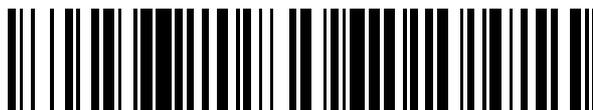


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 566 569**

51 Int. Cl.:

G06F 17/30 (2006.01)

G10L 25/54 (2013.01)

G10L 15/08 (2006.01)

G10L 15/26 (2006.01)

G10L 15/32 (2013.01)

G10L 15/02 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **28.06.2012 E 12382256 (1)**

97 Fecha y número de publicación de la concesión europea: **06.01.2016 EP 2680165**

54 Título: **Sistema y método para realizar consultas textuales en comunicaciones de voz**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
13.04.2016

73 Titular/es:

JAJAH LTD (50.0%)
9 H'Atsiya Street
Ra'Anana, IL y
TELEFÓNICA, S.A. (50.0%)

72 Inventor/es:

NEYSTADT, JOHN y
URDIALES DELGADO, DIEGO

74 Agente/Representante:

ARIZTI ACHA, Monica

ES 2 566 569 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).

Sistema y método para realizar consultas textuales en comunicaciones de voz

DESCRIPCIÓN

5 **Campo técnico**

La presente invención se refiere a sistemas y métodos para realizar consultas textuales en comunicaciones de voz.

10 **Antecedentes**

10 Muchos sistemas de comunicación ofrecen a los usuarios la capacidad de realizar búsquedas de texto en las comunicaciones textuales archivadas, tales como correo electrónico y mensajes de texto. Normalmente, el usuario proporciona al motor de búsqueda uno o más términos de búsqueda y el motor de búsqueda devuelve una lista de comunicaciones textuales (por ejemplo, correo electrónico y mensajes de texto) que contienen el término o términos de búsqueda.

15 El usuario puede realizar cualquier acción estándar en las comunicaciones mencionadas textuales, tales como leer /ver la comunicación, reenviar la comunicación, y copiar el texto de la comunicación.

20 Ha existido durante mucho tiempo la necesidad de proporcionar a los usuarios la capacidad de realizar búsquedas de texto similares en las comunicaciones de voz archivadas, tales como llamadas de voz y mensajes de voz.

25 Una patente relacionada en el campo es la solicitud de patente US-A1-20120029918 que desvela un sistema para grabar, buscar y compartir ficheros multimedia entre una pluralidad de usuarios. En dicha solicitud de patente el sistema incluye un servidor que está configurado para recibir, indexar y almacenar una pluralidad de ficheros multimedia, que se reciben mediante el servidor desde una pluralidad de fuentes, en al menos una base de datos en comunicación con el servidor. Además, el servidor está configurado para hacer a uno o más de los ficheros multimedia accesibles a una o más personas distintas de las fuentes originales de tales ficheros multimedia y para transcribir los ficheros multimedia en texto; recibir y publicar comentarios asociados con los ficheros multimedia en una interfaz de usuario gráfica de un sitio web; y permitir a usuarios consultar y reproducir porciones extraídas de tales ficheros multimedia. En dicha solicitud de patente, al contrario de la presente invención, no se calcula una suma ponderada de una puntuación de coincidencia para cada comunicación de voz localizada mediante la búsqueda. En la presente invención la ponderación puede basarse en el conjunto de datos de contenido de audio particulares usados para identificar la comunicación de voz.

35 Por lo tanto, hay una necesidad de un sistema para hacer frente a las deficiencias de los sistemas actuales.

Las realizaciones descritas a continuación no se limitan a las implementaciones que resuelven cualquiera o todas las desventajas de los procedimientos y sistemas conocidos.

40

Resumen

45 Este resumen se presenta para introducir una selección de conceptos de una forma simplificada que se describe más adelante en la descripción detallada. Este resumen no está destinado a identificar las características clave o características esenciales de la materia objeto reivindicada, ni está destinado a ser utilizado como una ayuda en la determinación del alcance de la materia objeto reivindicada.

50 En un primer aspecto, se proporciona un sistema para realizar consultas textuales en comunicaciones de voz de acuerdo con la reivindicación independiente 1.

En un segundo aspecto se proporciona un método implementado por ordenador para realizar consultas textuales en comunicaciones de voz de acuerdo con la reivindicación independiente 9.

55 Los métodos descritos en este punto pueden realizarse por software en forma legible por máquina en un medio de almacenamiento tangible, por ejemplo, en forma de un programa informático que comprende medios de código de programa informático adaptados para realizar todas las etapas de cualquiera de los métodos descritos en este documento, cuando el programa se ejecuta en un ordenador y donde el programa de ordenador puede incorporarse en un medio legible por ordenador. Ejemplos de medios de almacenamiento tangibles (o no transitorios) incluyen discos, memorias USB, tarjetas de memoria, etc., y no incluyen las señales propagadas. El software puede ser adecuado para su ejecución en un procesador paralelo o un procesador en serie de tal manera que las etapas del procedimiento pueden llevarse a cabo en cualquier orden adecuado, o simultáneamente.

60 Se reconoce así que el firmware y el software pueden ser productos valiosos comercializables por separado. Se pretende abarcar software, que se ejecuta en o controla "vuelca" o hardware convencional, para llevar a cabo las

funciones deseadas. También se pretende abarcar software que "describe" o define la configuración de hardware, tales como el software HDL (lenguaje de descripción de hardware), tal como se utiliza para diseñar los chips de silicio, o para configurar chips programables universales, para llevar a cabo las funciones deseadas. Las características preferidas se pueden combinar según sea apropiado, como sería evidente para un experto, y pueden combinarse con cualquiera de los aspectos de la invención.

Breve descripción de los dibujos

Las realizaciones de la invención se describirán, a modo de ejemplo, con referencia a los siguientes dibujos, en los cuales:

La figura 1 es un diagrama de bloques esquemático de un sistema para realizar consultas textuales en las comunicaciones de voz;

La figura 2 es un diagrama esquemático de bloques del servicio de análisis del habla de la figura 1;

La figura 3 es un esquema de una interfaz de usuario de búsqueda para el dispositivo de usuario final de la figura 1;

La figura 4 es un método para generar conjuntos de datos de contenidos de audio indexados empleando el sistema de la figura 1;

La figura 5 es un procedimiento para buscar texto en comunicaciones de voz empleando el sistema de la figura 1; y

La figura 6 es un diagrama de bloques de un ejemplo de dispositivo de usuario final ejemplar de la figura 1.

En las figuras se utilizan números de referencia comunes para indicar características similares.

Descripción detallada

Las realizaciones de la presente invención se describen a continuación solamente a modo de ejemplo. Estos ejemplos representan las mejores maneras de poner en práctica la invención que actualmente conoce el solicitante, aunque no son las únicas formas en que esto se podría lograr. La descripción expone las funciones del ejemplo y la secuencia de etapas para implementar y llevar a cabo el ejemplo. Sin embargo, las mismas funciones y secuencias o equivalentes pueden efectuarse mediante diversos ejemplos.

La figura 1 muestra un diagrama de bloques de un sistema 100 para buscar texto en comunicaciones de voz. El sistema 100 incluye un dispositivo de usuario final 102 para hacer o recibir una comunicación de voz, un servidor de comunicaciones de voz 104 para recibir datos que representan el habla relacionada con la comunicación de voz, un servicio de análisis del habla 106 para analizar los datos para generar una pluralidad de conjuntos de datos de contenidos de audio para la comunicación de voz, un servicio de índices 108 para indexar la pluralidad de conjuntos de datos de contenidos de audio, y un servicio de búsqueda 110 para recibir uno o más términos de búsqueda del dispositivo de usuario final 102 y buscar los conjuntos de datos de contenidos de audio indexados para localizar una o más comunicaciones de voz que contienen el o los términos de búsqueda. Las comunicaciones de voz localizadas por el servicio de búsqueda 110 pueden suministrarse entonces al dispositivo de usuario final 102 donde se muestran al usuario. En ciertas realizaciones se pueden utilizar diferentes dispositivos de usuario final para hacer las comunicaciones y, posteriormente, buscar las comunicaciones.

El dispositivo de usuario final 102 permite al usuario realizar o recibir comunicaciones de voz sobre una red de telecomunicaciones 112. El término "comunicación de voz" se utiliza aquí para indicar cualquier comunicación que utilice el habla y está destinada a incluir, pero no limitarse a, las llamadas telefónicas, llamadas de teleconferencia, llamadas de vídeo, y mensajes de voz. Como será evidente, cuando las técnicas aquí descritas se aplican a las llamadas de vídeo, se aplican a la parte de audio de las comunicaciones.

El dispositivo de usuario final 102 puede ser un teléfono inteligente o puede ser un sistema informático configurado para proporcionar funciones de telefonía a un usuario (por ejemplo, un 'softphone' proporcionado por una aplicación que se ejecuta en un ordenador personal (PC) o dispositivo informático portátil, o a través de un navegador web en el dispositivo de este tipo). Las funciones de telefonía suelen proporcionarse por las aplicaciones de comunicaciones que se ejecutan en el dispositivo de usuario final 102 que pueden ser aplicaciones instaladas por terceros o aplicaciones nativas proporcionadas como parte del sistema operativo del dispositivo. Por ejemplo, el dispositivo de usuario final 102 puede proporcionarse con un cliente de voz sobre IP (VoIP) para realizar llamadas VoIP. Como se apreciará, se pueden utilizar varias combinaciones de tipos de telefonía conocidas en la técnica (por ejemplo, la telefonía convencional para las llamadas VoIP).

5 La red de telecomunicaciones 112 puede ser cualquier red, o la combinación de redes, capaz de establecer una comunicación de voz entre los usuarios finales. Por ejemplo, la red de telecomunicaciones 112 puede ser una red telefónica pública conmutada (PSTN), una red de telefonía móvil, una red de datos (como Internet), una red de VoIP o cualquier combinación de las mismas.

10 Cuando una comunicación de voz se hace hacia o desde un dispositivo de usuario final 102 que está suscrito al servicio correspondiente, el servidor de comunicación de voz 104 recibe datos representativos del habla de la comunicación de voz de la red de telecomunicaciones 112. Una vez que el servidor de comunicaciones de voz 104 ha recibido los datos reenvía los datos al servicio de análisis del habla 106 para su análisis.

15 En general, el servidor de comunicación de voz 104 recibe las señales de audio para la comunicación de voz desde la red de telecomunicaciones 112. En algunos casos, el servidor de comunicaciones de voz 104 reenvía las señales de audio al servicio de análisis del habla 106 para que el servicio de análisis del habla 106 pueda realizar el análisis en tiempo real. En otros casos, el servidor de comunicaciones de voz 104 registra las señales de audio recibidas como archivos de datos de audio y envía los archivos de datos de audio al servicio de análisis del habla 106. El servidor de comunicación de voz 104 puede configurarse para obtener el permiso del usuario y de cualquier otra parte en la comunicación de voz antes de grabar las señales de audio recibidas. Para permitir un análisis independiente de las partes cada lado de una comunicación de voz puede grabarse por separado.

20 Una copia de los datos grabados puede archivarse también en un dispositivo de almacenamiento, tal como la base de datos 114, para referencia futura. Cuando éste sea el caso, los datos también se almacenan en la base de datos 114 para identificar la comunicación de voz y vincularla con el usuario apropiado. También puede almacenarse en la base de datos 114 más información sobre la comunicación de voz, por ejemplo la temporización y los participantes.

25 Dependiendo de la configuración del usuario, los datos relativos a la totalidad o sólo un subconjunto de las comunicaciones de voz del usuario pueden reenviarse al motor de análisis del habla 106. Por ejemplo, el usuario puede configurar el sistema 100 para analizar únicamente comunicaciones de voz salientes, únicamente comunicaciones de voz entrantes o únicamente comunicaciones de voz a y/o de un usuario específico o grupo de usuarios. Además, el sistema 100 puede configurarse para analizar únicamente un lado de una comunicación de voz.

30 El servicio de análisis del habla 106 recibe los datos de voz generados por el servidor de comunicaciones de voz 104 y analiza los datos para generar una pluralidad de conjuntos de datos de contenidos de audio para la comunicación de voz. El término "conjunto de datos de contenidos de audio" se utiliza aquí para indicar cualquier dato que representa el contenido de una comunicación de voz y está destinado a incluir, pero no se limita a, una transcripción textual, un entramado de fonemas, un entramado de palabras, y una lista de palabras clave. El término "conjunto de datos de contenidos de audio" no está destinado a incluir datos que representan la forma de onda de audio.

35 Preferentemente, el servicio de análisis del habla 106 comprende una pluralidad de motores de análisis del habla y cada motor de análisis del habla utiliza una técnica diferente para analizar los datos para generar un conjunto de datos de contenidos de audio. Por ejemplo, el servicio de análisis del habla 106 puede comprender un primer motor de análisis del habla configurado para usar una primera técnica para generar un primer conjunto de datos de contenidos de audio, un segundo motor de análisis del habla configurado para utilizar una segunda técnica para generar un segundo conjunto de datos de contenidos de audio y un tercer motor de análisis del habla configurado para utilizar una tercera técnica para generar un tercer conjunto de datos de contenidos de audio.

40 El primer motor de análisis del habla puede generar el primer conjunto de datos de contenidos de audio empleando una técnica de conversión voz-a-texto de modo que el primer conjunto de datos de contenidos de audio es una transcripción textual que representa el texto más coincidente, posiblemente complementada con las mejores coincidencias siguientes. El segundo motor de análisis del habla 204 puede generar el segundo conjunto de datos de contenidos de audio empleando una técnica de entramado de fonemas de modo que el segundo conjunto de datos de contenidos de audio es un entramado de fonemas. El tercer motor de análisis del habla 206 puede generar un tercer conjunto de datos de contenidos de audio empleando una técnica de identificación de palabras clave de modo que el tercer conjunto de datos de contenidos de audio es una transcripción textual de palabras clave de tipo anotadas (es decir, palabras clave y una categoría) en la comunicación de voz. Cada una de estas técnicas se describe con referencia a la figura 2. Un ejemplo de servicio de análisis del habla 106 también se describirá con referencia a la figura 2.

45 Una vez que el servicio de análisis del habla 106 ha generado la pluralidad de conjuntos de datos de contenidos de audio, los conjuntos de datos de contenidos de audio se proporcionan al servicio de índices 108 para indexar. El servicio de análisis del habla 106 también puede archivar los conjuntos de datos de contenidos de audio en un dispositivo de almacenamiento, tal como una base de datos 114, por lo que se pueden ver y/o analizar en una fecha

posterior.

El servicio de índices 108 recibe los conjuntos de datos de contenidos de audio generados por el servicio de análisis del habla 106 e indexa cada conjunto de datos de contenidos de audio para permitir una búsqueda rápida. Los conjuntos de datos de contenidos de audio pueden indexarse empleando cualquier técnica de indexación conocida tal como RDBMS (sistema de gestión de base de datos relacional). El servicio de índices 108 almacena entonces los conjuntos de datos de contenidos de audio y los índices asociados en su memoria. El servicio de índices 108 también puede archivar los conjuntos de datos de contenidos de audio indexados en un dispositivo de almacenamiento, tal como una base de datos 114, por lo que se pueden ver y/o analizar en una fecha posterior.

El servicio de búsqueda 110 está en comunicación con el servicio de índices 108 y el dispositivo de usuario final 102 para permitir a un usuario realizar búsquedas de texto en comunicaciones de voz archivadas. Por ejemplo, una aplicación o navegador web en el dispositivo de usuario final 102 puede permitir al usuario especificar criterios de búsqueda que se transmiten al servicio de búsqueda 110. El servicio de búsqueda 110 busca entonces cada conjunto de datos de contenidos de audio indexado generado por el servicio de índices 108 para localizar comunicaciones de voz que coinciden con los criterios de búsqueda.

Típicamente los criterios de búsqueda comprenden una o más palabras claves y el servicio de búsqueda 110 busca los conjuntos de datos de contenidos de audio indexados para localizar comunicaciones de voz que contienen la palabra o palabras clave especificadas. Junto con, o en lugar de, las palabras claves, los criterios de búsqueda puede incluir restricciones sobre el tipo de palabra clave que se están buscando, por ejemplo, un criterio puede ser buscar solamente nombres de personas. Otro ejemplo de conjunto criterios puede ser: ("John", solamente nombres de personas).

Cuando el conjunto de datos de contenidos de audio es un entramado de fonemas generados por una técnica de entramado de fonemas, buscar el entramado de fonemas indexado puede comprender primero convertir las palabras claves especificadas en representaciones de fonemas y a continuación utilizar las representaciones de fonemas de las palabras clave buscar los entramados de fonemas indexados. La representación de fonema puede convertirse adicionalmente en un entramado de fonemas antes de la búsqueda.

La información que identifica las comunicaciones de voz localizadas por el servicio de búsqueda 110 se entrega a continuación al dispositivo de usuario final 102. La información que identifica las comunicaciones de voz puede incluir, pero no se limita a, la fecha de la comunicación de voz, el número de teléfono o identificación (por ejemplo, ID de Skype™) de la otra parte, el nombre de la otra parte, el tipo de comunicación de voz (por ejemplo, llamada telefónica, llamada de video, llamada de teleconferencia, mensaje de voz). El servicio de búsqueda 110 también puede proporcionar al dispositivo del usuario final 102 información que identifique la comunicación de voz grabada asociada a cada comunicación de voz localizada de manera que el usuario puede recuperar, transmitir y/o reproducir la comunicación de voz grabada.

En algunos casos, las comunicaciones de voz coincidentes se clasifican o se ordenan antes de proporcionarlas al dispositivo de usuario final 102. Por ejemplo, para cada comunicación de voz coincidente, se puede asignar una puntuación de coincidencia. Se puede utilizar cualquier método conocido para generar una puntuación de coincidencia para buscar resultados mediante el servicio de búsqueda 110. Por ejemplo, la puntuación de coincidencia puede basarse en cuántas veces se encuentran las palabras clave en el conjunto de datos de contenidos de audio indexado.

Entonces se puede calcular una suma ponderada de las puntuaciones de coincidencia para cada comunicación de voz localizada por cada búsqueda. La ponderación puede basarse en el conjunto de datos de contenidos de audio particular que se utilizó para identificar la comunicación de voz. Por ejemplo, el servicio de búsqueda 110 puede localizar una comunicación de voz particular dos veces o una vez empleando el conjunto de datos de contenidos de audio voz-a-texto (por ejemplo transcripción textual) y una vez empleando el entramado de conjunto de datos de contenidos de audio de fonemas (por ejemplo entramado de fonemas). A continuación se asignarían dos puntuaciones de coincidencia a la comunicación de voz, una basada en el conjunto de datos de contenidos de audio voz-a-texto (por ejemplo transcripción textual) y una basada en el conjunto de datos de contenidos de audio de entramado de fonemas (por ejemplo entramado de fonemas). A continuación se genera una suma ponderada para la comunicación de voz ponderando las dos puntuaciones de coincidencia basadas en el conjunto de datos de contenidos de audio particular. Por ejemplo, el peso asignado a una puntuación de coincidencia para un conjunto de datos de contenidos de audio particular puede basarse en la precisión del conjunto de datos de contenidos de audio. Por ejemplo, la técnica de palabra clave puede ser la más precisa y se le da la más alta ponderación, y el voz-a-texto ser la menos precisa y se le da la menor ponderación. La ponderación se puede ajustar para diferentes entornos de sonido, lugares, ruido de fondo, e idiomas.

Entonces los resultados de la búsqueda pueden ordenarse basándose en la suma ponderada. Por ejemplo, los resultados de búsqueda pueden ordenarse en orden ascendente o descendente basándose en las sumas

ponderadas.

El dispositivo de usuario final 102 recibe la información que identifica las comunicaciones de voz localizadas desde el servicio de búsqueda 110 y envía la información de identificación al usuario. Por ejemplo, la información de identificación se puede mostrar en una interfaz de usuario de búsqueda en un dispositivo de visualización del dispositivo de usuario final 102. Tal como se ha descrito anteriormente, la información de identificación puede incluir, pero no se limita a, la fecha de la comunicación de voz, el número de teléfono o identificación (por ejemplo, ID de Skype) de la otra parte, el nombre de la otra parte, el tipo de comunicación de voz (por ejemplo, llamada telefónica, llamada de video, llamada de teleconferencia y mensaje de voz). Adicionalmente, la información de identificación puede incluir pistas de por qué la comunicación de voz encontrada ha sido producida en los resultados de la búsqueda, por ejemplo, cuáles de las palabras clave de entrada se encontraron en la comunicación de voz y que tipos, si los había, estaban asociados con esas palabras claves. La interfaz de usuario de búsqueda puede permitir al usuario llevar a cabo una o más acciones en las comunicaciones de voz localizadas. Por ejemplo, la interfaz de usuario de búsqueda puede permitir al usuario ordenar las comunicaciones de voz localizadas. La interfaz de usuario de búsqueda también puede permitir al usuario recuperar, transmitir y/o reproducir las comunicaciones de voz localizadas. Un ejemplo de interfaz de usuario de búsqueda para un dispositivo de usuario final 102 se describe con referencia a la figura 3.

Además, es posible ajustar la ponderación basándose en la retroalimentación del usuario, basándose en sus acciones, mediante la recopilación de una señal explícita (por ejemplo, permitiendo a los usuarios votar coincidencias precisas/no precisas) o las implícitas, observando, sobre qué coincidencias elige actuar un usuario.

El término "servicio" se utiliza para indicar un programa o sistema que proporciona ciertas funciones. Por ejemplo, un servicio de índice puede proporcionar la función de indexación de los datos. El servicio puede adoptar la forma de un programa que se ejecuta en un sistema informático único, o puede proporcionarse por una pluralidad de sistemas informáticos configurados para funcionar conjuntamente para proporcionar las funciones. Las funciones pueden ser visitadas y utilizada a través de un sistema de red, por ejemplo Internet. Cualquier servicio puede implementarse de cualquier manera conocida por la persona experta en la técnica. Aunque los diversos servicios se han descrito por separado, uno o más de los servicios pueden proporcionarse como parte de un único servicio, o por un único programa de un sistema informático.

Cuando en este documento se ha utilizado el término "conectado", no se pretende requerir una conexión permanente, siempre activa. Más bien se utiliza en el sentido de que las entidades conectadas están conectadas, cuando se requiera, para intercambiar datos. Por ejemplo, dos entidades se conectarían mediante la transmisión de datos desde una entidad a otra a través de una red IP.

Se hace referencia ahora a la figura 2 que ilustra un diagrama de bloques de un ejemplo de servicio de análisis del habla 106 para generar una pluralidad de conjuntos de datos de contenidos de audio. Tal como se ha descrito anteriormente, cada conjunto de datos de contenidos de audio representa el contenido de la comunicación de voz asociada. El servicio de análisis del habla 106 de la figura 2 comprende tres motores de análisis del habla 202, 204, y 206 para la generación de los tres conjuntos de datos de contenidos de audio. Cada motor de análisis del habla 202, 204 y 206 utiliza una técnica diferente para generar un conjunto de datos de contenidos de audio.

El primer motor de análisis del habla 202 utiliza una técnica de conversión voz a texto para generar un primer conjunto de datos de contenidos de audio. En una técnica de conversión voz a texto se utiliza un sistema de reconocimiento automático de voz (ASR) para generar una transcripción textual de la comunicación de voz. La transcripción textual típicamente comprende las mejores coincidencias de texto. También puede incluir las siguientes mejores coincidencias para frases específicas. Por lo tanto en el servicio de análisis del habla 106 de la figura 3 el primer conjunto de datos de contenidos de audio es textual. A continuación se puede realizar una búsqueda textual en la transcripción. La técnica de conversión voz a texto se enfrenta a dos problemas principales. En primer lugar, los sistemas de ASR típicamente producen transcripciones con un cierto porcentaje de palabras erróneamente transcritas. Estos errores pueden deberse a una fluctuación de un hablante, la articulación o acento. Los errores pueden deberse también a factores externos tales como el ruido de fondo. Cuanto más difíciles sean las condiciones acústicas mayor es la tasa de error.

En segundo lugar, los sistemas ASR típicamente utilizan un diccionario finito de palabras para generar la transcripción que limita las palabras que pueden ser transcritas. Por ejemplo, un diccionario ASR puede no incluir los nombres propios, palabras poco comunes, y los términos específicos de la industria. En consecuencia, los sistemas de conversión de voz a texto no pueden tratar adecuadamente con palabras que no están en el diccionario. Estas palabras se conocen como palabras fuera del vocabulario (OOV). Los efectos de este segundo problema se discuten en P.C. Woodland, S.E. Johnson, P. Jurlin y K. Sparck Jones, "Effects of out of Vocabulary Words in Spoken Document Retrieval", en Proc. SIGIR 2000.

El segundo motor de análisis del habla 204 utiliza una técnica de entramado de fonemas para generar un segundo

conjunto de datos de contenidos de audio. En una técnica de entramado de fonemas se genera un entramado de fonemas para la comunicación de voz. Cada nodo del entramado está asociado a un punto en el tiempo durante el audio, y a cada borde se le asigna una hipótesis de fonema y una puntuación que representa la probabilidad de esta hipótesis. De esta forma, se puede utilizar el entramado para el almacenamiento de múltiples secuencias modelo
 5 potenciales. Por lo tanto, en el servicio de análisis del habla 106 de la figura 3 el segundo conjunto de datos de contenidos de audio es un entramado de fonemas. Una vez que se ha generado el entramado, la búsqueda de palabras clave se realiza haciendo coincidir la pronunciación palabra frente al entramado.

Un ejemplo de técnica de entramado de fonemas se describe en D.A. James, S.J. Young, "A fast lattice-based approach to vocabulary independent wordspotting", *icassp*, vol. 1, pág. 377-380, Acoustics, Speech, and Signal Processing, 1994. ICASSP-94 Vol 1., 1994 IEEE International Conference on, 1994. En James et al. el entramado de fonemas se genera a partir de un reconocedor de voz de Viterbi modificado basado en el Paradigma de Paso de Testigo.
 10

La técnica de coincidencia de fonemas ofrece varias ventajas sobre la técnica de conversión de voz a texto, incluyendo su capacidad para encontrar palabras que no están en el diccionario ASR (las palabras OOV) y su capacidad para encontrar coincidencias parciales entre la palabra o palabras clave de búsqueda y la comunicación de voz. En general, sin embargo, la coincidencia de fonemas presenta problemas de velocidad.
 15

El tercer motor de análisis del habla 206 utiliza una técnica de identificación de palabras clave para generar un tercer conjunto de datos de contenidos de audio. En una técnica de identificación de palabras claves se identifican palabras y/o frases en el habla y a continuación se transcriben en texto las palabras y frases identificadas. Las palabras relevantes se pueden generar a partir de una lista de palabras clave estática, una lista de palabras clave dinámica, una lista de patrones, tales como fechas, horas o expresiones regulares, o cualquier combinación de los mismos. Por lo tanto en el servicio de análisis del habla 106 de la figura 3 el tercer conjunto de datos de contenidos de audio es un conjunto de palabras clave transcritas. Se describen ejemplos de métodos para la identificación de palabras clave en un documento del habla en la Solicitud de Patente Española N.º P201130858 y en la Solicitud de Patente Internacional No. PCT/EP2012/059832.
 20
 25

Mientras que cada técnica ofrece ciertas ventajas, hay situaciones en las que cada técnica por sí sola no produce resultados precisos y fiables. Mediante la combinación de los resultados de estas tres técnicas pueden obtenerse resultados más exactos y fiables. En otras realizaciones, el servicio de análisis del habla puede comprender motores de análisis de voz adicionales para generar conjuntos de datos de contenidos de audio adicionales empleando diferentes técnicas. Por ejemplo, el servicio de análisis del habla puede comprender un cuarto motor de análisis del habla que genera un cuarto conjunto de datos de contenidos de audio empleando una técnica de coincidencia de palabras. La técnica de coincidencia de palabras es similar a la técnica de coincidencia de fonemas en el hecho de que se genera un entramado que representa el contenido de la comunicación de voz. Sin embargo, en lugar de generar un entramado de fonemas, se genera un entramado de palabras. La palabra entramado representa todas las secuencias de palabras posibles que pueden obtenerse a partir de la comunicación de voz.
 30
 35
 40

Los sistemas de coincidencia de palabras proporcionan una ventaja sobre los sistemas de conversión de voz a texto porque proporcionan una pluralidad de coincidencias posibles de palabras coincidentes en lugar de solamente la mejor coincidencia. El sistema de coincidencia de palabras puede superar a los sistemas de coincidencia de entramado en términos de velocidad, pero dado que los sistemas de coincidencia de palabras se basan en un diccionario fijo, no pueden tratar adecuadamente con palabras OOV.
 45

Se hace referencia ahora a la figura 3 que ilustra una interfaz de usuario de búsqueda 302 para un dispositivo de usuario final 102. La interfaz de usuario de búsqueda 302 se divide en dos secciones, una sección de búsqueda 304 y una sección de resultados de búsqueda 306. La sección de búsqueda de 304 permite a un usuario iniciar una búsqueda de las comunicaciones de voz archivadas relacionadas con ese usuario. Por ejemplo, en la figura 3, la sección de búsqueda 304 comprende un campo de criterios de entrada de búsqueda 308 y un botón de activación de búsqueda 310. El campo de criterios de entrada de búsqueda 308 permite al usuario introducir los criterios de una búsqueda. Típicamente, los criterios de búsqueda comprenden una o más palabras clave. Por ejemplo, en la figura 3, el usuario ha introducido la búsqueda de palabras clave "musical". El botón de activación de búsqueda 310 permite al usuario iniciar o activar la búsqueda usando los criterios especificados. Por ejemplo, en la figura 3, al pulsar el botón "IR" 310 se inicia una búsqueda para las comunicaciones de voz que contienen la palabra "musical". El usuario podría realizar una nueva búsqueda mediante la introducción de nuevas palabras clave en el campo de criterios de entrada de búsqueda 308 y a continuación presionando el botón de activación de búsqueda 310. Será evidente para una persona experta en la técnica que la sección de búsqueda 304 puede tener otros componentes y configuraciones, por ejemplo los necesarios para introducir criterios para filtrar palabras clave por tipo, o comunicaciones de voz por el llamante.
 50
 55
 60

La sección de resultados de búsqueda 306 muestra los resultados de la búsqueda. Por ejemplo, en la figura 3, la sección de resultados de búsqueda 306 tiene cuatro columnas de información 312, 314, 316, y 318 que muestran

información acerca de los resultados de búsqueda. En concreto, la primera columna 312 muestra información acerca de cuándo se hizo la comunicación de voz. Esto puede incluir información como la fecha y hora. La segunda columna 314 muestra información sobre el número o la identificación de la otra parte. Por ejemplo, cuando la comunicación de voz es una llamada telefónica la segunda columna 314 puede mostrar el número de teléfono de la otra parte. Sin embargo, cuando la comunicación de voz es una llamada de Skype™, la segunda columna 314 puede mostrar el ID de usuario de Skype™ de la otra parte. La tercera columna de 316 identifica a la otra parte con, por ejemplo, un nombre (es decir, "Marie") o una etiqueta (es decir, "Hijo" o "Jefe"). Esta información puede obtenerse de la lista de contactos del usuario, por ejemplo. La cuarta columna 318 muestra la información sobre el tipo de comunicación por voz (por ejemplo, llamada telefónica, llamada de vídeo, llamada de teleconferencia o mensaje de voz). Será evidente para una persona experta en la materia que la información sobre los resultados de búsqueda se puede mostrar en cualquier forma adecuada y que la sección de resultados de búsqueda 306 de la figura 3 es meramente un ejemplo de una forma adecuada de mostrar la información. También será evidente que otra información acerca de las comunicaciones de voz, tales como la longitud de la comunicación de voz, se puede mostrar además o en lugar de la información que se muestra en las cuatro columnas de información 312, 314, 316 y 318 de la figura 3.

La sección de resultados de búsqueda 306 también puede configurarse para permitir al usuario realizar una o más acciones sobre los resultados de búsqueda. Por ejemplo, en la figura 3, la sección de resultados de búsqueda 306 tiene una columna de acción 320 que permite a un usuario reproducir las comunicaciones de voz. En concreto, se proporciona un botón de reproducción para cada comunicación de voz. En algunos casos, después de que un usuario haya pulsado el botón de reproducción para una comunicación de voz particular, el dispositivo de usuario final 102 envía una solicitud a la base de datos 114 para el audio grabado para la comunicación de voz seleccionada. En respuesta, la base de datos 114 proporciona una copia del audio grabado para la comunicación de voz al dispositivo de usuario final 102. El audio grabado puede suministrarse al dispositivo de usuario final 102, por ejemplo, como un flujo de audio o un archivo de audio completo. Será evidente para un experto en la materia que la sección de resultados de búsqueda 306 puede permitir al usuario realizar otras acciones, tales como reenviar la comunicación de voz, además de o en lugar de permitir al usuario reproducir la comunicación de voz.

En la figura 3, la búsqueda de las comunicaciones de voz que contienen la palabra "musical" ha dado tres resultados. El primer resultado de búsqueda 322 es una llamada que se hizo el 12 de octubre 2009, a/desde el jefe del usuario (123-44-55). El segundo resultado de la búsqueda 324 es la llamada que se hizo el 9 de noviembre de 2009, a/de Marie (234-11-22). El tercer resultado de búsqueda 326 es una llamada que se realizó hoy a/desde el hijo del usuario (234-55-66). El segundo resultado de la búsqueda 324 puede contener, por ejemplo, el audio siguiente:

"Me voy a Londres, ¿cuál era el nombre de ese musical que le encantó a Marie?"

"Realmente me encantó el musical Chicago, de Bob Foss, ¡tienes que verlo! Puedes conseguir billetes baratos en Leicester Square".

Cuando se ordenan los resultados de la búsqueda mediante el servicio de búsqueda 110 antes de proporcionarlos al dispositivo de usuario final 102, la sección de resultados de búsqueda 306 puede mostrar los resultados de la búsqueda en orden o clasificados. Por ejemplo, el primer resultado de búsqueda 322 puede tener la clasificación más alta y los terceros resultados de la búsqueda 326 pueden tener la clasificación más baja.

Se hace referencia ahora a la figura 4 que ilustra un método 400 para generar una pluralidad de conjuntos de datos de contenidos de audio indexados para una comunicación de voz empleando el sistema de la figura 1. En la etapa 402, el servicio de análisis del habla 106 recibe datos que representan el habla de una comunicación de voz. Tal como se ha descrito anteriormente, la comunicación de voz puede ser, por ejemplo, una llamada telefónica, una llamada de teleconferencia, correo de voz, o una llamada de video. Los datos pueden ser una grabación de la comunicación de voz o un canal de audio en vivo.

En la etapa 404 un primer motor de análisis del habla 202 analiza los datos para generar un primer conjunto de datos de contenidos de audio que representa el contenido de la comunicación de voz. Tal como se ha descrito anteriormente, el primer motor de análisis del habla puede generar el primer conjunto de datos de contenidos de audio empleando una técnica de conversión voz-a-texto de modo que el primer conjunto de datos de contenidos de audio es una transcripción textual que representa el texto que mejor coincide.

En la etapa 406, un segundo motor de análisis del habla 204 analiza los datos para generar un segundo conjunto de datos de contenidos de audio que representa el contenido de la comunicación de voz. Tal como se ha descrito anteriormente, el segundo motor de análisis del habla 204 puede generar el segundo conjunto de datos de contenidos de audio empleando una técnica de entramado de fonemas de modo que el segundo conjunto de datos de contenidos de audio es un entramado de fonemas.

En la etapa 408, un tercer motor de análisis del habla 206 analiza los datos para generar un tercer conjunto de datos

de contenidos de audio que representa el contenido de la comunicación de voz. Tal como se ha descrito anteriormente, el tercer motor de análisis del habla 206 puede generar un tercer conjunto de datos de contenidos de audio empleando una técnica de identificación de palabras clave de modo que el tercer conjunto de datos de contenidos de audio es una transcripción textual de palabra claves en la comunicación de voz.

5 Será evidente para una persona experta en la técnica que las etapas 404 a 408 del método 400 pueden realizarse en cualquier orden o simultáneamente.

10 En la etapa 410, el servicio de índices 108 indexa y almacena cada uno de los conjuntos de datos de contenidos de audio para la recuperación eficaz. Tal como se ha descrito anteriormente, los conjuntos de datos de contenidos de audio pueden indexarse empleando cualquier técnica conocida tal como RDBMS.

15 Se hace referencia ahora a la figura 5 que ilustra un método 500 para realizar consultas textuales en comunicaciones de voz empleando el sistema de la figura 1. En la etapa 502 el servicio de búsqueda 110 recibe los criterios de búsqueda del usuario. Los criterios de búsqueda pueden suministrarse al servicio de búsqueda 110 desde un dispositivo de usuario final 102 a través de una interfaz de usuario de búsqueda, tal como una interfaz de usuario de búsqueda 302. Tal como se ha descrito anteriormente, los criterios de búsqueda típicamente comprenden una o más palabra clave. Los criterios de búsqueda también pueden comprender información que identifica al usuario de modo que solo se buscan las comunicaciones de voz del usuario. Los criterios de búsqueda también pueden comprender otra información que se utiliza para limitar adicionalmente la búsqueda, tales como fechas, horas, número/ID y/o nombre de la parte.

25 En la etapa 504, el servicio de búsqueda 110 busca cada uno de los conjuntos de datos de contenidos de audio indexados producidos por el servicio de índices 108 para encontrar comunicaciones de voz que coinciden con los criterios de búsqueda. Buscar los conjuntos de datos de contenidos de audio indexados puede comprender la traducción de las palabras clave de búsqueda en representaciones fonéticas y el uso de las representaciones fonéticas para buscar en los índices. En algunos casos, las representaciones fonéticas pueden traducirse adicionalmente a un entramado de fonemas antes de la búsqueda. Como se señaló anteriormente, los criterios de búsqueda pueden comprender información de identificación del usuario. Esta información puede utilizarse por el servicio de búsqueda 110 para limitar la búsqueda a aquellas comunicaciones de voz que están asociadas con el usuario específico. Si los criterios de búsqueda comprenden otra información de limitación de búsqueda (por ejemplo, fechas, tiempos, etc.) esto también puede utilizarse para limitar la búsqueda.

35 En la etapa 506, el servicio de búsqueda 110 asigna a cada resultado de búsqueda (cada comunicación de voz localizada por la búsqueda en la etapa 504) una puntuación de coincidencia basándose en una comparación de la comunicación de voz y los criterios de búsqueda (por ejemplo lo bien que coincide la comunicación de voz con los criterios de búsqueda). Cualquier método conocido para generar una puntuación de coincidencia para resultados de búsqueda puede utilizarse por el servicio de búsqueda 110. Por ejemplo, la puntuación de coincidencia puede basarse en cuántas veces las palabras clave se encuentran en el correspondiente conjunto de datos de contenidos de audio.

45 En la etapa 508, el servicio de búsqueda 110 genera una suma ponderada de las puntuaciones de coincidencia para cada comunicación de voz localizada por la búsqueda. La ponderación puede basarse en el conjunto de datos de contenidos de audio particular que se utilizó para identificar la comunicación de voz.

50 Por ejemplo, cuando se utiliza el servicio de análisis del habla 106 de la figura 2, podría identificarse una comunicación de voz particular tres veces durante la búsqueda - una vez por el conjunto de datos de contenidos de audio voz-a-texto, una vez por el conjunto de datos de contenidos de audio de entramado de fonemas, y una vez por el conjunto de datos de contenidos de audio de identificación de palabras clave. Se podría entonces asignar a la comunicación de voz tres puntuaciones de coincidencia - una para el conjunto de datos de contenidos de audio voz-a-texto, una para el conjunto de datos de contenidos de audio de entramado de fonemas y una para el conjunto de datos de identificación de palabras clave.

55 Se genera entonces una suma ponderada de puntuaciones de coincidencia para la comunicación de voz. El peso asignado a una puntuación de coincidencia particular puede basarse en la precisión del conjunto de datos de contenidos de audio correspondiente. Por ejemplo, se puede asignar a una puntuación de coincidencia de entramado de fonemas un mayor peso de ponderación que una puntuación de coincidencia de voz-a-texto basándose en que la técnica de entramado de fonemas suele dar resultados más precisos que la técnica voz-a-texto.

60 En la etapa 510, el servicio de búsqueda 110 clasifica u ordena los resultados de búsqueda (las comunicaciones de voz localizadas por la búsqueda en la etapa 504) basándose en las sumas ponderadas generadas en la etapa 508. Por ejemplo, el servicio de búsqueda 110 puede ordenar los resultados de búsqueda (las comunicaciones de voz localizadas por la búsqueda en la etapa 504) en orden creciente o decreciente basado en las sumas ponderadas.

Pueden entonces proporcionarse los resultados de la búsqueda ordenados o clasificados al dispositivo del usuario final 102, donde se emiten al usuario. Por ejemplo, el dispositivo de usuario final 102 puede mostrar los resultados de la búsqueda en una interfaz de usuario de búsqueda, tal como la interfaz de usuario de búsqueda 302.

5 Se hace referencia ahora a la figura 6 que ilustra un ejemplo de dispositivo de usuario final 102 que puede implementarse como cualquier forma de dispositivo informático y/o electrónico, y en el que pueden implementarse las realizaciones de la descripción anterior.

10 El dispositivo de usuario final 102 comprende uno o más procesadores 602 que pueden ser microprocesadores, controladores o cualquier otro tipo adecuado de procesadores para procesar las instrucciones ejecutables por ordenador para controlar el funcionamiento del dispositivo con el fin de ejecutar las aplicaciones, tales como aplicaciones de comunicación y aplicaciones de análisis de elementos accionables a los que se refiere el presente documento. En algunos ejemplos, por ejemplo cuando se utiliza una arquitectura de sistema en un chip, los procesadores 602 pueden incluir uno o más bloques de función fija (también llamados aceleradores) que implementan una parte del método de funcionamiento del dispositivo de usuario final 102 o aplicaciones que se describen en este documento en hardware (en lugar de software o firmware). Se puede proporcionar un software de plataforma que incluye un sistema operativo 604 o cualquier otro software de plataforma adecuado en el dispositivo de usuario final 102 para permitir al software de aplicación 606 ejecutarse en el dispositivo.

20 Las instrucciones ejecutables por ordenador se pueden proporcionar utilizando cualquier medio legible por ordenador que sea accesible a un dispositivo de usuario final 102. Los medios legibles por ordenador pueden incluir, por ejemplo, medios de almacenamiento tales como memoria 608 y medios de comunicación. Los medios de almacenamiento informático, tal como la memoria 608, incluyen medios volátiles y no volátiles, extraíbles y no extraíbles implementados en cualquier método o tecnología para el almacenamiento de la información tales como instrucciones legibles por ordenador, estructuras de datos, módulos de programas u otros datos. Los medios de almacenamiento informáticos incluyen, aunque no se limitan a, RAM, ROM, EPROM, EEPROM, memoria flash u otras tecnologías de memoria, CD-ROM, discos versátiles digitales (DVD) u otro almacenamiento óptico, cassetes magnéticos, cinta magnética, almacenamiento en disco magnético u otros dispositivos de almacenamiento magnético, o cualquier otro medio de no transmisión que se puede utilizar para almacenar información para el acceso por un dispositivo informático. En contraste, los medios de comunicación pueden incorporar instrucciones legibles por ordenador, estructuras de datos, módulos de programa, u otros datos en una señal de datos modulada, tales como una onda portadora u otro mecanismo de transporte. Tal como se define en este documento, los medios de almacenamiento informático no incluyen medios de comunicación. Aunque los medios de almacenamiento informático (memoria 608) se muestran dentro del dispositivo del usuario final 102, se apreciará que el almacenamiento puede distribuirse o situarse de forma remota y accederse a través de una red u otro enlace de comunicación (por ejemplo, utilizando la interfaz de comunicación 610).

40 El dispositivo de usuario final 102 también comprende un controlador de entrada/salida 612 dispuesto para emitir información de visualización a un dispositivo de pantalla 614 que puede estar separado o integrado en el dispositivo de usuario final 102. La información de pantalla puede proporcionar una interfaz gráfica de usuario. El controlador de entrada/salida 612 también puede estar dispuesto para recibir y procesar la entrada desde uno o más dispositivos, tales como un dispositivo de entrada de usuario 616 (por ejemplo un ratón o un teclado). Esta entrada de usuario se puede utilizar para las aplicaciones de comunicaciones y de elementos accionables. En una realización, el dispositivo de visualización 614 también puede actuar como dispositivo de entrada de usuario 616 si se trata de un dispositivo de pantalla táctil. El dispositivo de usuario final 102 también puede estar provisto de otra funcionalidad como es conocido para estos dispositivos. Por ejemplo, la interfaz de comunicación 610 puede comprender una interfaz de radio a un teléfono móvil u otros sistemas de comunicación inalámbricos, y puede proporcionarse micrófono, altavoz y cámara para llamadas de voz y vídeo.

50 El término "ordenador" se utiliza aquí para referirse a cualquier dispositivo con capacidad de procesamiento de modo que pueda ejecutar instrucciones. Los expertos en la materia se darán cuenta de que tales capacidades de procesamiento se incorporan en muchos dispositivos diferentes y por lo tanto, el término "ordenador" incluye PC, servidores, teléfonos móviles, asistentes personales digitales y muchos otros dispositivos.

55 Los expertos en la técnica se darán cuenta de que los dispositivos de almacenamiento utilizados para almacenar instrucciones de programa pueden estar distribuidos a través de una red. Por ejemplo, un ordenador remoto puede almacenar un ejemplo del proceso descrito como software. Un ordenador local o terminal puede acceder al ordenador remoto y descargar una parte o la totalidad del software para ejecutar el programa. Como alternativa, el ordenador local puede descargar las piezas de software según sea necesario, o ejecutar algunas instrucciones software en el terminal local y algunas en el ordenador remoto (o red de ordenadores). Los expertos en la técnica también se darán cuenta de que mediante la utilización de técnicas convencionales conocidas por los expertos en la técnica que todo, o una parte de las instrucciones de software pueden llevarse a cabo por un circuito especializado, tal como un DSP, una matriz lógica programable, o similares.

Cualquier valor de rango o dispositivo dado en este documento puede ser ampliado o modificado sin perder el efecto pretendido, como será evidente para el experto.

- 5 Se comprenderá que los beneficios y ventajas descritos anteriormente pueden referirse a una realización o pueden referirse a varias realizaciones. Las realizaciones no están limitadas a las que resuelven cualquiera o todos los problemas indicados o aquellas que tienen uno o todos de los beneficios y ventajas indicados.

- 10 Cualquier referencia a 'un' elemento se refiere a uno o más de esos elementos. El término 'que comprende' se utiliza aquí para significar que incluye los bloques de procedimiento o elementos identificados, pero que tales bloques o elementos no comprenden una lista exclusiva y un método o aparato puede contener bloques o elementos adicionales.

- 15 Las etapas de los métodos descritos aquí pueden llevarse a cabo en cualquier orden adecuado, o simultáneamente, donde sea apropiado. Adicionalmente, los bloques individuales se pueden eliminar de cualquiera de los métodos sin apartarse del alcance de la materia objeto descrita en este documento. Aspectos de cualquiera de los ejemplos descritos anteriormente se pueden combinar con los aspectos de cualquiera de los otros ejemplos descritos para formar ejemplos adicionales, sin perder el efecto buscado.

- 20 Se comprenderá que la descripción anterior de una realización preferida se da a modo de ejemplo solamente y que pueden realizarse varias modificaciones por los expertos en la técnica. Aunque diversas realizaciones se han descrito anteriormente con un cierto grado de particularidad, o con referencia a una o más realizaciones individuales, los expertos en la técnica podrían hacer numerosas modificaciones a las realizaciones descritas sin apartarse del alcance de esta invención.

REIVINDICACIONES

1. Un sistema para realizar consultas textuales en comunicaciones de voz, comprendiendo el sistema:

5 un servicio de índices (108) para almacenar una pluralidad de conjuntos de datos de contenidos de audio para una pluralidad de comunicaciones de voz, comprendiendo la pluralidad de conjuntos de datos de contenidos de audio al menos dos conjuntos de datos de contenidos de audio para cada comunicación de voz, seleccionándose los al menos dos conjuntos de datos de contenidos de audio desde un conjunto de datos de contenidos de audio voz-a-texto generado usando una técnica de conversión voz-a-texto,
 10 un conjunto de datos de contenidos de audio de fonemas generado usando una técnica de entramado de fonemas, y un conjunto de datos de contenidos de audio de palabras clave generado usando una técnica de identificación de palabras clave; y
 un motor de búsqueda (110) configurado para:

15 recibir criterios de búsqueda desde un usuario, comprendiendo los criterios de búsqueda al menos una palabra clave;
 buscar cada uno de los al menos dos conjuntos de datos de contenidos de audio para al menos una porción de la pluralidad de comunicaciones de voz para identificar comunicaciones de voz que coinciden con los criterios de búsqueda; y
 20 combinar las comunicaciones de voz identificadas por cada búsqueda para producir una lista combinada de comunicaciones de voz identificadas,

dicho sistema estando **caracterizado por que** combinar las comunicaciones de voz identificadas por cada búsqueda comprende:

25 asignar a cada comunicación de voz identificada una puntuación de coincidencia basándose en una comparación de la comunicación de voz y los criterios de búsqueda; y
 generar para cada comunicación de voz identificada una suma ponderada de las puntuaciones de coincidencia para la comunicación de voz, estando basada la ponderación de cada puntuación de coincidencia en el conjunto de datos de contenidos de audio particular que identifican la comunicación de voz.

2. El sistema de acuerdo con la reivindicación 1, en el que el peso de cada puntuación de coincidencia se actualiza utilizando retroalimentación de búsquedas anteriores, dicha retroalimentación estando basada en dichas acciones de usuario, recopilando una señal explícita o unas implícitas, observando, sobre cuáles de las múltiples coincidencias el usuario elige actuar.

3. El sistema de acuerdo con la reivindicación 1 o 2, en el que el motor de búsqueda (110) está además configurado para clasificar la lista combinada de comunicaciones de voz identificadas a partir de las sumas ponderadas.

4. El sistema de acuerdo con cualquier reivindicación anterior, que comprende además:

un primer motor de análisis del habla (202) configurado para recibir datos que representan el habla de una comunicación de voz y analizar los datos utilizando la técnica voz-a-texto para generar el conjunto de datos de contenidos de audio voz-a-texto para esta comunicación de voz;
 45 un segundo motor de análisis del habla (204) configurado para recibir los datos y analizar los datos usando la técnica de entramado de fonemas para generar el conjunto de datos de contenidos de audio de fonemas para esta comunicación de voz; y
 un tercer motor de análisis del habla (206) configurado para recibir los datos y analizar los datos usando la técnica de identificación de palabras clave para generar el tercer conjunto de datos de contenidos de audio para esta comunicación de voz.

5. El sistema de acuerdo con la reivindicación 3, en el que el servicio de índices está además configurado para indexar los conjuntos de datos de contenidos de audio para generar conjuntos de datos de contenidos de audio indexados; y buscar los conjuntos de datos de contenidos de audio comprende buscar los conjuntos de datos de contenidos de audio indexados.

6. El sistema de acuerdo con cualquier reivindicación anterior, que comprende además un dispositivo de usuario final (102) configurado para recibir los criterios de búsqueda del usuario, transmitir los criterios de búsqueda del dispositivo de usuario final (102) al servicio de búsqueda, recibir la lista combinada de comunicaciones de voz identificadas del servicio de búsqueda (110), y emitir la lista combinada de comunicaciones de voz identificadas al usuario.

7. El sistema de acuerdo con cualquier reivindicación anterior, que comprende además una base de datos (114) configurada para almacenar datos para cada comunicación de voz de la pluralidad de comunicaciones de voz,

representando los datos el habla de la comunicación de voz.

5 8. El sistema de acuerdo con la reivindicación 7, en el que la base de datos (114) está además configurada para recibir una solicitud de un usuario para datos para una comunicación de voz identificada particular y para proporcionar los datos al usuario.

9. Un método implementado por ordenador para realizar consultas textuales en comunicaciones de voz, comprendiendo el método:

10 almacenar una pluralidad de conjuntos de datos de contenidos de audio usando un servicio de índices (108) para una pluralidad de comunicaciones de voz, comprendiendo la pluralidad de conjuntos de datos de contenidos de audio al menos dos conjuntos de datos de contenidos de audio para cada comunicación de voz, seleccionándose los al menos dos conjuntos de datos de contenidos de audio desde un conjunto de datos de contenidos de audio voz-a-texto generado usando una técnica de conversión voz-a-texto, un conjunto de datos de contenidos de audio de fonemas generado usando una técnica de entramado de fonemas, y un conjunto de datos de contenidos de audio de palabras clave generado usando una técnica de identificación de palabras clave; y
 15 recibir en un servicio de búsqueda (110) criterios de búsqueda desde un usuario, comprendiendo los criterios de búsqueda al menos una palabra clave;
 buscar usando el servicio de búsqueda cada uno de los al menos dos conjuntos de datos de contenidos de audio para al menos una porción de comunicaciones de voz para identificar comunicaciones de voz que coinciden con los criterios de búsqueda; y
 20 combinar las comunicaciones de voz identificadas por cada búsqueda para producir una lista combinada de comunicaciones de voz identificadas,
 el método estando **caracterizado por que** combinar las comunicaciones de voz identificadas por cada búsqueda comprende:

asignar a cada comunicación de voz identificada una puntuación de coincidencia basándose en una comparación de la comunicación de voz y los criterios de búsqueda; y
 30 generar para cada comunicación de voz identificada una suma ponderada de las puntuaciones de coincidencia para la comunicación de voz, estando el peso de cada puntuación de coincidencia basado en el conjunto de datos de contenidos de audio particular a partir del cual la comunicación de voz se identificó.

35 10. El método de acuerdo con la reivindicación 9, que comprende además clasificar usando el motor de búsqueda (110) la lista combinada de comunicaciones de voz identificadas basándose en las sumas ponderadas.

11. El método de acuerdo con una cualquiera de las reivindicaciones 9 a 10, que comprende además recibir en un primer motor de análisis del habla (202) datos que representan el habla de una comunicación de voz y analizar los datos usando la técnica voz-a-texto para generar el primer conjunto de datos de contenidos de audio para esta comunicación de voz;
 40 recibir en un segundo motor de análisis del habla (204) los datos y analizar los datos usando la técnica de entramado de fonemas para generar el segundo conjunto de datos de contenidos de audio para esta comunicación de voz; y
 recibir en un tercer motor de análisis del habla (206) los datos y analizar los datos usando la técnica de identificación de palabras clave para generar el segundo conjunto de datos de contenidos de audio para esta comunicación de voz.
 45 voz.

12. El método de acuerdo con una cualquiera de las reivindicaciones 9 a 11, que comprende además:

50 transmitir la lista combinada de comunicaciones de voz identificadas a un dispositivo de usuario final (102); y emitir al usuario usando el dispositivo de usuario final (102) la lista combinada de comunicaciones de voz identificadas.

13. El método de acuerdo con la reivindicación 12, que comprende además:

55 recibir en el dispositivo de usuario final (102) la entrada desde un usuario que selecciona una de las comunicaciones de voz identificadas;
 transmitir una solicitud desde el dispositivo de usuario final (102) a un dispositivo de almacenamiento que solicita datos que representan el habla de la comunicación de voz seleccionada; y
 60 recibir en el dispositivo de usuario final (102) los datos.

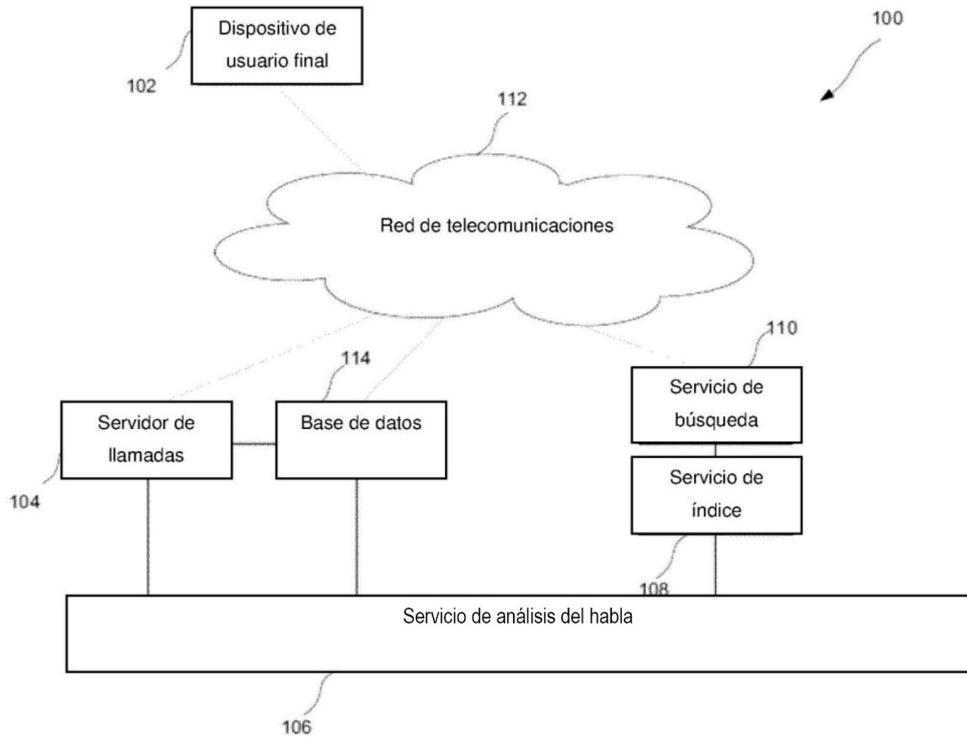


Figura 1

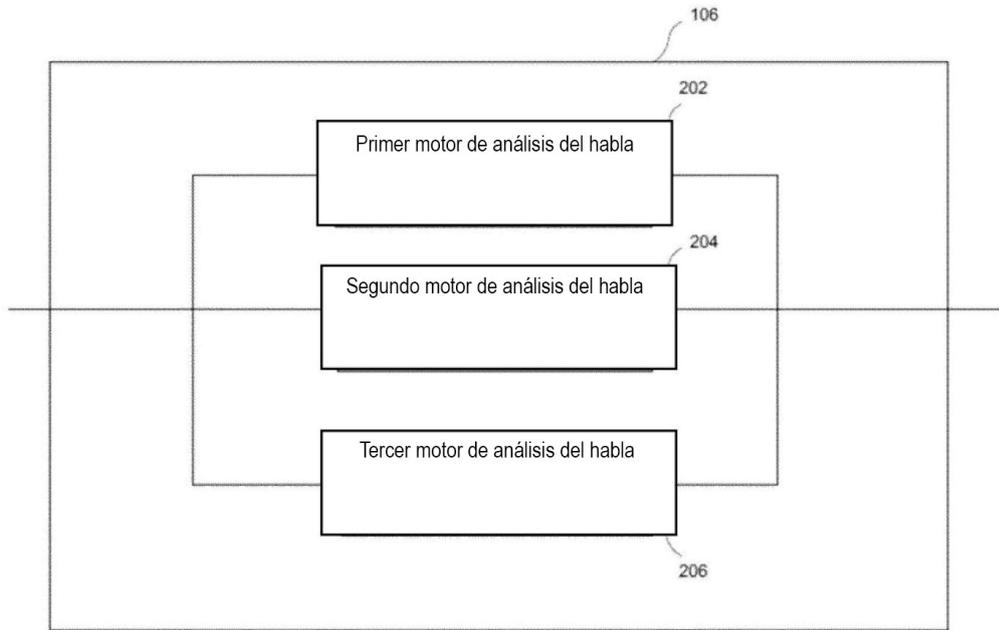


Figura 2

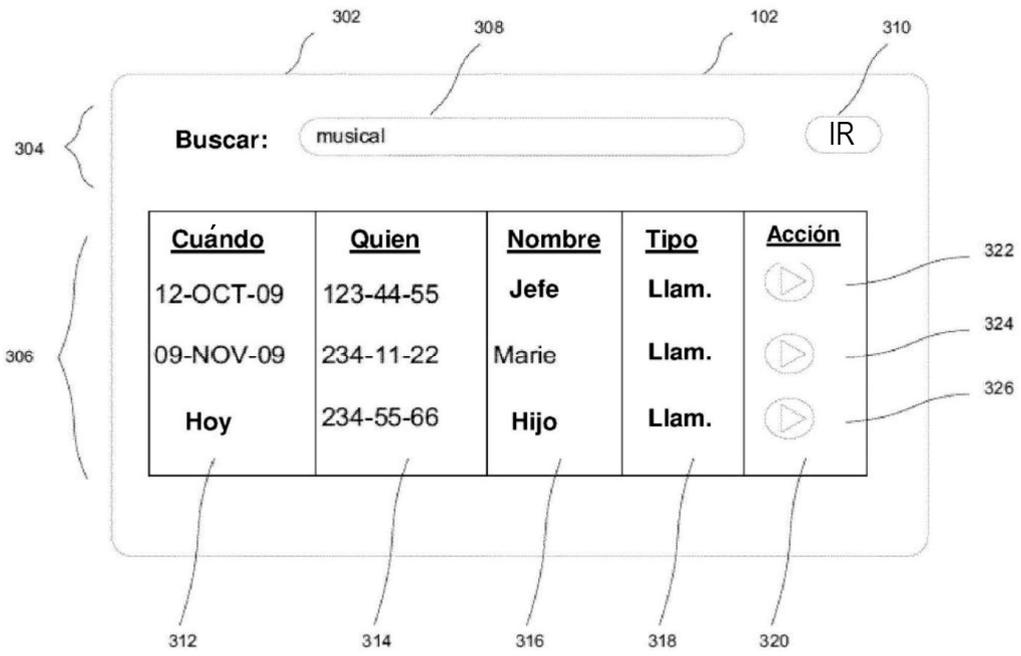


Figura 3

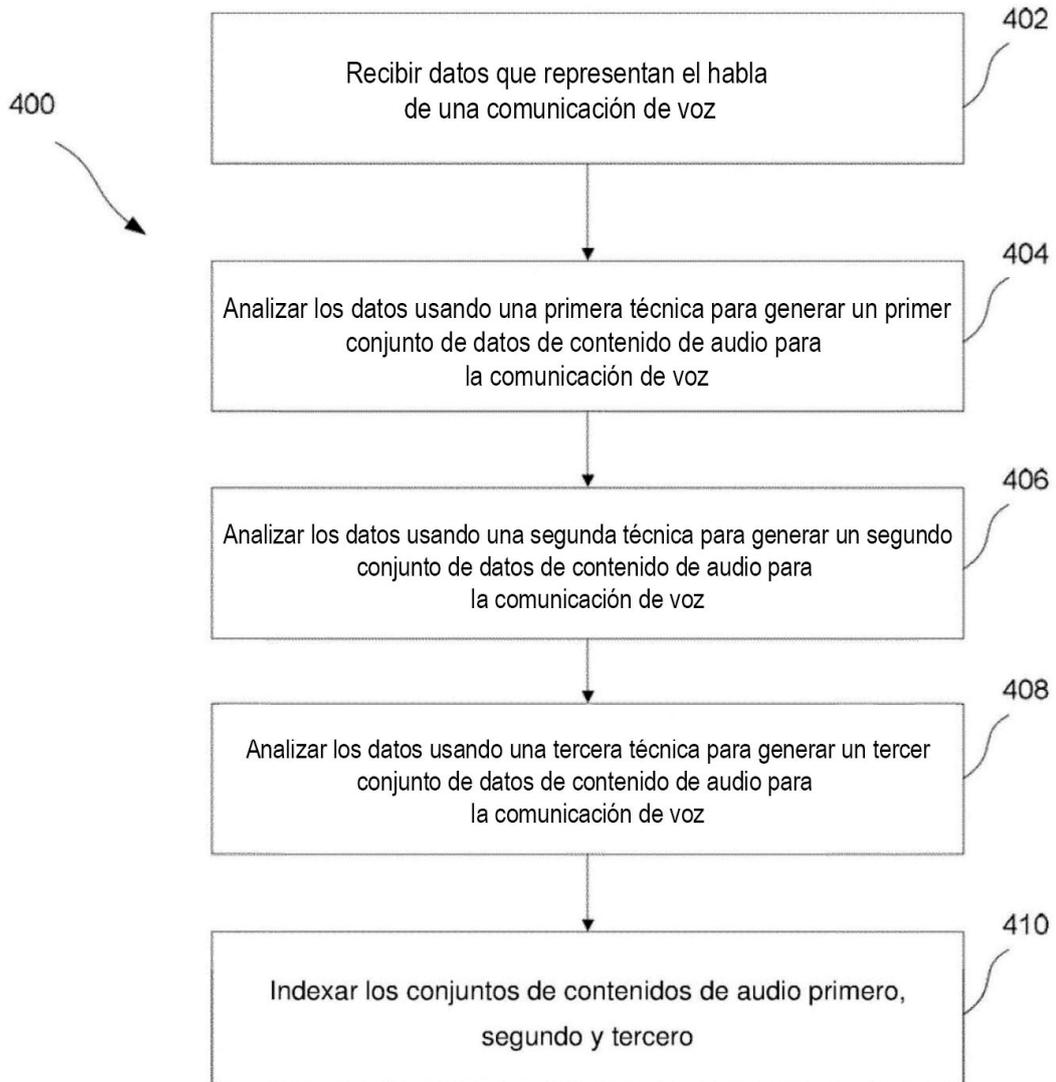


Figura 4

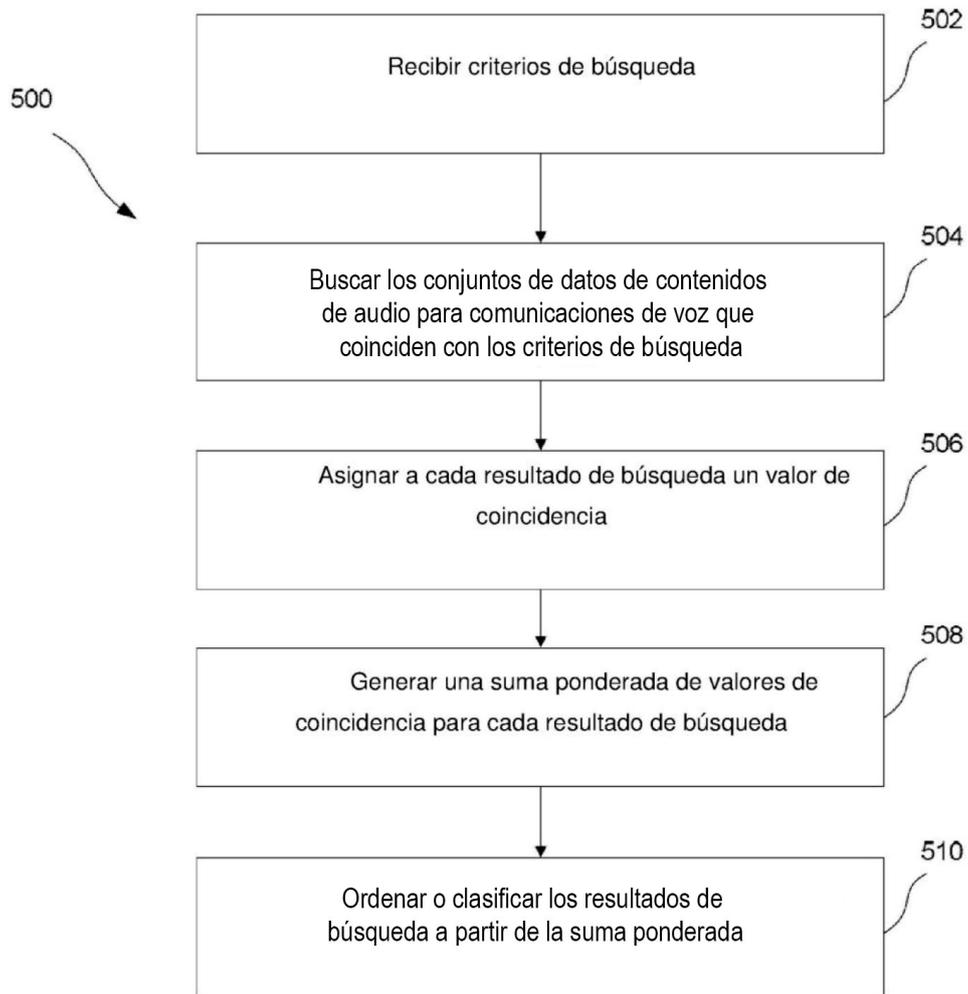


Figura 5

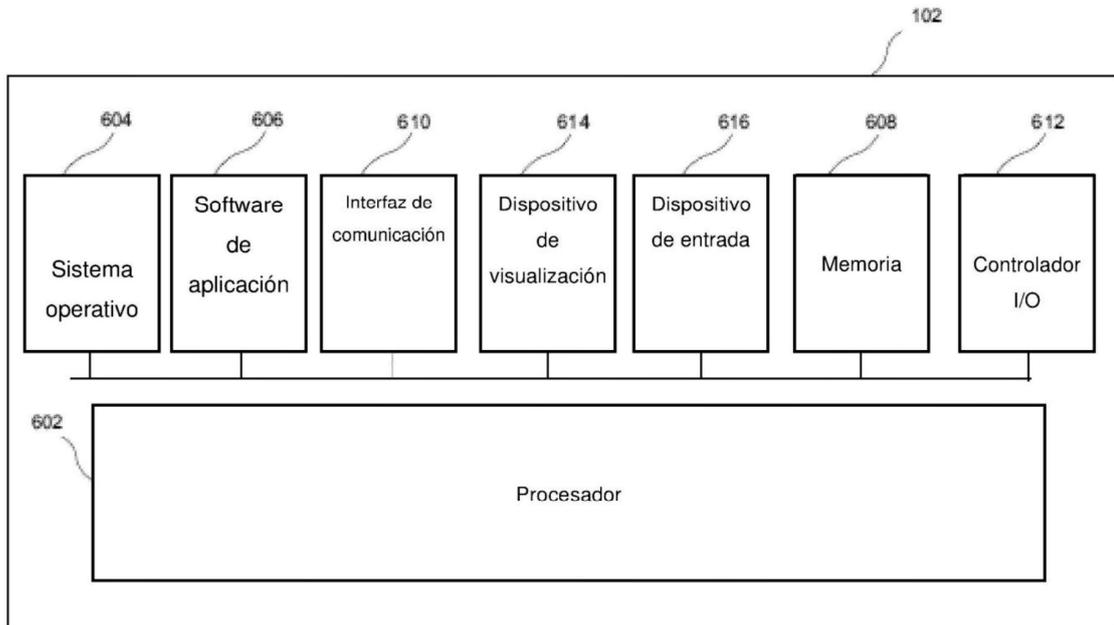


Figura 6