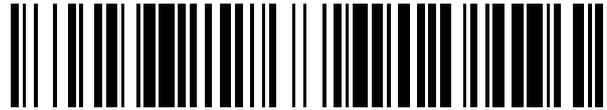


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 770 681**

51 Int. Cl.:

**G10L 19/20** (2013.01)

**G10L 25/27** (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **30.03.2016 PCT/US2016/025049**

87 Fecha y número de publicación internacional: **13.10.2016 WO16164231**

96 Fecha de presentación y número de la solicitud europea: **30.03.2016 E 16715227 (1)**

97 Fecha y número de publicación de la concesión europea: **06.11.2019 EP 3281198**

54 Título: **Selección de codificador**

30 Prioridad:

**05.04.2015 US 201562143155 P**  
**29.03.2016 US 201615083867**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**02.07.2020**

73 Titular/es:

**QUALCOMM INCORPORATED (100.0%)**  
**5775 Morehouse Drive**  
**San Diego, CA 92121-1714, US**

72 Inventor/es:

**ATTI, VENKATRAMAN, S.;**  
**CHEBIYYAM, VENKATA, SUBRAHMANYAM,**  
**CHANDRA, SEKHAR;**  
**RAJENDRAN, VIVEK y**  
**SUBASINGHA, SUBASINGHA, SHAMINDA**

74 Agente/Representante:

**FORTEA LAGUNA, Juan José**

ES 2 770 681 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

Selección de codificador

5 **I. Referencia cruzada a solicitudes relacionadas**

[0001] La presente solicitud reivindica el beneficio de la Solicitud de Patente de los EE. UU. número 15/083.867, titulada "ENCODER SELECTION [SELECCIÓN DE CODIFICADOR]", presentada el 29 de marzo de 2016, y la Solicitud de Patente Provisional de los EE. UU. número 62/143.155, titulada "ENCODER SELECTION [SELECCIÓN DE CODIFICADOR]", presentada el 5 de abril de 2015.

15 **II. Campo**

[0002] La presente divulgación se refiere, en general, a la selección de un codificador.

20 **III. Descripción de la técnica relacionada**

[0003] La grabación y transmisión de audio por técnicas digitales está muy extendida. Por ejemplo, el audio se puede transmitir en aplicaciones de larga distancia y radio telefonía digital. Los dispositivos, como los teléfonos inalámbricos, pueden enviar y recibir señales representativas de la voz humana (por ejemplo, voz) y no de voz (por ejemplo, música u otros sonidos).

[0004] En algunos dispositivos, están disponibles múltiples tecnologías de codificación. Por ejemplo, un codificador-descodificador de audio (CÓDEC) de un dispositivo puede usar un enfoque de codificación conmutada para codificar una variedad de contenido. Para ilustrar, el dispositivo puede incluir un codificador de voz, como un codificador de predicción lineal excitado por código algebraico (ACELP), y un codificador no de voz, como un codificador de excitación codificada por transformación (TCX) (por ejemplo, un codificador de dominio de transformación). El codificador de voz puede ser competente para codificar contenido de voz y el codificador no de voz, como un codificador de música, puede ser competente para codificar contenido inactivo y de música. Cabe señalar que, como se usa en el presente documento, un "codificador" podría referirse a uno de los modos de codificación de un codificador conmutado. Por ejemplo, el codificador ACELP y el codificador TCX podrían ser dos modos de codificación separados dentro de un codificador conmutado.

[0005] El dispositivo puede utilizar uno de los múltiples enfoques para clasificar una trama de audio y seleccionar un codificador. Por ejemplo, una trama de audio puede clasificarse como una trama de voz o como una trama no de voz (por ejemplo, una trama de música). Si la trama de audio se clasifica como una trama de voz, el dispositivo puede seleccionar el codificador de voz para codificar la trama de audio. De forma alternativa, si la trama de audio se clasifica como una trama no de voz (por ejemplo, una trama de música), el dispositivo puede seleccionar el codificador no de voz para codificar la trama de audio.

[0006] Un primer enfoque que puede usar el dispositivo para clasificar la trama de audio puede incluir un modelo de mezcla gaussiana (GMM) que se basa en las características de la voz. Por ejemplo, el GMM puede usar características de la voz, como tono, forma espectral, una métrica de correlación, etc., de la trama de audio para determinar si la trama de audio es más probable que sea una trama de voz o no de voz. El GMM puede ser competente en la identificación de tramas de voz, pero puede no funcionar tan bien para identificar tramas que no sean de voz (por ejemplo, tramas de música).

[0007] Un segundo enfoque puede incluir un clasificador en bucle abierto. El clasificador en bucle abierto puede predecir qué codificador (por ejemplo, el codificador de voz o el codificador no de voz) es más adecuado para codificar una trama de audio. El término "bucle abierto" se utiliza para indicar que la trama de audio no está codificada explícitamente antes de predecir qué codificador seleccionar. El clasificador en bucle abierto puede ser competente para identificar tramas que no son de voz, pero puede no funcionar tan bien para identificar tramas de voz.

[0008] Un tercer enfoque que puede usar el dispositivo para clasificar la trama de audio puede incluir un clasificador basado en modelos y un clasificador en bucle abierto. El clasificador basado en modelos puede enviar una decisión al clasificador en bucle abierto, que puede usar la decisión al clasificar la trama de audio.

[0009] El dispositivo puede analizar una señal de audio entrante trama por trama y puede decidir si codifica una trama de audio particular utilizando el codificador de voz o el codificador no de voz, como un codificador de música. Si la trama de audio particular está mal clasificada (por ejemplo, se clasifica incorrectamente como una trama de voz o como una trama no de voz), pueden producirse artefactos, mala calidad de la señal o una combinación de los mismos.

El documento US 2011/0202337 A1 divulga un procedimiento para clasificar segmentos de audio y voz de una señal.

65 **IV. Sumario**

**[0010]** La presente invención está definida por las reivindicaciones independientes adjuntas. Las reivindicaciones dependientes definen rasgos característicos opcionales de la invención.

5 **[0011]** Otros aspectos, ventajas y características de la presente divulgación resultarán evidentes después de revisar la solicitud, que incluye las siguientes secciones: Breve descripción de los dibujos, Descripción detallada y Reivindicaciones.

**V. Breve descripción de los dibujos**

10 **[0012]**

La FIG. 1 es un diagrama de bloques de un aspecto ilustrativo particular de un sistema que es operable para seleccionar un codificador;

15 la FIG. 2 es un diagrama de bloques de un aspecto ilustrativo particular de un selector del sistema de la FIG. 1;

la FIG. 3 es un diagrama de bloques de un aspecto ilustrativo particular de un primer clasificador del sistema de la FIG. 1;

20 la FIG. 4 es un diagrama de bloques de un aspecto ilustrativo particular de un segundo clasificador del sistema de la FIG. 1;

la FIG. 5 es un diagrama de flujo que ilustra un procedimiento para seleccionar un codificador;

25 la FIG. 6 es un diagrama de flujo que ilustra un procedimiento para seleccionar un valor de un parámetro de ajuste para desviar una selección hacia un codificador particular;

30 la FIG. 7 es un diagrama de bloques de un aspecto ilustrativo particular de un dispositivo que es operable para seleccionar un codificador; y

la FIG. 8 es un diagrama de bloques de un aspecto ilustrativo particular de una estación base que es operable para seleccionar un codificador.

35 **VI. Descripción detallada**

**[0013]** A continuación, se describen aspectos particulares de la presente divulgación con respecto a los dibujos. En la descripción, las características comunes se designan mediante números de referencia comunes. Como se usa en el presente documento, diversa terminología se usa con el propósito de describir implementaciones particulares solamente y no pretende ser limitativa. Por ejemplo, las formas en singular "un", "una", "el" y "la" pretenden incluir también las formas en plural, a menos que el contexto lo indique claramente de otro modo. Puede entenderse, además, que los términos "comprende" y "que comprende" pueden usarse indistintamente con "incluye" o "que incluye". Además, se entenderá que el término "en el que" se puede usar indistintamente con "dónde". Como se usa en el presente documento, un término ordinal (por ejemplo, "primero", "segundo", "tercero", etc.) usado para modificar un elemento, tal como una estructura, un componente, una operación, etc. no indica por sí mismo ninguna prioridad u orden del elemento con respecto a otro elemento, sino más bien meramente distingue el elemento de otro elemento que tenga un mismo nombre (excepto para el uso del término ordinal). Como se usa en el presente documento, el término "conjunto" se refiere a una agrupación de uno o más elementos, y el término "pluralidad" se refiere a múltiples elementos.

50 **[0014]** En la presente divulgación, se describen técnicas para seleccionar un codificador o un modo de codificación. Un dispositivo puede recibir una trama de audio y puede seleccionar un codificador particular de múltiples codificadores (o modos de codificación) para usar para codificar la trama de audio. Las técnicas descritas en el presente documento pueden usarse para establecer un valor de un parámetro de ajuste (por ejemplo, una métrica de histéresis) que se usa para desviar una selección hacia un codificador particular (por ejemplo, un codificador de voz o un codificador no de voz/música) o un modo de codificación particular. El parámetro de ajuste puede usarse para proporcionar una clasificación más precisa de la trama de audio, lo que puede dar como resultado una selección mejorada de un codificador para codificar la trama de audio.

60 **[0015]** Para ilustrar, el dispositivo puede recibir una trama de audio y puede usar múltiples clasificadores, como un primer clasificador y un segundo clasificador, para identificar un codificador que se seleccionará para codificar la trama de audio. El primer clasificador puede generar primeros datos de decisión basados en un modelo de voz (por ejemplo, circuitería de modelo de voz), basados en un modelo no de voz (por ejemplo, circuitería de modelo no de voz), o una combinación de los mismos. Los primeros datos de decisión pueden indicar si la trama de audio es una trama similar a la voz o una trama no de voz (por ejemplo, música, ruido de fondo, etc.). El contenido de voz puede designarse como que incluye voz activa, voz inactiva, voz con ruido o una combinación de las mismas, como ejemplos ilustrativos,

no limitativos. El contenido que no es de voz puede designarse como que incluye contenido musical, contenido similar a la música (por ejemplo, música en espera, tonos de llamada, etc.), ruido de fondo o una combinación de los mismos, como ejemplos ilustrativos y no limitativos. En otras implementaciones, el dispositivo puede clasificar la voz inactiva, la voz con ruido o una combinación de las mismas como contenido no de voz si un codificador particular asociado con voz (por ejemplo, un codificador de voz) tiene dificultades para descodificar la voz inactiva o la voz con ruido. En otra implementación, el ruido de fondo puede clasificarse como contenido de voz. Por ejemplo, el dispositivo puede clasificar el ruido de fondo como contenido de voz si un codificador particular asociado con voz (por ejemplo, un codificador de voz) es competente para descodificar ruido de fondo.

**[0016]** En algunas implementaciones, el primer clasificador puede estar asociado con un algoritmo de máxima verosimilitud (por ejemplo, basado en modelos de mezcla gaussiana, basado en modelos ocultos de Markov o basado en redes neuronales). Para generar los primeros datos de decisión, el primer clasificador puede generar uno o más valores de probabilidad, como un primer valor de probabilidad (por ejemplo, primeros datos de probabilidad) asociado con una primera probabilidad de que la trama de audio sea una trama de voz, un segundo valor de probabilidad (por ejemplo, datos de segunda probabilidad), asociado con una segunda probabilidad de que la trama de audio sea una trama no de voz, o una combinación de las mismas. El primer clasificador puede incluir una máquina de estados que recibe los primeros datos de probabilidad, los segundos datos de probabilidad, o una combinación de los mismos, y que genera los primeros datos de decisión. Los primeros datos de decisión pueden ser emitidos por la máquina de estados y recibidos por el segundo clasificador.

**[0017]** El segundo clasificador puede configurarse para generar segundos datos de decisión asociados con (por ejemplo, que indican) una selección de un codificador particular de múltiples codificadores para codificar la trama de audio. Los segundos datos de decisión pueden corresponder a una clasificación actualizada o modificada de la trama de audio (por ejemplo, los segundos datos de decisión pueden indicar una clasificación diferente de los primeros datos de decisión). En algunas implementaciones, los primeros datos de decisión pueden indicar la misma clasificación que los segundos datos de decisión. Además, o de forma alternativa, los segundos datos de decisión pueden corresponder a una "decisión final" (por ejemplo, si la trama de audio tiene una clasificación de trama de voz, se selecciona el codificador de voz). El segundo clasificador puede ser un clasificador basado en modelos, puede ser un clasificador no basado únicamente en modelos (por ejemplo, un clasificador en bucle abierto), o puede basarse en un conjunto de parámetros de codificación. Los parámetros de codificación pueden incluir un indicador clave, un modo de codificación, un tipo de codificador, una decisión clave de paso bajo, un valor de tono, una estabilidad de tono o una combinación de los mismos, como ejemplos ilustrativos y no limitativos.

**[0018]** El segundo clasificador puede generar los segundos datos de decisión basándose en los primeros datos de decisión, los primeros datos de probabilidad, los segundos datos de probabilidad, o una combinación de los mismos. En algunas implementaciones, el segundo clasificador puede usar uno o más del conjunto de parámetros de codificación para generar los segundos datos de decisión. Además, el segundo clasificador puede generar los segundos datos de decisión basándose en una o más condiciones asociadas con la trama de audio. Por ejemplo, el segundo clasificador puede determinar si se cumple un conjunto de condiciones asociadas con la trama de audio, como se describe en el presente documento. En respuesta a una o más condiciones del conjunto de condiciones que se satisfacen (o no se satisfacen), el segundo clasificador puede determinar un valor de un parámetro de ajuste para desviar (por ejemplo, influir) una selección hacia un primer codificador (por ejemplo, el codificador de voz) o un segundo codificador (por ejemplo, el codificador no de voz). En otras implementaciones, el segundo clasificador puede determinar un valor de un parámetro de ajuste para desviar (por ejemplo, influir en) una selección hacia un modo de codificación particular de un codificador conmutable que tiene múltiples modos de codificación, como un codificador conmutado. El parámetro de ajuste puede funcionar como una métrica de histéresis (por ejemplo, una métrica basada en el tiempo) que puede ser utilizada por el segundo clasificador para mejorar la selección de un codificador para la trama de audio. Por ejemplo, la métrica de histéresis puede "suavizar" un flujo de audio codificado que incluye la trama de audio codificada al retardar o reducir la conmutación entre dos codificadores hasta que se haya identificado un umbral de tramas de audio secuenciales con una clasificación particular.

**[0019]** El conjunto de condiciones puede incluir una primera condición de que al menos uno de los codificadores esté asociado con una primera frecuencia de muestreo (por ejemplo, una frecuencia de muestreo de audio). En algunas implementaciones, la primera frecuencia de muestreo puede ser una frecuencia de muestreo de audio baja, como 12,8 kilohercios (kHz), como un ejemplo ilustrativo no limitante. En otras implementaciones, la primera frecuencia de muestreo puede ser mayor o menor que 12,8 kHz, como 14,4 kHz u 8 kHz. En una implementación particular, la primera frecuencia de muestreo puede ser menor que otras frecuencias de muestreo utilizadas por los codificadores. El conjunto de condiciones puede incluir una segunda condición de que la primera decisión esté asociada con la clasificación de la trama de audio como una trama de voz. El conjunto de condiciones puede incluir una tercera condición de que un primer valor de ganancia de codificación estimada asociado con el primer codificador que se usa para codificar la trama de audio es mayor o igual que un primer valor, donde el primer valor está asociado con una diferencia entre un segundo valor de ganancia de codificación estimada y un segundo valor.

**[0020]** En algunas implementaciones, si una trama clasificada más recientemente se asocia con contenido de voz, el conjunto de condiciones puede incluir una condición asociada con una determinación de que el primer valor de probabilidad es mayor o igual que el segundo valor de probabilidad. De forma alternativa, si cada trama de múltiples

tramas recientemente clasificadas está asociada con contenido de voz, el conjunto de condiciones puede incluir otra condición que está asociada con una determinación de que el primer valor de probabilidad es mayor o igual que un tercer valor, donde el tercer valor está asociado con una diferencia entre el segundo valor de probabilidad y un cuarto valor.

5 **[0021]** En algunas implementaciones, el conjunto de condiciones puede incluir una condición asociada con un valor promedio de voz de múltiples subtramas de la trama de audio que es mayor o igual que un primer umbral. Adicionalmente o de forma alternativa, el conjunto de condiciones puede incluir una condición asociada con un valor de no estacionaridad asociado con que la trama de audio sea mayor que un segundo umbral. Adicionalmente o de forma alternativa, el conjunto de condiciones puede incluir una condición asociada con un valor de desplazamiento asociado con que la trama de audio sea inferior a un tercer umbral.

15 **[0022]** Con referencia a la FIG. 1, se divulga un ejemplo ilustrativo particular de un sistema 100 operable para seleccionar un codificador. El sistema 100 incluye un dispositivo 102 configurado para recibir una señal de audio que puede incluir voz de entrada 110, tal como una señal de audio muestreada recibida a través de un micrófono que está acoplado o incluido en el dispositivo 102. El dispositivo 102 está configurado para seleccionar entre un primer codificador 132 y un segundo codificador 134 para codificar todo o parte de la voz de entrada 110. Aunque el primer codificador 132 y el segundo codificador 134 se ilustran como codificadores separados, en otras implementaciones, el primer codificador 132 y el segundo codificador 134 pueden incluirse en un único codificador (por ejemplo, un codificador conmutado). Por ejemplo, el primer codificador 132 y el segundo codificador 134 pueden corresponder a diferentes modos de codificación de un codificador conmutado. La codificación de la voz de entrada 110 puede generar una serie de tramas de audio codificadas, tales como una trama de audio codificada 114, que pueden enviarse a uno o más dispositivos distintos, como a través de una red inalámbrica. Por ejemplo, el dispositivo 102 puede participar en una llamada de voz, tal como una llamada de voz sobre protocolo de Internet (VoIP), con un dispositivo remoto. En algunas implementaciones, el primer codificador 132, el segundo codificador 134, o ambos, pueden configurarse para operar de acuerdo con uno o más protocolos/estándares, tal como de acuerdo (por ejemplo, en cumplimiento) con un protocolo/estándar de servicios de voz mejorados (EVS) del Proyecto de Colaboración de Tercera Generación (3GPP), como un ejemplo ilustrativo, no limitante.

25 **[0023]** El dispositivo 102 incluye un codificador 104 que incluye un selector 120, un conmutador 130 y múltiples codificadores que incluyen el primer codificador 132 y el segundo codificador 134. El codificador 104 está configurado para recibir tramas de audio de la señal de audio que incluye la voz de entrada 110, tal como una trama de audio 112. La señal de audio puede incluir datos de voz, datos no de voz (por ejemplo, música o ruido de fondo) o ambos. El selector 120 puede estar configurado para determinar si se va a codificar cada trama de la señal de audio mediante el primer codificador 132 o el segundo codificador 134. Por ejemplo, el primer codificador 132 puede incluir un codificador de voz, tal como un codificador ACELP, y el segundo codificador 134 puede incluir un codificador no de voz, tal como un codificador de música. En una implementación particular, el segundo codificador 134 incluye un codificador TCX. El conmutador 130 responde al selector 120 para enrutar la trama de audio 112 a uno seleccionado del primer codificador 132 o el segundo codificador 134 para generar la trama de audio codificada 114.

35 **[0024]** El selector 120 puede incluir un primer clasificador 122 y/o un segundo clasificador 124. El primer clasificador 122 puede configurarse para recibir la trama de audio 112 o una parte de la trama de audio 112, tal como un conjunto de características descrito con respecto a las FIGS. 2 y 3. El primer clasificador 122 puede estar configurado para emitir primeros datos de decisión 146 que indican una clasificación de la trama de audio 112 como una trama de voz o una trama no de voz. Los primeros datos de decisión 146 pueden determinarse en base a los primeros datos de probabilidad 142 asociados con una primera probabilidad de que la trama de audio sea una trama de voz y en base a los segundos datos de probabilidad 144 asociados con una segunda probabilidad de que la trama de audio sea una trama no de voz. Por ejemplo, el primer clasificador 122 puede incluir o corresponder a un clasificador basado en modelos, un circuito GMM (por ejemplo, un módulo GMM), o ambos. Una implementación particular del primer clasificador 122 se describe con más detalle con respecto a la FIG. 3.

40 **[0025]** El segundo clasificador 124 está acoplado al primer clasificador 122 y configurado para emitir los segundos datos de decisión 148 en base a los primeros datos de probabilidad 142, los segundos datos de probabilidad 144 y los primeros datos de decisión 146. Los segundos datos de decisión 148 indican una selección de un codificador particular de múltiples codificadores (por ejemplo, el primer codificador 132 o el segundo codificador 134) que está disponible para codificar la trama de audio 112. En algunas implementaciones, el segundo clasificador 124 puede configurarse para recibir la trama de audio 112. El segundo clasificador 124 puede recibir la trama de audio 112 desde el primer clasificador 122, desde el codificador 104, o desde otro componente del dispositivo 102. Adicionalmente o de forma alternativa, el segundo clasificador 124 puede configurarse para generar un parámetro de ajuste. Un valor del parámetro de ajuste puede desviar (por ejemplo, influir en) los segundos datos de decisión 148 para indicar un codificador particular de los múltiples codificadores (por ejemplo, el primer codificador 132 o el segundo codificador 134). Por ejemplo, un primer valor del parámetro de ajuste puede aumentar la probabilidad de seleccionar el codificador particular. El segundo clasificador 124 puede incluir o corresponder a un clasificador en bucle abierto. Una implementación particular del segundo clasificador 124 se describe con más detalle con respecto a la FIG. 4.

65

**[0026]** El conmutador 130 está acoplado al selector 120 y puede configurarse para recibir los segundos datos de decisión 148. El conmutador 130 puede configurarse para seleccionar el primer codificador 132 o el segundo codificador 134 de acuerdo con los segundos datos de decisión 148. El conmutador 130 puede configurarse para proporcionar la trama de audio 112 al primer codificador 132 o al segundo codificador 134 de acuerdo con (por ejemplo, en base a) los segundos datos de decisión 148. En otras implementaciones, el conmutador 130 proporciona o dirige una señal a un codificador seleccionado para activar o habilitar una salida del codificador seleccionado.

**[0027]** El primer codificador 132 y el segundo codificador 134 pueden estar acoplados al conmutador 130 y configurados para recibir la trama de audio 112 desde el conmutador 130. En otras implementaciones, el primer codificador 132 o el segundo codificador 134 pueden configurarse para recibir la trama de audio 112 desde otro componente del dispositivo 102. El primer codificador 132 y el segundo codificador 134 pueden configurarse para generar la trama de audio codificada 114 en respuesta a la recepción de la trama de audio 112.

**[0028]** Durante el funcionamiento, la voz de entrada 110 puede procesarse trama por trama, y puede extraerse un conjunto de características de la voz de entrada 110 en el codificador 104 (por ejemplo, en el selector 120). El conjunto de características puede ser utilizado por el primer clasificador 122. Por ejemplo, el primer clasificador 122 (por ejemplo, un clasificador basado en modelos) puede generar y emitir los primeros datos de probabilidad 142 y los segundos datos de probabilidad 144, tales como una probabilidad de voz a corto plazo ("lps") y una probabilidad de música a corto plazo ("lpm"), respectivamente. Como se describe con respecto a la FIG. 3, los valores lps y lpm correspondientes a una trama particular pueden proporcionarse a una máquina de estados en el primer clasificador 122 que realiza un seguimiento de uno o más estados (por ejemplo, parámetros de estado) del codificador 104 para generar una decisión de voz o música ("decisión spaud") para la trama particular. El uno o más estados del codificador 104 pueden incluir valores de parámetros a largo plazo, tales como un recuento de tramas inactivas, un recuento de tramas de voz, un recuento de tramas de música, etc. La máquina de estados también puede recibir parámetros tales como un decisión de actividad de voz de un detector de actividad de voz (VAD), una energía de una trama actual, etc. Aunque el VAD se describe como un detector de actividad de "voz", debe entenderse que el VAD es un discriminador entre una señal activa (que puede incluir voz o música) y una señal inactiva, como ruido de fondo.

**[0029]** El segundo clasificador 124 puede usar características a corto plazo extraídas de la trama para estimar dos estimaciones o medidas de ganancia de codificación, denominadas proporción señal/ruido para la codificación ACELP ("snr\_acelp") y una proporción señal/ruido para codificación TCX ("snr\_tcx"). Aunque se conocen como proporciones SNR, snr\_acelp y snr\_tcx pueden ser estimaciones de ganancia de codificación u otras estimaciones o medidas que pueden corresponder a la probabilidad de que una trama actual sea voz o música, respectivamente, o que pueda corresponder a un grado estimado de efectividad del primer codificador 132 (por ejemplo, un codificador ACELP) o el segundo codificador 134 (por ejemplo, un codificador TCX) en la codificación de la trama. El segundo clasificador 124 puede modificar (por ejemplo, ajustar un valor de) snr\_acelp, snr\_tcx, o ambos, en base a información a largo plazo, como los primeros datos de decisión 146 (por ejemplo, "sp\_aud\_decision") y, además, en base a datos adicionales del primer clasificador 122, tal como los primeros datos de probabilidad 142 (por ejemplo, "lps"), los segundos datos de probabilidad 144 (por ejemplo, "lpm"), uno o más parámetros, o una combinación de los mismos.

**[0030]** Por lo tanto, el selector 120 puede desviar (por ejemplo, influir en) la decisión de qué codificador (por ejemplo, el primer codificador 132 o el segundo codificador 134) se aplicará a una trama particular en base a parámetros a largo y corto plazo que pueden generarse en cualquiera de los clasificadores 122, 124 y como se muestra en la FIG. 2. Al seleccionar un codificador basándose en datos adicionales (por ejemplo, los primeros datos de probabilidad 142, los segundos datos de probabilidad 144, o ambos) del primer clasificador 122, el selector 120 puede reducir una cantidad de falsos positivos y una cantidad de detecciones perdidas para seleccionar tramas que serán codificadas por el primer codificador 132 o el segundo codificador 134 en comparación con una implementación que utiliza una decisión de un primer clasificador (por ejemplo, un clasificador basado en modelos o un clasificador en bucle abierto) para seleccionar el primer codificador 132 o el segundo codificador 134 para cada trama.

**[0031]** Además, cabe destacar que, aunque la FIG. 1 ilustra el primer codificador 132 y el segundo codificador 134, no se debe considerar esto limitante. En ejemplos alternativos, se pueden incluir dos codificadores, circuitos u otros módulos. Por ejemplo, el codificador 104 puede incluir uno o más circuitos o módulos "centrales" de banda baja (LB) (por ejemplo, un núcleo TCX, un núcleo ACELP, uno o más núcleos distintos, o cualquier combinación de los mismos) y uno o más circuitos o módulos de extensión de ancho de banda (BWE)/de banda alta (HB). Se puede proporcionar una parte de banda baja de una trama de audio 112 para la codificación de voz a un módulo o circuito central de banda baja particular para su codificación, dependiendo de las características de la trama (por ejemplo, si la trama contiene voz, ruido, música, etc.). Se puede proporcionar la parte de banda alta de cada trama a un módulo o circuito HB/BWE particular.

**[0032]** La FIG. 2 es un diagrama de bloques de un ejemplo ilustrativo particular 200 del selector 120 de la FIG. 1. En el ejemplo 200, el selector 120 está configurado para recibir una trama de entrada (por ejemplo, la trama de audio 112 de la FIG. 1) y los datos correspondientes a un estado a largo plazo del codificador 104 y emitir una decisión de voz/música (por ejemplo, los primeros datos de decisión 146 de la FIG. 1). Un extractor de características a corto plazo 226 está configurado para recibir la trama de entrada y generar un conjunto de características extraídas de la

trama de entrada. Para ilustrar, el extractor de características a corto plazo 226 puede configurarse para generar características a corto plazo basadas en la trama de entrada.

**[0033]** El primer clasificador 122 se representa como un clasificador basado en modelos que está configurado para recibir el conjunto de características desde el extractor de características a corto plazo 226 y los datos de estado a largo plazo. El primer clasificador 122 está configurado para generar un indicador de una probabilidad de voz a corto plazo ("lps") (por ejemplo, los primeros datos de probabilidad 142 de la FIG. 1), un indicador de probabilidad de música a corto plazo ("lpm") (por ejemplo, los segundos datos de probabilidad 144 de la FIG. 1) y la decisión de voz/música ("sp\_aud\_decision") (por ejemplo, los primeros datos de decisión 146 de la FIG. 1). En algunas implementaciones, el primer clasificador 122 puede configurarse para recibir la trama de entrada.

**[0034]** El segundo clasificador 124 se representa como un clasificador en bucle abierto que está configurado para recibir la trama de entrada y los datos de estado a largo plazo. El segundo clasificador 124 también puede configurarse para recibir las características a corto plazo del extractor de características a corto plazo 226 y para recibir el indicador de una probabilidad de voz a corto plazo ("lps"), el indicador de una probabilidad a corto plazo de música ("lpm"), y la decisión de voz/música ("sp\_aud\_decision") del primer clasificador 122. El segundo clasificador 124 está configurado para emitir una decisión de clasificación actualizada (o modificada) (por ejemplo, los segundos datos de decisión 148 de la FIG. 1). El segundo clasificador 124 puede emitir los segundos datos de decisión a un conmutador (por ejemplo, el conmutador 130 de la FIG. 1) o a un codificador conmutado. Adicionalmente o de forma alternativa, el segundo clasificador 124 puede configurarse para recibir el conjunto de características del extractor de características a corto plazo 226.

**[0035]** Los detalles del primer clasificador 122 se ilustran de acuerdo con un ejemplo particular 300 que se representa en la FIG. 3. En el ejemplo 300, el primer clasificador 122 incluye un modelo de voz 370 (por ejemplo, circuitos de modelo de voz), un modelo de música 372 (por ejemplo, circuitos de modelo de música) y una máquina de estados 374. El modelo de voz 370 está configurado para calcular el indicador de una probabilidad de voz a corto plazo ("lps") basándose en el conjunto de características recibido del extractor de características a corto plazo 226 de la FIG. 2. El modelo de música 372 está configurado para calcular el indicador de una probabilidad de música a corto plazo ("lpm") basándose en el conjunto de características recibido del extractor de características a corto plazo 226. En otras implementaciones, el primer clasificador 122 puede recibir la trama de entrada y puede determinar el conjunto de características.

**[0036]** La máquina de estados 374 puede configurarse para recibir los primeros datos de probabilidad (por ejemplo, el indicador de probabilidad de voz ("lps") a corto plazo emitido del modelo de voz 370, correspondiente a los primeros datos de probabilidad 142 de la FIG. 1). La máquina de estados 374 puede configurarse para recibir segundos datos de probabilidad (por ejemplo, el indicador de una probabilidad de música a corto plazo ("lpm") emitido del modelo de música 372, correspondiente a los segundos datos de probabilidad 144 de la FIG. 1). La máquina de estados 374 puede configurarse para generar la decisión de voz/música ("sp\_aud\_decision") (por ejemplo, los primeros datos de decisión 146 de la FIG. 1) basándose en los primeros datos de probabilidad y los segundos datos de probabilidad.

**[0037]** Los detalles del segundo clasificador 124 se ilustran de acuerdo con un ejemplo particular 400 que se representa en la FIG. 4. En el ejemplo 400, el segundo clasificador 124 incluye un estimador de probabilidad de voz a corto plazo 410, un estimador de probabilidad de música a corto plazo 412, una unidad de desviación de decisión a largo plazo 414, un generador de parámetros de ajuste 416 y un generador de decisión de clasificación 418.

**[0038]** El estimador de probabilidad de voz a corto plazo 410 está configurado para recibir el conjunto de características a corto plazo extraídas de la trama de entrada (por ejemplo, del extractor de características a corto plazo 226 de la FIG. 2) y la trama de entrada (por ejemplo, la trama de audio 112 de la FIG. 1). El estimador de probabilidad de voz a corto plazo 410 está configurado para generar un primer valor de ganancia de codificación estimada (por ejemplo, "snr\_acelp") correspondiente a una ganancia de codificación estimada o eficiencia de codificación de la trama de entrada usando un codificador ACELP (por ejemplo, el primer codificador 132 de FIG. 1).

**[0039]** El estimador de probabilidad de música a corto plazo 412 está configurado para recibir el conjunto de características a corto plazo extraídas de la trama de entrada (por ejemplo, del extractor de características a corto plazo 226 de la FIG. 2) y la trama de entrada. El estimador de probabilidad de música a corto plazo 412 está configurado para generar un segundo valor de ganancia de codificación estimada (por ejemplo, "snr\_tcx") correspondiente a una ganancia o eficiencia de codificación para codificar la trama de entrada usando un codificador TCX (por ejemplo, el segundo codificador 134 de FIG. 1).

**[0040]** La unidad de desviación de decisión a largo plazo 414 está configurada para recibir el primer valor de ganancia de codificación estimada (por ejemplo, "snr\_acelp"), el segundo valor de ganancia de codificación estimada (por ejemplo, "snr\_tcx"), la decisión de voz/música ("sp\_aud\_decision") generada por el primer clasificador 122 como se representa en la FIG. 3, y los datos del estado a largo plazo. La unidad de desviación de decisión a largo plazo 414 está configurada para generar una salida basada en uno o más de los valores introducidos en la unidad de desviación de decisión a largo plazo 414.

**[0041]** El generador de parámetros de ajuste 416 está configurado para recibir los primeros datos de probabilidad (por ejemplo, "lps") emitidos del modelo de voz 370 de la FIG. 3, los segundos datos de probabilidad (por ejemplo, "lpm") emitidos del modelo de música 372 de la FIG. 3, los datos de estado a largo plazo y la salida de la unidad de desviación de decisión a largo plazo 414. El generador de parámetros de ajuste 416 está configurado para establecer un valor de un parámetro de ajuste (denominado "dsnr") que utiliza el generador de decisiones de clasificación 418 para desviar una decisión de voz/música hacia un codificador de voz o hacia un codificador de música. Aunque el parámetro de ajuste está etiquetado como "dsnr" en la FIG. 4 y en los Ejemplos descritos a continuación, el parámetro de ajuste puede corresponder o no a una proporción señal/ruido. Por ejemplo, en algunas implementaciones, el valor de ajuste puede representar un desplazamiento de una proporción señal/ruido (por ejemplo, un "delta snr"), mientras que en otras implementaciones el parámetro de ajuste puede corresponder a un desplazamiento de un valor de ganancia de codificación o una proporción de ganancia de codificación (por ejemplo, "delta de ganancia de codificación"), un desplazamiento de una estimación de ganancia de codificación o de uno o más valores físicos o parámetros del modelo, o puede ser un valor numérico que no tiene una correspondencia directa con un valor físico o parámetro del modelo. Por lo tanto, debe entenderse que la etiqueta "dsnr" se usa solo por conveniencia y no impone ninguna limitación sobre el contenido o el uso del parámetro de ajuste.

**[0042]** El generador de decisión de clasificación 418 está configurado para recibir el primer valor de ganancia de codificación estimada (por ejemplo, "snr\_aceip"), el segundo valor de ganancia de codificación estimada (por ejemplo, "snr\_tcx"), el parámetro de ajuste (por ejemplo, "dsnr"), el conjunto de características a corto plazo del extractor de características a corto plazo 226 de la FIG. 2, los datos de estado a largo plazo y la decisión de voz/música ("sp\_aud\_decision") generada por el primer clasificador 122 como se representa en la FIG. 3. En base a los valores de entrada recibidos, el generador de decisión de clasificación 418 está configurado para emitir la decisión de clasificación actualizada (o modificada), que puede corresponder a los segundos datos de decisión 148 de la FIG. 1.

**[0043]** Un valor del parámetro de ajuste ("dsnr") desvía la decisión de voz/música del generador de decisión de clasificación 418. Por ejemplo, un valor positivo del parámetro de ajuste puede hacer que el generador de decisión de clasificación 418 sea más probable que seleccione un codificador de voz para la trama de entrada, y un valor negativo del parámetro de ajuste puede hacer que el generador de decisión de clasificación 418 sea más probable que seleccione un codificador no de voz para la trama de entrada.

**[0044]** Como se describe con respecto a la FIG. 4, varios parámetros están disponibles y pueden usarse para influir en o desviar la decisión de voz/música hacia voz o no voz. Por ejemplo, la probabilidad de voz a corto plazo ("lps"), la probabilidad de música a corto plazo ("lpm"), o una combinación de ambas, calculada por el modelo de voz y el modelo de música como parámetros intermedios para obtener la decisión del primer clasificador 122 ("sp\_aud\_decision") puede usarse para desviar la decisión de voz/música del segundo clasificador 124.

**[0045]** Como otro ejemplo, la decisión a largo plazo del primer clasificador 122 ("sp\_aud\_decision") puede usarse para desviar la decisión de voz/música del segundo clasificador 124. Como otro ejemplo, una cercanía (por ejemplo, similitud numérica) de las estimaciones de ganancia de codificación a corto plazo (por ejemplo, "snr\_aceip" y "snr\_tcx") puede usarse para desviar la decisión de voz/música del segundo clasificador 124.

**[0046]** Como otro ejemplo, una serie de tramas consecutivas pasadas que se eligieron como ACELP/voz (por ejemplo, en los datos de estado a largo plazo) pueden usarse para desviar la decisión de voz/música del segundo clasificador 124. De forma alternativa, se puede usar una medida del número de tramas ACELP/voz elegidas entre un subconjunto de las tramas anteriores (un ejemplo de esto podría ser el porcentaje de tramas ACELP/voz en las últimas 50 tramas) para desviar la decisión de voz/música del segundo clasificador 124.

**[0047]** Como otro ejemplo, una decisión de trama previa entre ACELP/voz y TCX/música (por ejemplo, en los datos de estado a largo plazo) puede usarse para desviar la decisión de voz/música del segundo clasificador 124. Como otro ejemplo, una medida no estacionaria de la energía de la voz ("non\_staX") puede estimarse como una suma de las proporciones de la energía de una trama actual y la energía de la trama pasada entre diferentes bandas de frecuencia. La medida de no estacionaridad puede incluirse en el conjunto de características proporcionadas por el extractor de características a corto plazo 226 de la FIG. 2. La medida de no estacionaridad puede usarse para desviar la decisión de voz/música del segundo clasificador 124.

**[0048]** Como otro ejemplo, la voz promedio (por ejemplo, media o aritmética media) entre todas (o un subconjunto de) las subtramas de una trama de entrada puede usarse para desviar la decisión de voz/música del segundo clasificador 124. La voz promedio puede incluir una medida de la correlación normalizada de voz en las subtramas con una versión desplazada de voz. Una cantidad de desplazamiento de la versión desplazada puede corresponder a un desfase de tono calculado de la subtrama. Una voz alta indica que la señal es altamente repetitiva coincidiendo sustancialmente el intervalo de repetición con el retardo de tono. La voz promedio se puede incluir en el conjunto de características proporcionadas por el extractor de características a corto plazo 226 de la FIG. 2.

**[0049]** Como otro ejemplo, se puede usar un parámetro de desplazamiento para desviar la decisión de voz/música del segundo clasificador 124. Por ejemplo, si se usa un codificador TCX para codificar segmentos de música, el parámetro de desplazamiento puede incorporarse al desviar la decisión de voz/música. El parámetro de

desplazamiento puede corresponder a una medida inversa de la ganancia de codificación TCX. El parámetro de desplazamiento puede estar inversamente relacionado con el segundo valor de ganancia de codificación estimada ("snr\_tcx"). En una implementación particular, se puede determinar si un valor del parámetro de desplazamiento es menor que un umbral (por ejemplo, desplazamiento < 74,0) para imponer un criterio mínimo correspondiente al segundo valor de ganancia de codificación estimada ("snr\_tcx"). Verificar que el parámetro de desplazamiento no sea menor que el umbral, además de verificar que el primer valor de ganancia de codificación estimada ("snr\_acelp") excede otro umbral (por ejemplo, snr\_acelp > snr\_tcx-4), puede indicar si uno o ambos codificadores son insuficientes para codificar la trama de entrada. Si ambos codificadores son insuficientes para codificar la trama de entrada, se puede usar un tercer codificador para codificar la trama de entrada. Aunque se enumeran anteriormente varios parámetros que pueden usarse para desviar una selección de codificador, debe entenderse que algunas implementaciones pueden excluir uno o más de los parámetros enumerados, incluir uno o más parámetros, o cualquier combinación de los mismos.

**[0050]** Al modificar (por ejemplo, ajustar un valor de) estimaciones o medidas de ganancia de codificación basándose en datos adicionales (por ejemplo, datos del primer clasificador 122 de la FIG. 1), el segundo clasificador 124 puede reducir una cantidad de falsos positivos y una cantidad de detecciones perdidas para seleccionar tramas para codificar en comparación con una implementación que utiliza una decisión de un primer clasificador (por ejemplo, un clasificador basado en modelos o un clasificador en bucle abierto) para seleccionar el primer codificador 132 o el segundo codificador 134 para cada trama. Al usar el codificador seleccionado para codificar tramas de audio, los artefactos y la mala calidad de la señal que resultan de una clasificación errónea de las tramas de audio y del uso del codificador incorrecto para codificar las tramas de audio pueden reducirse o eliminarse.

**[0051]** Varios ejemplos de código informático que ilustran posibles implementaciones de aspectos descritos con respecto a las FIGS. 1-4 se presentan a continuación. En los ejemplos, el término "st->" indica que la variable que sigue al término es un parámetro de estado (por ejemplo, un estado del codificador 104 de la FIG. 1, un estado del selector 120 de la FIG. 1, o una combinación de los mismos). Por ejemplo, "st->lps" indica que la probabilidad a corto plazo de que la trama de entrada sea una trama de voz ("lps") es un parámetro de estado. Los siguientes ejemplos corresponden a una implementación basada en el sistema 100 de la FIG. 1, los ejemplos de las FIGS. 2-4, o ambos, y donde el primer clasificador 122 es un clasificador basado en modelos, el segundo clasificador 124 es un clasificador en bucle abierto, el primer codificador 132 incluye un codificador ACELP y el segundo codificador 134 incluye un codificador TCX.

**[0052]** El código informático incluye comentarios que no son parte del código ejecutable. En el código informático, el comienzo de un comentario se indica mediante una barra diagonal y un asterisco (por ejemplo, "/\*") y el final del comentario se indica mediante un asterisco y una barra diagonal (por ejemplo, "\*/"). Para ilustrar, un comentario "COMENTARIO" puede aparecer en el pseudocódigo como /\*COMENTARIO\*/.

**[0053]** En los ejemplos proporcionados, el operador "==" indica una comparación de igualdad, de modo que "A==B" tiene un valor de VERDADERO cuando el valor de A es igual al valor de B y tiene un valor de FALSO de lo contrario. El operador "&&" indica una operación AND lógica. El operador "||" indica una operación OR lógica. El operador ">" (mayor que) representa "mayor que", el operador ">=" representa "mayor o igual que", y el operador "<" indica "menor que". El término "f" que sigue a un número indica un formato de número de coma flotante (por ejemplo, decimal). Como se señaló previamente, el término "st->A" indica que A es un parámetro de estado (es decir, los caracteres "st->" no representan una operación lógica o aritmética).

**[0054]** En los ejemplos proporcionados, "\*" puede representar una operación de multiplicación, "+" o "sum" puede representar una operación de suma, "-" puede indicar una operación de resta y "/" puede representar una operación de división. El operador "=" representa una asignación (por ejemplo, "a=1" asigna el valor de 1 a la variable "a"). Otras implementaciones pueden incluir una o más condiciones además o en lugar del conjunto de condiciones del Ejemplo 1.

**[0055]** La condición "st->lps > st->lpm" indica que la probabilidad a corto plazo de que la trama actual sea similar a voz es mayor que la probabilidad a corto plazo de que la trama actual sea similar a música, según lo calculado por el clasificador basado en modelos. Estos son parámetros intermedios cuyos valores pueden proporcionarse o pasarse al segundo clasificador 124 antes de que el procesamiento en la máquina de estados 374 tenga lugar en el primer clasificador 122 (por ejemplo, el clasificador basado en modelos).

**[0056]** Por ejemplo, lps puede corresponder a la probabilidad logarítmica de voz dadas las características observadas, y lpm puede corresponder a la probabilidad logarítmica de música dadas las características observadas. Por ejemplo,

$$[\text{Ecuación 1}]: \text{lps} = \log(p(\text{voz}|\text{características}) * p(\text{características})) = \log(p(\text{características}|\text{voz}) + \log(\text{voz}),$$

y

$$\text{[Ecuación 2]: } \log(p(\text{música}|\text{características}) * p(\text{características})) = \log(p(\text{características}|\text{música})) + \log(\text{música}),$$

5 donde  $p(x)$  indica una probabilidad de  $x$  y  $p(x|y)$  indica la probabilidad de  $x$ , dado  $y$ . En algunas implementaciones, cuando se realizan comparaciones relativas entre  $lps$  y  $lpm$ , se puede ignorar  $p(\text{características})$  porque es un término común. El término  $p(\text{características}|\text{voz})$  es la probabilidad de que el conjunto de características observadas suponga que las características pertenecen a voz. El término  $p(\text{características}|\text{voz})$  se puede calcular en base a un modelo de voz. El término  $p(\text{voz})$  es la probabilidad previa de voz. En general,  $p(\text{voz}) > p(\text{música})$  para aplicaciones de comunicación móvil porque la probabilidad de que alguien esté hablando por teléfono puede ser mayor que la probabilidad de que se escuche música en el teléfono. Sin embargo, en casos de uso alternativo,  $p(\text{voz})$  y  $p(\text{música})$  podrían estar relacionadas arbitrariamente.

15 **[0057]** Los parámetros  $lps$  y  $lpm$  pueden indicar la probabilidad de que un conjunto de características observadas sean voz y música, respectivamente, con información sobre modelos de voz, modelos de música o una combinación de los mismos, junto con probabilidades anteriores de voz y música.

20 **[0058]** La condición " $st \rightarrow sr\_core == 12800$ " puede indicar un codificador o un modo de funcionamiento del codificador (por ejemplo, una frecuencia de muestreo central ACELP de 12,8 kHz). Por ejemplo, en algunas implementaciones, un modo operativo de codificador de 12,8 kHz puede exhibir una mayor predicción errónea de voz/música en comparación con los modos operativos de codificador de frecuencia de muestreo más alta.

25 **[0059]** La condición " $sp\_aud\_decision0 == 0$ " puede indicar que la decisión de voz/música del primer clasificador 122 indica que la trama de entrada es una trama de voz. La decisión de voz/música del primer clasificador 122 se genera después de que se calculan los parámetros basados en modelos  $lps$  y  $lpm$  y después de que se completa el procesamiento de la máquina de estados 374 (que considera la información a largo plazo para que  $sp\_aud\_decision$  evite la conmutación frecuente).

30 **[0060]** El término " $st \rightarrow acelpFramesCount$ " indica un recuento del número de tramas consecutivas pasadas que se decidió que eran ACELP (o voz). Este recuento puede usarse para desviar la decisión hacia la voz cuando el número de tramas ACELP consecutivas pasadas es relativamente alto. El uso de este recuento para desviar la decisión puede proporcionar un mayor efecto de desviación en casos límite, como cuando  $lps$  tiene un valor similar al valor de  $lpm$  y cuando  $snr\_acelp$  tiene un valor similar al valor de  $snr\_tcx$ . Esto también evita el cambio frecuente entre ACELP/TCX.

35 **[0061]** Se puede evaluar un conjunto de condiciones para determinar si se debe desviar una decisión de voz/música estableciendo un valor del parámetro de ajuste " $dsnr$ " como se indica en el Ejemplo 1.

```
40 if((st->sr_core == 12800) && ((st->lps > st->lpm) || (st->acelpFramesCount >= 6
&& (st->lps > st->lpm - 1.5f))) && (sp_aud_decision0 == 0) && (snr_acelp >=
snr_tcx - 4) && st->acelpFramesCount >= 1)
```

```
{
    dsnr = 4.0f; /*Para desviar la decisión hacia ACELP*/
}
```

#### 45 Ejemplo 1

50 **[0062]** Cabe señalar que  $st \rightarrow acelpFramesCount \geq 1$  indica que la última trama (es decir, la trama que precede a la trama que se está evaluando actualmente) se determinó que era una trama ACELP (por ejemplo, los segundos datos de decisión 148 indican una selección del primer codificador 132). Si se determinó que la última trama (la trama anterior) es una trama ACELP, el conjunto de condiciones del Ejemplo 1 también incluye una comprobación de  $st \rightarrow lps > st \rightarrow lpm$ . Sin embargo, si se determinó que las últimas 6 tramas consecutivas son tramas ACELP, el conjunto de condiciones del Ejemplo 1 permite ajustar el parámetro de ajuste " $dsnr$ " para una trama actual para desviar una selección hacia la trama actual que es una trama ACELP incluso si  $st \rightarrow lps$  es menor que  $st \rightarrow lpm$ , siempre y cuando el valor de  $st \rightarrow lps$  esté dentro de 1,5 del valor de  $st \rightarrow lpm$ . También se debe tener en cuenta que  $st \rightarrow acelpFramesCount \geq 6$  indica que al menos las últimas 6 tramas se determinaron como tramas ACELP (por ejemplo, los segundos datos de decisión 148 indican una selección del primer codificador 132) e indica implícitamente que se determinó que la última trama (es decir, la trama que precede a la trama que se está evaluando actualmente) es una trama ACELP. Para ilustrar, en algunas implementaciones, un valor de  $st \rightarrow lps$  puede estar típicamente entre -27 y 27, y un valor de  $st \rightarrow lpm$  puede estar típicamente entre -16 y 23.

60 **[0063]** Cabe señalar que incluso después de la modificación del parámetro de ajuste (por ejemplo,  $dsnr = 4,0f$ ) como se aplica en el Ejemplo 1, en algunas implementaciones el valor del parámetro de ajuste se puede ajustar más (por ejemplo, aumentar o disminuir) antes de aplicarlo durante la decisión de voz/música del generador de decisión de

clasificación 418. Por lo tanto, la modificación del parámetro de ajuste "dsnr" en el Ejemplo 1 aumenta la probabilidad de seleccionar voz/ACELP, aunque no necesariamente lo garantiza, cuando se cumple el conjunto de condiciones del Ejemplo 1.

5 **[0064]** Otras implementaciones pueden incluir una o más condiciones además o en lugar del conjunto de condiciones del Ejemplo 1. Por ejemplo, el parámetro "non\_staX" puede indicar una medida de la varianza absoluta en las energías en varias bandas de frecuencia entre la trama actual y la pasada. En el dominio logarítmico, non\_staX puede ser la suma de las diferencias absolutas de energía logarítmica entre las tramas actuales y pasadas entre las diferentes bandas. En el ejemplo 2 se proporciona un ejemplo de cálculo de un valor del parámetro non\_staX.

```
10 for(band_i = band_start; i < band-stop; i++)
/*bucle desde band_start a band-stop*/
{
15     log_enr = log(enr[band_i]);
*non_staX = *non_staX + abs(log_enr - st->past_log_enr[band_i]);
st->past_log_enr[band_i] = log_enr;
}
```

Ejemplo 2

20 **[0065]** Las señales de música, especialmente las señales instrumentales (por ejemplo, violín) tienen un grado muy alto de estacionaridad en todas las bandas de frecuencia, pero a veces pueden confundirse con voz sonora debido a su alta armonía. Se puede usar una condición de no estacionaridad relativamente alta para reducir la probabilidad de codificar señales instrumentales estacionarias como voz (por ejemplo, con un codificador ACELP).

25 **[0066]** Como otro ejemplo, una condición basada en voz promedio "mean(voicing\_fr, 4) >= 0.3" puede cumplirse cuando una media aritmética de valores del parámetro voicing\_fr dentro de las cuatro subtramas de la trama actual es mayor o igual que 0,3. Aunque se consideran cuatro subtramas, que pueden corresponder a todas las subtramas de una trama, en otras implementaciones se pueden considerar menos de cuatro subtramas. El parámetro voicing\_fr se puede determinar como:

[Ecuación 3]:  $voicing\_fr[i] =$

$$\frac{\sum_{k=0}^{l\_subfr} speech\_subfr_i[k - \tau_i].speech\_subfr_i[k]}{\sum_{k=0}^{l\_subfr} (speech\_subfr_i[k])^2}$$

35 **[0067]** En la ecuación 3,  $\tau_i$  es el período de tono estimado en la subtrama i. Voicing\_fr[i] es el parámetro de voz para la subtrama i. Voicing\_fr[i] que tiene un valor de 1 indica que una correlación entre la voz en la subtrama actual y el conjunto de muestras  $\tau_i$  es muy alta y un valor 0 significa que la correlación es muy baja. Voicing\_fr puede ser una medida de la repetición de voz. Una trama sonora es altamente repetitiva y la condición "mean(voicing\_fr, 4) > 0.3" puede cumplirse para señales de voz.

40 **[0068]** Como otro ejemplo, se puede usar una condición basada en el parámetro de desplazamiento "desplazamiento < 74,0f" cuando se determina si desviar la decisión de voz/música hacia voz. El parámetro de desplazamiento está inversamente relacionado con snr\_tcx, lo que significa que un aumento en el valor de desplazamiento conduciría a una disminución en snr\_tcx y viceversa, y restringir el parámetro de desplazamiento para que tenga un valor bajo restringe indirectamente a que snr\_tcx tenga un nivel que exceda un nivel inferior para la codificación TCX efectiva. Cabe señalar que el parámetro de desplazamiento se calcula dentro del segundo clasificador basándose en estado a largo plazo, las características a corto plazo, etc. En una implementación, la relación entre snr\_tcx y desplazamiento puede ser:

[Ecuación 4]:  $snr\_tcx = \alpha \frac{\sum_{i=0}^{63} [s_h(n)]^2}{10^{\left(\frac{desplazamiento}{10}\right)}}$  (donde  $S_h$  es la voz ponderada

50 y donde se realiza la ponderación en los LPC de la voz de entrada)

o

$$\text{[Ecuación 5]: } snr_{tcx} = \alpha \frac{\sum_{i=0}^3 \log(\text{energy\_subframe}[i])}{10^{\left(\frac{\text{desplazamiento}}{10}\right)}}.$$

5 **[0069]** Como otro ejemplo, una decisión de voz/música puede estar desviada hacia música cuando “sp\_aud\_decision0 == 1” (por ejemplo, los primeros datos de decisión 146 indican una trama de música) para reducir la aparición de tramas ACELP en una señal de música, como se ilustra en el Ejemplo 3.

```

10 if((st->sr_core == 12800) && sp_aud_decision0 == 1)
    {
        /* Aumenta la probabilidad de elegir TCX, desviando la decisión hacia TCX*/
    } dsnr = -2.0f;

```

### Ejemplo 3

15 **[0070]** En el Ejemplo 4 se proporciona un conjunto ampliado de condiciones propuestas en comparación con el Ejemplo 1 para desviar la decisión del segundo clasificador 124 hacia ACELP o TCX.

```

    if((st->sr_core == 12800) && ((st->lps > st->lpm) && mean(voicing_fr, 4) >= 0.3f ||
    (st->acelpFramesCount >= 6 && (st->lps > st->lpm - 1.5f))) && (sp_aud_decision0
    == 0) && (non_staX > 5.0f) && (snr_acelp >= snr_tcx - 4) && st-
    >acelpFramesCount >= 1 && (offset < 74.0f))
    {
        /*Desvía la decisión hacia ACELP/VOZ */
        dsnr = 4.0f;
    }
    else if((st->sr_core == 12800) && sp_aud_decision0 == 1)
    {
        /* Aumenta la probabilidad de elegir TCX, desviando la decisión hacia TCX*/
        dsnr = -2.0f;
    }

```

### Ejemplo 4

20 **[0071]** En el Ejemplo 5 se proporciona otro conjunto de condiciones propuestas para desviar la decisión del segundo clasificador 124 hacia ACELP o TCX. En el Ejemplo 5, la mean(voicing\_fr, 4) que es mayor que 0,3 es una condición independiente.

```

if( (st->sr_core == 12800) && mean(voicing_fr, 4) >= 0.3f && ((st->lps > st->lpm) ||
(st->acelpFramesCount >= 6 && (st->lps > st->lpm - 1.5f))) && (sp_aud_decision0
== 0) && (non_staX > 5.0f) && (snr_acelp >= snr_tcx - 4) && st-
>acelpFramesCount >= 1 && (offset < 74.0f))
{
    /*Desvía la decisión hacia ACELP/VOZ */
    dsnr = 4.0f;
}
else if((st->sr_core == 12800) && sp_aud_decision0 == 1)
{
    /* Aumenta la probabilidad de elegir TCX, desviando la decisión hacia TCX*/
    dsnr = -2.0f;
}

```

#### Ejemplo 5

- 5 **[0072]** Aunque los ejemplos 1 y 3-5 proporcionan ejemplos de conjuntos de condiciones correspondientes a los valores de configuración del parámetro de ajuste "dsnr", otras implementaciones pueden excluir una o más condiciones, incluir una o más condiciones distintas, o cualquier combinación de las mismas. Por ejemplo, aunque los ejemplos 1 y 3-5 incluyen la condición "st->sr\_core == 12800", que indica un modo de funcionamiento del codificador (por ejemplo, una frecuencia de muestreo de 12,8 kHz) que puede exhibir un aumento de la predicción errónea de voz/música, en otras implementaciones, uno o más modos de codificador, o modo no de codificador, pueden incluirse en un conjunto de condiciones para establecer el parámetro de ajuste. Aunque los valores numéricos (por ejemplo, 74,0f) se proporcionan en algunos de los ejemplos, dichos valores se proporcionan solo como ejemplos, y se pueden determinar otros valores para proporcionar predicciones erróneas reducidas en otras implementaciones. Además, las indicaciones de parámetros (por ejemplo, "lps", "lpm", etc.) utilizadas en el presente documento son solo ilustrativas.
- 10 En otras implementaciones, los parámetros pueden ser referidos por diferentes nombres. Por ejemplo, el parámetro de probabilidad de voz puede ser referido por "prob\_s" o "lp\_prob\_s". Además, se han descrito parámetros promediados en el tiempo (paso bajo) (denominados "lp"), las FIGS. 1-4 y los Ejemplos 1 y 3-5 podrían usar otros parámetros (por ejemplo, "prob\_s", "prob\_m", etc.) en lugar de un parámetro promediado en el tiempo o de paso bajo.
- 15
- 20 **[0073]** La FIG. 5 es un diagrama de flujo que ilustra un procedimiento 500 para determinar una selección de un codificador. El procedimiento 500 puede realizarse en un codificador o mediante un codificador que selecciona entre codificación de voz y codificación no de voz. Por ejemplo, se puede realizar el procedimiento 500 en el codificador 104 de la FIG. 1.
- 25 **[0074]** El procedimiento 500 incluye recibir, desde un primer clasificador, primeros datos de probabilidad y segundos datos de probabilidad en un segundo clasificador, en 502. Los primeros datos de probabilidad están asociados con una primera probabilidad de que una trama de audio sea una trama de voz y los segundos datos de probabilidad están asociados con una segunda probabilidad de que la trama de audio sea una trama no de voz. Para ilustrar, los primeros datos de probabilidad 142 y los segundos datos de probabilidad 144 se reciben en el segundo clasificador 124 desde el primer clasificador 122 de la FIG. 1. Por ejemplo, el primer clasificador puede estar asociado con un clasificador basado en modelos, y el segundo clasificador puede estar asociado con un modelo en bucle abierto o un clasificador en bucle abierto.
- 30 **[0075]** Los primeros datos de decisión pueden recibirse desde el primer clasificador en el segundo clasificador, los primeros datos de decisión que indican una clasificación de la trama de audio como una trama de voz o una trama no de voz, en 504. Los primeros datos de decisión pueden recibirse en el segundo clasificador desde una máquina de estados del primer clasificador. Por ejemplo, los primeros datos de decisión pueden corresponder a los primeros datos de decisión 146 de la FIG. 1.
- 35 **[0076]** El procedimiento 500 también incluye determinar, en el segundo clasificador, los segundos datos de decisión basándose en los primeros datos de probabilidad, los segundos datos de probabilidad y los primeros datos de decisión, en 506. Los segundos datos de decisión se configuran para indicar una selección de un codificador particular de múltiples codificadores para codificar la trama de audio. Por ejemplo, los múltiples codificadores pueden incluir un
- 40

primer codificador y un segundo codificador, como el primer codificador 132 y el segundo codificador 134 de la FIG. 1, respectivamente. El primer codificador puede incluir un codificador de voz, y el segundo codificador puede incluir un codificador no de voz. Para ilustrar, el codificador no de voz puede incluir un codificador de música, como un codificador TCX.

5 **[0077]** El procedimiento 500 puede incluir proporcionar los segundos datos de decisión desde una salida del segundo clasificador a un conmutador configurado para seleccionar un codificador particular de los múltiples codificadores. La trama de audio se codifica utilizando el codificador seleccionado. Por ejemplo, el segundo clasificador 124 de la FIG. 1 puede emitir los segundos datos de decisión 148 que se proporcionan al conmutador 130 para seleccionar uno del primer codificador 132 o el segundo codificador 134.

15 **[0078]** El procedimiento 500 puede incluir determinar un primer valor de ganancia de codificación estimada asociado con un primer codificador de los múltiples codificadores que se usa para codificar la trama de audio y determinar un segundo valor de ganancia de codificación estimada asociado con un segundo codificador de los múltiples codificadores que se usa para codificar la trama de audio. Por ejemplo, el primer valor de ganancia de codificación estimada puede corresponder a un valor (por ejemplo, `snr_acelp`) emitido por el estimador de probabilidad de voz a corto plazo 410 de la FIG. 4 y el segundo valor de ganancia de codificación estimada puede corresponder a un valor (por ejemplo, `snr_tcx`) emitido por el estimador de probabilidad de música a corto plazo 412. El procedimiento 500 puede incluir ajustar el primer valor de ganancia de codificación estimada basándose en un valor de un parámetro de ajuste. Por ejemplo, un valor del parámetro de ajuste "dsnr" en la FIG. 4 puede ser emitido por el generador de parámetros de ajuste 416 de la FIG. 4 y puede ser utilizado por el generador de decisión de clasificación 418 para ajustar un valor de `snr_acelp`. La selección del uno o más codificadores puede basarse en el primer valor de ganancia de codificación estimada ajustado y el segundo valor de ganancia de codificación estimada.

25 **[0079]** El procedimiento 500 puede incluir seleccionar un valor del parámetro de ajuste (por ejemplo, "dsnr"). El valor puede seleccionarse basándose en al menos uno de los primeros datos de probabilidad (por ejemplo, `lps`), los segundos datos de probabilidad (por ejemplo, `lpm`), datos de estado a largo plazo o la primera decisión (por ejemplo, `sp_aud_decision`). Por ejemplo, el generador de parámetros de ajuste 416 de la FIG. 4 puede seleccionar un valor del parámetro de ajuste. Los segundos datos de decisión pueden determinarse basándose en el valor del parámetro de ajuste, tal como la salida del generador de decisión de clasificación 418 que responde al valor del parámetro de ajuste que se recibe del generador de parámetros de ajuste 416. El valor del parámetro de ajuste puede seleccionarse para desviar la selección hacia un primer codificador asociado con voz o un segundo codificador asociado con no voz.

35 **[0080]** El procedimiento 500 puede incluir determinar si se cumple un conjunto de condiciones asociadas con una trama de audio y, en respuesta al conjunto de condiciones que se satisfacen, seleccionar un valor de un parámetro de ajuste para desviar la selección hacia un primer codificador asociado con voz. Se puede determinar que el conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que la trama de audio está asociada con una frecuencia de muestreo central de 12,8 kHz, como la condición "`st->sr_core == 12800`" en el Ejemplo 1. Se puede determinar que el conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que los primeros datos de decisión indican que la trama de audio se clasifica como trama de voz, como la condición "`sp_aud_decision0 == 0`" en el Ejemplo 1. Se puede determinar que el conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de un primer valor de ganancia de codificación estimada asociado con el primer codificador (por ejemplo, `snr_acelp`) que se utiliza para codificar la trama de audio es mayor o igual que un primer valor. El primer valor puede estar asociado con una diferencia entre un segundo valor de ganancia de codificación estimada (por ejemplo, `snr_tcx`) y un segundo valor (por ejemplo, 4), como la condición "`snr_acelp >= snr_tcx - 4`" en el Ejemplo 1. Se puede determinar que el conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que una trama clasificada más recientemente se clasifica como que incluye contenido de voz (por ejemplo, "`st->acelpFramesCount >= 1`" en el Ejemplo 1) y a la determinación de que un primer valor de probabilidad indicado por los primeros datos de probabilidad es mayor que un segundo valor de probabilidad indicado por la segunda probabilidad (por ejemplo, "`st->lps > st->lpm`" del Ejemplo 1).

55 **[0081]** Se puede determinar que el conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que cada trama correspondiente a un número de tramas clasificadas más recientemente se clasifica como que incluye contenido de voz (por ejemplo, "`st->acelpFramesCount >= 6`" en el Ejemplo 1) y a la determinación de que un primer valor de probabilidad indicado por los primeros datos de probabilidad (por ejemplo, "`st->lps`") es mayor o igual que un tercer valor (por ejemplo, "`st->lpm - 1,5`" en el Ejemplo 1). El tercer valor puede estar asociado con una diferencia entre un segundo valor de probabilidad indicado por los segundos datos de probabilidad (por ejemplo, "`st->lpm`") y un cuarto valor (por ejemplo, 1,5).

60 **[0082]** Se puede determinar que el conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que un valor promedio de voz de múltiples subtramas de la trama de audio es mayor o igual que un primer umbral (por ejemplo, "`mean(voicing_fr, 4) >= 0.3`" en el Ejemplo 4), la determinación de un valor de no estacionaridad asociado con la trama de audio es mayor que un segundo umbral (por ejemplo, "`non-staX > 5,0`" en el Ejemplo 4), y la determinación de un valor de desplazamiento asociado con la trama de audio es inferior a un tercer umbral (por ejemplo, "`desplazamiento < 74`" en el Ejemplo 4).

65

**[0083]** En un aspecto particular, el procedimiento 500 incluye determinar si se cumple un segundo conjunto de condiciones asociadas con una trama de audio y, en respuesta a que se cumple el segundo conjunto de condiciones, se selecciona un valor de un parámetro de ajuste para desviar la selección hacia un segundo codificador asociado con no voz, tal como se describe con respecto al Ejemplo 3. Se puede determinar que el segundo conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que la trama de audio está asociada con una frecuencia de muestreo central de 12,8 kHz (por ejemplo, "st->st\_core == 12800" en el Ejemplo 3). De forma alternativa o, además, se puede determinar que el segundo conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que los primeros datos de decisión indican que la trama de audio se clasifica como trama no de voz (por ejemplo, "sp\_aud\_decision0 == 1" en el Ejemplo 3).

**[0084]** El procedimiento 500 puede permitir una clasificación más precisa de una trama de audio particular y una selección mejorada de un codificador utilizado para codificar la trama de audio particular. Al usar los datos de probabilidad y los primeros datos de decisión del primer clasificador para determinar la selección, las tramas de audio se pueden clasificar con precisión como tramas de voz o tramas de música y se puede reducir un número de tramas de voz mal clasificadas en comparación con las técnicas de clasificación convencionales. En base a las tramas de audio clasificadas, se puede seleccionar un codificador (por ejemplo, un codificador de voz o un codificador no de voz) para codificar la trama de audio. Al usar el codificador seleccionado para codificar las tramas de voz, se pueden reducir los artefactos y la mala calidad de la señal que resultan de una clasificación errónea de las tramas de audio y del uso del codificador incorrecto para codificar las tramas de audio.

**[0085]** La FIG. 6 es un diagrama de flujo que ilustra un procedimiento 600 para seleccionar un valor de un parámetro de ajuste para desviar una selección hacia un codificador particular. El procedimiento 600 se puede realizar en o por un codificador que selecciona entre codificación de voz y codificación no de voz. Por ejemplo, se puede realizar el procedimiento 600 en el codificador 104 de la FIG. 1.

**[0086]** Los primeros datos de probabilidad y los primeros datos de decisión de un primer clasificador se reciben en un segundo clasificador, en 602. Los primeros datos de probabilidad están asociados con una primera probabilidad de que una trama de audio sea una trama de voz. Por ejemplo, los primeros datos de probabilidad pueden corresponder a los primeros datos de probabilidad 142, los segundos datos de probabilidad 144, o una combinación de los mismos, recibidos en el segundo clasificador 124 desde el primer clasificador 122 de la FIG. 1. Los primeros datos de decisión indican una clasificación de la trama de audio como una trama de voz o una trama no de voz, tal como los primeros datos de decisión 146 de la FIG. 1.

**[0087]** El procedimiento 600 también incluye determinar, en el segundo clasificador, si se cumple un conjunto de condiciones asociadas con la trama de audio, en 604. Una primera condición del conjunto de condiciones se basa en los primeros datos de probabilidad y una segunda condición del conjunto de condiciones se basa en los primeros datos de decisión. Por ejemplo, la primera condición puede corresponder a "st->lps > st->lpm" en el Ejemplo 1, y la segunda condición puede corresponder a "sp\_aud\_decision0 == 0" en el Ejemplo 1.

**[0088]** El procedimiento 600 incluye, además, en respuesta a la determinación de que se satisface el conjunto de condiciones, establecer un valor de un parámetro de ajuste para desviar una primera selección hacia un primer codificador de múltiples codificadores, en 606. Por ejemplo, el valor del parámetro de ajuste puede corresponder a un valor de una salida del generador de parámetros de ajuste 416 de la FIG. 4 que se proporciona al generador de decisiones de clasificación 418. Para ilustrar, establecer el valor del parámetro de ajuste para desviar la primera selección hacia el primer codificador puede corresponder a establecer (o actualizar) un valor del parámetro de ajuste, como "dnrsr = 4,0" en el Ejemplo 1. El primer codificador puede incluir o corresponder a un codificador de voz.

**[0089]** En un aspecto particular, se determina que el conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que la trama de audio está asociada con una frecuencia de muestreo de 12.800 kHz (por ejemplo, "st->sr\_core == 12800" en el Ejemplo 1). Se puede determinar que el conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que los primeros datos de decisión indican la clasificación de la trama de audio como trama de voz (por ejemplo, "spaud\_decision0 == 0" en el Ejemplo 1). Se puede determinar que el conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de un primer valor de ganancia de codificación estimada asociado con la codificación de la trama de audio en el primer codificador (por ejemplo, "snr\_acelp") es mayor o igual que un primer valor, el primer valor asociado con una diferencia entre un segundo valor de ganancia de codificación estimada (por ejemplo, "snr\_tcx") y un segundo valor (por ejemplo, "snr\_acelp >= snr\_tcx - 4" en el Ejemplo 1).

**[0090]** En un aspecto particular, se determina que el conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que una trama clasificada más recientemente se clasifica como que incluye contenido de voz (por ejemplo, "st->acelpFramesCount >= 1" en el Ejemplo 1). En un aspecto particular, se determina que el conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que un primer valor de probabilidad indicado por los primeros datos de probabilidad es mayor que un segundo valor de probabilidad indicado por los segundos datos de probabilidad (por ejemplo, "st->lps > st->lpm"), los segundos datos de probabilidad asociados con una segunda probabilidad de que la trama de audio sea una trama no de voz.

**[0091]** Se puede determinar que el conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que cada trama correspondiente a un número de tramas clasificadas más recientemente se clasifica como que incluye contenido de voz (por ejemplo, "st->acelpFramesCount >= 6"). Se puede determinar que el conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que un primer valor de probabilidad indicado por los primeros datos de probabilidad (por ejemplo, "st->lps") es mayor o igual que un tercer valor, el tercer valor asociado con una diferencia entre un segundo valor de probabilidad indicado por los segundos datos de probabilidad (por ejemplo, "st->lpm") y un cuarto valor, como la condición "st->lps > st-lpm - 1,5" en el Ejemplo 1. Los segundos datos de probabilidad pueden estar asociados con una segunda probabilidad de que la trama de audio sea una trama no de voz.

**[0092]** Se puede determinar que el conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que un valor promedio de voz de múltiples subtramas de la trama de audio es mayor o igual a un primer umbral (por ejemplo, "mean(voicing\_fr, 4) >= 0.3" en el Ejemplo 4). Se puede determinar que el conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que un valor de no estacionaridad asociado con la trama de audio es mayor que un segundo umbral (por ejemplo, "non\_staX > 5.0" en el Ejemplo 4). Se puede determinar que el conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que un valor de desplazamiento asociado con la trama de audio es menor que un tercer umbral (por ejemplo, "desplazamiento < 74.0" en el Ejemplo 4).

**[0093]** En algunas implementaciones, el procedimiento 600 puede incluir determinar si se cumple un segundo conjunto de condiciones asociadas con la trama de audio, tal como el conjunto de condiciones del Ejemplo 3. El procedimiento 600 también puede incluir, en respuesta a determinar que se cumple el segundo conjunto de condiciones, actualizar el valor del parámetro de ajuste del primer valor a un segundo valor para desviar una segunda selección hacia un segundo codificador de los múltiples codificadores, incluyendo el segundo codificador un codificador no de voz. Por ejemplo, la actualización del valor del parámetro de ajuste para desviar una segunda selección hacia el segundo codificador puede realizarse estableciendo un valor de la salida del generador de parámetros de ajuste 416 de la FIG. 4 (por ejemplo, "dsnr = -2.0" en el ejemplo 3). Para ilustrar, se puede determinar que el segundo conjunto de condiciones se cumple en respuesta a la determinación de que la trama de audio está asociada con una frecuencia de muestreo de 12,8 kHz y la determinación de que los primeros datos de decisión indican la clasificación de la trama de audio como trama no de voz (por ejemplo, "(st->sr\_core == 12800) && (sp\_aud\_decision0 == 1)" en el Ejemplo 3).

**[0094]** Al usar el parámetro de ajuste para determinar la selección, las tramas de audio se pueden clasificar como tramas de voz o tramas de música y se puede reducir un número de tramas de voz mal clasificados en comparación con las técnicas de clasificación convencionales. En base a las tramas de audio clasificadas, se puede seleccionar un codificador (por ejemplo, un codificador de voz o un codificador no de voz) para codificar la trama de audio. Al usar el codificador seleccionado para codificar las tramas de voz, se pueden reducir los artefactos y la mala calidad de la señal que resultan de una clasificación errónea de las tramas de audio y del uso del codificador incorrecto para codificar las tramas de audio.

**[0095]** En aspectos particulares, uno o más de los procedimientos de las FIGS. 5-6, los Ejemplos 1-5, o una combinación de los mismos, pueden implementarse mediante un dispositivo de matriz de puertas programables *in situ* (FPGA), un circuito integrado específico de la aplicación (ASIC), una unidad de procesamiento tal como una unidad de procesamiento central (CPU), un procesador de señales digitales (DSP), un controlador, otro dispositivo de hardware, dispositivo de firmware o cualquier combinación de los mismos. Como ejemplo, se pueden realizar uno o más de los procedimientos de las FIGS. 5-6, los ejemplos 1-5 o una combinación de los mismos, mediante un procesador que ejecute instrucciones, como se describe con respecto a la FIGS. 7 y 8. Para ilustrar, una parte del procedimiento 500 de la FIG. 5 puede combinarse con una segunda porción de la FIG. 6 o con una tercera porción del Ejemplo 1.

**[0096]** Con referencia a la FIG. 7, se representa un diagrama de bloques de un ejemplo ilustrativo particular de un dispositivo 700 (por ejemplo, un dispositivo de comunicación inalámbrica). En diversas implementaciones, el dispositivo 700 puede tener menos o más componentes de los ilustrados en la FIG. 7. En un ejemplo ilustrativo, el dispositivo 700 corresponder al sistema 102 de la FIG. 1. En un ejemplo ilustrativo, el dispositivo 700 puede funcionar de acuerdo con uno o más de los procedimientos de las FIGS. 5-6, uno o más de los ejemplos 1-5 o una combinación de los mismos.

**[0097]** En un ejemplo particular, el dispositivo 700 incluye un procesador 706 (por ejemplo, una CPU). El dispositivo 700 puede incluir uno o más procesadores adicionales, tales como un procesador 710 (por ejemplo, un DSP). El procesador 710 puede incluir un codificador-descodificador (CÓDEC) de audio 708. Por ejemplo, el procesador 710 puede incluir uno o más componentes (por ejemplo, circuitos) configurados para realizar operaciones del CÓDEC 708 de audio. Como otro ejemplo, el procesador 710 puede configurarse para ejecutar una o más instrucciones legibles por ordenador para realizar las operaciones del audio CÓDEC 708. Aunque se ilustra el CÓDEC de audio 708 como un componente de los procesadores 710, en otros ejemplos se pueden incluir uno o más componentes del CÓDEC de audio 708 en el procesador 706, el CÓDEC 734, otro componente de procesamiento o una combinación de los mismos.

**[0098]** El CÓDEC de audio 708 puede incluir un codificador vocoder (codificador de voz) 736. El codificador vocoder 736 puede incluir un selector de codificador 760, un codificador de voz 762 y un codificador no de voz 764. Por ejemplo, el codificador de voz 762 puede corresponder al primer codificador 132 de la FIG. 1, el codificador no de voz 764 puede corresponder al segundo codificador 134 de la FIG. 1, y el selector de codificador 760 puede corresponder al selector 120 de la FIG. 1.

**[0099]** El dispositivo 700 puede incluir una memoria 732 y un CÓDEC 734. La memoria 732, como un dispositivo de almacenamiento legible por ordenador, puede incluir las instrucciones 756. Las instrucciones 756 pueden incluir una o más instrucciones que son ejecutables por el procesador 706, el procesador 710, o una combinación de los mismos, para realizar uno o más de los procedimientos de las FIGS. 5-6, los Ejemplos 1-5, o una combinación de los mismos. El dispositivo 700 puede incluir un controlador inalámbrico 740 acoplado (por ejemplo, a través de un tranceptor) a una antena 742.

**[0100]** El dispositivo 700 puede incluir una pantalla 728 acoplada a un controlador de pantalla 726. Se puede acoplar un altavoz 741, un micrófono 746 o ambos al CÓDEC 734. El CÓDEC 734 puede incluir un convertidor de digital a analógico (DAC) 702 y un convertidor de analógico a digital (ADC) 704. En un aspecto particular, el CÓDEC 734 puede recibir señales analógicas desde el micrófono 746, convertir las señales analógicas a señales digitales usando el ADC 704 y proporcionar las señales digitales al CÓDEC de audio 708. El CÓDEC de audio 708 puede procesar las señales digitales. En algunas implementaciones, el CÓDEC de audio 708 puede proporcionar señales digitales al CÓDEC 734. El CÓDEC 734 puede convertir las señales digitales en señales analógicas usando el DAC 702 y puede proporcionar las señales analógicas al altavoz 741.

**[0101]** El selector de codificador 760 puede usarse para implementar una implementación de hardware de la selección de codificador, que incluye la desviación de la selección del codificador mediante el establecimiento (o actualización) de un valor de un parámetro de ajuste basado en uno o más conjuntos de condiciones, como se describe en el presente documento. De forma alternativa o, además, se puede implementar una implementación de software (o una implementación combinada de software/hardware). Por ejemplo, las instrucciones 756 pueden ser ejecutables por el procesador 710 u otra unidad de procesamiento del dispositivo 700 (por ejemplo, el procesador 706, el CÓDEC 734, o ambos). Para ilustrar, las instrucciones 756 pueden corresponder a operaciones descritas como realizadas con respecto al selector 120 de la FIG. 1.

**[0102]** En una implementación particular, el dispositivo 700 puede estar incluido en un dispositivo de sistema en paquete o de sistema en chip 722. En una implementación particular, la memoria 732, el procesador 706, el procesador 710, el controlador de pantalla 726, el CÓDEC 734 y el controlador inalámbrico 740 pueden estar incluidos en un dispositivo de sistema en paquete o de sistema en chip 722. En una implementación particular, un dispositivo de entrada 730 y una fuente de alimentación 744 están acoplados al dispositivo de sistema en chip 722. Además, en una implementación particular, como se ilustra en la FIG. 7, la pantalla 728, el dispositivo de entrada 730, el altavoz 741, el micrófono 746, la antena 742 y la fuente de alimentación 744 son externos con respecto al dispositivo de sistema en chip 722. En una implementación particular, cada uno de la pantalla 728, el dispositivo de entrada 730, el altavoz 741, el micrófono 746, la antena 742 y la fuente de alimentación 744 se pueden acoplar a un componente del dispositivo de sistema en un chip 722, tal como una interfaz o un controlador.

**[0103]** El dispositivo 700 puede incluir un dispositivo de comunicación, un codificador, un descodificador, un teléfono inteligente, un teléfono celular, un dispositivo de comunicación móvil, un ordenador portátil, un ordenador, una tableta, un asistente digital personal (PDA), un descodificador, un reproductor de vídeo, una unidad de entretenimiento, un dispositivo de visualización, un televisor, una consola de juegos, un reproductor de música, una radio, un reproductor de vídeo digital, un reproductor de disco de vídeo digital (DVD), un sintonizador, una cámara, un dispositivo de navegación, un sistema descodificador, un sistema codificador, una estación base, un vehículo o cualquier combinación de los mismos.

**[0104]** En una implementación ilustrativa, el procesador 710 se pueden operar para realizar todos o una porción de los procedimientos u operaciones descritos con respecto a las FIGS. 1-6, Ejemplos 1-5, o una combinación de los mismos. Por ejemplo, el micrófono 746 puede capturar una señal de audio correspondiente a una señal de voz del usuario. El ADC 704 puede convertir la señal de audio capturada de una forma de onda analógica en una forma de onda digital que comprende muestras de audio digitales. El procesador 710 pueden procesar las muestras de audio digitales.

**[0105]** El codificador vocoder 736 puede determinar, trama por trama, si cada trama recibida de las muestras de audio digitales corresponde a datos de audio de voz o no de voz y puede seleccionar un codificador correspondiente (por ejemplo, el codificador de voz 762 o el codificador no de voz 764) para codificar la trama. Los datos de audio codificados generados en el codificador vocoder 736 pueden proporcionarse al controlador inalámbrico 740 para la modulación y transmisión de los datos modulados a través de la antena 742.

**[0106]** Por lo tanto, el dispositivo 700 puede incluir un dispositivo de almacenamiento legible por ordenador (por ejemplo, la memoria 732) que almacena instrucciones (por ejemplo, las instrucciones 756) que, cuando son ejecutadas

por un procesador (por ejemplo, el procesador 706 o el procesador 710), hacen que el procesador realice operaciones que incluyen determinar los primeros datos de probabilidad (por ejemplo, los primeros datos de probabilidad 142 de la FIG. 1) asociada con una primera probabilidad de que una trama de audio sea una trama de voz y determinar los segundos datos de probabilidad (por ejemplo, los segundos datos de probabilidad 144 de la FIG. 1) asociada con una segunda probabilidad de que la trama de audio sea una trama no de voz. Las operaciones también pueden incluir la determinación de los primeros datos de decisión (por ejemplo, los primeros datos de decisión 146 de la FIG. 1) en base a los primeros datos de probabilidad y los segundos datos de probabilidad. Los primeros datos de decisión indican una clasificación de la trama de audio como una trama de voz o una trama no de voz. Las operaciones también pueden incluir determinar los segundos datos de decisión (por ejemplo, los segundos datos de decisión 148 de la FIG. 1) basándose en los primeros datos de probabilidad, los segundos datos de probabilidad y los primeros datos de decisión. Los segundos datos de decisión indican una selección de un codificador (por ejemplo, el codificador de voz 762 o el codificador no de voz 764) para codificar la trama de audio.

**[0107]** En referencia a la FIG. 8, se representa un diagrama de bloques de un ejemplo ilustrativo particular de una estación base 800. En diversas implementaciones, la estación base 800 puede tener más o menos componentes de los ilustrados en la FIG. 8. En un ejemplo ilustrativo, la estación base 800 puede incluir el dispositivo 102 de la FIG. 1. En un ejemplo ilustrativo, la estación base 800 puede funcionar de acuerdo con uno o más de los procedimientos de las FIGS. 5-6, uno o más de los Ejemplos 1-5, o una combinación de los mismos.

**[0108]** La estación base 800 puede ser parte de un sistema de comunicación inalámbrica. Los sistemas de comunicación inalámbrica pueden incluir múltiples estaciones base y múltiples dispositivos inalámbricos. El sistema de comunicación inalámbrica puede ser un sistema de evolución a largo plazo (LTE), un sistema de acceso múltiple por división de código (CDMA), un sistema global para comunicaciones móviles (GSM), un sistema de red inalámbrica de área local (WLAN), o algún otro sistema inalámbrico. Un sistema CDMA puede implementar CDMA de banda ancha (WCDMA), CDMA IX, evolución de datos optimizados (EVDO), CDMA síncrono por división de tiempo (TD-SCDMA) o alguna otra versión de CDMA.

**[0109]** Los dispositivos inalámbricos también pueden denominarse un equipo de usuario (UE), una estación móvil, un terminal, un terminal de acceso, una unidad de abonado, una estación, etc. Los dispositivos inalámbricos pueden ser un teléfono móvil, un teléfono inteligente, una tableta, un módem inalámbrico, un asistente digital personal (PDA), un dispositivo portátil, un ordenador portátil, un smartbook, un netbook, un teléfono inalámbrico, una estación de bucle local inalámbrico (WLL), un dispositivo Bluetooth, etc. Los dispositivos inalámbricos pueden incluir o corresponder con el dispositivo 700 de la FIG. 7.

**[0110]** Uno o más componentes de la estación base 800 (y/o en otros componentes no mostrados) pueden realizar varias funciones, tales como enviar y recibir mensajes y datos (por ejemplo, datos de audio). En un ejemplo particular, la estación base 800 incluye un procesador 806 (por ejemplo, una CPU). La estación base 800 puede incluir un transcodificador 810. El transcodificador 810 puede incluir un CÓDEC de audio 808. Por ejemplo, el transcodificador 810 puede incluir uno o más componentes (por ejemplo, circuitería) configurados para realizar operaciones del CÓDEC de audio 808. Como otro ejemplo, el transcodificador 810 puede configurarse para ejecutar una o más instrucciones legibles por ordenador para realizar las operaciones del CÓDEC de audio 808. Aunque se ilustra el CÓDEC de audio 808 como un componente del transcodificador 810, en otros ejemplos se pueden incluir uno o más componentes del CÓDEC de audio 808 en el procesador 806, el CÓDEC 834, otro componente de procesamiento o una combinación de los mismos. Por ejemplo, un descodificador vocoder 838 puede incluirse en un procesador de datos del receptor 864. Como otro ejemplo, se puede incluir un codificador vocoder 836 en un procesador de datos de transmisión 866.

**[0111]** El transcodificador 810 puede funcionar para transcodificar mensajes y datos entre dos o más redes. El transcodificador 810 puede configurarse para convertir mensajes y datos de audio de un primer formato (por ejemplo, un formato digital) a un segundo formato. Para ilustrar, el descodificador vocoder 838 puede descodificar señales codificadas que tienen un primer formato y el codificador vocoder 836 puede codificar las señales descodificadas en señales codificadas que tienen un segundo formato. Adicionalmente o de forma alternativa, el transcodificador 810 puede configurarse para realizar la adaptación de la velocidad de transferencia de datos. Por ejemplo, el transcodificador 810 puede reducir la velocidad de transferencia de datos o aumentar la velocidad de transferencia de datos sin cambiar el formato de los datos de audio. Para ilustrar, el transcodificador 810 puede reducir las señales de 64 kbit/s en señales de 16 kbit/s.

**[0112]** El CÓDEC de audio 808 puede incluir un codificador vocoder 836 y/o el descodificador vocoder 838. El codificador vocoder 836 puede incluir un selector de codificador, un codificador de voz y un codificador no de voz, como se describe con respecto a la FIG. 7. El descodificador vocoder 838 puede incluir un selector de descodificador, un descodificador de voz y un descodificador no de voz.

**[0113]** La estación base 800 puede incluir una memoria 832. La memoria 832, como un dispositivo de almacenamiento legible por ordenador, puede incluir instrucciones. Las instrucciones pueden incluir una o más instrucciones que son ejecutables por el procesador 806, el transcodificador 810, o una combinación de los mismos, para realizar uno o más de los procedimientos de las FIGS. 5-6, los Ejemplos 1-5, o una combinación de los mismos.

La estación base 800 puede incluir múltiples transmisores y receptores (por ejemplo, transceptores), tales como un primer transceptor 852 y un segundo transceptor 854, acoplados a un conjunto de antenas. El conjunto de antenas puede incluir una primera antena 842 y/o una segunda antena 844. El conjunto de antenas se puede configurar para comunicarse de forma inalámbrica con uno o más dispositivos inalámbricos, como el dispositivo 700 de la FIG. 7. Por ejemplo, la segunda antena 844 puede recibir un flujo de datos 814 (por ejemplo, un flujo de bits) desde un dispositivo inalámbrico. El flujo de datos 814 puede incluir mensajes, datos (por ejemplo, datos de voz codificados) o una combinación de los mismos.

**[0114]** La estación base 800 puede incluir una conexión de red 860, tal como una conexión de retorno. La conexión de red 860 puede configurarse para comunicarse con una red central o una o más estaciones base de la red de comunicación inalámbrica. Por ejemplo, la estación base 800 puede recibir un segundo flujo de datos (por ejemplo, mensajes o datos de audio) desde una red central a través de la conexión de red 860. La estación base 800 puede procesar el segundo flujo de datos para generar mensajes o datos de audio y proporcionar los mensajes o los datos de audio a uno o más dispositivos inalámbricos a través de una o más antenas del conjunto de antenas u otra estación base a través de la conexión de red 860. En una implementación particular, la conexión de red 860 puede ser una conexión de red de área amplia (WAN), como un ejemplo ilustrativo, no limitante.

**[0115]** La estación base 800 puede incluir un desmodulador 862 que está acoplado a los transceptores 852, 854, el procesador de datos del receptor 864 y el procesador 806, y el procesador de datos del receptor 864 puede estar acoplado al procesador 806. El desmodulador 862 puede configurarse para desmodular señales moduladas recibidas de los transceptores 852, 854 y para proporcionar datos desmodulados al procesador de datos del receptor 864. El procesador de datos del receptor 864 puede configurarse para extraer un mensaje o datos de audio de los datos desmodulados y enviar el mensaje o los datos de audio al procesador 806.

**[0116]** La estación base 800 puede incluir un procesador de datos de transmisión 866 y un procesador de múltiples entradas-múltiples salidas (MIMO) de transmisión 868. El procesador de datos de transmisión 866 puede estar acoplado al procesador 806 y al procesador MIMO de transmisión 868. El procesador MIMO de transmisión 868 puede estar acoplado a los transceptores 852, 854 y al procesador 806. El procesador de datos de transmisión 866 puede configurarse para recibir los mensajes o los datos de audio del procesador 806 y codificar los mensajes o los datos de audio en base a un esquema de codificación, tal como CDMA o multiplexación por división de frecuencia ortogonal (OFDM), como ejemplos ilustrativos, no limitativos. El procesador de datos de transmisión 866 puede proporcionar los datos codificados al procesador de transmisión MIMO 868.

**[0117]** Los datos codificados pueden multiplexarse con otros datos, como datos piloto, utilizando técnicas CDMA u OFDM para generar datos multiplexados. Los datos multiplexados pueden entonces ser modulados (es decir, correlacionados con símbolos) por el procesador de datos de transmisión 866 basado en un esquema de modulación particular (por ejemplo, codificación de desplazamiento de fase binaria ("BPSK"), codificación de desplazamiento de fase en cuadratura ("QSPK"), Modulación de desplazamiento de fase M-aria ("M-PSK"), modulación de amplitud de cuadratura M-aria ("M-QAM"), etc.) para generar símbolos de modulación. En una implementación particular, los datos codificados y otros datos pueden modularse usando diferentes esquemas de modulación. La velocidad de transferencia de datos, la codificación y la modulación para cada flujo de datos pueden determinarse mediante instrucciones realizadas mediante el procesador 806.

**[0118]** El procesador de transmisión MIMO 868 puede configurarse para recibir los símbolos de modulación del procesador de datos de transmisión 866 y puede procesar adicionalmente los símbolos de modulación y puede realizar la conformación de haces en los datos. Por ejemplo, el procesador de transmisión MIMO 868 puede aplicar ponderaciones de conformación de haz a los símbolos de modulación. Las ponderaciones de conformación de haz pueden corresponder a una o más antenas del conjunto de antenas desde las cuales se transmiten los símbolos de modulación.

**[0119]** Durante el funcionamiento, la segunda antena 844 de la estación base 800 puede recibir un flujo de datos 814. El segundo transceptor 854 puede recibir el flujo de datos 814 desde la segunda antena 844 y puede proporcionar el flujo de datos 814 al desmodulador 862. El desmodulador 862 puede desmodular señales moduladas del flujo de datos 814 y proporcionar datos desmodulados al procesador de datos del receptor 864. El procesador de datos del receptor 864 puede extraer datos de audio de los datos desmodulados y proporcionar los datos de audio extraídos al procesador 806.

**[0120]** El procesador 806 puede proporcionar los datos de audio al transcodificador 810 para la transcodificación. El descodificador vocoder 838 del transcodificador 810 puede descodificar los datos de audio de un primer formato en datos de audio descodificados y el codificador vocoder 836 puede codificar los datos de audio descodificados en un segundo formato. En algunas implementaciones, el codificador vocoder 836 puede codificar los datos de audio usando una velocidad de transferencia de datos más alta (por ejemplo, aumento) o una velocidad de datos más baja (por ejemplo, reducción) que la recibida del dispositivo inalámbrico. En otras implementaciones, los datos de audio pueden no ser transcodificados. Aunque la transcodificación (por ejemplo, descodificación y codificación) se ilustra como realizada por un transcodificador 810, las operaciones de transcodificación (por ejemplo, descodificación y codificación) pueden realizarse por múltiples componentes de la estación base 800. Por ejemplo, el procesador de

datos del receptor 864 puede realizar la descodificación y el procesador de datos de transmisión 866 puede realizar la codificación.

**[0121]** El descodificador vocoder 838 y el codificador vocoder 836 pueden determinar, trama por trama, si cada trama recibido del flujo de datos 814 corresponde a datos de audio de voz o no de voz y puede seleccionar un descodificador correspondiente (por ejemplo, un descodificador de voz o un descodificador no de voz) y un codificador correspondiente para transcodificar (por ejemplo, descodificar y codificar) la trama. Los datos de audio codificados generados en el codificador vocoder 836, tales como datos transcodificados, pueden proporcionarse al procesador de datos de transmisión 866 o a la conexión de red 860 a través del procesador 806.

**[0122]** Los datos de audio transcodificados del transcodificador 810 pueden proporcionarse al procesador de datos de transmisión 866 para codificar de acuerdo con un esquema de modulación, tal como OFDM, para generar los símbolos de modulación. El procesador de datos de transmisión 866 puede proporcionar los símbolos de modulación al procesador MIMO de transmisión 868 para su posterior procesamiento y conformación de haces. El procesador de transmisión MIMO 868 puede aplicar ponderaciones de conformación de haz y puede proporcionar los símbolos de modulación a una o más antenas del conjunto de antenas, tales como la primera antena 842 a través del primer transceptor 852. Por lo tanto, la estación base 800 puede proporcionar un flujo de datos transcodificados 816, que corresponde al flujo de datos 814 recibido desde el dispositivo inalámbrico, a otro dispositivo inalámbrico. El flujo de datos transcodificados 816 puede tener un formato de codificación, velocidad de transferencia de datos o ambos diferentes, que el flujo de datos 814. En otras implementaciones, el flujo de datos transcodificados 816 puede proporcionarse a la conexión de red 860 para su transmisión a otra estación base o una red central.

**[0123]** Por lo tanto, la estación base 800 puede incluir un dispositivo de almacenamiento legible por ordenador (por ejemplo, la memoria 832) que almacena instrucciones que, cuando son ejecutadas por un procesador (por ejemplo, el procesador 806 o el transcodificador 810), hacen que el procesador realice operaciones que incluyen determinar los primeros datos de probabilidad asociados con una primera probabilidad de que una trama de audio sea una trama de voz y determinar los segundos datos de probabilidad asociados con una segunda probabilidad de que la trama de audio sea una trama no de voz. Las operaciones también pueden incluir la determinación de los primeros datos de decisión basándose en los primeros datos de probabilidad y los segundos datos de probabilidad. Los primeros datos de decisión indican una clasificación de la trama de audio como una trama de voz o una trama no de voz. Las operaciones también pueden incluir determinar los segundos datos de decisión basándose en los primeros datos de probabilidad, los segundos datos de probabilidad y los primeros datos de decisión. Los segundos datos de decisión pueden indicar una selección de un codificador para codificar la trama de audio o una selección de un descodificador para descodificar la trama de audio.

**[0124]** Junto con los aspectos descritos, un aparato puede incluir medios para determinar los primeros datos de probabilidad asociados con una primera probabilidad de que una trama de audio sea una trama de voz. Por ejemplo, los medios para determinar los primeros datos de probabilidad pueden incluir el primer clasificador 122 de las FIGS. 1-3, el modelo de voz 370 de la FIG. 3, el selector de codificador 760 de la FIG. 7, el procesador 706 o el procesador 710 que ejecutan las instrucciones 756 de la FIG. 7, el procesador 806 o el transcodificador 810 de la FIG. 8, uno o más dispositivos distintos configurados para determinar los primeros datos de probabilidad asociados con una primera probabilidad de que una trama de audio sea una trama de voz, o cualquier combinación de los mismos.

**[0125]** El aparato puede incluir medios para determinar los segundos datos de probabilidad asociados con una segunda probabilidad de que la trama de audio sea una trama que no es de voz. Por ejemplo, los medios para determinar los segundos datos de probabilidad pueden incluir el primer clasificador 122 de las FIGS. 1-3, el modelo de música 372 de la FIG. 3, el selector de codificador 760 de la FIG. 7, el procesador 706 o el procesador 710 que ejecutan las instrucciones 756 de la FIG. 7, el procesador 806 o el transcodificador 810 de la FIG. 8, uno o más dispositivos distintos configurados para determinar los segundos datos de probabilidad asociados con una segunda probabilidad de que una trama de audio sea una trama que no es de voz, o cualquier combinación de los mismos.

**[0126]** El aparato puede incluir medios para determinar los primeros datos de decisión basados en los primeros datos de probabilidad y los segundos datos de probabilidad, incluyendo los primeros datos de decisión una primera indicación de una clasificación de la trama de audio como una trama de voz o una trama no de voz. Por ejemplo, los medios para determinar los primeros datos de decisión pueden incluir el primer clasificador 122 de las FIGS. 1-3, la máquina de estados 374 de la FIG. 3, el selector de codificador 760 de la FIG. 7, el procesador 706 o el procesador 710 que ejecutan las instrucciones 756 de la FIG. 7, el procesador 806 o el transcodificador 810 de la FIG. 8, uno o más dispositivos distintos configurados para determinar los primeros datos de decisión basados en los primeros datos de probabilidad y los segundos datos de probabilidad, o cualquier combinación de los mismos.

**[0127]** El aparato puede incluir medios para determinar segundos datos de decisión basados en los primeros datos de probabilidad, los segundos datos de probabilidad y los primeros datos de decisión, los segundos datos de decisión incluyen una segunda indicación de una selección de un codificador para codificar la trama de audio. Por ejemplo, los medios para determinar los segundos datos de decisión pueden incluir el segundo clasificador 124 de las FIGS. 1-2 y 4, la unidad de desviación de la decisión a largo plazo 414, el generador de parámetros de ajuste 416, el generador de decisión de clasificación 418, el selector de codificador 760, el procesador 706 o el procesador 710 que ejecutan

las instrucciones 756 de la FIG. 7, el procesador 806 o el transcodificador 810 de la FIG. 8, uno o más dispositivos distintos configurados para determinar los segundos datos de decisión basándose en los primeros datos de probabilidad, los segundos datos de probabilidad y los primeros datos de decisión, o cualquier combinación de los mismos. En una implementación particular, los medios para determinar los primeros datos de probabilidad, los medios para determinar los segundos datos de probabilidad y los medios para determinar los primeros datos de decisión se incluyen en la circuitería GMM, como se describe con respecto a la FIG. 1.

**[0128]** Los medios para determinar los primeros datos de probabilidad, los medios para determinar los segundos datos de probabilidad, los medios para determinar los primeros datos de decisión y los medios para determinar los segundos datos de decisión están integrados en un codificador, un descodificador, un reproductor de música, un reproductor de vídeo, una unidad de entretenimiento, un dispositivo de navegación, un dispositivo de comunicaciones, un PDA, un ordenador o una combinación de los mismos.

**[0129]** En los aspectos de la descripción descrita en el presente documento, diversas funciones realizadas por el sistema 100 de la FIG. 1, el ejemplo 200 de la FIG. 2, el ejemplo 300 de la FIG. 3, el ejemplo 400 de la FIG. 3, el dispositivo 700 de la FIG. 7, la estación base 800 de la FIG. 8, o una combinación de los mismos, se describen como realizados por ciertos circuitos o componentes. Sin embargo, esta división de componentes o circuitos sirve solo para ilustrar. En un ejemplo alternativo, una función realizada por un componente o circuito particular se puede dividir, en cambio, entre múltiples componentes o módulos. Además, en un ejemplo alternativo, dos o más componentes o circuitos de las FIGS. 1-4, 7 y 8 se pueden integrar en un único componente o circuito. Cada componente o circuito ilustrado en las FIGS. 1-4, 7 y 8 se puede implementar usando hardware (por ejemplo, un ASIC, un DSP, un controlador, un dispositivo FPGA, etc.), software (por ejemplo, lógica, módulos, instrucciones ejecutables por un procesador) o cualquier combinación de los mismos.

**[0130]** Los expertos apreciarán, además, que los diversos bloques lógicos, configuraciones, módulos, circuitos y etapas de algoritmo ilustrativos descritos en conexión con los ejemplos divulgados en el presente documento se pueden implementar como hardware electrónico, software informático ejecutado por un procesador o combinaciones de ambos. Se han descrito anteriormente diversos componentes, bloques, configuraciones, módulos, circuitos y etapas ilustrativos, en general, en lo que respecta a su funcionalidad. Que dicha funcionalidad se implemente como hardware o instrucciones ejecutables por procesador depende de la aplicación particular y de las restricciones de diseño impuestas al sistema global. Los expertos en la técnica pueden implementar la funcionalidad descrita de diferentes maneras para cada aplicación particular, pero no debe interpretarse que dichas decisiones de implementación suponen una desviación del alcance de la presente divulgación.

**[0131]** Las etapas de un procedimiento o algoritmo descrito en relación con los ejemplos divulgados en el presente documento se pueden incorporar directamente en hardware, en un módulo de software ejecutado por un procesador o en una combinación de los dos. Un módulo de software puede residir en una memoria de acceso aleatorio (RAM), memoria flash, memoria de solo lectura (ROM), memoria de solo lectura programable (PROM), memoria de solo lectura programable y borrable (EPROM), memoria de solo lectura programable y borrable eléctricamente (EEPROM), registros, disco duro, disco extraíble, disco compacto con memoria de solo lectura (CD-ROM) o en cualquier otra forma de medio de almacenamiento no transitorio conocido en la técnica. Un medio de almacenamiento a modo de ejemplo está acoplado al procesador de modo que el procesador pueda leer información de, y escribir información en, el medio de almacenamiento. De forma alternativa, el medio de almacenamiento puede estar integrado en el procesador. El procesador y el medio de almacenamiento pueden residir en un ASIC. El ASIC puede residir en un dispositivo informático o en un terminal de usuario. Como alternativa, el procesador y el medio de almacenamiento pueden residir como componentes discretos en un dispositivo informático o en un terminal de usuario.

**[0132]** La descripción anterior de los ejemplos divulgados se proporciona para permitir que un experto en la técnica elabore o use las implementaciones divulgadas. Diversas modificaciones de estos ejemplos resultarán fácilmente evidentes para los expertos en la técnica, y los principios definidos en el presente documento se pueden aplicar a otras implementaciones sin apartarse del alcance de la divulgación. Por tanto, la presente divulgación no pretende limitarse a los ejemplos mostrados en el presente documento, sino que se le ha de conceder el alcance más amplio posible consecuente con los principios y características novedosas, como se define por las reivindicaciones siguientes.

**REIVINDICACIONES**

1. Un dispositivo para seleccionar uno de un primer codificador y un segundo codificador para codificar una trama de audio, comprendiendo el dispositivo:

- 5 un primer clasificador que comprende:
  - medios (370) para determinar los primeros datos de probabilidad (lps) asociados con una primera probabilidad de que la trama de audio sea una trama de voz;
  - 10 medios (372) para determinar los segundos datos de probabilidad (lpm) asociados con una segunda probabilidad de que la trama de audio sea una trama no de voz; y
  - 15 medios (374) para determinar los primeros datos de decisión (346) que indican una clasificación de la trama de audio como una trama de voz o una trama no de voz, los primeros datos de decisión determinados en base a los primeros datos de probabilidad y en base a los segundos datos de probabilidad; y

un segundo clasificador acoplado para recibir los primeros datos de decisión, los primeros datos de probabilidad y los segundos datos de probabilidad del primer clasificador,

- 20 **caracterizado por que** el segundo clasificador comprende:
  - medios para determinar un primer valor de ganancia de codificación estimada asociado con el primer codificador, y un segundo valor de ganancia de codificación estimada asociado con el segundo codificador, en el que los valores de ganancia de codificación estimada primero y segundo son estimaciones de ganancia de codificación que corresponden a una probabilidad de que la trama de audio sea una trama de voz o una trama no de voz;
  - 25 un generador de parámetros de ajuste (416) configurado para generar un parámetro de ajuste basándose en si se cumple un conjunto de condiciones, comprendiendo el conjunto de condiciones una primera condición de que al menos uno de los codificadores esté asociado con una primera frecuencia de muestreo, una segunda condición sobre si los primeros datos de decisión indican la clasificación de la trama de audio como una trama de voz o una trama no de voz, una tercera condición de que el primer valor de ganancia de codificación estimada sea mayor o igual que un primer valor, donde el primer valor está asociado con una diferencia entre el segundo valor de ganancia de codificación estimada y un segundo valor, y una cuarta condición basada en los primeros datos de probabilidad y los segundos datos de probabilidad; y
  - 30 un generador de decisión de clasificación (418) configurado para generar segundos datos de decisión basados en el primer valor de ganancia de codificación estimada, el segundo valor de ganancia de codificación estimada y el parámetro de ajuste, en el que los segundos datos de decisión indican uno del primer codificador y el segundo codificador para codificar la trama de audio.

2. El dispositivo según la reivindicación 1,

- 45 en el que el primer codificador comprende un codificador de voz, y en el que el segundo codificador comprende un codificador que no es de voz; y/o
- en el que el segundo codificador comprende un codificador de música; y/o
- 50 en el que el primer codificador comprende un codificador de predicción lineal excitado por código algebraico (ACELP), y en el que el segundo codificador comprende un codificador de excitación codificada por transformación (TCX); y/o
- en el que el primer clasificador comprende un módulo modelo de mezcla gaussiana, y en el que el segundo clasificador comprende un clasificador en bucle abierto; y/o
- 55 en el que el primer clasificador incluye una máquina de estados, la máquina de estados configurada para recibir los primeros datos de probabilidad y los segundos datos de probabilidad y para generar los primeros datos de decisión basados en los primeros datos de probabilidad y los segundos datos de probabilidad.

3. El dispositivo según la reivindicación 1, que comprende, además, un codificador conmutado que está configurado para operar en múltiples modos de codificación, en el que el primer y segundo codificadores corresponden a los múltiples modos de codificación del codificador conmutado; y/o

- 60 que comprende, además, el primer y segundo codificadores, en el que el primer clasificador, el segundo clasificador y el primer y segundo codificadores están integrados en un dispositivo de comunicación móvil o una estación base.

4. El dispositivo según la reivindicación 1, que comprende, además:

un receptor configurado para recibir una señal de audio que incluye la trama de audio;

5 un desmodulador acoplado al receptor, el desmodulador configurado para desmodular la señal de audio;

un procesador acoplado al desmodulador; y

10 múltiples descodificadores, en el que el receptor, el desmodulador, el procesador y los múltiples descodificadores están integrados en un dispositivo de comunicación móvil o una estación base.

5. Un procedimiento para seleccionar uno de un primer codificador y un segundo codificador para codificar una trama de audio, comprendiendo el procedimiento:

15 determinar, en un primer clasificador, los primeros datos de probabilidad (lps) asociados con una primera probabilidad de que la trama de audio sea una trama de voz;

20 determinar, en el primer clasificador, los segundos datos de probabilidad (lpm) asociados con una segunda probabilidad de que la trama de audio sea una trama no de voz; y

25 determinar, en el primer clasificador, los primeros datos de decisión (346) que indican una clasificación de la trama de audio como una trama de voz o una trama no de voz, los primeros datos de decisión determinados en base a los primeros datos de probabilidad asociados con una primera probabilidad de que la trama de audio sea una trama de voz y en base a los segundos datos de probabilidad

30 recibir los primeros datos de probabilidad y los segundos datos de probabilidad en un segundo clasificador; recibir los primeros datos de decisión en el segundo clasificador;

**caracterizado por:**

35 determinar, en el primer clasificador, un primer valor de ganancia de codificación estimada asociado con el primer codificador, y un segundo valor de ganancia de codificación estimada asociado con el segundo codificador, en el que los valores de ganancia de codificación estimada primero y segundo son estimaciones de ganancia de codificación que corresponden a una probabilidad de que la trama de audio sea una trama de voz o una trama no de voz;

40 determinar, en el segundo clasificador, un parámetro de ajuste, basándose en si se cumple un conjunto de condiciones, comprendiendo el conjunto de condiciones una primera condición de que al menos uno de los codificadores esté asociado con una primera frecuencia de muestreo, una segunda condición sobre si los primeros datos de decisión indican la clasificación de la trama de audio como una trama de voz o una trama no de voz, una tercera condición de que el primer valor de ganancia de codificación estimada sea mayor o igual que un primer valor, donde el primer valor está asociado con una diferencia entre segundo valor de ganancia de codificación estimada y un segundo valor, y una cuarta condición basada en los primeros datos de probabilidad y los segundos datos de probabilidad;

45 generar segundos datos de decisión basados en el primer valor de ganancia de codificación estimada, el segundo valor de ganancia de codificación estimada y el parámetro de ajuste, en el que los segundos datos de decisión indican uno del primer codificador y el segundo codificador para codificar la trama de audio.

50 6. El procedimiento según la reivindicación 5, en el que los primeros datos de decisión se reciben en el segundo clasificador desde una máquina de estados del primer clasificador; y/o

55 que comprende, además, proporcionar los segundos datos de decisión desde una salida del segundo clasificador a un conmutador configurado para seleccionar el codificador particular de los múltiples codificadores, en el que la trama de audio se codifica utilizando el codificador particular. 5

60 7. El procedimiento según la reivindicación 6, que comprende, además, determinar si la trama de audio está asociada con una frecuencia de muestreo de 12.800 hercios, en el que se determina que el conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que la trama de audio está asociada con la frecuencia de muestreo de 12.800 hercios.

65 8. El procedimiento según la reivindicación 5, en el que se determina que el conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que los primeros datos de decisión indican que la trama de audio se clasifica como trama de voz.

9. El procedimiento según la reivindicación 5, en el que se determina que el conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que el primer valor de ganancia de codificación estimada es mayor o igual que el primer valor.

5 10. El procedimiento según la reivindicación 5, que comprende, además:

determinar si una trama clasificada más recientemente se clasifica como que incluye contenido de voz; y

10 determinar si un primer valor de probabilidad indicado por los primeros datos de probabilidad es mayor que un segundo valor de probabilidad indicado por los segundos datos de probabilidad,

15 en el que se determina que el conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que la trama clasificada más recientemente se clasifica como que incluye contenido de voz y en respuesta a la determinación de que el primer valor de probabilidad es mayor que el segundo valor de probabilidad.

11. El procedimiento según la reivindicación 5

20 que comprende, además:

determinar si cada trama correspondiente a un número de tramas clasificadas más recientemente se clasifica como que incluye contenido de voz; y

25 determinar si un primer valor de probabilidad indicado por los primeros datos de probabilidad es mayor o igual que un tercer valor, el tercer valor asociado con una diferencia entre un segundo valor de probabilidad indicado por los segundos datos de probabilidad y un cuarto valor,

30 en el que se determina que el conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que cada trama correspondiente al número de tramas clasificadas más recientemente se clasifica como que incluye contenido del voz y en respuesta a la determinación de que el primer valor de probabilidad es mayor que o igual al tercer valor.

12. El procedimiento según la reivindicación 5, que comprende, además:

35 determinar si un valor promedio de voz de múltiples subtramas de la trama de audio es mayor o igual que un primer umbral;

40 determinar si un valor de no estacionaridad asociado con la trama de audio es mayor que un segundo umbral; y

determinar si un valor de desplazamiento asociado con la trama de audio es menor que un tercer umbral,

45 en el que se determina que el conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que el valor promedio de voz es mayor o igual que el primer umbral, la determinación de que el valor de no estacionaridad es mayor que el segundo umbral y la determinación de que el valor de desplazamiento es menor que el tercer umbral.

13. El procedimiento según la reivindicación 5, que comprende, además:

50 determinar si se cumple un segundo conjunto de condiciones asociadas con una trama de audio; y

en respuesta a que se cumple el segundo conjunto de condiciones, seleccionar un valor de un parámetro de ajuste para desviar la selección hacia un segundo codificador asociado con la no voz.

55 14. El procedimiento según la reivindicación 13, que comprende, además, determinar si la trama de audio está asociada con una frecuencia de muestreo de 12.800 hercios, en el que se determina que el segundo conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que la trama de audio está asociada con la frecuencia de muestreo de 12.800 hercios; y/o

60 que comprende, además, determinar si los primeros datos de decisión indican que la trama de audio se clasifica como una trama no de voz, en el que se determina que el segundo conjunto de condiciones se cumple al menos en parte en respuesta a la determinación de que los primeros datos de decisión indican que la trama de audio es clasificada como una trama no de voz.

**15.** Un dispositivo de almacenamiento legible por ordenador que almacena instrucciones que, cuando se ejecutan mediante un procesador, provocan que el procesador realice el procedimiento de cualquiera de las reivindicaciones 5-14.

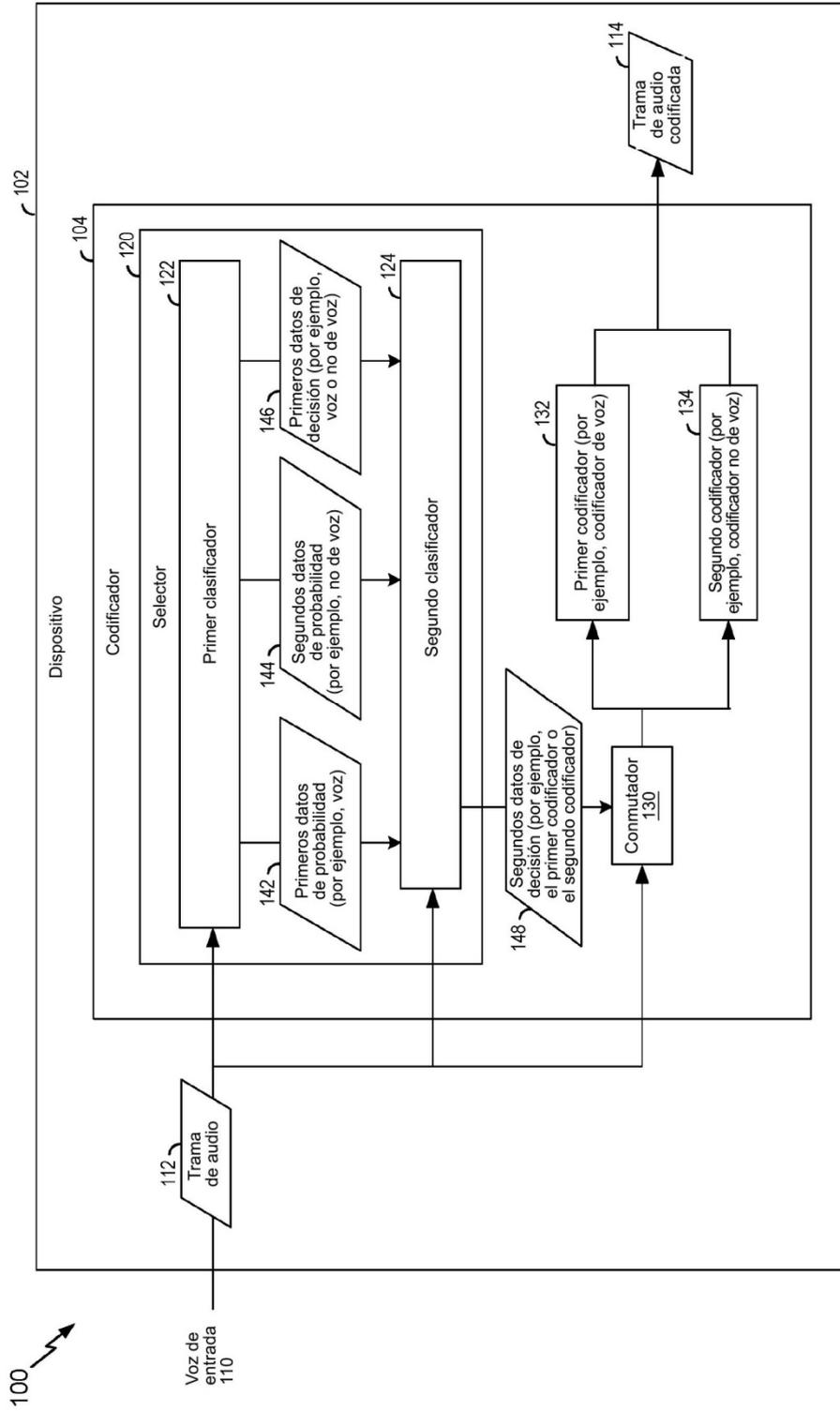


FIG. 1

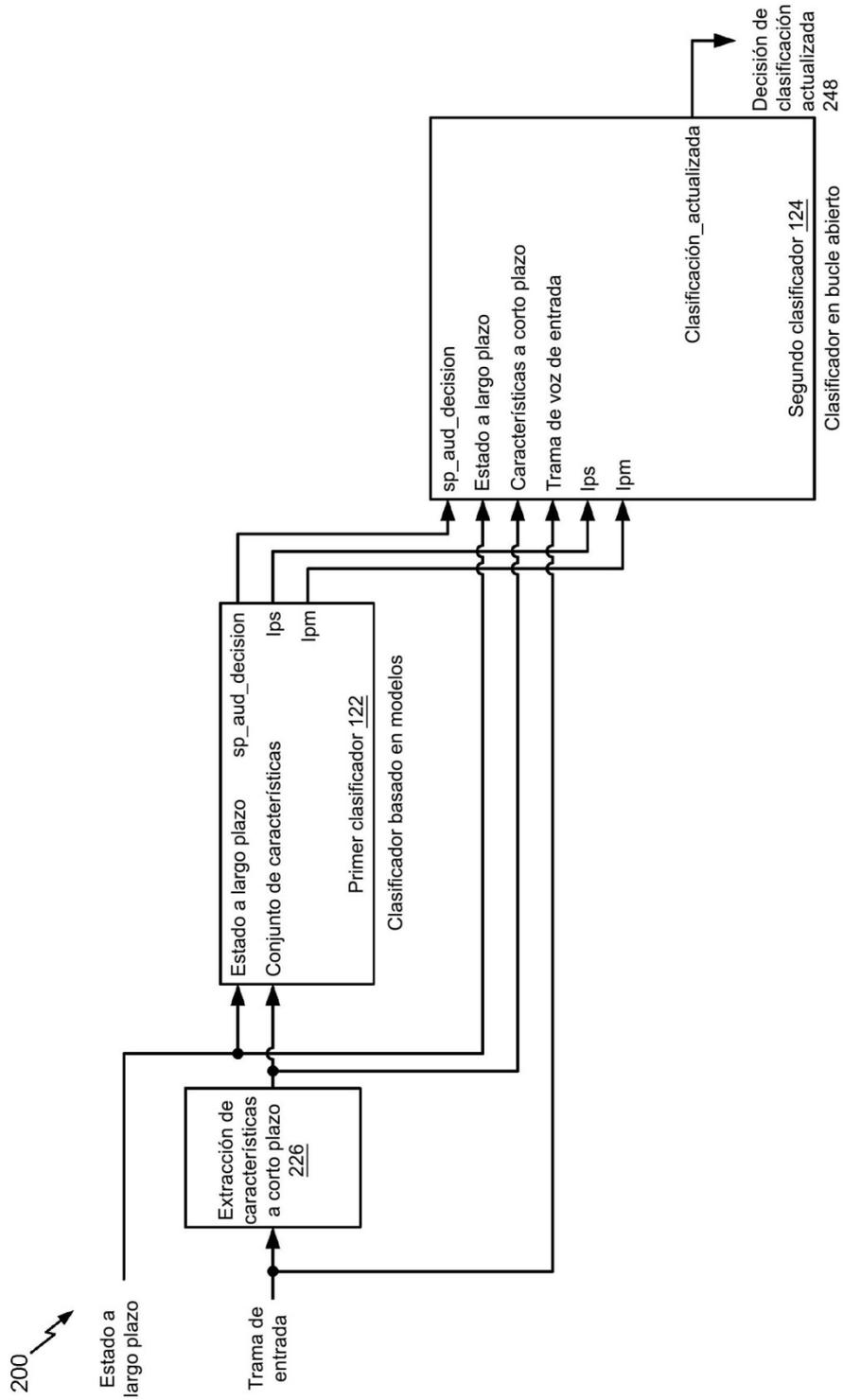
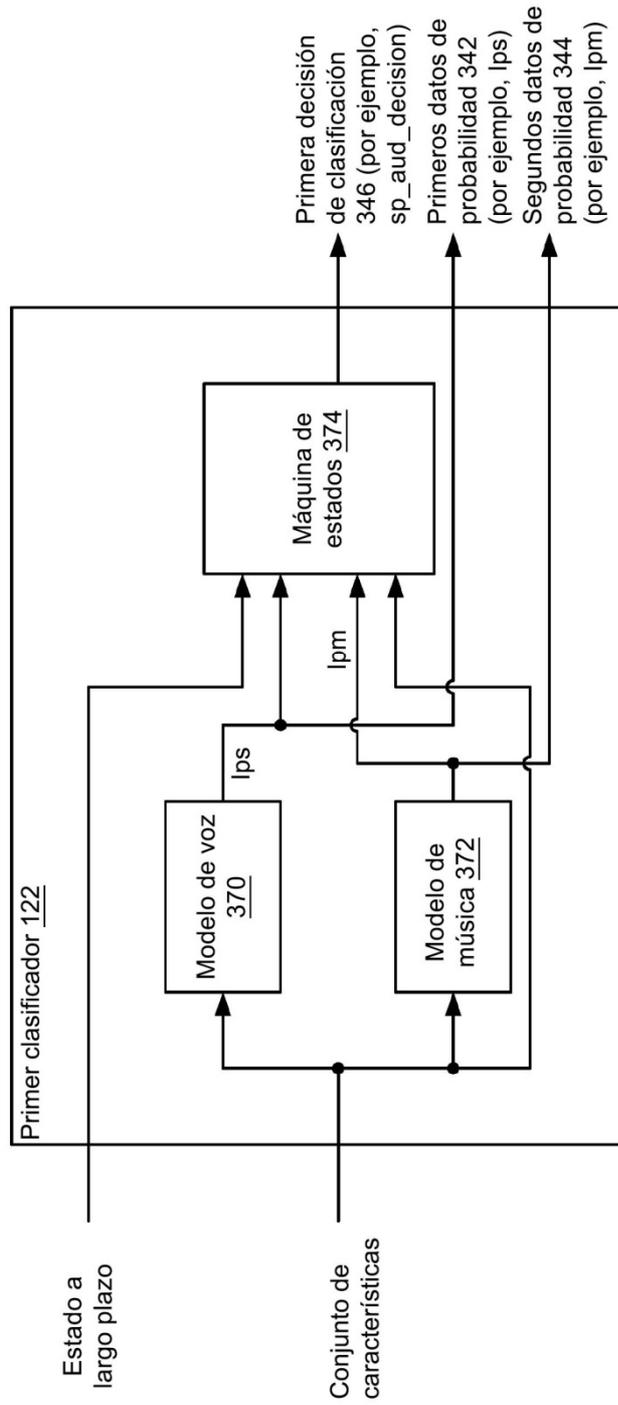


FIG. 2

300 ↗



**FIG. 3**

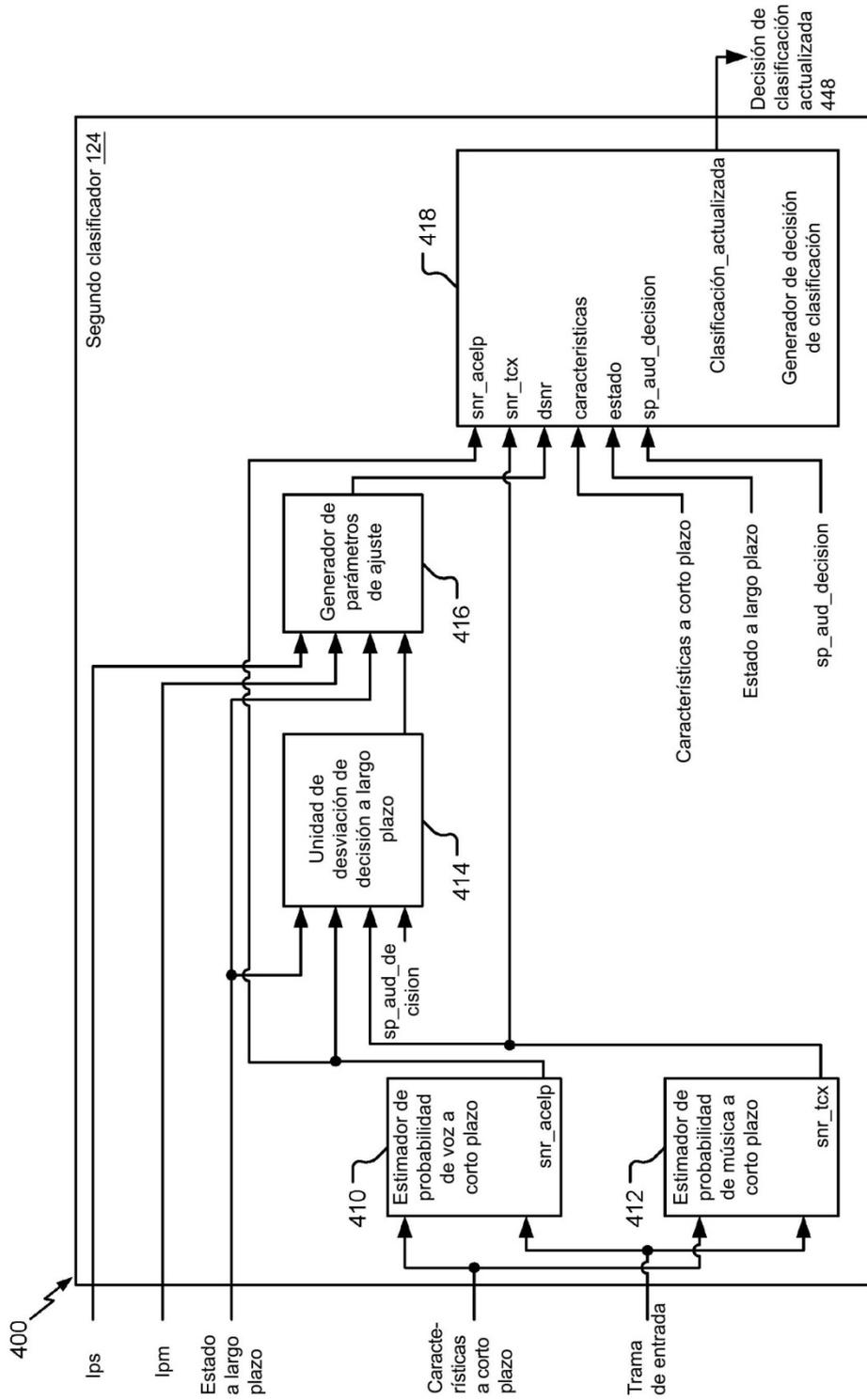
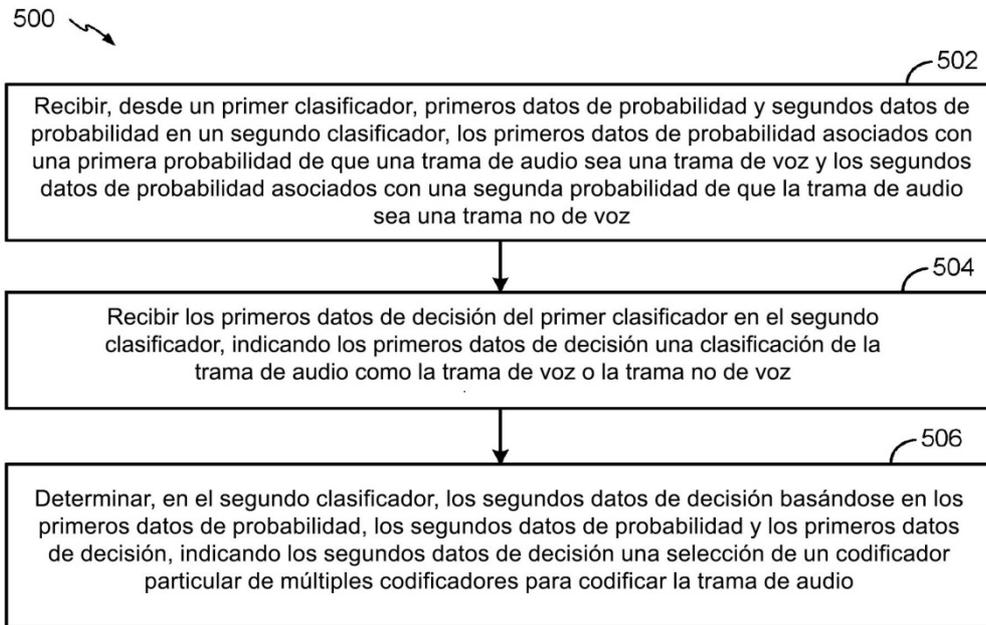
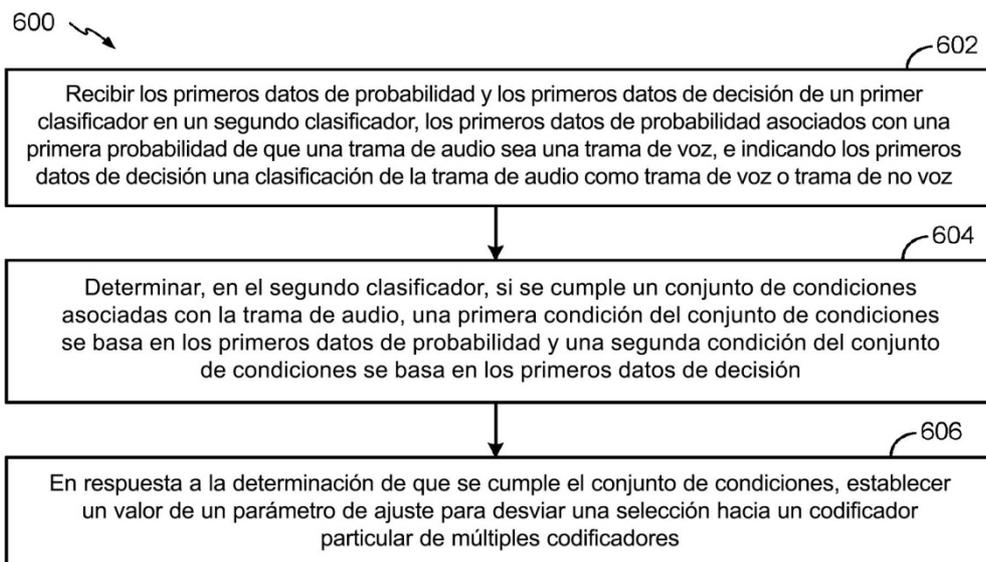


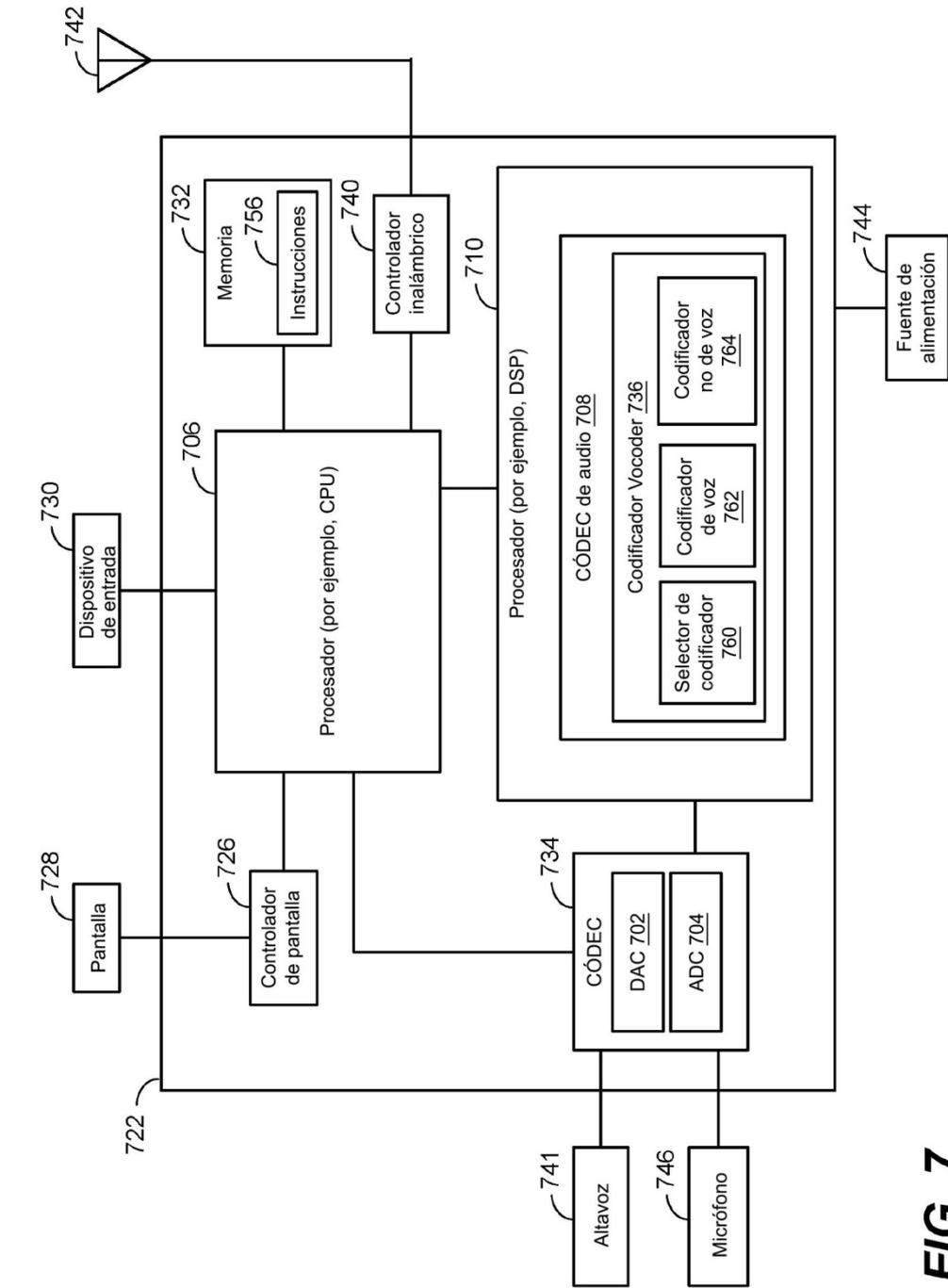
FIG. 4



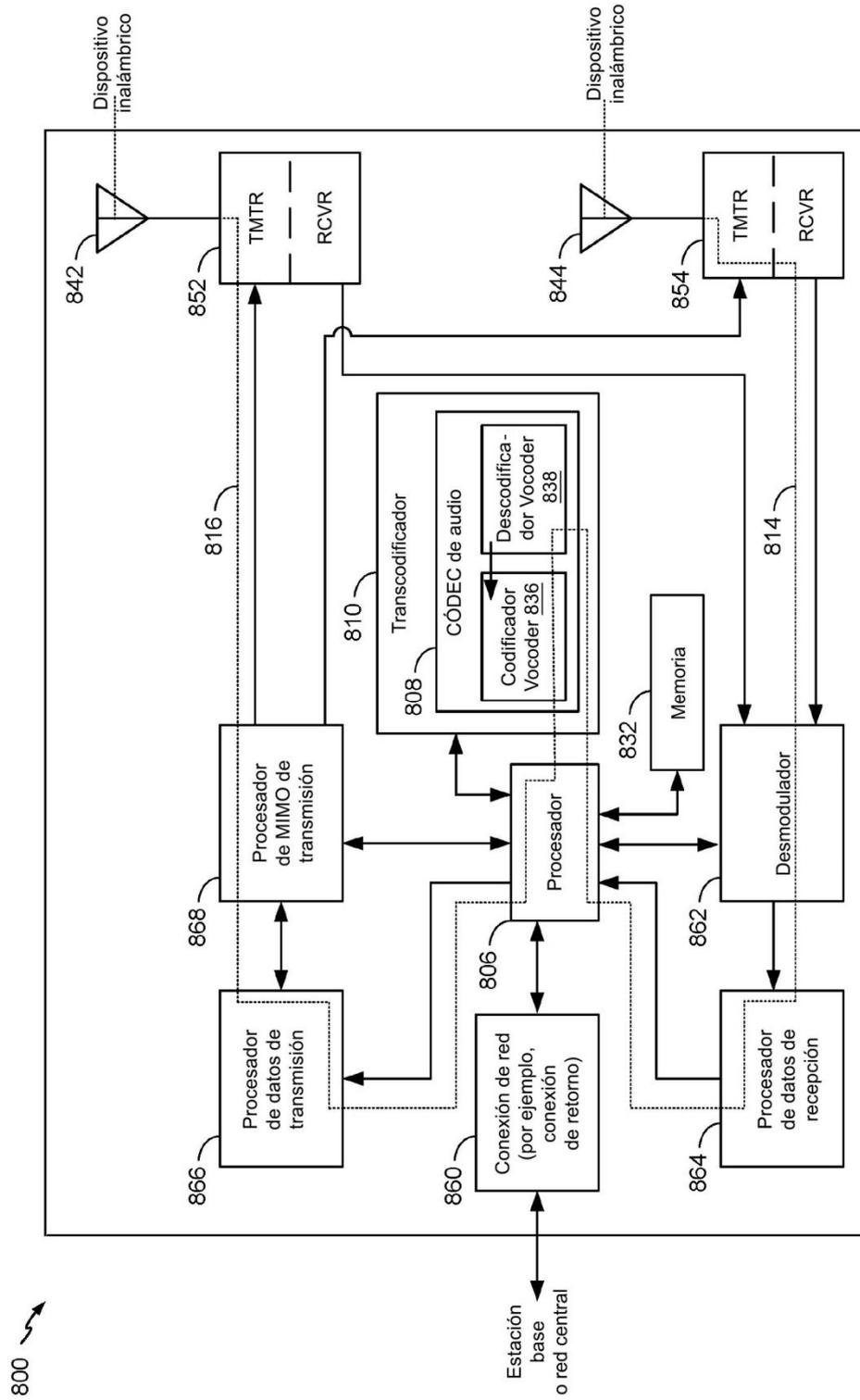
**FIG. 5**



**FIG. 6**



**FIG. 7**



**FIG. 8**