(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 373 511**

(51) Int. CI.:
G10L 11/02 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Número de solicitud europea: **08833863 .7**
(96) Fecha de presentación: **26.09.2008**
(97) Número de publicación de la solicitud: **2201563**
(97) Fecha de publicación de la solicitud: **30.06.2010**

(54) Título: **DETECTOR DE ACTIVIDAD VOCAL EN MÚLTIPLES MICRÓFONOS.**

(30) Prioridad:
28.09.2007 US 864897

(45) Fecha de publicación de la mención BOPI:
06.02.2012

(45) Fecha de la publicación del folleto de la patente:
06.02.2012

(73) Titular/es:
QUALCOMM Incorporated
Attn: International IP Administration 5775
Morehouse Drive
San Diego, CA 92121, US

(72) Inventor/es:
WANG, Song;
GUPTA, Samir Kumar y
CHOY, Eddie, L. T.

(74) Agente: **Carpintero López, Mario**

DESCRIPCIÓN

Detector de actividad vocal en múltiples micrófonos

Campo de la invención

La revelación versa acerca del campo del procesamiento de audio. En particular, la revelación versa acerca de una detección de actividad vocal utilizando múltiples micrófonos.

Antecedentes

Descripción de la técnica relacionada

Los detectores de actividad de señales, tales como los detectores de actividad vocal, pueden ser utilizados para minimizar la cantidad de procesamiento necesario en un dispositivo electrónico. El detector de actividad vocal puede controlar de forma selectiva una o más etapas de procesamiento de señales posteriores a un micrófono.

Por ejemplo, un dispositivo de grabación puede implementar un detector de actividad vocal para minimizar un procesamiento y una grabación de señales de ruido. El detector de actividad vocal puede desconectar o desactivar de otra manera un procesamiento y una grabación de señales durante periodos de actividad no vocal. De forma similar, un dispositivo de comunicaciones, tal como un teléfono móvil, una agenda electrónica, o un ordenador portátil, pueden implementar un detector de actividad vocal para reducir la potencia de procesamiento asignado a señales de ruido y para reducir las señales de ruido que son transmitidas o comunicadas de otra manera a un dispositivo de destino remoto. El detector de actividad vocal puede desconectar o desactivar el procesamiento y la transmisión de voz durante periodos de actividad no vocal.

La capacidad del detector de actividad vocal para operar de forma satisfactoria puede ser dificultada por condiciones variables de ruido y condiciones de ruido que tienen una energía significativa de ruido. El rendimiento de un detector de actividad vocal puede complicarse adicionalmente cuando la detección de actividad vocal está integrada en un dispositivo móvil, que está expuesto a un entorno dinámico de ruido. Un dispositivo móvil puede operar en entornos relativamente libres de ruido o puede operar en condiciones considerables de ruido, en las que la energía de ruido es del orden de la energía vocal.

La presencia de un entorno dinámico de ruido complica la decisión de actividad vocal. La indicación errónea de actividad vocal puede tener como resultado el procesamiento y la transmisión de señales de ruido. El procesamiento y la transmisión de señales de ruido pueden crear una experiencia deficiente para el usuario, en particular cuando hay intercalados periodos de transmisión de ruido con periodos de inactividad debido a una indicación de una ausencia de actividad vocal por medio del detector de actividad vocal.

Por el contado, una mala detección de actividad vocal puede tener como resultado la pérdida de porciones considerables de señales vocales. La pérdida de porciones iniciales de actividad vocal puede tener como resultado que un usuario necesite repetir a menudo porciones de una conversación, lo que es una condición no deseable.

Los algoritmos tradicionales de Detección de actividad de voz (VAD) solo utilizan una señal de micrófono. Los primeros algoritmos de VAD utilizan criterios basados en energía. Este tipo de algoritmo estima un umbral para tomar la decisión acerca de la actividad vocal. Una VAD en un único micrófono puede funcionar bien para ruido estacionario. Sin embargo, una VAD en un único micrófono tiene algo de dificultad para enfrentarse a ruido no estacionario.

Otra técnica de VAD cuenta el paso de señales por el cero y toma una decisión de actividad vocal en base a la tasa en el paso por el cero. Este procedimiento puede funcionar bien cuando el ruido de fondo son señales no vocales. Cuando la señal de fondo es una señal similar a la frecuencia vocal, este procedimiento no logra tomar una decisión fiable. Se pueden utilizar otras características, tales como el tono, la forma formante, el cepstrum y la periodicidad para la detección de actividad de voz. Estas características son detectadas y comparadas con la señal de frecuencia vocal para tomar una decisión de actividad de voz.

En vez de utilizar características de frecuencia vocal, también pueden utilizarse modelos estadísticos de presencia de frecuencia vocal y de ausencia de frecuencia vocal para tomar una decisión de actividad de voz. En tales implementaciones, se actualizan los modelos estadísticos y se toma una decisión de actividad de voz en base a la relación de probabilidad de los modelos estadísticos. Otro procedimiento utiliza una red de separación de la fuente de un único micrófono para preprocesar la señal. Se toma la decisión utilizando una señal de error filtrada de redes neurales de programación de Lagrange y un umbral adaptado a la actividad.

También han sido estudiados los algoritmos de VAD basados en múltiples micrófonos. Las realizaciones de múltiples micrófonos pueden combinar la supresión de ruido, la adaptación del umbral y la detección del tono para conseguir una detección robusta. Una realización utiliza un filtrado lineal para maximizar una relación de señal/interferencia (SIR). Entonces, se utiliza un procedimiento basado en un modelo estadístico para detectar la actividad vocal utilizando a señal realzada. Otra realización utiliza un conjunto de micrófono lineal y transformadas

de Fourier para generar una representación de dominio frecuencial del vector de salida del conjunto. Se pueden utilizar las representaciones de dominio frecuencial para estimar una relación de señal/ruido (SNR) y un umbral predeterminado para detectar una actividad de frecuencia vocal. Otra realización más sugiere utilizar la magnitud de coherencia cuadrada (MSC) y un umbral adaptativo para detectar la actividad vocal en un procedimiento de VAD basado en dos sensores. Se proporciona un ejemplo de tal realización en LE BOUQUIN-JEANNES R ET AL: "Study of a voice activity detector and its influence on a noise reduction system", SPEECH COMMUNICATION, ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, PAÍSES BAJOS, Vol. 16, nº 3, 1 de abril de 1995, páginas 245-254. Otra realización más, tal como el documento WO 2005/031703 A1, sugiere utilizar un micrófono para frecuencia vocal y un micrófono para ruido al igual que una medida de la variación de las señales entre los dos micrófonos para detectar la actividad de frecuencia vocal.

Muchos de los algoritmos de detección de actividad de voz requieren mucho cálculo y no son adecuados para aplicaciones móviles, en las que son motivo de preocupación un consumo de energía y la complejidad de cálculo. Sin embargo, las aplicaciones móviles también presentan entornos de detección de actividad de voz que suponen un reto debido en parte al entorno dinámico de ruido y a la naturaleza no estacionaria de las señales de ruido que inciden en un dispositivo móvil.

Breve resumen

La detección de actividad de voz utilizando múltiples micrófonos puede estar basada en una relación entre la energía en cada uno de un micrófono de referencia de frecuencia vocal y un micrófono de referencia de ruido. Se puede determinar el gasto de energía de cada uno del micrófono de referencia de frecuencia vocal y del micrófono de referencia de ruido. Se puede determinar una relación de energía de frecuencia vocal a ruido y se puede comparar con un umbral predeterminado de actividad vocal. En otra realización, se determinan el valor absoluto de la correlación de frecuencia vocal y la autocorrelación y/o el valor absoluto de la autocorrelación de las señales de referencia de ruido y se determina una relación en base a los valores de correlación. Las relaciones que superan el umbral predeterminado pueden indicar la presencia de una señal de frecuencias vocales. Se pueden determinar las energías o correlaciones de frecuencia vocal y de ruido utilizando una media ponderada o en un tamaño discreto de trama.

Los aspectos de la invención incluyen un procedimiento, un aparato y un medio legible por un ordenador como en las reivindicaciones 1, 7 y 14, respectivamente.

Breve descripción de los dibujos

Las características, los objetos, y las ventajas de las realizaciones de la revelación serán evidentes a partir de la descripción detallada definida a continuación cuando es tomada junto con los dibujos, en los que los elementos similares tienen números similares de referencia.

La Figura 1 es un diagrama simplificado de bloques funcionales de un dispositivo de múltiples micrófonos que opera en un entorno de ruido.

La Figura 2 es un diagrama simplificado de bloques funcionales de una realización de un dispositivo móvil con un detector calibrado de actividad de voz en múltiples micrófonos.

La Figura 3 es un diagrama simplificado de bloques funcionales de una realización de un dispositivo móvil con un detector de actividad de voz y una cancelación de eco.

La Figura 4A es un diagrama simplificado de bloques funcionales de una realización de un dispositivo móvil con un detector de actividad de voz con un realce de señales.

La Figura 4B es un diagrama simplificado de bloques funcionales de un realce de señales que utiliza una formación de haces.

La Figura 5 es un diagrama simplificado de bloques funcionales de una realización de un dispositivo móvil con un detector de actividad de voz con un realce de señales.

La Figura 6 es un diagrama simplificado de bloques funcionales de una realización de un dispositivo móvil con un detector de actividad de voz con una codificación de frecuencia vocal.

La Figura 7 es un diagrama de flujo de un procedimiento simplificado de detección de actividad de voz.

La Figura 8 es un diagrama simplificado de bloques funcionales de una realización de un dispositivo móvil con un detector calibrado de actividad de voz en múltiples micrófonos.

Descripción detallada de realizaciones de la invención

Se dan a conocer un aparato y procedimientos para la Detección de actividad de voz (VAD) utilizando múltiples micrófonos. El aparato y los procedimientos utilizan un primer conjunto o grupo de micrófonos configurados sustancialmente en un campo próximo de un punto de referencia boca (PRB), considerándose el PRB la posición de la fuente de las señales. Puede haber configurado un segundo conjunto o grupo de micrófonos sustancialmente en una ubicación reducida de voz. Idealmente, el segundo conjunto de micrófonos está colocado sustancialmente en el mismo entorno de ruido que el primer conjunto de micrófonos, pero no acopla sustancialmente ninguna de las señales de frecuencia vocal. Algunos dispositivos móviles no permiten esta configuración óptima, sino que permiten una configuración en la que la frecuencia vocal recibida en el primer conjunto de micrófonos es constantemente mayor que la frecuencia vocal recibida por el segundo conjunto de micrófonos.

El primer conjunto de micrófonos recibe y convierte una señal de frecuencia vocal que es normalmente de mejor calidad con respecto al segundo conjunto de micrófonos. Como tal, el primer conjunto de micrófonos puede ser considerado micrófonos de referencia de frecuencia vocal y el segundo conjunto de micrófonos puede ser considerado micrófonos de referencia de ruido.

Un módulo de VAD puede determinar inicialmente una característica en base a las señales en cada uno de los micrófonos de referencia de frecuencia vocal y de los micrófonos de referencia de ruido. Se utilizan los valores característicos correspondientes a los micrófonos de referencia de frecuencia vocal y a los micrófonos de referencia de ruido para tomar la decisión de actividad de voz.

Por ejemplo, un módulo de VAD puede estar configurado para calcular, estimar, o determinar de otra manera las energías de cada una de las señales procedentes de los micrófonos de referencia de frecuencia vocal y los micrófonos de referencia de ruido. Las energías pueden ser calculadas en instantes predeterminados de muestra de frecuencia vocal y de ruido o pueden ser calculadas en base a una trama de muestras de frecuencia vocal y de ruido.

En otro ejemplo, el módulo de VAD puede estar configurado para determinar una autocorrelación de las señales en cada uno de los micrófonos de referencia de frecuencia vocal y de los micrófonos de referencia de ruido. Los valores de autocorrelación pueden corresponderse con un instante predeterminado de muestra o pueden ser calculados en un intervalo predeterminado de trama.

El módulo de VAD puede calcular o determinar de otra manera una métrica de actividad en base, al menos en parte, a una relación de los valores característicos. En una realización, el módulo de VAD está configurado para determinar una relación de energía de los micrófonos de referencia de frecuencia vocal con respecto a la energía de los micrófonos de referencia de ruido. El módulo de VAD puede estar configurado para determinar una relación de autocorrelación de los micrófonos de referencia de frecuencia vocal con respecto a la autocorrelación de los micrófonos de referencia de ruido. En otra realización, se utiliza la raíz cuadrada de una de las relaciones descritas anteriormente como la métrica de actividad. La VAD compara la métrica de actividad con un umbral predeterminado para determinar la presencia o ausencia de actividad vocal.

La Figura 1 es un diagrama simplificado de bloques funcionales de un entorno operativo 100 que incluye un dispositivo móvil 110 de múltiples micrófonos que tiene una detección de actividad de voz. Aunque se describen en el contexto de un dispositivo móvil, es evidente que los procedimientos y el aparato de detección de actividad de voz dados a conocer en el presente documento no están limitados a la aplicación en dispositivos móviles, sino que pueden ser implementados en dispositivos estacionarios, dispositivos portátiles, dispositivos móviles, y pueden operar mientras que el dispositivo anfitrión es móvil o estacionario.

El entorno operativo 100 muestra un dispositivo móvil 110 de múltiples micrófonos. El dispositivo de múltiples micrófonos incluye al menos un micrófono 12 de referencia de frecuencia vocal, mostrado aquí en una cara frontal del dispositivo móvil 110, y al menos un micrófono 114 de referencia de ruido, mostrado aquí en un lado del dispositivo móvil 110 frente al micrófono 112 de referencia de frecuencia vocal.

Aunque el dispositivo móvil 110 de la Figura 1, y en general, las realizaciones mostradas en las figuras, muestran un micrófono 112 de referencia de frecuencia vocal y un micrófono 114 de referencia de ruido, el dispositivo móvil 110 puede implementar un grupo de micrófonos de referencia de frecuencia vocal y un grupo de micrófonos de referencia de ruido. Cada uno del grupo de micrófonos de referencia de frecuencia vocal y del grupo de micrófonos de referencia de ruido puede incluir uno o más micrófonos. El grupo de micrófonos de referencia de frecuencia vocal puede incluir un número de micrófonos que es distinto o igual que el número de micrófonos en el grupo de micrófonos de referencia de ruido.

Además, los micrófonos del grupo de micrófonos de referencia de frecuencia vocal son normalmente exclusivos de los micrófonos en el grupo de micrófonos de referencia de ruido, pero esta no es una limitación absoluta, dado que se pueden compartir uno o más micrófonos entre los dos grupos de micrófonos. Sin embargo, la unión del grupo de micrófonos de referencia de frecuencia vocal con el grupo de micrófonos de referencia de ruido incluye al menos dos micrófonos.

Se muestra que el micrófono 112 de referencia de frecuencia vocal está en una superficie del dispositivo móvil 110 que es generalmente frente a la superficie que tiene el micrófono 114 de referencia de ruido. La colocación del micrófono 112 de referencia de frecuencia vocal y del micrófono 114 de referencia de ruido no está limitada a ninguna orientación física. Normalmente, la colocación de los micrófonos está regida por la capacidad para aislar señales de frecuencia vocal del micrófono 114 de referencia de ruido.

En general, los micrófonos de los dos grupos de micrófonos están montados en distintas ubicaciones en el dispositivo móvil 110. Cada micrófono recibe su propia versión de combinación de frecuencia vocal deseada y ruido de fondo. Se puede suponer que la señal frecuencia vocal es de fuentes de campo próximo. El nivel de presión sonora (SPL) en los dos grupos de micrófonos puede ser distinta dependiendo de la ubicación de los micrófonos. Si un micrófono se encuentra más cercano al punto de referencia boca (PRB) o a una fuente 130 de frecuencia vocal, puede recibir un mayor SPL que otro micrófono colocado más lejos del PRB. El micrófono con el mayor SPL es denominado el micrófono 112 de referencia de frecuencia vocal o el micrófono primario, que genera una señal de referencia de frecuencia vocal, denotado como $s_{SP}(n)$. El micrófono que tiene el SPL reducido del PRB de la fuente 130 de frecuencia vocal es denominado el micrófono 114 de referencia de ruido o el micrófono secundario, que genera una señal de referencia de ruido, denotes como $s_{NS}(n)$. Se hace notar que la señal de referencia de frecuencia vocal contiene normalmente ruido de fondo, y la señal de referencia de ruido también puede contener frecuencia vocal deseada.

El dispositivo móvil 110 puede incluir una detección de actividad de voz, como se describe con más detalle a continuación, para determinar la presencia de una señal de frecuencia vocal procedente de la fuente 130 de frecuencia vocal. La operación de la detección de actividad de voz puede complicarse mediante el número y la distribución de las fuentes de ruido que puede haber en el entorno operativo 100.

El ruido incidente sobre el dispositivo móvil 110 puede tener un componente significativo de ruido blanco no correlacionado, pero también puede incluir una o más fuentes de ruido de color, por ejemplo 140-1 a 140-4. Además, el propio teléfono móvil 110 puede generar una interferencia, por ejemplo, en forma de una señal de eco que se acopla desde un transductor 120 de salida a uno del micrófono 112 de referencia de frecuencia vocal y del micrófono 114 de referencia de ruido, o a ambos.

La o las fuentes de ruido de color pueden generar señales de ruido que se originan cada una desde una ubicación distinta y una orientación relativa al dispositivo móvil 110. Cada una de las fuentes primera 140-1 de ruido y segunda 140-2 de ruido puede ser colocada más cerca del micrófono 112 de referencia de frecuencia vocal, o en una vía más directa al mismo, mientras que puede haber colocadas fuentes tercera y cuarta 140-3 y 140-4 de ruido más cerca del micrófono 114 de referencia de ruido, o en una vía más directa al mismo. Además, una o más fuentes de ruido, por ejemplo 140-4, pueden generar una señal de ruido que se refleja de una superficie 150 o que recorre de otra manera múltiples vías hasta el dispositivo móvil 110.

Aunque cada una de las fuentes de ruido puede contribuir una señal significativa a los micrófonos, cada una de las fuentes 140-1 a 140-4 de ruido está colocada normalmente en el campo distante, y por lo tanto, contribuye Niveles de presión sonora (SPL) sustancialmente similares a cada uno del micrófono 112 de referencia de frecuencia vocal y del micrófono 114 de referencia de ruido.

La naturaleza dinámica de la magnitud, de la posición, y de la respuesta de frecuencia asociadas con cada señal de ruido contribuye a la complejidad del procedimiento de detección de actividad de voz. Además, el dispositivo móvil 110 está alimentado por batería normalmente y, por lo tanto, el consumo de energía asociado con la detección de actividad de voz puede ser un motivo de preocupación.

El dispositivo móvil 110 puede llevar a cabo una detección de actividad de voz al procesar cada una de las señales procedentes del micrófono 112 de referencia de frecuencia vocal y el micrófono 114 de referencia de ruido para generar valores característicos correspondientes de frecuencia vocal y de ruido. El dispositivo móvil 110 puede generar una métrica de actividad vocal basado en parte en los valores característicos de frecuencia vocal y de ruido, y puede determinar una actividad vocal al comparar la métrica de actividad vocal con un valor umbral.

La Figura 2 es un diagrama simplificado de bloques funcionales de una realización de un dispositivo móvil 110 con un detector calibrado de actividad de voz en múltiples micrófonos. El dispositivo móvil 110 incluye un micrófono 112 de referencia de frecuencia vocal, que puede ser un grupo de micrófonos, y un micrófono 114 de referencia de ruido, que puede ser un grupo de micrófonos de referencia de ruido.

La salida del micrófono 112 de referencia de frecuencia vocal puede estar acoplada a un primer Convertidor 212 de analógico a digital (ADC). Aunque el dispositivo móvil 110 implementa normalmente un procesamiento analógico de las señales de los micrófonos, tal como filtrado y amplificación, no se muestra el procesamiento analógico de las señales de frecuencia vocal en aras de la claridad y brevedad.

La salida del micrófono 114 de referencia de ruido puede estar acoplada a un segundo ADC 214. Normalmente, el procesamiento analógico de las señales de referencia de ruido puede ser sustancialmente el mismo que el procesamiento analógico llevado a cabo en las señales de referencia de frecuencia vocal para mantener

sustancialmente la misma respuesta espectral. Sin embargo, la respuesta espectral de las porciones de procesamiento analógico no necesita ser la misma, dado que un calibrador 220 puede proporcionar alguna corrección. Además, se pueden implementar algunas de las funciones, o todas ellas, del calibrador 220 en las porciones de procesamiento analógico en vez del procesamiento digital mostrado en la Figura 2.

Cada uno de los ADC primero y segundo 212 y 214 convierte sus señales respectivas en una representación digital. Las salidas digitalizada de los ADC primero y segundo 212 y 214 están acopladas a un calibrador 220 que opera para igualar sustancialmente la respuesta espectral de los recorridos de las señales de frecuencia vocal y de ruido antes de la detección de actividad de voz.

El calibrador 220 incluye un generador 222 de calibración que está configurado para determinar una corrección selectiva de frecuencias y controlar un escalador/filtro 224 colocado en serie con uno del recorrido de señal de frecuencia vocal o del recorrido de la señal de ruido. El generador 222 de calibración puede estar configurado para controlar el escalador/filtro 224 para proporcionar una curva fija de respuesta de calibración, o el generador 222 de calibración puede estar configurado para controlar el escalador/filtro 224 para proporcionar una curva dinámica de respuesta de calibración. El generador 222 de calibración puede controlar el escalador/filtro 224 para proporcionar una curva variable de respuesta de calibración basada en uno o más parámetros operativos. Por ejemplo, el generador 222 de calibración puede incluir un detector (no mostrado) de potencia de la señal, o acceder al mismo de otra manera, y puede variar la respuesta del escalador/filtro 224 en respuesta a la intensidad de la frecuencia vocal o del ruido. Otras realizaciones pueden utilizar otros parámetros o combinación de parámetros.

El calibrador 220 puede estar configurado para determinar la calibración proporcionada por el escalador/filtro 224 durante un periodo de calibración. El dispositivo móvil 110 puede ser calibrado inicialmente, por ejemplo, durante su fabricación, o puede ser calibrado según un plan de calibración que puede iniciar la calibración tras uno o más eventos, tiempos, o una combinación de eventos y tiempos. Por ejemplo, el calibrador 220 puede iniciar una calibración cada vez que se enciende el dispositivo móvil, o durante el encendido solo si ha transcurrido un tiempo predeterminado desde la calibración más reciente.

Durante la calibración, el dispositivo móvil 110 puede estar en una condición en la que se encuentra en la presencia de fuentes de campo distante, y no experimenta señales de campo próximo ni en el micrófono 112 de referencia de frecuencia vocal ni en el micrófono 114 de referencia de ruido. El generador 222 de calibración monitoriza cada una de la señal de frecuencia vocal y la señal de ruido y determina la respuesta espectral relativa. El generador 222 de calibración genera o caracteriza de otra manera una señal de control de calibración que, cuando es aplicada al escalador/filtro 224, hace que el escalador/filtro 224 compense las diferencias relativas en la respuesta espectral.

El escalador/filtro 224 puede introducir amplificación, atenuación, filtrado, o algún otro procesamiento de señales que puede compensar sustancialmente las diferencias espectrales. Se muestra el escalador/filtro 224 colocado en el recorrido de la señal de ruido, lo que puede ser conveniente para evitar que el escalador/filtro distorsione las señales de frecuencia vocal. Sin embargo, porciones del escalador/filtro 224, o todo él, pueden estar colocadas en el recorrido de la señal de frecuencia vocal, y pueden estar distribuidas a través de los recorridos de las señales analógica y digital de uno del recorrido de señal de frecuencia vocal y del recorrido de señal de ruido, o de ambos.

El calibrador 220 acopla las señales calibradas de frecuencia vocal y de ruido a entradas respectivas de un módulo 230 de detección de actividad de voz (VAD). El módulo 230 de VAD incluye un generador 232 de valor característico de frecuencia vocal, un generador 234 de valor característico de ruido, un módulo 240 de métrica de actividad vocal que opera sobre valores característicos de frecuencia vocal y de ruido, y un comparador 250 configurado para determinar la presencia o ausencia de actividad vocal en base a la métrica de actividad vocal. El módulo 230 de VAD puede incluir opcionalmente un generador 236 de valor característico combinado configurado para generar una característica basada en una combinación tanto de la señal de referencia de frecuencia vocal como de la señal de referencia de ruido. Por ejemplo, el generador 236 de valor característico combinado puede estar configurado para determinar una correlación cruzada de las señales de frecuencia vocal y de ruido. Puede tomarse el valor absoluto de la correlación cruzada, o pueden elevarse al cuadrado los componentes de la correlación cruzada.

El generador 232 de valor característico de frecuencia vocal puede estar configurado para generar un valor que está basado al menos en parte en la señal de frecuencia vocal. El generador 232 de valor característico de frecuencia vocal puede estar configurado, por ejemplo, para generar un valor característico tal como una energía de la señal de frecuencia vocal en un instante específico de muestra (E_{SP}(n)), una autocorrelación de la señal de frecuencia vocal en un instante específico de muestra ($\rho_{SP}(n)$), o puede tomarse algún otro valor característico de la señal, como el valor absoluto de la autocorrelación de la señal de frecuencia vocal o los componentes de la autocorrelación.

El generador 234 de valor característico de ruido puede estar configurado para generar un valor característico complementario de ruido. Es decir, el generador 234 de valor característico de ruido puede estar configurado para generar un valor de energía de ruido en un instante específico (E_{NS}(n)) si el generador 232 de valor característico de frecuencia vocal genera un valor de energía de frecuencia vocal. De forma similar, el generador 234 de valor característico de ruido puede estar configurado para generar un valor de autocorrelación de ruido en un instante específico ($\rho_{NS}(n)$) si el generador 232 de valor característico de frecuencia vocal genera un valor de autocorrelación

de frecuencia vocal. También puede ser tomado el valor absoluto del valor de autocorrelación de ruido, o se puede tomar el valor de autocorrelación de ruido.

El módulo 240 de métrica de actividad vocal puede estar configurado para generar una métrica de actividad vocal en base al valor característico de frecuencia vocal, al valor característico de ruido, y opcionalmente, al valor de correlación cruzada. El módulo 240 de métrica de actividad vocal puede estar configurado, por ejemplo, para generar una métrica de actividad vocal que no es complejo de calcular. Por lo tanto, el módulo 230 de VAD puede generar una señal de detección de actividad de voz sustancialmente en tiempo real, y utilizando relativamente pocos recursos de procesamiento. En una realización, el módulo 240 de métrica de actividad vocal está configurado para determinar una relación de uno o más de los valores característicos o de una relación de uno o más de los valores característicos y el valor de correlación cruzada o una relación de uno o más de los valores característicos y el valor absoluto del valor de correlación cruzada.

El módulo 240 de métrica de actividad vocal acopla la métrica a un comparador 250 que puede estar configurado para determinar la presencia de actividad de frecuencia vocal al comparar la métrica de actividad vocal con uno o más umbrales. Cada uno de los umbrales puede ser un umbral fijo predeterminado, o uno o más de los umbrales pueden ser un umbral dinámico.

En una realización, el módulo 230 de VAD determina tres correlaciones distintas para determinar la actividad de frecuencia vocal. El generador 2323 de valor característico de frecuencia vocal genera una autocorrelación de la señal $\rho_{SP}(n)$ de referencia de frecuencia vocal, el generador 234 de valor característico de ruido genera una autocorrelación de la señal $\rho_{NS}(n)$ de referencia de ruido y el módulo 236 de correlación cruzada genera la correlación cruzada de valores absolutos de la señal de referencia de frecuencia vocal y de la señal $\rho_c(n)$ de referencia de ruido. Aquí, n representa un índice de tiempos. Para evitar un retraso excesivo, las correlaciones pueden ser calculadas aproximadamente utilizando un procedimiento de ventana exponencial utilizando las siguientes ecuaciones. Para una autocorrelación, la ecuación es:

$$\rho(n) = \alpha\rho(n-1) + s(n)^2 \quad o \quad \rho(n) = \alpha\rho(n-1) + (1-\alpha)s(n)^2.$$

Para la correlación cruzada, la ecuación es:

$$\rho_c(n) = \alpha\rho_c(n-1) + \left|s_{SP}(n)s_{NS}(n)\right| \quad o \quad \rho_c(n) = \alpha\rho_c(n-1) + (1-\alpha)\left|s_{SP}(n)s_{NS}(n)\right|.$$

En las anteriores ecuaciones, $\rho(n)$ es la correlación en el instante n. $s(n)$ es una de las señales de frecuencia vocal o de ruido en el instante n. α es una constante entre 0 y 1. $|\bullet|$ representa el valor absoluto. La correlación también puede ser calculada utilizando una ventana cuadrada con un tamaño N de ventana como sigue:

$$\rho(n) = \rho(n-1) + s(n)^2 - s(n-N)^2$$

o

$$\rho_c(n) = p_c(n-1) + \left|s_{SP}(n)s_{NS}(n)\right| - \left|s_{SP}(n-N)s_{NS}(n-N)\right|.$$

La decisión de VAD puede ser tomada en base a $\rho_{SP}(n)$, $\rho_{NS}(n)$ y $\rho_c(n)$. En general,

$$D(n) = vad(\rho_{SP}(n), \rho_{NS}(n), \rho_c(n)).$$

En los siguientes ejemplos, se describen dos categorías de la decisión de VAD. Una es un procedimiento de decisión de VAD en base a muestras. La otra es un procedimiento de decisión de VAD en base a tramas. En general, los procedimientos de decisión de VAD que están basados en el uso del valor absoluto de la autocorrelación o la correlación cruzada pueden permitir un intervalo dinámico menor de la correlación cruzada o de la autocorrelación. La reducción del intervalo dinámico puede permitir transiciones más estables en los procedimientos de decisión de VAD.

Decisión de VAD basada en muestras

El módulo de VAD puede tomar una decisión de VAD para cada par de muestras de frecuencia vocal y de ruido en el instante n en base a las correlaciones calculadas en el instante n. Como ejemplo, el módulo de métrica de actividad vocal puede estar configurado para determinar la métrica de actividad vocal en base a una relación entre los tres valores de correlación.

$$R(n) = f(\rho_{SP}(n), \rho_{NS}(n), \rho_C(n)).$$

Se puede determinar la cantidad *T(n)* en base a $\rho_{SP}(n)$, $\rho_{NS}(n)$ y *R(n)*, por ejemplo

$$T(n) = g(\rho_{SP}(n), \rho_{NS}(n), \rho_C(n), R(n)).$$

El comparador puede tomar la decisión de VAD en base a *R(n)* y *T(n)*, por ejemplo

$$D(n) = vad(R(n), T(n)).$$

Como ejemplo específico, la métrica *R(n)* de actividad vocal puede ser definida para que sea la relación entre el valor $\rho_{SP}(n)$ de autocorrelación de frecuencia vocal del generador 232 de valor característico de frecuencia vocal y la correlación cruzada $\rho_C(n)$ del módulo 236 de correlación cruzada. En el instante *n*, la métrica de actividad vocal puede ser la relación definida como:

$$R(n) = \frac{\rho_{SP}(n)}{\rho_C(n) + \delta},$$

En el anterior ejemplo de la métrica de actividad vocal, el módulo 240 de métrica de actividad vocal limita el valor. El módulo 240 de métrica de actividad vocal limita el valor al limitar el denominador a no menos de δ, siendo δ un número positivo pequeño para evitar la división por cero. Como otro ejemplo, *R(n)* puede ser definido ser entre $\rho_C(n)$ y $\rho_{NS}(n)$, por ejemplo

$$R(n) = \frac{\rho_C(n)}{\rho_{NS}(n) + \delta}.$$

Como ejemplo específico, la cantidad *T(n)* puede ser un umbral fijo. Sea $R_{SP}(n)$ la relación mínima cuando hay presente frecuencia vocal deseada hasta el instante *n*. Sea $R_{NS}(n)$ la relación máxima cuando la frecuencia vocal deseada está ausente hasta el instante *n*. El umbral *T(n)* puede ser determinado o seleccionado de otra manera para ser entre $R_{NS}(n)$ y $R_{SP}(n)$, o de forma equivalente:

$$R_{NS}(n) \leq Th(n) \leq R_{SP}(n).$$

El umbral también puede ser variable y puede variar en base al menos en parte al cambio de frecuencia vocal deseada y ruido de fondo. En tal caso, se pueden determinar $R_{SP}(n)$ y $R_{NS}(n)$ en base a las señales más recientes de micrófonos.

El comparador 250 compara el umbral con la métrica de actividad vocal, aquí la relación *R(n)*, para tomar una decisión acerca de una actividad vocal. En este ejemplo específico, la función de adopción de la decisión vad(•,•) puede estar definida como sigue

$$vad(R(n), T(n)) = \begin{cases} Activo & R(n) > T(n) \\ Inactivo & si\ no \end{cases}.$$

Decisión de VAD basada en tramas

También se puede tomar la decisión de VAD de forma que una trama completa de muestras genere y comparta una decisión de VAD. La trama de muestras puede ser generada o recibida de otra manera entre el instante *m* y el instante *m + M - 1*, en la que M representa el tamaño de la trama.

Como ejemplo, el generador 232 de valor característico de frecuencia vocal, el generador 234 de valor característico de ruido y el generador 236 de valor característico combinado pueden determinar las correlaciones para una trama completa de datos. En comparación con las correlaciones calculadas utilizando una ventana cuadrada, la correlación de la trama es equivalente a la correlación calculada en el instante *m + M - 1*, por ejemplo ρ(*m + M - 1*).

La decisión de VAD puede ser tomada en base a los valores de energía o de autocorrelación de las dos señales de micrófono. De forma similar, el módulo 240 de métrica de actividad vocal puede determinar la métrica de actividad en base a una relación *R(n)* como se ha descrito anteriormente en la realización basada en muestras. El comparador puede basar la decisión de actividad de voz en base a un umbral *T(n)*.

VAD basada en señales después de un realce de señales

Cuando la SNR de la señal de referencia de frecuencia vocal es baja, la decisión de VAD tiende a ser agresiva. Las partes de comienzo y de final de la frecuencia vocal pueden estar clasificadas como segmentos que no son de

frecuencia vocal. Si los niveles de señal del micrófono de referencia de frecuencia vocal y del micrófono de referencia de ruido son similares cuando hay presente la señal de frecuencia vocal deseada, el aparato y los procedimientos de VAD descritos anteriormente pueden no proporcionar una decisión fiable de VAD. En tales casos, se puede aplicar un realce adicional de señales a una o más de las señales de los micrófonos para ayudar a la VAD a tomar una decisión fiable.

Se puede implementar el realce de señales para reducir la cantidad de ruido de fondo en la señal de referencia de frecuencia vocal sin cambiar la señal de frecuencia vocal deseada. También se puede implementar el realce de señales para reducir el nivel o la cantidad de frecuencia vocal en la señal de referencia de ruido sin cambiar el ruido de fondo. En algunas realizaciones, el realce de señales puede llevar a cabo una combinación de realce de referencia de frecuencia vocal y de realce de referencia de ruido.

La Figura 3 es un diagrama simplificado de bloques funcionales de una realización de dispositivo móvil 110 con un detector de actividad vocal y una cancelación de eco. Se muestra el dispositivo móvil 110 sin el calibrador mostrado en la Figura 2, pero la implementación de la cancelación de eco en el dispositivo móvil 110 no es exclusiva de la calibración. Además, el dispositivo móvil 110 implementa la cancelación de eco en el dominio digital, pero parte de la cancelación de eco, o toda ella, puede ser llevada a cabo en el dominio analógico.

La porción de procesamiento de voz del dispositivo móvil 110 puede ser sustancialmente similar a la porción ilustrada en la Figura 2. Un micrófono 112 o un grupo de micrófonos de referencia de frecuencia vocal recibe una señal de frecuencia vocal y convierte el SPL de la señal de audio en una señal eléctrica de referencia de frecuencia vocal. El primer ADC 212 convierte la señal analógica de referencia de frecuencia vocal en una representación digital. El primer ADC 212 acopla la señal digitalizada de referencia de frecuencia vocal a una primera entrada de un primer combinador 352.

De forma similar, un micrófono 114 o grupo de micrófonos de referencia de ruido recibe las señales de ruido y genera una señal de referencia de ruido. El segundo ADC 214 convierte la señal analógica de referencia de ruido en una representación digital. El segundo ADC 214 acopla la señal digitalizada de referencia de ruido a una primera entrada de un segundo combinador 354.

Los combinadores primero y segundo 352 y 354 pueden ser parte de una porción de cancelación de eco del dispositivo móvil 110. Los combinadores primero y segundo 352 y 354 pueden ser, por ejemplo, sumadores de señales, restadores de señales, acopladores, moduladores, y similares, o algún otro dispositivo configurado para combinar señales.

El dispositivo móvil 110 puede implementar la cancelación de eco para eliminar de forma eficaz la señal de eco atribuible a la salida de audio del dispositivo móvil 110. El dispositivo móvil 110 incluye un convertidor digital a analógico (DAC) 310 de salida que recibe una señal digitalizada de salida de audio procedente una fuente (no mostrada) de señales tal como un procesador de banda base y convierte la señal digitalizada de audio en una representación analógica. La salida del DAC 310 puede estar acoplada a un transductor de salida, tal como un altavoz 320. El altavoz 320, que puede ser un receptor o un altavoz, puede estar configurado para convertir la señal analógica en una señal de audio. El dispositivo móvil 110 puede implementar una o más etapas de procesamiento de audio entre el DAC 310 y el altavoz 320. Sin embargo, las etapas de procesamiento de señales de salida no están ilustradas en aras de la brevedad.

La señal de salida digital también puede estar acoplada a entradas de un primer cancelador 342 de ecos y a un segundo cancelador 344 de ecos. El primer cancelador 342 de ecos puede estar configurado para generar una señal de cancelación de ecos que se aplica a la señal de referencia de frecuencia vocal, mientras que el segundo cancelador 344 de ecos puede estar configurado para generar una señal de cancelación de ecos que se aplica a la señal de referencia de ruido.

La salida del primer cancelador 342 de ecos puede estar acoplada a una segunda entrada del primer combinador 342. La salida del segundo cancelador 344 de ecos puede estar acoplada a una segunda entrada del segundo combinador 344. Los combinadores 352 y 354 acoplan las señales combinadas al módulo 230 de VAD. El módulo 230 de VAD puede estar configurado para operar de una forma descrita en relación con la Figura 2.

Cada uno de los canceladores 342 y 344 de ecos puede estar configurado para generar una señal de cancelación de ecos que reduce o elimina sustancialmente la señal de eco en las líneas respectivas de señales. Cada cancelador 342 y 344 de ecos puede incluir una entrada que muestrea o monitoriza de otra manera la señal de eco cancelado en la salida de los combinadores respectivos 352 y 354. La salida de los combinadores 352 y 354 opera como una señal de realimentación de errores que puede ser utilizada por los canceladores respectivos 342 y 344 de ecos para minimizar el eco residual.

Cada cancelador 342 y 344 de ecos puede incluir, por ejemplo, amplificadores, atenuadores, filtros, módulos de retraso, o alguna combinación de los mismos para generar la señal de cancelación de eco. La correlación alta entre la señal de salida y la señal de eco puede permitir que los canceladores 342 y 344 de ecos detecten y compensen más fácilmente la señal de eco.

En otras realizaciones, puede ser deseable un realce adicional de las señales porque no se cumpla la suposición de que los micrófonos de referencia de frecuencia vocal están colocados más cerca del punto de referencia boca. Por ejemplo, los dos micrófonos pueden estar colocados tan cerca entre sí que la diferencia entre las dos señales de los micrófonos es muy pequeña. En este caso, las señales no realzadas pueden no producir una decisión fiable de VAD. En este caso, se puede utilizar un realce de señales para ayudar a mejorar la decisión de VAD.

La Figura 4 es un diagrama simplificado de bloques funcionales de una realización del dispositivo móvil 110 con un detector de actividad vocal con un realce de señales. Como antes, se pueden implementar una o ambas técnicas y aparato de calibración y de cancelación de ecos descritos anteriormente en relación con las Figuras 2 y 3, además de un realce de las señales.

El dispositivo móvil 110 incluye un micrófono 112 o un grupo de micrófonos de referencia de frecuencia vocal configurado para recibir una señal de frecuencia vocal y convertir el SPL de la señal de audio en una señal eléctrica de referencia de frecuencia vocal. El primer ADC 212 convierte la señal analógica de referencia de frecuencia vocal en una representación digital. El primer ADC 212 acopla la señal digitalizada de referencia de frecuencia vocal a una primera entrada de un módulo 400 de realce de señales.

De forma similar, un micrófono 114 o grupo de micrófonos de referencia de ruido recibe las señales de ruido y genera una señal de referencia de ruido. El segundo ADC 214 convierte la señal analógica de referencia de ruido en una representación digital. El segundo ADC 214 acopla la señal digitalizada de referencia de ruido en una segunda entrada del módulo 400 de realce de señales.

El módulo 400 de realce de señales puede estar configurado para generar una señal realzada de referencia de frecuencia vocal y una señal realzada de referencia de ruido. El módulo 400 de realce de señales acopla las señales realzadas de referencia de frecuencia vocal y de ruido a un módulo 230 de VAD. El módulo 230 de VAD opera en las señales realzadas de referencia de frecuencia vocal y de ruido para tomar la decisión de actividad de voz.

VAD basada en señales después de la formación de haces o la separación de señales

El módulo 400 de realce de señales puede estar configurado para implementar la formación adaptativa de haces para producir una directividad de los sensores. El módulo 400 de realce de señales implementa la formación adaptativa de haces utilizando un conjunto de filtros y tratando los micrófonos como un conjunto de sensores. Esta directividad de los sensores puede ser utilizado para extraer una señal deseada cuando hay presentes múltiples fuentes de señales. Hay disponibles muchos algoritmos de formación de haces para conseguir una directividad de los sensores. Una instanciación de un algoritmo de formación de haces o de una combinación de algoritmos de formación de haces es denominada formador de haces. En comunicaciones de frecuencia vocal de dos micrófonos, el formador de haces puede ser utilizado para dirigir la dirección del sensor al punto de referencia boca para generar una señal realzada de referencia de frecuencia vocal en la que puede estar reducido el ruido de fondo. También puede generar una señal realzada de referencia de ruido en la que puede estar reducido la frecuencia vocal deseada.

La Figura 4B es un diagrama simplificado de bloques funcionales de una realización de un módulo 400 de realce de señales de formación de haces los micrófonos 112 y 114 de referencia de frecuencia vocal y de ruido.

El módulo 400 de realce de señales incluye un conjunto de micrófonos 112-1 a 112-n de referencia de frecuencia vocal que comprende un primer conjunto de micrófonos. Cada uno de los micrófonos 112-1 a 112-n de referencia de frecuencia vocal puede acoplar su salida a un filtro correspondiente 412-1 a 412-n. Cada uno de los filtros 412-1 a 412-n proporciona una respuesta que puede ser controlada por el primer controlador 420-1 de formación de haces. Cada filtro, por ejemplo 412-1, puede estar controlado para proporcionar un retraso variable, una respuesta espectral, una ganancia o algún otro parámetro.

El primer controlador 420-1 de formación de haces puede estar configurado con un conjunto predeterminado de señales de control de los filtros, correspondientes a un conjunto predeterminado de haces, o puede estar configurado para variar las respuestas de los filtros según un algoritmo predeterminado para orientar de forma eficaz el haz de forma continua.

Cada uno de los filtros 412-1 a 412 da salida a su señal filtrada a una entrada correspondiente de un primer combinador 430-1. La salida del primer combinador 430-1 puede ser una señal formada en haz de referencia de frecuencia vocal.

La señal de referencia de ruido puede estar formada en haz, de forma similar, utilizando un conjunto de micrófonos 114-1 a 114-k de referencia de ruido que comprende un segundo conjunto de micrófonos. El número, k, de micrófonos de referencia de ruido puede ser distinto del número, n, de micrófonos de referencia de frecuencia vocal, o puede ser el mismo.

Aunque el dispositivo móvil 110 de la Figura 4B ilustra micrófonos 112-1 a 112-n de referencia de frecuencia vocal y micrófonos 114-1 a 114-k de referencia de ruido distintos, en otras realizaciones, se pueden utilizar algunos de los

micrófonos 112-1 a 112-n de referencia de frecuencia vocal, o todos ellos, como los micrófonos 114-1 a 114-k de referencia de ruido. Por ejemplo, el conjunto de micrófonos 112-1 a 112-n de referencia de frecuencia vocal pueden ser los mismos micrófonos utilizados para el conjunto de micrófonos 114-1 a 114-k de referencia de ruido.

Cada uno de los micrófonos 114-1 a 114-k de referencia de ruido acopla su salida a un filtro correspondiente 414-1 a 414-k. Cada uno de los filtros 414-1 a 414-k proporciona una respuesta que puede estar controlada por el segundo controlador 420-2 de formación de haces. Cada filtro, por ejemplo 414-1, puede estar controlado para proporcionar un retraso variables, una respuesta espectral, una ganancia, o algún otro parámetro. El segundo controlador 420-2 de formación de haces puede controlar los filtros 414-1 a 414-k para proporcionar un número discreto predeterminado de configuraciones de haces, o puede estar configurado para orientar el haz de forma sustancialmente continua.

En el módulo 400 de realce de señales de la Figura 4B, se utilizan controladores distintos 420-1 y 420-2 de formación de haces para formar haces de forma independiente con las señales de referencia de frecuencia vocal y de ruido. Sin embargo, en otras realizaciones, se puede utilizar un único controlador de formación de haces para formar haces tanto con las señales de referencia de frecuencia vocal como con las señales de referencia de ruido.

El módulo 400 de realce de señales puede implementar una separación ciega de fuentes. La separación ciega de fuentes (BSS) es un procedimiento para restaurar señales de fuentes independientes utilizando mediciones de mezclas de estas señales. Aquí, el término “ciego” tiene un doble significado. El primero, que no son conocidas las señales originales ni las fuentes de las señales. El segundo, que puede no ser conocido el procedimiento de mezclado. Existen muchos algoritmos disponibles para conseguir la separación de señales. En comunicaciones de frecuencia vocal de dos micrófonos, se puede utilizar la BSS para separar la frecuencia vocal y el ruido de fondo. Después de la separación de señales, el ruido de fondo en la señal de referencia de frecuencia vocal puede estar algo reducido y la frecuencia vocal en la señal de referencia de ruido puede estar algo reducida.

El módulo 400 de realce de señales puede, por ejemplo, implementar uno de los procedimientos y aparatos de BSS descritos en uno cualquiera de S. Amari, A. Cichocki, y H. H. Yang, “A new learning algorithm for blind signal separation”, en Advances in Neural Information Processing Systems 8, MIT Press, 1996, L. Molgedey y H. G. Schuster, “Separation of a mixture of independent signals using time delayed correlations”, Phys. Rev. Lett., 72(23): 3634-3637, 1994, o L. Parra y C. Spence, “Convolutive blind source separation of non-stationary sources”, IEEE Trans. On Speech and Audio Processing, 8(3): 320-327, mayo de 2000.

VAD basada en un realce más agresivo de señales

A veces el nivel de ruido de fondo es tan elevado que la SNR de la señal sigue sin ser buena después de la formación de haces o la separación de señales. En este caso, se puede realzar adicionalmente la SNR de la señal en la señal de referencia de frecuencia vocal. Por ejemplo, el módulo 400 de realce de señales puede implementar una sustracción espectral para realzar adicionalmente la SNR de la señal de referencia de frecuencia vocal. La señal de referencia de ruido puede o no necesitar ser realzada en este caso.

El módulo 400 de realce de señales puede, por ejemplo, implementar uno de los procedimientos y aparatos de sustracción espectral descritos en uno cualquiera de S. F. Boll, “Suppression of Acoustic Noise in Speech Using Spectral Subtraction”, IEEE Trans. Acoustics, Speech and Signal Processing, 27(2): 112-120, abril de 1979, R. Mukai, S. Araki, H. Sawada y S. Makino, “Removal of residual crosstalk components in blind source separation using LMS filters”, en Proc. Of 12th IEEE Workshop on Neural Networks for Signal Processing, pp. 435-444, Matigny, Suiza, septiembre de 2002, o R. Mukai, S. Araki, H. Sawada y S. Makino, “Removal of residual cross-talk components in blind source separation using time-delayed spectral subtraction”, en Proc. of ICASSP 2002, pp. 1789-1792, mayo de 2002.

Aplicaciones potenciales

Se pueden utilizar los procedimientos y el aparato de VAD descritos en el presente documento para eliminar el ruido de fondo. Los ejemplos proporcionados a continuación no son exhaustivos de posibles aplicaciones y no limitan la aplicación del aparato y de los procedimientos de VAD de múltiples micrófonos descritos en el presente documento. Se pueden utilizar potencialmente los procedimientos y el aparato de VAD descritos en cualquier aplicación en la que sea necesaria una decisión de VAD y haya disponibles señales de múltiples micrófonos. La VAD es adecuada para un procesamiento de señales en tiempo real pero no está limitada por una implementación potencial en aplicaciones de procesamiento de señales fuera de línea.

La Figura 5 es un diagrama simplificado de bloques funcionales de una realización de un dispositivo móvil 110 con un detector de actividad vocal con un realce opcional de señales. Se puede utilizar la decisión de VAD del módulo 230 de VAD para controlar la ganancia de un amplificador 510 de ganancia variable.

El módulo 230 de VAD puede acoplar la señal de detección de actividad de voz de salida a la entrada de un generador 520 o controlador de ganancia, que está configurado para controlar la ganancia aplicada a la señal de referencia de frecuencia vocal. En una realización, el generador 520 de ganancia está configurado para controlar la

ganancia aplicada por un amplificador 510 de ganancia variable. Se muestra el amplificador 510 de ganancia variable implementado en el dominio digital, y puede estar implementado, por ejemplo, como un escalador, un multiplicador, un registrador de impulsos, un rotador de registros, y similar, o alguna combinación de los mismos.

Como ejemplo, se puede aplicar una ganancia escalar controlada por la VAD de dos micrófonos a la señal de referencia de frecuencia vocal. Como ejemplo específico, se puede establecer la ganancia del amplificador 510 de ganancia variable como I cuando se detecta una frecuencia vocal. Se puede establecer la ganancia del amplificador 510 de ganancia variable menor que I cuando no se detecta una frecuencia vocal.

Se muestra el amplificador 510 de ganancia variable en el dominio digital, pero la ganancia variable puede ser aplicada directamente a una señal procedente del micrófono 112 de referencia de frecuencia vocal. La ganancia variable también puede ser aplicada a la señal de referencia de frecuencia vocal en el dominio digital o a la señal realzada de referencia de frecuencia vocal obtenida del módulo 400 de realce de señales, como se muestra en la Figura 5.

También se pueden utilizar los procedimientos y el aparato de VAD descritos en el presente documento para ayudar en la codificación de frecuencia vocal en módem. La Figura 6 es un diagrama simplificado de bloques funcionales de una realización de un dispositivo móvil 110 con un detector de actividad de voz que controla la codificación de la frecuencia vocal.

En la realización de la Figura 6, el módulo 230 de VAD acopla la decisión de VAD a una entrada de control de un codificador 600 de frecuencia vocal.

En general, los codificadores de frecuencia vocal en módem pueden tener detectores internos de actividad vocal, que utilizan tradicionalmente la señal o la señal realzada de un micrófono. Al utilizar un realce de señal de dos micrófonos, tal como se proporciona por el módulo 400 de realce de señales, la señal recibida por la VAD interna puede tener una SNR mejor que la señal original del micrófono. Por lo tanto, es probable que la VAD interna que utiliza la señal realzada pueda tomar una decisión más fiable. Al combinar la decisión de la VAD interna y la VAD externa, que utiliza dos señales, es posible obtener una decisión de VAD aún más fiable. Por ejemplo, el codificador 600 de frecuencia vocal puede estar configurado para llevar a cabo una combinación lógica de la decisión de VAD interna y de la decisión de VAD del módulo 230 de VAD. El codificador 600 de frecuencia vocal puede, por ejemplo, operar en la lógica Y o en la lógica O de las dos señales.

La Figura 7 es un diagrama de flujo de un procedimiento simplificado 700 de detección de actividad vocal. El procedimiento 700 puede ser implementado por el dispositivo móvil de la Figura 1 con uno o una combinación del aparato y técnicas descritos en relación con las Figuras 2-6.

Se describe el procedimiento 700 con varias etapas opcionales que pueden ser omitidas en implementaciones particulares. Además, se describe el procedimiento 700 como llevado a cabo en un orden particular únicamente para fines ilustrativos, y se pueden llevar a cabo algunas etapas en un orden distinto.

El procedimiento comienza en el bloque 710, en el que el dispositivo móvil lleva a cabo inicialmente una calibración. El dispositivo móvil puede, por ejemplo, introducir una ganancia selectiva de frecuencia, una atenuación, o un retraso para igualar sustancialmente la respuesta de los recorridos de las señales de referencia de frecuencia vocal y de referencia de ruido.

Después de la calibración, el dispositivo móvil avanza hasta el bloque 722 y recibe una señal de referencia de frecuencia vocal procedente de los micrófonos de referencia. La señal de referencia de frecuencia vocal puede incluir la presencia o la ausencia de actividad vocal.

El dispositivo móvil avanza hasta el bloque 724 y recibe al mismo tiempo una señal calibrada de referencia de ruido procedente del módulo de calibración basada en una señal procedente de un micrófono de referencia de ruido. Normalmente, el micrófono de referencia de ruido acopla un nivel reducido de señal de frecuencias vocales con respecto a los micrófonos de referencia de frecuencia vocal, pero no se requiere que lo haga.

El dispositivo móvil avanza hasta el bloque opcional 728 y lleva a cabo una cancelación de eco en las señales recibidas de frecuencia vocal y de ruido, por ejemplo, cuando el dispositivo móvil da salida a una señal de audio que puede ser acoplada a una de las señales de referencia de frecuencia vocal y de ruido, o a ambas.

El dispositivo móvil avanza hasta el bloque 730 y lleva a cabo, opcionalmente, un realce de señales de las señales de referencia de frecuencia vocal y de las señales de referencia de ruido. El dispositivo móvil puede incluir un realce de señales en dispositivos que no pueden separar de forma significativa el micrófono de referencia de frecuencia vocal del micrófono de referencia de ruido, por ejemplo, debido a limitaciones físicas. Si la estación móvil lleva a cabo un realce de señales, el procesamiento subsiguiente puede llevarse a cabo en la señal realzada de referencia de frecuencia vocal y la señal realzada de referencia de ruido. Si se omite el realce de señales, el dispositivo móvil puede operar en la señal de referencia de frecuencia vocal y la señal de referencia de ruido.

El dispositivo móvil avanza hasta el bloque 742 y determina, calcula, o genera de otra manera un valor característico de frecuencia vocal en base a la señal de referencia de frecuencia vocal. El dispositivo móvil puede estar configurado para determinar un valor característico de frecuencia vocal que es relevante para una muestra particular, en base a una pluralidad de muestras, en base a una media ponderada de muestras previas, en base a una disminución exponencial de muestras anteriores, o en base a una ventana predeterminada de muestras.

En una realización, el dispositivo móvil está configurado para determinar una autocorrelación de la señal de referencia de frecuencia vocal. En otra realización, el dispositivo móvil está configurado para determinar una energía de la señal recibida.

El dispositivo móvil avanza hasta el bloque 744 y determina, calcula, o genera de otra manera un valor característico complementario de ruido. Normalmente, la estación móvil determina el valor característico de ruido utilizando las mismas técnicas utilizadas para generar el valor característico de frecuencia vocal. Es decir, si el dispositivo móvil determina un valor característico de frecuencia vocal basado en tramas, el dispositivo móvil determina de la misma manera un valor característico de ruido basado en tramas. De forma similar, si el dispositivo móvil determina una autocorrelación como el valor característico de frecuencia vocal, el dispositivo móvil determina una autocorrelación de la señal de ruido como el valor característico de ruido.

La estación móvil puede avanzar opcionalmente hasta el bloque 746 y determinar, calcular, o generar de otra manera un valor característico combinado complementario, basado al menos en parte tanto en la señal de referencia de frecuencia vocal como en la señal de referencia de ruido. Por ejemplo, el dispositivo móvil puede estar configurado para determinar una correlación cruzada de las dos señales. En otras realizaciones, el dispositivo móvil puede omitir la determinación de un valor característico combinado, por ejemplo, tal como cuando la métrica de actividad vocal no está basada en un valor característico combinado.

El dispositivo móvil avanza hasta el bloque 750 y determina, calcula, o genera de otra manera una métrica de actividad vocal basada al menos en parte en uno o más del valor característico de frecuencia vocal, del valor característico de ruido, y del valor característico combinado. En una realización, el dispositivo móvil está configurado para determinar una relación del valor de autocorrelación de frecuencia vocal con respecto al valor de correlación cruzada combinado. En otra realización, el dispositivo móvil está configurado para determinar una relación del valor de energía de frecuencia vocal con respecto al valor de energía de ruido. El dispositivo móvil puede determinar, de forma similar, otra métrica de actividad utilizando otras técnicas.

El dispositivo móvil avanza hasta el bloque 760 y toma la decisión de actividad de voz o determina de otra manera el estado de actividad vocal. Por ejemplo, el dispositivo móvil puede tomar la determinación de actividad vocal al comparar la métrica de actividad vocal con uno o más umbrales. Los umbrales pueden ser fijos o dinámicos. En una realización, el dispositivo móvil determina la presencia de actividad vocal si la métrica de actividad vocal supera un umbral predeterminado.

Después de determinar el estado de actividad vocal, el dispositivo móvil avanza hasta el bloque 770 y varía, ajusta, o modifica de otra manera uno o más parámetros o controles en base en parte al estado de actividad vocal. Por ejemplo, el dispositivo móvil puede establecer una ganancia de un amplificador de señal de referencia de frecuencia vocal en base al estado de actividad vocal, puede utilizar el estado de actividad vocal para controlar un codificador de frecuencia vocal, o puede utilizar el estado de actividad vocal en combinación con otra decisión de VAD para controlar un estado del codificador de la frecuencia vocal.

El dispositivo móvil avanza hasta el bloque 780 de decisión para determinar si se desea una recalibración. El dispositivo móvil puede llevar a cabo una calibración tras el paso de uno o más eventos, periodos de tiempo, y similares, o alguna combinación de los mismos. Si se desea una recalibración, el dispositivo móvil vuelve al bloque 710. De lo contrario, el dispositivo móvil puede volver al bloque 722 para continuar monitorizando las señales de referencia de frecuencia vocal y de ruido en busca de actividad vocal.

La Figura 8 es un diagrama simplificado de bloques funcionales de una realización de un dispositivo móvil 800 con un detector calibrado de actividad vocal en múltiples micrófonos y un realce de señales. El dispositivo móvil 800 incluye micrófonos 812 y 814 de referencia de frecuencia vocal y de ruido, un medio para convertir las señales de referencia de frecuencia vocal y de ruido en representaciones digitales, 822 y 824, y medios para cancelar ecos en las señales 842 y 844 de referencia de frecuencia vocal y de ruido. Los medios para cancelar los ecos operan junto con medios para combinar una señal 832 y 834 con la salida procedente del medio de cancelación.

Las señales de referencia de frecuencia vocal y de ruido de eco cancelado pueden ser acopladas a un medio para calibrar 850 una respuesta espectral de un recorrido de la señal de referencia de frecuencia vocal para que sea sustancialmente similar a una respuesta espectral de un recorrido de la señal de referencia de ruido. Las señales de referencia de frecuencia vocal y de ruido también pueden estar acopladas a un medio 856 para realzar al menos una de la señal de referencia de frecuencia vocal o la señal de referencia de ruido. Si se utiliza el medio 856 para realzar, la métrica de actividad vocal está basada al menos en parte en una de una señal realzada de referencia de frecuencia vocal o una señal realzada de referencia de ruido.

Un medio para detectar 860 actividad vocal puede incluir un medio para determinar una autocorrelación basada en la señal de referencia de frecuencia vocal, un medio para determinar una correlación cruzada basada en la señal de referencia de frecuencia vocal y en la señal de referencia de ruido, un medio para determinar una métrica de actividad vocal basada en parte en una relación de la autocorrelación de la señal de referencia de frecuencia vocal con respecto a la correlación cruzada, y un medio para determinar un estado de actividad vocal al comparar la métrica de actividad vocal con al menos un umbral.

En el presente documento se describen los procedimientos y el aparato para una detección de actividad vocal y para variar la operación de una o más porciones de un dispositivo móvil en base al estado de actividad vocal. Los procedimientos y el aparato de VAD presentados en el presente documento pueden ser utilizados por sí solos, pueden ser combinados con procedimientos y aparatos tradicionales de VAD para tomar decisiones de VAD más fiables. Como ejemplo, el procedimiento de VAD dado a conocer puede ser combinado con un procedimiento de paso por cero para tomar una decisión más fiable de actividad vocal.

Se debería hacer notar que una persona con un nivel normal de dominio de la técnica reconocerá que un circuito puede implementar algunas de las funciones descritas anteriormente, o todas ellas. Puede haber un circuito que implemente todas las funciones. También puede haber múltiples secciones de un circuito en combinación con un segundo circuito que puede implementar todas las funciones. En general, si se implementan múltiples funciones en el circuito, puede ser un circuito integrado. Con las tecnologías actuales de plataforma móvil, un circuito integrado comprende al menos un procesador de señales digitales (DSP), y al menos un procesador ARM para controlar y/o comunicarse con el al menos un DSP. Se puede describir un circuito por secciones. A menudo, se reutilizan secciones para llevar a cabo distintas funciones. Por lo tanto, al describir qué circuitos comprenden algunas de las anteriores descripciones, una persona con un nivel normal de dominio de la técnica comprenderá que una primera sección, una segunda sección, una tercera sección, una cuarta sección, y una quinta sección de un circuito pueden ser el mismo circuito, o pueden ser distintos circuitos que son parte de un circuito mayor o de un conjunto de circuitos.

Un circuito puede estar configurado para detectar la actividad vocal, comprendiendo el circuito una primera sección adaptada para recibir una señal de referencia de frecuencia vocal de salida procedente de un micrófono de referencia de frecuencia vocal. El mismo circuito, un circuito distinto, o una segunda sección del mismo circuito, o diferente, puede estar configurado para recibir una señal de referencia de salida procedente de un micrófono de referencia de ruido. Además, puede haber un mismo circuito, un distinto circuito, o una tercera sección del mismo circuito, o distinto, que comprende un generador de valor característico de frecuencia vocal acoplado a la primera sección configurada para determinar un valor característico de frecuencia vocal. Una cuarta sección que comprende un generador de valor característico combinado acoplado a la primera sección y a la segunda sección configuradas para determinar un valor característico combinado también pueden ser parte del circuito integrado. Además, una quinta sección que comprende un módulo de métrica de actividad vocal configurado para determinar una métrica de actividad vocal en base, al menos en parte, al valor característico de frecuencia vocal y al valor característico combinado puede ser parte del circuito integrado. Para comparar la métrica de actividad vocal con un umbral y una salida de un estado de actividad vocal se puede utilizar un comparador. En general, cualquiera de las secciones (primera, segunda, tercera, cuarta o quinta) puede ser parte del circuito integrado, o ser independiente del mismo. Es decir, cada una de las secciones puede ser parte de un circuito mayor, o cada una puede ser un circuito integrado individual o una combinación de los dos.

Como se ha descrito anteriormente, el micrófono de referencia de frecuencia vocal comprende una pluralidad de micrófonos y el generador de valor característico de frecuencia vocal puede estar configurado para determinar una autocorrelación de la señal de referencia de frecuencia vocal y/o determinar una energía de la señal de referencia de frecuencia vocal, y/o determinar una media ponderada en base a una disminución exponencial de anteriores valores característicos de frecuencia vocal. Como se ha descrito anteriormente, las funciones del generador de valor característico de frecuencia vocal pueden ser implementadas en una o más secciones de un circuito.

Según se utiliza en el presente documento, el término acoplado o conectado se utiliza con el significado de un acoplamiento indirecto al igual que un acoplamiento o una conexión directos. Cuando hay acoplados dos o más bloques, módulos, dispositivos o aparatos, puede haber uno o más bloques interpuestos entre los dos bloques acoplados.

Los diversos bloques lógicos, módulos, y circuitos ilustrativos descritos en conexión con las realizaciones dadas a conocer en el presente documento pueden ser implementados o llevados a cabo con un procesador para uso general, un procesador de señales digitales (DSP), un procesador de un Ordenador con grupo reducido de instrucciones (RISC), un Circuito integrado para aplicaciones específicas (ASIC), una Matriz de puertas de campo programable, u otro dispositivo lógico programable, puerta discreta o lógica de transistor, componentes discretos de *hardware*, o cualquier combinación de los mismos diseñada para llevar a cabo las funciones descritas en el presente documento. Un procesador para uso general puede ser un microprocesador, pero de forma alternativa, el procesador puede ser cualquier procesador, controlador, microcontrolador, o máquina de estado. También se puede implementar un procesador 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 ese tipo.

Las etapas de un procedimiento, proceso, o algoritmo descrito en conexión con las realizaciones dadas a conocer en el presente documento pueden ser implementadas directamente en el *hardware*, en un módulo de *software* ejecutado por un procesador, o en una combinación de los dos. Las diversas etapas o acciones en un procedimiento o proceso pueden ser llevadas a cabo en el orden mostrado, o pueden ser llevadas a cabo en otro orden. Además, se pueden omitir una o más etapas del proceso o del procedimiento o se pueden añadir una o más etapas del proceso o del procedimiento a los procedimientos o procesos. Se puede añadir una etapa, un bloque, o una acción adicional al comienzo, al final, o interpuesto entre elementos existentes de los procedimientos y procesos.

Se proporciona la anterior descripción de las realizaciones dadas a conocer para permitir a cualquier persona con un nivel normal de dominio de la técnica realizar o utilizar la revelación. Serán inmediatamente evidentes diversas modificaciones a estas realizaciones para las personas con un nivel normal de dominio de la técnica, y los principios genéricos definidos en el presente documento pueden ser aplicados a otras realizaciones sin alejarse del alcance de la revelación siempre que que se encuentren dentro del alcance de las reivindicaciones adjuntas.

REIVINDICACIONES

1. Un procedimiento para detectar actividad vocal, comprendiendo el procedimiento:

recibir (722) una señal de referencia de frecuencia vocal procedente de un micrófono (112) de referencia de frecuencia vocal;

recibir (724) una señal de referencia de ruido procedente de un micrófono (114) de referencia de ruido distinto del micrófono (112) de referencia de frecuencia vocal;

determinar (742) un valor característico de frecuencia vocal en base, al menos en parte, a la señal de referencia de frecuencia vocal;

determinar (746) un valor característico combinado en base, al menos en parte, a la señal de referencia de frecuencia vocal y a la señal de referencia de ruido;

determinar (750) una métrica de actividad vocal en base, al menos en parte, al valor característico de frecuencia vocal y al valor característico combinado,

en el que determinar (742) el valor característico de frecuencia vocal comprende determinar un valor absoluto de una autocorrelación de la señal de referencia de frecuencia vocal y determinar (746) el valor característico combinado comprende determinar una correlación cruzada en base a la señal de referencia de frecuencia vocal y a la señal de referencia de ruido, y

en el que determinar (750) la métrica de actividad vocal comprende determinar una relación del valor absoluto de la autocorrelación de la señal de referencia de frecuencia vocal con respecto a la correlación cruzada; y

determinar (760) un estado de actividad vocal en base a la métrica de actividad vocal.

2. El procedimiento de la reivindicación 1, que comprende, además:

formar un haz con al menos una de la señal de referencia de frecuencia vocal o la señal de referencia de ruido;

llevar a cabo una Separación ciega de fuentes, BSS, en la señal de referencia de frecuencia vocal y la señal de referencia de ruido para realizar un componente de señal de frecuencia vocal en la señal de referencia de frecuencia vocal;

llevar a cabo una sustracción espectral en al menos una de la señal de referencia de frecuencia vocal o la señal de referencia de ruido; o

determinar un valor característico de ruido en base, al menos en parte, a la señal de referencia de ruido, y en el que la métrica de actividad vocal está basada, al menos en parte, en el valor característico de ruido.

3. El procedimiento de la reivindicación 1, en el que la señal de referencia de frecuencia vocal incluye la presencia o la ausencia de actividad vocal, y preferentemente:

la autocorrelación comprende una suma ponderada de una autocorrelación anterior con una energía de referencia de frecuencia vocal en un instante temporal particular;

determinar el valor característico de frecuencia vocal comprende determinar una energía de la señal de referencia de frecuencia vocal;

determinar el valor característico combinado comprende determinar una correlación cruzada en base a la señal de referencia de frecuencia vocal y a la señal de referencia de ruido; o

determinar el estado de actividad vocal comprende comparar la métrica de actividad vocal con un umbral.

4. El procedimiento de la reivindicación 1, en el que:

el micrófono (112) de referencia de frecuencia vocal comprende al menos un micrófono de frecuencia vocal;

el micrófono (114) de referencia de ruido comprende al menos un micrófono de ruido distinto del al menos un micrófono de frecuencia vocal;

determinar (742) el valor característico de frecuencia vocal comprende determinar una autocorrelación en base a la señal de referencia de frecuencia vocal; y

determinar (760) el estado de actividad vocal comprende comparar la métrica de actividad vocal con al menos un umbral.

5. El procedimiento de la reivindicación 4, que comprende, además:

llevar a cabo (730) un realce de señal de al menos una de la señal de referencia de frecuencia vocal o de la señal de referencia de ruido, y en el que la métrica de actividad vocal está basada, al menos en parte, en una de una señal realzada de referencia de frecuencia vocal o una señal realzada de referencia de ruido; o

variar (770) un parámetro operativo en base al estado de actividad vocal.

6. El procedimiento de la reivindicación 5, en el que el parámetro operativo comprende:

una ganancia aplicada a la señal de referencia de frecuencia vocal; o

un estado de un codificador de frecuencia vocal que opera en la señal de referencia de frecuencia vocal.

7. Un aparato configurado para detectar actividad vocal, comprendiendo el aparato:

un medio (112) para recibir una señal de referencia de frecuencia vocal;

un medio (114) para recibir una señal de referencia de ruido;

un medio (232) para determinar un valor característico de frecuencia vocal en base a la señal de referencia de frecuencia vocal al determinar un valor absoluto de una autocorrelación de la señal de referencia de frecuencia vocal;

un medio (236) para determinar un valor característico combinado al determinar una correlación cruzada en base a la señal de referencia de frecuencia vocal y a la señal de referencia de ruido;

un medio (240) para determinar una métrica de actividad vocal al determinar una relación del valor absoluto de la autocorrelación de la señal de referencia de frecuencia vocal con respecto a la correlación cruzada; y

un medio (250) para determinar un estado de actividad vocal al comparar la métrica de actividad vocal con al menos un umbral.

8. El aparato de la reivindicación 7, que comprende, además:

un micrófono de referencia de frecuencia vocal configurado para dar salida a una señal de referencia de frecuencia vocal; y

un micrófono de referencia de ruido configurado para dar salida a una señal de referencia de ruido.

9. El aparato de la reivindicación 7, que comprende, además, un medio para calibrar una respuesta espectral de un recorrido de la señal de referencia de frecuencia vocal para que sea sustancialmente similar a una respuesta espectral de un recorrido de la señal de referencia de ruido.

10. El aparato de la reivindicación 8, en el que:

el micrófono de referencia de frecuencia vocal comprende una pluralidad de micrófonos; o

el medio para determinar un valor característico de frecuencia vocal está configurado para determinar una media ponderada en base a una disminución exponencial de valores característicos anteriores de frecuencia vocal.

11. El aparato de la reivindicación 8, en el que el medio para determinar una métrica de actividad vocal está configurado para determinar una relación del valor característico de frecuencia vocal con respecto a un valor característico de ruido determinado en base a la señal de referencia de ruido.

12. El aparato de la reivindicación 7, que comprende un circuito configurado para detectar actividad vocal, en el que:

el medio para recibir una señal de referencia de frecuencia vocal comprende una primera sección del circuito adaptada para recibir una señal de referencia de frecuencia vocal de salida procedente de un micrófono de referencia de frecuencia vocal;

el medio para recibir una señal de referencia de ruido comprende una segunda sección del circuito adaptada para recibir una señal de referencia de ruido de salida procedente de un micrófono de referencia de ruido;

el medio para determinar un valor característico de frecuencia vocal comprende una tercera sección del circuito que comprende un generador de valor característico de frecuencia vocal acoplado a la primera sección configurada para determinar un valor característico de frecuencia vocal, en el que determinar el valor característico de frecuencia vocal comprende determinar un valor absoluto de la autocorrelación de la señal de referencia de frecuencia vocal;

el medio para determinar un valor característico combinado comprende una cuarta sección del circuito que comprende un generador de valor característico combinado acoplado a la primera sección y a la segunda sección configuradas para determinar un valor característico combinado, en el que determinar el valor característico combinado comprende determinar una correlación cruzada en base a la señal de referencia de frecuencia vocal y a la señal de referencia de ruido;

el medio para determinar una métrica de actividad vocal comprende una quinta sección del circuito que comprende un módulo de métrica de actividad vocal configurado para determinar una métrica de actividad vocal al determinar una relación del valor absoluto de la autocorrelación de la señal de referencia de frecuencia vocal con respecto a la correlación cruzada; y

el medio para determinar un estado de actividad vocal comprende un comparador configurado para comparar la métrica de actividad vocal con un umbral y dar salida a un estado de actividad vocal.

13. El aparato de la reivindicación 12, en el que cualesquiera dos secciones en un grupo consistente en la primera sección, la segunda sección, la tercera sección, la cuarta sección, y la quinta sección del circuito comprenden circuitería similar.

14. Un medio legible por un ordenador que incluye instrucciones que, cuando son ejecutadas por un procesador, tienen como resultado la realización de etapas de procedimiento de cualquiera de las reivindicaciones 1 a 6.

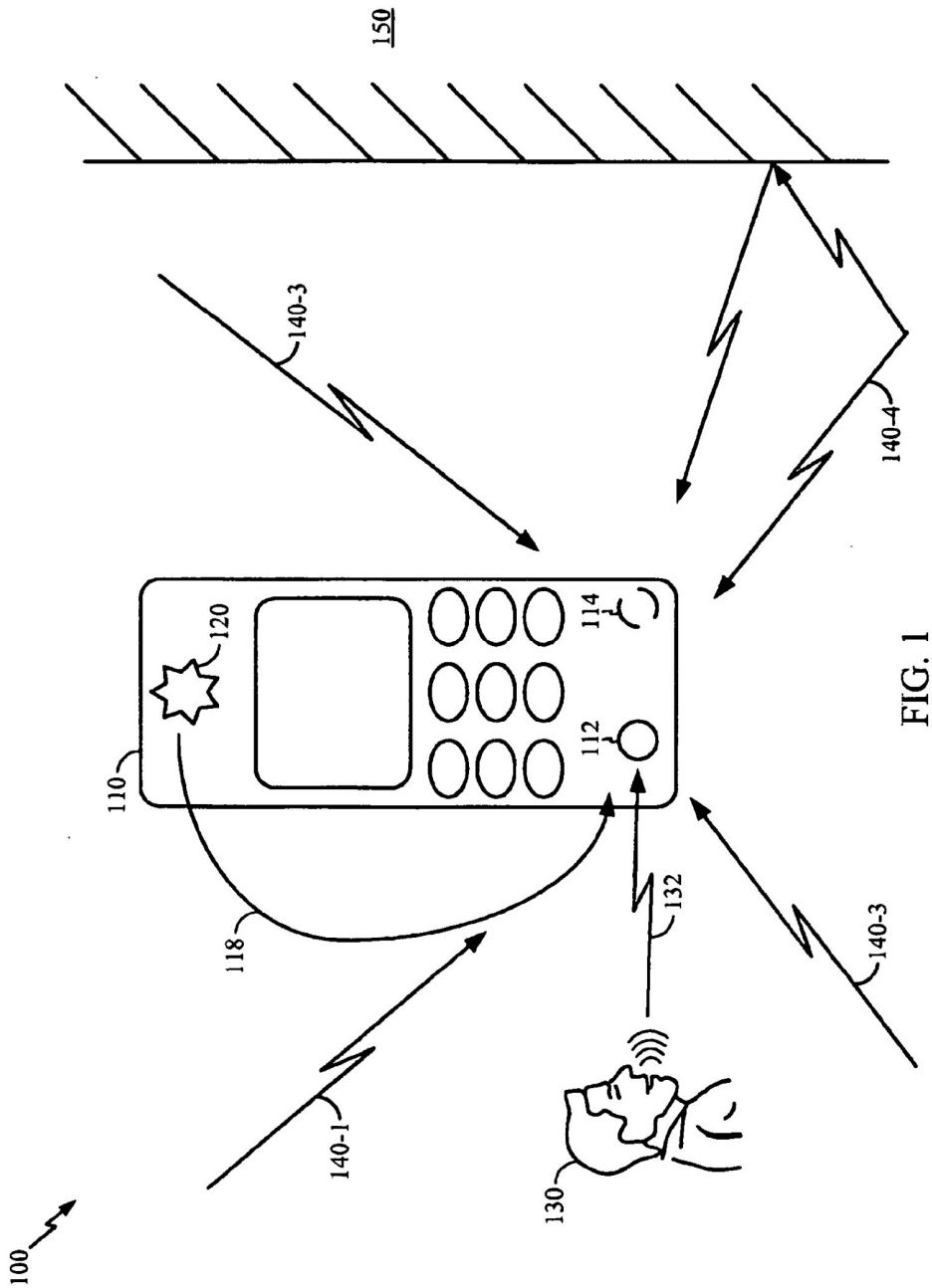

FIG. 1

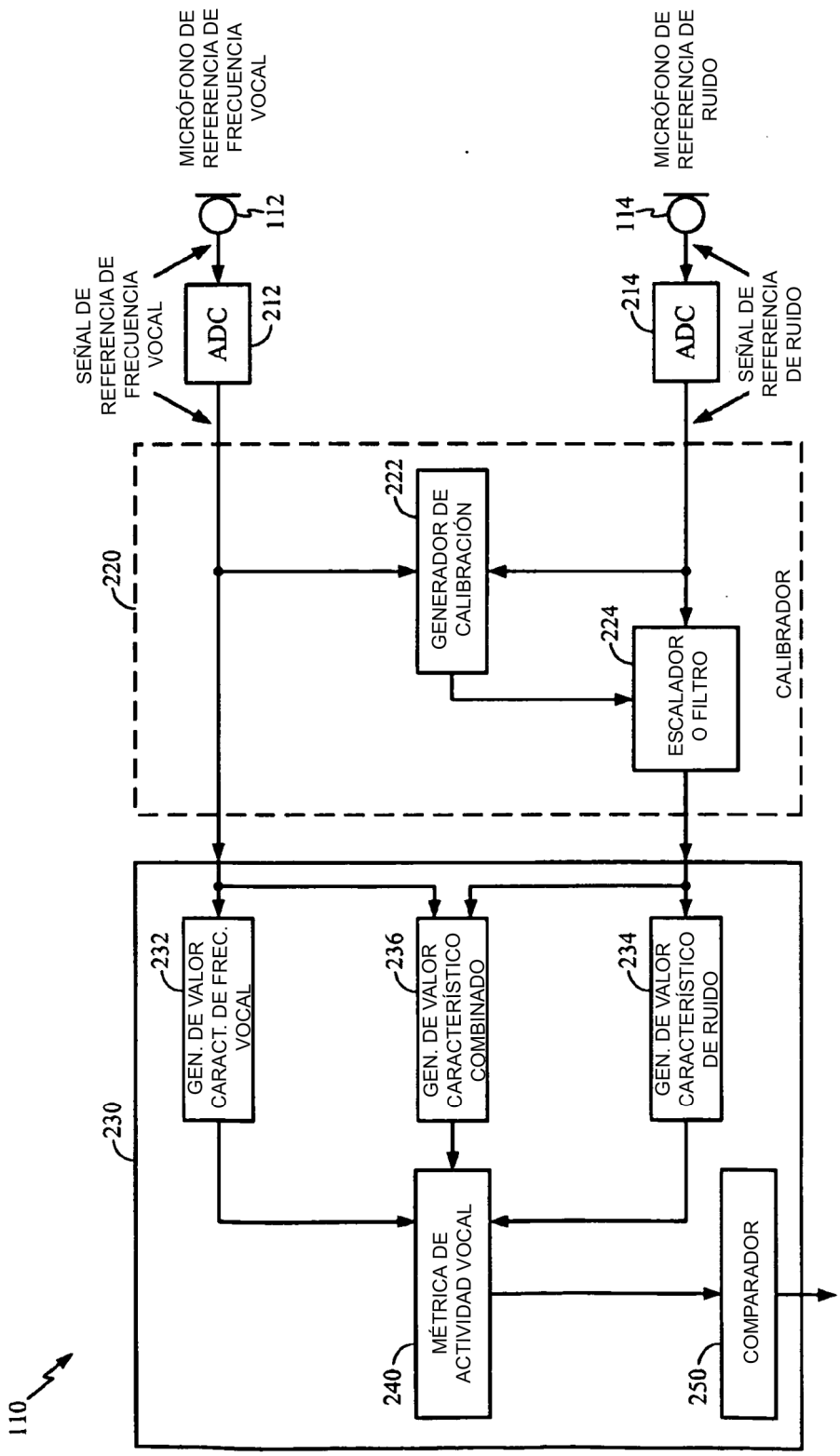
MICRÓFONO DE REFERENCIA DE FRECUENCIA VOCAL
112
MICRÓFONO DE REFERENCIA DE RUIDO
114
SEÑAL DE REFERENCIA DE FRECUENCIA VOCAL
ADC
212
214
ADC
SEÑAL DE REFERENCIA DE RUIDO
220
222
GENERADOR DE CALIBRACIÓN
224
ESCALADOR O FILTRO
CALIBRADOR
230
232
GEN. DE VALOR CARACT. DE FREC. VOCAL
236
GEN. DE VALOR CARACTERÍSTICO COMBINADO
234
GEN. DE VALOR CARACTERÍSTICO DE RUIDO
MÉTRICA DE ACTIVIDAD VOCAL
240
COMPARADOR
250
110

FIG. 2

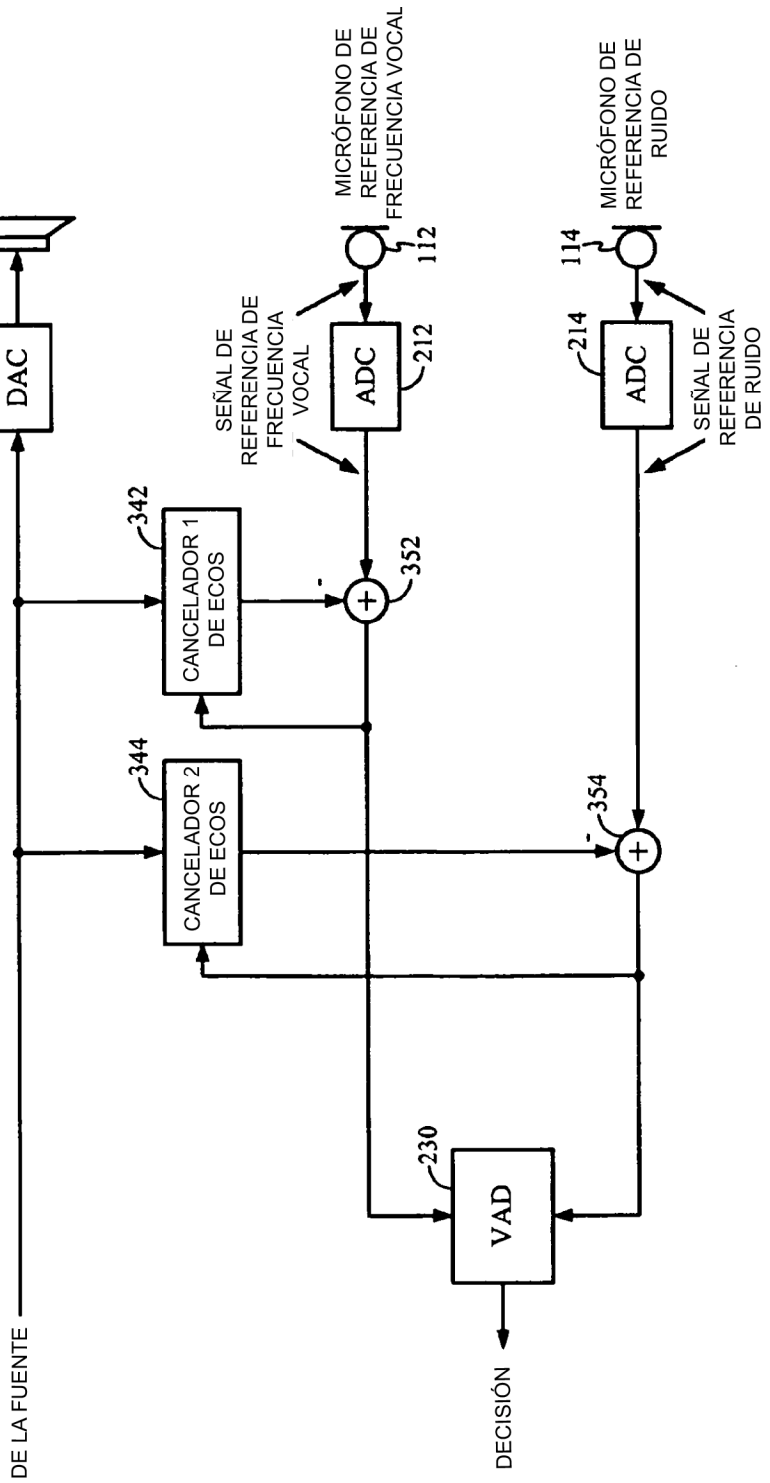
320
310
DAC
MICRÓFONO DE REFERENCIA DE FRECUENCIA VOCAL
112
SEÑAL DE REFERENCIA DE FRECUENCIA VOCAL
ADC
212
MICRÓFONO DE REFERENCIA DE RUIDO
114
214
ADC
SEÑAL DE REFERENCIA DE RUIDO
342
CANCELADOR 1 DE ECOS
352
344
CANCELADOR 2 DE ECOS
354
230
VAD
DE LA FUENTE
DECISIÓN

FIG. 3

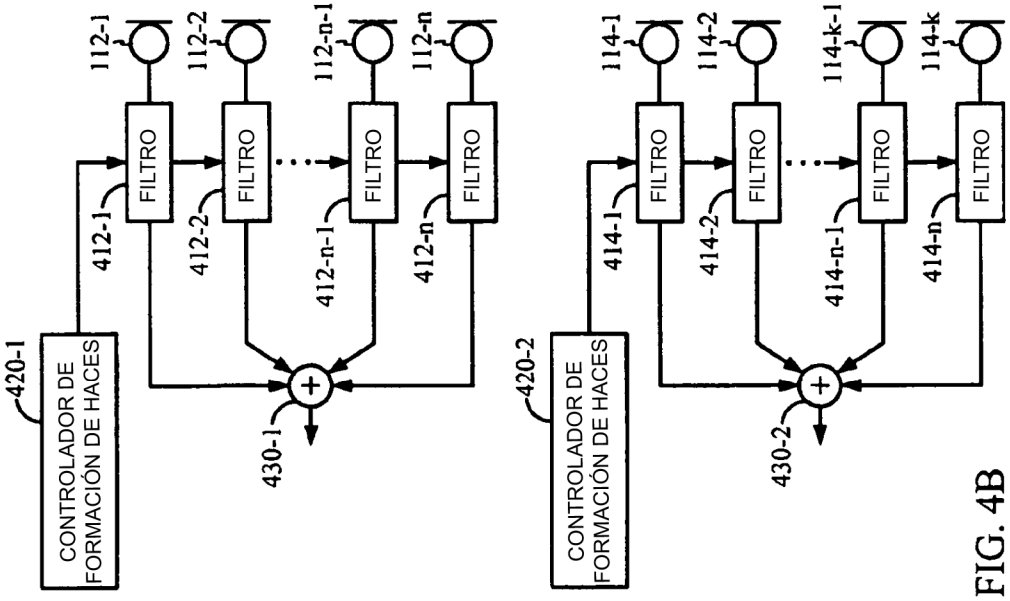
400
CONTROLADOR DE FORMACIÓN DE HACES
420-1
412-1
412-2
412-n-1
412-n
FILTRO
112-1
112-2
112-n-1
112-n
430-1
CONTROLADOR DE FORMACIÓN DE HACES
420-2
414-1
414-2
414-n-1
414-n
114-1
114-2
114-k-1
114-k
430-2

FIG. 4B

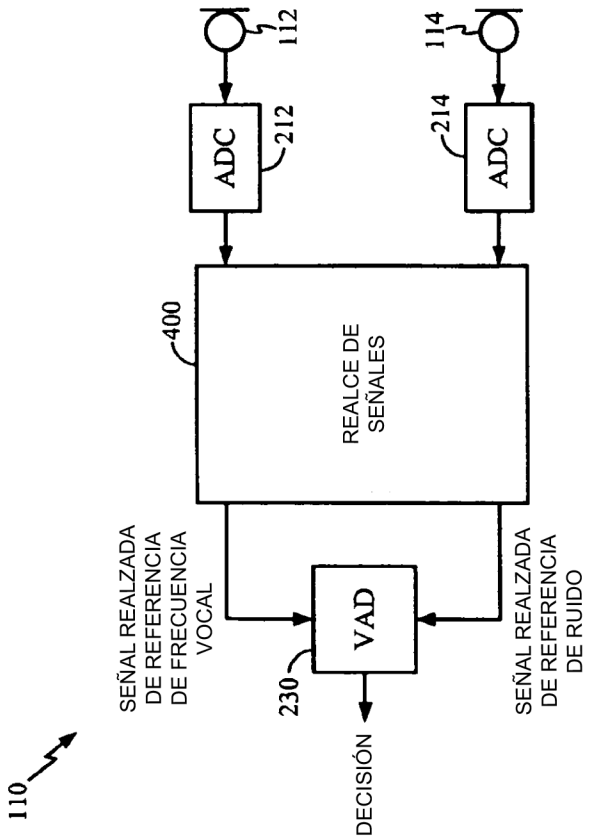
110
400
REALCE DE SEÑALES
ADC
212
214
112
114
VAD
230
DECISIÓN
SEÑAL REALZADA DE REFERENCIA DE FRECUENCIA VOCAL
SEÑAL REALZADA DE REFERENCIA DE RUIDO

FIG. 4A

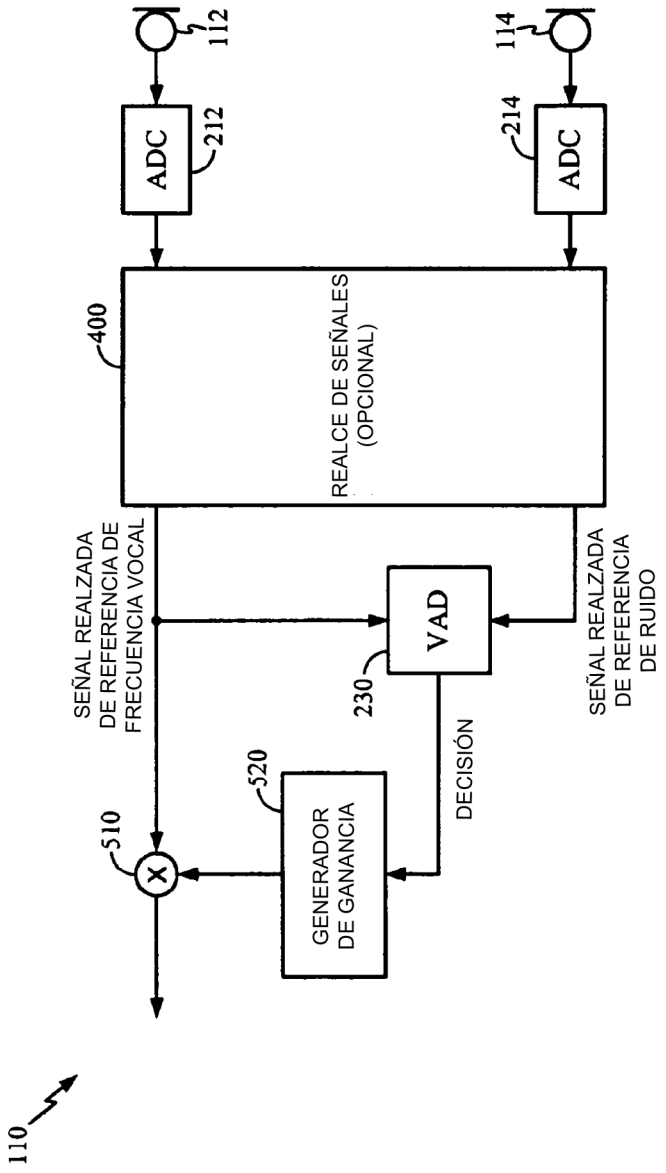
110
112
114
ADC
212
214
ADC
400
REALCE DE SEÑALES
(OPCIONAL)
SEÑAL REALZADA
DE REFERENCIA DE
FRECUENCIA VOCAL
SEÑAL REALZADA
DE REFERENCIA
DE RUIDO
VAD
230
DECISIÓN
520
GENERADOR
DE GANANCIA
510

FIG. 5

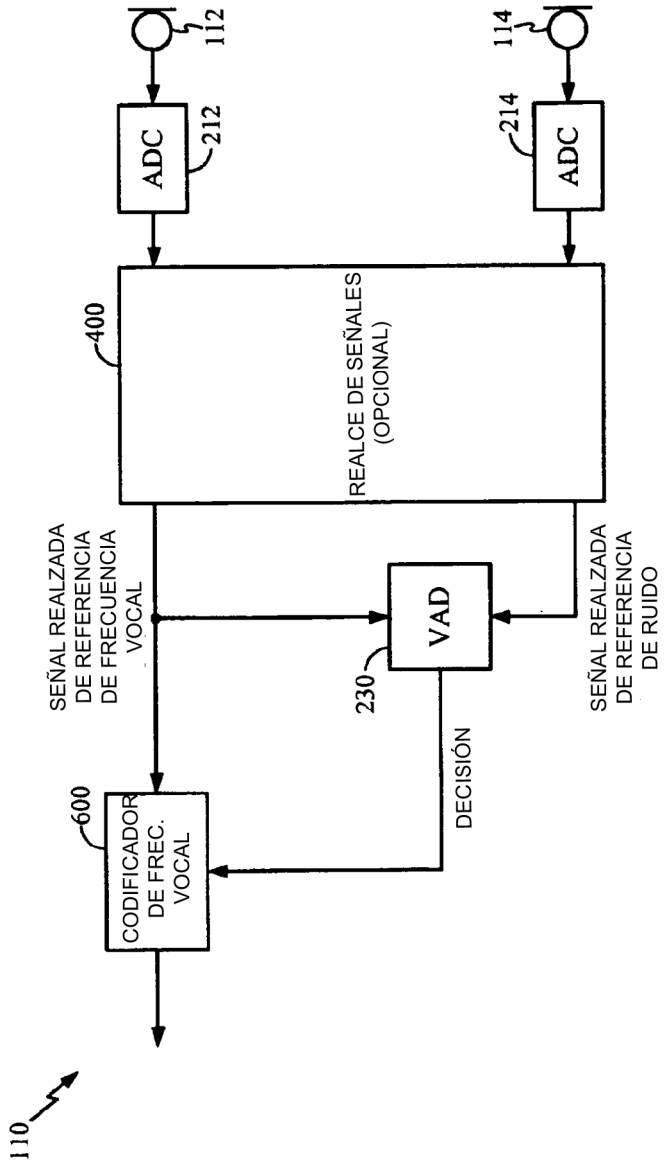
112
114
ADC
212
214
ADC
400
REALCE DE SEÑALES
(OPCIONAL)
SEÑAL REALZADA
DE REFERENCIA
DE FRECUENCIA
VOCAL
VAD
230
SEÑAL REALZADA
DE REFERENCIA
DE RUIDO
DECISIÓN
600
CODIFICADOR
DE FREC.
VOCAL
110

FIG. 6

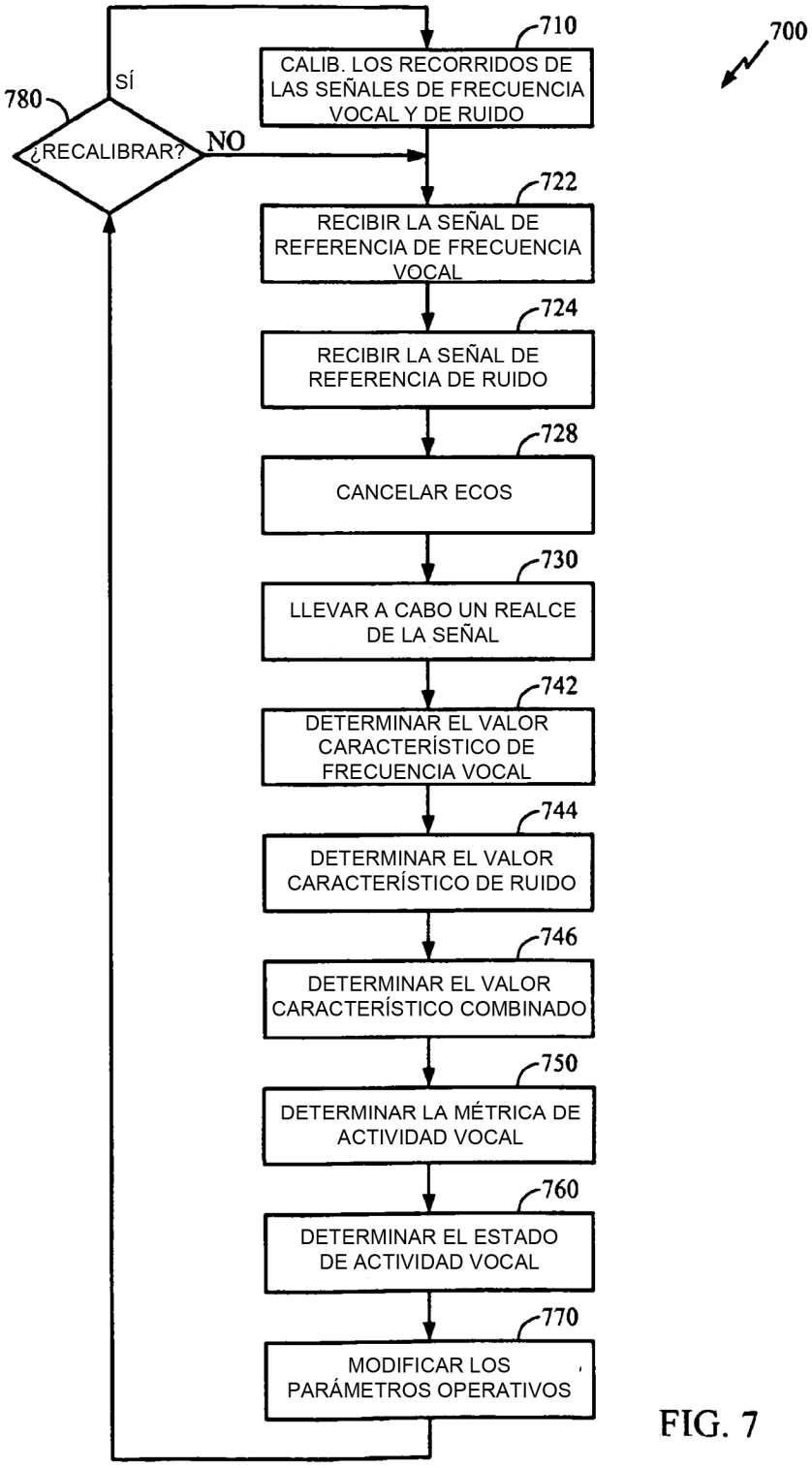
700
710
CALIB. LOS RECORRIDOS DE LAS SEÑALES DE FRECUENCIA VOCAL Y DE RUIDO
780
SÍ
¿RECALIBRAR?
NO
722
RECIBIR LA SEÑAL DE REFERENCIA DE FRECUENCIA VOCAL
724
RECIBIR LA SEÑAL DE REFERENCIA DE RUIDO
728
CANCELAR ECOS
730
LLEVAR A CABO UN REALCE DE LA SEÑAL
742
DETERMINAR EL VALOR CARACTERÍSTICO DE FRECUENCIA VOCAL
744
DETERMINAR EL VALOR CARACTERÍSTICO DE RUIDO
746
DETERMINAR EL VALOR CARACTERÍSTICO COMBINADO
750
DETERMINAR LA MÉTRICA DE ACTIVIDAD VOCAL
760
DETERMINAR EL ESTADO DE ACTIVIDAD VOCAL
770
MODIFICAR LOS PARÁMETROS OPERATIVOS

FIG. 7

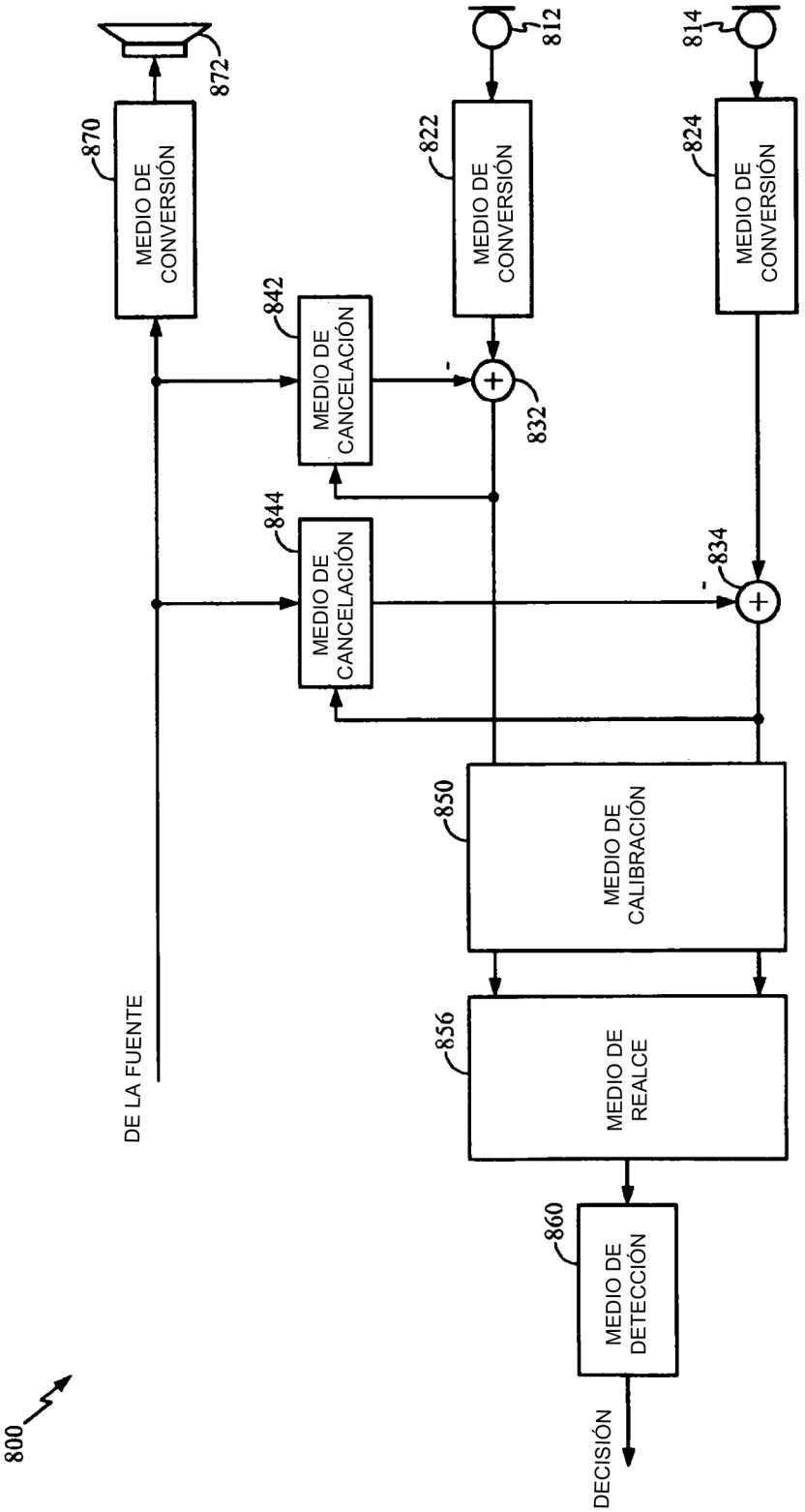
800
DE LA FUENTE
870
MEDIO DE CONVERSIÓN
872
822
MEDIO DE CONVERSIÓN
812
824
MEDIO DE CONVERSIÓN
814
842
MEDIO DE CANCELACIÓN
844
MEDIO DE CANCELACIÓN
832
834
850
MEDIO DE CALIBRACIÓN
856
MEDIO DE REALCE
860
MEDIO DE DETECCIÓN
DECISIÓN

FIG. 8