

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 811 771**

51 Int. Cl.:

G10L 15/22 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **06.03.2015** **E 15157919 (0)**

97 Fecha y número de publicación de la concesión europea: **20.05.2020** **EP 3065131**

54 Título: **Método y sistema de postratamiento de un resultado de reconocimiento de voz**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
15.03.2021

73 Titular/es:
ZETES INDUSTRIES S.A. (100.0%)
Rue de Strasbourg 3
1130 Bruxelles, BE

72 Inventor/es:
FORSTER, JEAN-LUC

74 Agente/Representante:
FLORES DREOSTI, Lucas

ES 2 811 771 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

Método y sistema de postratamiento de un resultado de reconocimiento de voz

Campo de la invención

5 **[0001]** Según un primer aspecto, la invención se refiere a un método de postratamiento de un resultado de reconocimiento de voz. Según un segundo aspecto, la invención se refiere a un sistema (o dispositivo) de postratamiento de un resultado de reconocimiento de voz. Según un tercer aspecto, la invención se refiere a un programa. Según un cuarto aspecto, la invención se refiere a un medio de almacenamiento que comprende unas instrucciones (por ejemplo: memoria USB, disco de tipo CD-ROM o DVD).

Estado de la técnica

10 **[0002]** Un motor de reconocimiento de voz permite generar, a partir de un audio o mensaje hablado, un resultado que generalmente es en forma de texto o código ejecutable por una máquina. Actualmente esta tecnología está ampliamente extendida y se considera muy útil. Distintas aplicaciones del reconocimiento de voz se exponen especialmente en el documento US6,754,629B1.

15 **[0003]** Existen estudios para mejorar los resultados obtenidos por un motor de reconocimiento de voz. Por ejemplo, el documento US2014/0278418A1 propone aprovechar la identidad de un orador para adaptar en consecuencia los algoritmos de reconocimiento de voz de un motor de reconocimiento de voz. Esta adaptación de los algoritmos se realiza, por tanto, en el propio motor de reconocimiento de voz; por ejemplo, modificando su diccionario fonético para tener en cuenta la manera de hablar del orador o del usuario.

20 **[0004]** Un resultado de reconocimiento de voz comprende generalmente una serie de elementos, por ejemplo, palabras, separados por silencios. El resultado está caracterizado por un inicio y un fin y sus elementos están distribuidos temporalmente entre este inicio y este fin.

25 **[0005]** Un resultado obtenido por un motor de reconocimiento de voz puede utilizarse, por ejemplo, para introducir una información en un sistema informático, por ejemplo, cualquier número de artículo o instrucción que se quiera realizar. En lugar de utilizar un resultado bruto de reconocimiento de voz, este resultado se somete en ocasiones a una o varias operaciones de postratamiento para extraer una solución postratada. Por ejemplo, es posible recorrer un resultado de reconocimiento de voz desde el inicio hasta el fin y conservar, por ejemplo, los cinco primeros elementos considerados como válidos, si se sabe que la información útil no comprende más de cinco elementos (un elemento es, por ejemplo, una palabra). De hecho, al saber que la información útil (un código, por ejemplo) no comprende más de cinco palabras (cinco cifras, por ejemplo), en ocasiones se decide, por lo tanto, conservar únicamente los cinco primeros elementos válidos de un resultado de reconocimiento de voz. Cualquier elemento posterior adicional se considera superfluo en relación con la información esperada y, por lo tanto, se considera como no válido. El documento US7181399 procede de manera similar. Dicho método de postratamiento no proporciona siempre soluciones aceptables. Así, los inventores han descubierto que dicho método puede conducir a generar una solución postratada falsa en ciertos casos, es decir, una solución que no se corresponde con la información que debe ser proporcionada realmente por el orador. Este método de postratamiento no es, por lo tanto, suficientemente fiable.

40 **[0006]** El documento US 2007/050190 describe un método de tratamiento de una señal vocal según el cual dicha señal vocal se descompone en primer lugar en secciones vocales en función de los tiempos muertos entre dos tramos vocales (por ejemplo, dos palabras) pronunciados por el orador. Un mecanismo de prioridad selecciona posteriormente un subconjunto limitado de estas secciones vocales en función de criterios predefinidos. Solo las secciones vocales de este subconjunto limitado se envían a continuación a un motor de reconocimiento de voz. A diferencia del método según la invención, en este caso se trata de un método de pretratamiento de una señal vocal, es decir, un método que interviene antes de que los datos vocales se envíen a un motor de reconocimiento de voz para obtener entonces un resultado de reconocimiento de voz. Además, este método de pretratamiento no proporciona siempre soluciones aceptables, ya que los tiempos muertos entre dos tramos vocales pueden variar entre un orador y otro y/o entre una expresión y otra de un mismo orador.

Resumen de la invención

50 **[0007]** Según un primer aspecto, uno de los objetivos de la invención es proporcionar un método de postratamiento de un resultado de reconocimiento de voz que sea más fiable. A tal efecto, los inventores proponen el siguiente método. Método de postratamiento de un resultado de reconocimiento de voz, comprendiendo dicho resultado un inicio, un fin y una pluralidad de elementos distribuidos entre dicho inicio y dicho fin, comprendiendo dicho método de postratamiento las siguientes etapas:

i. recibir dicho resultado;

55 ii. aislar un elemento de dicha pluralidad de elementos que no se ha sometido a la prueba de validación de la etapa iii.a.;

iii. entonces,

- a. si se ha aislado un elemento en la etapa ii., determinar si es válido utilizando una prueba de validación,
- b. si no, pasar directamente a la etapa v.;

iv. repetir las etapas ii. y iii. (en el orden siguiente: etapa ii. y después etapa iii.);

5 v. si al menos un elemento se ha determinado como válido en la etapa iii.a, determinar una solución postratada utilizando al menos un elemento determinado como válido en la etapa iii.a;

caracterizado por que cada elemento aislado en la etapa ii. se selecciona de dicho fin del resultado a dicho inicio del resultado de manera consecutiva y por que la etapa iii.a. comprende, además, una instrucción para pasar directamente a la etapa v. si el elemento (113) sometido a la prueba de validación de la etapa iii.a no ha sido determinado como válido.

10 **[0008]** Con el método de la invención, se recorre un resultado de reconocimiento de voz desde el inicio hasta el fin. De hecho, los inventores han descubierto que una persona que dicta un mensaje a un motor de reconocimiento de voz tenía más tendencia a dudar y/o a equivocarse en el inicio que en el fin. Al tratar un resultado de reconocimiento de voz por el fin en lugar que por el inicio, el método de la invención favorece la parte del resultado que tiene más posibilidades de presentar la información correcta. Por lo tanto, al final, este método es más fiable.

15 Además, una solución postratada para la que se haya determinado al menos un elemento como válido en la etapa iii.a comprende únicamente elementos válidos consecutivos del resultado del motor de reconocimiento de voz. Por consiguiente, la fiabilidad del método se mejora, ya que solo se mantiene una serie de elementos válidos consecutivos.

20 **[0009]** Tomemos el siguiente ejemplo. Imaginemos que un código para leer es: 4531. El operador, al leerlo, dice: «5, 4, pues, 4, 5, 3, 1». Por lo general, un motor de reconocimiento de voz proporcionará como resultado o «5, 4, 2, 4, 5, 3, 1» o «5, 4, 4, 5, 3, 1». En el primer caso, «pues» se asocia a «dos»; en el segundo caso, el motor no proporciona ningún resultado para «pues». Pongamos como hipótesis que un sistema de postratamiento (que puede estar integrado en un motor de reconocimiento de voz) sabe que el resultado no debe tener más de cuatro elementos correctos (cifras, en este caso). Un sistema de postratamiento que recorre el resultado de inicio a fin

25 del resultado proporcionará como solución postratada: 5425 o 5445 (y no 4531). El método de la invención proporcionará 4531, es decir, la solución correcta.

30 **[0010]** Los inventores han observado que la situación representada en este ejemplo, es decir, el hecho de que un operador tenga una mayor tendencia a dudar o a equivocarse en el inicio que en el fin de la secuencia registrada, es más frecuente que la contraria. Por tanto, en conjunto, el método de la invención resulta más fiable debido a que proporciona menos resultados incorrectos. Las posibilidades de obtener una solución postratada correcta son también más altas con el método de la invención. Por consiguiente, también es más eficaz.

35 **[0011]** El método de la invención presenta otras ventajas. Es fácil de implementar. En concreto, no requiere muchas etapas de implementación. Además, las etapas de implementación son simples. Estos aspectos facilitan la integración del método de la invención, por ejemplo, en un sistema informático que utilice un resultado de reconocimiento de voz, o, por ejemplo, en un motor de reconocimiento de voz.

[0012] El método de postratamiento de acuerdo con la invención se puede considerar un método de filtrado de un resultado de reconocimiento de voz: de hecho, los elementos no válidos no se utilizan para determinar la solución postratada.

40 **[0013]** Un resultado de reconocimiento de voz se encuentra generalmente en forma de texto o de código ejecutable por una máquina. Un elemento de un resultado representa una información del resultado delimitada por dos tiempos distintos a lo largo de una escala temporal, t, asociada al resultado, y que no se considera como un silencio o un ruido de fondo. Por lo general, un elemento es un grupo de fonemas. Un fonema resulta conocido para los expertos en la materia. Preferentemente, un elemento es una palabra. Un elemento también puede ser un grupo o combinación de palabras. Un ejemplo de combinación de palabras es «anular operación».

45 **[0014]** En el contexto de la invención, un resultado de reconocimiento de voz puede ser de diferentes tipos. Según un primer ejemplo posible, un resultado de reconocimiento de voz representa una hipótesis proporcionada por un motor de reconocimiento de voz a partir de un mensaje expresado por un usuario u orador. En general, un motor de reconocimiento de voz proporciona varias (por ejemplo, tres) hipótesis a partir de un mensaje expresado por un usuario. En este caso, por lo general también proporciona una puntuación (que se puede expresar en distintas unidades en función del tipo de motor de reconocimiento de voz) para cada hipótesis. Preferentemente, el método de postratamiento de la invención comprende, por lo tanto, una etapa preliminar para seleccionar únicamente la o las hipótesis que tengan una puntuación superior o igual a una puntuación predeterminada. Por ejemplo, si el motor de reconocimiento de voz empleado es el modelo VoCon® 3200 V3.14 de Nuance, dicha puntuación predeterminada es igual a 4000. Por lo tanto, las etapas anteriormente descritas (etapas i, ii, iii, iv, v) solo se aplican

50

55 a los resultados que tienen una puntuación superior o igual a dicha puntuación predeterminada.

[0015] Según otro ejemplo posible, un resultado de reconocimiento de voz es una solución que comprende generalmente una pluralidad de elementos, obtenidos a partir de una o varias operaciones de postratamiento aplicada(s) a una o varias hipótesis proporcionada(s) por un motor de reconocimiento de voz. En este último

ejemplo, el resultado de reconocimiento de voz procede, por lo tanto, de un módulo de reconocimiento de voz y procede de uno o varios módulo(s) de postratamiento de una o varias hipótesis proporcionada(s) por un motor de reconocimiento de voz.

5 **[0016]** Si no se ha determinado ningún elemento como válido en la etapa iii.a, la etapa v comprende preferiblemente una subetapa de proporcionar otra solución postratada. Preferiblemente, esta otra solución postratada corresponde a una solución postratada que no comprende el elemento de dicho resultado. En esta variante preferida, y cuando no se haya determinado ningún elemento válido en la etapa iii.a, distintos ejemplos de solución postratada son: mensaje vacío, es decir, que no comprenda ningún elemento (por ejemplo, ninguna palabra), mensaje que mencione que el postratamiento no ha tenido éxito. Según otra variante posible, esta otra solución postratada corresponde al resultado de reconocimiento de voz si no se ha determinado ningún elemento válido en la etapa iii.a (sin filtrado del resultado).

[0017] A lo largo de una escala temporal t asociada al resultado (véanse las figuras 1 y 2, por ejemplo), el inicio del resultado es anterior al fin del resultado.

15 **[0018]** Preferentemente, un elemento es una palabra. Ejemplos de palabras son: uno, dos, coche, paraguas. Según esta variante preferida, el método de la invención produce todavía mejores resultados. Cada palabra se determina a partir de un mensaje expresado por un usuario mediante un motor de reconocimiento de voz que usa un diccionario. Unas reglas gramaticales permiten reducir potencialmente la elección de palabras posibles dentro de un diccionario.

20 **[0019]** Preferentemente, el método de la invención comprende, además, la siguiente etapa: vi. determinar si dicha solución postratada de la etapa v. cumple una regla gramatical. Al utilizar una regla gramatical, se puede aumentar además la fiabilidad del método de la invención. En particular, puede filtrar mejor un resultado absurdo. Un ejemplo de regla gramatical es un intervalo de números de palabras admitido por la solución postratada. Por ejemplo, se podría definir como regla gramatical que la solución postratada debe contener entre tres y seis palabras.

25 **[0020]** Preferentemente, cuando se utiliza una regla gramatical, el método de la invención comprende, además, la siguiente etapa: vii.

a. si la respuesta a la prueba de la etapa vi. es positiva, proporcionar dicha solución postratada,

b. si no, proporcionar dicho resultado de reconocimiento de voz.

[0021] Según otra variante posible, el método de la invención comprende la siguiente etapa cuando se utiliza una regla gramatical: vii.

30 a. si la respuesta a la prueba de la etapa vi. es positiva (es decir, la solución postratada cumple la regla gramatical), proporcionar dicha solución postratada,

b. si la respuesta a la prueba de la etapa vi. es negativa (es decir, la solución postratada no cumple la regla gramatical), no proporcionar una solución postratada, o proporcionar un mensaje vacío, o proporcionar un mensaje que exprese que no se ha podido determinar ninguna solución postratada satisfactoria.

35 **[0022]** Se pueden concebir distintas pruebas de validación de la etapa iii.a. Por ejemplo, la prueba de validación de la etapa iii.a. puede comprender una etapa de consideración de un elemento válido si su duración es superior o igual a un umbral de duración inferior. A cada elemento del resultado le corresponde una duración o intervalo temporal que, por lo general, es proporcionado por el motor de reconocimiento de voz. Con esta forma de realización preferida, es posible prescindir de manera más eficaz de los elementos que sean de corta duración, como, por ejemplo, un ruido parásito que pueda proceder de una máquina.

40 **[0023]** Según otro ejemplo, la prueba de validación de la etapa iii.a. comprende una etapa de consideración de un elemento válido si su duración es inferior o igual a un umbral de duración superior. Con esta forma de realización preferida, es posible prescindir de manera más eficaz de los elementos que sean de larga duración, como, por ejemplo, una duda de un orador que diga por ejemplo «pues» pero para la que el motor de reconocimiento de voz proporcione la palabra «dos» (por ejemplo, porque emplea una regla gramatical predefinida que le impone que solo proporcione cifras). Al utilizar esta forma de realización preferida, se podrá eliminar más fácilmente esta palabra «dos» no válida.

45 **[0024]** Según otro ejemplo, dicha prueba de validación de la etapa iii.a. comprende una etapa de consideración de un elemento válido si su índice de confianza es mayor o igual a un índice de confianza mínimo. La fiabilidad del método se incrementa aún más en este caso.

50 **[0025]** Según otro ejemplo, dicha prueba de validación de la etapa iii.a. comprende una etapa de consideración de un elemento válido si un intervalo temporal que lo separa de otro elemento directamente adyacente hacia dicho fin del resultado es superior o igual a un intervalo de tiempo mínimo. Gracias a esta variante preferida, es posible rechazar de manera más eficaz elementos que no sean producidos por un ser humano, sino por una máquina, por ejemplo, y que sean temporalmente muy próximos.

[0026] Preferiblemente, dicha prueba de validación de la etapa iii.a. comprende una etapa de consideración de un elemento válido si un intervalo temporal que lo separa de otro elemento directamente adyacente hacia dicho fin del resultado es inferior o igual a un intervalo de tiempo máximo. Gracias a esta variante, es posible rechazar de manera más eficaz elementos que estén temporalmente muy separados entre sí.

5 **[0027]** Según otra variante posible del método de la invención, la prueba de validación de la etapa iii.a. comprende una etapa de consideración de un elemento válido si un intervalo temporal que lo separa de otro elemento directamente adyacente hacia dicho fin del resultado es mayor a un intervalo (temporal) mínimo.

10 **[0028]** Según otra variante posible del método de la invención, la prueba de validación de la etapa iii.a. comprende una etapa de consideración de un elemento válido si un intervalo temporal que lo separa de otro elemento directamente adyacente hacia dicho fin del resultado es menor a un intervalo (temporal) máximo.

15 **[0029]** Preferiblemente, dicha prueba de validación de la etapa iii.a. comprende una etapa de consideración, para un orador determinado, de un elemento de dicho resultado válido, si una estadística asociada a este elemento se ajusta, en un intervalo cercano, a una estadística preestablecida para un mismo elemento y para este orador determinado. La estadística (o estadística de reconocimiento de voz) asociada a dicho elemento es proporcionada generalmente por el motor de reconocimiento de voz. Ejemplos de estadística asociada a un elemento son: la duración del elemento, su índice de confianza. Son posibles otros ejemplos. Se pueden registrar dichas estadísticas para diferentes elementos y para distintos oradores (u operadores), por ejemplo, durante una etapa previa de registro de datos. Si luego se conoce la identidad del orador que ha registrado una declaración a la que corresponde un resultado proporcionado por un motor de reconocimiento de voz, se pueden comparar estadísticas asociadas a distintos elementos de dicho resultado con estadísticas preestablecidas para estos elementos y para este orador. En este caso, por lo tanto, el método de la invención comprende preferiblemente una etapa complementaria que permite determinar la identidad del orador. Gracias a esta forma de realización preferida, la fiabilidad y la eficacia se incrementan aún más, ya que se pueden tener en cuenta particularidades vocales del orador.

25 **[0030]** Preferiblemente, se recuperan todos los elementos determinados como válidos en la etapa iii.a para determinar dicha solución postratada en la etapa v.

[0031] Los inventores también proponen un método de optimización para proporcionar una solución optimizada a partir de un primer y un segundo resultado de reconocimiento de voz y que comprende las siguientes etapas:

30 A. aplicar un método de postratamiento de acuerdo con cualquiera de las reivindicaciones anteriores a dicho primer resultado;

B. aplicar un método de postratamiento de acuerdo con cualquiera de las reivindicaciones anteriores a dicho segundo resultado;

35 C. determinar dicha solución optimizada a partir de uno o varios elementos perteneciente(s) a uno o varios resultados de dichos primer y segundo resultado y que han sido determinados como válidos mediante la prueba de validación de la etapa iii.a.

40 **[0032]** Según un segundo aspecto, la invención se refiere a un programa (preferiblemente, un programa informático) para permitir tratar un resultado de reconocimiento de voz, comprendiendo dicho resultado un inicio, un fin y una pluralidad de elementos distribuidos entre dicho inicio y dicho fin, comprendiendo dicho programa un código para permitir que un dispositivo (por ejemplo, un motor de reconocimiento de voz, un ordenador apto para comunicarse con un motor de reconocimiento de voz) lleve a cabo las siguientes etapas:

i. leer dicho resultado de reconocimiento de voz,

ii. aislar un elemento de dicha pluralidad de elementos que no se ha sometido a la prueba de validación de la etapa iii.a.,

iii. entonces,

45 a. si se ha aislado un elemento en la etapa ii., determinar si es válido utilizando una prueba de validación,

b. si no, pasar directamente a la etapa v.,

iv. repetir las etapas ii. y iii.;

v. si al menos un elemento se ha determinado como válido en la etapa iii.a, determinar una solución postratada recuperando al menos un elemento determinado como válido en la etapa iii.a;

50 caracterizado por que cada elemento aislado en la etapa ii se selecciona de dicho fin del resultado a dicho inicio del resultado de manera consecutiva y por que la etapa iii.a. comprende, además, una instrucción para pasar directamente a la etapa v. si el elemento sometido a la prueba de validación de la etapa iii.a no ha sido determinado como válido.

[0033] Las ventajas asociadas al método de acuerdo con el primer aspecto de la invención se aplican al programa de la invención, *mutatis mutandis*. Así, en concreto, se puede presentar una solución postratada más fiable con el programa de la invención. Del mismo modo, se puede presentar un programa más eficaz para determinar una solución postratada correcta. Las distintas formas de realización expuestas para el método de acuerdo con el primer aspecto de la invención se aplican al programa de la invención, *mutatis mutandis*.

[0034] Si no se ha determinado ningún elemento como válido en la etapa iii.a, la etapa v comprende preferiblemente la siguiente subetapa: determinar una solución postratada que no comprenda ningún elemento de dicho resultado. En esta variante preferida, y cuando no se haya determinado ningún elemento válido en la etapa iii.a, distintos ejemplos de solución postratada son, por lo tanto: mensaje vacío, es decir, que no comprenda ningún elemento (por ejemplo, ninguna palabra), mensaje que mencione que el postratamiento no ha tenido éxito, resultado proporcionado por el motor de reconocimiento de voz.

[0035] Según un tercer aspecto, la invención se refiere a un medio de almacenamiento (o medio de registro) que puede estar conectado a un dispositivo (por ejemplo, un motor de reconocimiento de voz, un ordenador que se pueda comunicar con un motor de reconocimiento de voz) y que comprende unas instrucciones que, al leerlas, permiten que dicho dispositivo trate un resultado de reconocimiento de voz, comprendiendo dicho resultado un inicio, un fin y una pluralidad de elementos distribuidos entre dicho inicio y dicho fin, permitiendo dichas instrucciones imponer que dicho dispositivo lleve a cabo las siguientes etapas:

i. leer dicho resultado;

ii. aislar un elemento de dicha pluralidad de elementos que no se ha sometido a la prueba de validación de la etapa iii.a.,

iii. entonces,

a. si se ha aislado un elemento en la etapa ii., determinar si es válido utilizando una prueba de validación,

b. si no, pasar directamente a la etapa v.,

iv. repetir las etapas ii. y iii.;

v. si al menos un elemento se ha determinado como válido en la etapa iii.a, determinar una solución postratada recuperando al menos un elemento determinado como válido en la etapa iii.a;

caracterizado por que cada elemento aislado en la etapa ii. se selecciona de dicho fin del resultado a dicho inicio del resultado de manera consecutiva y por que la etapa iii.a. comprende, además, una instrucción para pasar directamente a la etapa v. si el elemento sometido a la prueba de validación de la etapa iii.a no ha sido determinado como válido.

[0036] Las ventajas asociadas al método de acuerdo con el primer aspecto de la invención se aplican al medio de almacenamiento de la invención, *mutatis mutandis*. Por lo tanto, en concreto, se puede presentar una solución postratada más fiable. Del mismo modo, se puede determinar más eficazmente una solución postratada correcta. Las distintas formas de realización expuestas para el método de acuerdo con el primer aspecto de la invención se aplican al medio de almacenamiento de la invención, *mutatis mutandis*.

[0037] Si no se ha determinado ningún elemento como válido en la etapa iii.a, la etapa v comprende preferiblemente la siguiente subetapa: determinar una solución postratada que no comprenda ningún elemento de dicho resultado. En esta variante preferida, y cuando no se haya determinado ningún elemento válido en la etapa iii.a, distintos ejemplos de solución postratada son, por lo tanto: mensaje vacío, es decir, que no comprenda ningún elemento (por ejemplo, ninguna palabra), mensaje que mencione que el postratamiento no ha tenido éxito, resultado proporcionado por el motor de reconocimiento de voz.

Breve descripción de las figuras

[0038] Estos aspectos, así como otros aspectos de la invención, quedarán más claros en la descripción detallada de formas de realización concretas de la invención, haciéndose referencia a los dibujos de las figuras, en las cuales:

la figura 1 muestra esquemáticamente a un orador que expresa un mensaje que es tratado por un motor de reconocimiento de voz;

la figura 2 muestra esquemáticamente un ejemplo de resultado de reconocimiento de voz;

la figura 3 muestra esquemáticamente diferentes etapas, y su interacción, de una variante preferida del método de la invención;

la figura 4 muestra esquemáticamente un ejemplo de sistema de postratamiento.

[0039] Los dibujos de las figuras no están trazados a escala. Por lo general, unos elementos similares se indican con referencias similares en las figuras. La presencia de números de referencia en los dibujos no se puede considerar limitativa, incluso cuando estos números se indiquen en las reivindicaciones.

Descripción detallada de formas de realización particulares

[0040] La figura 1 muestra a un orador 40 (o usuario 40) expresando un mensaje 50 a través de un micrófono 5. Este mensaje 50 se transfiere entonces hacia un motor de reconocimiento de voz 10 conocido por los expertos en la materia. En el mercado hay disponibles distintos modelos y diferentes marcas. Por lo general, el micrófono 5 forma parte del motor de reconocimiento de voz 10. Este último procesa el mensaje 50 con algoritmos de reconocimiento de voz basados, por ejemplo, en un modelo oculto de Markov (MMC). Da como resultado un resultado 100 de reconocimiento de voz. Un ejemplo de resultado 100 es una hipótesis generada por el motor de reconocimiento de voz 10. Otro ejemplo de resultado 100 es una solución obtenida a partir de algoritmos de reconocimiento de voz y a partir de operaciones de postratamiento que se aplican, por ejemplo, a una o varias hipótesis generadas por el motor de reconocimiento de voz 10. Unos módulos de postratamiento para proporcionar dicha solución pueden formar parte del motor de reconocimiento de voz 10. El resultado 100 se produce normalmente en forma de texto que puede ser descifrado, por ejemplo, por una máquina, un ordenador o una unidad de tratamiento. El resultado 100 se caracteriza por un inicio 111 y un fin 112. El inicio 111 es anterior a dicho fin 112 a lo largo de una escala temporal, t. El resultado 100 comprende una pluralidad de elementos 113 distribuidos temporalmente entre el inicio 111 y el fin 112. Un elemento 113 representa una información comprendida entre dos tiempos distintos a lo largo de la escala temporal, t. Por lo general, los distintos elementos 113 están separados por porciones del resultado 100 que representan un silencio, un ruido de fondo o un intervalo temporal durante el cual ningún elemento 113 (por ejemplo, palabra) es reconocido por el motor de reconocimiento de voz 10.

[0041] El método de la invención se refiere al postratamiento de un resultado 100 de reconocimiento de voz. En otras palabras, la entrada del método de la invención corresponde a un resultado 100 obtenido a partir de algoritmos de reconocimiento de voz aplicados a un mensaje 50 expresado por un orador 40 (o usuario 40). La figura 2 muestra un resultado 100 de reconocimiento de voz 100. Entre su inicio 111 y su fin 112, el resultado 100 comprende varios elementos 113, siete en el caso expuesto en la figura 2. En esta figura, los elementos 113 se representan en función del tiempo, t (abscisa). La ordenada, C, representa un nivel o índice de confianza. Los expertos en la materia conocen este concepto. Se trata de una propiedad o estadística asociada generalmente a cada elemento 113 y que puede ser proporcionada por lo general por un motor de reconocimiento de voz 100. Un índice de confianza representa, normalmente, una probabilidad de que un elemento del resultado de reconocimiento de voz, determinado por un motor de reconocimiento de voz 10 a partir de un elemento hablado, es el correcto. Los expertos en la materia conocen esta propiedad. Un ejemplo de motor de reconocimiento de voz es el modelo VoCon® 3200 V3.14 de Nuance. En este caso, el índice de confianza varía entre 0 y 10 000. Un valor de 0 se refiere a un valor mínimo de un índice de confianza (probabilidad muy baja de que el elemento del resultado de reconocimiento de voz es el correcto) y 10 000 representa un valor máximo de un índice de confianza (probabilidad muy alta de que el elemento del resultado de reconocimiento de voz es el correcto). En función de la altura de un elemento 113 en la figura 2, su índice de confianza 160 es más o menos elevado.

[0042] La primera etapa del método de la invención, la etapa i., consiste en recibir el resultado 100. A continuación, partiendo del fin 112, el método aísla un primer elemento 113. Por lo tanto, el método de la invención aísla en primer lugar el último elemento 113 del resultado a lo largo de la escala temporal, t. Una vez escogido este elemento 113, el método determina si es válido empleando una prueba de validación. Más abajo se presentan diferentes ejemplos de pruebas de validación. A continuación, se pasa al segundo elemento 113 partiendo del fin 112 y así sucesivamente. Se deja de recorrer los elementos 113 del resultado 100 a lo largo de la flecha en la parte superior de la figura 2 tan pronto como se detecta que un elemento 113 no es válido. A continuación, se determina una solución postratada 200 recuperando unos elementos 113 que han sido determinados como válidos, preferiblemente, utilizando todos los elementos 113 que han sido determinados como válidos. Durante la determinación de la solución postratada 200, es necesario mantener el orden correcto de los distintos elementos 113 seleccionados a lo largo de una escala temporal, t. Por consiguiente, hay que tener en cuenta que el primer elemento 113 tratado mediante el método de la invención representa el último elemento 113 del mensaje 100 y, por lo tanto, debe encontrarse al final en la solución postratada 200 si se ha determinado como válido. En general, un motor de reconocimiento de voz 10 proporciona, con los distintos elementos 113 del mensaje 100, unas informaciones temporales asociadas, por ejemplo, el inicio y el fin de cada elemento 113. Estas informaciones temporales asociadas se pueden utilizar para clasificar el orden correcto de los elementos determinados como válidos en la etapa iii.a., es decir, en un orden cronológico creciente.

[0043] Preferentemente, el método de la invención comprende una etapa de verificación de que la solución postratada 200 cumple una regla gramatical. Un ejemplo de regla gramatical es un número de palabras. Si la solución postratada 200 no cumple dicha regla gramatical, se puede decidir que no se proporcione. En este caso, en ocasiones se prefiere proporcionar el resultado 100 del motor de reconocimiento de voz 10. Si la solución postratada 200 cumple dicha regla gramatical, se preferirá entonces proporcionarla.

[0044] La figura 3 presenta esquemáticamente una versión preferida del método de la invención en la que:

- se deja de aislar (o escoger) un elemento 113 adicional para someterlo a la prueba de validación cuando se detecta un elemento 113 no válido, donde
- se verifica que la solución postratada 200 cumple una regla gramatical (etapa vi.), donde

- se proporciona la solución postratada 200 si esta cumple dicha regla gramatical, y donde
- se proporciona el resultado 100 del motor de reconocimiento de voz 10 si la solución postratada 200 no cumple dicha regla gramatical.

5 **[0045]** La etapa iii.a consiste en determinar si un elemento 113 seleccionado en la etapa ii. es válido utilizando una prueba de validación. Esta última puede adoptar diversas formas.

10 **[0046]** Un elemento 113 se caracteriza por un inicio y un fin. Este tiene por tanto una determinada duración 150. Según una variante posible, la prueba de validación comprende una etapa de consideración de un elemento 113 válido si su duración 150 es superior o igual a un umbral de duración inferior. El umbral de duración inferior está comprendido, por ejemplo, entre 50 y 160 milisegundos. Preferentemente, el valor del umbral de duración inferior es de 120 milisegundos. El umbral de duración inferior puede adaptarse dinámicamente. Según otra variante posible, la prueba de validación comprende una etapa de consideración de un elemento 113 válido si su duración 150 es inferior o igual a un umbral de duración superior. El umbral de duración superior está comprendido, por ejemplo, entre 400 y 800 milisegundos. Preferentemente, el valor del umbral de duración superior es de 600 milisegundos. El umbral de duración superior puede adaptarse dinámicamente. Preferentemente, el umbral de duración inferior y/o el umbral de duración superior es/son determinado(s) mediante una gramática.

20 **[0047]** En general, un índice de confianza 160 se asocia a cada elemento 113. Según otra variante posible, la prueba de validación comprende una etapa de consideración de un elemento 113 válido si su índice de confianza 160 es superior o igual a un índice de confianza mínimo 161. Preferentemente, este índice de confianza mínimo 161 puede variar de forma dinámica. En tal caso, es posible entonces que el índice de confianza mínimo 161 utilizado para determinar si un elemento 113 es válido sea distinto del utilizado para determinar si otro elemento 113 es válido o no. Los inventores han descubierto que un índice de confianza mínimo 161 comprendido entre 3500 y 5000 proporcionaba buenos resultados, siendo todavía preferido un valor de 4000 (valores para el modelo VoCon® 3200 V3.14 de Nuance, pero que se pueden extrapolar a otros modelos de motores de reconocimiento de voz).

25 **[0048]** Según otra variante posible, la prueba de validación comprende una etapa de consideración de un elemento 113 válido si un intervalo temporal 170 que lo separa de otro elemento 113 directamente adyacente hacia el fin 112 del resultado 100 es superior o igual a un intervalo de tiempo mínimo. Dicho intervalo de tiempo mínimo está comprendido, por ejemplo, entre cero y 50 milisegundos. Según otra variante posible, la prueba de validación comprende una etapa de consideración de un elemento 113 válido si un intervalo temporal 170 que lo separa de otro elemento 113 directamente adyacente hacia el fin 112 del resultado 100 es inferior o igual a un intervalo de tiempo máximo. Dicho intervalo de tiempo máximo está comprendido, por ejemplo, entre 300 y 600 milisegundos y un valor preferido es 400 ms. Para estos dos ejemplos de prueba de validación, se considera, por tanto, el intervalo de tiempo 170 que separa un elemento 113 de su vecino inmediato a la derecha en la figura 2. En otras palabras, se observa el intervalo de tiempo que separa un elemento 113 de su vecino inmediato a la derecha, es decir, su vecino posterior a lo largo de la escala temporal, t. Un intervalo de tiempo que separa dos elementos 113 es, por ejemplo, un intervalo de tiempo durante el cual un motor de reconocimiento de voz 10 no reconoce ningún elemento 113, por ejemplo, ninguna palabra.

40 **[0049]** Según otra variante posible, la prueba de validación se adapta al orador 40 (un usuario) que ha registrado el mensaje 50. Cada persona pronuncia unos elementos 113 o palabras de un modo particular. Por ejemplo, algunas personas pronuncian palabras lentamente, mientras que otras las pronuncian rápidamente. Del mismo modo, un índice de confianza 160 asociado a una palabra y proporcionado por un motor de reconocimiento de voz 10 depende, en general, del orador 40 que ha pronunciado esa palabra. Si se conocen una o varias estadísticas asociadas a diferentes elementos 113 para un orador 40 determinado, pueden emplearse durante la prueba de validación de la etapa iii.a. para determinar si un elemento 113 es válido o no. Por ejemplo, se puede considerar que un elemento 113, expresado por un orador 40 determinado, es válido si una o varias estadísticas asociada(s) a este elemento 113 es/son acorde(s), en un intervalo de error cercano (10 %, por ejemplo), a la misma o mismas estadística(s) preestablecida(s) por este mismo elemento 113 para este mismo orador 40. Esta variante preferida de la prueba de validación requiere que se conozca la identidad del orador 40. Esta puede ser proporcionada, por ejemplo, por el motor de reconocimiento de voz 10. Según otra posibilidad, el método de postratamiento de la invención comprende una etapa de identificación del orador 40.

55 **[0050]** En la figura 2, unos elementos 113 considerados como válidos están delimitados por unos trazos continuos, mientras que unos elementos no considerados como válidos están delimitados por unos trazos discontinuos. El cuarto elemento 113 partiendo desde el fin 112 se considera como no válido, por ejemplo, porque su duración 150 es menor que un umbral de duración inferior. El quinto elemento 113 partiendo desde el fin 112 se considera como no válido, por ejemplo, porque su índice de confianza 160 es inferior a un índice de confianza mínimo 161.

[0051] Asimismo, los inventores proponen un método para generar una solución optimizada a partir de un primer y un segundo resultado 100 de reconocimiento de voz y que comprende las siguientes etapas:

- A. aplicar un método de postratamiento de acuerdo con el primer aspecto de la invención a dicho primer resultado 100;

B. aplicar un método de postratamiento de acuerdo con el primer aspecto de la invención a dicho segundo resultado 100;

C. determinar dicha solución optimizada a partir de uno o varios elementos 113 perteneciente(s) a uno o varios resultados 100 de dichos primer y segundo resultado 100 y que han sido determinados como válidos mediante la prueba de validación de la etapa iii.a.

[0052] El siguiente ejemplo se refiere a un sistema de postratamiento 11 o dispositivo de postratamiento de un resultado 100 de reconocimiento de voz. La figura 4 representa esquemáticamente dicho sistema de postratamiento 11 en combinación con un motor de reconocimiento de voz 10 y una pantalla 20. En esta figura, el sistema de postratamiento 11 y el motor de reconocimiento de voz 10 son dos dispositivos separados. Según otra versión posible, el sistema de postratamiento 11 está integrado en un motor de reconocimiento de voz 10, de manera que no sea posible diferenciarlos. En dicho caso, un motor de reconocimiento de voz 10 tradicional se modifica o se adapta para poder realizar las funciones del sistema de postratamiento 11 descritas a continuación.

[0053] Ejemplos de sistema de postratamiento 11 son: un ordenador, un motor de reconocimiento de voz 10 adaptado o programado para poder llevar a cabo un método postratamiento de acuerdo con el primer aspecto de la invención, un módulo físico (o *hardware*) de un motor de reconocimiento de voz 10, un módulo físico capaz de comunicarse con un motor de reconocimiento de voz 10. No obstante, son posibles otros ejemplos. El sistema de postratamiento 11 comprende unos medios de obtención 12 para recibir y leer un resultado 100 de reconocimiento de voz. Ejemplos de medios de obtención 12 son: un puerto de entrada del sistema de postratamiento 11, por ejemplo, un puerto USB, un puerto Ethernet, un puerto inalámbrico (por ejemplo, wifi). No obstante, son posibles otros ejemplos de medios de obtención 12. El sistema de postratamiento 11 comprende, además, unos medios de tratamiento 13 para llevar a cabo las siguientes etapas recursivamente: aislar, desde el fin 112 hasta el inicio 111 del resultado 100, un elemento 113 del resultado 100 y que no se haya sometido anteriormente a una prueba de validación de los medios de tratamiento 13, determinar si es válido empleando una prueba de validación, determinar una solución postratada 200 recuperando al menos un elemento 113 determinado como válido por dichos medios de tratamiento 13. Preferentemente, dichos medios de tratamiento 13 determinan una solución postratada 200 recuperando todos los elementos 113 determinados como válidos por dichos medios de tratamiento 13. Preferentemente, el sistema de postratamiento 11 es capaz de enviar la solución postratada 200 a una pantalla 20 para su visualización.

[0054] Ejemplos de medios de tratamiento 13 son: una unidad de control, un procesador o unidad central de procesamiento, un controlador, un chip, un microchip, un circuito integrado, un procesador de múltiples núcleos. No obstante, son posibles otros ejemplos conocidos por los expertos en la materia. De acuerdo con una versión posible, los medios de tratamiento 13 comprenden distintas unidades para llevar a cabo las diferentes etapas anteriormente mencionadas y en relación con estos medios de tratamiento 13 (aislar un elemento 113, determinar si es válido, determinar una solución postratada 200).

[0055] Según un segundo aspecto, la invención se refiere a un programa, preferiblemente un programa informático. Preferiblemente, este programa forma parte de una interfaz de voz hombre-máquina.

[0056] Según un tercer aspecto, la invención se refiere a un medio de almacenamiento que puede estar conectado a un dispositivo, por ejemplo, a un ordenador que puede comunicarse con un motor de reconocimiento de voz 10. Según otra variante posible, este dispositivo es un motor de reconocimiento de voz 10. Ejemplos de medios de almacenamiento de acuerdo con la invención son: una memoria USB, un disco duro externo, un disco de tipo CD-ROM. No obstante, son posibles otros ejemplos.

[0057] La presente invención ha sido descrita en relación con unas formas de realización específicas, que tienen un valor meramente ilustrativo y no deben considerarse como limitativas. En general, la presente invención no se limita a los ejemplos anteriormente descritos y/o representados. El uso de los verbos «comprender», «incluir», «contener» o cualquier otra variante, así como sus conjugaciones, no puede excluir en absoluto la presencia de elementos distintos a los mencionados. El uso del artículo indefinido «un», «una» o del artículo definido «el» o «la» para introducir un elemento no excluye la presencia de una pluralidad de estos elementos. Los números de referencia en las reivindicaciones no limitan su alcance.

REIVINDICACIONES

1. Método de postratamiento de un resultado (100) de reconocimiento de voz, comprendiendo dicho resultado (100) un inicio (111), un fin (112) y una pluralidad de elementos (113) distribuidos entre dicho inicio (111) y dicho fin (112), comprendiendo dicho método de postratamiento las siguientes etapas:
- 5 i. recibir dicho resultado (100);
- ii. aislar un elemento (113) de dicha pluralidad de elementos (113) que no se ha sometido a la prueba de validación de la etapa iii.a.;
- iii. entonces,
- 10 a. si se ha aislado un elemento (113) en la etapa ii., determinar si es válido utilizando una prueba de validación,
- b. si no, pasar directamente a la etapa v.;
- iv. repetir las etapas ii. y iii.;
- v. si al menos un elemento (113) se ha determinado como válido en la etapa iii.a, determinar una solución postratada (200) utilizando al menos un elemento (113) determinado como válido en la etapa iii.a;
- 15 **caracterizado por que** cada elemento (113) aislado en la etapa ii. se selecciona desde dicho fin (112) del resultado (100) a dicho inicio (111) del resultado (100) de manera consecutiva y **por que** la etapa iii.a. comprende, además, una instrucción para pasar directamente a la etapa v. si el elemento (113) sometido a la prueba de validación de la etapa iii.a no ha sido determinado como válido.
2. Método de acuerdo con la reivindicación 1 **caracterizado por que** dichos elementos (113) son palabras.
- 20 3. Método de acuerdo con cualquiera de las reivindicaciones anteriores **caracterizado por que** comprende, además, la siguiente etapa:
- vi. determinar si dicha solución postratada (200) de la etapa v. cumple una regla gramatical.
4. Método de acuerdo con la reivindicación anterior **caracterizado por que** comprende, además, la siguiente etapa:
- 25 vii.
- a. si la respuesta a la prueba de la etapa vi. es positiva, proporcionar dicha solución postratada (200),
- b. si no, proporcionar dicho resultado (100) de reconocimiento de voz.
5. Método de acuerdo con cualquiera de las reivindicaciones anteriores **caracterizado por que** dicha prueba de validación de la etapa iii.a. comprende una etapa de consideración de un elemento (113) válido si su duración es superior o igual a un umbral de duración inferior.
- 30 6. Método de acuerdo con cualquiera de las reivindicaciones anteriores **caracterizado por que** dicha prueba de validación de la etapa iii.a. comprende una etapa de consideración de un elemento (113) válido si su duración es inferior o igual a un umbral de duración superior.
7. Método de acuerdo con cualquiera de las reivindicaciones anteriores **caracterizado por que** cada elemento (113) de dicho resultado (100) está **caracterizado por** un índice de confianza (160) y **por que** dicha prueba de validación de la etapa iii.a. comprende una etapa de consideración de un elemento (113) válido si su índice de confianza (160) es superior o igual a un índice de confianza mínimo (161).
- 35 8. Método de acuerdo con cualquiera de las reivindicaciones anteriores **caracterizado por que** dicha prueba de validación de la etapa iii.a. comprende una etapa de consideración de un elemento (113) válido si un intervalo temporal (170) que lo separa de otro elemento (113) directamente adyacente hacia dicho fin (112) del resultado (100) es superior o igual a un intervalo de tiempo mínimo.
- 40 9. Método de acuerdo con cualquiera de las reivindicaciones anteriores **caracterizado por que** dicha prueba de validación de la etapa iii.a. comprende una etapa de consideración, para un orador (40) determinado, de un elemento (113) de dicho resultado (100) válido, si una estadística asociada a este elemento (113) se ajusta, en un intervalo cercano, a una estadística preestablecida para un mismo elemento (113) y para este orador (40) determinado.
- 45 10. Método de acuerdo con cualquiera de las reivindicaciones anteriores **caracterizado por que** todos los elementos (113) determinados como válidos en la etapa iii.a se recuperan para determinar dicha solución postratada (200) en la etapa v.
- 50 11. Método para determinar una solución optimizada a partir de un primer y un segundo resultado (100) de reconocimiento de voz y comprendiendo las siguientes etapas:

A. aplicar un método de postratamiento de acuerdo con cualquiera de las reivindicaciones anteriores a dicho primer resultado (100);

B. aplicar un método de postratamiento de acuerdo con cualquiera de las reivindicaciones anteriores a dicho segundo resultado (100);

5 C. determinar dicha solución optimizada a partir de uno o varios elementos (113) perteneciente(s) a uno o varios resultados (100) de dichos primer y segundo resultado (100) y que han sido determinados como válidos mediante la prueba de validación de la etapa iii.a.

12. Programa que permite tratar un resultado (100) de reconocimiento de voz, comprendiendo dicho resultado (100) un inicio (111), un fin (112) y una pluralidad de elementos (113) distribuidos entre dicho inicio (111) y dicho fin (112), comprendiendo dicho programa un código para permitir que un dispositivo lleve a cabo las siguientes etapas:

i. leer dicho resultado (100) de reconocimiento de voz.

ii. aislar un elemento (113) de dicha pluralidad de elementos (113) que no se ha sometido a la prueba de validación de la etapa iii.a.,

15 iii. entonces,

a. si se ha aislado un elemento (113) en la etapa ii., determinar si es válido utilizando una prueba de validación,

b. si no, pasar directamente a la etapa v.

iv. repetir las etapas ii. y iii.;

20 v. si al menos un elemento (113) se ha determinado como válido en la etapa iii.a, determinar una solución postratada (200) recuperando al menos un elemento (113) determinado como válido en la etapa iii.a;

caracterizado por que cada elemento (113) aislado en la etapa ii se selecciona de dicho fin (112) del resultado (100) a dicho inicio (111) del resultado (100) de manera consecutiva y **por que** la etapa iii.a. comprende, además, una instrucción para pasar directamente a la etapa v. si el elemento (113) sometido a la prueba de validación de la etapa iii.a no ha sido determinado como válido.

13. Medio de almacenamiento que puede estar conectado a un dispositivo y que comprende unas instrucciones que, al leerse, permiten que dicho dispositivo trate un resultado (100) de reconocimiento de voz, comprendiendo dicho resultado (100) un inicio (111), un fin (112) y una pluralidad de elementos (113) distribuidos entre dicho inicio (111) y dicho fin (112), permitiendo dichas instrucciones que se obligue a dicho dispositivo a llevar a cabo las siguientes etapas:

i. leer dicho resultado (100);

ii. aislar un elemento (113) de dicha pluralidad de elementos (113) que no se ha sometido a la prueba de validación de la etapa iii.a.,

iii. entonces,

35 a. si se ha aislado un elemento (113) en la etapa ii., determinar si es válido utilizando una prueba de validación,

b. si no, pasar directamente a la etapa v.,

iv. repetir las etapas ii. y iii.;

40 v. si al menos un elemento (113) se ha determinado como válido en la etapa iii.a, determinar una solución postratada (200) recuperando al menos un elemento (113) determinado como válido en la etapa iii.a;

caracterizado por que cada elemento (113) aislado en la etapa ii. se selecciona de dicho fin (112) del resultado (100) a dicho inicio (111) del resultado (100) de manera consecutiva y **por que** la etapa iii.a. comprende, además, una instrucción para pasar directamente a la etapa v. si el elemento (113) sometido a la prueba de validación de la etapa iii.a no ha sido determinado como válido.

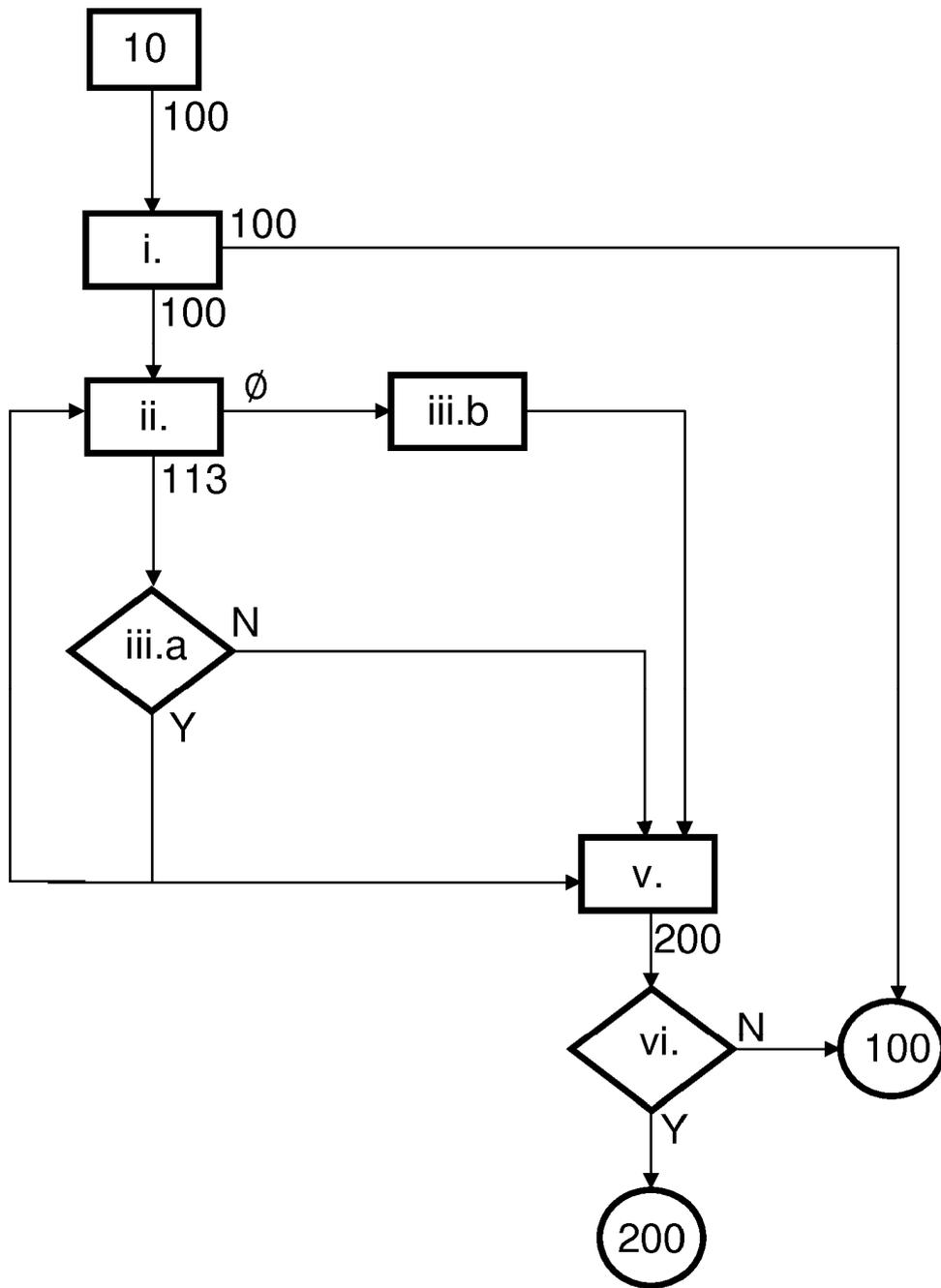


Fig. 3

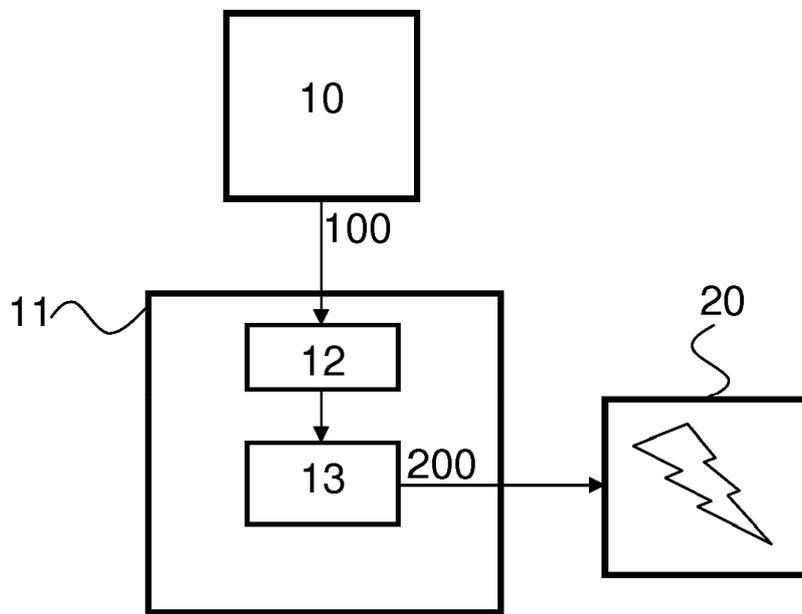


Fig. 4