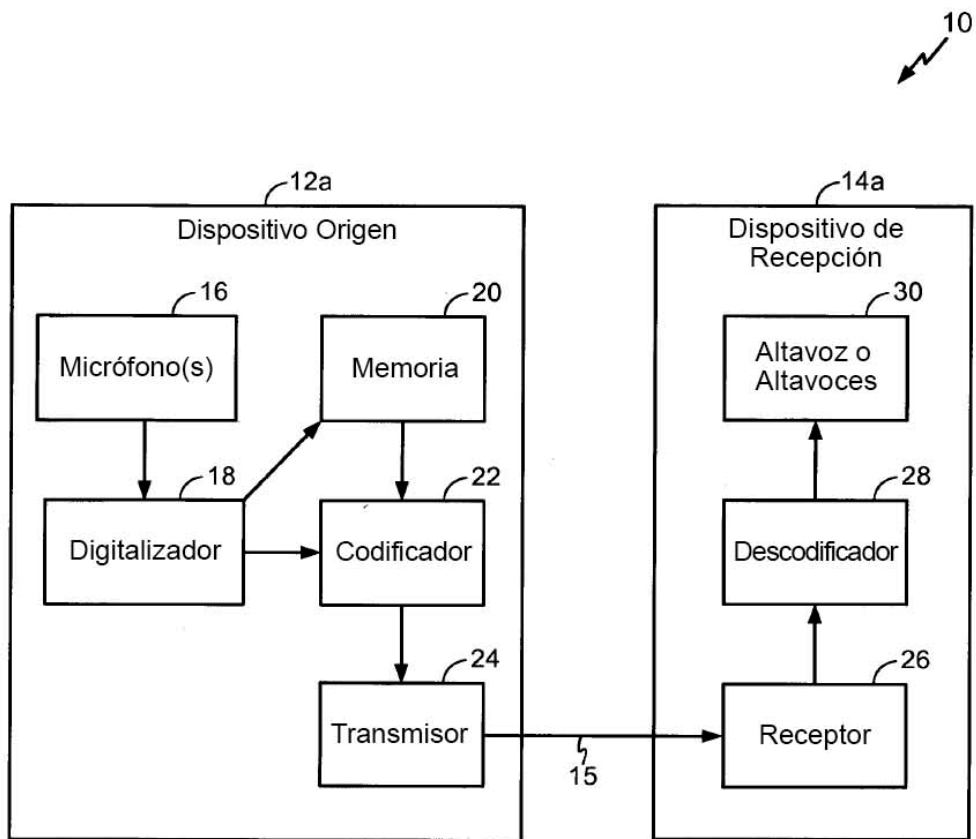


FIG. 1A

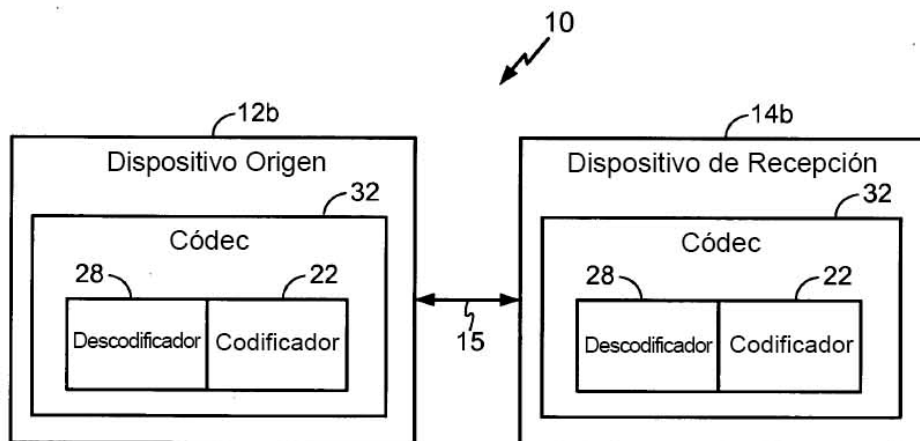


FIG. 1B

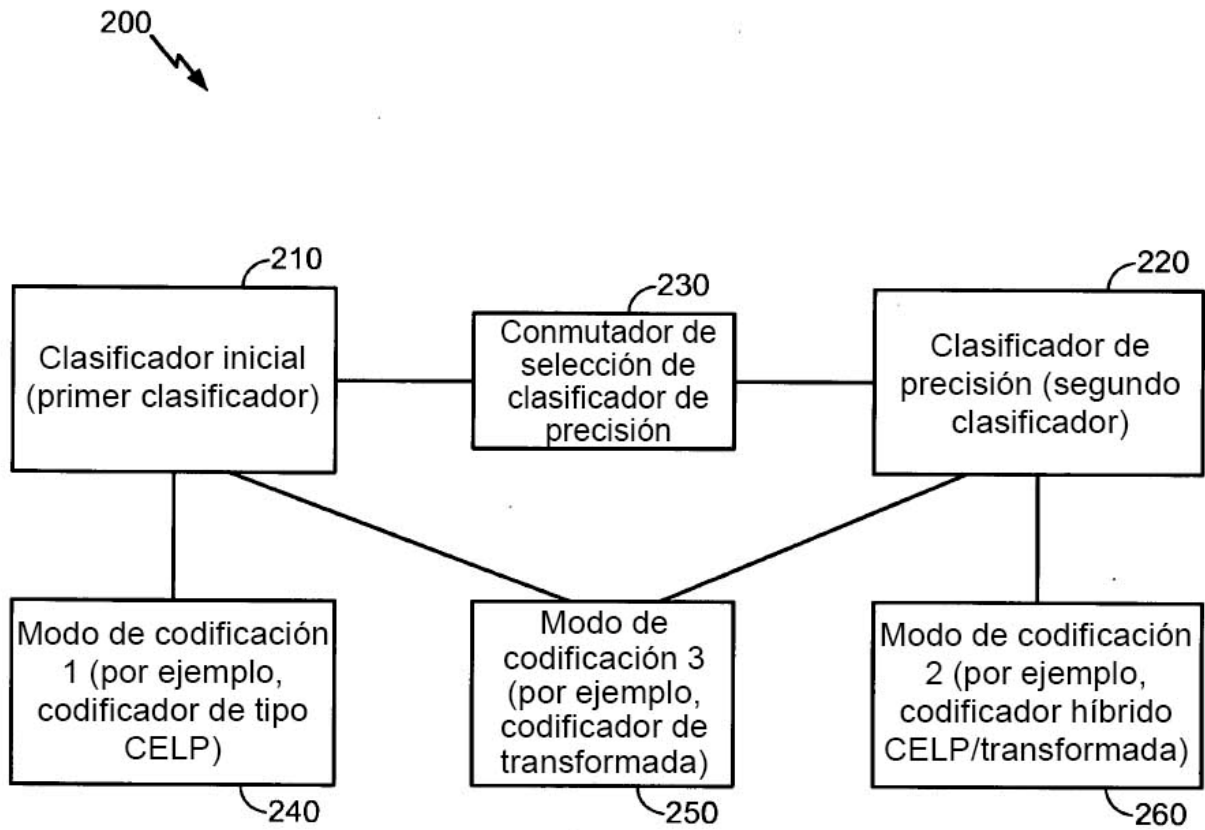


FIG. 2

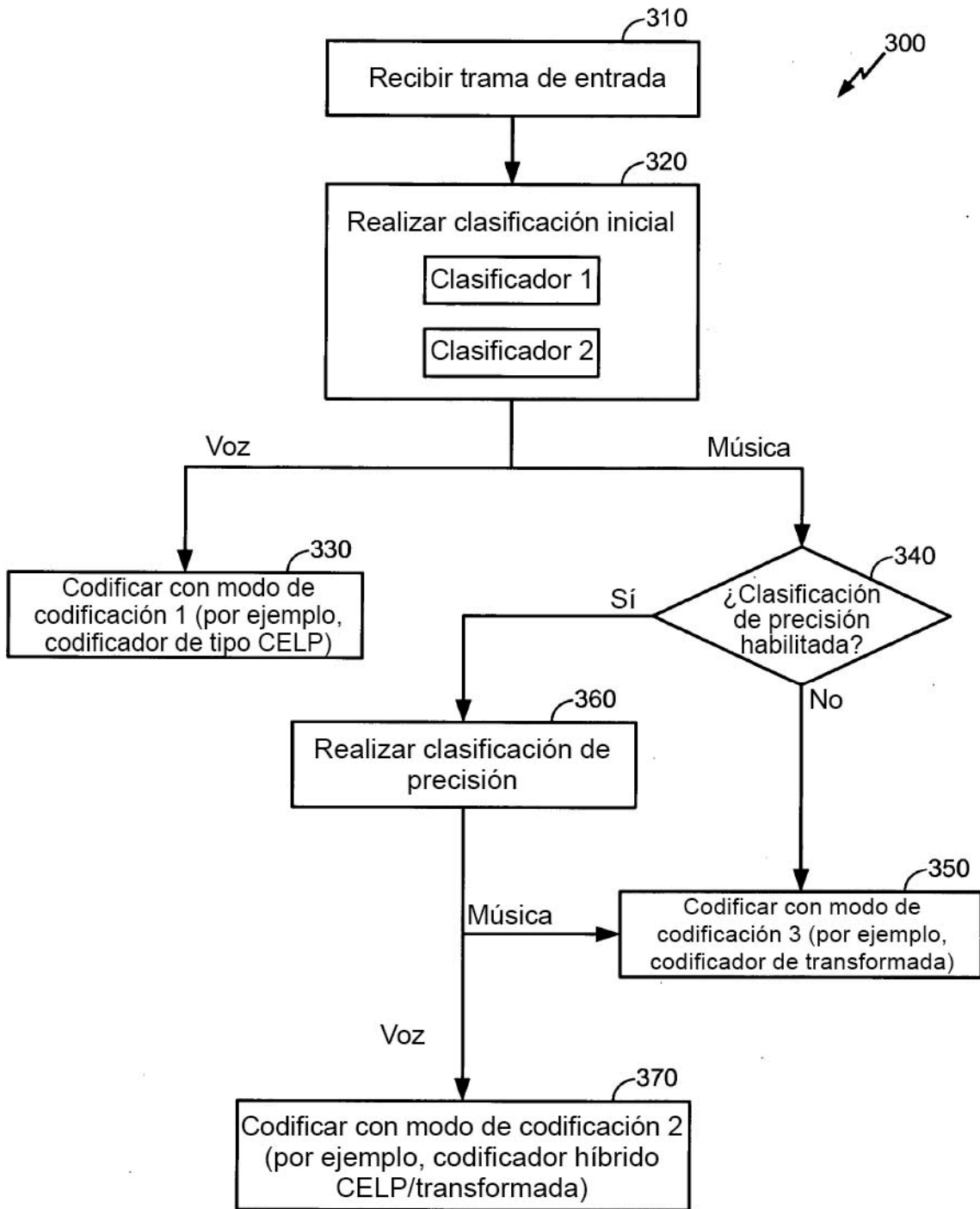


FIG. 3

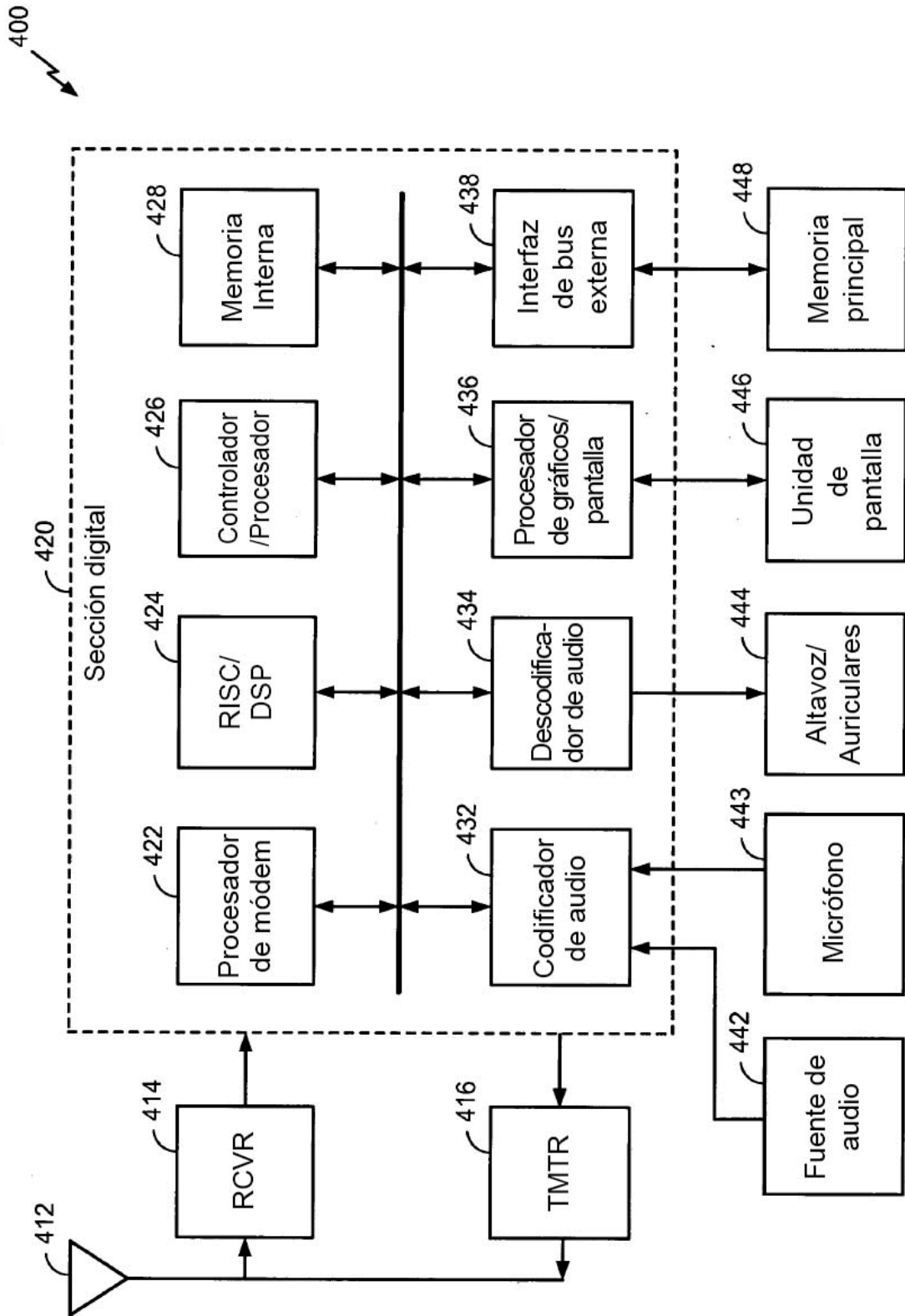


FIG. 4

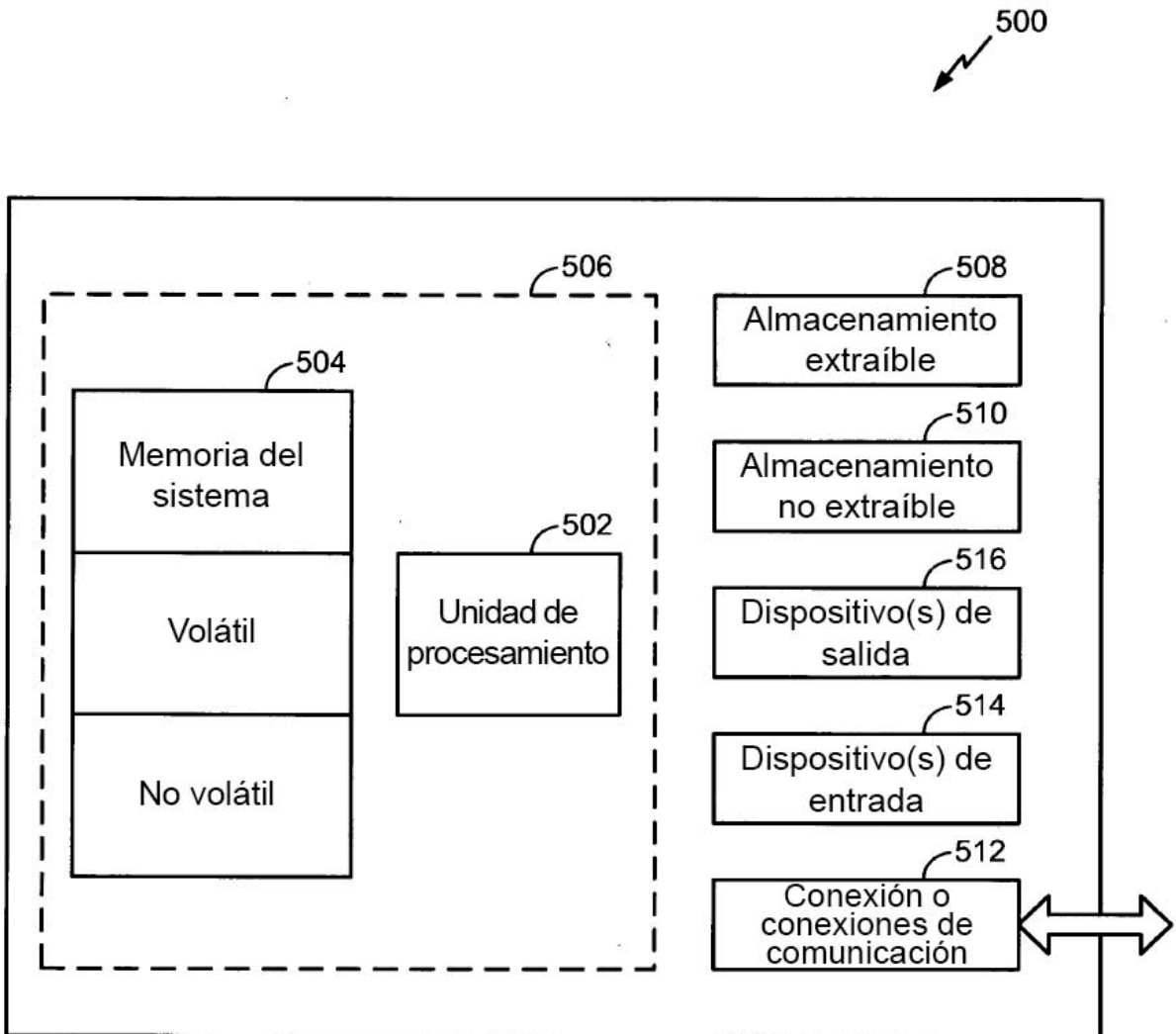


FIG. 5