

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 668 969**

51 Int. Cl.:

G10L 15/06 (2013.01)

G10L 15/18 (2013.01)

G10L 17/04 (2013.01)

G10L 17/26 (2013.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **17.06.2009 PCT/IL2009/000604**

87 Fecha y número de publicación internacional: **23.12.2009 WO09153788**

96 Fecha de presentación y número de la solicitud europea: **17.06.2009 E 09766323 (1)**

97 Fecha y número de publicación de la concesión europea: **24.01.2018 EP 2304718**

54 Título: **Creación de una base de datos de referencia de parámetros de habla para clasificar expresiones del habla**

30 Prioridad:

17.06.2008 US 140413

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

23.05.2018

73 Titular/es:

**VOICESENSE LTD. (100.0%)
P.O.Box 8844 8 Hamachshev street
42507 Netanya, IL**

72 Inventor/es:

**DEGANI, YOAV y
ZAMIR, YISHAI**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 668 969 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Creación de una base de datos de referencia de parámetros de habla para clasificar expresiones del habla

La presente invención se refiere al campo del análisis del habla, y más particularmente, al uso de rasgos prosódicos del habla para determinar las características actuales de comportamiento, psicológicas y de estilo de habla de una persona en una situación específica.

Antecedentes de la invención

Análisis del habla es un término general dado a los métodos computarizados que procesan las expresiones del habla humana con el fin de descubrir la información que transportan. El análisis del habla se clasifica como parte de la disciplina fonética dentro de las ciencias lingüísticas.

El análisis del habla se puede dividir en dos planteamientos principales: el primer planteamiento se centra en revelar el contenido del habla aprendiendo cómo se pronuncian las palabras, las sílabas y los fonemas y cómo se ordenan las frases. Muchas de las aplicaciones de reconocimiento de voz, tales como 'habla a texto' o 'detección de palabras', usan este planteamiento con el fin de extraer el contenido del habla.

El segundo planteamiento, el análisis de la prosodia del habla, se centra en la manera en que se dice el discurso, analizando los rasgos no segmentarios (sin palabras, sin contenido) del habla tales como la entonación, el tempo, la intensidad, el énfasis y el ritmo.

La manera en que se dice el discurso proporciona la "melodía" del habla, que añade una visión significativa del significado y contexto generales del habla. Por ejemplo, las personas perciben una frase como una frase interrogativa por la entonación ascendente al final de la frase. Los acentos son un ejemplo clásico de cómo los parámetros del habla prosódica alteran la pronunciación de las palabras. Los actores modifican la prosodia de su habla para que suene como un cierto personaje. También hay diferencias de género en la prosodia, por ejemplo, las mujeres típicamente tienen frecuencias más altas que los hombres. Muchas veces, los rasgos prosódicos cambian el significado del habla: interpretamos la misma frase como que tiene un significado cínico, sarcástico o simple por los cambios en la manera (prosodia) en que se pronuncia.

Las características prosódicas de las expresiones del habla también transportan información acerca del estado emocional del orador. Esto ha sido demostrado a través de los años en muchos trabajos en la bibliografía. También se entiende intuitivamente –una persona excitada que llama a un servicio de emergencia típicamente tendría un ritmo de habla rápido, su voz sería intensa, entrecortada, fluctuante, etc. Por otra parte, una persona en un estado triste, melancólico tendería a hablar lentamente, con escasa energía, pausas largas y similares. Estas características se reflejan en los rasgos prosódicos del habla.

Se han publicado distintas patentes con respecto a usar rasgos prosódicos del habla con el fin de analizar automáticamente el estado emocional de los oradores. Entre ellas, la Patente de EE.UU. Nº 6151571 de Pertrushin, la Patente de EE.UU. Nº 6173260 de Slaney, para clasificar diferentes estados emocionales, y la Patente Europea Nº EP 1423846 de Degani y Zamir, para determinar la excitación emocional en general.

Se hicieron pocos intentos para correlacionar entre la personalidad del orador y las características prosódicas del habla. La Patente de EE.UU. Nº 6006188 de Bogdashevsky describe un método para determinar los rasgos del habla de personas con tipos de personalidad similares (según inventarios psicológicos conocidos), y luego usando los rasgos detectados para clasificación automática de tipos de personalidad. La idea de que la personalidad se refleja en el habla tiene sentido y también se puede entender intuitivamente: uno puede imaginar, por ejemplo, los patrones del habla suave y vacilante de una persona introvertida en oposición a los patrones de habla fuerte e impulsiva de una persona extrovertida. Vincular los patrones de personalidad y de habla refleja bien la comprensión de que el habla expresa amplios aspectos de las características personales. No obstante, la debilidad de este concepto radica en su implementación práctica. La personalidad representa características estables de una persona; por lo tanto, también se debería medir más bien de manera consistente. Esto significa que los patrones del habla del orador, que reflejan su personalidad, deberían permanecer bastante consistentes durante situaciones cambiantes, estados internos variables y diferentes contextos. Este no es el caso en realidad: los patrones del habla tienden a estar fuertemente afectados por factores situacionales, como es evidente, por ejemplo, a partir de la relación firme y probada entre los estados emocionales y la prosodia del habla. Incluso la persona introvertida del ejemplo mencionado anteriormente se enfada de vez en cuando, y cuando lo hace, sus patrones del habla (y sus parámetros prosódicos del habla) cambian significativamente y se asemejan mucho más a los patrones del habla de una persona abierta, extrovertida. Quizás estadísticamente, si midiéramos los patrones del habla de la persona introvertida en muchas ocasiones diferentes, habría una correlación significativa entre su personalidad y sus patrones del habla. También puede ser fiable si muestreamos el habla de una persona bajo condiciones muy similares a las condiciones en las que se tomaron los datos de referencia, que representan los patrones del habla de una cierta personalidad. Pero este no sería el caso cuando sea autoritario un factor situacional interviniente. En la vida real, los factores situacionales están afectando frecuentemente al habla. Por lo tanto, la medición de la personalidad a través de la prosodia del habla no se puede percibir como un método independiente de la situación.

Breve compendio

De esta manera, la presente invención proporciona un método implementado por ordenador de creación de una base de datos de referencia de parámetros del habla para clasificar las expresiones del habla según distintas características de comportamiento, psicológicas y de estilos de habla, como se expone en la reivindicación 1, así como un programa de ordenador correspondiente y un aparato, tal como se expone en las reivindicaciones 2 y 3.

Breve descripción de los dibujos

La materia objeto considerada como la invención, como se define por las reivindicaciones 1-3 adjuntas, llegará a ser comprendida más claramente a la luz de la descripción que sigue de las realizaciones en la presente memoria, dadas a modo de ejemplo y con propósitos de discusión ilustrativa de la presente invención solamente, con referencia a los dibujos anexos (Figuras, o simplemente "FIG."), en donde:

la FIG. 1 es un diagrama de flujo de alto nivel que muestra las etapas de obtención y formación de los datos de referencia de parámetros de habla para clasificar diferentes características de comportamiento, psicológicas y de estilos de habla, según el método de la presente invención;

la FIG. 2 es un diagrama de flujo de alto nivel que muestra las etapas de análisis de expresiones del habla actuales de un orador y la determinación de las características de comportamiento, psicológicas y de estilo de habla del orador en la situación dada mientras que se usa la base de datos de referencia;

la FIG. 3 es un diagrama de flujo de alto nivel que muestra algunas de las etapas opcionales que se pueden llevar a cabo además del método de la presente invención;

la FIG. 4 es un diagrama de bloques esquemático de un sistema de procesamiento de datos; y

la FIG. 5 es un diagrama de bloques esquemático de un aparato.

Los dibujos junto con la descripción hacen evidente para los expertos en la técnica cómo se puede incorporar en la práctica la invención.

Descripción detallada de la invención

Antes de exponer la invención en detalle, puede ser útil exponer definiciones de ciertos términos que se usarán en lo sucesivo.

El término "situacional" como se usa en la presente memoria significa características pertenecientes o relevantes al estado actual del orador. Por el contrario, una caracterización situacional puede no ser válida para el orador en general. Por ejemplo, una caracterización situacional de un orador puede ser: "El orador suena asertivo" (ahora), o "El orador suena reservado" o "El orador suena calculador" y similares. Esto no supone decir que el orador es, por naturaleza, asertivo, reservado o calculador en general, sino más bien que el orador suena de esa forma en la expresión del habla analizada actual. Claramente, el mismo orador puede sonar de manera diferente en diferentes situaciones.

El término "contextual" como se usa en la presente memoria en esta solicitud se define como que representa características generales de la situación (que pueden afectar al habla) que se derivan, atribuyen o influyen o bien por los antecedentes biográficos del orador, que comprenden factores tales como cultura, idioma, acento, edad y similares, o bien por atributos especiales del establecimiento de la situación, tales como conversación o monólogo, texto escrito leído en alto o habla libre, condiciones normales de habla o habla durante la actividad física, durante la actuación en una obra, mientras que da una conferencia, mientras que da un discurso, bajo situaciones estresantes – combate, emergencia, juego de pelota y similares.

El término "estilos de habla" como se usa en la presente memoria significa patrones de habla típicos, que representan agrupaciones de características de comportamiento y psicológicas, que tienden a ser usados en ciertas situaciones. Por ejemplo, estilo de habla expresivo, estilo de habla calculador, estilo de habla directo y similares. Las personas tienden a cambiar sus estilos de habla según la situación y el contexto específicos.

El término "sistema de comunicación" como se usa en la presente memoria pertenece a cualquier sistema de comunicación de voz que permite la transferencia de expresiones de voz. Estos sistemas pueden ser, por ejemplo, sistemas de telefonía fija o inalámbrica, red de comunicación celular, sistemas de comunicación basados en web, otros sistemas de comunicación basados en voz sobre IP, sistemas de centro de llamadas y similares.

En la siguiente descripción detallada, se exponen muchos detalles específicos con el fin de proporcionar una comprensión minuciosa de la invención. No obstante, se entiende que las enseñanzas de la presente invención se pueden poner en práctica sin estos detalles específicos. En otros casos, métodos, procedimientos, componentes y circuitos bien conocidos no se han descrito en detalle para no oscurecer las enseñanzas de la presente invención.

La presente invención se basa en el hecho de que los patrones del habla reflejan un amplio espectro de características de comportamiento y psicológicas del orador más allá de sólo los estados emocionales. Uno puede pensar que muchas de tales características de comportamiento o psicológicas que no son emociones, sino que se expresan a través de patrones del habla: por ejemplo, franqueza, asertividad, sensibilidad, sinceridad, expresividad, ser vacilante, reservado, autoritario, controlado, calculador, paciente, impaciente y empático, y así sucesivamente, etc. La presente invención está dirigida a la capacidad de analizar estas características de comportamiento y psicológicas y similares a partir de expresiones del habla.

Según las realizaciones de la invención, la implementación de la presente invención tiene en cuenta que las reflexiones de estas características en los patrones del habla son principalmente de naturaleza situacional más que de una naturaleza estable y constante. Algunas veces, tal característica de comportamiento estaría relacionada con el contexto de la situación, algunas veces con el estado emocional del orador, algunas veces con la personalidad del orador y muchas veces sería una combinación de todos los factores. Por ejemplo, un patrón del habla directa y asertiva puede caracterizar una conversación de entrenador-jugador estándar en medio de un partido (contexto) o puede estar relacionado con la indignación de un cliente insatisfecho en una conversación con el servicio al cliente (emoción). Aún en otros casos puede reflejar una personalidad asertiva general del orador. Los parámetros de prosodia del habla (que reflejan el ritmo, entonación, intensidad y similares) pueden indicar que actualmente el habla suena asertiva, pero no pueden señalar su contexto específico ni diferenciar si está relacionado con la situación o la personalidad.

Por lo tanto, estas son las dos bases principales de la invención actual: las características prosódicas del habla reflejan una amplia gama de características de comportamiento y psicológicas del orador; y la manifestación de estas características en los patrones del habla tiene una fuerte naturaleza situacional, es decir, es indicativa principalmente del estado actual del orador.

Por lo tanto, es el objeto de la presente invención proporcionar una base de datos de referencia de parámetros del habla para un método general, independiente del orador, para determinar las características actuales de comportamiento y psicológicas de una persona analizando los rasgos del habla de la persona en una situación específica.

Ventajosamente, el método independiente del orador antes mencionado se puede usar para proporcionar realimentación a personas de sus características de comportamiento y psicológicas en diferentes situaciones, como parte de diversas aplicaciones de formación con propósitos vocacionales, educativos, terapéuticos y recreativos. También se puede usar para proporcionar una visión a otras personas de cómo comunicarse mejor con la persona analizada, en base a la comprensión de las características actuales de comportamiento y psicológicas de la persona analizada para diversas aplicaciones comerciales, así como con propósitos vocacionales, educativos, terapéuticos y recreativos.

Uno de los desafíos abordados por la presente invención es extraer y detectar parámetros indicativos de las características de comportamiento y psicológicas, que son independientes del orador. Las personas difieren significativamente en sus características prosódicas individuales del habla – tono típico, ritmo, entonación, etc. También pueden diferir en la forma individual en que expresan ciertas características de comportamiento y psicológicas. Por ejemplo, uno puede expresar asertividad principalmente cambiando la intensidad del habla y otro alterando el ritmo del habla, etc. Alcanzar un procesamiento independiente del orador requiere superar estas diferencias individuales en la expresión prosódica de las características de comportamiento y psicológicas.

La presente invención logra un procesamiento independiente del orador mediante el uso de extracciones del habla secundarias únicas de los parámetros del habla de amplitud y tono primario, para el análisis de los rasgos prosódicos. Estos parámetros secundarios son una variedad de características de indicadores estadísticos de los segmentos de tono ascendente, tono descendente y tono igual dentro de la expresión del habla. Estando menos afectados por las diferencias individuales del habla, estos parámetros secundarios únicos son, por lo tanto, más indicativos de las características de comportamiento y psicológicas por sí mismos.

Un aspecto adicional de la presente invención son los estilos del habla. Cuando se examinan empíricamente los patrones del habla de muchas características de comportamiento y psicológicas, parece evidente que distintas agrupaciones de tales características se pueden agrupar juntas para formar estilos del habla diferenciados. Esto llega a ser evidente ya cuando se escuchan las expresiones del habla y se clasifican: ciertas características de comportamiento y psicológicas tienden a aparecer con frecuencia juntas dentro de las mismas expresiones (misma situación), implicando que pueden estar relacionadas. Además, cuando se analizan sus rasgos prosódicos, ciertas características de comportamiento y psicológicas son más similares entre sí en comparación con otras, implicando de nuevo su relación.

Los siguientes son unos pocos ejemplos de tales estilos de habla potenciales. Un estilo se puede denominar estilo de habla 'pasivo' o 'estable', comprendiendo características de comportamiento y psicológicas tales como lentitud, quietud, estabilidad, calma, indecisión, etc. Otro estilo se puede denominar 'comunicativo', comprendiendo características como sensibilidad, sinceridad, locuacidad, expresividad, etc. Otro estilo se puede denominar 'asertivo', que comprende franqueza, autoridad, ritmo rápido, volumen, actividad, etc. Otro estilo aún se puede

denominar 'calculador', que comprende características tales como ser analítico, no emocional, controlado, reservado, sistemático, etc. Estos estilos de habla se traen aquí solamente a modo de ejemplo. Naturalmente, otros diversos estilos de habla se pueden representar mediante el uso de los rasgos prosódicos.

5 Es importante enfatizar que estos estilos de habla tienen una fuerte naturaleza situacional también. Aunque de vez en cuando pueden representar una característica más estable del orador, muchas veces representarían un patrón actual de interacción, o un estado interno actual del orador pertinente para la situación o contexto específico. A menudo, la misma persona cambiaría entre tales estilos de habla en diferentes contextos.

10 La determinación de tales estilos de habla a través del análisis del habla puede ser muy ventajosa para las aplicaciones potenciales mencionadas anteriormente de la presente invención, en la medida que puede proporcionar una visión más exhaustiva de la situación analizada.

Por lo tanto, también es el objeto de la presente invención proporcionar una base de datos de referencia de parámetros del habla para un método general, independiente del orador para determinar los estilos del habla actuales de una persona analizando los rasgos prosódicos del habla de la persona en una situación específica.

15 Las realizaciones de la presente invención permiten aprovecharse de tres aspectos principales que las diferencian de los métodos tradicionales de análisis prosódico del habla. El primer aspecto es ampliar significativamente el alcance de las características personales de un orador que se determinan a través del análisis prosódico del habla. Una amplia variedad de características de comportamiento, psicológicas y de estilo de habla se demandan que sean determinadas por el análisis, más allá del estado emocional del orador, que fue la principal aplicación demostrada en el pasado. El segundo aspecto diferenciador es el vínculo entre los atributos actuales, situacionales y de comportamiento de un orador y sus características de habla correspondientes. El tercer aspecto diferenciador es el uso de extracciones secundarias únicas de los parámetros de frecuencia y amplitud fundamentales del habla, para lograr un procesamiento independiente del orador. El uso de estos parámetros secundarios únicos y sus combinaciones permite superar las diferencias individuales en la expresión del habla, de manera que aspectos de comportamiento y psicológicos similares se representan de manera similar en las características del habla de diferentes personas.

25 La FIG. 1 ilustra un diagrama de flujo de alto nivel que muestra los pasos del método de creación de una base de datos de referencia de parámetros del habla para clasificar diversas características de comportamiento, psicológicas y de estilos de habla, según algunas realizaciones de la invención. El método comprende: seleccionar manualmente un cierto contexto de habla, por ejemplo, el establecimiento de un cierto idioma, cultura y situación única 1010; seleccionar manualmente las características de comportamiento, psicológicas y de estilo de habla a ser analizadas en el contexto seleccionado 1020; obtener una pluralidad de expresiones del habla de personas en el contexto seleccionado 1030; y agrupar manualmente las expresiones del habla en grupos que representen características similares de comportamiento, psicológicas y de estilos de habla 1040. Esto se realiza usando métodos de evaluación válidos y conocidos, tales como clasificaciones de expertos, clasificación por pares, autoevaluaciones, pruebas objetivas y similares, o cualquier otro método de evaluación usado y conocido en las ciencias sociales y del comportamiento. Tales características situacionales pueden ser, a modo de ejemplo solamente, franqueza, asertividad, sensibilidad, sinceridad, expresividad, ser vacilante, reservado, autoritario, controlador, calculador, paciente, empático, o cualquier otra característica de comportamiento y psicológica que se pueda expresar o reflejar en los patrones del habla. Tal estilo de habla puede ser, a modo de ejemplo solamente, un estilo de habla 'pasivo' y 'estable', comprendiendo varias características de comportamiento y psicológicas tales como lentitud, quietud, estabilidad, calma, indecisión, etc., o un estilo de habla 'comunicativo' comprendiendo características tales como sensibilidad, sinceridad, locuacidad, expresividad, etc., o un estilo de habla 'asertivo' comprendiendo franqueza, autoridad, ritmo rápido, volumen, actividad, etc., o un estilo de habla 'calculador' comprendiendo características tales como ser analítico, no emocional, controlador, reservado, sistemático, etc. Según algunas realizaciones de la invención, la base de datos puede contener parámetros del habla representativos de cualquier número (uno o más) de características de comportamiento, psicológicas y de estilos de habla. El método también incluye: procesamiento previo de cada una de las expresiones del habla en dichos grupos de expresiones que representan características similares de comportamiento, psicológicas y de estilos de habla, en segmentos de habla silenciosos y activos 1050; dividir los segmentos de habla activos en cadenas de bloques de igual longitud, teniendo dichos bloques parámetros del habla primarios incluyendo los parámetros de tono y amplitud 1060; derivar una pluralidad de parámetros del habla secundarios a partir de los parámetros del habla primarios, siendo dichos parámetros secundarios seleccionados de la suma de segmentos por unidad de tiempo de tendencias de tono igual, tono ascendente y tono descendente en dichas cadenas de bloques y las longitudes promedio de segmentos de tono igual, tono ascendente y tono descendente en dichas cadenas de bloques 1070; y determinar parámetros del habla secundarios únicos, combinaciones de parámetros y valores de parámetros que son comunes a cada grupo de expresiones y representan las características típicas de comportamiento, psicológicas o de estilos de habla de cada grupo, a partir de dichos parámetros secundarios derivados 1080. Ha de ser entendido que el uso de una base de datos es meramente una realización de la presente invención y que el conjunto o el diccionario de parámetros del habla y sus características de comportamiento correspondientes en un contexto específico se pueden almacenar juntos de cualquier manera que representen los parámetros del habla para referencia. La base de datos o el conjunto de parámetros se pueden almacenar por diferentes medios o ser usados para referencia por un programa de ordenador sin el uso de una base de datos de software real.

La FIG. 2 ilustra un diagrama de flujo de alto nivel que muestra los pasos del método implementado por ordenador de análisis de expresiones del habla de un orador en una situación y contexto dados y determinar las características de comportamiento, psicológicas y de estilo de habla del orador en la situación dada.

5 El método comprende: crear una base de datos de referencia de parámetros del habla para clasificar las expresiones del habla según diversas características de comportamiento, psicológicas y de estilo de habla del orador 2010; obtener expresiones del habla de un orador en una situación y contexto específicos 2020; procesar previamente las expresiones obtenidas en segmentos de habla silenciosos y activos y dividir los segmentos de habla activos en cadenas de bloques de igual longitud, teniendo dichos bloques parámetros del habla primarios que incluyen parámetros de tono y amplitud 2030; derivar una pluralidad de parámetros del habla secundarios a partir de los parámetros primarios, comprendiendo dichos parámetros secundarios la suma de segmentos por unidad de tiempo de tendencias de tono igual, tono ascendente y tono decreciente en las cadenas de bloques y las longitudes promedio de segmentos con tendencias de tono igual, tono ascendente y tono decreciente en dichas cadenas de bloques 2040; calcular parámetros de habla únicos, combinaciones de parámetros y valores de parámetros representativos de las características situacionales de comportamiento, psicológicas y de estilos de habla, a partir de dichos parámetros secundarios en la expresión del habla 2050; determinar y calificar las características situacionales de comportamiento, psicológicas y de estilo de habla en la expresión del habla comparando los parámetros calculados con la base de datos de referencia predefinida de parámetros del habla 2060; y emitir los resultados del análisis 2070.

20 La etapa de derivación de una pluralidad de parámetros del habla secundarios se dirige además a parámetros seleccionados de una lista que comprende: longitud promedio de pausa, suma de pausas por unidad de tiempo, longitud promedio de silencios cortos, suma de silencios cortos por unidad de tiempo, longitud promedio de segmentos de tono igual, suma de segmentos de tono iguales por unidad de tiempo, longitud promedio de segmentos de tono ascendente, suma de segmentos de tono ascendente por unidad de tiempo, longitud promedio de segmentos de tono decreciente, suma de segmentos de tono descendente por unidad de tiempo, dispersión de amplitud promedio dentro de segmentos de tono igual, dispersión de amplitud promedio dentro de segmentos de tono ascendente, dispersión de amplitud promedio dentro de segmentos de tono descendente, variación de tono y rango por unidad de tiempo, pendiente de tono promedio por unidad de tiempo, patrones de forma espectral y envolvente espectral de la señal de habla en el tiempo, patrones de orden secuencial de tendencias de tono ascendente, descendente e igual en el tiempo.

30 La FIG. 3 ilustra un diagrama de flujo de alto nivel que muestra pasos opcionales del método implementado por ordenador de análisis de las expresiones del habla de un orador en una situación y contexto dados y determinación de las características de comportamiento, psicológicas y de estilo de habla del orador en la situación dada.

35 El método de análisis puede comprender además la etapa de inclusión de muestras de habla obtenidas previamente del orador analizado actualmente dentro de la base de datos de referencia de parámetros de habla para el reconocimiento de los patrones de habla únicos del orador, de manera que el análisis actual del orador sería dependiente del orador más que independiente del orador, para añadir precisión a la determinación y clasificación de las características situacionales de comportamiento, psicológicas y de estilo de habla del orador 3010.

40 El método de análisis puede comprender además la etapa de analizar repetidamente el mismo orador en varias situaciones diferentes, y la extracción adicional de los patrones de habla que son repetitivos y comunes en las diferentes situaciones, con el fin de determinar las características de personalidad del orador y diferenciarlas de las características situacionales 3020.

45 El método de análisis puede comprender además usar detección de emociones a través del análisis del habla para diferenciar el estado emocional actual de las características situacionales, con el propósito de añadir precisión para determinar y calificar las características situacionales de comportamiento, psicológicas y de estilo de habla del orador 3030.

El método de análisis puede comprender además la etapa de usar el análisis del habla de palabras y contenido, a través de técnicas de reconocimiento de voz, complementarias con el análisis prosódico del habla, con el propósito de añadir precisión para determinar y calificar las características situacionales de comportamiento, psicológicas y de estilo de habla 3040.

50 El método de análisis puede comprender además usar análisis del lenguaje corporal, a través del análisis visual del orador, complementario con el análisis del habla, con el propósito de añadir precisión para determinar y calificar las características situacionales de comportamiento, psicológicas y de estilo de habla 3050.

55 El método de análisis puede comprender además proporcionar rasgos de sistema de aprendizaje usando análisis continuo de las características de comportamiento, psicológicas y de estilo de habla de los oradores con el propósito de mejorar la base de datos de referencia de parámetros del habla y el proceso de clasificación.

El método de análisis se puede dirigir además a expresiones del habla que comprenden una pluralidad de oradores en donde el método comprende además separar la pluralidad de voces del orador y en donde cada etapa de dicho método se realiza por separado para cada orador.

La FIG. 4 ilustra un diagrama de bloques esquemático de alto nivel que muestra un sistema de procesamiento de datos para analizar una expresión del habla de un orador y determinar las características de comportamiento, psicológicas y de estilo de habla del orador en una situación y contexto dados. El sistema 4000 comprende: una unidad de entrada de voz 4010 dispuesta para muestrear las expresiones de voz del orador o para obtener expresiones de voz muestreadas; una unidad de procesamiento previo 4020 acoplada a la unidad de entrada de voz 4010 y dispuesta para procesar previamente muestras de voz desde la unidad de entrada de voz; una unidad de procesamiento principal 4030 acoplada a la unidad de procesamiento previo 4020; una unidad de salida de indicadores principales 4040; y una base de datos de referencia de oradores 4050 acoplada a la unidad de procesamiento principal 4030.

En operación, la unidad de entrada de voz 4010 está dispuesta para muestrear u obtener expresiones del habla muestreadas de un orador en una situación y contexto específicos; la unidad de procesamiento previo 4020 está dispuesta para procesar previamente la expresión muestreada en segmentos de habla silenciosos y activos y dividir los segmentos de habla activos en cadenas de bloques de igual longitud. Los bloques teniendo parámetros de habla primarios que incluyen parámetros de tono y amplitud. Además, la base de datos 4050 está dispuesta para almacenar clasificaciones de comportamiento, psicológicas y de estilos de habla y sus parámetros del habla secundarios representativos y combinaciones de parámetros, en donde tales parámetros se seleccionan a partir de una lista que comprende: longitud de pausa promedio, suma de pausas por unidad de tiempo, longitud promedio de silencios cortos, suma de silencios cortos por unidad de tiempo, longitud promedio de segmentos de tono igual, suma de segmentos de tono igual por unidad de tiempo, longitud promedio de segmentos de tono ascendente, suma de segmentos de tono ascendente por unidad de tiempo, longitud promedio de segmentos de tono descendente, suma de segmentos de tono descendente por unidad de tiempo, dispersión promedio de amplitud dentro de segmentos de tono igual, dispersión promedio de amplitud dentro de segmentos de tono ascendente, dispersión promedio de amplitud dentro de segmentos de tono decreciente, variación de tono y rango por unidad de tiempo, pendiente promedio de tono por unidad de tiempo, patrones de forma espectral y envolvente espectral de la señal de habla en el tiempo, patrones de orden secuencial de tendencias de tono ascendente, descendente e igual en el tiempo; además, la unidad de procesamiento principal 4030 está dispuesta para derivar una pluralidad de parámetros del habla secundarios a partir de la expresión del habla analizada actualmente, seleccionada de una lista que comprende dichos parámetros del habla secundarios; la unidad de procesamiento principal 4030 está dispuesta además para calcular los parámetros del habla únicos, las combinaciones de parámetros y los valores de parámetros representativos de las características situacionales de comportamiento, psicológicas y de estilo de habla, a partir de dichos parámetros secundarios en la expresión del habla; Se dispone además para determinar y calificar las características situacionales de comportamiento, psicológicas y de estilo de habla en la expresión del habla comparando dichos parámetros calculados con la base de datos de referencia predefinida de los parámetros del habla. Finalmente, la unidad de indicadores principales 4040 está dispuesta para emitir los resultados del análisis.

La unidad de salida 4040 puede comprender al menos uno de los siguientes: visual, audio, almacenamiento, texto, interfaz a un dispositivo de monitorización, interfaz a internet, interfaz a un dispositivo u ordenador local o uno remoto, interfaz a otra red, interfaz a un teléfono fijo, inalámbrico o celular, interfaz a un juego de ordenador, interfaz a un juguete, interfaz a un organizador personal electrónico, e interfaz a un equipo de salida electrónico.

El sistema 4000 se puede implementar dentro de un sistema de comunicación seleccionado a partir de un grupo que comprende: un sistema de comunicación de voz, sistema de comunicación cableado o inalámbrico, red de comunicación celular, sistema de comunicación basado en web, otro sistema de comunicación basado en voz sobre IP, y sistema de comunicación de centros de llamadas.

El sistema 4000 se puede configurar para su uso en al menos uno de los siguientes usos: auto realimentación, conciencia de sí mismo, entretenimiento, entrenamiento de comportamiento, mejora de ventas, mejora de servicio al cliente, inteligencia de clientes y de negocios, aplicaciones de defensa y seguridad, negociación comercial, publicidad, emparejamiento, citas, recreo, juegos, juguetes, asesoramiento, aplicaciones de recursos humanos, entrenamiento de ejecutivos, entrenamiento de empleados y vocacional, aplicaciones psicológicas y educativas, y aplicaciones médicas.

La base de datos 4050 puede comprender además muestras de habla obtenidas previamente de un orador analizado actualmente para el reconocimiento de los patrones de habla únicos del orador, de manera que el análisis actual del orador sería dependiente del orador más que independiente del orador, para añadir precisión para determinar y calificar las características situacionales de comportamiento, y psicológicas y de estilo de habla del orador.

El sistema 4000 puede estar dispuesto además para analizar repetidamente el mismo orador en varias situaciones diferentes, y extraer los patrones de habla que son repetitivos y comunes en las diferentes situaciones, con el fin de determinar las características de personalidad del orador y diferenciarlas de las características situacionales.

El sistema 4000 puede estar dispuesto además para usar detección de emociones a través de análisis del habla, para la diferenciación del estado emocional actual a partir de las características situacionales, con el propósito de añadir precisión para determinar y calificar las características situacionales de comportamiento, psicológicas y de estilo de habla del orador.

El sistema 4000 puede estar dispuesto para usar análisis del habla de palabras y contenido, a través de técnicas de reconocimiento de voz, complementario con análisis prosódico del habla, con el propósito de añadir precisión a determinar y calificar las características situacionales de comportamiento, psicológicas y de estilo de habla.

5 El sistema 4000 puede estar dispuesto para usar análisis del lenguaje corporal, a través de análisis visual del orador, complementario con el análisis del habla, con el propósito de añadir precisión para determinar y calificar las características situacionales de comportamiento, psicológicas y de estilo de habla.

El sistema 4000 puede estar dispuesto para proporcionar rasgos de sistema de aprendizaje usando análisis continuo de las características de comportamiento, psicológicas y de estilo de habla de los oradores con el propósito de mejorar la base de datos de referencia de parámetros del habla y el proceso de clasificación.

10 El sistema 4000 se puede dirigir a expresiones del habla que comprenden una pluralidad de oradores y en donde el sistema está dispuesto para separar la pluralidad de voces de orador y en donde el sistema está dispuesto además para analizar a cada orador por separado.

15 La FIG. 5 ilustra un diagrama de bloques esquemático de alto nivel que muestra un aparato para analizar una expresión del habla de un orador y determinar características de comportamiento, psicológicas y de estilo de habla del orador en una situación y contexto dados.

20 El aparato 5000 comprende: una unidad de entrada de voz 5010 dispuesta para muestrear las expresiones de voz del orador o para obtener expresiones de voz muestreadas; una unidad de procesamiento previo 5020 acoplada a la unidad de entrada de voz 5010 y dispuesta para procesar previamente muestras de voz de la unidad de entrada de voz 5010; una unidad de procesamiento principal 5030 acoplada a la unidad de procesamiento previo 5020; y una unidad de salida de indicadores principales 5040 acoplada a la unidad de procesamiento principal 5030.

25 En operación, la unidad de entrada de voz 5010 está dispuesta para muestrear u obtener expresiones del habla muestreadas de un orador en una situación y contexto específicos; la unidad de proceso previo 5020 está dispuesta para procesar previamente la expresión muestreada en segmentos del habla silenciosos y activos y dividir los segmentos del habla activos en cadenas de bloques de igual longitud. Los bloques tienen parámetros de habla primarios, incluyendo parámetros de tono y amplitud. Además, el aparato 5000 está asociado operativamente con una base de datos de referencia 5050 estando dispuesta para almacenar clasificaciones de comportamiento, psicológicas y de estilos de habla y sus parámetros del habla secundarios representativos y combinaciones de parámetros, en donde tales parámetros se seleccionan a partir de una lista que comprende: longitud de pausa promedio, suma de pausas por unidad de tiempo, longitud promedio de silencios cortos, suma de silencios cortos por unidad de tiempo, longitud promedio de segmentos de tono igual, suma de segmentos de tono igual por unidad de tiempo, longitud promedio de segmentos de tono ascendente, suma de segmentos de tono ascendente por unidad de tiempo, longitud promedio de segmentos de tono descendente, suma de segmentos de tono descendente por unidad de tiempo, dispersión de amplitud promedio dentro de segmentos de tono igual, dispersión de amplitud promedio dentro segmentos de tono ascendente, dispersión de amplitud promedio dentro de segmentos de tono descendente, variación y rango de tono por unidad de tiempo, pendiente de tono promedio por unidad de tiempo, patrones de forma espectral y envolvente espectral de la señal de habla en el tiempo, patrones de orden secuencial de tendencias de tono ascendente, y descendente e igual en el tiempo.

35 Además, la unidad de procesamiento principal 5030 está dispuesta para derivar una pluralidad de parámetros del habla secundarios de la expresión del habla analizada actualmente, seleccionada de una lista que comprende dichos parámetros del habla secundarios. La unidad de procesamiento principal 5030 está dispuesta además para calcular parámetros del habla únicos, combinaciones de parámetros y valores de parámetros representativos de características situacionales de comportamiento, psicológicas y de estilos de habla, a partir de dichos parámetros secundarios en la expresión del habla; está dispuesta además para determinar y calificar las características situacionales de comportamiento, psicológicas y de estilo de habla en la expresión del habla comparando los parámetros calculados con la base de datos de referencia predefinida de parámetros del habla. Finalmente, la unidad de indicadores principales 5040 está dispuesta para emitir los resultados del análisis.

40 El aparato 5000 se puede configurar para al menos uno de los siguientes usos: auto realimentación, conciencia de sí mismo, entretenimiento, entrenamiento de comportamiento, mejora de ventas, mejora de servicio al cliente, inteligencia de clientes y de negocios, aplicaciones de defensa y seguridad, negociación comercial, publicidad, emparejamiento, citas, recreo, juegos, juguetes, asesoramiento, aplicaciones de recursos humanos, entrenamiento de ejecutivos, entrenamiento de empleados y vocacional, aplicaciones psicológicas y educativas, y aplicaciones médicas.

55 El aparato 5000 se puede seleccionar a partir de una lista que comprende un dispositivo de comunicación de voz, un dispositivo de voz cableado o inalámbrico, un dispositivo de comunicación celular, un accesorio de dispositivo personal (PDA) interconectado (o no interconectado), un dispositivo de grabación de voz, un ordenador personal y un servidor.

La unidad de salida 5040 puede comprender al menos uno de los siguientes: visual, audio, almacenamiento, texto, interfaz a un dispositivo de monitorización, interfaz a internet, interfaz a un dispositivo u ordenador local o remoto,

interfaz a otra red, interfaz a un teléfono fijo, inalámbrico o celular, interfaz a un juego de ordenador, interfaz a un juguete, e interfaz a un organizador personal electrónico o interfaz a cualquier otro equipo de salida electrónico.

5 Según algunas realizaciones, el método implementado por ordenador, el programa informático y el aparato de la presente invención se pueden implementar dentro de un sistema de comunicación de línea inalámbrico o cableado. Específicamente, la invención se puede implementar en un sistema de comunicación celular. La invención se puede configurar en algunas realizaciones para operar en un entorno en tiempo real o alternativamente como un sistema de análisis del habla fuera de línea.

10 Según algunas realizaciones, la invención se puede implementar en una pluralidad de arquitecturas. Por ejemplo, se puede proporcionar un módulo de análisis central que realice las etapas mencionadas anteriormente sobre los datos de voz de las llamadas telefónicas que pasan a través de los servidores centrales de un operador celular en cualquier ubicación física central.

15 Según otras realizaciones, la presente invención se puede implementar como un módulo autónomo integrado como software o hardware en un aparato de comunicación tal como un dispositivo de telecomunicación de habla. Por ejemplo, se puede proporcionar un módulo de análisis que realiza las etapas mencionadas anteriormente sobre los datos de voz de llamadas telefónicas que pasan a través de un dispositivo de teléfono celular de usuario final específico.

20 Opcionalmente, las realizaciones se pueden utilizar para, pero no se limitan a, analizar uno de los oradores en una conversación, o ambos, con los propósitos de auto realimentación, conocimiento de sí mismo, entretenimiento, entrenamiento de comportamiento, mejora de ventas, mejora de servicio al cliente, inteligencia de clientes y de negocios, aplicaciones de defensa y seguridad, negociación comercial, publicidad, emparejamiento, citas, recreo, juegos, juguetes, asesoramiento, aplicaciones de recursos humanos, entrenamiento de ejecutivos, entrenamiento de empleados y vocacional, aplicaciones psicológicas y educativas, aplicaciones médicas, y similares.

25 Según algunas realizaciones, el método implementado por ordenador, aparato y programa de ordenador de la presente invención se pueden implementar dentro de un análisis del habla en tiempo real o fuera de línea de conversaciones de centros de llamadas y centros de contacto para la mejora del servicio, ventas, atención al cliente, entrenamiento y otros propósitos de organización.

Según algunas realizaciones, la presente invención puede proporcionar análisis del habla en tiempo real y fuera de línea de conversaciones de teléfono fijo, analizando uno de los oradores en una conversación, o ambos, para los mismos propósitos que se han descrito anteriormente.

30 Según algunas realizaciones, la presente invención puede proporcionar análisis del habla en tiempo real y fuera de línea de conversaciones basadas en la web, analizando una de las partes o ambas partes, con los mismos propósitos que se han descrito anteriormente.

35 Según algunas realizaciones, la presente invención puede proporcionar análisis del habla en tiempo real y fuera de línea de conversaciones cara a cara analizando una de las partes o ambas partes, con los mismos propósitos que se han descrito anteriormente.

Según algunas realizaciones, la presente invención puede proporcionar análisis del habla en tiempo real y fuera de línea de cualquier habla grabada, ya sea conversación, auto discurso o discurso frente a una audiencia, analizando una de las partes o ambas partes, con los mismos propósitos que se han descrito anteriormente.

40 Ventajosamente, la evaluación de las características comunicacionales, psicológicas, personales y de comportamiento de las personas proporciona una tremenda importancia por diversas razones. En primer lugar, proporciona datos de evaluación para el orador en sí mismo, alentando de esta manera la conciencia, el desarrollo y diferentes propósitos de aprendizaje. Además, tiene una importancia significativa para otras personas que puedan querer aprender acerca del orador, para ayudarles a interactuar mejor con él, con diversos propósitos – personal, comercial, educativo, vocacional, médico, y psicológico, y similares

45 La invención se puede implementar en circuitería electrónica digital, o en hardware, microprograma, software de ordenador o en combinaciones de los mismos.

50 La invención se puede implementar ventajosamente en uno o más programas de ordenador que son ejecutables en un sistema programable que incluye al menos un procesador programable acoplado para recibir datos e instrucciones de, y para transmitir datos e instrucciones a, un sistema de almacenamiento de datos, al menos un dispositivo de entrada y al menos un dispositivo de salida. Un programa de ordenador es un conjunto de instrucciones que se pueden usar, directa o indirectamente, en un ordenador para realizar una cierta actividad u ocasionar un cierto resultado. Un programa de ordenador se puede escribir en cualquier forma de lenguaje de programación, incluyendo lenguajes compilados o interpretados, y se puede desplegar de cualquier forma, incluyendo como programa autónomo o como módulo, componente, subrutina u otra unidad adecuada para su uso

55 en un entorno informático.

- Procesadores adecuados para la ejecución de un programa de instrucciones incluyen, a modo de ejemplo, procesadores de señal digital (DSP) pero también microprocesadores de propósito general, y el único procesador o uno de múltiples procesadores de cualquier clase de ordenador. De manera general, un procesador recibirá instrucciones y datos de una memoria de sólo lectura o una memoria de acceso aleatorio o ambas. Los elementos esenciales de un ordenador son un procesador para ejecutar instrucciones y una o más memorias para almacenar instrucciones y datos. De manera general, un ordenador también incluirá, o estará acoplado operativamente para comunicarse con uno o más dispositivos de almacenamiento masivo para almacenar archivos de datos; tales dispositivos incluyen discos magnéticos, tales como discos duros internos y discos extraíbles; discos magneto-ópticos; y discos ópticos. Dispositivos de almacenamiento adecuados para incorporar tangiblemente instrucciones de programa de ordenador y datos incluyen todas las formas de memoria no volátil, incluyendo a modo de ejemplo dispositivos de memoria de semiconductores, tales como EPROM, EEPROM y dispositivos de memoria rápida; discos magnéticos tales como discos duros internos y discos extraíbles; discos magneto-ópticos; y discos CD-ROM y DVD-ROM. El procesador y la memoria se pueden complementar por, o incorporar en, ASIC (circuitos integrados de aplicaciones específicas).
- En la descripción anterior, una realización es un ejemplo o implementación de la invención. Las diversas apariciones de "una realización" o "algunas realizaciones" no necesariamente se refieren todas a las mismas realizaciones.
- Aunque se pueden describir diversos rasgos de la invención en el contexto de una única realización, los rasgos también se pueden proporcionar por separado o en cualquier combinación adecuada. Por el contrario, aunque la invención se puede describir en la presente memoria en el contexto de realizaciones separadas por claridad, la invención también se puede implementar en una única realización.
- Una referencia en la especificación a "algunas realizaciones", "una realización" u "otras realizaciones" significa que un rasgo, estructura o característica particular descrita en conexión con las realizaciones se incluye en al menos algunas realizaciones, pero no necesariamente en todas las realizaciones, de la invención.
- Ha de ser entendido que la fraseología y la terminología empleadas en la presente memoria no han de ser interpretadas como limitativas y son solamente con propósito descriptivo.
- Los principios y usos de las enseñanzas de la presente invención se pueden entender mejor con referencia a la descripción, figuras y ejemplos anexos.
- Ha de ser entendido que los detalles expuestos en la presente memoria no representan una limitación a una aplicación de la invención.
- Además, ha de ser entendido que la invención se puede llevar a cabo o poner en práctica de diversas formas y que la invención se puede implementar en realizaciones distintas de las perfiladas en la descripción anterior.
- Ha de ser entendido que los términos "incluyendo", "comprendiendo", "consistiendo" y variantes gramaticales de los mismos no excluyen la adición de uno o más componentes, rasgos, pasos o enteros o grupos de los mismos y que los términos han de ser interpretados como componentes, rasgos, pasos o enteros especificadores.
- Si la especificación o las reivindicaciones se refieren a un elemento "adicional", eso no excluye que haya más de uno del elemento adicional.
- Ha de ser entendido que donde las reivindicaciones o la especificación hacen referencia a "un" elemento, tal referencia no ha de ser interpretada como que hay solamente uno de ese elemento.
- Ha de ser entendido que donde la especificación expone que un componente, rasgo, estructura o característica se "puede" o "podría" incluir, ese componente, rasgo, estructura o característica particular no se requiere que sea incluida.
- Cuando sea aplicable, aunque se pueden usar diagramas de estado, diagramas de flujo o ambos para describir realizaciones, la invención no está limitada a esos diagramas o a las descripciones correspondientes. Por ejemplo, el flujo no necesita moverse a través de cada cuadro o estado ilustrado, o exactamente en el mismo orden que se ilustra y describe.
- Los métodos según la presente invención se pueden implementar realizando o completando manualmente, automáticamente, o una combinación de los mismos, los pasos o tareas seleccionados.
- El término "método" se puede referir a maneras, medios, técnicas y procedimientos para realizar una tarea dada incluyendo, pero no limitada a, esas maneras, medios, técnicas y procedimientos o bien conocidos o bien desarrollados fácilmente a partir de maneras, medios, técnicas y procedimientos conocidos por los profesionales de la técnica a la que pertenece la invención.
- Las descripciones, ejemplos, métodos y materiales presentados en la especificación no han de ser interpretados como limitativos, sino más bien como ilustrativos solamente.

Los significados de términos técnicos y científicos usados en la presente memoria han de ser entendidos comúnmente por un experto en la técnica a la que pertenece la invención, a menos que se defina de otro modo.

La presente invención se puede implementar en las pruebas o en la puesta en práctica con métodos y materiales equivalentes o similares a los descritos en la presente memoria.

- 5 La mención o la identificación de cualquier referencia en la descripción de algunas realizaciones de la invención no se interpretarán como una admisión de que tal referencia está disponible como la técnica anterior a la presente invención.

- 10 Aunque la invención se ha descrito con respecto a un número limitado de realizaciones, éstas no se deberían interpretar como limitaciones sobre el alcance de la invención, sino más bien como ejemplificaciones de algunas de las realizaciones preferidas. Otras posibles variaciones, modificaciones y aplicaciones también están dentro del alcance de la invención. Por consiguiente, el alcance de la invención no se debería limitar por lo que se ha descrito hasta ahora, sino por las reivindicaciones adjuntas y sus equivalentes legales.

REIVINDICACIONES

1. Un método implementado por ordenador de creación de una base de datos de referencia de parámetros de habla para clasificar expresiones del habla según diversas características de comportamiento, psicológicas y de estilos de habla, dicho método implementado por ordenador que comprende:

- 5 seleccionar manualmente un cierto contexto de habla;
- seleccionar manualmente las características de comportamiento, psicológicas y de estilos de habla a ser analizadas en el contexto seleccionado;
- obtener una pluralidad de expresiones del habla de personas en el contexto seleccionado;
- 10 agrupar manualmente las expresiones del habla en grupos que representan características similares de comportamiento, psicológicas y de estilos de habla;
- procesar previamente cada una de las expresiones del habla en dichos grupos de expresiones que representan características similares de comportamiento, psicológicas y de estilos de habla, en segmentos de habla silenciosos y activos;
- 15 dividir los segmentos de habla activos en cadenas de bloques de igual longitud, dichos bloques que tienen parámetros del habla primarios que incluyen parámetros de tono y amplitud;
- derivar una pluralidad de parámetros del habla secundarios a partir de los parámetros del habla primarios, en donde dicha pluralidad de parámetros del habla secundarios seleccionados de: la suma de segmentos por unidad de tiempo de tendencias de tono igual, tono ascendente, y tono descendente en dichas cadenas de bloques, y las longitudes promedio de segmentos tendencias de tono igual, tono ascendente y tono descendente
- 20 tiende en dichas cadenas de bloques; y
- determinar parámetros del habla secundarios únicos, combinaciones de parámetros y valores de parámetros que son comunes a cada grupo de expresiones a partir de dichos parámetros secundarios, en donde dichos parámetros del habla secundarios únicos representan las características típicas de comportamiento, psicológicas y de estilos de habla de cada grupo.
- 25 2. Un programa de ordenador que comprende un conjunto de instrucciones que se pueden usar directa o indirectamente, en un ordenador, para realizar el método de la reivindicación 1.
- 3. Un aparato operable para realizar el método de la reivindicación 1.

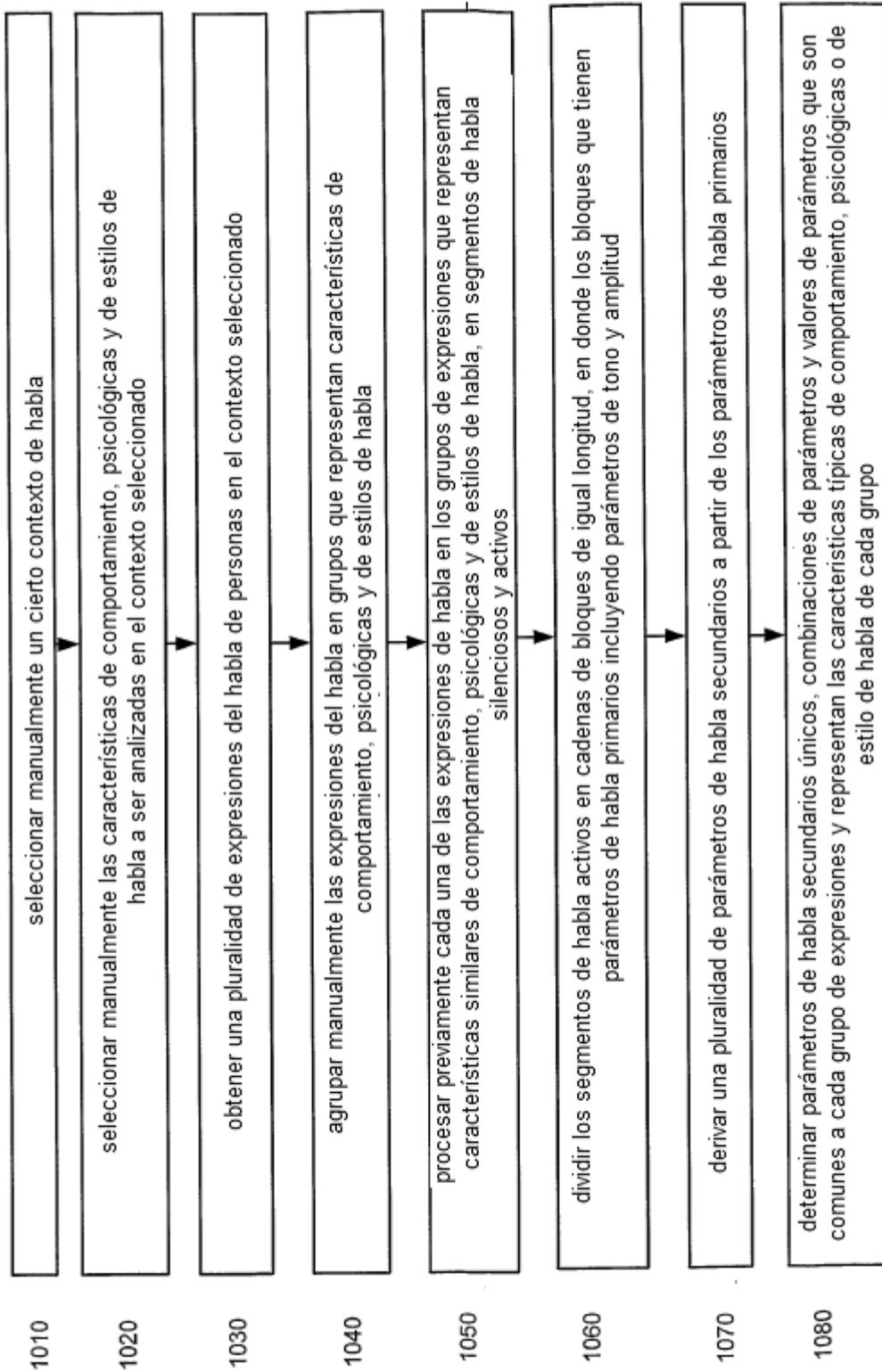


Figura 1

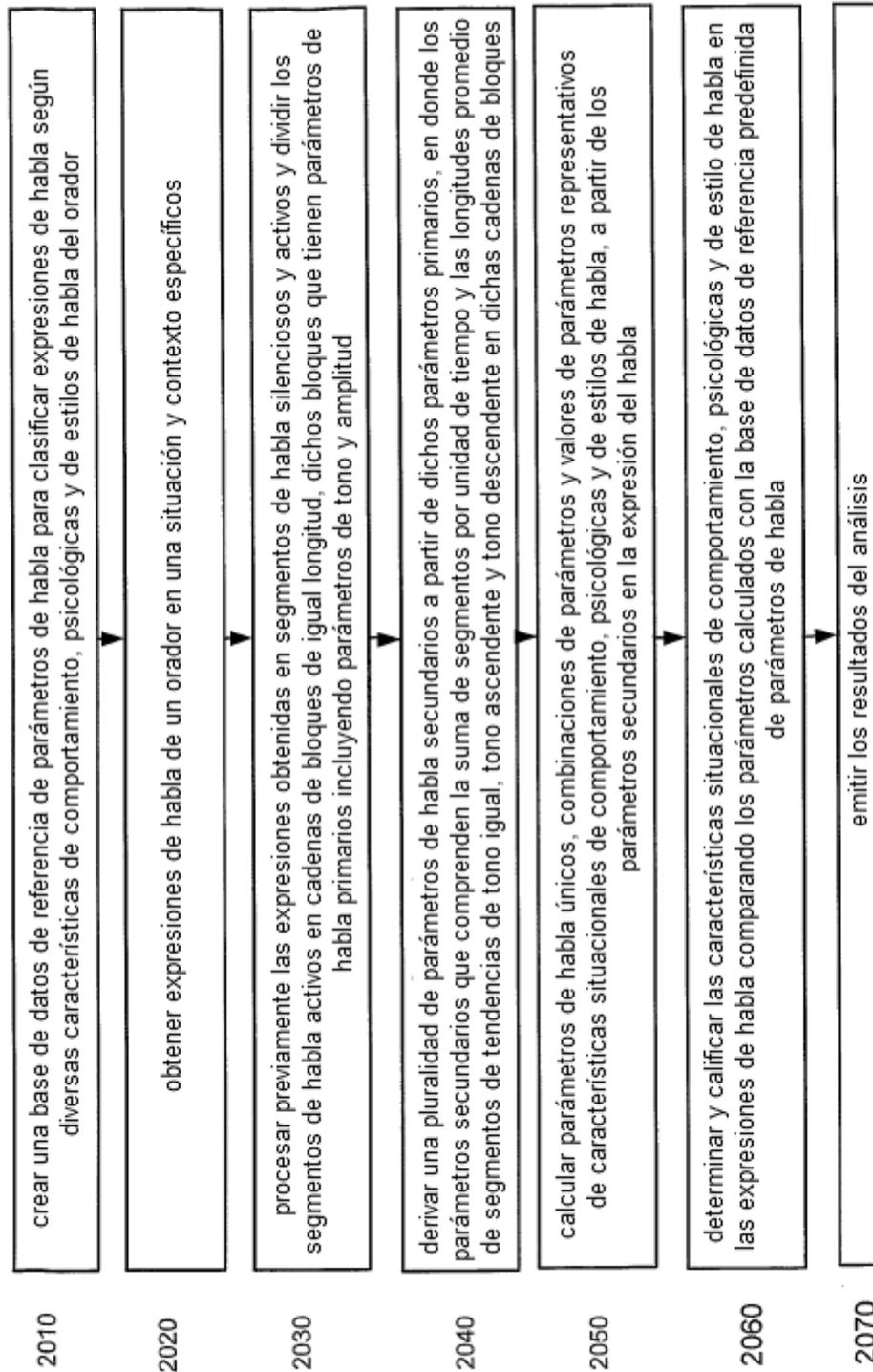


Figura 2

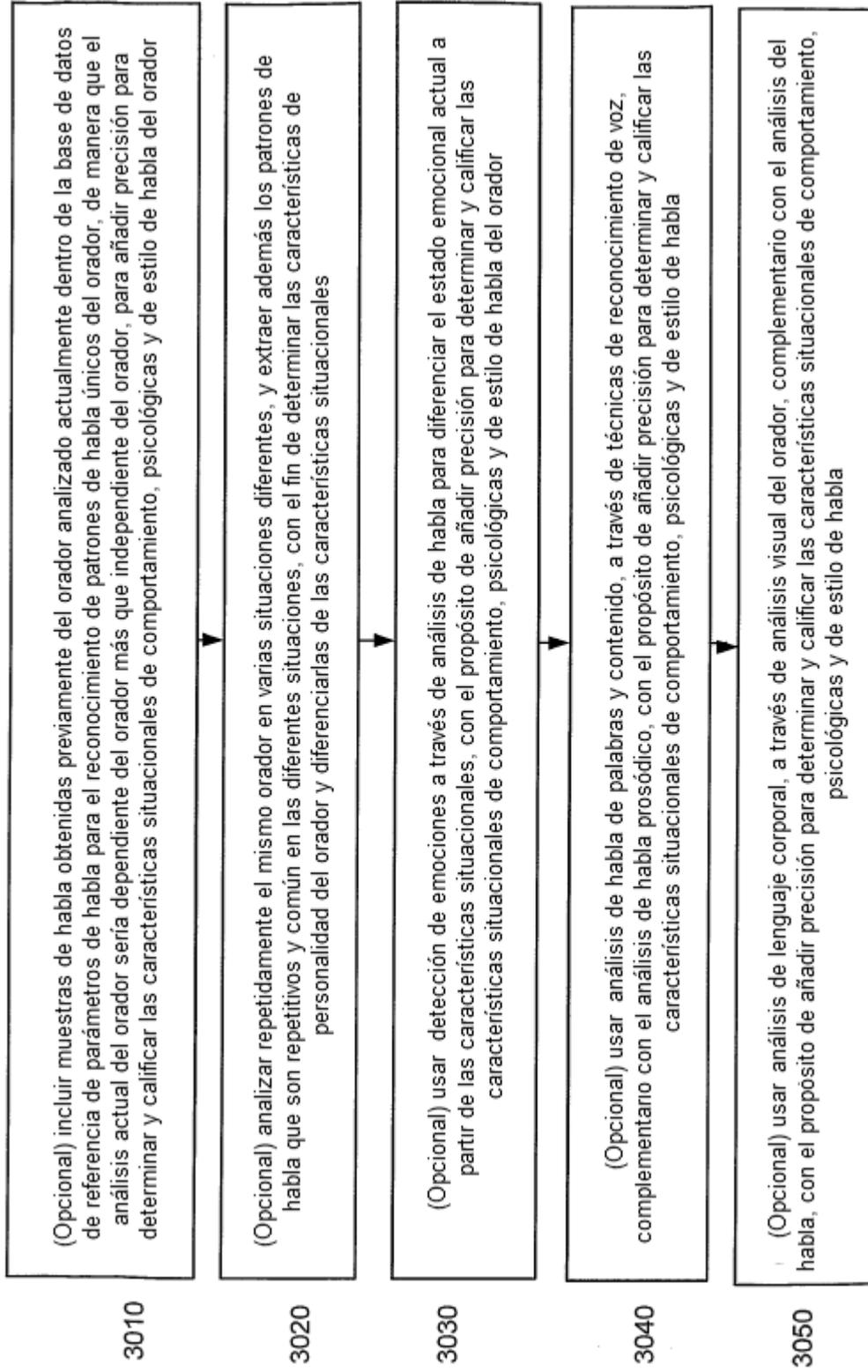


Figura 3

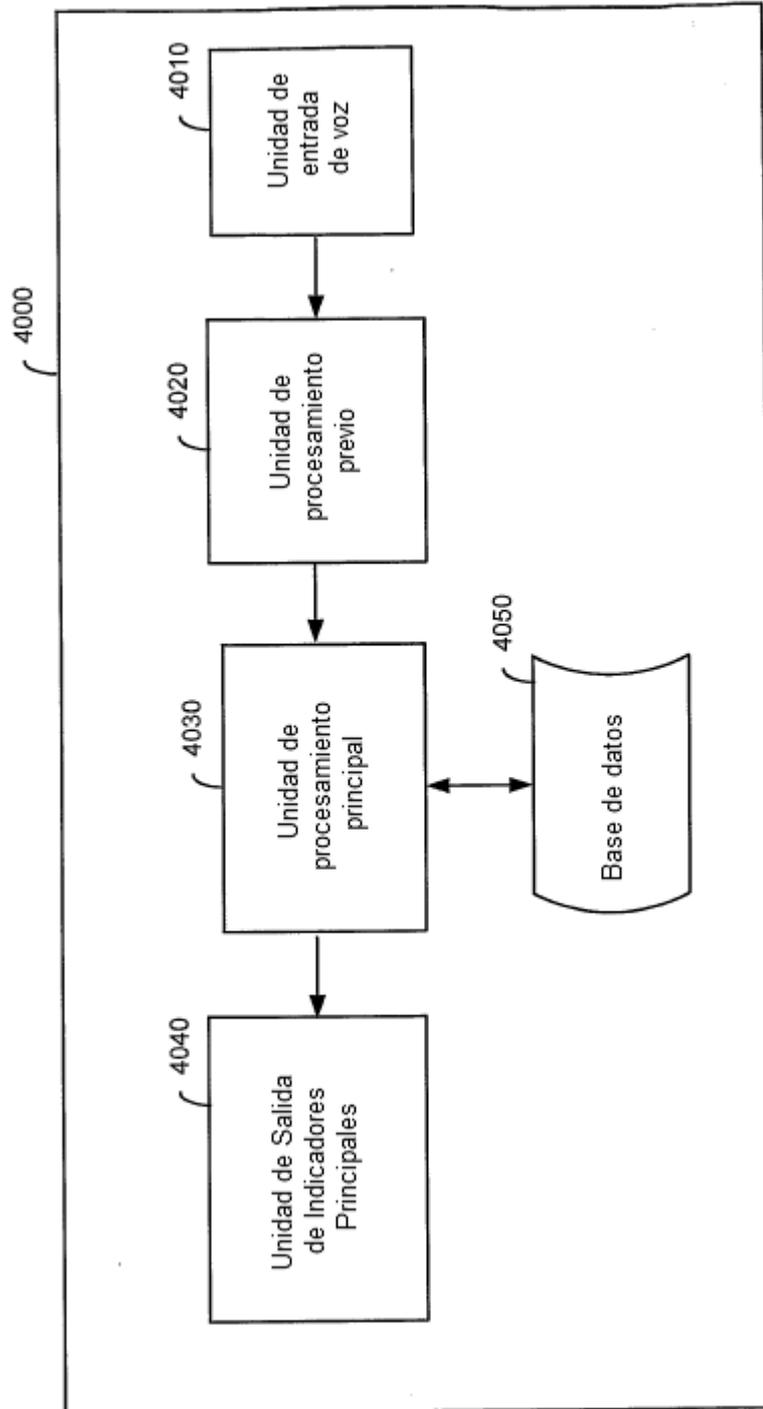


Figura 4

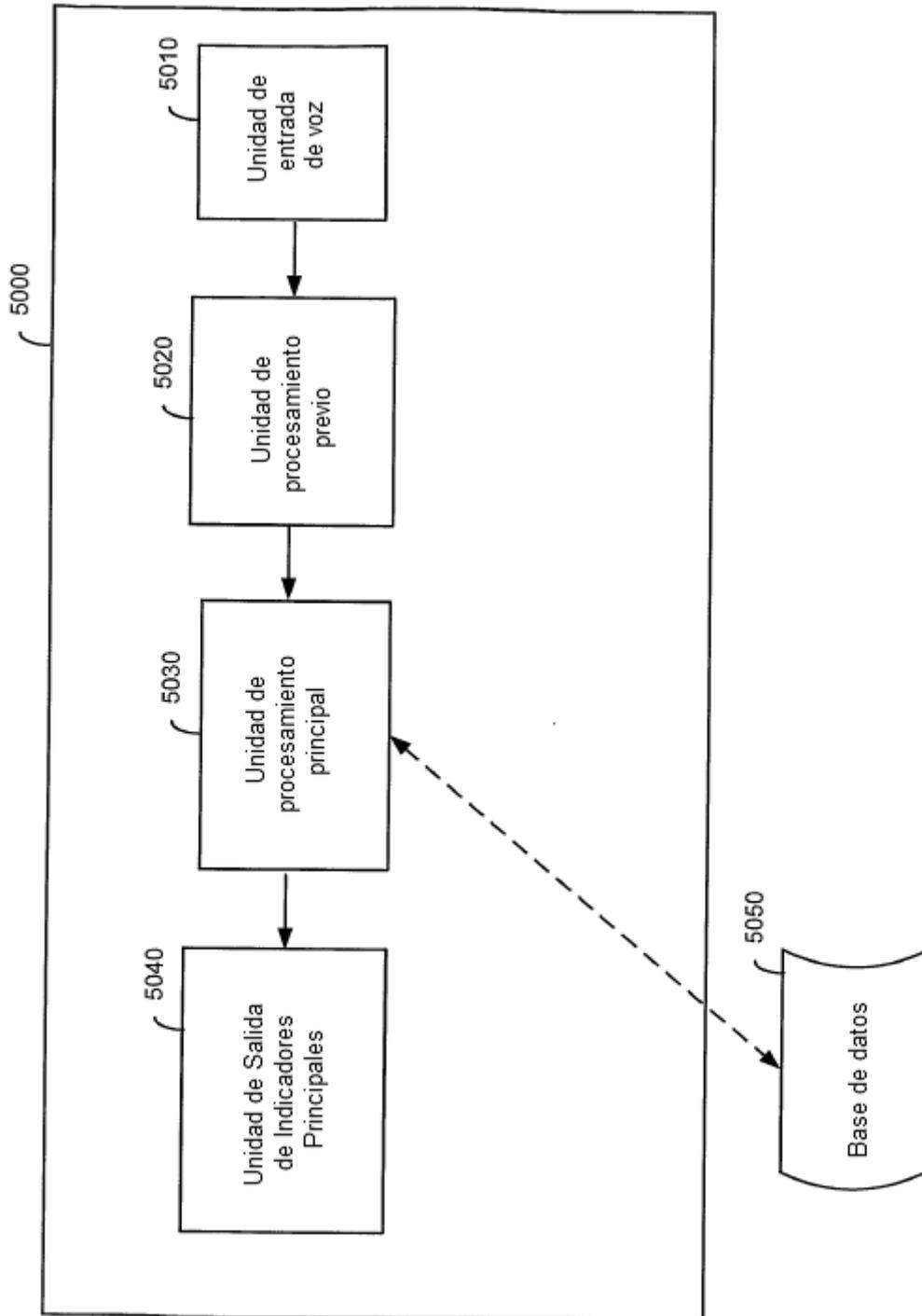


Figura 5