

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 817 841**

51 Int. Cl.:

**G10L 15/22** (2006.01)

**G06F 3/16** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **08.09.2014 PCT/US2014/054540**

87 Fecha y número de publicación internacional: **26.03.2015 WO15041882**

96 Fecha de presentación y número de la solicitud europea: **08.09.2014 E 14784130 (8)**

97 Fecha y número de publicación de la concesión europea: **10.06.2020 EP 3047482**

54 Título: **Procedimiento y aparato para ajustar umbral de detección para activar función de asistente de voz**

30 Prioridad:  
**17.09.2013 US 201314029131**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**08.04.2021**

73 Titular/es:  
**QUALCOMM INCORPORATED (100.0%)  
5775 Morehouse Drive  
San Diego, CA 92121-1714, US**

72 Inventor/es:  
**LEE, MINSUB;  
KIM, TAESU;  
HWANG, KYU WOONG y  
JIN, MINHO**

74 Agente/Representante:  
**FORTEA LAGUNA, Juan José**

ES 2 817 841 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

Procedimiento y aparato para ajustar umbral de detección para activar función de asistente de voz

### 5 **CAMPO**

[0001] La presente divulgación se refiere en general a un procedimiento para activar una función de asistente de voz en un dispositivo móvil. Más específicamente, la presente divulgación se refiere al ajuste de un umbral para detectar una palabra clave objetivo en el dispositivo móvil.

10

### **ANTECEDENTES**

[0002] En los últimos años, el uso de dispositivos móviles tales como teléfonos inteligentes y *tablets* se ha generalizado. Dichos dispositivos móviles permiten en general comunicaciones de voz y datos a través de redes inalámbricas. Típicamente, estos dispositivos incluyen características o aplicaciones adicionales, que proporcionan una variedad de funciones diseñadas para mejorar la comodidad del usuario.

15

[0003] Entre dichas aplicaciones o características, una aplicación de asistente de voz permite que un dispositivo móvil reciba un comando de voz de un usuario para hacer funcionar diversas funciones u otras aplicaciones en respuesta al comando de voz. Por ejemplo, una aplicación de asistente de voz puede permitir al usuario pronunciar un comando de voz para llamar al número de teléfono deseado, reproducir un archivo de audio, hacer una foto, buscar en Internet u obtener información meteorológica, sin hacer funcionar físicamente el dispositivo móvil con la mano. En consecuencia, el usuario puede optar por usar un comando de voz para controlar diversas operaciones del dispositivo móvil en una variedad de situaciones donde el usuario puede no querer o no puede hacer funcionar físicamente el dispositivo móvil.

20

25

[0004] Las aplicaciones convencionales de asistente de voz a menudo se activan en respuesta a una entrada física de un usuario (por ejemplo, a mano). Sin embargo, dado que uno de los propósitos al usar una aplicación de asistente de voz puede ser hacer funcionar un dispositivo móvil sin una entrada física, el usuario puede encontrar inconveniente o difícil activar físicamente la aplicación de asistente de voz. Por tanto, algunos dispositivos móviles permiten que un usuario active la aplicación de asistente de voz al pronunciar un comando de voz.

30

[0005] Sin embargo, dichos dispositivos móviles se activan a menudo erróneamente en respuesta a diversos sonidos de entrada capturados por los dispositivos o no se activan en respuesta a un comando de voz adecuado. Por ejemplo, un dispositivo móvil puede reconocer erróneamente el enunciado de otra persona como un comando de voz adecuado y activar una aplicación de asistente de voz. Por otro lado, cuando un dispositivo móvil está en un entorno ruidoso, el dispositivo móvil puede no reconocer un comando de voz adecuado debido al ruido de fondo y, por tanto, no puede activar la aplicación de asistente de voz.

35

[0006] En un artículo titulado "OK Google... Now We're Talking!" publicado en opusresearch.net (<https://opusresearch.net/wordpress/2013/08/02/ok-google-now-were-talking/>) Dan Miller describe la posibilidad de activar la función de asistente de voz de un teléfono móvil mediante la detección de una palabra clave objetivo.

40

### **BREVE EXPLICACIÓN**

45

[0007] La presente divulgación proporciona procedimientos y dispositivos para activar una función de asistente de voz mediante la detección de una palabra clave objetivo en un flujo de sonido de entrada en base a un umbral que puede ajustarse de acuerdo con los contextos de los dispositivos.

50

[0008] De acuerdo con un aspecto de la presente divulgación, se divulga un procedimiento para activar una función de asistente de voz en un dispositivo móvil. El procedimiento incluye recibir un flujo de sonido de entrada por un sensor de sonido y determinar un contexto del dispositivo móvil. En este procedimiento, un umbral para activar la función de asistente de voz se ajusta en base al contexto. El procedimiento detecta una palabra clave objetivo del flujo de sonido de entrada en base al umbral ajustado. Si se detecta la palabra clave objetivo, el procedimiento activa la función de asistente de voz. La presente divulgación también describe un aparato, un dispositivo, una combinación de medios y un medio legible por ordenador relacionado con este procedimiento.

55

[0009] De acuerdo con otro aspecto de la presente divulgación, un dispositivo móvil para activar una función de asistente de voz incluye un sensor de sonido, una unidad de ajuste de umbral y una unidad de activación por voz. El sensor de sonido está configurado para recibir un flujo de sonido de entrada. La unidad de ajuste de umbral está configurada para ajustar un umbral para activar la función de asistente de voz determinando un contexto del dispositivo móvil. La unidad de activación por voz está configurada para activar la función de asistente de voz si se detecta una palabra clave objetivo del flujo de sonido de entrada en base al umbral ajustado.

60

### **BREVE DESCRIPCIÓN DE LOS DIBUJOS**

65

**[0010]** Los modos de realización de los aspectos inventivos de la presente divulgación se entenderán con referencia a la siguiente descripción detallada, cuando se lea junto con los dibujos adjuntos.

5 La **FIG. 1** ilustra un dispositivo móvil configurado para ajustar un umbral de detección para activar una función de asistente de voz cuando se determina que un contexto del dispositivo móvil está en un vehículo, de acuerdo con un modo de realización de la presente divulgación.

10 La **FIG. 2** ilustra un dispositivo móvil configurado para ajustar un umbral de detección para activar una función de asistente de voz si se determina que un contexto del dispositivo móvil es por la mañana cuando es probable que un usuario esté despierto, de acuerdo con un modo de realización de la presente divulgación.

15 La **FIG. 3** ilustra un dispositivo móvil configurado para ajustar un umbral de detección para activar una función de asistente de voz si se determina que un contexto del dispositivo móvil es por la noche cuando es probable que un usuario esté dormido, de acuerdo con un modo de realización de la presente divulgación.

La **FIG. 4** ilustra un dispositivo móvil configurado para ajustar un umbral de detección para activar una función de asistente de voz cuando se determina que un contexto del dispositivo móvil está en una bolsa, de acuerdo con un modo de realización de la presente divulgación.

20 La **FIG. 5** ilustra un dispositivo móvil configurado para ajustar un umbral de detección para activar una función de asistente de voz en un contexto de recepción de un mensaje de texto, de acuerdo con un modo de realización de la presente divulgación.

25 La **FIG. 6** ilustra un diagrama de bloques de un dispositivo móvil configurado para detectar una palabra clave objetivo a partir de un flujo de sonido de entrada ajustando un umbral para detectar la palabra clave objetivo en base al contexto del dispositivo móvil, de acuerdo con un modo de realización de la presente divulgación.

30 La **FIG. 7** ilustra un diagrama de bloques de una unidad de sensor en un dispositivo móvil configurado para emitir un flujo de sonido y datos de sensor para su uso en la determinación de un contexto del dispositivo móvil, de acuerdo con un modo de realización de la presente divulgación.

35 La **FIG. 8** ilustra un diagrama de bloques de una unidad de ajuste de umbral en un dispositivo móvil configurado para determinar un umbral de detección para activar una función de asistente de voz infiriendo un contexto del dispositivo móvil, de acuerdo con un modo de realización de la presente divulgación.

La **FIG. 9** es un diagrama de flujo de un procedimiento, realizado en un dispositivo móvil, para determinar un umbral de detección para activar una función de asistente de voz infiriendo un contexto del dispositivo móvil, de acuerdo con un modo de realización de la presente divulgación.

40 La **FIG. 10** es un diagrama de flujo de un procedimiento, realizado en un dispositivo móvil, para generar un umbral de detección ajustado y restablecer el umbral de detección a un umbral de detección predeterminado después de un período de tiempo especificado, de acuerdo con un modo de realización de la presente divulgación.

45 La **FIG. 11** ilustra un diagrama de bloques de una unidad de activación por voz en un dispositivo móvil configurado para activar una función de asistente de voz mediante la detección de una palabra clave objetivo en un flujo de sonido de entrada en base a un umbral de detección ajustado, de acuerdo con un modo de realización de la presente divulgación.

50 La **FIG. 12** es un diagrama de flujo de un procedimiento, realizado en un dispositivo móvil, para activar una función de asistente de voz mediante la detección de una palabra clave objetivo en un flujo de sonido de entrada en base a un umbral de detección ajustado, de acuerdo con un modo de realización de la presente divulgación.

55 La **FIG. 13** ilustra un diagrama de bloques de un dispositivo móvil en un sistema de comunicación inalámbrica en el cual los procedimientos y los aparatos para detectar una palabra clave objetivo en un flujo de sonido de entrada se basan en un umbral de detección que se ajusta de acuerdo con el contexto del dispositivo móvil, de acuerdo con un modo de realización de la presente divulgación.

60 **DESCRIPCIÓN DETALLADA**

**[0011]** Ahora se hará referencia en detalle a diversos modos de realización de ejemplo, ejemplos de los cuales se ilustran en los dibujos adjuntos. En la siguiente descripción detallada, se exponen numerosos detalles específicos para proporcionar un entendimiento exhaustivo de la presente materia objeto. Sin embargo, será evidente para un experto en la técnica que la presente materia objeto puede llevarse a la práctica sin estos detalles

específicos. En otros casos, los procedimientos, sistemas y componentes bien conocidos no se han descrito en detalle para no complicar innecesariamente aspectos de los diversos modos de realización.

5 [0012] La FIG. 1 ilustra un dispositivo móvil 110 configurado para ajustar un umbral de detección para activar una función de asistente de voz cuando se determina que un contexto del dispositivo móvil 110 está en un vehículo 130, de acuerdo con un modo de realización de la presente divulgación. En este modo de realización ilustrado, un usuario 120 del dispositivo móvil 110 está conduciendo el vehículo 130 mientras que el dispositivo móvil 110 está colocado en un asiento de pasajero del vehículo 130. El dispositivo móvil 110 de la FIG. 1 puede ser un teléfono inteligente equipado con una función de reconocimiento de habla para reconocer el habla de un usuario en un sonido de entrada.

10 [0013] Cuando se activa la función de asistente de voz en el dispositivo móvil 110, el dispositivo móvil 110 puede responder a diversas solicitudes verbales, instrucciones, comandos y similares del usuario 120 usando la función de reconocimiento de habla. Por ejemplo, el usuario 120 puede pronunciar una instrucción tal como "LLAMAR A SUSAN" cuando se haya activado la función de asistente de voz. En respuesta, el dispositivo móvil 110 reconoce la instrucción y llama a Susan usando un número de teléfono asociado con el nombre reconocido. La función de asistente de voz en el dispositivo móvil 110 puede activarse en respuesta a una entrada por medio de un botón o de una pantalla táctil del dispositivo móvil 110.

15 [0014] En algunos modos de realización, el dispositivo móvil 110 almacena una palabra clave objetivo predeterminada para activar la función de asistente de voz cuando la palabra clave objetivo se detecte en un sonido de entrada del usuario 120. Como se usa en el presente documento, el término "palabra clave objetivo" se refiere a cualquier representación digital o analógica de una o más palabras o sonido que puede usarse para activar una función o una aplicación en el dispositivo móvil 110. Por ejemplo, una palabra clave predeterminada tal como "HOLA, ASISTENTE DE VOZ" o "INICIAR ASISTENTE DE VOZ" puede almacenarse en el dispositivo móvil 110 como una palabra clave objetivo para activar la función de asistente de voz. Cuando se reciba un sonido de entrada del usuario 120, el dispositivo móvil 110 puede detectar la palabra clave objetivo en el sonido de entrada realizando una función de reconocimiento de voz en el sonido de entrada recibido.

20 [0015] El dispositivo móvil 110 puede configurarse para detectar una palabra clave objetivo en base a un umbral de detección para garantizar un nivel de confianza deseado. En un modo de realización, el dispositivo móvil 110 calcula un grado de similitud (o un puntaje de coincidencia de palabra clave) entre un sonido de entrada y la palabra clave objetivo y compara el grado de similitud con el umbral de detección. Al calcular el grado de similitud, puede usarse un modelo de sonido preestablecido para la palabra clave objetivo. Si el grado de similitud entre el sonido de entrada y la palabra clave objetivo (o el modelo de sonido) excede el umbral de detección, el dispositivo móvil 110 detecta la palabra clave objetivo en el sonido de entrada y activa la función de asistente de voz. Por otro lado, si el grado de similitud no excede el umbral de detección, el dispositivo móvil 110 determina que la palabra clave objetivo no se detecta en el sonido de entrada.

25 [0016] El umbral de detección afecta la precisión al detectar una palabra clave objetivo en un sonido de entrada. Por ejemplo, si el umbral de detección es demasiado alto, el dispositivo móvil 110 puede no detectar una palabra clave objetivo a partir de un sonido de entrada que incluya la palabra clave objetivo. En dicho caso, el dispositivo móvil 110 puede "omitir" la palabra clave objetivo en el sonido de entrada, y conducir a un aumento de la tasa de omisión para la palabra clave objetivo. Por otro lado, si el umbral de detección es demasiado bajo, el dispositivo móvil 110 puede detectar erróneamente una palabra clave objetivo en un sonido de entrada que no incluya la palabra clave objetivo. En dicho caso, el dispositivo móvil 110 puede activar erróneamente la función de asistente de voz (es decir, "activación falsa") y causar un aumento de la tasa de activación falsa para la función de asistente de voz.

30 [0017] En algunos modos de realización, el umbral de detección se establece inicialmente en un valor predeterminado (por ejemplo, un valor predeterminado) y luego se ajusta en base a un cambio en un contexto del dispositivo móvil 110. Con referencia a la FIG. 1, el dispositivo móvil 110 está configurado para determinar un contexto en base a datos de contexto de diversos sensores tales como un sensor de sonido, un sistema de posicionamiento global (GPS) y un acelerómetro. Por ejemplo, el sensor de sonido del dispositivo móvil 110 puede capturar ruidos de tráfico ambiental, incluyendo un sonido de motor y ruido de carretera, como un sonido de entrada. Además, el GPS y el acelerómetro pueden monitorear la localización, la velocidad y/o la aceleración del dispositivo móvil 110. En base a los datos de contexto de los sensores, el dispositivo móvil 110 puede identificar el contexto del dispositivo móvil 110 como si estuviera en un vehículo.

35 [0018] Cuando se determina que el contexto del dispositivo móvil 110 está en un vehículo, puede suponerse que el usuario 120 del dispositivo móvil 110 en la FIG. 1 es más probable que use la función de asistente de voz que hacer funcionar físicamente el dispositivo móvil 110 a mano. Por consiguiente, el dispositivo móvil 110 puede ajustar el umbral de detección a un valor inferior al umbral de detección predeterminado que está predeterminado para contextos normales o no identificados. El umbral de detección reducido aumenta la probabilidad de detectar la palabra clave objetivo (por ejemplo, una disminución de la tasa de fallos) para reducir la probabilidad de que el usuario tenga que hacer funcionar el dispositivo móvil 110 a mano.

[0019] También puede suponerse que la probabilidad de usar la función de asistente de voz es alta en algunos otros contextos. La **FIG. 2** ilustra el dispositivo móvil 110 configurado para ajustar el umbral de detección para activar la función de asistente de voz si se determina que un contexto del dispositivo móvil 110 es por la mañana cuando es probable que un usuario 210 esté despierto, de acuerdo con un modo de realización de la presente divulgación. El dispositivo móvil 110 se coloca cerca del usuario 210, por ejemplo, se coloca sobre una mesa 230 cerca de una cama 220 como se ilustra en la **FIG. 2**.

[0020] En este ejemplo, el usuario 210 se ha despertado por la mañana pero todavía está sentado en la cama 220. En este caso, un sensor de iluminación en el dispositivo móvil 110 puede generar datos de contexto que indican un entorno brillante del dispositivo móvil 110. Además, una unidad de reloj en el dispositivo móvil 110 también puede proporcionar una hora actual (por ejemplo, 7:00 AM) como datos de contexto. En algunos modos de realización, el tiempo actual puede compararse con un tiempo establecido en el dispositivo móvil 110 por el usuario 210 que indica cuándo se despierta típicamente. Además, un sensor de sonido en el dispositivo móvil 110 también captura un sonido ambiental silencioso que tiene una intensidad pequeña como sonido de entrada, que se proporciona como datos de contexto. Además, una unidad GPS en el dispositivo móvil 110 puede determinar que el dispositivo móvil 110 está en casa del usuario 210, que también se usa como datos de contexto.

[0021] En base a los datos de contexto generados anteriormente, el dispositivo móvil 110 puede determinar el contexto del dispositivo móvil 110 para estar por la mañana cuando es probable que el usuario 210 esté despierto. En este caso, puede suponerse que es más probable que el usuario 210 del dispositivo móvil 110 use la función de asistente de voz que haga funcionar físicamente el dispositivo móvil 110 a mano. Por ejemplo, el usuario 210 puede desear verificar el clima o el horario del día usando la función de asistente de voz en el dispositivo móvil 110 mientras está sentado en la cama 220. En consecuencia, el dispositivo móvil 110 puede ajustar el umbral de detección a un valor inferior al umbral de detección predeterminado, lo lleva a aumentar la probabilidad de detectar la palabra clave objetivo (por ejemplo, una disminución de la tasa de omisión).

[0022] Por el contrario, puede suponerse que la probabilidad de usar la función de asistente de voz es baja. La **FIG. 3** ilustra el dispositivo móvil 110 configurado para ajustar el umbral de detección para activar la función de asistente de voz si se determina que un contexto del dispositivo móvil 110 es de noche cuando es probable que un usuario 310 esté dormido (es decir, un contexto de sueño), de acuerdo con un modo de realización de la presente divulgación. En la **FIG. 3**, el dispositivo móvil 110 se coloca cerca del usuario 310, por ejemplo, se coloca sobre una mesa 330 cerca de una cama 320.

[0023] En el ejemplo de la **FIG. 3**, el usuario 310 está durmiendo en la cama 320 por la noche. En este caso, un sensor de iluminación en el dispositivo móvil 110 puede detectar un entorno oscuro del dispositivo móvil 110 y proporcionar datos de contexto que indiquen el entorno oscuro. Además, una unidad de reloj en el dispositivo móvil 110 también puede proporcionar una hora actual (por ejemplo, 3:00 AM) como datos de contexto. En algunos modos de realización, el tiempo actual puede compararse con un intervalo de tiempo establecido en el dispositivo móvil 110 por el usuario 310, que indica cuándo él o ella típicamente se va a dormir y se despierta. Además, un sensor de sonido en el dispositivo móvil 110 también puede capturar un sonido ambiental silencioso que tenga una intensidad pequeña como sonido de entrada y proporcionarlo como datos de contexto. Además, la localización del dispositivo móvil 110 puede determinarse mediante una unidad GPS en el dispositivo móvil 110 y usarse como datos de contexto (por ejemplo, en el hogar del usuario 310).

[0024] En base a los datos de contexto generados anteriormente, el dispositivo móvil 110 puede determinar que el contexto del dispositivo móvil 110 es por la noche cuando es probable que el usuario 310 esté dormido. En este caso, puede suponerse que es menos probable que el usuario 310 del dispositivo móvil 110 use la función de asistente de voz que cuando el usuario 310 está despierto durante el día. En consecuencia, el dispositivo móvil 110 puede ajustar el umbral de detección a un valor superior al umbral de detección predeterminado. El aumento del umbral de detección lleva a reducir la probabilidad de activar erróneamente la función de asistente de voz (por ejemplo, una disminución de la tasa de activación falsa).

[0025] También puede suponerse que la probabilidad de usar la función de asistente de voz es baja en algunos otros contextos. La **FIG. 4** ilustra el dispositivo móvil 110 configurado para ajustar el umbral de detección para activar la función de asistente de voz cuando se determina que el contexto del dispositivo móvil 110 está en una bolsa 410, de acuerdo con un modo de realización de la presente divulgación. En el ejemplo ilustrado, aunque el dispositivo móvil 110 se coloca en la bolsa 410 (como se indica mediante un círculo punteado), puede colocarse dentro de cualquier otro tipo de recipientes, tales como un bolsillo de ropa, una bolsita y similares.

[0026] En el ejemplo de la **FIG. 4**, se usan diversos sensores para generar datos de contexto que indiquen el contexto del dispositivo móvil 110. Por ejemplo, un sensor de iluminación en el dispositivo móvil 110 puede detectar un entorno oscuro del dispositivo móvil 110 dentro de la bolsa 410 y generar datos de contexto que indiquen el entorno oscuro. Además, un sensor de proximidad en el dispositivo móvil 110, que está configurado para detectar objetos localizados cerca del sensor, puede usarse para generar datos de contexto que indiquen que existe un objeto cerca del dispositivo móvil 110. En este ejemplo, el sensor de proximidad puede detectar la existencia de

objetos cercanos, por ejemplo, una superficie interna de la bolsa 410. Además, un sensor de sonido del dispositivo móvil 110 puede capturar el sonido de raspado o pulsación como un sonido de entrada, lo que puede causarse por la fricción entre el dispositivo móvil 110 y la superficie interna circundante de la bolsa 410. El sonido de entrada capturado también se usa como datos de contexto para determinar el contexto del dispositivo móvil 110.

**[0027]** En base a los datos de contexto generados anteriormente, el dispositivo móvil 110 puede determinar que el contexto está en un recipiente tal como la bolsa 410. En este contexto, puede suponerse que es menos probable que un usuario del dispositivo móvil 110 en la **FIG. 4** use la función de asistente de voz que cuando el dispositivo móvil 110 está colocado fuera de dicho recipiente (por ejemplo, la bolsa 410). En consecuencia, el dispositivo móvil 110 puede ajustar el umbral de detección a un valor superior al umbral de detección predeterminado. El aumento del umbral de detección lleva a reducir la probabilidad de activar erróneamente la función de asistente de voz (por ejemplo, una disminución de la tasa de activación falsa).

**[0028]** En algunos modos de realización, el dispositivo móvil 110 puede detectar un contexto donde un estado operativo del dispositivo móvil 110 cambia de un estado inactivo a un estado activo. La **FIG. 5** ilustra el dispositivo móvil 110 configurado para ajustar un umbral de detección para activar la función de asistente de voz en un contexto de recepción de un mensaje de texto, de acuerdo con un modo de realización de la presente divulgación. Inicialmente, el dispositivo móvil 110 puede estar en el estado inactivo en el que un procesador de aplicaciones puede estar en un estado inactivo para conservar energía para el dispositivo móvil 110. Cuando el dispositivo móvil 110 recibe el mensaje de texto de una red externa durante el estado inactivo, el procesador de la aplicación puede activarse para ejecutar una aplicación de mensaje y emitir una notificación de un remitente (es decir, "JOHN DOE") y el mensaje de texto (es decir, "HOLA, ¿PUEDES HABLAR AHORA?") en una pantalla 510 del dispositivo móvil 110. El estado operativo del dispositivo móvil 110 también puede cambiar del estado inactivo al estado activo cuando el dispositivo móvil 110 reciba un mensaje de correo electrónico, un mensaje multimedia, una notificación automática y similares. Además, el dispositivo móvil 110 también puede cambiar del estado inactivo al estado activo en respuesta a un acontecimiento de alarma (por ejemplo, una alarma de temporizador).

**[0029]** Cuando el estado operativo del dispositivo móvil 110 o el procesador de la aplicación cambie del estado inactivo al estado activo, es más probable que un usuario del dispositivo móvil 110 use la función de asistente de voz para realizar una operación. En el ejemplo de la **FIG. 5**, el usuario puede usar el dispositivo móvil 110 para llamar al remitente del mensaje de texto a través de la función de asistente de voz. En consecuencia, puede suponerse que la probabilidad de usar la función de asistente de voz es mayor que cuando el dispositivo móvil 110 está en el estado inactivo. Por tanto, cuando el dispositivo móvil 110 detecta un cambio en el estado operativo desde un estado inactivo a un estado activo, ajusta el umbral de detección para que sea inferior al umbral de detección predeterminado. El umbral de detección reducido aumenta la probabilidad de detectar la palabra clave objetivo (por ejemplo, una disminución de la tasa de omisión).

**[0030]** La **FIG. 6** ilustra un diagrama de bloques del dispositivo móvil 110 configurado para detectar una palabra clave objetivo a partir de un flujo de sonido de entrada ajustando un umbral para detectar la palabra clave objetivo en base a un contexto del dispositivo móvil 110, de acuerdo con un modo de realización de la presente divulgación. Como se usa en el presente documento, el término "flujo de sonido" se refiere a una secuencia de una o más señales de sonido o datos de sonido. Como se ilustra en la **FIG. 6**, el dispositivo móvil 110 puede incluir una unidad de sensor 610, una unidad I/O (entrada/salida) 620, una unidad de almacenamiento 630, una unidad de comunicación 640 y un procesador 650. El dispositivo móvil 110 puede ser cualquier dispositivo adecuado equipado con una capacidad de captura y de procesamiento de sonido tal como un teléfono celular, un teléfono inteligente, un ordenador portátil, un ordenador personal tipo *tablet*, un dispositivo de videojuegos, un reproductor multimedia, etc.

**[0031]** El procesador 650 puede incluir un procesador de señales digitales (DSP) 680, una unidad de asistente de voz 688, una unidad de reloj 690 y una unidad de programación 692, y puede ser un procesador de aplicaciones o una unidad central de procesamiento (CPU) para gestionar y hacer funcionar el dispositivo móvil 110. La unidad de reloj 690 implementa una función de reloj, una función de temporizador y similares. Por ejemplo, la unidad de reloj 690 puede emitir una hora actual a un usuario del dispositivo móvil 110 o notificar al usuario cuando se alcance una hora establecida (por ejemplo, una alarma de temporizador) a través de una unidad de pantalla y/o una unidad de altavoz de la unidad I/O 620. Además, la unidad de programación 692 almacena y gestiona programaciones (por ejemplo, citas, eventos, etc.) que se introducen por el usuario. La unidad de programación 692 puede realizar un seguimiento de los tiempos de las programaciones almacenadas y proporcionar notificaciones asociadas al usuario a través de la unidad I/O 620.

**[0032]** El DSP 680 puede incluir un detector de habla 682, una unidad de ajuste de umbral 684 y una unidad de activación por voz 686. En un modo de realización, el DSP 680 es un procesador de baja potencia para reducir el consumo de energía en el procesamiento de un flujo de sonido. En esta configuración, la unidad de ajuste de umbral 684 está configurada para ajustar el umbral de detección en base al contexto del dispositivo móvil 110. La unidad de activación por voz 686 está configurada para activar la unidad de asistente de voz 688 cuando se detecte la palabra clave objetivo en el flujo de sonido en base al umbral de detección ajustado.

**[0033]** La unidad de sensor 610 puede incluir un sensor de sonido 660 y un sensor de contexto 670. El sensor de sonido 660 está configurado para recibir y proporcionar el flujo de sonido de entrada al detector de habla 682 y la unidad de ajuste de umbral 684 en el DSP 680. En algunos modos de realización, el sensor de sonido 660 activa el detector de habla 682 y la unidad de ajuste de umbral 684 en el DSP 680 cuando el flujo de sonido de entrada recibido es mayor en intensidad que una intensidad de sonido de umbral. Tras la activación, el flujo de sonido de entrada también se proporciona desde el sensor de sonido 660 al detector de habla 682 y a la unidad de ajuste de umbral 684.

**[0034]** El sensor de contexto 670 en la unidad de sensor 610 puede incluir una variedad de sensores para generar datos de sensor indicativos del contexto del dispositivo móvil 110. Los datos de sensor se proporcionan luego a la unidad de ajuste de umbral 684. En un modo de realización, el sensor de contexto 670 también puede activar la unidad de ajuste de umbral 684 antes de que se proporcionen los datos de sensor. El sensor de sonido 660 y el sensor de contexto 670 se describirán con más detalle con referencia a la **FIG. 7** a continuación. La unidad de comunicación 640, la unidad de reloj 690 y la unidad de programación 692 pueden configurarse para proporcionar a la unidad de ajuste de umbral 684 una notificación que indique un cambio en un estado operativo del procesador 650 desde un estado inactivo a un estado activo, como se describirá con más detalle con referencia a la **FIG. 8** a continuación.

**[0035]** La unidad de ajuste de umbral 684 está configurada para determinar un contexto del dispositivo móvil 110 en base al flujo de sonido de entrada recibido, los datos de sensor y/o una o más notificaciones, que pueden denominarse colectivamente datos de contexto. Después de determinar el contexto del dispositivo móvil 110, la unidad de ajuste de umbral 684 ajusta el umbral de detección de acuerdo con el contexto determinado. Por ejemplo, si el contexto del dispositivo móvil 110 indica que es probable que se use la unidad de asistente de voz 688 (por ejemplo, en los contextos como se describe anteriormente con referencia a las **FIGS. 1, 2 y 5**), la unidad de ajuste de umbral 684 ajusta el umbral de detección para que sea bajo, por ejemplo, inferior que el umbral de detección predeterminado. Por otro lado, si el contexto del dispositivo móvil 110 indica que no es probable que se use la unidad de asistente de voz 688 (por ejemplo, en los contextos descritos anteriormente con referencia a las **FIGS. 3 y 4**), la unidad de ajuste de umbral 684 ajusta el umbral de detección para que sea alto, por ejemplo, superior que el umbral de detección predeterminado.

**[0036]** En algunos modos de realización, los valores de umbral de detección pueden determinarse para diferentes contextos y almacenarse en la unidad de almacenamiento 630 del dispositivo móvil 110. Por ejemplo, los contextos y los valores de umbral de detección asociados pueden almacenarse como una tabla de búsqueda o una estructura de base de datos. En el caso de la tabla de búsqueda, la unidad de ajuste de umbral 684 puede acceder a la tabla usando un contexto determinado como índice para obtener el valor umbral de detección asociado. La unidad de ajuste de umbral 684 proporciona entonces el valor umbral de detección obtenido como el umbral de detección ajustado a la unidad de activación por voz 686. La unidad de almacenamiento 630 en el dispositivo móvil 110 puede implementarse usando cualquier dispositivo de almacenamiento o de memoria adecuado tal como una RAM (memoria de acceso aleatorio), una ROM (memoria de solo lectura), una EEPROM (memoria de solo lectura programable y borrable eléctricamente), una memoria flash o una SSD (unidad de estado sólido).

**[0037]** Para su uso en la detección de la palabra clave objetivo, la unidad de almacenamiento 630 también puede almacenar la palabra clave objetivo. En un modo de realización, la palabra clave objetivo puede dividirse en una pluralidad de unidades de sonido básicas tales como teléfonos, fonemas o subunidades de las mismas, y puede generarse una pluralidad de porciones que representen la palabra clave objetivo en base a las unidades básicas de sonido. Cada porción de la palabra clave objetivo se asocia entonces con un estado bajo un modelo de cadena de Markov, tal como un modelo oculto de Markov (HMM), un modelo semiMarkov (SMM) o una combinación de ambos. En este caso, la unidad de almacenamiento 630 puede almacenar información de estado en una pluralidad de estados asociados con la pluralidad de porciones de la palabra clave objetivo, que incluye información de transición de cada uno de los estados al siguiente estado, incluyendo ella misma.

**[0038]** El detector de habla 682 en el DSP 680, cuando se activa, recibe el flujo de sonido de entrada desde el sensor de sonido 660. A continuación, el detector de habla 682 extrae una pluralidad de características de sonido del flujo de sonido de entrada recibido y determina si las características de sonido extraídas son indicativas de un sonido de interés tal como el habla usando cualquier procedimiento de clasificación de sonido adecuado tal como un clasificador basado en el modelo de mezclas gaussianas (GMM), una red neuronal artificial, un HMM, un modelo gráfico, una máquina de vectores de soporte (SVM) y similares. Si se determina que el flujo de sonido de entrada recibido es un sonido de interés, el detector de habla 682 activa la unidad de activación por voz 686 y el flujo de sonido de entrada recibido se proporciona a la unidad de activación por voz 686. En algunos otros modos de realización, el detector de habla 682 puede omitirse en el DSP 680. En este caso, cuando el flujo de sonido de entrada recibido es mayor en intensidad que el umbral de intensidad, el sensor de sonido 660 activa la unidad de activación por voz 686 y proporciona el flujo de sonido de entrada recibido directamente a la unidad de activación por voz 686.

**[0039]** Como se describe anteriormente, la unidad de activación por voz 686 está provista del umbral de

detección ajustado y del flujo de sonido de entrada desde la unidad de ajuste de umbral 684 y el detector de habla 682. Al recibir el umbral de detección ajustado y el flujo de sonido de entrada, la unidad de activación por voz 686 detecta la palabra clave objetivo del flujo de sonido de entrada en base al umbral de detección. A medida que se recibe el flujo de sonido de entrada, la unidad de activación por voz 686 puede extraer secuencialmente una pluralidad de características de sonido del flujo de sonido de entrada. Además, la unidad de activación por voz 686 puede procesar cada una de la pluralidad de características de sonido extraídas, y obtener la información de estado que incluye la pluralidad de estados y la información de transición para la palabra clave objetivo desde la unidad de almacenamiento 630. Para cada característica de sonido procesada, puede determinarse un puntaje de observación para cada uno de los estados mediante el uso de cualquier modelo de probabilidad adecuado tal como un GMM, una red neuronal y un SVM.

**[0040]** A partir de la información de transición, la unidad de activación por voz 686 puede obtener puntajes de transición de cada uno de los estados a un estado siguiente en una pluralidad de secuencias de estado que son posibles para la palabra clave objetivo. Después de determinar los puntajes de observación y obtener los puntajes de transición, la unidad de activación por voz 686 determina los puntajes para las posibles secuencias de estado. En un modo de realización, el puntaje más alto entre los puntajes determinados puede usarse como un puntaje de palabra clave para la característica de sonido procesada. Si el puntaje de palabra clave para la característica de sonido procesada es mayor que el umbral de detección (que también puede denominarse puntaje de palabra clave de umbral), la unidad de activación por voz 686 detecta el flujo de sonido de entrada como la palabra clave objetivo. Al detectar la palabra clave objetivo, la unidad de activación por voz 686 genera y transmite una señal de activación para encender la unidad de asistente de voz 688, que está asociada con la palabra clave objetivo.

**[0041]** La unidad de asistente de voz 688 se activa en respuesta a la señal de activación de la unidad de activación por voz 686. Una vez activada, la unidad de asistente de voz 688 puede emitir un mensaje tal como "¿PUEDO AYUDARTE?" en la unidad de pantalla y/o a través de la unidad de altavoz de la unidad I/O 620. En respuesta, un usuario puede pronunciar comandos de voz para activar diversas funciones asociadas del dispositivo móvil 110. Por ejemplo, cuando se recibe un comando de voz para la búsqueda en Internet, la unidad de asistente de voz 688 puede reconocer el comando de voz como un comando de búsqueda y realizar una búsqueda web a través de la unidad de comunicación 640 a través de una red externa 694. Aunque el modo de realización ilustrado activa la unidad de asistente de voz 688 en respuesta a la detección de la palabra clave objetivo, también puede activar cualquier otra aplicación o característica en respuesta a la detección de una palabra clave objetivo asociada.

**[0042]** La **FIG. 7** ilustra un diagrama de bloques más detallado de la unidad de sensor 610 configurada para emitir un flujo de sonido y datos de sensor para su uso en la determinación de un contexto del dispositivo móvil 110, de acuerdo con un modo de realización de la presente divulgación. El sensor de sonido 660 recibe un flujo de sonido de entrada, que luego se proporciona al detector de habla 682 en el DSP 680. El flujo de sonido de entrada recibido también se proporciona a la unidad de ajuste de umbral 684 en el DSP 680 como datos de contexto. El sensor de sonido 660 puede incluir uno o más micrófonos o cualquier otro tipo de sensores de sonido que puedan usarse para recibir, captar, detectar y/o percibir el flujo de sonido de entrada. Además, el sensor de sonido 660 puede emplear cualquier software y/o hardware adecuado para realizar dichas funciones. En un modo de realización, el sensor de sonido 660 puede configurarse para recibir el flujo de sonido de entrada periódicamente de acuerdo con un ciclo de trabajo.

**[0043]** Como se ilustra en la **FIG. 7**, el sensor de contexto 670 puede incluir un sensor de localización 710, un acelerómetro 720, un sensor de proximidad 730 y un sensor de iluminación 740. El sensor de localización 710, que puede implementarse mediante un dispositivo GPS, está configurado para generar datos de sensor asociados con el dispositivo móvil 110 tales como una localización, una dirección, una velocidad a la cual está moviéndose el dispositivo móvil 110 o similares. El acelerómetro 720 genera datos de sensor que indican aceleración, orientación y similares del dispositivo móvil 110.

**[0044]** El sensor de proximidad 730 está configurado para detectar un objeto localizado cerca del dispositivo móvil 110, por ejemplo, usando tecnologías magnéticas, ópticas o capacitivas. Aunque un sensor de proximidad 730 se ilustra en la **FIG. 7**, el sensor de contexto 670 puede incluir cualquier número adecuado de sensores de proximidad para detectar objetos en múltiples localizaciones del dispositivo móvil 110 (por ejemplo, en los lados frontal, trasero, derecho e izquierdo del dispositivo móvil 110). El sensor de iluminación 740 es un detector de luz configurado para detectar una condición de iluminación, es decir, brillo, del entorno del dispositivo móvil 110. Los datos de sensor generados por el sensor de localización 710, el acelerómetro 720, el sensor de proximidad 730 y el sensor de iluminación 740 se proporcionan a la unidad de ajuste de umbral 684 y se usan como datos de contexto para determinar el contexto del dispositivo móvil 110.

**[0045]** La **FIG. 8** ilustra un diagrama de bloques de la unidad de ajuste de umbral 684 configurada para ajustar un umbral de detección para activar la unidad de asistente de voz 688 determinando un contexto del dispositivo móvil 110, de acuerdo con un modo de realización de la presente divulgación. La unidad de ajuste de umbral 684 puede incluir una unidad de inferencia de contexto 810 y una unidad de determinación de umbral 820. La unidad de inferencia de contexto 810 determina el contexto del dispositivo móvil 110 y la unidad de determinación de umbral 820 ajusta el umbral de detección en base al contexto del dispositivo móvil 110. El umbral de detección



ajustado se proporciona luego a la unidad de activación por voz 686 para su uso en la detección de una palabra clave objetivo a partir de un flujo de sonido de entrada.

5 **[0046]** Para determinar el contexto del dispositivo móvil 110, la unidad de inferencia de contexto 810 recibe el flujo de sonido de entrada del sensor de sonido 660 y los datos de sensor del sensor de contexto 670. La unidad de inferencia de contexto 810 también está configurada para recibir una o más notificaciones de la unidad de reloj 690, la unidad de programación 692 y la unidad de comunicación 640. Además, la unidad de reloj 690 puede proporcionar una hora actual a la unidad de inferencia de contexto 810. El flujo de sonido de entrada, los datos de sensor, las notificaciones y la hora actual se usan por la unidad de inferencia de contexto 810 como datos de contexto para determinar el contexto del dispositivo móvil 110.

15 **[0047]** Como se describe con referencia a la **FIG. 6** anteriormente, las notificaciones proporcionadas a la unidad de inferencia de contexto 810 pueden indicar información de estado de un dispositivo móvil, tal como un cambio en un estado operativo del procesador 650 de un estado inactivo a un estado activo. Por ejemplo, la unidad de comunicación 640 puede proporcionar dicha notificación al recibir un mensaje electrónico a través de la red 694 en la **FIG. 6**. El mensaje electrónico puede incluir cualquier mensaje o notificación recibida a través de la red 694, tal como un mensaje de correo electrónico, un mensaje de texto, un mensaje multimedia, una notificación automática y similares. La unidad de reloj 690 puede configurarse para alertar a un usuario a una hora especificada (por ejemplo, una alarma). Cuando se alcanza el tiempo especificado, la unidad de reloj 690 activa el procesador 650 y se proporciona una notificación del cambio en el estado operativo del procesador 650 a la unidad de inferencia de contexto 810. Además, la unidad de programación 692 puede realizar un seguimiento de los tiempos de las programaciones o acontecimientos almacenados en el dispositivo móvil 110 y proporcionar una notificación a la unidad de inferencia de contexto 810 en dichos momentos o en un momento predeterminado antes de dichos momentos.

25 **[0048]** La unidad de inferencia de contexto 810 analiza los datos de contexto recibidos y determina el contexto del dispositivo móvil 110. Al determinar el contexto del dispositivo móvil 110, la unidad de inferencia de contexto 810 puede inferir el contexto a partir del sonido de entrada, los datos de sensor, las notificaciones, la hora actual o cualquier combinación de los mismos. Por ejemplo, puede inferirse que el contexto del dispositivo móvil 110 está en un vehículo en base al flujo de sonido de entrada y a los datos de sensor que pueden incluir una localización, una velocidad y/o una aceleración del dispositivo móvil 110, como se describe con referencia a la **FIG. 1** anteriormente.

35 **[0049]** En algunos modos de realización, diversos contextos del dispositivo móvil 110 pueden determinarse y almacenarse de acuerdo con una base de datos de contexto que asocie los contextos con datos de contexto que pueden recibirse por la unidad de inferencia de contexto 810. Entre los contextos predeterminados en la base de datos de contexto, la unidad de inferencia de contexto 810 puede seleccionar un contexto que corresponda a los datos de contexto recibidos como contexto actual del dispositivo móvil 110. La unidad de inferencia de contexto 810 genera entonces un indicador de contexto para el contexto seleccionado y proporciona el indicador de contexto a la unidad de determinación de umbral 820.

45 **[0050]** Al recibir el indicador de contexto de la unidad de inferencia de contexto 810, la unidad de determinación de umbral 820 ajusta el umbral de detección en base al indicador de contexto recibido. En algunos modos de realización, los valores para el umbral de detección pueden preestablecerse para los contextos predeterminados y almacenarse en la unidad de almacenamiento 630. La unidad de determinación de umbral 820 puede acceder a la unidad de almacenamiento 630 para obtener un valor de umbral de detección asociado con el contexto identificado por el indicador de contexto.

50 **[0051]** En un modo de realización, el valor umbral de detección puede establecerse en base a la probabilidad de usar la unidad de asistente de voz 688 en cada uno de los contextos. Por ejemplo, si la probabilidad de usar la unidad de asistente de voz 688 es alta (por ejemplo, en los contextos descritos anteriormente con referencia a las **FIGS. 1, 2 y 5**), la unidad de ajuste de umbral 684 puede disminuir el umbral de detección (por ejemplo, ser inferior al umbral de detección predeterminado que está predeterminado para contextos normales o no identificados). Por otro lado, si la probabilidad de usar la unidad de asistente de voz 688 es baja (por ejemplo, en los contextos descritos anteriormente con referencia a las **FIGS. 3 y 4**), la unidad de ajuste de umbral 684 puede aumentar el umbral de detección (por ejemplo, para ser superior que el umbral de detección predeterminado). Como resultado, el umbral de detección puede ajustarse en asociación con la probabilidad de usar la unidad de asistente de voz 688 en el contexto determinado del dispositivo móvil 110. El umbral de detección ajustado se proporciona luego a la unidad de activación por voz 686 para usarse en la detección de una palabra clave objetivo para activar la unidad de asistente de voz 688.

65 **[0052]** La **FIG. 9** es un diagrama de flujo de un procedimiento 900, realizado en la unidad de ajuste de umbral 684 del dispositivo móvil 110, para determinar el umbral de detección para activar la unidad de asistente de voz 688 infiriendo un contexto del dispositivo móvil 110, de acuerdo con un modo de realización de la presente divulgación. La unidad de inferencia de contexto 810 en la unidad de ajuste de umbral 684 recibe datos de contexto, en 910. En base a los datos de contexto recibidos, el contexto del dispositivo móvil 110 puede determinarse en la

unidad de inferencia de contexto 810, en 920. La unidad de inferencia de contexto 810 también puede generar un indicador de contexto asociado con el contexto determinado, que se proporciona a la unidad de determinación de umbral 820.

5 **[0053]** La unidad de determinación de umbral 820, en 930, ajusta el umbral de detección de acuerdo con el contexto del dispositivo móvil 110 que se identifica por el indicador de contexto recibido de la unidad de inferencia de contexto 810. Para ajustar el umbral de detección, la unidad de determinación de umbral 820 puede acceder a la unidad de almacenamiento 630 para obtener un valor de umbral de detección asociado con el contexto del dispositivo móvil 110. Una vez que se ajusta el umbral de detección, la unidad de determinación de umbral 820  
10 puede proporcionar el umbral de detección ajustado a la unidad de activación por voz 686, en 940. El umbral de detección ajustado se usa para detectar la palabra clave objetivo por la unidad de activación por voz 686.

**[0054]** En algunos modos de realización, el umbral de detección ajustado puede restablecerse al umbral de detección predeterminado. La **FIG. 10** es un diagrama de flujo de un procedimiento 1000, realizado en la unidad de ajuste de umbral 684 del dispositivo móvil 110, para generar un umbral de detección ajustado y restablecer el umbral de detección a un umbral de detección predeterminado después de un período de tiempo especificado, de acuerdo con un modo de realización de la presente divulgación. De la misma manera que se describe con respecto a la **FIG. 9** anteriormente, la unidad de inferencia de contexto 810 recibe datos de contexto en 910, y determina un contexto del dispositivo móvil 110 en 920. La unidad de determinación de umbral 820 ajusta el umbral de detección en base al contexto en 930, y proporciona el umbral de detección ajustado a la unidad de activación por voz 686 en 940.

**[0055]** Puede establecerse un temporizador para un período de tiempo específico, en 1010, por ejemplo, usando la unidad de reloj 690 en el procesador 650. El período de tiempo puede estar predeterminado. La unidad de ajuste de umbral 684 monitorea el temporizador y determina si ha transcurrido el período de tiempo especificado, en 1020. De forma alternativa, la unidad de reloj 690 puede detectar si ha transcurrido el período de tiempo especificado y notificar a la unidad de ajuste de umbral 684 cuándo ha transcurrido el período de tiempo. Si ha transcurrido el período de tiempo especificado, la unidad de ajuste de umbral 684 puede restablecer el umbral de detección al umbral de detección predeterminado (por ejemplo, el umbral de detección predeterminado), en 1030. De lo contrario, si el período de tiempo especificado no ha transcurrido, el procedimiento 1000 puede volver a 1020 para determinar si el período de tiempo especificado ha transcurrido o no.

**[0056]** La **FIG. 11** ilustra un diagrama de bloques de la unidad de activación por voz 686 configurada para activar la unidad de asistente de voz 688 mediante la detección de la palabra clave objetivo en un flujo de sonido de entrada en base al umbral de detección ajustado, de acuerdo con un modo de realización de la presente divulgación. La unidad de activación por voz 686 puede incluir un extractor de características 1110, una unidad de cálculo de puntaje de palabra clave 1120 y una unidad de detección de palabras clave 1130. Cuando la unidad de detección de palabras clave 1130 en la unidad de activación por voz 686 detecta la palabra clave objetivo del flujo de sonido de entrada, genera una señal de activación para encender la unidad de asistente de voz 688.

**[0057]** Cuando el detector de habla 682 determina que un flujo de sonido de entrada es voz humana, el extractor de características 1110 recibe el flujo de sonido de entrada y extrae una o más características de sonido del flujo de sonido de entrada. En algunos modos de realización, el extractor de características 1110 segmenta el flujo de sonido de entrada en una pluralidad de tramas secuenciales de un período de tiempo igual. Por ejemplo, el flujo de sonido de entrada puede recibirse y segmentarse secuencialmente en tramas de 10 ms. En este caso, el extractor de características 1110 extrae una característica de sonido de cada una de las tramas. En un modo de realización, el extractor de características 1110 puede extraer las características de sonido de las tramas usando cualquier procedimiento de extracción de características adecuado tal como el procedimiento MFCC (coeficientes cepstrales de frecuencia de Mel). Por ejemplo, en el caso del procedimiento MFCC, los componentes de un vector n-dimensional se calculan a partir de cada una de las tramas segmentadas y el vector se usa como una característica de sonido.

**[0058]** Las características de sonido extraídas se proporcionan desde el extractor de características 1110 a la unidad de cálculo de puntaje de palabra clave 1120. Al recibir cada característica de sonido, la unidad de cálculo de puntaje de palabra clave 1120 determina un puntaje de palabra clave para la característica de sonido. El puntaje de palabra clave puede determinarse de la manera descrita anteriormente con referencia a la **FIG. 6**. El puntaje de palabra clave determinado se proporciona a la unidad de detección de palabras clave 1130.

**[0059]** La unidad de detección de palabras clave 1130 recibe el puntaje de palabra clave de la unidad de cálculo de puntaje de palabra clave 1120. La unidad de detección de palabras clave 1130 también recibe el umbral de detección ajustado desde la unidad de ajuste de umbral 684. El puntaje de palabra clave recibida se compara con el umbral de detección ajustado. Cuando el puntaje de palabra clave supera el puntaje de umbral ajustado, la unidad de detección de palabras clave 1130 genera la señal de activación para encender la unidad de asistente de voz 688. De lo contrario, la unidad de detección de palabras clave 1130 no activa la unidad de asistente de voz 688. En este caso, la unidad de detección de palabras clave 1130 puede recibir un puntaje de palabra clave posterior, que está determinada por la unidad de cálculo de puntaje de palabra clave 1120 para una próxima

característica de sonido extraída, y determinar si el puntaje de palabra clave posterior excede el umbral de detección ajustado.

5 **[0060]** La **FIG. 12** es un diagrama de flujo de un procedimiento 1200, realizado en la unidad de activación por voz 686, para activar la unidad de asistente de voz 688 mediante la detección de una palabra clave objetivo en un flujo de sonido de entrada en base a un umbral de detección ajustado, de acuerdo con un modo de realización de la presente divulgación. El procedimiento 1200 comienza en 1210 y continúa hasta 1220, donde el extractor de características 1110 en la unidad de activación por voz 686 recibe el flujo de sonido de entrada desde el detector de habla 682. El extractor de características 1110 puede extraer una característica de sonido del flujo de sonido de entrada, en 1230. La característica de sonido extraída se proporciona a la unidad de cálculo de puntaje de palabra clave 1120 en la unidad de activación por voz 686.

15 **[0061]** En un modo de realización, el flujo de sonido de entrada se segmenta secuencialmente en una pluralidad de tramas, de los cuales se extrae una pluralidad de características de sonido. En este caso, la pluralidad de características de sonido se proporciona secuencialmente a la unidad de cálculo de puntaje de palabra clave 1120. Al recibir cada característica de sonido del extractor de características 1110, la unidad de puntaje de palabra clave 1120 puede determinar un puntaje de palabra clave para la característica de sonido, en 1240. Como se describe anteriormente con referencia a la **FIG. 6**, la característica de palabra clave puede determinarse usando cualquier tecnología de reconocimiento de habla adecuada tal como un modelo de cadena de Markov.

20 **[0062]** En el procedimiento ilustrado 1200, la unidad de detección de palabras clave 1130 recibe el umbral de detección ajustado desde la unidad de ajuste de umbral 684, en 1250. En un modo de realización, el umbral de detección ajustado puede recibirse en 1250, mientras que el flujo de sonido de entrada se recibe en 1220, la característica de sonido se extrae en 1230 o el puntaje de palabra clave se determina en 1240. En otro modo de realización, la unidad de detección de palabras clave 1130 puede recibir el umbral de detección ajustado antes de que el extractor de características 1110 reciba el flujo de sonido de entrada en 1220 o después de que la unidad de cálculo de puntaje de palabra clave 1120 determine el puntaje de palabra clave en 1240.

30 **[0063]** La unidad de detección de palabras clave 1130 puede determinar, en 1260, si el puntaje de palabra clave excede el umbral de detección recibido. Si se determina que el puntaje de palabra clave no excede el umbral de detección en 1260, el procedimiento puede proceder a 1270 para recibir una próxima característica de sonido extraída por el extractor de características 1110 y determinar un puntaje de palabra clave para la próxima característica de sonido en 1240. Por otro lado, si se determina que el puntaje de palabra clave excede el umbral de detección en 1260, la unidad de detección de palabras clave 1130 puede generar y transmitir una señal de activación para encender la unidad de asistente de voz 688, en 1280. El procedimiento puede terminar entonces en 1290.

40 **[0064]** La **FIG. 13** ilustra un diagrama de bloques de un dispositivo móvil 1300 en un sistema de comunicación inalámbrica en el cual los procedimientos y aparatos para detectar una palabra clave objetivo en un flujo de sonido de entrada en base a un umbral de detección que se ajusta de acuerdo con un contexto del dispositivo móvil 1300, de acuerdo con un modo de realización de la presente divulgación. El dispositivo móvil 1300 puede ser un teléfono celular, un terminal, un teléfono, un asistente digital personal (PDA), un módem inalámbrico, un teléfono sin cable, etc. El sistema de comunicación inalámbrica puede ser un sistema de Acceso Múltiple por División de Código (CDMA), un Sistema Global para Comunicaciones Móviles (GSM), un sistema CDMA de Banda Ancha (W-CDMA), un sistema de Evolución a Largo Plazo (LTE), un sistema LTE Avanzado, y así sucesivamente.

50 **[0065]** El dispositivo electrónico 1300 puede proporcionar comunicación bidireccional por medio de una ruta de recepción y una ruta de transmisión. En la ruta de recepción, las señales transmitidas por las estaciones base se reciben por una antena 1312 y se proporcionan a un receptor (RCVR) 1314. El receptor 1314 puede condicionar y digitalizar la señal recibida, y proporcionar la señal digital condicionada y digitalizada a una sección digital 1320 para su procesamiento adicional. En la ruta de transmisión, un transmisor (TMTR) recibe datos que se van a transmitir desde una sección digital 1320, procesa y condiciona los datos, y genera una señal modulada, que se transmite por medio de la antena 1312 a las estaciones base. El receptor 1314 y el transmisor 1316 forman parte de un transceptor que soporta CDMA, GSM, W-CDMA, LTE, LTE avanzada, etc.

55 **[0066]** La sección digital 1320 incluye diversas unidades de procesamiento, interfaz y memoria, tales como, por ejemplo, un procesador de módem 1322, un ordenador con conjunto reducido de instrucciones/procesador de señales digitales (RISC/DSP) 1324, un controlador/procesador 1326, una memoria interna 1328, un codificador de audio generalizado 1332, un decodificador de audio generalizado 1334, un procesador de gráficos/pantalla 1336, y una interfaz de bus externo (EBI) 1338. El procesador de módem 1322 puede realizar un procesamiento para la transmisión y la recepción de datos, por ejemplo, codificación, modulación, demodulación y decodificación. El RISC/DSP 1324 puede realizar un procesamiento general y especializado para el dispositivo electrónico 1300. El controlador/procesador 1326 controla la operación de diversas unidades de procesamiento e interfaz dentro de la sección digital 1320. La memoria interna 1328 puede almacenar datos y/o instrucciones para diversas unidades dentro de la sección digital 1320.

**[0067]** El codificador de audio generalizado 1332 realiza la codificación para señales de entrada desde una fuente de audio 1342, un micrófono 1343, etc. El decodificador de audio generalizado 1334 realiza la decodificación de datos de audio codificados y proporciona señales de salida a un altavoz/auricular 1344. Cabe señalar que el codificador de audio generalizado 1332 y el decodificador de audio generalizado 1334 no se requieren necesariamente para la interfaz con la fuente de audio, el micrófono 1343 y el altavoz/auricular 1344 y por tanto no se muestran en el dispositivo móvil 1300. El procesador de gráficos/pantalla 1336 puede realizar un procesamiento de gráficos, vídeos, imágenes y textos, que pueden presentarse a una unidad de pantalla 1346. La EBI 1338 puede facilitar la transferencia de datos entre la sección digital 1320 y una memoria principal 1348.

**[0068]** La sección digital 1320 se implementa con uno o más procesadores, DSP, microprocesadores, RISC, etc. La sección digital 1320 también puede fabricarse en uno o más circuitos integrados específicos de la aplicación (ASIC) y/o en algún otro tipo de circuitos integrados (IC).

**[0069]** En general, cualquier dispositivo descrito en el presente documento es indicativo de diversos tipos de dispositivos, tales como un teléfono inalámbrico, un teléfono celular, un ordenador portátil, un dispositivo multimedia inalámbrico, una tarjeta de ordenador personal (PC) de comunicación inalámbrica, un PDA, un dispositivo externo o módem interno, un dispositivo que se comunica a través de un canal inalámbrico, etc. Un dispositivo puede tener diversos nombres, tales como un terminal de acceso (AT), una unidad de acceso, una unidad de abonado, una estación móvil, un dispositivo cliente, una unidad móvil, un teléfono móvil, una estación remota, un terminal remoto, una unidad remota, un dispositivo de usuario, un equipo de usuario, un dispositivo portátil, etc. Cualquier dispositivo descrito en el presente documento puede tener una memoria para almacenar instrucciones y datos, así como hardware, software, firmware o combinaciones de los mismos.

**[0070]** Las técnicas descritas en el presente documento pueden implementarse mediante diversos medios. Por ejemplo, estas técnicas se implementan en hardware, firmware, software o en una combinación de los mismos. Los expertos en la técnica apreciarán además que los diversos bloques lógicos, módulos, circuitos y etapas de algoritmo ilustrativos descritos en relación con la divulgación en el presente documento pueden implementarse como hardware electrónico, software informático o combinaciones de ambos. Para ilustrar claramente esta intercambiabilidad de hardware y software, anteriormente se han descrito diversos componentes, bloques, módulos, circuitos y etapas ilustrativos, en general, en términos de su funcionalidad. Que dicha funcionalidad se implemente como hardware o software depende de la aplicación en particular y de las restricciones de diseño impuestas al sistema global. Los expertos en la técnica pueden implementar la funcionalidad descrita de distintas formas para cada solicitud en particular, pero no se debe interpretar que dichas decisiones de implementación suponen apartarse del alcance de la presente divulgación.

**[0071]** Para una implementación en hardware, las unidades de procesamiento usadas para realizar las técnicas se implementan dentro de uno o más ASIC, DSP, dispositivos de procesamiento de señales digitales (DSPD), dispositivos de lógica programable (PLD), matrices de compuertas programables por campo (FPGA), procesadores, controladores, microcontroladores, microprocesadores, dispositivos electrónicos, otras unidades electrónicas diseñadas para realizar las funciones descritas en el presente documento, un ordenador o una combinación de los mismos.

**[0072]** Por tanto, los diversos bloques lógicos, módulos y circuitos ilustrativos descritos en relación con la divulgación en el presente documento pueden implementarse o realizarse con un procesador de uso general, un DSP, un ASIC, una FPGA o con otro dispositivo de lógica programable, lógica de transistor o de compuertas discretas, componentes de hardware discretos, o con cualquier combinación de los mismos diseñada para realizar las funciones descritas en el presente documento. Un procesador de uso general puede ser un microprocesador pero, de forma alternativa, el procesador puede ser cualquier procesador, controlador, microcontrolador o máquina de estados convencional. Un procesador también puede implementarse como una combinación de dispositivos informáticos, por ejemplo, una combinación de un DSP y un microprocesador, una pluralidad de microprocesadores, uno o más microprocesadores junto con un núcleo de DSP o cualquier otra configuración de este tipo.

**[0073]** Si se implementan en software, las funciones pueden almacenarse en, o transmitir a través de, un medio legible por ordenador como una o más instrucciones o código. Los medios legibles por ordenador incluyen tanto medios de almacenamiento informático como medios de comunicación incluyendo cualquier medio que facilite la transferencia de un programa informático de un lugar a otro. Un medio de almacenamiento puede ser cualquier medio disponible al que se pueda acceder mediante un ordenador. A modo de ejemplo y no limitado a ello, dichos medios legibles por ordenador pueden comprender RAM, ROM, EEPROM, CD-ROM u otros dispositivos de almacenamiento en disco óptico, almacenamiento en disco magnético u otro almacenamiento magnético, o cualquier otro medio que pueda usarse para transportar o almacenar el código de programa deseado en forma de instrucciones o estructuras de datos y al que pueda accederse mediante un ordenador. Además, cualquier conexión recibe apropiadamente la denominación de medio legible por ordenador. Por ejemplo, si el software se transmite desde una página web, un servidor u otra fuente remota usando un cable coaxial, un cable de fibra óptica, un par trenzado, una línea de abonado digital (DSL) o tecnologías inalámbricas tales como infrarrojos, radio y microondas, entonces el cable coaxial, el cable de fibra óptica, el par trenzado, la DSL o las tecnologías

5 inalámbricas, tales como infrarrojos, radio y microondas se incluyen en la definición de medio. Los discos, como se usan en el presente documento, incluyen el disco compacto (CD), el disco láser, el disco óptico, el disco versátil digital (DVD), el disco flexible y el disco Blu-ray, donde algunos discos habitualmente reproducen datos de forma magnética, mientras que otros discos reproducen los datos de forma óptica con láser. Las combinaciones de lo anterior también se deben incluir dentro del alcance de los medios legibles por ordenador.

10 **[0074]** La descripción previa de la divulgación se proporciona para permitir que cualquier experto en la técnica realice o use la divulgación. Diversas modificaciones a la divulgación serán fácilmente evidentes para los expertos en la técnica, y los principios genéricos definidos en el presente documento pueden aplicarse a otras variaciones sin apartarse del espíritu o alcance de la divulgación. Por tanto, la divulgación no está prevista para los ejemplos descritos en el presente documento, sino que se le debe conceder el alcance más amplio conforme a los principios y características novedosas divulgados en el presente documento.

15 **[0075]** Aunque las implementaciones ejemplares pueden referirse a usar aspectos de la presente materia objeto divulgada en el contexto de uno o más sistemas informáticos autónomos, la materia objeto no está limitada, sino que en su lugar puede implementarse en relación con cualquier entorno informático, tal como una red o un entorno informático distribuido. Aún más, pueden implementarse aspectos de la presente materia objeto divulgada en o a través de una pluralidad de chips o dispositivos de procesamiento, y el almacenamiento puede realizarse de forma similar en una pluralidad de dispositivos. Dichos dispositivos pueden incluir PC, servidores de red y dispositivos de  
20 auricular.

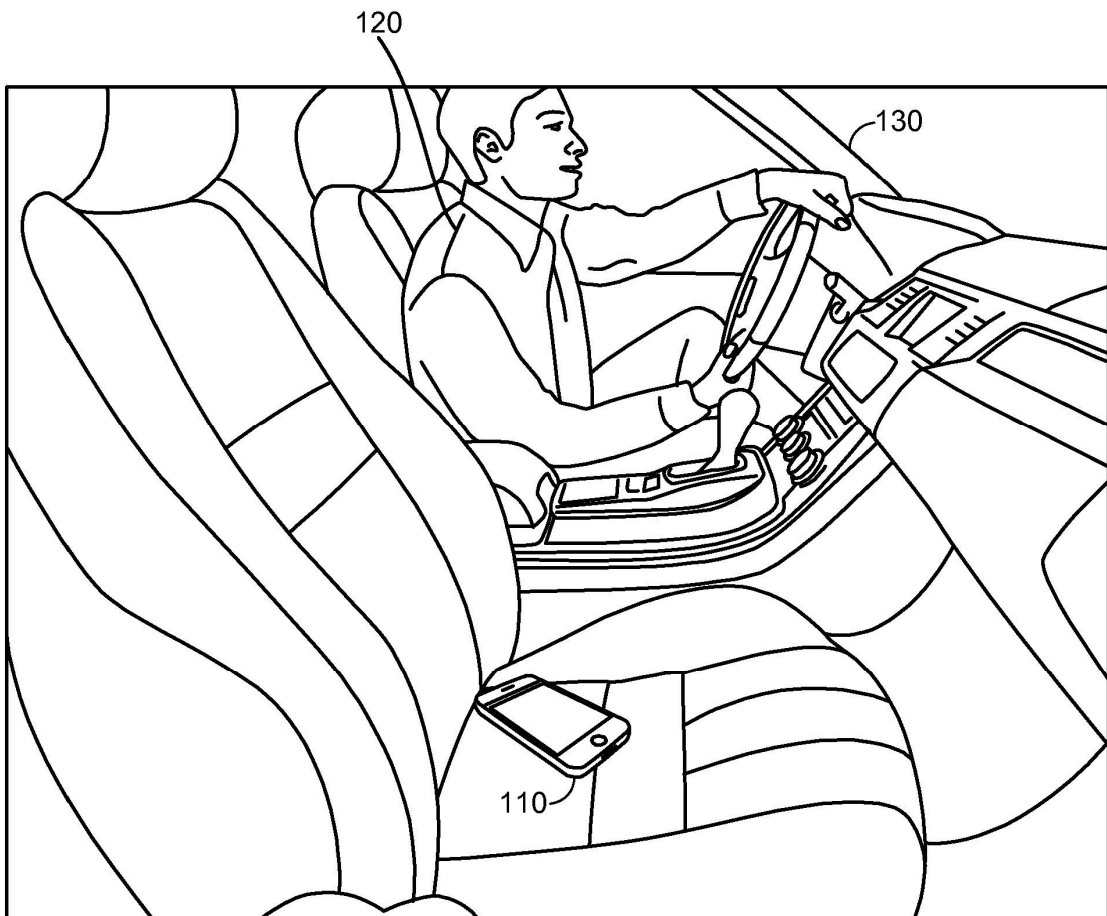
25 **[0076]** Aunque la materia objeto se ha descrito en un lenguaje específico con respecto a las características estructurales y/o acciones metodológicas, se entenderá que la materia objeto definida en las reivindicaciones adjuntas no se limita necesariamente a las características específicas o acciones que se han descrito anteriormente. En su lugar, las características y acciones específicas descritas anteriormente se divulgan como formas de ejemplo de implementación de las reivindicaciones.

**REIVINDICACIONES**

- 5           1.    Un procedimiento para activar una función de asistente de voz en un dispositivo móvil, comprendiendo el procedimiento:
- recibir un flujo de sonido de entrada por un sensor de sonido;
- determinar un contexto del dispositivo móvil;
- 10           ajustar un umbral para activar la función de asistente de voz en base al contexto;
- detectar una palabra clave objetivo del flujo de sonido de entrada en base a una característica de sonido extraída del flujo de sonido de entrada y en base al umbral ajustado; y
- 15           activar la función de asistente de voz si se detecta la palabra clave objetivo;
- caracterizado por que** ajustar el umbral comprende ajustar el umbral en base a una probabilidad de usar la función de asistente de voz en el contexto del dispositivo móvil.
- 20           2.    El procedimiento de la reivindicación 1, en el que determinar el contexto del dispositivo móvil comprende identificar el contexto del dispositivo móvil en base al flujo de sonido de entrada, y en el que la detección de la palabra clave objetivo se basa además en realizar una comparación entre una característica de sonido procesada correspondiente a la característica de sonido y al umbral ajustado.
- 25           3.    El procedimiento de la reivindicación 1, en el que determinar el contexto del dispositivo móvil comprende:
- obtener datos indicativos del contexto del dispositivo móvil de al menos uno de un sensor de aceleración, un sensor de localización, un sensor de iluminación, un sensor de proximidad, una unidad de reloj y una unidad de calendario en el dispositivo móvil; e
- 30           identificar el contexto del dispositivo móvil en base a los datos.
4.    El procedimiento de la reivindicación 1, en el que determinar el contexto del dispositivo móvil comprende:
- 35           obtener información de estado del dispositivo móvil; e
- identificar el contexto del dispositivo móvil en base a la información de estado.
- 40           5.    El procedimiento de la reivindicación 4, en el que la información de estado indica un cambio en un estado operativo de un procesador en el dispositivo móvil de un estado inactivo a un estado activo.
6.    El procedimiento de la reivindicación 1, que comprende además establecer el umbral en un umbral predeterminado cuando transcurra un período de tiempo predeterminado después de ajustar el umbral.
- 45           7.    El procedimiento de la reivindicación 1, en el que la detección de palabras clave objetivo del flujo de sonido de entrada comprende:
- calcular un grado de similitud entre el flujo de sonido de entrada y un modelo de sonido de la palabra clave objetivo;
- 50           comparar el grado de similitud con el umbral; y
- detectar la palabra clave objetivo del flujo de sonido de entrada si el grado de similitud excede el umbral.
- 55           8.    Un dispositivo móvil para activar una función de asistente de voz, que comprende:
- un sensor de sonido configurado para recibir un flujo de sonido de entrada;
- 60           una unidad de ajuste de umbral configurada para ajustar un umbral para activar la función de asistente de voz determinando un contexto, y ajustar el umbral para activar la función de asistente de voz en base al contexto; y
- una unidad de activación por voz configurada para activar la función de asistente de voz si se detecta una palabra clave objetivo del flujo de sonido de entrada en base a una característica de sonido extraída del flujo de sonido de entrada y en base al umbral ajustado;
- 65

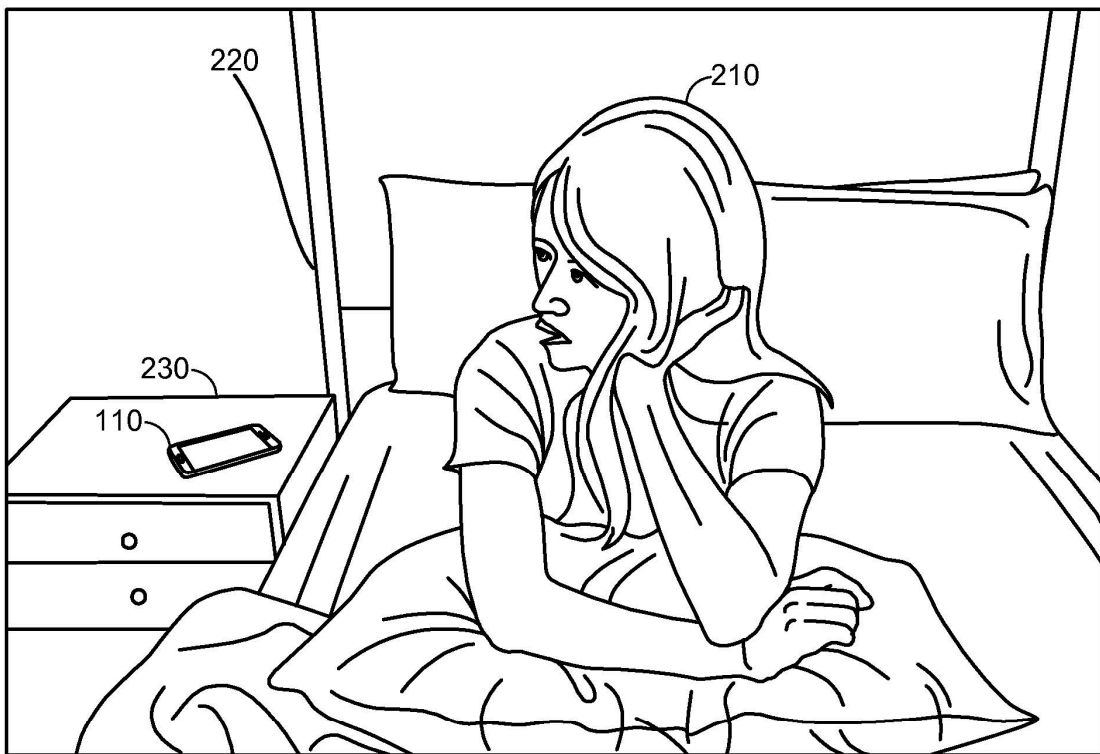
**caracterizado por que** la unidad de ajuste de umbral está configurada además para ajustar el umbral en base a una probabilidad de usar la función de asistente de voz en el contexto del dispositivo móvil.

- 5
9. El dispositivo móvil de la reivindicación 8, en el que la unidad de ajuste de umbral comprende una unidad de inferencia de contexto configurada para identificar el contexto del dispositivo móvil en base al flujo de sonido de entrada, y en el que la unidad de activación por voz está configurada además para activar la función de asistente de voz si la palabra clave objetivo se detecta en base a una comparación entre una característica de sonido procesada correspondiente a la característica de sonido y el umbral ajustado.
- 10
10. El dispositivo móvil de la reivindicación 8, en el que la unidad de ajuste de umbral comprende una unidad de determinación de contexto configurada para obtener datos indicativos del contexto del dispositivo móvil de al menos uno de un sensor de aceleración, un sensor de localización, un sensor de iluminación, un sensor de proximidad, una unidad de reloj y una unidad de calendario en el dispositivo móvil, e identificar el contexto del dispositivo móvil en base a los datos.
- 15
11. El dispositivo móvil de la reivindicación 8, en el que la unidad de ajuste de umbral comprende una unidad de determinación de contexto configurada para obtener información de estado del dispositivo móvil e identificar el contexto del dispositivo móvil en base a la información de estado.
- 20
12. El dispositivo móvil de la reivindicación 11, en el que la información de estado indica un cambio en un estado operativo de un procesador en el dispositivo móvil de un estado inactivo a un estado activo.
- 25
13. El dispositivo móvil de la reivindicación 8, en el que la unidad de ajuste de umbral está configurada además para establecer el umbral en un umbral predeterminado cuando transcurra un período de tiempo predeterminado después de ajustar el umbral.
- 30
14. El dispositivo móvil de la reivindicación 8, en el que la unidad de activación por voz está configurada además para calcular un grado de similitud entre el flujo de sonido de entrada y un modelo de sonido de la palabra clave objetivo, comparar el grado de similitud con el umbral y detectar la palabra clave objetivo desde el flujo de sonido de entrada si el grado de similitud excede el umbral.
- 35
15. Un medio de almacenamiento legible por ordenador que comprende instrucciones para activar una función de asistente de voz en un dispositivo móvil, causando las instrucciones que un procesador del dispositivo móvil realice operaciones de acuerdo con el procedimiento de las reivindicaciones 1 a 7.

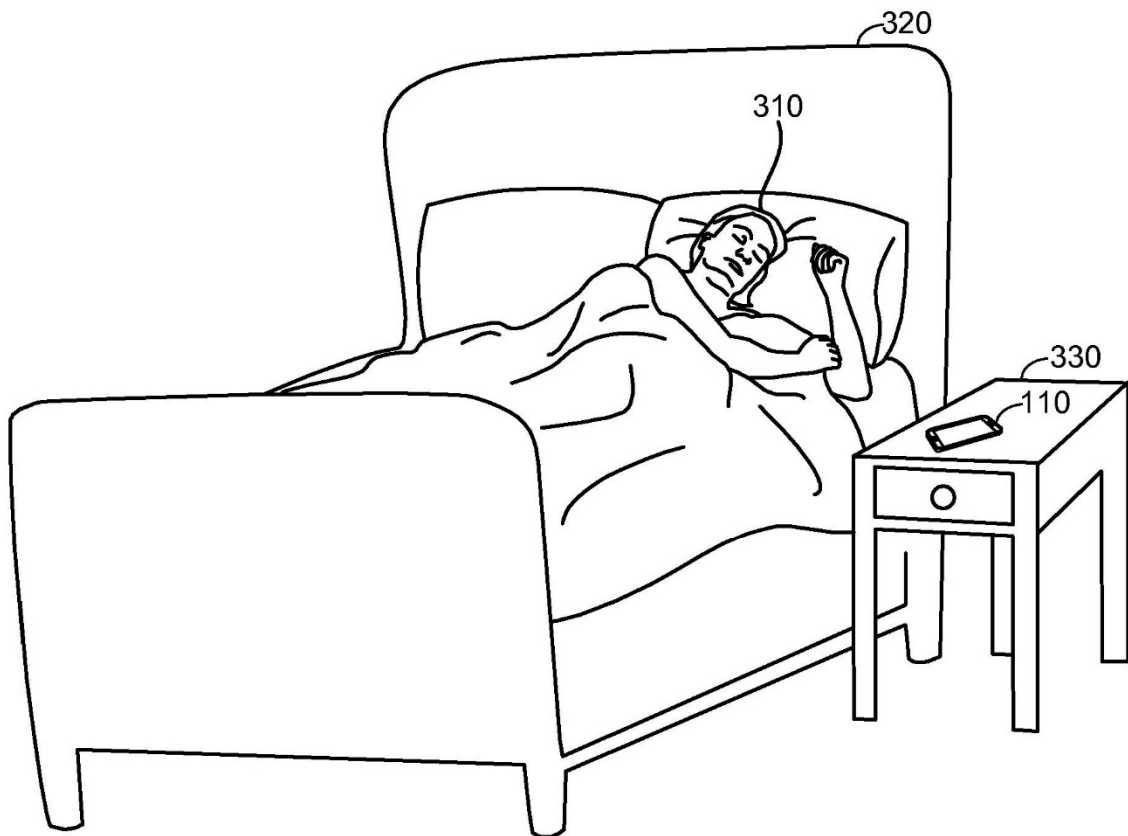


**FIG. 1**





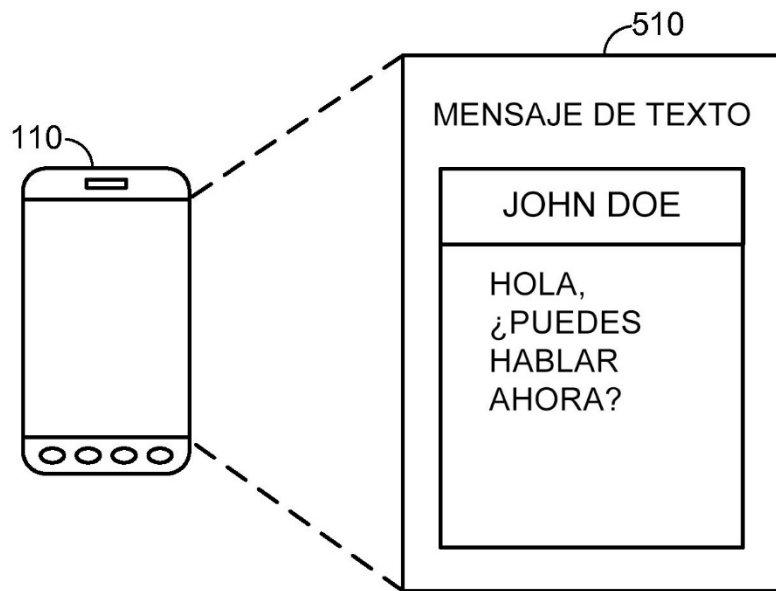
**FIG. 2**



**FIG. 3**



**FIG. 4**



**FIG. 5**

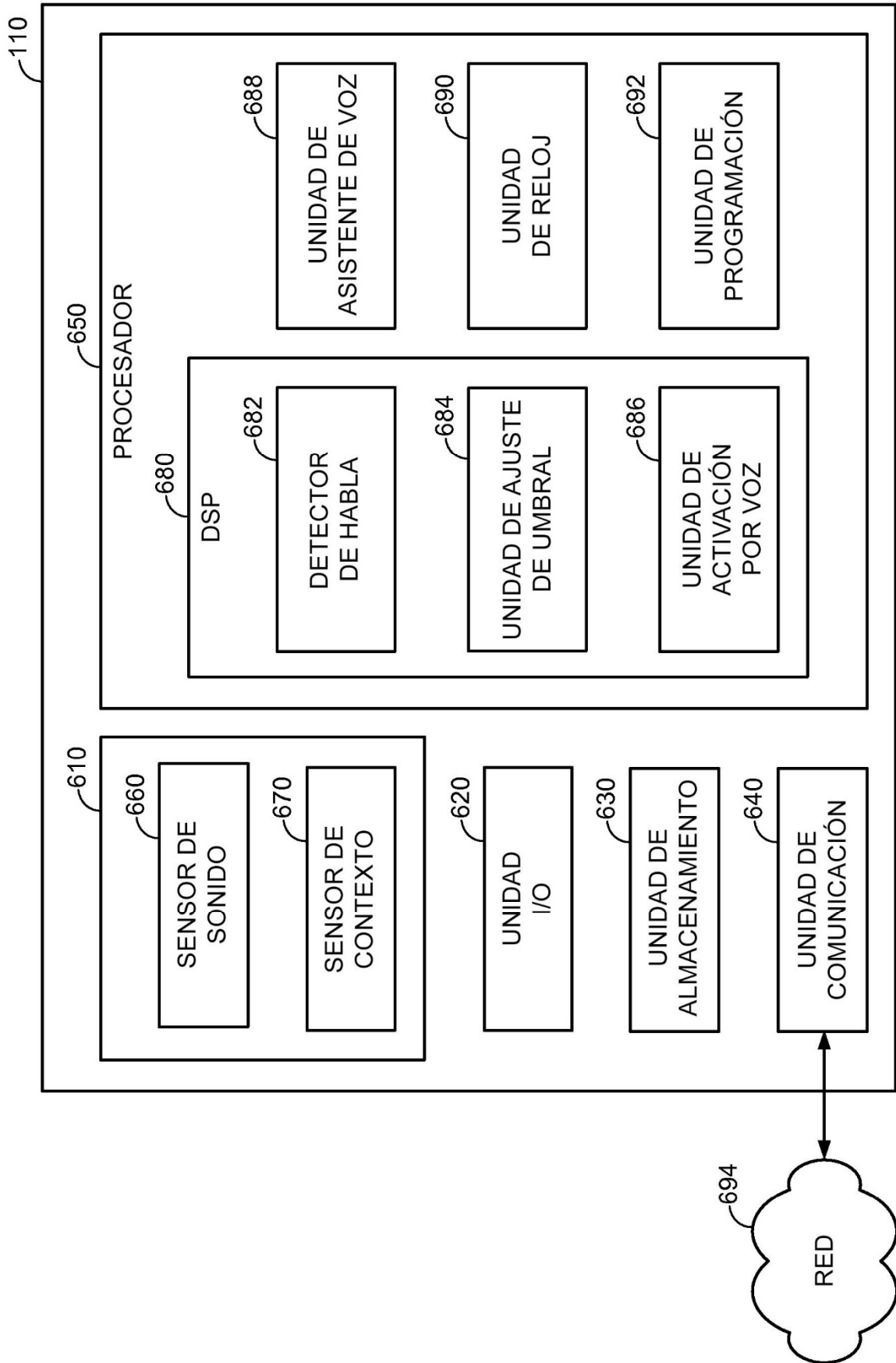
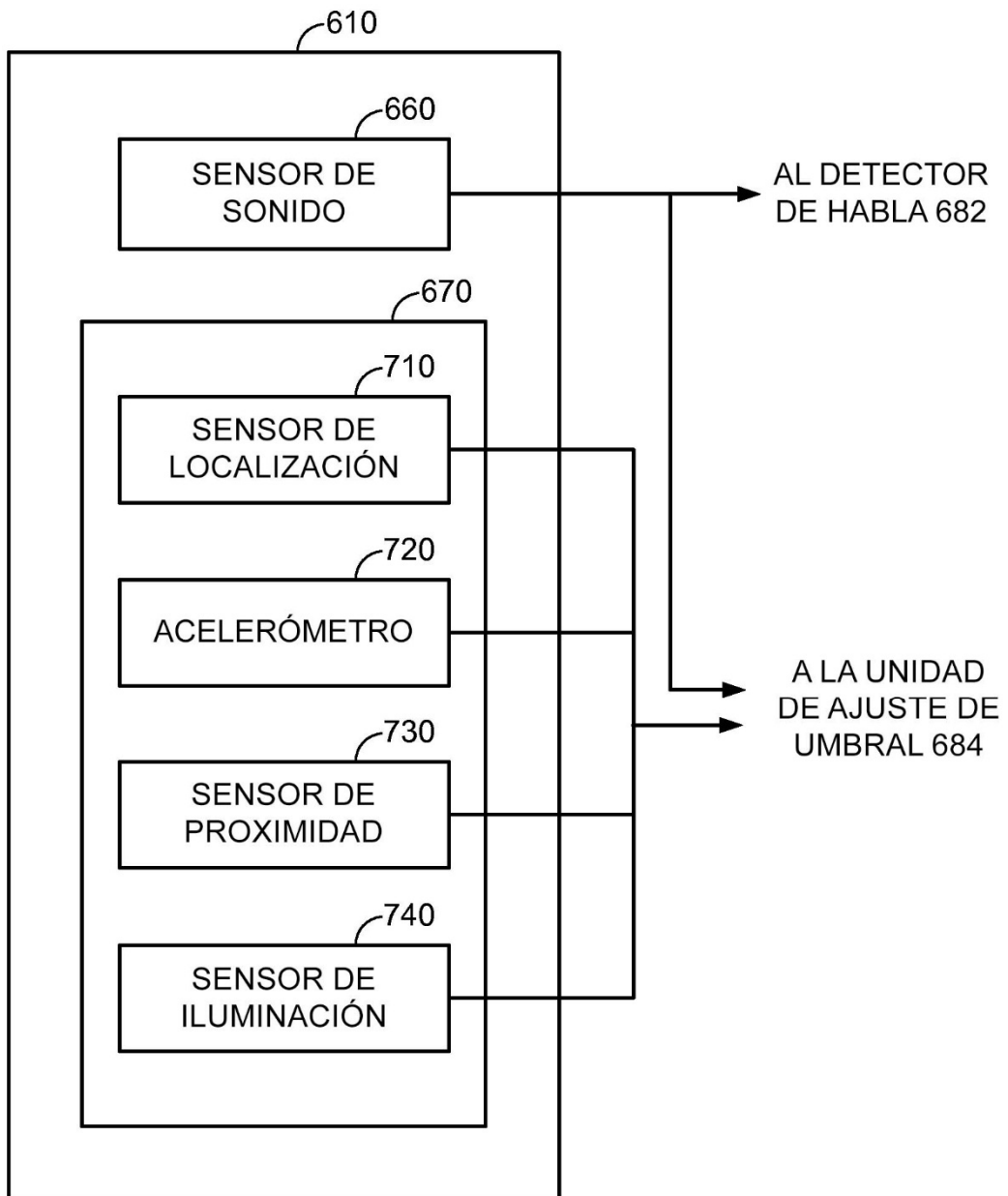
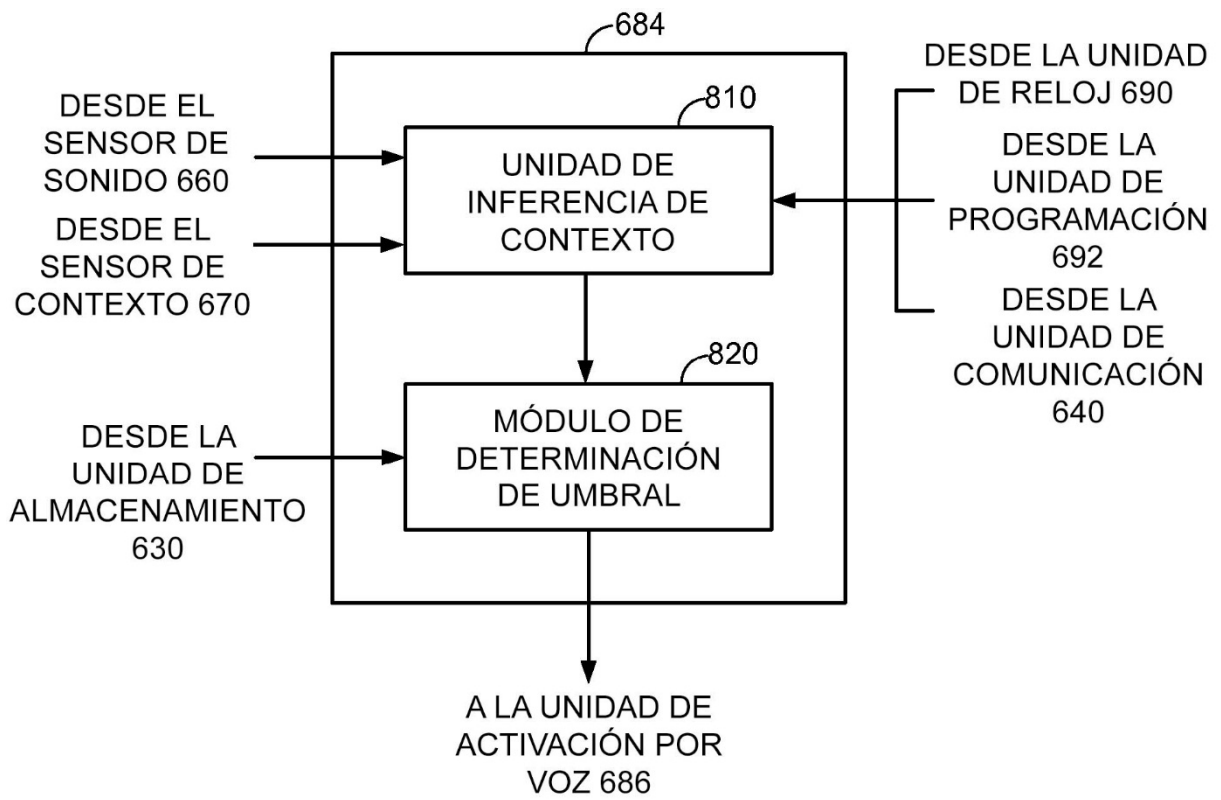


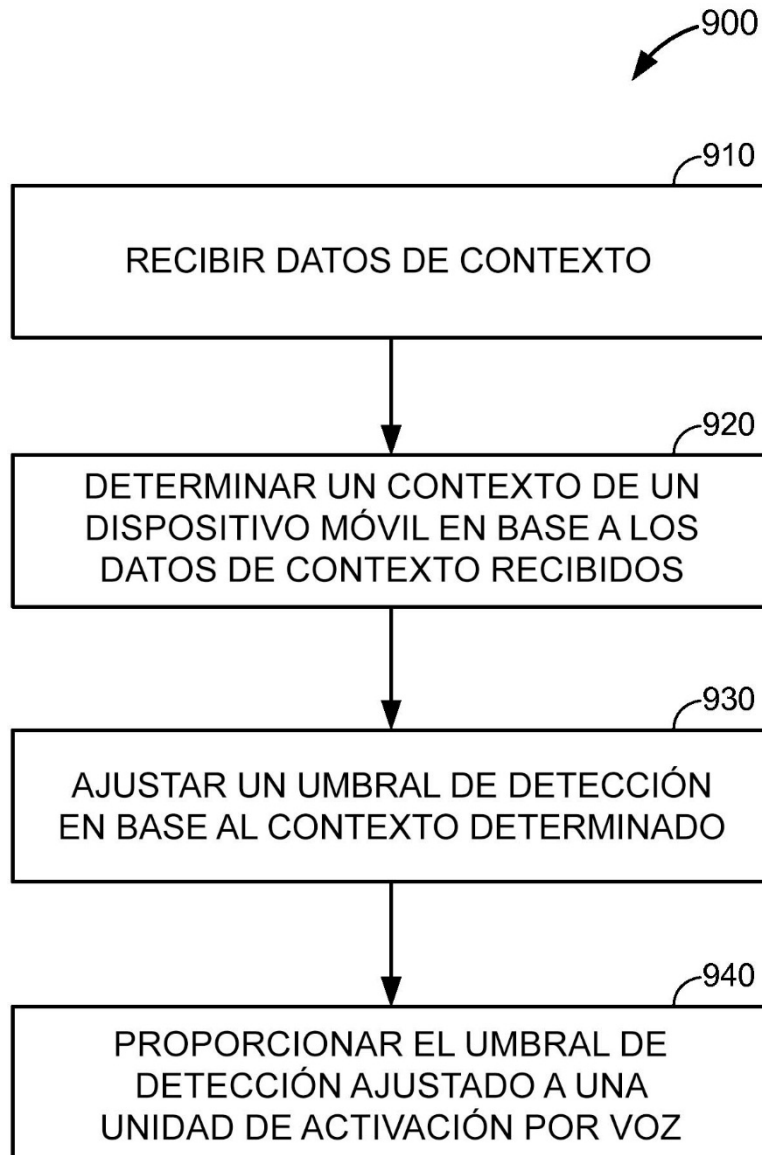
FIG. 6



**FIG. 7**

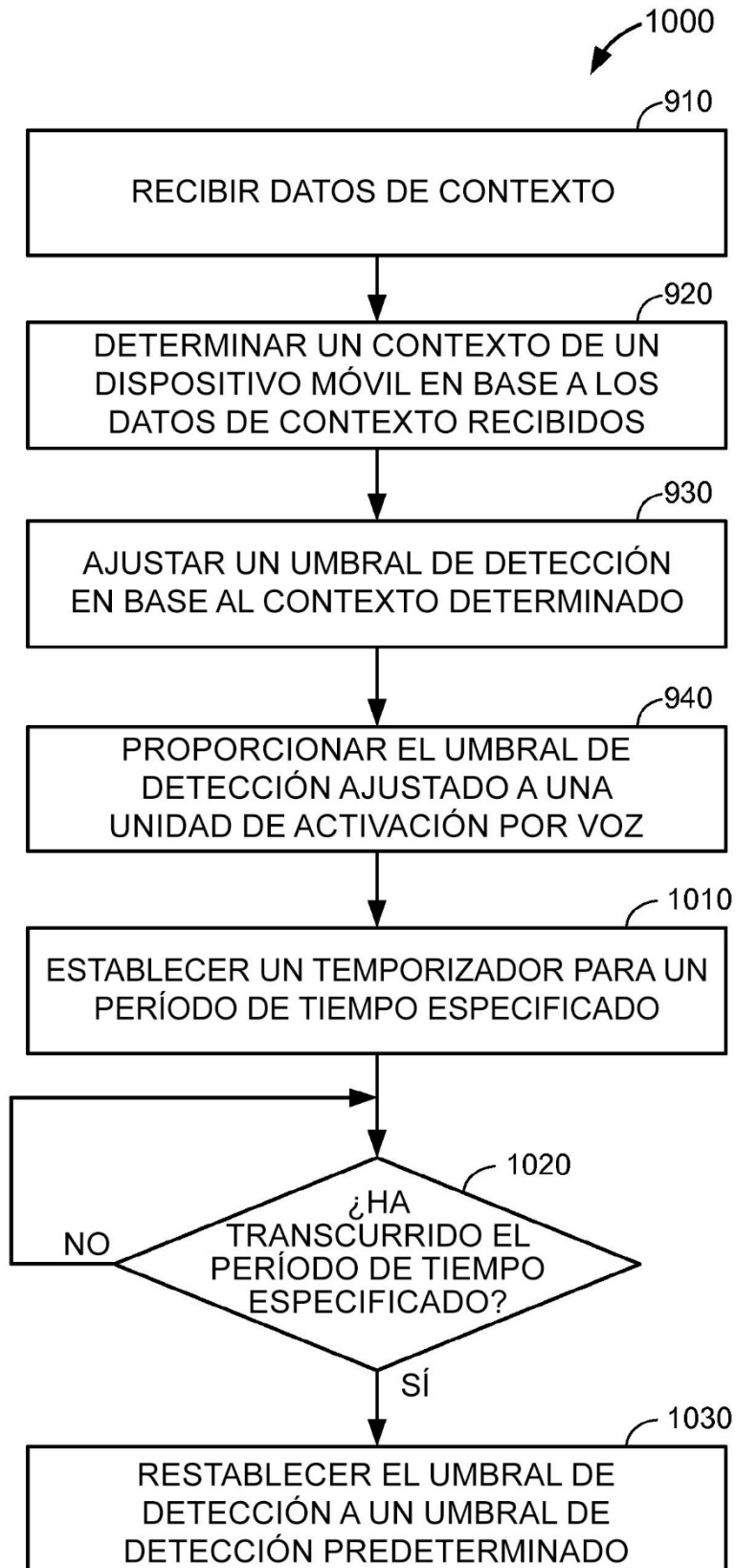


**FIG. 8**

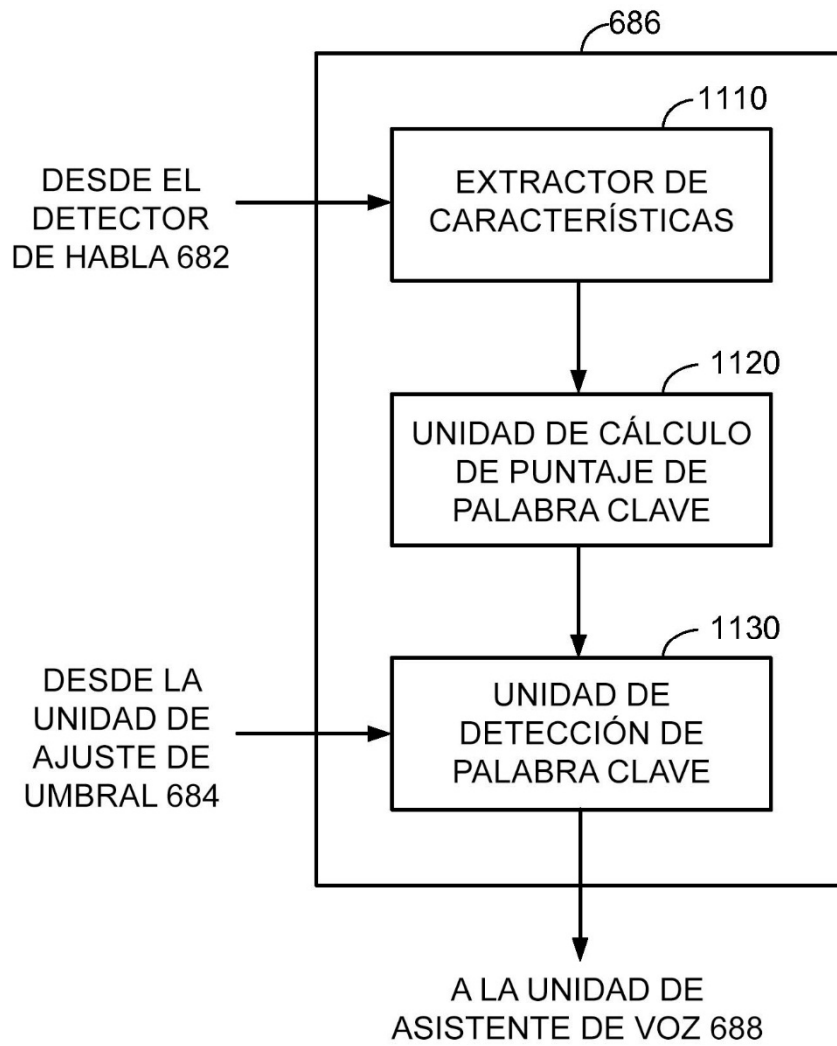


**FIG. 9**

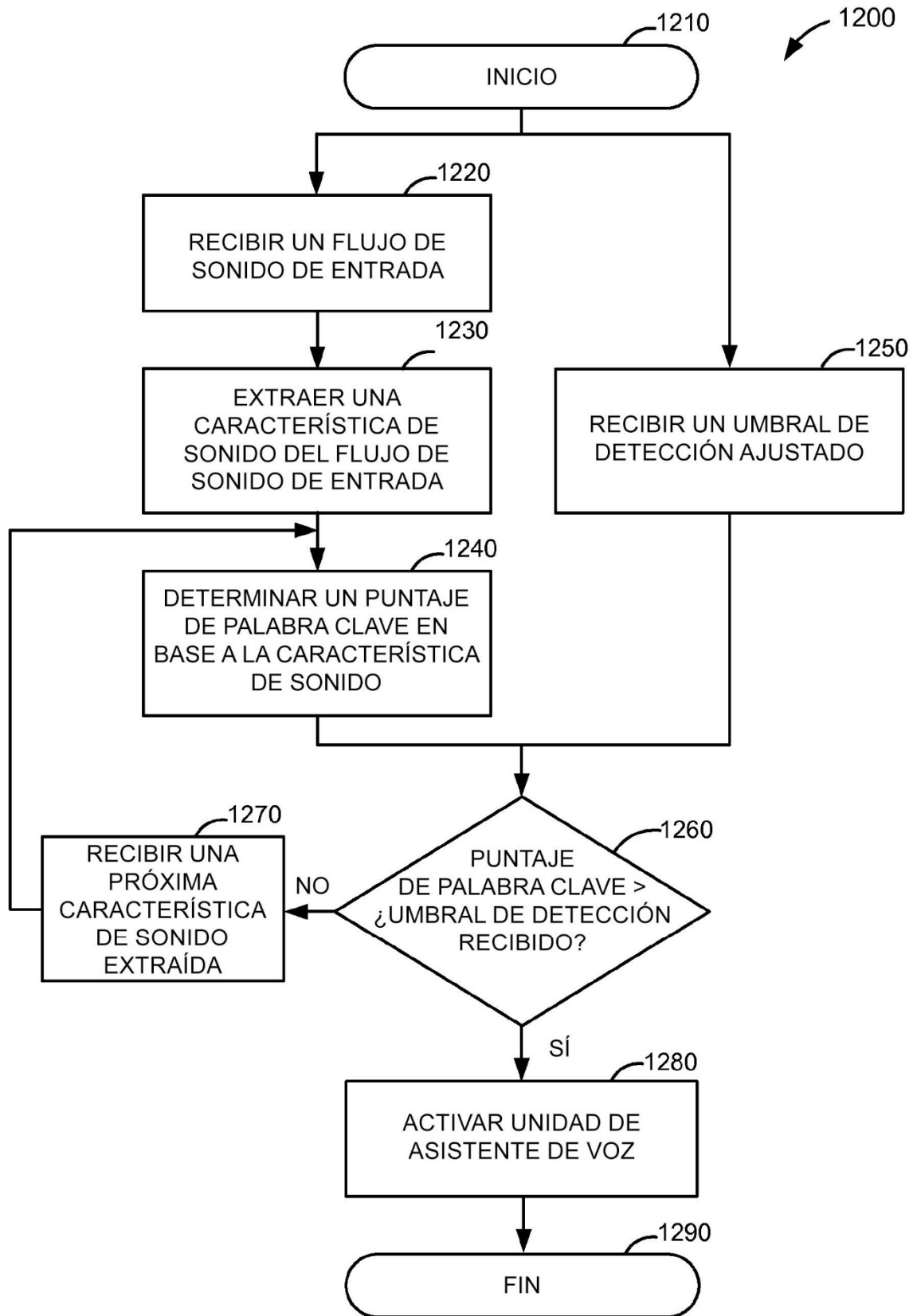




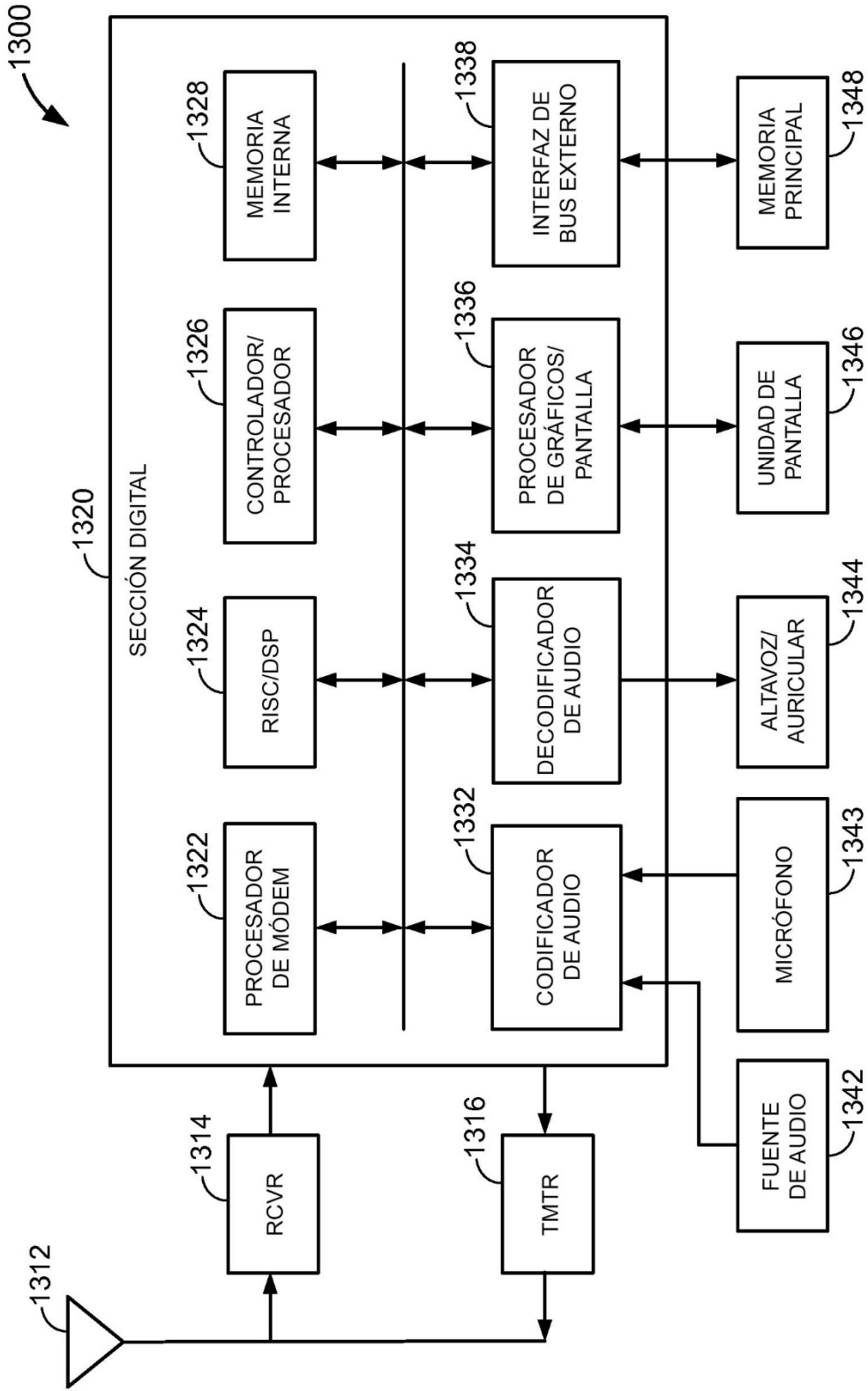
**FIG. 10**



**FIG. 11**



**FIG. 12**



**FIG. 13**