

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 818 085**

51 Int. Cl.:

G10L 15/22 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **02.04.2015 PCT/US2015/024155**

87 Fecha y número de publicación internacional: **29.10.2015 WO15164047**

96 Fecha de presentación y número de la solicitud europea: **02.04.2015 E 15717353 (5)**

97 Fecha y número de publicación de la concesión europea: **17.06.2020 EP 3134896**

54 Título: **Procedimiento y aparato para activación de aplicación mediante entrada de voz**

30 Prioridad:

21.04.2014 US 201414257814

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

09.04.2021

73 Titular/es:

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

72 Inventor/es:

**KIM, TAESU y
LEE, MINSUB**

74 Agente/Representante:

FORTEA LAGUNA, Juan José

ES 2 818 085 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 activación de aplicación mediante entrada de voz

5 **REIVINDICACIÓN DE PRIORIDAD**

[0001] La presente solicitud reivindica la prioridad de la solicitud de Estados Unidos n.º 14/257 814, presentada el 21 de abril de 2014, titulada "PROCEDIMIENTO Y APARATO PARA ACTIVAR LA APLICACIÓN MEDIANTE ENTRADA DE VOZ".

10 **CAMPO TÉCNICO**

[0002] La presente divulgación se refiere en general a la activación de una aplicación en un dispositivo electrónico, y más específicamente, a la activación de una aplicación en un dispositivo electrónico para realizar una función mediante el almacenamiento de la entrada de voz.

15 **ANTECEDENTES**

[0003] Recientemente, el uso de dispositivos electrónicos tales como teléfonos inteligentes, ordenadores tipo tablet, y ordenadores que se pueden llevar encima ha ido en aumento entre los consumidores. Estos dispositivos pueden proporcionar una variedad de capacidades tales como procesamiento y comunicación de datos, comunicación de voz, navegador de Internet, reproductor multimedia, reproductor de juegos, etc. Además, dichos dispositivos electrónicos pueden incluir una variedad de aplicaciones capaces de realizar diversas funciones para los usuarios.

[0004] Para comodidad del usuario, los dispositivos electrónicos convencionales a menudo incluyen una función de reconocimiento de voz para reconocer la voz de los usuarios. En dichos dispositivos electrónicos, un usuario puede pronunciar un comando de voz para realizar una función específica en lugar de navegar manualmente a través de un dispositivo de E/S, como una pantalla táctil o un teclado. A continuación, el comando de voz del usuario puede reconocerse y la función especificada puede realizarse en los dispositivos electrónicos.

[0005] En dispositivos electrónicos convencionales equipados con una función de reconocimiento de voz, un usuario típicamente dice una predeterminada palabra clave para iniciar una aplicación de procesamiento de voz que puede reconocer un comando posterior de voz y realizar una función asociada con el comando de voz. Por ejemplo, el usuario puede pronunciar la palabra clave predeterminada para iniciar la aplicación de procesamiento de voz y esperar a que se cargue la aplicación. Después de cargar la aplicación, el usuario puede pronunciar un comando de voz que especifica una función de la aplicación (por ejemplo, una búsqueda en Internet) a realizar. Dado que el tiempo necesario para reconocer la palabra clave y cargar la aplicación conlleva un retraso en el uso del comando de voz por parte del usuario, usar una palabra clave y un comando de voz conjuntamente para realizar una función en un dispositivo electrónico puede no ser muy conveniente para el usuario.

[0006] En algunos casos, un dispositivo electrónico equipado con una función de reconocimiento de voz puede iniciar una aplicación accidentalmente. Por ejemplo, un usuario puede pronunciar una palabra clave como parte de una conversación o análisis con otra persona sin tener la intención de iniciar una aplicación asociada. Sin embargo, el dispositivo electrónico puede reconocer la palabra clave pronunciada por el usuario e iniciar la aplicación. Por lo tanto, el usuario puede necesitar finalizar manualmente la aplicación, lo cual da como resultado inconvenientes para el usuario. Por otro lado, si el usuario no nota el inicio de la aplicación, la aplicación puede continuar ejecutándose en el dispositivo electrónico.

50 **BREVE EXPLICACIÓN**

[0007] La presente divulgación proporciona procedimientos y aparatos para activar una aplicación objetivo en respuesta a la detección de una palabra clave de activación para realizar una función de la aplicación objetivo que está asociada con un comando de voz. de acuerdo con las características de las reivindicaciones independientes.

55 **DESCRIPCIÓN DETALLADA**

[0008] Ahora se hará referencia en detalle a diversos modos de realización, 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 se puede llevar 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.

[0009] La **FIG. 1** ilustra un dispositivo móvil 120 que activa una aplicación de asistente de voz 130 en respuesta a una palabra clave de activación en un flujo de sonido de entrada, de acuerdo con un modo de realización de la presente divulgación. Inicialmente, el dispositivo móvil 120 puede almacenar una palabra clave de activación para activar una

aplicación tal como la aplicación de asistente de voz 130 en el dispositivo móvil 120. En el modo de realización ilustrado, cuando un usuario 110 pronuncia la palabra clave de activación tal como "INICIAR ASISTENTE DE VOZ" al dispositivo móvil 120, el dispositivo móvil 120 puede capturar un flujo de sonido de entrada y detectar la palabra clave de activación en el flujo de sonido de entrada. Como se usa en el presente documento, el término "flujo de sonido" puede referirse a una secuencia de una o más señales de sonido o datos de sonido, y puede incluir señales o datos analógicos, digitales y acústicos.

[0010] Al detectar la palabra clave de activación, el dispositivo móvil 120 puede almacenar en memoria intermedia ("búfer") una parte del flujo de sonido de entrada en una memoria intermedia del dispositivo móvil 120. Por ejemplo, el usuario 110 puede pronunciar un comando de voz asociado con una función de la aplicación de asistente de voz 130. La aplicación de asistente de voz 130 puede configurarse para realizar cualquier número adecuado de funciones tales como realizar una búsqueda en Internet, establecer alarmas, gestionar horarios y similares, cada una de las cuales puede estar asociada con un comando de voz. En el modo de realización ilustrado, el usuario 110 dice "QUÉ TIEMPO HACE HOY" como un comando de voz, y al menos una parte del comando de voz en el flujo de sonido de entrada puede almacenarse en la memoria intermedia.

[0011] Una vez que se almacena en memoria intermedia la parte del flujo de sonido de entrada que incluye al menos una parte del comando de voz, el dispositivo móvil 120 puede verificar si activar la aplicación de asistente de voz 130 basándose en la parte almacenada en memoria intermedia del flujo de sonido de entrada, como se describirá a continuación con más detalle con referencia a la FIG. 4. Por ejemplo, el dispositivo móvil 120 puede verificar si se activa la aplicación de asistente de voz 130 basándose en de las características de la voz y/o una relación señal/ruido (SNR) de la parte almacenada en memoria intermedia del flujo de sonido de entrada. Si la activación de la aplicación de asistente de voz 130 no se verifica basándose en de la parte almacenada en memoria intermedia en el flujo de sonido de entrada, el dispositivo móvil 120 puede proceder a recibir un siguiente flujo de sonido de entrada para detectar la palabra clave de activación.

[0012] Por otra parte, si se verifica la activación de la aplicación de asistente de voz 130, el dispositivo móvil 120 puede activar la aplicación de asistente de voz 130 para reconocer el comando de voz en la parte almacenada en memoria intermedia del flujo de sonido de entrada. En el modo de realización ilustrado, la aplicación de asistente de voz 130 puede reconocer el comando de voz "QUÉ TIEMPO HACE HOY" de la parte almacenada en memoria intermedia del flujo de sonido de entrada. Al reconocer el comando de voz, la aplicación de asistente de voz 130 puede realizar una función de búsqueda del tiempo a través de Internet y mostrar información del tiempo para el usuario 110. En algunos modos de realización, el dispositivo móvil 120 puede almacenar una pluralidad de palabras clave de activación para su uso en la detección de las palabras clave de activación a partir de un sonido de entrada de un usuario. En este caso, cada una de las palabras clave de activación puede configurarse para activar una aplicación asociada con la palabra clave de activación en el dispositivo móvil 120.

[0013] La FIG. 2 ilustra un diagrama de bloques de un dispositivo electrónico 200 configurado para activar una aplicación de asistente de voz 262 almacenando una parte de un flujo de sonido de entrada que incluye un comando de voz al detectar una palabra clave de activación en el flujo de sonido de entrada, de acuerdo con un modo de realización de la presente divulgación. Como se usa en el presente documento, el término "palabra clave de activación" puede referirse a una o más palabras o sonidos emitidos por un hablante indicativo de la aplicación de asistente de voz 262 para realizar una función en el dispositivo móvil 120, y el término "comando de voz" puede referirse a una o más palabras o sonidos emitidos desde un altavoz indicativos de una función que puede ser realizada por la aplicación de asistente de voz 262. El dispositivo electrónico 200 puede incluir un sensor de sonido 210, una unidad de E/S (entrada/salida) 220, una unidad de comunicación 230, un procesador 240 y una unidad de almacenamiento 260. El dispositivo electrónico 200 puede ser cualquier dispositivo adecuado equipado con capacidades de captura y procesamiento de sonido, como un teléfono celular, un teléfono inteligente (por ejemplo, el dispositivo móvil 120), un ordenador personal, un ordenador portátil, una tablet, un televisor inteligente, un dispositivo de juegos, un reproductor multimedia, gafas inteligentes, un ordenador que se puede llevar encima, etc.

[0014] El procesador 240 puede ser un procesador de aplicaciones (AP), una unidad de procesamiento central (CPU), o una unidad de microprocesador (MPU) para la gestión y el funcionamiento del dispositivo electrónico 200 y puede incluir un procesador de señal digital (DSP) 250. El DSP 250 puede incluir una unidad de activación de voz 252 y una memoria intermedia 254. En un modo de realización, el DSP 250 puede ser un procesador de baja potencia para reducir el consumo de energía en el procesamiento de flujos de sonido. En esta configuración, la unidad de activación de voz 252 en el DSP 250 puede configurarse para activar la aplicación de asistente de voz 262, que se almacena en la unidad de almacenamiento 260, en respuesta a una palabra clave de activación en un flujo de sonido de entrada. Según un modo de realización, la unidad de activación de voz 252 puede activar el procesador 240, que a su vez puede activar la aplicación de asistente de voz 262 (como se indica con una flecha en la FIG. 2) cargando e iniciando la aplicación de asistente de voz 262. Aunque la unidad de activación de voz 252 está configurada para activar la aplicación de asistente de voz 262 en el modo de realización ilustrado, también puede activar cualquier función o aplicación del dispositivo electrónico 200 que pueda estar asociada con una palabra clave de activación.

[0015] El sensor de sonido 210 se puede configurar para recibir un flujo de sonido de entrada y proporcionar el flujo de sonido de entrada recibido al DSP 250. El sensor de sonido 210 puede incluir uno o más micrófonos u otros tipos

de sensores de sonido que se puedan usar para recibir, captar, detectar y/o percibir sonido. Además, el sensor de sonido 210 puede emplear cualquier software y/o hardware adecuado para realizar dichas funciones.

5 **[0016]** Para reducir el consumo de energía, el sensor de sonido 210 se puede configurar para recibir el flujo de sonido de entrada periódicamente de acuerdo con un ciclo de trabajo. Por ejemplo, el sensor de sonido 210 puede funcionar en un ciclo de trabajo del 10 % de modo que el flujo de sonido de entrada se reciba el 10 % del tiempo (por ejemplo, 20 ms en un período de 200 ms). En este caso, el sensor de sonido 210 puede detectar sonido determinando si la parte recibida del flujo de sonido de entrada excede un umbral de intensidad de sonido predeterminado. Por ejemplo, se puede determinar una intensidad de sonido de la parte recibida del flujo de sonido de entrada y compararla con la intensidad de sonido umbral predeterminada. Si la intensidad de sonido de la parte recibida excede el umbral de intensidad de sonido, el sensor de sonido 210 puede deshabilitar la función de ciclo de trabajo para continuar recibiendo una parte restante del flujo de sonido de entrada. Además, el sensor de sonido 210 puede activar el DSP 250 y proporcionar la parte recibida del flujo de sonido de entrada, incluida la parte restante al DSP 250.

15 **[0017]** Cuando el sensor de sonido 210 activa el DSP 250, la unidad de activación de voz 252 puede configurarse para recibir continuamente el flujo de sonido de entrada desde el sensor de sonido 210 y detectar una palabra clave de activación (por ejemplo, "INICIAR ASISTENTE DE VOZ") en el flujo de sonido de entrada recibido para activar la aplicación de asistente de voz 262. Para detectar la palabra clave de activación, la unidad de activación de voz 252 puede emplear cualquier procedimiento de detección de palabras clave adecuado basado en un modelo de cadena de Markov tal como un modelo de Markov oculto (HMM), un modelo de semi-Markov (SMM), o una combinación de los mismos.

25 **[0018]** En un modo de realización, la unidad de almacenamiento 260 puede almacenar una pluralidad de palabras para una o más palabras clave de activación y comandos de voz que pueden ser utilizados para activar la aplicación de asistente de voz 262 y realizar una función de la aplicación de asistente de voz 262. Además, la unidad de almacenamiento 260 puede almacenar información de estado en una pluralidad de estados asociados con una pluralidad de partes de las palabras. Según un modo de realización, cada una de las palabras para las palabras clave de activación y los comandos de voz se puede dividir en una pluralidad de unidades básicas de sonido, como fonos, fonemas o subunidades de las mismas, y se puede generar una pluralidad de partes de cada una de las palabras basándose en las unidades básicas de sonido. A continuación, cada parte de cada una de las palabras puede asociarse con un estado bajo un modelo de cadena de Markov, como un HMM, un SMM o una combinación de los mismos.

35 **[0019]** La unidad de almacenamiento 260 también puede almacenar la aplicación de asistente de voz 262 a la cual se puede acceder mediante el procesador 240. Aunque la unidad de almacenamiento 260 almacena la aplicación de asistente de voz 262 en el modo de realización ilustrado, también puede almacenar cualquier otra aplicación o función del dispositivo electrónico 200 que pueda ser ejecutada por el procesador 240. La unidad de almacenamiento 260 puede implementarse utilizando cualquier dispositivo de almacenamiento o memoria adecuado, tal como una RAM (memoria de acceso aleatorio), una ROM (memoria de solo lectura), una EEPROM (memoria de solo lectura programable y borrrable eléctricamente), una memoria flash o una SSD (unidad de estado sólido).

45 **[0020]** Al detectar la palabra clave de activación, la unidad de activación de voz 252 puede configurarse para almacenar en memoria intermedia (o almacenar temporalmente) una parte del flujo de sonido de entrada recibido desde el sensor de sonido 210 en la memoria intermedia 254 del DSP 250. La memoria intermedia 254 puede implementarse usando cualquier esquema de almacenamiento o memoria adecuado en un procesador tal como una memoria local o una memoria caché. Aunque el DSP 250 incluye la memoria intermedia 254 en el modo de realización ilustrado, la memoria intermedia 254 puede implementarse como un área de memoria en la unidad de almacenamiento 260. En algunos modos de realización, la memoria intermedia 254 puede implementarse usando una pluralidad de áreas de memoria física o una pluralidad de áreas de memoria lógica.

50 **[0021]** Como se recibe la parte del flujo de sonido de entrada desde el sensor de sonido 210, la unidad de activación de voz 252 puede almacenar en memoria intermedia al menos una parte del comando de voz en la memoria intermedia 254. Por ejemplo, la parte almacenada en memoria intermedia puede incluir todo o la mayor parte del comando de voz indicativo de una función de la aplicación de asistente de voz 262 (por ejemplo, "¿QUÉ TIEMPO HACE HOY?"). En algunos modos de realización, la parte recibida del flujo de sonido de entrada puede continuar almacenándose en la memoria intermedia 254 hasta que se detecte un final del comando de voz.

60 **[0022]** Accediendo a la parte almacenada en memoria intermedia del flujo de sonido de entrada en la memoria intermedia 254, la unidad de activación de voz 252 puede verificar si se activa la aplicación de asistente de voz 262 en respuesta a la detección de la palabra clave de activación. Por ejemplo, la aplicación de asistente de voz 262 puede verificar si se activa la aplicación de asistente de voz 262 basándose en de las características de sonido de la parte almacenada en memoria intermedia del flujo de sonido de entrada. Si se verifica la activación de la aplicación de asistente de voz 262, la unidad de activación de voz 252 puede activar el procesador 240, que a su vez puede activar la aplicación de asistente de voz 262 accediendo a la unidad de almacenamiento 260.

65

[0023] Una vez que la aplicación de asistente de voz 262 se activa mediante la unidad de activación de voz 252, la aplicación de asistente de voz 262 puede realizar una función de asistente de voz mediante el reconocimiento del comando de voz usando cualquiera de los procedimientos de reconocimiento de voz adecuados basándose en un HMM, un SMM, o similares. En un modo de realización, la aplicación de asistente de voz 262 puede reconocer el comando de voz desde la parte almacenada en memoria intermedia del flujo de sonido de entrada que se almacena en la memoria intermedia 254. En algunos modos de realización, la parte almacenada en memoria intermedia del flujo de sonido de entrada puede copiarse desde la memoria intermedia 254 a la unidad de almacenamiento 260 o una memoria local (no mostrada) en el procesador 240, a la que puede acceder la aplicación de asistente de voz activada 262 para reconocer el comando de voz.

[0024] Al reconocer el comando de voz a partir de la parte almacenada en memoria intermedia del flujo de sonido de entrada, la aplicación de asistente de voz 262 puede realizar una función asociada con el comando de voz reconocido. Por ejemplo, cuando se recibe un comando de voz para una búsqueda en Internet, la aplicación de asistente de voz 262 puede realizar una búsqueda en la web a través de la unidad de comunicación 230 a través de la red 270. En este caso, los resultados de búsqueda para el comando de voz pueden aparecer en una pantalla de visualización de la unidad de E/S 220.

[0025] La **FIG. 3** ilustra un diagrama de bloques detallado de la memoria intermedia 254 que incluye una pluralidad de unidades de memoria intermedia 310 y 320 para almacenar en memoria intermedia un flujo de sonido de entrada, de acuerdo con un modo de realización de la presente divulgación. Las unidades de memoria intermedia 310 y 320 pueden implementarse como bloques de memoria física o lógica en la memoria intermedia 254. Aunque la memoria intermedia 254 incluye las dos unidades de memoria intermedia 310 y 320 en el modo de realización ilustrado, la memoria intermedia 254 puede incluir cualquier número adecuado de unidades de memoria intermedia.

[0026] Si las unidades de memoria intermedia 310 y 320 se implementan con bloques de memoria lógicos, las unidades de memoria intermedia 310 y 320 pueden compartir un rango de direcciones de memoria predeterminada de una memoria física en la que cualquier información o datos adecuados (por ejemplo, una parte de un flujo de sonido de entrada recibido) para las unidades de memoria intermedia 310 y 320 pueden almacenarse en memoria intermedia (o almacenarse). En este caso, se puede asignar una pluralidad de punteros para referirse a una pluralidad de direcciones o ubicaciones de datos, respectivamente, en el rango de direcciones de memoria. Se puede asignar un puntero para referirse a la información o los datos almacenados en memoria intermedia para la unidad de memoria intermedia 310, mientras que se puede asignar un puntero diferente para referirse a la información o los datos almacenados en memoria intermedia para la unidad de memoria intermedia 320. De forma alternativa, se puede asignar una misma dirección de memoria para referirse a diferentes datos almacenados en las unidades de memoria intermedia 310 y 320 y se puede realizar una operación de gestión de datos predeterminada para almacenar, mantener, acceder y/o recuperar los datos para cada una de las unidades de memoria intermedia 310 y 320.

[0027] Dado que se recibe un flujo de sonido de entrada, el sensor de sonido 210 puede proporcionar el flujo de sonido de entrada al DSP 250, que puede comenzar a almacenar en memoria intermedia el flujo de sonido de entrada en la unidad de memoria intermedia 310. La unidad de memoria intermedia 310 puede configurarse para almacenar en memoria intermedia (o almacenar temporalmente) el flujo de sonido de entrada que incluye al menos una parte de una palabra clave de activación para usar en la detección de la palabra clave de activación. Desde la parte almacenada en memoria intermedia del flujo de sonido de entrada en la unidad de memoria intermedia 310, la unidad de activación de voz 252 puede detectar la palabra clave de activación. Al detectar la palabra clave de activación, el DSP 250 puede dejar de almacenar el flujo de sonido de entrada en la unidad de memoria intermedia 310.

[0028] Cuando se detecta la palabra clave de activación, el DSP 250 puede continuar recibiendo el flujo de sonido de entrada desde el sensor de sonido 210 y comenzar a almacenar en memoria intermedia el flujo de sonido de entrada en la unidad de memoria intermedia 320. En un modo de realización, el DSP 250 puede recibir y almacenar en memoria intermedia (o almacenar temporalmente) una parte del flujo de sonido de entrada en la unidad de memoria intermedia 320 que puede incluir al menos una parte de un comando de voz. A medida que el flujo de sonido de entrada está almacenado en la unidad de memoria intermedia 320, la unidad de activación de voz 252 puede acceder a la parte almacenada en memoria intermedia del flujo de sonido de entrada en la unidad de memoria intermedia 320 y verificar si se activará la aplicación de asistente de voz 262 basándose en de la parte almacenada en memoria intermedia.

[0029] La **FIG. 4** ilustra un diagrama de bloques detallado de la unidad de activación de voz 252 del dispositivo electrónico 200 que está configurado para verificar si se activa la aplicación de asistente de voz 262 basándose en una parte almacenada en memoria intermedia de un flujo de sonido de entrada, de acuerdo con un modo de realización de la presente divulgación. La unidad de activación de voz 252 puede incluir una unidad de detección de palabras clave 410 y una unidad de verificación 420. Como se muestra, la unidad de activación de voz 252 puede recibir un flujo de sonido de entrada desde el sensor de sonido 210 para detectar una palabra clave de activación y verificar si se activa la aplicación de asistente de voz 262. La memoria intermedia 254 puede almacenar en memoria intermedia al menos una parte del flujo de sonido de entrada para usar en la detección de la palabra clave de activación y verificar si se activa la aplicación de asistente de voz 262. En un modo de realización, la memoria intermedia 254 puede incluir las unidades de memoria intermedia 310 y 320 para almacenar en memoria intermedia al menos una parte de la

palabra clave de activación y al menos una parte de un comando de voz, respectivamente, como se describe anteriormente con referencia a la **FIG. 3**.

5 **[0030]** Dado que se recibe el flujo de sonido de entrada y se almacena en memoria intermedia en la memoria intermedia 254, la unidad de detección de palabras clave 410 en la unidad de activación de voz 252 puede acceder a la memoria intermedia 254 y detectar una palabra clave de activación desde la parte almacenada en memoria intermedia del flujo de sonido de entrada utilizando cualquier procedimiento de detección de palabras clave adecuado basado en un HMM, un SMM o similar. En algunos modos de realización, a medida que se recibe y almacena el flujo de sonido de entrada, la unidad de detección de palabras clave 410 puede determinar una pluralidad de puntuaciones de palabras clave para la parte almacenada en memoria intermedia del flujo de sonido de entrada usando cualquier modelo de probabilidad adecuado tal como un modelo de mezcla Gaussiana (GMM), una red neuronal, una máquina de vectores de soporte (SVM) y similares. La unidad de detección de palabras clave 410 puede comparar cada una de las puntuaciones de palabras clave con una puntuación de umbral de detección predeterminada y cuando una de las puntuaciones de palabras clave excede la puntuación de umbral de detección, la palabra clave de activación puede detectarse desde la parte almacenada en memoria intermedia del flujo de sonido de entrada.

20 **[0031]** Al detectar la palabra clave de activación, el DSP 250 puede continuar recibiendo el flujo de sonido de entrada del sensor de sonido 210 y empezar a almacenar en memoria intermedia el flujo de sonido de entrada recibido en la memoria intermedia 254. En este caso, la parte del flujo de sonido de entrada que se almacena en la memoria intermedia 254 puede incluir al menos una parte del comando de voz de un usuario del dispositivo electrónico 200. En algunos modos de realización, el DSP 250 puede detectar un final del comando de voz desde la parte almacenada en la memoria intermedia del comando de voz y terminar almacenando en memoria intermedia el flujo de sonido de entrada en la memoria intermedia 254. El final del comando de voz puede detectarse utilizando cualquier procedimiento de detección de actividad de voz (VAD) adecuado. De acuerdo con un modo de realización, el final del comando de voz puede detectarse analizando una intensidad de sonido de la parte almacenada en memoria intermedia del flujo de sonido de entrada. Por ejemplo, cuando la intensidad del sonido disminuye por debajo de un valor de intensidad predeterminado, se puede detectar el final del comando de voz. De forma adicional o alternativa, el final del comando de voz puede detectarse cuando se detecta una disminución rápida en la intensidad del sonido (por ejemplo, un borde en el cambio de la intensidad del sonido). Al detectar el final del comando de voz, el DSP 250 puede ordenar a la unidad de verificación 420 en la unidad de activación de voz 252 que verifique si se activa la aplicación de asistente de voz 262.

35 **[0032]** Cuando se detecta el final del comando de voz, la unidad de verificación 420 puede acceder a la parte del flujo de sonido de entrada que incluye al menos una parte del comando de voz almacenado en la memoria intermedia 254 y verificar si se activa la aplicación de asistente de voz 262 basándose en la parte almacenada en memoria intermedia. En un modo de realización, la unidad de verificación 420 puede determinar si la parte almacenada en memoria intermedia del flujo de sonido de entrada incluye voz usando cualquier procedimiento de clasificación de sonido adecuado tal como un clasificador basado en GMM, una red neuronal, un HMM, un modelo gráfico, una técnica SVM, y similares. Si se determina que la parte almacenada en memoria intermedia del flujo de sonido de entrada incluye voz, la unidad de verificación 420 puede verificar la activación de la aplicación de asistente de voz 262.

45 **[0033]** En un modo de realización, la unidad de verificación 420 puede determinar además si la voz en la parte almacenada en memoria intermedia del flujo de sonido de entrada es indicativa de un altavoz de la palabra clave de activación. En este caso, las características de sonido tales como las características de sonido y/o las huellas digitales de audio pueden extraerse de la palabra clave de activación y la voz en las partes almacenadas en memoria intermedia en el flujo de sonido de entrada. A continuación, las características de sonido de la palabra clave de activación pueden compararse con las características de sonido del comando de voz, y la activación de la aplicación de asistente de voz 262 puede verificarse si se determina que las características de sonido son iguales o similares.

50 **[0034]** En otro modo de realización, la unidad de verificación 420 puede determinar una relación de señal a ruido (SNR) de la parte almacenada en memoria intermedia del flujo de sonido de entrada que incluye la parte del comando de voz. En este caso, cuando se determina que la SNR de la parte almacenada en memoria intermedia es mayor que una SNR de umbral predeterminado, la unidad de verificación 420 puede verificar la activación de la aplicación de asistente de voz 262. En los modos de realización anteriores, la unidad de verificación 420 puede acceder a la memoria intermedia 254 para verificar si se activa la aplicación de asistente de voz 262 antes o después de que se detecte el final del comando de voz.

60 **[0035]** De acuerdo con algunos modos de realización, el comando de voz puede incluir una o más palabras predeterminadas indicativas de revocación de la palabra clave de activación (por ejemplo, "CANCELAR", "NO IMPORTA" y "OLVÍDALO"). Por ejemplo, una persona o un usuario cerca del dispositivo electrónico 200 puede pronunciar accidental o involuntariamente una palabra clave de activación (por ejemplo, "INICIAR ASISTENTE DE VOZ") y la persona o el usuario cerca del dispositivo electrónico 200 puede revocar la palabra clave de activación al decir de manera predeterminada palabras para evitar la activación de la aplicación de asistente de voz 262. La una o más palabras de revocación también pueden denominarse comando de revocación. A medida que un flujo de sonido de entrada que incluye una o más palabras de revocación es recibido y almacenado en la memoria intermedia 254, la unidad de verificación 420 puede acceder al flujo de sonido de entrada almacenado incluyendo las palabras de

revocación y determinar si el flujo de sonido de entrada almacenado incluye tales palabras predeterminadas. Si se determina que el flujo de sonido de entrada almacenado incluye alguna de las palabras predeterminadas indicativas de revocar la palabra clave de activación, el dispositivo electrónico 200 puede proceder a recibir un siguiente flujo de sonido de entrada para detectar la palabra clave de activación sin activar la aplicación de asistente de voz 262. Por otro lado, si se determina que el flujo de sonido de entrada almacenado en memoria intermedia no incluye tales palabras de revocación, se puede inferir que el hablante o el usuario pronunció intencionalmente la palabra clave de activación y la unidad de verificación 420 puede verificar la activación de la aplicación de asistente de voz 262.

[0036] La FIG. 5 ilustra un diagrama de flujo de un procedimiento 500 para activar la aplicación de asistente de voz 262 almacenando en memoria intermedia una parte de un flujo de sonido de entrada que incluye un comando de voz al detectar una palabra clave de activación en el flujo de sonido de entrada, de acuerdo con un modo de realización de la presente divulgación. El dispositivo electrónico 200 puede recibir el flujo de sonido de entrada que incluye la palabra clave de activación para activar la aplicación de asistente de voz 262 y el comando de voz indicativo de una función de la aplicación de asistente de voz 262, en 510. Por ejemplo, la palabra clave de activación puede ser "INICIAR ASISTENTE DE VOZ" y el comando de voz puede ser "¿QUÉ TIEMPO HACE HOY?".

[0037] En un modo de realización, una parte del flujo de sonido de entrada recibido incluyendo al menos una parte de la palabra clave de activación se puede almacenar en memoria intermedia en la unidad de memoria intermedia 310 de la memoria intermedia 254. La unidad de activación de voz 252 en el DSP 250 del dispositivo electrónico 200 puede detectar la palabra clave de activación desde la parte del flujo de sonido de entrada almacenado en la unidad de memoria intermedia 310, en 520. Si se detecta la palabra clave de activación, el dispositivo electrónico 200 puede almacenar en memoria intermedia una parte del flujo de sonido de entrada en la unidad de memoria intermedia 320 de la memoria intermedia 254, en 530. La parte del flujo de sonido de entrada almacenada en la unidad de memoria intermedia 320 puede incluir al menos una parte del comando de voz.

[0038] El dispositivo electrónico 200 puede activar la aplicación de asistente de voz 262 que está asociada con la palabra clave de activación, a 540. De acuerdo con un modo de realización, la unidad de activación de voz 252 puede verificar si se activa la aplicación de asistente de voz 262 basándose en la parte del flujo de sonido de entrada almacenado en memoria intermedia en la unidad de memoria intermedia 320. En este modo de realización, la unidad de activación de voz 252 puede activar la aplicación de asistente de voz 262 cuando se verifica la activación de la aplicación de asistente de voz 262. Una vez que se activa la aplicación de asistente de voz 262, la aplicación de asistente de voz 262 puede acceder a la parte del flujo de sonido de entrada almacenada en memoria intermedia en la unidad de memoria intermedia 320 para reconocer el comando de voz desde la parte almacenada en memoria intermedia y realizar la función de la aplicación de asistente de voz 262 asociada con el comando de voz reconocido.

[0039] La FIG. 6 es un diagrama de temporización 600 de almacenamiento en memoria intermedia ("buffering") de una parte 642, 644 o 646 de un flujo de sonido de entrada 610 que incluye un comando de voz 630 cuando se detecta una palabra clave de activación 620 en el flujo de sonido de entrada 610, de acuerdo con un modo de realización de la presente divulgación. Como se muestra, el dispositivo electrónico 200 puede recibir el flujo de sonido de entrada 610 que incluye la palabra clave de activación 620 y el comando de voz 630. En el diagrama de temporización 600, la palabra clave de activación 620 en el flujo de sonido de entrada 610 puede recibirse desde el momento T_1 al momento T_2 . Así, el momento T_1 y el momento T_2 pueden corresponder a un principio y un fin, respectivamente, de la palabra clave de activación 620 en el flujo de sonido de entrada 610. Posteriormente, el comando de voz 630 en el flujo de sonido de entrada 610 puede recibirse desde el momento T_3 hasta el momento T_4 . En este caso, el momento T_3 y el momento T_4 pueden corresponder a un comienzo y un final, respectivamente, del comando de voz 630 en el flujo de sonido de entrada 610.

[0040] El flujo de sonido de entrada 610 puede ser recibido por el sensor de sonido 210 en el dispositivo electrónico 200 y proporcionado al DSP 250 para la detección de la palabra clave de activación 620 y la verificación de si activar la aplicación de asistente de voz 262. En un modo de realización, el sensor de sonido 210 del dispositivo electrónico 200 puede configurarse para determinar si el flujo de sonido de entrada 610 incluye sonido que excede una intensidad de sonido predeterminada basada en un ciclo de trabajo. En este caso, si se determina que el flujo de sonido de entrada recibido 610 incluye un sonido que excede la intensidad de sonido predeterminada, la función de ciclo de trabajo del sensor de sonido 210 puede deshabilitarse (es decir, 100 % de ciclo de trabajo) para que el sensor de sonido 210 pueda recibir continuamente y proporcionar el flujo de sonido de entrada 610 al DSP 250.

[0041] A medida que el flujo de sonido de entrada 610 se recibe desde el sensor de sonido 210, el DSP 250 puede almacenar en memoria intermedia el flujo de sonido de entrada recibido 610 en la unidad de memoria intermedia 310 de la memoria intermedia 254 y detectar la palabra clave de activación 620 del flujo de sonido de entrada 610 que está almacenado en la unidad de memoria intermedia 310. Como se describió anteriormente con referencia a la FIG. 4, el DSP 250 puede detectar la palabra clave de activación 620 determinando una pluralidad de puntuaciones de palabras clave para la parte almacenada en memoria intermedia del flujo de sonido de entrada 610 en la unidad de memoria intermedia 310 y comparando las puntuaciones de palabras clave con una puntuación de umbral de detección. En el diagrama de temporización 600, la palabra clave de activación 620 puede detectarse desde la parte almacenada en memoria intermedia del flujo de sonido de entrada 610 en el momento T_D antes de recibir la palabra clave de activación 620 completa en el flujo de sonido de entrada 610. En otros modos de realización, la palabra clave de activación 620

se puede detectar en el momento T_2 al recibir toda la palabra clave de activación 620 en el flujo de sonido de entrada 610.

[0042] Cuando se detecta la palabra clave de activación 620, el DSP 250 puede comenzar a almacenar en memoria intermedia el flujo de sonido de entrada 610 recibido desde el sensor de sonido 210 en la unidad de memoria intermedia 320 de la memoria intermedia 254. En un modo de realización, el DSP 250 puede comenzar a almacenar en memoria intermedia el flujo de sonido de entrada 610 en la unidad de memoria intermedia 320 al detectar la palabra clave de activación en el momento T_D . El DSP 250 puede continuar almacenando el flujo de sonido de entrada 610 en la unidad de memoria intermedia 320 hasta que se detecte el final del comando de voz 630 en el momento T_4 . Como resultado, la parte 642 del flujo de sonido de entrada 610 puede almacenarse en la unidad de memoria intermedia 320 durante un período desde el momento T_D hasta el momento T_4 .

[0043] En otro modo de realización, el DSP 250 puede empezar a almacenar en memoria intermedia el flujo de sonido de entrada 610 en la unidad de memoria intermedia 320 al detectar el final de la palabra clave de activación de voz en el momento T_2 . Por ejemplo, el DSP 250 puede detectar el final de la palabra clave de activación 620 determinando una pluralidad de puntuaciones de palabras clave para la parte almacenada en memoria intermedia del flujo de sonido de entrada 610 en la unidad de memoria intermedia 310 y comparando las puntuaciones de palabras clave con una puntuación de umbral final, que se describirá a continuación en detalle con referencia a la **FIG. 7**. En algunos modos de realización, la unidad de activación de voz 252 en el DSP 250 puede configurarse para detectar el final de la palabra clave de activación 620. El flujo de sonido de entrada 610 puede almacenarse en memoria intermedia hasta que se detecte el final del comando de voz 630 en el momento T_4 . En consecuencia, la parte 644 del flujo de sonido de entrada 610 puede almacenarse en memoria intermedia en la unidad de memoria intermedia 320 durante un período desde el momento T_2 hasta el momento T_4 .

[0044] De acuerdo con algunos modos de realización, el DSP 250 puede empezar a almacenar en memoria intermedia el flujo de sonido de entrada 610 en la unidad de memoria intermedia 320 al detectar el inicio del comando de voz 630 en el momento T_3 . En este caso, después de detectar el final de la palabra clave de activación 620 en el momento T_2 , el DSP 250 puede detectar el comienzo del comando de voz 630 utilizando cualquier procedimiento VAD adecuado. Por ejemplo, después de detectar el final de la palabra clave de activación 620 en el momento T_2 , el DSP 250 puede almacenar en memoria intermedia el flujo de sonido de entrada 610 que se recibe después del momento T_2 en la unidad de memoria intermedia 310. Dado que el flujo de sonido de entrada 610 se está almacenando en memoria intermedia después del momento T_2 , el DSP 250 puede determinar periódicamente o aperiódicamente una intensidad de sonido del flujo de sonido de entrada almacenado en memoria intermedia 610 y comparar la intensidad de sonido con un umbral predeterminado de intensidad de sonido. Si la intensidad del sonido excede el umbral de la intensidad del sonido, el DSP 250 puede detectar el comienzo del comando de voz 630. De forma adicional o alternativa, el final del comando de voz 630 puede detectarse cuando se detecta un aumento rápido en la intensidad del sonido (por ejemplo, un borde en el cambio de la intensidad del sonido). Por ejemplo, dicho aumento puede detectarse cuando una intensidad de sonido actual es mayor que una intensidad de sonido anterior en más de una intensidad de sonido predeterminada. En algunos modos de realización, la unidad de activación de voz 252 en el DSP 250 puede configurarse para detectar el comienzo del comando de voz 630. En el diagrama de temporización ilustrado 600, cuando se detecta el comienzo del comando de voz 630 en el momento T_3 , el DSP 250 puede almacenar en memoria intermedia la parte 646 del flujo de sonido de entrada 610 en la unidad de memoria intermedia 320 durante un período desde el momento T_3 hasta el momento T_4 .

[0045] Aunque el diagrama de temporización 600 ilustra el almacenamiento en memoria intermedia de la parte 642, 644, o 646 del flujo de sonido de entrada 610 en la unidad de memoria intermedia 320 hasta que se detecta el final del comando de voz 630, el DSP 250 puede terminar el almacenamiento en memoria intermedia del flujo de sonido de entrada 610 en la unidad de memoria intermedia 320 antes de que se detecte el final del comando de voz 630, como se describirá a continuación con más detalle con referencia a la **FIG. 8**. Por consiguiente, la parte 642, 644 o 646 del flujo de sonido de entrada 610 almacenado en la unidad de memoria intermedia 320 puede incluir al menos una parte del comando de voz 630. Una vez que la parte 642, 644 o 646 del flujo de sonido de entrada 610 está almacenada en la unidad de memoria intermedia 320, la unidad de activación de voz 252 puede verificar si se activa la aplicación de asistente de voz 262 basándose en la parte almacenada en memoria intermedia 642, 644 o 646 del flujo de sonido de entrada 610.

[0046] La **FIG. 7** es un gráfico que ilustra el cambio en una puntuación de palabra clave de un flujo de sonido de entrada que se determina para detectar una palabra clave de activación a partir del flujo de sonido de entrada y un final de la palabra clave de activación en el flujo de sonido de entrada, de acuerdo con un modo de realización de la presente divulgación. El flujo de sonido de entrada puede ser recibido por el sensor de sonido 210 en el dispositivo electrónico 200 y proporcionado al DSP 250. A medida que el flujo de sonido de entrada se recibe desde el sensor de sonido 210, el DSP 250 puede almacenar en memoria intermedia el flujo de sonido de entrada recibido en la unidad de almacenamiento en memoria intermedia 310 de la memoria intermedia 254. Para detectar la palabra clave de activación a partir del flujo de sonido de entrada y el final de la palabra clave de activación en el flujo de sonido de entrada, la unidad de detección de palabras clave 410 en la unidad de activación de voz 252 del DSP 250 puede determinar periódicamente o aperiódicamente una pluralidad de puntuaciones de palabras clave para flujo de sonido de entrada almacenado en la unidad de memoria de entrada 310.

[0047] En un modo de realización, dado que el flujo de sonido de entrada se recibe y se almacena en memoria intermedia, la unidad de detección de palabras clave 410 puede extraer secuencialmente una pluralidad de características de sonido (por ejemplo, huellas digitales de audio o vectores MFCC (coeficientes cepstrales de frecuencia Mel)) del flujo de sonido de entrada almacenado en memoria intermedia. En este caso, se puede determinar una puntuación de palabra clave para una o más de las características de sonido utilizando cualquier procedimiento de detección de palabras clave adecuado basándose en un modelo de cadena de Markov como un HMM, un SMM o similar. A medida que se recibe y almacena en memoria intermedia una parte más de la palabra clave de activación en el flujo de sonido de entrada, la puntuación de palabra clave puede aumentar y puede alcanzar una puntuación de palabra clave máxima cuando se recibe una parte sustancialmente completa de la palabra clave de activación. En el ejemplo ilustrado en la **FIG. 7**, la puntuación de palabra clave alcanza la puntuación de palabra clave máxima en el momento T_M cuando una parte sustancialmente completa de la palabra clave de activación en el flujo de sonido de entrada es recibida y almacenada, y luego disminuye.

[0048] Cada una de las puntuaciones de palabras clave puede compararse con una puntuación de umbral de detección predeterminada. La unidad de detección de palabras clave 410 puede detectar la palabra clave de activación cuando al menos una de las puntuaciones de palabras clave excede la puntuación de umbral de detección. La puntuación de umbral de detección puede indicar una puntuación para garantizar la detección de la palabra clave de activación que se incluye en el flujo de sonido de entrada almacenado en la unidad de memoria intermedia 310. En algunos modos de realización, la puntuación de umbral de detección puede establecerse de modo que la palabra clave de activación se detecte cuando una parte de la palabra clave de activación se incluye en el flujo de sonido de entrada. Por ejemplo, mientras se recibe la palabra clave de activación "INICIAR ASISTENTE DE VOZ", la unidad de detección de palabras clave 410 puede detectar la palabra clave de activación de una primera parte de la palabra clave de activación, por ejemplo, "INICIAR ASISTENTE DE VOZ" que está almacenada en memoria intermedia en la unidad de memoria intermedia 310. En otros modos de realización, la puntuación de umbral de detección puede establecerse de modo que la puntuación de palabra clave de activación se detecte cuando la palabra clave de activación completa se incluye en el flujo de sonido de entrada, por ejemplo, cuando se recibe un final de la palabra clave de activación. En el ejemplo ilustrado en la **FIG. 7**, la puntuación de umbral de detección se establece de modo que la unidad 410 de detección de palabras clave puede detectar la palabra clave de activación en el momento T_D cuando se recibe y se almacena en memoria intermedia una primera parte de la palabra clave de activación.

[0049] Además, el DSP 250 puede detectar un final de la palabra clave de activación en el flujo de sonido de entrada. Al detectar el final de la palabra clave de activación, el DSP 250 puede comenzar a almacenar en memoria intermedia una parte del flujo de sonido de entrada, que puede incluir al menos una parte de un comando de voz, en la unidad de memoria intermedia 320 de la memoria intermedia 254 como se describió anteriormente con referencia a la **FIG. 6**. En un modo de realización, después de detectar la palabra clave de activación en el momento T_D , el DSP 250 puede continuar almacenando el flujo de sonido de entrada, que se recibe desde el sensor de sonido 210, en la unidad de almacenamiento en memoria intermedia 310 de la memoria intermedia 254. El DSP 250 puede detectar el final de la palabra clave de activación basándose en una o más puntuaciones de palabras clave determinadas para el flujo de sonido de entrada almacenado en la unidad de memoria intermedia 310.

[0050] En un modo de realización, el DSP 250 puede calcular una pluralidad de puntuaciones de palabras clave basadas en características de sonido extraídas a partir de partes consecutivas del flujo de sonido de entrada almacenado en memoria intermedia en la unidad de memoria intermedia 310. Además, el DSP 250 puede determinar una puntuación de palabra clave máxima entre la pluralidad de puntuaciones de palabras clave para detectar el final de la palabra clave de activación. Por ejemplo, el DSP 250 puede determinar el final de la palabra clave de activación que se detectará en un momento cuando transcurre un período predeterminado de tiempo desde el momento en que se recibe una parte del flujo de sonido de entrada correspondiente a la puntuación de palabra clave máxima. En el ejemplo ilustrado en la **FIG. 7**, el DSP 250 puede determinar el final de la palabra clave de activación que se detectará en el momento T_{E1} , cuando transcurre un período predeterminado de tiempo P ya que la puntuación de palabra clave máxima se calcula para una parte del flujo de sonido de entrada recibido en el momento T_M .

[0051] En otro modo de realización, después de determinar la máxima puntuación de palabra clave, el DSP 250 puede detectar el final de la palabra clave de activación mediante la comparación de las puntuaciones de palabras clave determinados posteriormente con un umbral de puntuación final predeterminado. En este caso, el final de la palabra clave de activación puede detectarse cuando se determina que una de las puntuaciones de palabras clave determinadas posteriormente es menor que la puntuación de umbral final. En el ejemplo de la **FIG. 7**, el DSP 250 determina la puntuación de palabra clave máxima que se calcula para una parte del flujo de sonido de entrada recibido en el momento T_M , y a continuación detecta el final de la palabra clave de activación en el momento T_{E2} , desde el cual la puntuación de palabra clave está por debajo de la puntuación de umbral final.

[0052] La **FIG. 8** es un diagrama de temporización 800 de activación de la aplicación de asistente de voz 262 en respuesta a verificar si se activa la aplicación de asistente de voz 262, de acuerdo con un modo de realización de la presente divulgación. De forma similar al diagrama de temporización 600 en la **FIG. 6**, el dispositivo electrónico 200 puede recibir un flujo de sonido de entrada 810 que incluye una palabra clave de activación 820 y un comando de voz 830. En el diagrama de temporización 800, la palabra clave de activación 820 en el flujo de sonido de entrada 810

puede recibirse desde el momento T_1 hasta el momento T_2 . Así, el momento T_1 y el momento T_2 pueden corresponder a un principio y un fin, respectivamente, de la palabra clave de activación 820 en el flujo de sonido de entrada 810. Posteriormente, el comando de voz 830 en el flujo de sonido de entrada 810 puede recibirse desde el momento T_3 hasta el momento T_4 . Por lo tanto, el momento T_3 y el momento T_4 pueden corresponder a un principio y un final, respectivamente, del comando de voz 830 en el flujo de sonido de entrada 810.

[0053] A medida que el sensor de sonido 210 recibe el flujo de sonido de entrada 810 y proporciona el DSP 250 en el procesador 240 con el flujo de sonido de entrada recibido, el DSP 250 puede almacenar en memoria intermedia una parte del flujo de sonido de entrada 810, que puede incluir al menos una parte de la palabra clave de activación 820, en la unidad de memoria intermedia 310 de la memoria intermedia 254. Además, el DSP 250 puede almacenar en memoria intermedia otra parte del flujo de sonido de entrada 810, que puede incluir al menos una parte del comando de voz 830, en la unidad de memoria intermedia 320 de la memoria intermedia 254. El DSP 250 puede terminar el almacenamiento en memoria intermedia de la parte del flujo de sonido de entrada 810 en la unidad de memoria intermedia 320 cuando se detecta el final del comando de voz 830 en el momento T_4 . A continuación, la unidad de activación de voz 252 en el DSP 250 puede verificar si activar la aplicación de asistente de voz 262 basándose en la parte del flujo de sonido de entrada 810 almacenado en la unidad de memoria intermedia 320. Si se verifica la activación de la aplicación de asistente de voz 262, la unidad de activación de voz 252 puede activar la aplicación de asistente de voz 262.

[0054] En un modo de realización, cuando el DSP 250 termina el almacenamiento en memoria intermedia de la parte del flujo de sonido de entrada 810 en la unidad de memoria intermedia 320 en el momento T_4 y la unidad de activación de voz 252 verifica la activación de la aplicación de asistente de voz 262 basándose en la parte almacenada en memoria intermedia del flujo de sonido de entrada 810 en la unidad de memoria intermedia 320, la unidad de activación de voz 252 puede activar el procesador 240, que a su vez puede cargar la aplicación de asistente de voz 262 (como se indica con una flecha 840 en la **FIG. 8**). Según algunos modos de realización, la aplicación de asistente de voz 262 puede almacenarse inicialmente en la unidad de almacenamiento 260, por ejemplo, una ROM, una EEPROM o una memoria flash. Además, el procesador 240 puede cargar la aplicación de asistente de voz almacenada 262 en una memoria temporal (o una memoria de trabajo), por ejemplo, una memoria caché o una RAM, a la que puede acceder rápidamente el procesador 240.

[0055] Una vez que se carga la aplicación de asistente de voz 262, el procesador 240 inicia la aplicación de asistente de voz 262 para reconocer el comando de voz 830 y realizar una función de la aplicación de asistente de voz 262 que está asociada con el comando de voz reconocido 830. El procesador 240 puede iniciar la aplicación de asistente de voz 262 ejecutando una o más instrucciones codificadas en la aplicación de asistente de voz 262 que se carga en la memoria temporal. Además, la unidad de E/S 220 puede emitir el estado actual de la aplicación de asistente de voz iniciada 262 para mostrarlo a un usuario del dispositivo electrónico 200.

[0056] En otro modo de realización, la unidad de activación de voz 252 puede activar el procesador 240 para comenzar la carga de la aplicación de asistente de voz 262 al detectar la palabra clave de activación 820 en el momento T_D (como se indica con una flecha 850 en la **FIG. 8**). En este caso, el DSP 250 puede finalizar el almacenamiento en memoria intermedia de la parte del flujo de sonido de entrada 810 en la unidad de memoria intermedia 320 en el momento T_4 y a continuación verificar si se activa la aplicación de asistente de voz cargada 262 basándose en la parte almacenada en memoria intermedia del flujo de sonido de entrada 810 en la unidad de memoria intermedia 320. Si se verifica la activación de la aplicación de asistente de voz 262, el procesador 240 puede iniciar la aplicación de asistente de voz 262 después de que se complete la carga de la aplicación de asistente de voz 262.

[0057] En el modo de realización anterior, la carga de la aplicación de asistente de voz 262 mediante el procesador 240 puede comenzar tras la detección de la palabra clave de activación 820 en el momento T_D , y puede completarse antes de que el DSP 250 termine el almacenamiento en memoria intermedia de la parte del flujo de sonido de entrada 810 en la unidad de memoria intermedia 320 en el momento T_4 . Por ejemplo, como se muestra en la **FIG. 8**, la carga de la aplicación de asistente de voz 262 puede completarse en el momento T_L antes de que se detecte el final del comando de voz 830 en el momento T_4 . En tal caso, la carga de la aplicación de asistente de voz 262 puede completarse antes de que la unidad de activación de voz 252 verifique si se activa la aplicación de asistente de voz 262 basándose en la parte almacenada en memoria intermedia del flujo de sonido de entrada 810. En consecuencia, si se verifica la activación de la aplicación de asistente de voz 262, el procesador 240 puede iniciar la aplicación de asistente de voz 262 sin esperar a que se cargue la aplicación de asistente de voz 262.

[0058] De acuerdo con algunos modos de realización, si se ha completado la carga de la aplicación de asistente de voz 262 mediante el procesador 240 (por ejemplo, en el momento T_L) antes de que se reciba el final del comando de voz 830, el DSP 250 puede terminar almacenando en memoria intermedia una parte del flujo de sonido de entrada 810 en la unidad de memoria intermedia 320 al completar la carga de la aplicación de asistente de voz 262. Después de que el DSP 250 finaliza el almacenamiento en memoria intermedia de la parte del flujo de sonido de entrada 810 en la unidad de memoria intermedia 320, el procesador 240 puede comenzar a almacenar en memoria intermedia otra parte del flujo de sonido de entrada 810 que se recibe posteriormente desde el sensor de sonido 210, por ejemplo, en la unidad de almacenamiento 260. Como resultado, la parte del flujo de sonido de entrada 810 almacenada en la unidad de memoria intermedia 320 puede incluir una primera parte del comando de voz 830, y la parte del flujo de

sonido de entrada 810 almacenada en la unidad de almacenamiento 260 puede incluir una segunda parte del comando de voz 830 que sigue a la primera parte. Las partes del flujo de sonido de entrada 810 almacenadas en la unidad de memoria intermedia 320 y la unidad de almacenamiento 260 pueden usarse para reconocer el comando de voz 830 mediante la aplicación de asistente de voz 262 para realizar una función asociada con el comando de voz 830.

[0059] Cuando el DSP 250 termina el almacenamiento en memoria intermedia de la parte del flujo de sonido de entrada 810 en la unidad de memoria intermedia 320, la unidad de activación de voz 252 puede verificar si activar la aplicación de asistente de voz 262 basándose en la parte del flujo de sonido de entrada 810 almacenada en memoria intermedia en la unidad de memoria intermedia 320 que incluye la primera parte del comando de voz 830. Si se verifica la activación de la aplicación de asistente de voz 262, el procesador 240 puede iniciar la aplicación de asistente de voz 262, que puede acceder a ambas partes del flujo de sonido de entrada 810 almacenado en memoria intermedia en la unidad de memoria intermedia 320 y la unidad de almacenamiento 260. De manera adicional o alternativa, la parte del flujo de sonido de entrada 810 almacenada en la unidad de memoria intermedia 320 puede transmitirse o copiarse a la unidad de almacenamiento 260, de modo que la aplicación de asistente de voz 262 pueda acceder a la primera y segunda parte del comando de voz 830 desde unidad de almacenamiento 260.

[0060] La **FIG. 9** es un gráfico que ilustra el cambio en una puntuación de palabra clave de un flujo de sonido de entrada que se determina para cargar la aplicación de asistente de voz 262 basándose en una primera puntuación de umbral de detección y verificar la carga de la aplicación de asistente de voz 262 basándose en una segunda puntuación de umbral de detección, de acuerdo con un modo de realización de la presente divulgación. El flujo de sonido de entrada puede ser recibido por el sensor de sonido 210 en el dispositivo electrónico 200 y proporcionado al DSP 250, que puede almacenar una parte del flujo de sonido de entrada recibido desde el sensor de sonido 210 en la unidad de almacenamiento en memoria intermedia 310 de la memoria intermedia 254. La parte almacenada en memoria intermedia del flujo de sonido de entrada en la unidad de memoria intermedia 310 puede incluir al menos una parte de una palabra clave de activación. La unidad de detección de palabras clave 410 en la unidad de activación de voz 252 del DSP 250 puede extraer una pluralidad de características de sonido de la parte almacenada en memoria intermedia del flujo de sonido de entrada y determinar una pluralidad de puntuaciones de palabras clave para la pluralidad de características de sonido.

[0061] En el modo de realización ilustrado, la unidad de detección de palabras clave 410 puede estar configurado para comparar cada una de las puntuaciones de la palabra clave con la primera puntuación de umbral de detección. La primera puntuación de umbral de detección puede indicar una puntuación de palabra clave para garantizar la detección de la palabra clave de activación que se incluye en la parte almacenada en memoria intermedia del flujo de sonido de entrada, con una confianza relativamente baja. Si se determina que una de la pluralidad de puntuaciones de palabras clave (por ejemplo, determinada para una parte del flujo de sonido de entrada recibido en el momento T_{D1} en la **FIG. 9**) excede la primera puntuación de umbral de detección, se puede inferir con la confianza relativamente baja que la palabra clave de activación se detecta desde el flujo de sonido de entrada. Al detectar la palabra clave de activación, la unidad de detección de palabras clave 410 puede activar el procesador 240 para comenzar a cargar la aplicación de asistente de voz 262 desde la unidad de almacenamiento 260 en una memoria temporal (o una memoria de trabajo) a la que puede acceder rápidamente el procesador 240.

[0062] Mientras se carga la aplicación de asistente de voz 262, el DSP 250 puede continuar recibiendo el flujo de sonido de entrada del sensor de sonido 210 y almacenar en memoria intermedia una parte del flujo de sonido de entrada, que puede incluir al menos una parte de la palabra clave de activación, en la unidad de memoria intermedia 310. La unidad de detección de palabras clave 410 puede verificar la carga de la aplicación de asistente de voz 262 basándose en la parte del flujo de sonido de entrada que está almacenada en la unidad de memoria intermedia 310 mientras se carga la aplicación de asistente de voz 262. Por ejemplo, la unidad de detección de palabras clave 410 puede determinar una pluralidad de puntuaciones de palabras clave para la parte almacenada en memoria intermedia del flujo de sonido de entrada y comparar cada una de la pluralidad de puntuaciones de palabras clave con la segunda puntuación de umbral de detección. La segunda puntuación de umbral de detección puede indicar una puntuación de palabra clave para garantizar la detección de la palabra clave de activación que se incluye en el flujo de sonido de entrada, con una confianza relativamente alta. En consecuencia, la segunda puntuación de umbral de detección puede establecerse para que sea mayor que la primera puntuación de umbral de detección como se ilustra en el gráfico de la **FIG. 9**.

[0063] Si se determina que una de la pluralidad de puntuaciones de palabras clave excede la segunda puntuación de umbral de detección (por ejemplo, en el momento T_{D2} en la **FIG. 9**), se puede inferir con la confianza relativamente alta de que la palabra clave de activación se detecta desde flujo de sonido de entrada, y la carga de la aplicación de asistente de voz 262 puede verificarse. En este caso, la unidad de detección de palabras clave 410 puede indicar al procesador 240 que continúe cargando la aplicación de asistente de voz 262. Posteriormente, el DSP 250 puede continuar recibiendo el flujo de sonido de entrada desde el sensor de sonido 210 y almacenar una parte del flujo de sonido de entrada, que puede incluir al menos una parte de un comando de voz, en la unidad de memoria intermedia 320 de la memoria intermedia 254. A continuación, la unidad de verificación 420 en la unidad de activación de voz 252 puede verificar si activar la aplicación de asistente de voz 262 basándose en la parte del flujo de sonido de entrada almacenado en memoria intermedia en la unidad de memoria intermedia 320 para iniciar la aplicación de asistente de voz 262 de la manera descrita anteriormente con respecto a la **FIG. 8**. Por otro lado, si se determina que ninguna de

la pluralidad de puntuaciones de palabras clave excede la segunda puntuación de umbral de detección, puede inferirse que el flujo de sonido de entrada no incluye la palabra clave de activación. En este caso, el dispositivo electrónico 200 puede proceder a recibir un siguiente flujo de sonido de entrada para detectar la palabra clave de activación sin iniciar la aplicación de asistente de voz 262.

5
 [0064] En algunos modos de realización, cuando un usuario dice una palabra clave de activación para activar una aplicación objetivo de un dispositivo electrónico, el dispositivo electrónico puede verificar si se activa la aplicación objetivo basándose en la información ambiental generada por uno o más sensores de contexto en el dispositivo electrónico. La FIG. 10 ilustra el dispositivo móvil 120 que activa una aplicación de navegador web 1010 en respuesta a la detección de una palabra clave de activación en un flujo de sonido de entrada, de acuerdo con un modo de realización de la presente divulgación. De forma similar al modo de realización de la FIG. 1, el dispositivo móvil 120 puede almacenar previamente una palabra clave de activación para activar la aplicación de navegador web 1010 en el dispositivo móvil 120. En el modo realización ilustrado, cuando el usuario 110 pronuncia una palabra clave de activación tal como "INICIAR NAVEGADOR" al dispositivo móvil 120, el dispositivo móvil 120 puede capturar un flujo de sonido de entrada y detectar la palabra clave de activación en el flujo de sonido de entrada.

10
 15
 20
 25
 30
 [0065] El dispositivo móvil 120 puede incluir uno o más sensores contextuales, cada uno configurado para generar datos de sensor indicativos de un contexto del dispositivo móvil 120. Como se usa en el presente documento, el término "contexto" de un dispositivo electrónico puede referirse a un entorno en el que se encuentra el dispositivo, como una calle, una oficina, un vehículo, un restaurante, una instalación de transporte y similares. El término "contexto" también puede referirse a actividades o situaciones en las que el dispositivo o su usuario está involucrado, como estar en un ambiente oscuro, dormir en casa, mirar un programa de televisión, escuchar música, jugar fútbol en un patio de recreo y similares. Por ejemplo, el uno o más sensores de contexto pueden incluir un sensor de imagen 1020 tal como una cámara que está montada en una superficie del dispositivo móvil 120 que incluye una pantalla de visualización 1030, como se muestra en la FIG. 10. El sensor de imagen 1020 puede configurarse para generar datos del sensor capturando una imagen de entrada. Por ejemplo, cuando el usuario 110 ve la pantalla de visualización 1030, la imagen de entrada capturada puede incluir una parte de una cara del usuario 110 (por ejemplo, los ojos del usuario 110). En algunos modos de realización, al detectar la palabra clave de activación, el sensor de imagen 1020 puede capturar una imagen de entrada y el dispositivo móvil 120 puede almacenar en memoria intermedia la imagen de entrada capturada como datos del sensor en una unidad de memoria.

35
 40
 [0066] Cuando la palabra clave de activación para activar la aplicación de navegador web 1010 se detecta desde el flujo de sonido de entrada, el dispositivo móvil 120 puede verificar si se activa la aplicación de navegador web 1010 basándose en los datos del sensor generados por el sensor de imagen 1020. Por ejemplo, el dispositivo móvil 120 puede verificar si se activa la aplicación de navegador web 1010 determinando si la imagen de entrada capturada por el sensor de imagen 1020 incluye al menos un ojo de una persona (por ejemplo, el usuario 110). Si se determina que la imagen de entrada incluye al menos un ojo de una persona, el dispositivo móvil 120 verifica la activación de la aplicación del navegador web 1010 y activa la aplicación del navegador web 1010. Una vez activada, la aplicación de navegador web 1010 puede mostrarse en la pantalla de visualización 1030 para que el usuario 110 pueda ver una página web mostrada por la aplicación de navegador web 1010. Por otro lado, si se determina que la imagen de entrada no incluye ningún ojo de una persona, la activación de la aplicación del navegador web 1010 no se verifica y el dispositivo móvil 120 puede proceder a recibir un siguiente flujo de sonido de entrada para detectar la palabra clave de activación.

45
 50
 55
 [0067] La FIG. 11 ilustra un diagrama de bloques de un dispositivo electrónico 1100 configurado para activar una aplicación de navegador web 1130 basándose en datos del sensor generados por un sensor de contexto 1120 en el dispositivo electrónico 1100, de acuerdo con un modo de realización de la presente divulgación. El dispositivo electrónico 1100 puede incluir una unidad de sensor 1110, la unidad de E/S 220, la unidad de comunicación 230, el procesador 240 y la unidad de almacenamiento 260. La unidad de sensor 1110 puede incluir el sensor de sonido 210 y el sensor de contexto 1120. El procesador 240 puede incluir el DSP 250 que puede incluir la unidad de activación de voz 252 y la memoria intermedia 254. En el modo de realización ilustrado, el sensor de sonido 210, la unidad de E/S 220, la unidad de comunicación 230, el procesador 240, el DSP 250, la unidad de activación de voz 252, la memoria intermedia 254 y la unidad de almacenamiento 260 pueden tener una unidad similar configuración o función a las incluidas en el dispositivo electrónico 200 de la FIG. 2. Además, el dispositivo electrónico 1100 puede ser cualquier dispositivo adecuado equipado con capacidades de captura y procesamiento de sonido, como un teléfono celular, un teléfono inteligente (por ejemplo, el dispositivo móvil 120), un ordenador personal, un ordenador portátil, una tablet, una televisión inteligente, un dispositivo de juego, un reproductor multimedia, gafas inteligentes, un ordenador que se puede llevar encima, etc.

60
 65
 [0068] El sensor de sonido 210 puede recibir un flujo de sonido de entrada y proporcionar el flujo de sonido de entrada recibido al DSP 250. A medida que el flujo de sonido de entrada se recibe desde el sensor de sonido 210, la unidad de activación de voz 252 en el DSP 250 puede detectar una palabra clave de activación (por ejemplo, "INICIAR NAVEGADOR") para activar la aplicación del navegador web 1130 desde el flujo de sonido de entrada recibido. Para detectar la palabra clave de activación, la unidad de activación de voz 252 puede emplear cualquier procedimiento de detección de palabras clave adecuado basándose en un modelo de cadena de Markov tal como un HMM, un SMM o una combinación de los mismos.

[0069] En un modo de realización, la unidad de almacenamiento 260 puede almacenar cualquier información y/o datos adecuados para la detección de palabras clave tales como la información de estado para una pluralidad de estados de la palabra clave de activación bajo un modelo de cadena de Markov tal como un HMM, un SMM, o similares. La unidad de almacenamiento 260 también puede almacenar la aplicación de navegador web 1130 a la que puede acceder el procesador 240. Aunque la unidad de almacenamiento 260 almacena la aplicación de navegador web 1130 en el modo de realización ilustrado, también puede almacenar cualquier otra aplicación o función del dispositivo electrónico 1100, como una aplicación de cámara, una aplicación de notas, una aplicación de calendario, una aplicación de reproductor de música y similares.

[0070] El contexto sensor 1120 en la unidad de sensor 1110 puede incluir uno o más sensores, cada uno configurado para generar datos de sensor indicativos de un contexto de dispositivo electrónico 1100. Por ejemplo, el sensor de contexto 1120 puede incluir un sensor de cámara que está configurado para capturar una imagen de entrada como datos del sensor. El sensor de la cámara puede montarse en una superficie del dispositivo electrónico 1100 que incluye una pantalla de visualización de la unidad de E/S 220, de modo que el sensor de la cámara puede capturar una imagen que incluye uno o dos ojos de una persona (por ejemplo, el usuario 110) frente a la pantalla de visualización. De forma adicional o alternativa, el sensor de contexto 1120 puede incluir cualquier otro sensor adecuado, como un sensor de iluminación, un sensor de proximidad, un sensor de ubicación (por ejemplo, un sistema de posicionamiento global (GPS)), un acelerómetro, un sensor de movimiento, un sensor de contacto, y similares. En algunos modos de realización, el sensor de sonido 210 puede configurarse adicionalmente para generar datos de sensor indicativos de un contexto del dispositivo electrónico 1100 mediante la captura de sonido ambiental.

[0071] Al detectar la palabra clave de activación, el sensor de contexto 1120 puede proporcionar los datos del sensor al DSP 250, que puede almacenar los datos del sensor en la memoria intermedia 254. La unidad de activación de voz 252 puede verificar si se activa la aplicación de navegador web 1130 basándose en los datos del sensor almacenados en la memoria intermedia. Por ejemplo, el DSP 250 puede recibir la imagen de entrada como datos del sensor del sensor de la cámara del sensor de contexto 1120. Además, la unidad de activación de voz 252 puede verificar si se activa la aplicación de navegador web 1130 basándose en la imagen de entrada.

[0072] En un modo de realización, la unidad de activación de voz 252 puede verificar si se activa la aplicación de navegador web 1130 mediante la determinación de si la imagen de entrada incluye al menos un ojo de una persona. Si se determina que la imagen de entrada incluye al menos un ojo de una persona, la unidad de activación de voz 252 puede activar el procesador 240 que a su vez puede activar la aplicación de navegador web 1130. Por otro lado, si se determina que la imagen de entrada no incluye ningún ojo de una persona, la activación de la aplicación del navegador web 1130 no se verifica y el dispositivo electrónico 1100 puede proceder a recibir un próximo flujo de sonido de entrada para detectar la palabra clave de activación.

[0073] En algunos modos de realización, las operaciones de almacenamiento en memoria intermedia de los datos de sensor en la memoria intermedia 254 y la activación de la aplicación de navegador web 1130 pueden llevarse a cabo de maneras similares a las descritas anteriormente con referencia a las **FIGS. 6 y 8**. Además, aunque el dispositivo electrónico 1100 activa la aplicación del navegador web 1130 verificando la activación de la aplicación del navegador web 1130 basándose en de los datos del sensor, como una imagen de entrada capturada por un sensor de cámara, el dispositivo electrónico 1100 también puede configurarse para activar cualquier otra aplicación adecuada verificando si activar las aplicaciones basándose en de los datos del sensor generados a partir de cualquier otro sensor de contexto adecuado. En un modo de realización, el dispositivo electrónico 1100 puede activar una aplicación de cámara cuando la activación de la aplicación de cámara se verifica basándose en los datos del sensor de un sensor de iluminación en el sensor de contexto 1120. Por ejemplo, la unidad de activación de voz 252 puede verificar la activación de la aplicación de la cámara cuando los datos del sensor del sensor de iluminación indican que el entorno del dispositivo electrónico 1100 es lo suficientemente brillante como para ser visible. Si se verifica la activación de la aplicación de la cámara, la unidad de activación de voz 252 puede activar el procesador 240 para activar la aplicación de la cámara.

[0074] La **FIG. 12** ilustra un diagrama de flujo de un procedimiento 1200 para activar una aplicación objetivo detectando una palabra clave de activación y verificando si activar la aplicación objetivo basándose en de los datos del sensor indicativos de un contexto del dispositivo electrónico 1100, de acuerdo con un modo de realización de la presente divulgación. El dispositivo electrónico 1100 puede recibir un flujo de sonido de entrada que incluye la palabra clave de activación para activar la aplicación objetivo, en 1210. Por ejemplo, la palabra clave de activación puede ser "INICIAR NAVEGADOR" y la aplicación objetivo puede ser la aplicación de navegador web 1130 como se describió anteriormente con referencia a la **FIG. 11**.

[0075] En un modo de realización, una parte del flujo de sonido de entrada recibido incluyendo al menos una parte de la palabra clave de activación se puede almacenar en memoria intermedia en la memoria intermedia 254 en el DSP 250. La unidad de activación de voz 252 en el DSP 250 puede detectar la palabra clave de activación desde la parte del flujo de sonido de entrada almacenado en la memoria intermedia 254, a 1220. Si se detecta la palabra clave de activación, uno o más sensores de contexto en el dispositivo electrónico 1100 pueden generar los datos del sensor indicativos del contexto del dispositivo electrónico 1100, en 1230. Los datos del sensor generados pueden almacenarse en la memoria intermedia 254.

[0076] La unidad de activación de voz 252 puede verificar si se activa la aplicación objetivo basándose en los datos del sensor, en 1240. Si la unidad de activación de voz 252 verifica la activación de la aplicación objetivo (es decir, "SÍ" en 1240), la unidad de activación de voz 252 puede activar la aplicación objetivo, en 1250. Por otro lado, no se verifica la activación de la aplicación objetivo (es decir, "NO" en 1240), el procedimiento 1200 pasa a 1210 para que el dispositivo electrónico 1100 pueda recibir un flujo de sonido de entrada siguiente para detectar una palabra clave de activación.

[0077] La **FIG. 13** es un diagrama de bloques de un dispositivo electrónico a modo de ejemplo 1300 en el que los procedimientos y aparatos para activar una aplicación objetivo al detectar una palabra clave de activación pueden implementarse de acuerdo con algunos modos de realización de la presente divulgación. La configuración del dispositivo electrónico 1300 puede implementarse en los dispositivos electrónicos de acuerdo con los modos de realización anteriores descritos con referencia a las **FIGS. 1 a 12**. El dispositivo electrónico 1300 puede ser un teléfono celular, un teléfono inteligente, una tablet, un ordenador portátil, un terminal, un teléfono, un asistente digital personal (PDA), un módem inalámbrico, un teléfono inalámbrico, etc. El sistema de comunicación inalámbrica puede ser un sistema de acceso múltiple por división de código (CDMA), un sistema de radiodifusión para comunicaciones móviles (GSM), un sistema de banda ancha CDMA (WCDMA), un sistema de evolución a largo plazo (LTE), un sistema LTE avanzado, etc. Además, el dispositivo electrónico 1300 puede comunicarse directamente con otro dispositivo móvil, por ejemplo, usando Wi-Fi Direct o Bluetooth.

[0078] 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 mediante una antena 1312 y se proporcionan a un receptor (RCVR) 1314. El receptor 1314 condiciona y digitaliza la señal recibida, y proporciona muestras como la señal digital condicionada y digitalizada a una sección digital para su procesamiento adicional. En la vía de transmisión, un transmisor (TMTR) 1316 recibe datos que se van a transmitir desde la sección digital 1320, procesa y acondiciona los datos, y genera una señal modulada, que se transmite a través de la antena 1312 a las estaciones base. El receptor 1314 y el transmisor 1316 pueden formar parte de un transceptor que puede admitir CDMA, GSM, W-CDMA, LTE, LTE avanzada, etc.

[0079] 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/vídeo 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 recepción de datos, por ejemplo, codificación, modulación, desmodulación y decodificación. El RISC/DSP 1324 puede realizar un procesamiento general y especializado para el dispositivo electrónico 1300. El controlador/procesador 1326 puede realizar 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.

[0080] El codificador de audio/vídeo generalizado 1332 puede realizar la codificación para señales de entrada desde una fuente de audio/vídeo 1342, un micrófono 1344, un sensor de imagen 1346, etc. El decodificador de audio generalizado 1334 puede realizar la decodificación de datos de audio codificados y puede proporcionar señales de salida a un altavoz/auriculares 1348. El procesador de gráficos/pantalla 1336 puede realizar un procesamiento de gráficos, vídeos, imágenes y textos, que se pueden presentar a una unidad de pantalla 1350. La EBI 1338 puede facilitar la transferencia de datos entre la sección digital 1320 y una memoria principal 1352.

[0081] La sección digital 1320 se puede implementar con uno o más procesadores, DSP, microprocesadores, RISC, etc. La sección digital 1320 también se puede fabricar en uno o más circuitos integrados específicos de la aplicación (ASIC) y/o algún otro tipo de circuitos integrados (IC).

[0082] En general, cualquier dispositivo descrito en el presente documento puede representar diversos tipos de dispositivos, tales como un teléfono inalámbrico, un teléfono móvil, un ordenador portátil, un dispositivo multimedia inalámbrico, una tarjeta de ordenador personal (PC) de comunicación inalámbrica, un PDA, un módem externo o interno, un dispositivo que se comunica a través de un canal inalámbrico, etc. Un dispositivo puede tener diversos nombres, tales como terminal de acceso (AT), unidad de acceso, unidad de abonado, estación móvil, dispositivo móvil, unidad móvil, teléfono móvil, móvil, estación remota, terminal remoto, unidad remota, dispositivo de usuario, equipo de usuario, dispositivo 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.

[0083] Las técnicas descritas en el presente documento se pueden implementar por diversos medios. Por ejemplo, estas técnicas pueden implementarse en hardware, firmware, software o una combinación de los mismos. Los expertos en la técnica apreciarán además que los diversos bloques lógicos, módulos, circuitos y pasos de algoritmo ilustrativos descritos en relación con la divulgación en el presente documento se pueden implementar 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 pasos ilustrativos, en

general, en términos de su funcionalidad. Que dicha funcionalidad se implemente como hardware o software depende de las restricciones de aplicación y diseño en particular impuestas al sistema global. Los expertos en la técnica pueden implementar la funcionalidad descrita de distintas formas para cada aplicación en particular, pero no se debe interpretar que dichas decisiones de implementación suponen apartarse del alcance de la presente divulgación.

[0084] Para una implementación en hardware, las unidades de procesamiento usadas para realizar las técnicas se pueden implementar dentro de uno o más ASIC, DSP, dispositivos de procesamiento de señales digitales (DSPD), dispositivos de lógica programable (PLD), matrices de puertas programables *in situ* (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.

[0085] Por tanto, los diversos bloques lógicos, módulos y circuitos ilustrativos descritos en relación con la divulgación en el presente documento se implementan o realizan con un procesador de propósito general, un DSP, un ASIC, una FPGA u otro dispositivo de lógica programable, lógica de transistor o de puertas discretas, componentes de hardware discretos, o con cualquier combinación de los mismos diseñada para realizar las funciones descritas en el presente documento. Un procesador de 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 estar implementado 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.

[0086] Si se implementan en software, las funciones, como una o más instrucciones o código, pueden almacenarse en, o transmitirse por, un medio legible por ordenador. 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 se pueda usar para transportar o almacenar el código de programa deseado en forma de instrucciones o estructuras de datos y al que se pueda acceder 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 inalámbricas, tales como infrarrojos, radio y microondas, están incluidos en la definición de medio. El término discos, como se usa en el presente documento, incluye el disco compacto (CD), el disco láser, el disco óptico, el disco versátil digital (DVD), el disco flexible y el disco Blu-ray, donde los discos flexibles reproducen habitualmente datos magnéticamente, mientras que los demás discos reproducen los datos ópticamente con láseres. Las combinaciones de los anteriores también se deben incluir dentro del alcance de los medios legibles por ordenador.

[0087] La descripción previa de la divulgación se proporciona para permitir que cualquier experto en la técnica realice o use la divulgación. Varias modificaciones de la divulgación serán fácilmente evidentes para los expertos en la técnica, y los principios genéricos definidos en el presente documento se aplican a otras variaciones sin apartarse del espíritu o alcance de la divulgación. Por tanto, la divulgación no está limitada a 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.

[0088] Aunque las implementaciones a modo de ejemplo se refieren a utilizar aspectos de la materia objeto divulgada actualmente en el contexto de uno o más sistemas informáticos autónomos, la materia objeto no está limitada a ello, sino que en su lugar se puede implementar en relación con cualquier entorno informático, tal como una red o un entorno informático distribuido. Aún más, se pueden implementar aspectos de la materia objeto divulgada actualmente en o a través de una pluralidad de chips o dispositivos de procesamiento, y el almacenamiento se puede realizar de forma similar a través de una pluralidad de dispositivos. Dichos dispositivos pueden incluir PC, servidores de red y dispositivos portátiles.

[0089] Aunque la materia objeto se ha descrito en un lenguaje específico con respecto a los rasgos característicos estructurales y/o acciones metodológicas, se debe entender que la materia objeto definida en las reivindicaciones adjuntas no se limita necesariamente a los rasgos característicos o acciones específicas descritos anteriormente. En su lugar, los rasgos característicos y acciones específicas descritas anteriormente se divulgan como formas de ejemplo de implementación de las reivindicaciones.

REIVINDICACIONES

1. Un procedimiento, realizado en un dispositivo electrónico, para activar una aplicación de asistente de voz, siendo la aplicación de asistente de voz una aplicación configurada para reconocer un comando de voz y realizar la función asociada con el comando de voz, con el procedimiento que comprende:
 - recibir un flujo de sonido de entrada que incluye una palabra clave de activación para activar la aplicación de asistente de voz y una parte que sigue a la palabra clave de activación que puede incluir al menos una parte de un comando de voz indicativo de una función de la aplicación de asistente de voz;
 - detectar la palabra clave de activación a partir del flujo de sonido de entrada;
 - en respuesta a la detección de la palabra clave de activación, iniciar almacenamiento en memoria intermedia ("buffering"), en una memoria intermedia, de una parte del flujo de sonido de entrada que incluye al menos una parte que sigue a la palabra clave de activación; y
 - después de iniciar el almacenamiento den memoria intermedia, verificar si se activa la aplicación de asistente de voz basándose en la parte almacenada en memoria intermedia en el flujo de sonido de entrada, y si se verifica la activación de la aplicación de asistente de voz, activar la aplicación de asistente de voz para realizar la función de la aplicación de asistente de voz.
2. El procedimiento según la reivindicación 1, en el que la detección de la palabra clave de activación comprende detectar la palabra clave de activación a partir de una parte de la palabra clave de activación en el flujo de sonido de entrada, en el que el almacenamiento en memoria intermedia de la parte del flujo de sonido de entrada comprende el almacenamiento de la parte del flujo de sonido de entrada en respuesta a detectar la palabra clave de activación a partir de la parte de la palabra clave de activación.
3. El procedimiento según la reivindicación 1, que comprende además detectar un final de la palabra clave de activación en el flujo de sonido de entrada, y en el que almacenar en memoria intermedia la parte del flujo de sonido de entrada comprende almacenar en memoria intermedia la parte del flujo de sonido de entrada en respuesta a la detección del final de la palabra clave de activación.
4. El procedimiento según la reivindicación 1, que comprende además detectar un comienzo del comando de voz en el flujo de sonido de entrada, y en el que el almacenamiento en memoria intermedia de la parte del flujo de sonido de entrada comprende almacenar en memoria intermedia la parte del flujo de sonido de entrada en respuesta a la detección del comienzo del comando de voz.
5. El procedimiento según la reivindicación 1, en el que verificar si activar la aplicación de asistente de voz comprende determinar si la parte almacenada en memoria intermedia del flujo de sonido de entrada incluye voz.
6. El procedimiento según la reivindicación 5, en el que verificar si activar la aplicación de asistente de voz comprende además determinar si la voz en la parte almacenada en memoria intermedia del flujo de sonido de entrada es indicativa de un hablante de la palabra clave de activación.
7. El procedimiento según la reivindicación 1, en el que verificar si activar la aplicación de asistente de voz comprende determinar si una relación de señal a ruido (SNR) de la parte almacenada en memoria intermedia del flujo de sonido de entrada excede un valor umbral predeterminado.
8. El procedimiento según la reivindicación 1, en el que verificar si activar la aplicación de asistente de voz comprende determinar si la parte almacenada en memoria intermedia del flujo de sonido de entrada incluye un comando de revocación predeterminado indicativo de revocar la palabra clave de activación.
9. El procedimiento según la reivindicación 2, en el que detectar la palabra clave de activación comprende detectar la palabra clave de activación a partir de una parte de la palabra clave de activación en el flujo de sonido de entrada, y en el que activar la aplicación de asistente de voz comprende:
 - cargar la aplicación de asistente de voz en respuesta a la detección de la palabra clave de activación a partir de la parte de la palabra clave de activación; e
 - iniciar la aplicación de asistente de voz en respuesta a la verificación de activación de la aplicación de asistente de voz.
10. El procedimiento según la reivindicación 1, en el que la activación de la aplicación de asistente de voz comprende:

reconocer el comando de voz a partir de la parte almacenada en memoria intermedia del flujo de sonido de entrada; y

realizar la función de la aplicación de asistente de voz en respuesta al comando de voz.

5 11. Un medio de almacenamiento legible por ordenador no transitorio que comprende instrucciones para activar una aplicación de asistente de voz, con las instrucciones que hacen que un procesador de un dispositivo electrónico realice el procedimiento de acuerdo con cualquiera de las reivindicaciones anteriores.

10 12. Un dispositivo electrónico para activar una aplicación de asistente de voz, siendo la aplicación de asistente de voz una aplicación configurada para reconocer un comando de voz y realizar la función asociada con el comando de voz, que comprende:

15 medios para recibir un flujo de sonido de entrada que incluye una palabra clave de activación para activar la aplicación de asistente de voz y un comando de voz indicativo de una función de la aplicación de asistente de voz;

20 medios para detectar la palabra clave de activación a partir del flujo de sonido de entrada y verificar si se activa la aplicación de asistente de voz para realizar la función de la aplicación de asistente de voz;

medios para iniciar el almacenamiento antes de la verificación de si activar la aplicación de asistente de voz de una parte del flujo de sonido de entrada después de la palabra clave de activación, que puede incluir al menos una parte del comando de voz en respuesta a la detección de la palabra clave de activación

25 medios para activar la aplicación de asistente de voz después de verificar la activación de la aplicación de asistente de voz.

30 13. El dispositivo electrónico según la reivindicación 12, en el que los medios para detectar la palabra clave de activación y activar la aplicación de asistente de voz están configurados para detectar la palabra clave de activación a partir de una parte de la palabra clave de activación en el flujo de sonido de entrada, y en el que los medios para almacenar la parte del flujo de sonido de entrada están configurados para almacenar la parte del flujo de sonido de entrada en respuesta a la detección de la palabra clave de activación a partir de la parte de la palabra clave de activación.

35 14. El dispositivo electrónico según la reivindicación 12, en el que los medios para detectar la palabra clave de activación y activar la aplicación de asistente de voz están configurados para:

40 verificar si activar la aplicación de asistente de voz basándose en la parte almacenada del flujo de sonido de entrada;

detectar la palabra clave de activación a partir de una parte de la palabra clave de activación en el flujo de sonido de entrada;

45 cargar la aplicación de asistente de voz en respuesta a la detección de la palabra clave de activación a partir de la parte de la palabra clave de activación; e

iniciar la aplicación de asistente de voz en respuesta a la verificación de la activación de la aplicación de asistente de voz.

50 15. El dispositivo electrónico según la reivindicación 12, en el que los medios para recibir, los medios para detectar y los medios para iniciar están integrados en un dispositivo de comunicación móvil.

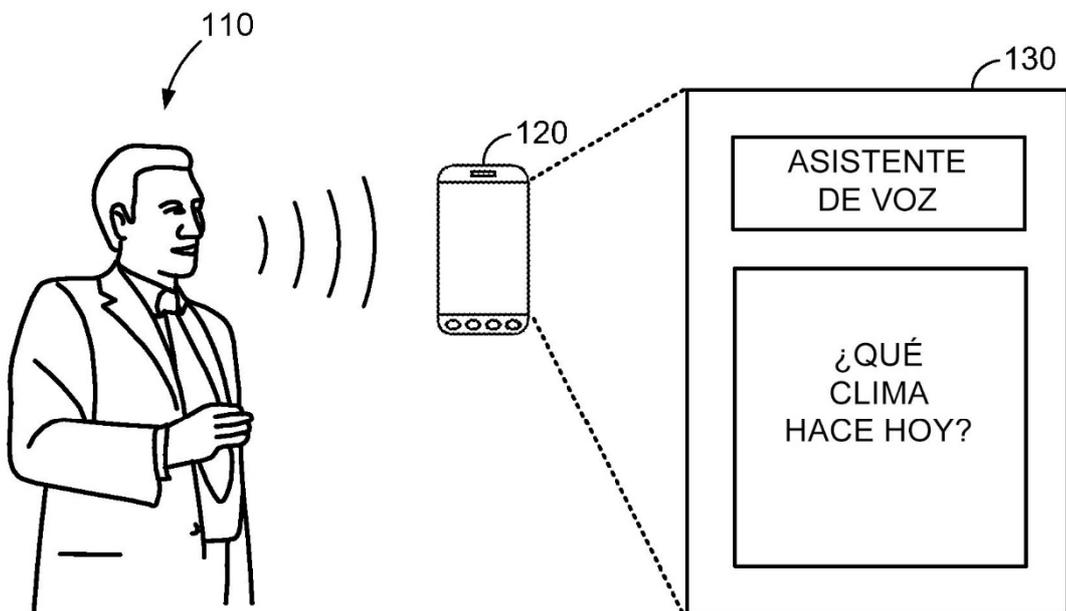


FIG. 1

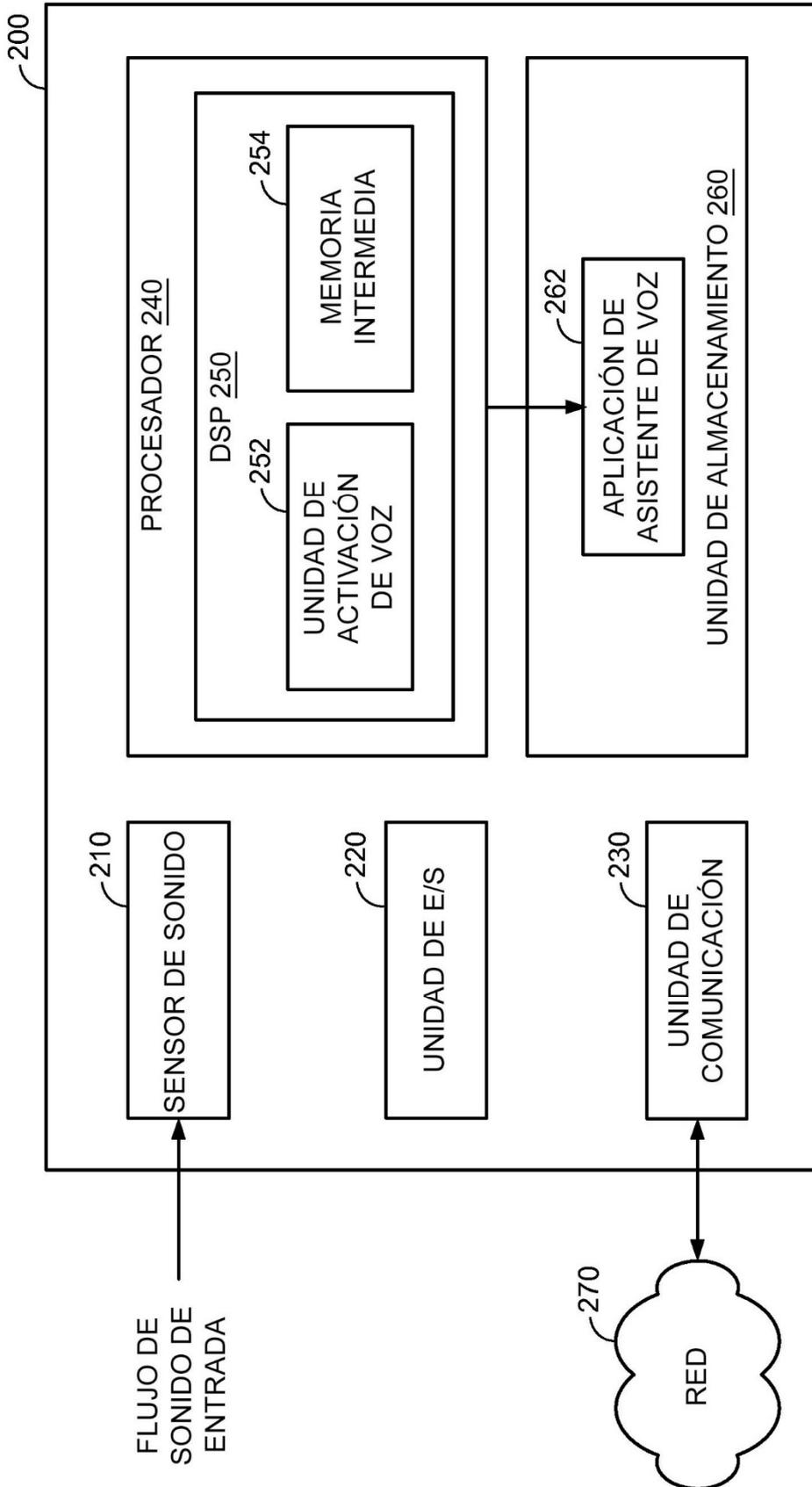


FIG. 2

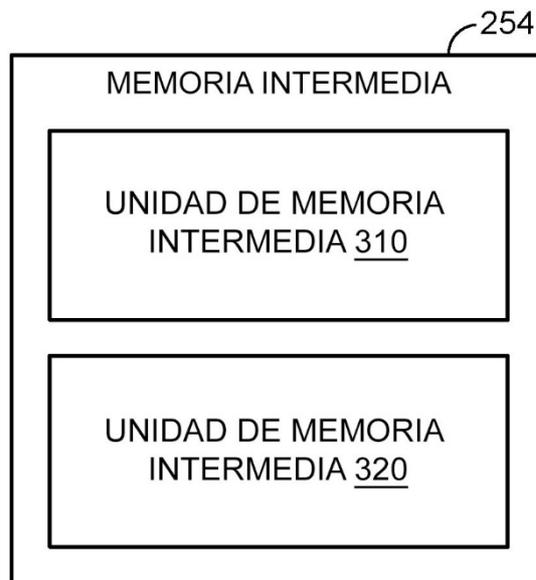


FIG. 3

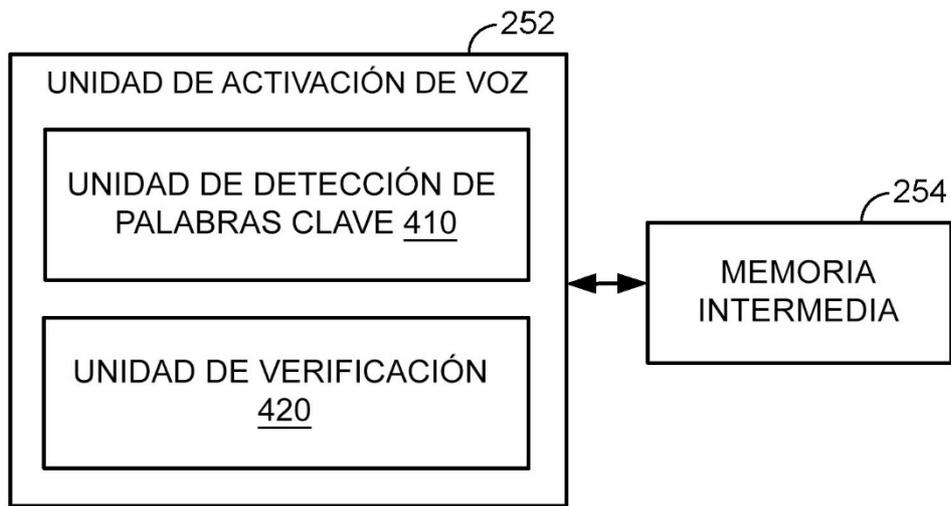


FIG. 4

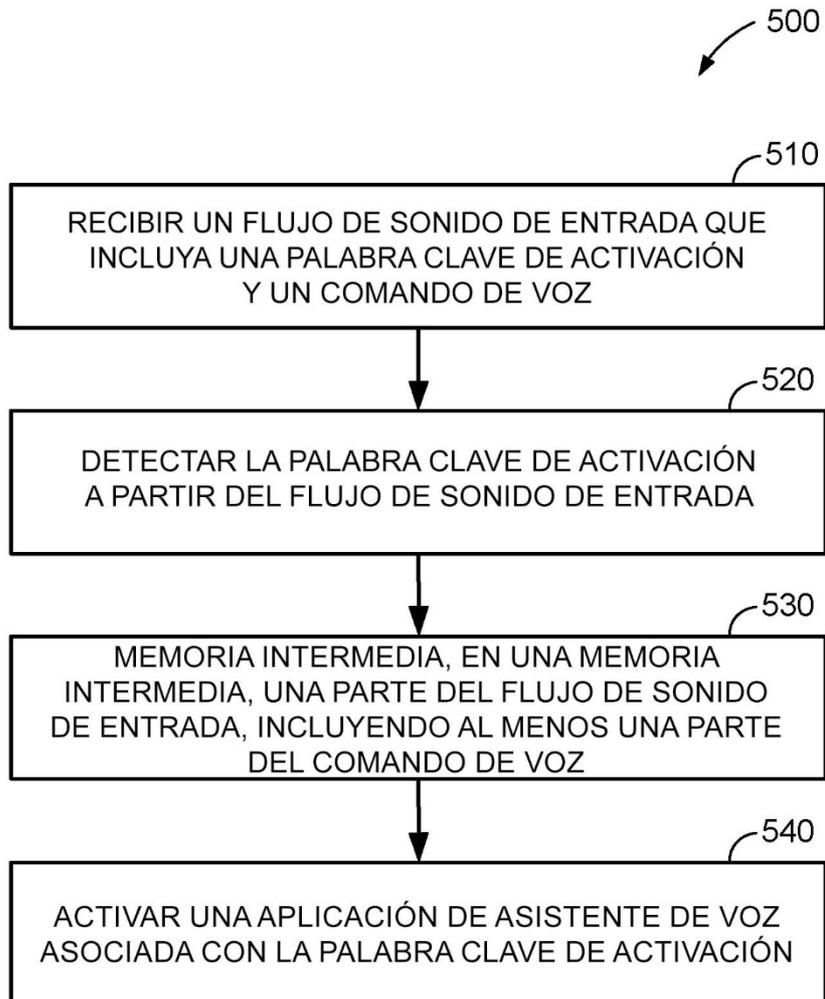


FIG. 5

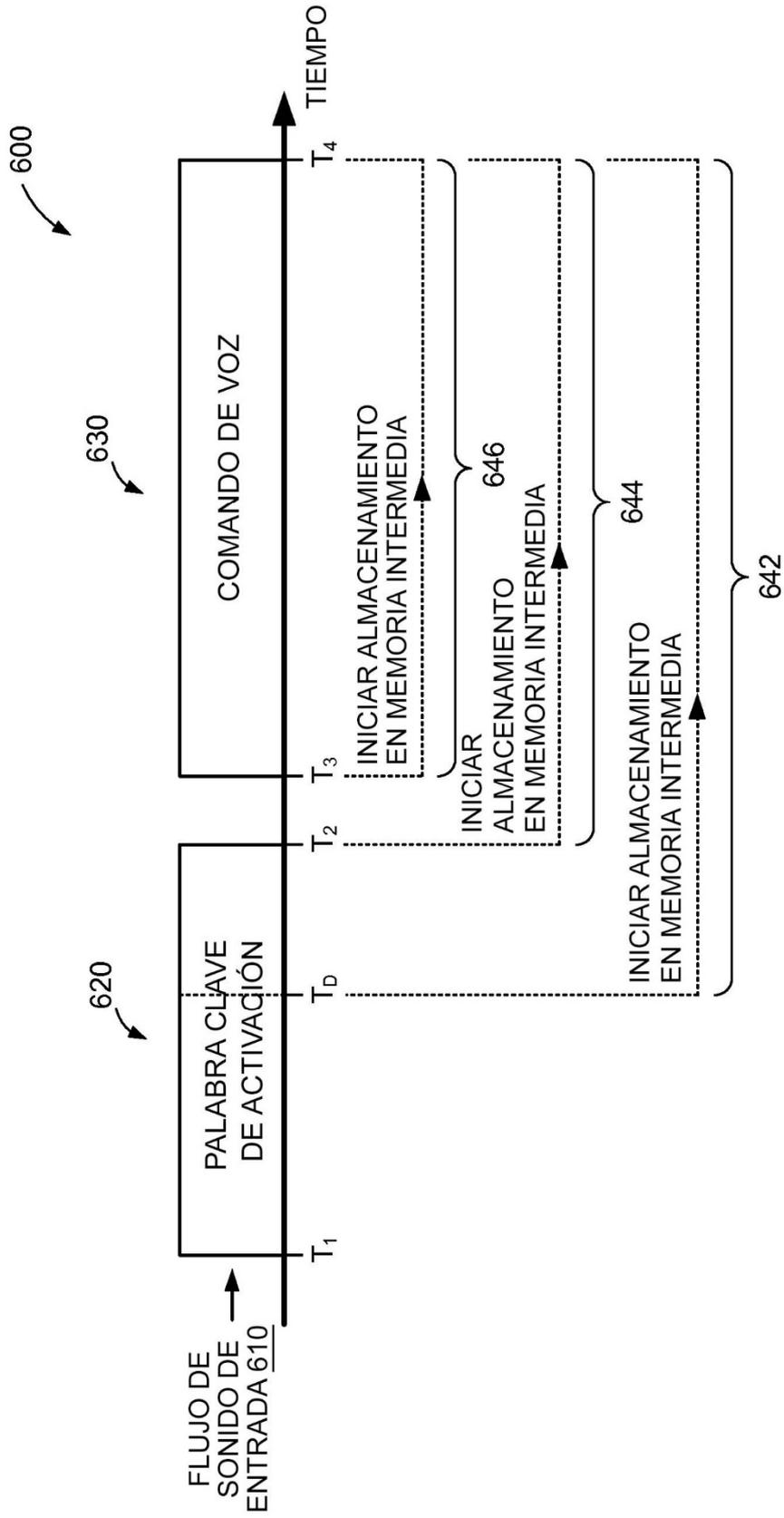


FIG. 6

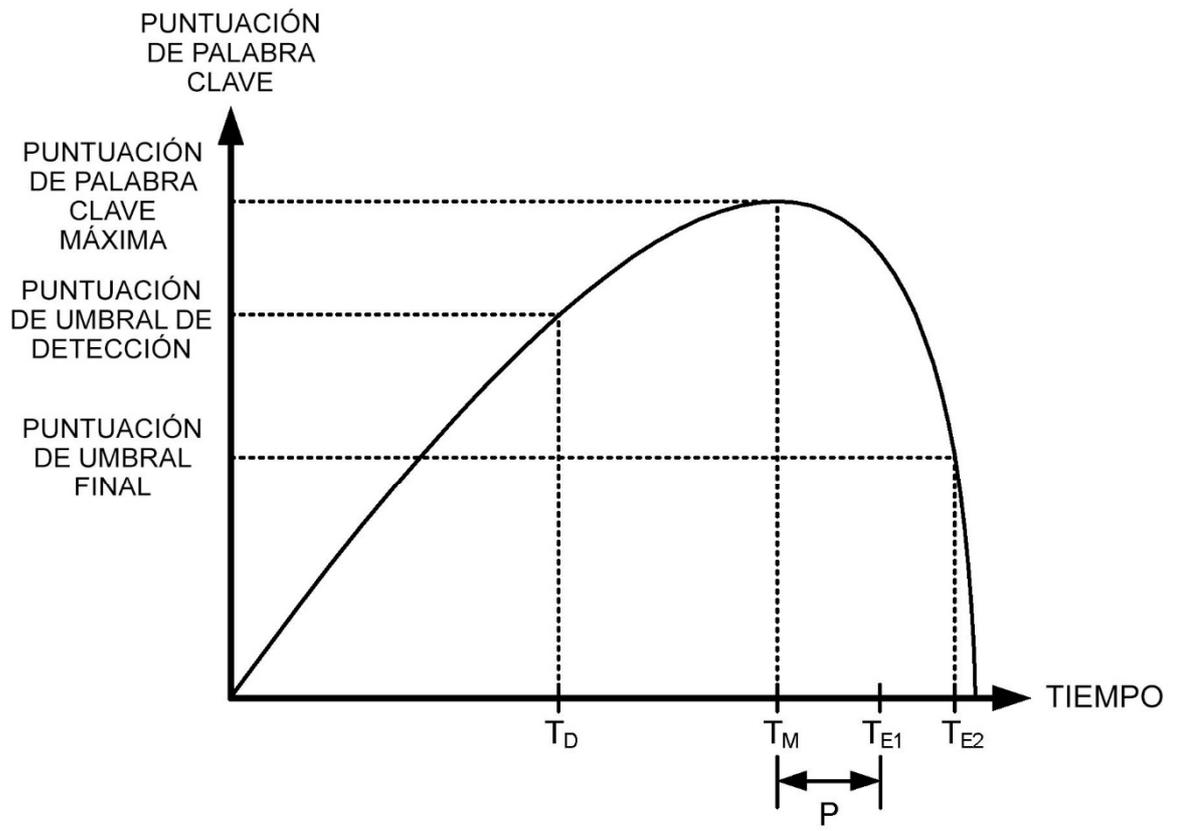


FIG. 7

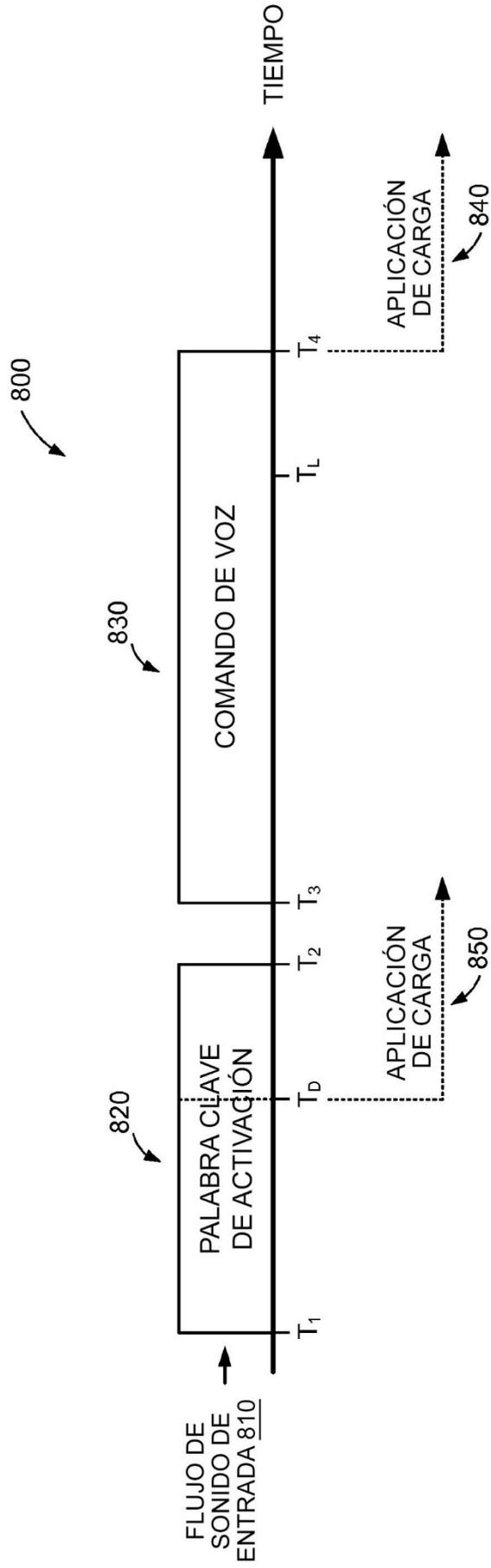


FIG. 8

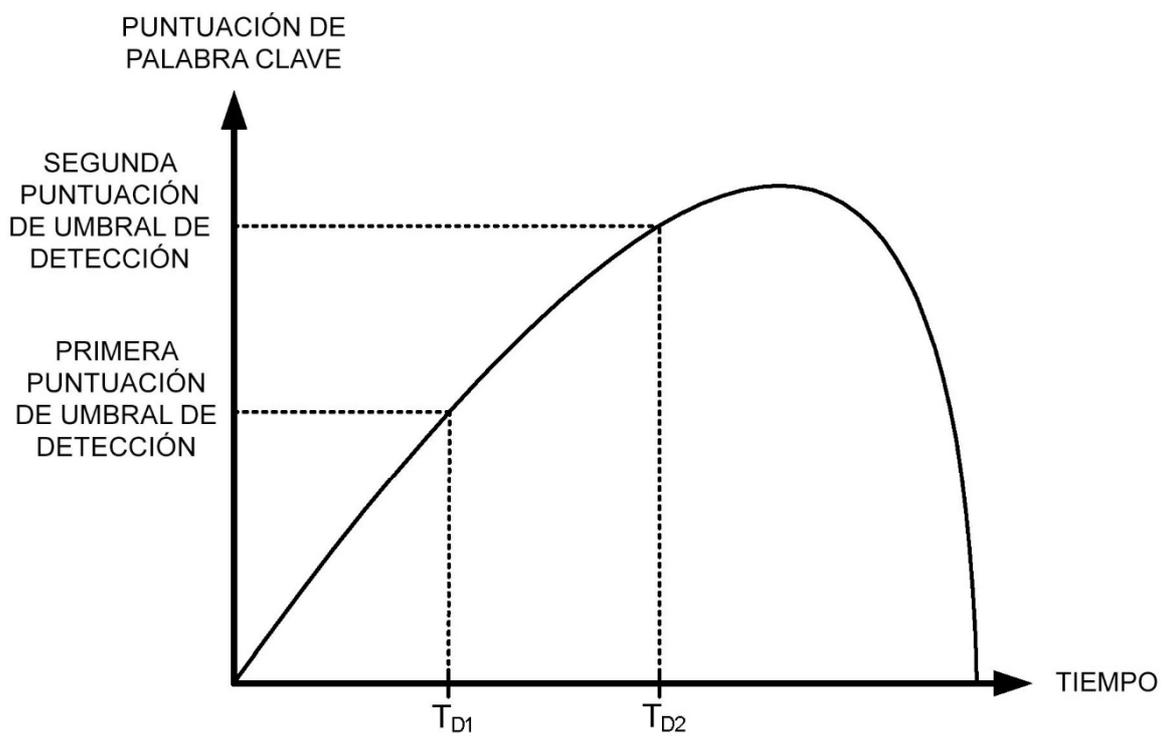


FIG. 9

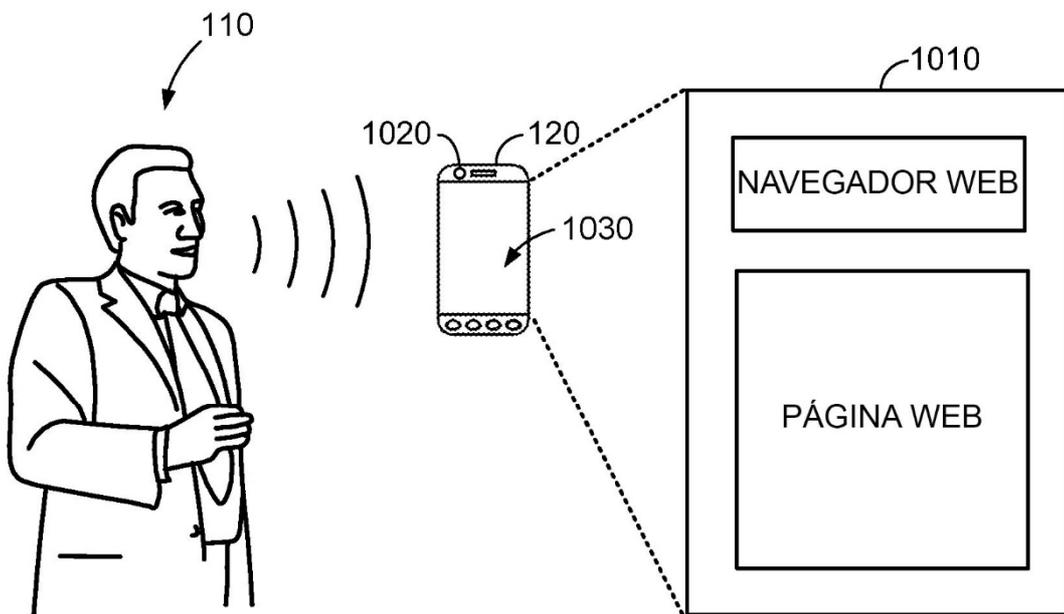


FIG. 10

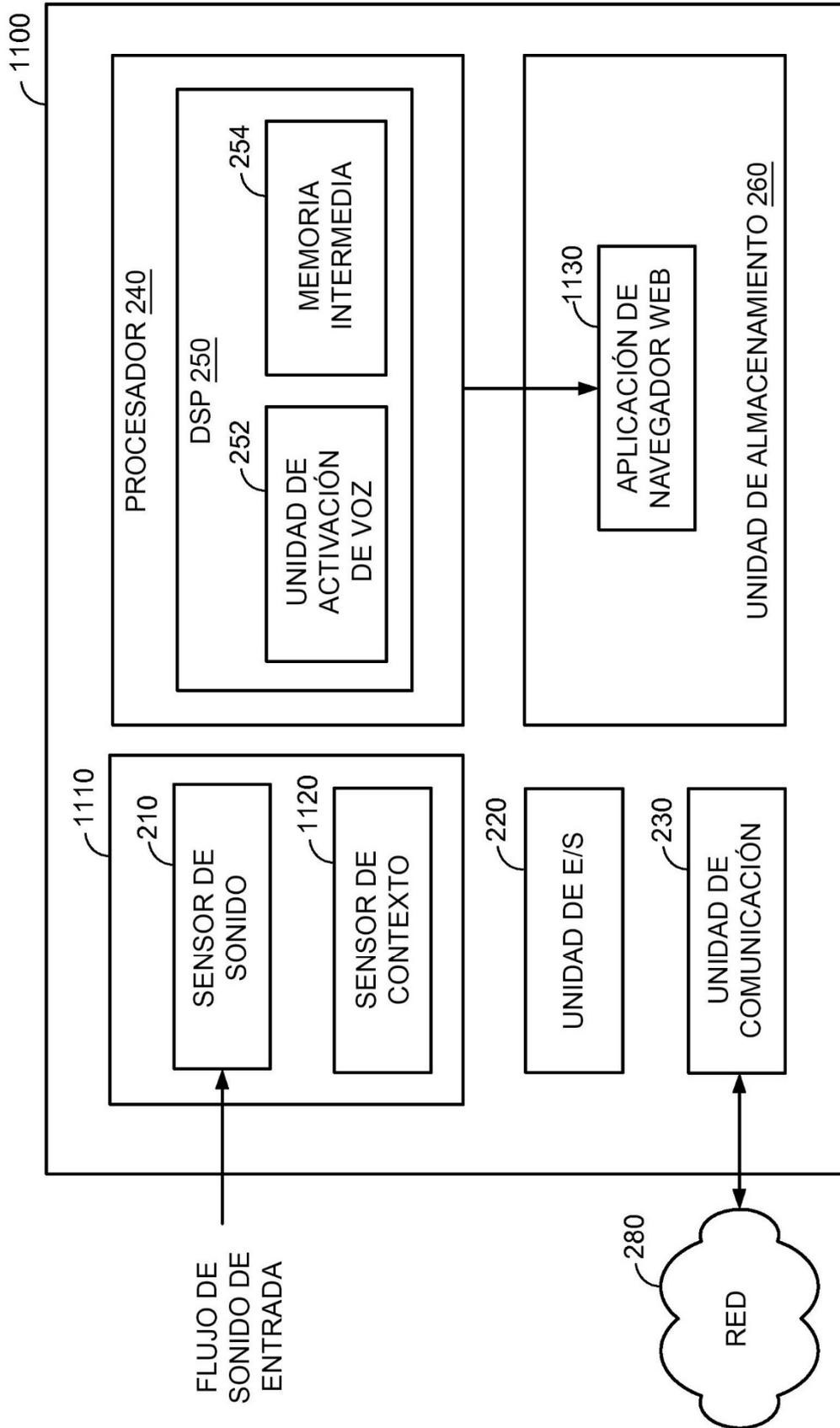


FIG. 11

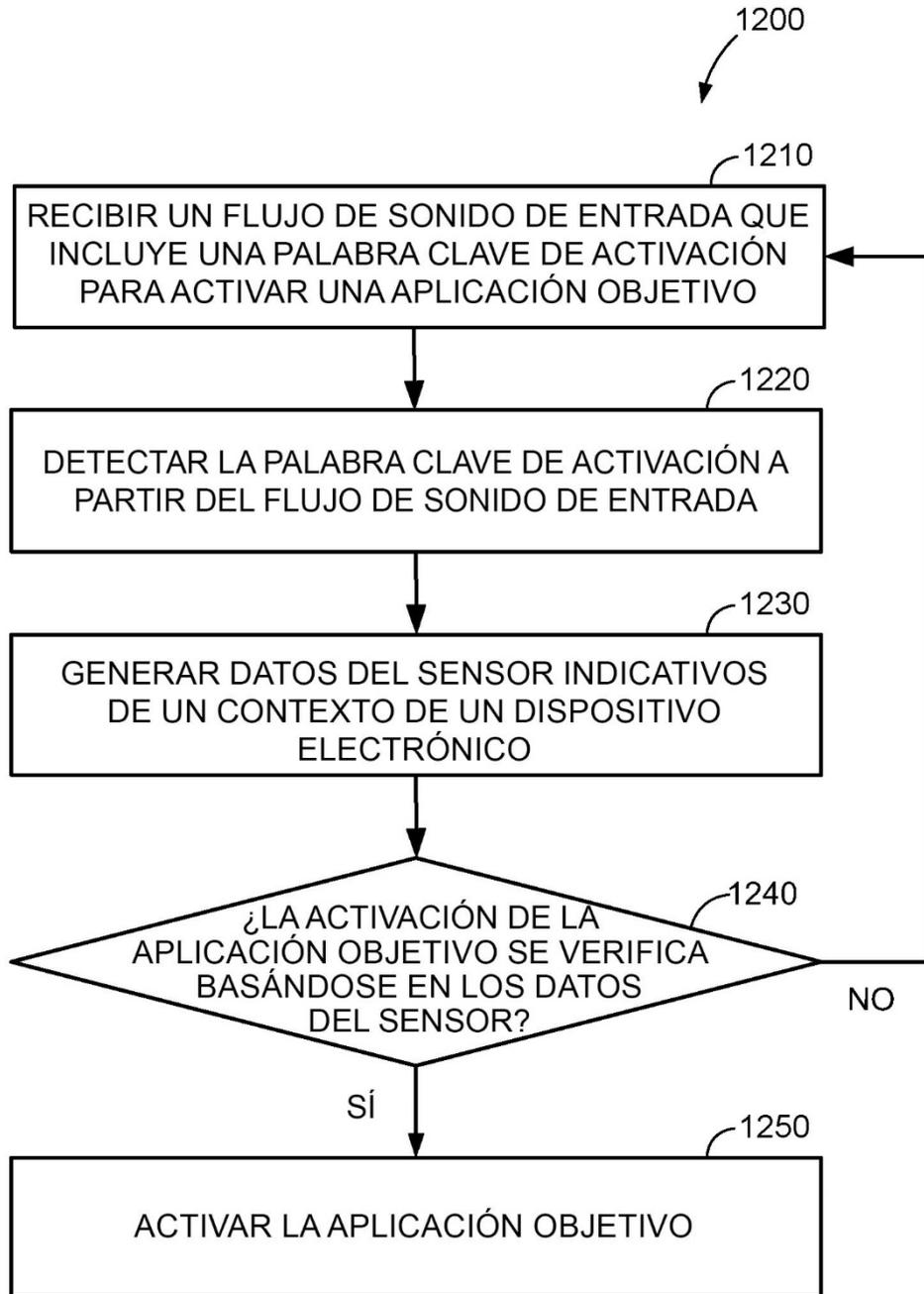


FIG. 12

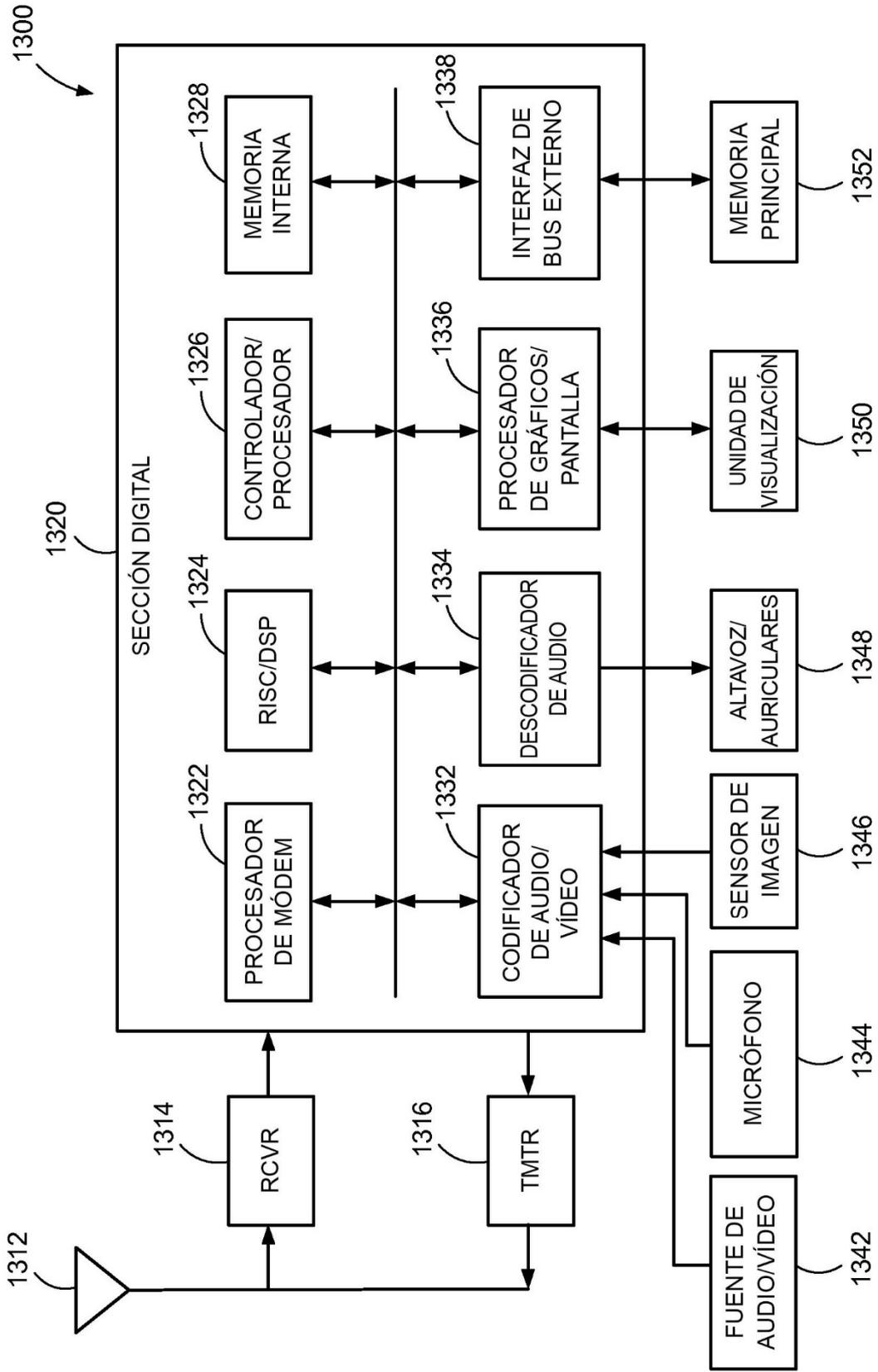


FIG. 13