



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 312 851**

51 Int. Cl.:
G10L 13/08 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **03799483 .7**

96 Fecha de presentación : **16.12.2003**

97 Número de publicación de la solicitud: **1721311**

97 Fecha de publicación de la solicitud: **15.11.2006**

54

Título: **Procedimiento y sistema texto a voz y el programa informático asociado.**

45

Fecha de publicación de la mención BOPI:
01.03.2009

45

Fecha de la publicación del folleto de la patente:
01.03.2009

73

Titular/es: **LOQUENDO S.p.A.**
Via Arrigo Olivetti 6
10100 Torino, IT

72

Inventor/es: **Badino, Leonardo;**
Barolo, Claudia y
Quazza, Silvia

74

Agente: **Ponti Sales, Adelaida**

ES 2 312 851 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).

ES 2 312 851 T3

DESCRIPCIÓN

Procedimiento y sistema texto a voz y el programa informático asociado.

5 **Campo de la técnica**

La presente invención se refiere a técnicas de texto a voz, concretamente las técnicas que permiten que un texto escrito se transforme en una señal de voz inteligible.

10 **Descripción de la técnica relacionada**

Los sistemas de texto a voz se conocen por estar basados en la denominada “síntesis concatenativa por selección de unidades”. Esto requiere una base de datos que incluya frases pregrabadas pronunciadas por hablantes nativos. La base de datos vocálica es de una sola lengua, ya que todas las frases se escriben y se pronuncian en la lengua del hablante.

15 Los sistemas de texto a voz de este tipo pueden, por tanto, “leer” correctamente solo los textos escritos en la lengua del hablante, mientras que cualquier palabra extranjera incluida en el texto solo se podrá pronunciar de manera inteligible si se incluye (junto con su fonetización correcta) en el léxico proporcionado como refuerzo del sistema texto a voz. Consecuentemente, los textos plurilingües solo pueden leerse correctamente en dichos sistemas cambiando la voz del hablante en el caso de un cambio en la lengua. Esto provoca un efecto que no suele ser agradable, y resulta cada vez más evidente cuando los cambios en la lengua ocurren frecuentemente y son de corta duración.

25 Adicionalmente, un hablante corriente que tenga que pronunciar palabras extranjeras de un texto en su propia lengua, normalmente tendrá tendencia a pronunciar esas palabras de un modo diferente, y también significativamente, a la manera correcta de pronunciar las mismas palabras cuando se incluyen en un texto completo en la lengua extranjera correspondiente.

30 A modo de ejemplo, un hablante británico o americano que tenga que pronunciar, por ejemplo, un nombre o apellido italiano incluido en un texto en inglés adoptará, en condiciones generales, una pronunciación bastante diferente a la pronunciación tomada por un hablante nativo de italiano a la hora de pronunciar el mismo nombre y apellido. Correspondientemente, una persona que hable inglés y que escuche el mismo texto hablado, generalmente pensará que el nombre y apellido italianos son más fáciles de entender (aproximadamente como mínimo) si han sido pronunciados como esperaba de forma “tergiversada” por un hablante inglés que si los hubiera pronunciado con la correcta pronunciación italiana.

35 De forma similar, la pronunciación de, por ejemplo, el nombre de una ciudad en Reino Unido o Estados Unidos incluida en un texto italiano leído por un hablante italiano adoptando la pronunciación de inglés británico o americano correcta normalmente se considerará como una sofisticación excesiva y, como tal, se rechazará en el uso común.

40 El problema de leer un texto plurilingüe ya se ha abordado en el pasado adoptando dos enfoques diferentes.

Por una parte, se realizaron intentos de producir bases de datos plurilingües recurriendo a hablantes bilingües o plurilingües. Un ejemplo de dicho enfoque es el artículo de C. Traber *et al.*: “From multilingual to polyglot speech synthesis” - Proceedings of the Eurospeech, páginas 835-838, 1999.

45 Este enfoque se basa en supuestos (esencialmente, la disponibilidad de un hablante plurilingüe) que son difíciles de encontrar y reproducir. Además, dicho enfoque no suele solucionar el problema asociado generalmente a las palabras extranjeras incluidas en un texto que se espera que se pronuncien de manera diferente (posiblemente notable) a la correcta pronunciación en la lengua correspondiente.

50 Otro enfoque es adoptar un transcriptor de lengua extranjero y los fonemas producidos en su producción que, para ser pronunciados, se asignan a los fonemas de las lenguas de la voz hablante. Un ejemplo de este último enfoque son los trabajos de W.N. Campbell “Foreign-language speech synthesis” Proceedings ESCA/COCSDA ETRW on Speech Synthesis, Jenolan Caves, Australia, 1998 and “Talking Foreign. Concatenative Speech Synthesis and Language Barrier”, Proceedings of the Eurospeech Scandinavia, pages 337 - 340, 2001.

55 Los trabajos de Campbell tienen como objetivo esencial sintetizar un texto bilingüe por ejemplo, en inglés y japonés, basándose en la voz generada partiendo de una base de datos japonesa monolingüe. Si la voz hablante es japonesa y el texto de entrada inglés, se activa un transcriptor inglés para producir fonemas ingleses. La similitud se evalúa según las categorías fonético-articulatorias. Un módulo de asignación fonético asigna cada fonema inglés a un fonema japonés correspondiente, similar. Esta similitud se evalúa basándose en las categorías fonético-articulatorias. La asignación se lleva a cabo mediante una búsqueda en una tabla de consulta que proporciona una correspondencia entre los fonemas japoneses y los ingleses.

65 Como etapa subsiguiente, las diversas unidades acústicas destinadas a componer la lectura realizada por un voz japonesa se seleccionan de la base de datos japonesa basándose en sus similitudes acústicas con las señales generadas al sintetizar el mismo texto con una voz inglesa.

ES 2 312 851 T3

La parte principal del procedimiento propuesto por Campbell es una tabla de consulta que expresa la correspondencia entre fonemas en las dos lenguas. Dicha tabla se crea manualmente investigando las características de las dos lenguas en cuestión.

5 En principio, dicho enfoque es aplicable a cualquier otro par de lenguas, pero cada par de lenguas requiere un análisis explícito de la correspondencia entre ellos. Dicho enfoque es de algún modo incómodo, y de hecho prácticamente inviable en el caso de un sistema de síntesis que incluya más de dos lenguas, ya que el número de parejas de lenguas a tener en cuenta sería demasiado alto.

10 Adicionalmente, se suele utilizar más de un hablante para cada lengua, teniendo sistemas fonológicos al menos algo diferentes. Para que cualquier voz hablante esté en condiciones de hablar todas las lenguas disponibles, sería necesaria una tabla para cada voz-par de lenguas.

15 En el caso de un sistema de síntesis que incluya N lenguas y M voces hablantes (evidentemente, M es igual o mayor que N), con tablas de consulta para la primera etapa de asignación fonética, si los fonemas de una voz hablante se asignan a los de una sola voz para cada lengua extranjera, entonces se tendrán que crear N-1 tablas diferentes para cada voz hablante, resultando así en un total de $N*(M-1)$ tablas de consulta.

20 En el caso de un sistema de síntesis que funcione con quince lenguas y dos voces hablantes para cada lengua (que se corresponde con la disposición actual adoptada en el sistema texto a voz de Loquendo TTS desarrollado por el cesionario de la presente invención), se necesitarían 435 tablas de consulta. Este dato es bastante importante, especialmente si se tiene en cuenta la posible necesidad de generar dichas tablas de consulta manualmente.

25 Ampliar dicho sistema para incluir solo una voz hablante que hable una lengua nueva requeriría añadir $M+N=45$ nuevas tablas. En ese sentido, se debe tener en cuenta que frecuentemente se añaden nuevos fonemas a los sistemas de texto a voz para una o más lenguas, siendo un caso frecuente que el nuevo fonema añadido sea un alófono de un fonema que ya existe en el sistema. En ese caso, existirá la necesidad de revisar y modificar todas las tablas de consulta que pertenezcan a la(s) lengua (s) en las que se ha(n) añadido el nuevo fonema.

30 Objeto y resumen de la invención

En vista de lo explicado anteriormente, existe la necesidad de mejorar los sistemas texto a voz para que prescindan de los inconvenientes de la técnica anterior en cuanto a las disposiciones descritas anteriormente. Más específicamente, el objeto de la presente invención es proporcionar un sistema texto a voz multilingüe que:

- 35
- pueda prescindir del requisito de confiar en hablantes plurilingües, y
 - pueda implementarse recurriendo a estructuras simples, requiriendo memorias moderadas, y que al mismo tiempo prescinda también de la necesidad de generar (posiblemente de forma manual) un número relevante de tablas de consulta, especialmente cuando el sistema se mejora con la adición de un fonema nuevo para una o varias lenguas.
- 40

Según la presente invención, este objeto se consigue mediante un procedimiento que tenga las características expuestas en la reivindicación 1. La invención también hace referencia al sistema texto a voz correspondiente, como se expone en la reivindicación 7, y un producto de un programa informático que se puede cargar en la memoria de al menos un ordenador y que comprenda porciones de código de software para realizar las etapas del procedimiento de la invención cuando el producto se ejecuta en un ordenador, como se expone en la reivindicación 13. Como se utiliza aquí, la referencia a dicho producto de programa informático se intenta que sea equivalente a la referencia de soportes legibles por ordenador que contengan instrucciones para controlar un sistema informático para coordinar la realización del procedimiento de la invención. La referencia "al menos un ordenador" está evidentemente dirigida a resaltar la posibilidad de que el sistema de la invención se implemente de forma distribuida.

Una realización preferida de la invención es, por tanto, una disposición para la conversión texto a voz de un texto en una primera lengua incluyendo secciones en como mínimo una segunda lengua, incluyendo:

- 55
- un grafema/morfema transcriptor para convertir dichas secciones de dicha segunda lengua en fonemas de dicha segunda lengua,
 - un módulo de asignación configurado para asignar como mínimo una parte de dichos fonemas de dicha segunda lengua a conjuntos de fonemas de dicha primera lengua.
- 60
- un módulo de síntesis de voz adaptado para ser alimentado por un flujo de fonemas resultante que incluya dichos conjuntos de fonemas de dicha primera lengua resultantes de dicha asignación y del flujo de fonemas de dicha primera lengua representativos de dicho texto, y para generar una señal de voz a partir de dicho flujo de fonemas resultante; el módulo de asignación está configurado para:
- 65
- llevar a cabo pruebas de similitud entre cada uno de dichos fonemas de dicha segunda lengua que están siendo asignados y un conjunto de fonemas de asignación candidatos de dicha primera lengua,

ES 2 312 851 T3

- asignar las respectivas puntuaciones a los resultados de dichas pruebas, y
- asignar dicho fonema de dicha segunda lengua en un conjunto de fonemas de asignación de dicha primera lengua seleccionada fuera de dichos fonemas de asignación candidatos como una función de dichas puntuaciones.

5

Preferentemente, el módulo de asignación está configurado para asignar dicho fonema de dicha segunda lengua a un conjunto de fonemas de asignación de dicha primera lengua seleccionado entre:

10

- un conjunto de fonemas de dicha primera lengua incluyendo tres, dos o un fonema de dicha primera lengua, o
- un conjunto vacío, mediante el que no se incluye ningún fonema en dicho flujo resultante para dicho fonema en dicha segunda lengua.

15

Generalmente, la asignación a dicho conjunto vacío de fonemas de dicha primera lengua ocurre en los fonemas de dicha segunda lengua en los que ninguna de dichas puntuaciones puede alcanzar un valor umbral.

El flujo de fonemas resultante puede ser, por tanto, pronunciado por una voz hablante de dicha primera lengua.

20

Esencialmente, la disposición aquí descrita está basada en una disposición de asignación fonética donde cada una de las voces hablantes incluidas en el sistema es capaz de leer un texto plurilingüe sin modificar la base de datos vocálica. Específicamente, una realización preferida de la disposición descrita busca, entre los fonemas presentes en la tabla para la lengua de la voz hablante, el fonema que resulta más parecido al fonema de la lengua extranjera recibido como entrada. El grado de similitud entre los dos fonemas se puede expresar sobre la base de rasgos fonético-articulatorios, como se define, por ejemplo, según el AFI estándar. Una tabla de asignación fonética cuantifica el grado de afinidad/similitud de las categorías fonéticas y la importancia de cada una de ellas en la comparación ente fonemas.

25

La disposición aquí descrita no incluye ninguna comparación “acústica” entre los segmentos incluidos en la base de datos de la lengua de la voz hablante y la señal sintetizada mediante la voz hablante de la lengua extranjera. Consecuentemente, toda la disposición es menos incómoda desde el punto de vista computacional y prescinde de la necesidad de que el sistema tenga una voz hablante disponible para la lengua extranjera: solo el transcriptor grafemamorfema será suficiente.

30

Adicionalmente, la asignación fonética es independiente de la lengua. La comparación entre fonemas se refiere exclusivamente al vector de los rasgos fonéticos asociados a cada fonema, siendo de hecho estos rasgos independientes de la lengua. El módulo de asignación, por tanto, no es “consciente” de las lenguas implicadas, lo que significa que no existen requisitos para ninguna actividad específica que se tenga que realizar (posiblemente de forma manual) para cada par de lenguas (o para cada par voz-lengua) en el sistema. Además, la incorporación de nuevas lenguas o nuevos fonemas al sistema no precisa modificaciones en el módulo de asignación fonética.

35

Sin pérdidas en cuestiones de efectividad, la disposición aquí descrita lleva a una simplificación apreciable en comparación con el sistema de la técnica anterior, mientras que también implica un mayor grado de generalización respecto a las soluciones previas.

40

Los experimentos llevados a cabo muestran que el objeto de utilizar una voz hablante monolingüe para hablar lenguas extranjeras de manera inteligible se consigue por completo.

45

Breve descripción de los dibujos adjuntos

50

La invención se describirá ahora solo mediante un ejemplo refiriéndose a las figuras anexas del dibujo, donde:

- la figura 1 es un diagrama de bloques de un sistema texto a voz adaptado para incorporar las mejoras aquí descritas, y

55

- las figuras 2 a 8 son diagramas de flujo que ejemplifican un posible funcionamiento del sistema texto a voz de la figura 1.

Descripción detallada de las realizaciones preferidas de la invención

60

El diagrama de bloques de la figura 1 muestra la arquitectura general de un sistema texto a voz de tipo plurilingüe.

Esencialmente, el sistema de la figura 1 está adaptado para recibir como su texto de llegada que califica básicamente como texto plurilingüe.

65

En el contexto de la invención, la importancia de la definición de plurilingüe es doble:

- en primer lugar, el texto de entrada es plurilingüe porque se corresponde con un texto escrito en cualquiera de una variedad de lenguas diferentes T1,..., Tn como por ejemplo, quince lenguas diferentes, y

ES 2 312 851 T3

- en segundo lugar, cada uno de los textos T1,..., Tn es *per se* plurilingüe, ya que puede incluir palabras o frases en una o más lenguas distintas a la lengua básica del texto.

El texto T1,..., Tn se proporciona al sistema (generalmente designado 10) en formato de texto electrónico.

El texto originalmente disponible en formas diferentes (por ejemplo, como copias impresas de un texto impreso) se puede pasar fácilmente a formato electrónico recurriendo a técnicas como la lectura de escáner OCR. Estos métodos son muy conocidos en la técnica, siendo por tanto innecesario proporcionar una descripción detallada.

Un primer bloque en el sistema 10 está representado por un módulo de reconocimiento del lenguaje 20 adaptado para reconocer tanto el lenguaje básico de un texto de entrada en el sistema como la(s) lengua(s) de cualquier palabra o frase extranjera incluida en el texto básico.

De nuevo, los módulos adaptados para realizar automáticamente dicha función de reconocimiento de la lengua son muy conocidos en la técnica (por ejemplo, por los correctores ortográficos de los sistemas de procesamiento de texto), siendo por tanto innecesario proporcionar una descripción detallada.

A continuación, al describir una realización ejemplar de la invención, se hará referencia a una situación en la que el texto de entrada básico es un texto italiano que incluye palabras o frases cortas en inglés. La voz hablante también será italiana.

A continuación del módulo de reconocimiento del lenguaje 20 hay tres módulos 30, 40 y 50, uno detrás del otro.

Específicamente, el módulo 30 es un transcriptor grafema/morfema adaptado para segmentar el texto recibido como una entrada en grafemas (por ejemplo, letras o grupos de letras) y convertirlo en un flujo de fonemas correspondiente. El módulo 30 puede ser un transcriptor grafema/fonema de un tipo conocido incluido en el sistema texto a voz Loquendo TTS, ya mencionado anteriormente.

Esencialmente, la salida del módulo 30 será un flujo de fonemas incluyendo fonemas en el lenguaje básico del texto de entrada (por ejemplo, italiano) que se hayan descompuesto en "ráfagas" de fonemas en la(s) lengua(s) (por ejemplo, inglés) consistiendo en las palabras o frases cortas de la lengua extranjera incluidas en el texto básico.

La referencia 40 designa una tabla de asignación cuya estructura y funcionamiento se detallará más adelante. Básicamente, el módulo 40 convierte el flujo mixto de fonemas que salen del módulo 30, comprendiendo tanto los fonemas del lenguaje básico (italiano) del texto de entrada como los fonemas de la lengua extranjera (inglés), en un flujo de fonemas que incluyen solo fonemas de la primera lengua básica, concretamente italiano en el ejemplo considerado.

Finalmente, el módulo 50 es un módulo de síntesis de voz adaptado para generar a partir del flujo de fonemas (italianos) procedentes del módulo 40 una señal de voz sintetizada que se proporcionará a un altavoz 60 para que genere una señal de voz acústica correspondiente para que sea percibida, escuchada y entendida por los hombres.

Un módulo de síntesis de señal de voz como el módulo 60 mostrado aquí es un componente básico de cualquier señal texto a voz, siendo por tanto innecesario proporcionar aquí una descripción detallada.

A continuación se describe el funcionamiento del módulo 40.

Esencialmente, el módulo 40 está compuesto de una primera y segunda porción designada 40a y 40b, respectivamente.

La primera porción 40a está configurada esencialmente para pasar al módulo 50 aquellos fonemas que ya son fonemas de la lengua básica (italiano, en el ejemplo considerado).

El segundo punto 40b incluye una tabla de fonemas de la voz hablante (italiano) y recibe como una entrada el flujo de fonemas en lengua extranjera (inglés) que se tienen que asignar a los fonemas de la lengua de la voz hablante (italiano) para permitir que dicha voz los pronuncie.

Como se indica anteriormente, el módulo 20 indica al módulo 40 cuando, dentro del marco de un texto en una lengua dada, aparece una palabra o frase en una lengua extranjera. Esto ocurre gracias a un "cambio de señal" enviado por el módulo 20 al módulo 40 por encima de la línea 24.

De nuevo se recuerda que la referencia al italiano y al inglés como dos lenguas implicadas en el proceso de conversión texto a voz tiene simplemente una naturaleza ejemplar. De hecho, una ventaja básica de la disposición aquí descrita reside en que la asignación fonética, como se realiza en la porción 40b del módulo 40, es independiente del lenguaje. El módulo 40 no es consciente de las lenguas implicadas, lo que significa que no existen requisitos para que se lleve a cabo (posiblemente de forma manual) cualquier actividad específica para cada par de lenguas (o cada par voz-lengua) en el sistema.

ES 2 312 851 T3

Esencialmente, en el módulo 40 cada fonema de lengua extranjera se compara con todos los fonemas presentes en la tabla (que también puede incluir fonemas que, *per se*, no son fonemas de la lengua básica).

Consecuentemente, un número variable de fonemas de salida puede corresponder a cada fonema de entrada: por ejemplo, tres fonemas, dos fonemas, un fonema o ningún fonema.

Por ejemplo, un diptongo extranjero se comparará con los diptongos de la voz hablante así como con parejas de vocales.

Se asocia una puntuación con cada comparación realizada.

Los fonemas finalmente elegidos serán aquellos que tengan la mayor puntuación y un valor mayor que el valor umbral. Si ninguno de los fonemas en la voz hablante alcanza el valor umbral, el fonema de la lengua extranjera se asignará a un fonema cero y, por tanto, no se producirá ningún sonido para ese fonema.

Cada fonema está definido de una manera inequívoca mediante un vector de n categorías fonético-articulatorias de longitudes variables. Las categorías, definidas según el estándar AFI, son las siguientes:

- (a) las dos categorías básicas vocal y consonante;
- (b) la categoría diptongo;
- (c) las características vocálicas (por ejemplo, vocal) acentuada/no acentuada, no silábica, larga, nasal, rótica, redondeada;
- (d) las categorías vocálicas anterior, central, posterior;
- (e) las categorías vocálicas cerrada, cerrada-mediocerrada, mediocerrada, mediocerrada, media, medioabierta, abierta-medioabierta, abierta;
- (f) las categorías del modo consonántico oclusiva, nasal, vibrante múltiple, vibrante simple, fricativa, fricativa lateral, aproximante, lateral, africada;
- (g) las categorías del lugar consonántico bilabial, labiodental, dental, alveolar, postalveolar, retrofleja, palatal, velar, uvular, faríngea, glotal; y
- (h) las otras categorías consonánticas sonora, larga, silábica, aspirada, oclusiva, sorda, semiconsonante.

Actualmente, la categoría semiconsonante no es un rasgo AFI estándar. Esta categoría es una categoría redundante utilizada para simplemente denotar una consonante aproximante/alveolar/palatal o una consonante aproximante-velar.

Las categorías (d) y (e) también describen el segundo componente de un diptongo.

Cada vector contiene una categoría (a), una o ninguna categoría (b) si el fonema es una vocal, una categoría (c) como mínimo si el fonema es una vocal, una categoría (e) si el fonema es una vocal, una categoría (f) si el fonema es una consonante, como mínimo una categoría (g) si el fonema es una consonante y como mínimo una categoría (h) si el fonema es una consonante.

La comparación entre fonemas se realiza mediante la comparación de los vectores correspondientes, repartiendo las puntuaciones respectivas a dichas comparaciones vector por vector.

La comparación entre vectores se realiza comparando las categorías correspondientes, estando dichas puntuaciones respectivas agregadas para generar dichas puntuaciones.

Cada comparación categoría por categoría ha asociado un peso diferenciado, de modo que diferentes comparaciones categoría por categoría pueden tener pesos diferentes a la hora de generar la puntuación correspondiente.

Por ejemplo, una puntuación máxima obtenida comparando categorías (f) siempre será menor que la puntuación obtenida al comparar categorías (g) (por ejemplo, el peso asociado a la comparación de la categoría (f) es mayor que el peso asociado a la comparación de la categoría (g)). Como consecuencia, la afinidad entre vectores (puntuación) estará principalmente influenciada por la similitud entre categorías (f), comparada con la similitud entre categorías (g).

El proceso descrito a continuación utiliza un conjunto de constantes que tiene preferiblemente el siguiente valor:

- MaxCount = 100
- Kopen = 14

ES 2 312 851 T3

- Sstep = 1-
- Mstep = 2* Lstep
- 5 - Lstep = 4* Mstep
- Kmode = Kopen + (Lstep * 2)
- Thr = Kmode
- 10 - Kplace3 = 1
- Kplace2 = (Kplace3 * 2) + 1
- 15 - Kplace1 = ((Kplace2) * 2) + 1
- DecrOPen = 5

20 El funcionamiento del sistema aquí ejemplificado se describirá ahora haciendo referencia a los diagramas de flujo de las figuras 2 a 8 asumiendo que un solo fonema se lleva a la entrada del módulo 40. Si se proporciona una variedad de fonemas como entrada al módulo 40, el proceso descrito a continuación se repetirá para cada fonema de entrada.

25 A continuación, el fonema que tenga la categoría consonántica de diptongo o africada se denominará fonema divisible.

Al definir las categorías modales y de lugar de un fonema, se intenta que sean unívocas, a menos que se indique lo contrario.

30 Por ejemplo, si un fonema extranjero (por ejemplo, PhonA) se califica como fricativo-uvular, significa que solo tiene una categoría modal (fricativo) y una categoría de lugar (uvular).

35 Al hacer referencia primero al diagrama de flujo de la figura 2 en la etapa 100, el índice (Indx) que escanea una tabla de la lengua de la voz hablante (designado a continuación TbB) se pone a cero, concretamente colocado como el primer fonema de la tabla.

El valor de la puntuación (Score) se pone al valor inicial cero, como es el caso de las variables TmpScrMax, FirstMaxScore, Loop y Continue. Los fonemas BestPhon, FirstBest y FirstBestCmp se ponen en el fonema cero.

40 En una etapa 104 el vector de las categorías para el fonema extranjero (PhonA) se compara con el vector del fonema para la lengua de la voz hablante (PhonB).

Si los dos vectores son idénticos, los dos fonemas serán idénticos y en la etapa 108 la puntuación se retrasa hasta el valor MaxCount y las etapas subsiguientes es la etapa 144.

45 Si los vectores son diferentes, las categorías básicas (a) se comparan en la etapa 112.

Existen tres alternativas: ambos fonemas son consonantes (128), ambos son vocales (116) o diferentes (140).

50 En el paso 116 se realiza una comprobación para saber si PhonA es un diptongo. Si así fuera, en una etapa 124 las funciones descritas en el diagrama de flujo de la figura 4 se activan, como se detalla a continuación.

Si no es un diptongo, en una etapa 120, la función descrita en el diagrama de flujo de la figura 5 se activa para comparar una vocal con otra vocal.

55 Se aprecia que las dos etapas 120 y 124 pueden llevar a una modificación de la puntuación, como se detalla a continuación.

Posteriormente, el proceso continúa hacia la etapa 144.

60 En una etapa 128 (comparación entre consonantes), se realiza una comprobación para saber si el PhonA es africado. Si así fuera, en una etapa 136 la función descrita en el diagrama de flujo de la figura 7 se activa. Si no, en una etapa 132, la función descrita en la figura 6 se activa para comparar las dos consonantes.

65 En una etapa 140 las funciones descritas en el diagrama de flujo de la figura 8 se activan, como se detalla a continuación.

De modo similar a continuación se detallan más ampliamente los criterios en los que se basa para poder modificar la puntuación en las etapas 132 y 136.

ES 2 312 851 T3

Posteriormente, el sistema pasa a la etapa 144.

Los resultados de la comparación convergen hacia la etapa 144, donde se lee el valor de la puntuación (Score).

5 En una etapa 148, el valor de la puntuación se compara con un valor denominado MaxCount. Si el valor de la puntuación iguala el MaxCount, la búsqueda finaliza, lo que significa que se ha encontrado un fonema correspondiente en la lengua de la voz hablante para el PhonA (etapa 152).

10 Si el valor de la puntuación es inferior que el MaxCount (comprobado en al etapa 148), el proceso avanza a la etapa 156, como se describe en el diagrama de flujo de la figura 3.

15 En una etapa 160, el valor Continue se compara con el valor 1. Si el resultado es positivo (concretamente Continue igual a 1), el sistema retrocede a la etapa 104 después de establecer el valor Loop al valor 1 y reajustar Continue, Indx y Score a cero. Si no, el sistema pasa a la etapa 164.

Desde aquí, si el PhonA es nasal o rótico y el fonema o fonemas seleccionados no son de ninguno de estos modos, el sistema pasa a la etapa 168, donde el fonema(s) seleccionado(s) está complementado por una forma consonántica TabB, cuyas características fonético-articulatorias permiten simular el sonido nasal o rótico del PhonA.

20 En una etapa 172, el fonema (o fonemas) seleccionados se envían hacia el módulo de asignación fonética de salida 40 para que sirvan de suministro al módulo 50.

La etapa 200 de la figura 3 se alcanza desde la etapa 156 del diagrama de flujo de la figura 2.

25 A partir de la etapa 200, el sistema pasa a la etapa 224 siempre que se cumpla una de las dos condiciones:

- El PhonA es un diptongo que se tiene que asignar a dos vocales;
- El PhonA es africado, el PhonB no es una consonante africada, pero puede ser el componente de una africada.

30 El parámetro Loop indica el número de veces que la tabla TabB se ha explorado de arriba a abajo. Su valor puede ser 0 ó 1.

35 El Loop se ajustará al valor 1 solo si el PhonA es un diptongo o una africada, ya que no es posible alcanzar la etapa 204 con un Loop igual a 1. En la etapa 204 se comprueba la Maximum Condition. Se consigue si el valor de la puntuación (Score) es mayor o igual que ello y el grupo de n rasgos fonéticos para PhonB es menor que el grupo de BestPhon.

40 Si se cumple la condición, el sistema pasa a la etapa 208, donde el MaxScore se reduce al valor de la puntuación. Y el PhonB se convierte en BestPhon.

45 En una etapa 212, Indx se compara con TbLen (el número de fonemas en TabB).

Si Indx es mayor o igual que TbLen, el sistema pasa a la etapa 284, que se describe a continuación.

50 Si Indx es menor, entonces PhonB no es el último fonema de la tabla y el sistema pasa a la etapa 220, donde Indx aumenta en 1.

Si PhonB es el último fonema en la tabla, entonces la búsqueda se termina y el BestPhon (habiendo asociado la puntuación MaxScore) es el fonema candidato para sustituir al PhonA.

55 En una etapa 224, se comprueba el valor para Loop.

Si Loop es igual a 0, entonces el sistema pasa a la etapa 228, donde se realiza una comprobación para saber si el PhonB es un diptongo o una africada.

60 Si la comprobación es positiva (por ejemplo, si el PhonB es un diptongo o africado), la siguiente etapa es la etapa 232.

En este punto, en una etapa 232 se comprueba la Maximum Condition entre Score y MaxScore.

65 Si se alcanza la condición, (por ejemplo, Score es mayor que MaxScore), en una etapa 236 el MaxScore se reduce al valor de Score y el PhonB se convierte en BestPhon.

En una etapa 240 (que se alcanza si la comprobación de la etapa 228 muestra que el PhonB no es diptongo no africado), se realiza un comprobación para saber si existe una Maximum Condition ente Score y TmpScrMAX (con

ES 2 312 851 T3

el FirstBestComp en el lugar de BestPhon). Si se consigue (por ejemplo, si Score es mayor que TmpScrMAX) en una etapa 244, TmpScrMax se retrasa mediante el Score y FirstBestComp mediante el PhonB.

5 En una etapa 248, se realiza una comprobación para saber si el PhonB es el último fonema en TabB (entonces Indx es igual a TabLen).

10 Si la comprobación es positiva (252), el valor para MaxScore se guarda como la variable FirstMaxScore, y el BestPhon se guarda como FirstBest y, posteriormente, en una etapa 256, Indx se ajusta a 0, mientras que Continue se ajusta a 1 (de modo que también el segundo componente para PhonA se buscará), y Score se ajusta a 0.

Una etapa 260 se alcanza a partir de la etapa 224 si Loop es igual a 1, concretamente si PhonB se examina como un posible segundo componente para PhonA. En una etapa 260, se realiza una comprobación para saber si se consigue la máxima condición en comparación con Score y MaxScore (que pertenece a BestPhon).

15 En una etapa 264, Score se almacena en MaxScore y PhonB en BestPhon en el caso de que se consiga la máxima condición. En una etapa 268, se realiza una comprobación para saber si PhonB es el último fonema de la tabla y, si así es, el sistema pasa a la etapa 272.

20 En la etapa 272, un fonema muy similar a PhonA se puede seleccionar entre un fonema divisible o una pareja de fonemas en la voz de la lengua del hablante si se ha cumplido la condición que FirstMaxScore es mayor o igual que (TmpScrMax + MaxScore). El valor más alto de los miembros de la relación se almacena como un MaxScore. En el caso de que la elección sea un par de fonemas, será FirstBestCmp y BestPhon. Si no, solo se considerará FirstBest.

25 Cabe destacar que BestPhon (encontrado en la segunda iteración) no puede ser diptongo o africado. En una etapa 276 Indx se aumenta en 1 y Score se ajusta a 0.

De la etapa 280, el sistema retrocede a la etapa 104.

30 La etapa 284 se alcanza desde la etapa 272 (o la etapa 2112), cuando se completa la búsqueda. En la etapa 284, se realiza una comparación entre MaxScore y una constante umbral Thr. Si MaxScore es mayor, entonces el fonema candidato (o el par de fonemas) es el sustituto para PhonA. Si no, PhonA se asigna al fonema cero.

El diagrama de flujo de la figura 4 es una descripción detallada del bloque 124 del diagrama de la figura 2.

35 Una etapa 300 se alcanza si PhonA es un diptongo.

40 En una etapa 302, se realiza una comprobación para saber si PhonB es un diptongo y Loop es igual a 0. Si se cumple, el sistema pasa a la etapa 302 donde, tras comprobar los rasgos para PhonA, el sistema pasa a la etapa 306 si PhonA es un diptongo que se tiene que asignar a una sola vocal.

Los diptongos de este tipo tienen un primer componente que es media y central y el segundo componente es cerrado-semicerrado y posterior.

45 De la etapa 306, el sistema pasa a la etapa 144. [0117] En una etapa 308, se requiere la función de comparar dos diptongos.

En una etapa 310, las categorías (b) de los dos fonemas se comparan mediante dicha función y Score se aumenta en 1 por cada rasgo común encontrado.

50 En una etapa 312, los primeros componentes de los dos diptongos se comparan y en una etapa 314 una función denominada F_CasiSpec_Voc se requiere para los dos componentes.

Esta función realiza tres comprobaciones que se cumplen si:

55 - los componentes de los dos diptongos son indistintamente vocales abiertas, o vocales abierta-medioabierta, anterior y no redondeada;

60 - el componente de PhonA es abierto y central, y en TabB no existen fonemas que tengan ambas categorías, y PhonB es semicerrado y anterior;

- el componente de PhonA es cerrado, anterior y no redondeado, o cerrado-semicerrado, anterior y redondeado, y en TabB ningún fonema tiene dichos rasgos mientras que PhonB es cerrado posterior y redondeado o cerrado-semicerrado, posterior y redondeado.

65 Si se cumple alguna de las tres condiciones, en una etapa 316 el valor de Score se retrasa añadiendo (KOpen * 2) al mismo.

Si no, en una etapa 318, se requiere una función F_ValPlace_Voc para los dos componentes.

ES 2 312 851 T3

Dicha función compara las categorías anterior, central y posterior (categoría (d)).

Si son idénticas, Score se aumenta en Kopen; si son diferentes, se añade un valor a Score que está compuesto de Kopen menos la constante DecrOpen si la distancia entre las dos categorías es 1, mientras que Score no se aumenta si la distancia es 2.

Una distancia igual a uno existe entre central y anterior y entre central y posterior, mientras que una distancia igual a dos existe entre anterior y posterior.

10 En una etapa 320 se requiere una función F_ValOpen_Voc para compara los dos componentes del diptongo. Específicamente, F_ValOpen_Voc funciona de manera cíclica comparando los primeros y los segundos componentes en dos iteraciones posteriores.

15 La función compara las categorías (e) y añade a Score la constante Kopen menos el valor de la distancia entre las categorías incluida posteriormente en la tabla 1.

La matriz es simétrica, mediante la cual solo se registró la porción superior.

20 Al realizar un ejemplo numérico, si PhonA es una vocal cerrada y PhonB es una vocal mediocerrada, un valor igual a $(KOpen - (6 * Lstep))$ se añadirá a Score que, considerando el valor de las constantes, es igual a 8.

En una etapa 322, si ambos componentes tienen el rasgo redondeo, la constante $(KOpen + 1)$ se añade a Score. De forma inversa, si solo uno de los dos es redondeado, entonces Score se reduce por Kopen.

25 De la etapa 324, el sistema retrocede a la etapa 314 si los dos primeros componentes se han comparado; de forma inversa, una etapa 326 se alcanza cuando también los segundos componentes se han comparado.

En la etapa 326, la comparación de los dos diptongos se termina y el sistema retrocede a la etapa 144.

30 En una etapa 328, se realiza una comprobación para saber si PhonB es un diptongo y Loop es igual a 1. Si así ocurre, el sistema pasa a la etapa 306.

En una etapa 330, se realiza una comprobación para saber si PhonA es un diptongo que se tiene que asociar a una sola vocal. Si así ocurre, en una etapa 331 Loop se comprueba y, si es igual a 1, se alcanza la etapa 306.

35 En una etapa 332, se crea un fonema TmpPhonA.

TmpPhonA es una vocal sin características de diptongo y que tiene rasgos de mediocerrada, posterior y redondeada.

40 Posteriormente, el sistema pasa a una etapa 334, donde TmpPhonA y PhonB se comparan. La comparación se efectúa forzando la comparación entre dos fonemas vocálicos sin categoría de diptongo.

Dicha función, presente también en la etapa 120 del diagrama de flujo de la figura 2, se describe detalladamente en la figura 5.

45 En una etapa 336, la función se reclama para realizar una comparación entre un componente de PhonA y PhonB: como consecuencia, en una etapa 338, si Loop es igual a 9, el primer componente de PhonA se compara con PhoB (en una etapa 344). De forma inversa, si Loop es igual a 1, el segundo componente de PhonA se compara con PhonB (en una etapa 340).

50 En la etapa 340, se hace referencia a las categorías nasal y rótica, aumentando Score por cada identidad encontrada.

En una etapa 342, si PhonA tiene un acento en su primer componente y PhonB es una vocal acentuada, Score aumenta en 2. En todos los otros casos se reduce en 2.

55 En una etapa 344, si PhonA tiene acento en su segundo componente y PhonB es una vocal acentuada, entonces Score aumenta en 2; de lo contrario, se reduce en 2 en el resto de los casos.

60 En 348, las categorías (d) y (e) del primer o segundo componente de PhonA (dependiendo de si Loop es igual a 0 o 1, respectivamente) se comparan con PhonB.

La comparación de los vectores de rasgos y la actualización de Score está realizada basándose en los mismos principios ya descritos en las etapas de 314 a 322.

65 Una etapa 350 marca el retroceso a la etapa 144.

El diagrama de flujo de la figura 5 describe detalladamente la etapa 120 del diagrama de la figura 2, concretamente la comparación entre dos vocales que no son diptongos.

ES 2 312 851 T3

En una etapa 400 se realiza una comparación para saber si PhonB es un diptongo. Si así es, el sistema pasa directamente a la etapa 470.

5 En una etapa 410, se realiza una comparación sobre la base de las categorías (b) aumentado Score en 1 por cada categoría idéntica encontrada.

De forma inversa, en una etapa 420, la función F_CasiSpec_Voc ya descrita anteriormente se reclama para comprobar si se alcanza una de las condiciones de la función.

10 Si así ocurre, Score se aumenta en la cantidad ($K_{Open} * 2$) en una etapa 430.

En el caso de un resultado negativo, en una etapa 440 se reclama la función F_ValPlace_Voc.

15 Posteriormente, en una etapa 450, se reclama la función F_ValOpen_Voc.

En una etapa 460, si ambas vocales tienen la categoría redondeada, Score aumenta en la constante ($K_{Open} + 1$); si, por el contrario, solo se encuentra un fonema con la categoría redondeada, entonces Score aumenta en K_{Open} .

20 Una etapa 470 marca el final de la comparación, tras la cual el sistema retrocede a la etapa 144.

El diagrama de flujo de la figura 6 describe detalladamente el bloque 132 del diagrama de la figura 1.

25 En una etapa 500, se comparan las dos consonantes, mientras que la variable TmpKP se ajusta a 0 y la función F_CasiSpec_Cons se reclama en una etapa 504.

La función en cuestión comprueba si se cumple alguna de las condiciones siguientes:

- 30 1.0 PhonA uvular fricativo y en TabB no hay fonemas con estas características y PhonB es vibrante simple -alveolar;
- 35 1.1 PhonA uvular fricativo y en TabB no hay fonemas con estas características y PhonB es aproximante-alveolar;
- 1.2 PhonA uvular fricativo y en TabB no hay fonemas con estas características y PhonB es uvular-vibrante simple;
- 40 1.3 PhonA uvular fricativo y en TabB no hay fonemas con estas características y o con las de PhonB de 1.0 o 1.1 o 1.2, y PhonB es lateral-alveolar;
- 45 2.0 PhonA glotal fricativo y en TabB no hay fonemas con estas características y PhonB es fricativo-velar;
- 3.0 PhonA fricativo-velar y en TabB no hay fonemas con estas características y PhonB es fricativo-glotal o oclusivo-velar;
- 50 4.0 PhonA vibrante simple-alveolar y en TabB no hay fonemas con estas características y PhonB es fricativo-uvular;
- 4.1 PhonA vibrante simple-alveolar y en TabB no hay fonemas con estas características y PhonB es aproximante-alveolar;
- 55 4.2 PhonA vibrante simple-alveolar y en TabB no hay fonemas con estas características o con las de PhonB de 4.0 y 4.1, y PhonB es lateral-alveolar;
- 60 5.0 PhonA nasal-velar y en TabB no hay fonemas con estas características y PhonB es nasal-alveolar;
- 5.1 PhonA nasal-velar y en TabB no hay fonemas con estas características o con las de PhonB de 5.0 y PhonB es nasal-bilabial;
- 65 6.0 PhonA es fricativo-dental-sordo y en TabB no hay fonemas con estas características y PhonB es aproximante-dental;
- 6.1 PhonA es fricativo-dental-sordo y en TabB no hay fonemas con estas características o con las de PhonB de 6.0, y PhonB es oclusivo-dental;
- 6.2 PhonA es fricativo-dental-sordo y en TabB no hay fonemas con estas características o con las de PhonB de 6.0 y PhonB es oclusivo-alveolar;

ES 2 312 851 T3

- 7.0 PhonA es fricativo-dental-sonoro y en TabB no hay fonemas con estas características y PhonB es aproximante-dental;
- 5 7.1 PhonA es fricativo-dental-sonoro y en TabB no hay fonemas con estas características o con las de PhonB de 7.0 y PhonB es oclusivo-dental;
- 7.2 PhonA es fricativo-dental-sonoro y en TabB no hay fonemas con estas características o las de PhonB de 7.0 y PhonB es oclusivo-alveolar;
- 10 8.0 PhonA es fricativo-palatal-alveolar-sordo y en TabB no hay fonemas con estas características y PhonB es fricativo-postalveolar;
- 8.1 PhonA es fricativo-palatal-alveolar-sordo y en TabB no hay fonemas con estas características o las de PhonB de 8.0 y PhonB es fricativo-palatal;
- 15 9.0 PhonA es fricativo-postalveolar y en TabB no hay fonemas con estas características o fricativas-retroflejas y PhonB es fricativo-alveolar-palatal;
- 20 10.0 PhonA es fricativo-postalveolar-velar y en TabB no hay fonemas con estas características y PhonB es fricativo-alveolar-palatal;
- 10.1 PhonA es fricativo-postalveolar-velar y en TabB no hay fonemas con estas características y PhonB es fricativo-palatal;
- 25 10.2 PhonA es fricativo-postalveolar-velar y en TabB no hay fonemas con estas características o las de 10.0 o 10.1 y PhonB es fricativo-postalveolar;
- 11.0 PhonA es oclusivo-palatal y en TabB no hay fonemas con estas características y PhonB es lateral-palatal;
- 30 11.1 PhonA es oclusivo-palatal y en TabB no hay fonemas con estas características o las de PhonB de 11.0 y PhonB es fricativo-palatal o aproximante-palatal;
- 12.0 PhonA es fricativo-bilabial-dental-sonoro y en TabB no hay fonemas con estas características y PhonB es aproximante-bilabial-sonoro;
- 35 13.0 PhonA es fricativo-palatal-sonoro y en TabB no hay fonemas con estas características y PhonB es oclusivo-palatal-sonoro o aproximante-palatal-sonoro;
- 40 14.0 PhonA es lateral-palatal y en TabB no hay fonemas con estas características y PhonB es oclusivo-palatal;
- 14.1 PhonA es lateral-palatal y en TabB no hay fonemas con estas características o las de PhonB de 14.0 y PhonB es fricativo-palatal o aproximante-palatal;
- 45 15.0 PhonA es aproximante-dental y en TabB no hay fonemas con estas características y PhonB es oclusivo-dental o oclusivo-alveolar;
- 16.0 PhonA es aproximante-bilabial y en TabB no hay fonemas con estas características y PhonB es oclusivo-bilabial;
- 50 17.0 PhonA es aproximante-velar y en TabB no hay fonemas con estas características y PhonB es oclusivo-velar;
- 18.0 PhonA es aproximante-alveolar y en TabB no hay fonemas con estas características y PhonB es vibrante simple-alveolar o fricativo-uvular o vibrante simple-uvular;
- 55 18.1 PhonA es aproximante-alveolar y en TabB no hay fonemas con estas características o las de PhonB en 18.0 y PhonB es lateral-alveolar.

60 Si cualquiera de estas condiciones se cumple, el sistema pasa a la etapa 508, donde TmpPhonB se substituye por PhonB durante todo el proceso de comparación hasta una etapa 552.

Si no se cumple ninguna de las condiciones anteriores, el sistema pasa directamente a la etapa 512, donde se comparan las categorías (f) modales.

65 Si PhonA y PhonB tienen la misma categoría, entonces Score aumenta en KMode.

ES 2 312 851 T3

En una etapa 516, una función F_CompPen_Consse reclama para controlar si se cumple la siguiente condición:

-PhonA es fricativo-postalveolar y PhonB (o TmpPhonB) es fricativo-postalveolar-velar.

5 Y si se cumple la condición, entonces Score se reduce en KPlace1.

En una etapa 520, una función F_ValPlace_Cons se reclama para incrementar basándose en lo que se registra en la tabla 2.

10 En la tabla en cuestión, las categorías para PhonA están en el eje vertical y las de PhonB en el eje horizontal. Cada celda incluye un valor de bonificación para añadirlo a Score.

15 Al considerar, como por ejemplo, que PhonA tiene la categoría labiodental y PhonB la categoría dental solo, entonces, al examinar la línea de labiodental y cruzar la columna para dental, se encuentra que el valor Kplace2 se tiene que añadir a Score.

20 En una etapa 524, se realiza una comprobación para saber si PhonA es aproximante-semivocal y PhonB (o TmpPhonB) es aproximante. Si la comprobación da un resultado positivo, el sistema pasa a la etapa 528, donde se realiza una prueba sobre TmpKP.

Dicha prueba se realiza para asegurar que, en el caso de que los dos fonemas que se estén comparando sean aproximantes y con categorías de lugar idénticas, su Score es mayor que en el caso de una comparación consonante-vocal.

25 Si dicha variable es mayor o igual a KPlace1, entonces en una etapa 532 TmpKP se aumenta en KMode. En caso negativo, TmpKP se ajusta a cero en una etapa 536.

En una etapa 540, la cantidad TmpKP se añade a Score.

30 En una etapa 544, se realiza una comprobación para saber si Score es mayor que KMode.

Si así ocurre, en una etapa 548 las categorías (h) se comparan con la excepción de la categoría semiconsonante. Por cada identidad encontrada, Score aumenta en 1.

35 Una etapa 552 marca el final de la comparación, tras la cuál el sistema retrocede a la etapa 144 de la figura 1.

El diagrama de flujo de la figura 7 hace referencia a la comparación entre fonemas cuando PhonA es una consonante africada (etapa 136 de la figura 2).

40 En una etapa 600, comienza la comparación y en una etapa 604 se realiza una comprobación para saber si PhonB es africado y Loop igual a 0.

Si así ocurre, el sistema pasa a la etapa 608, que causa a su vez que el sistema retroceda a la etapa 132.

45 En una etapa 612, se realiza una comprobación para saber si PhonB es africado y Loop igual a 1.

Si así ocurre, se alcanza directamente una etapa 600.

50 En una etapa 616, se realiza una comprobación para saber si se puede considerar PhonB está compuesto por una africada.

Esto no puede ocurrir si Loop es igual a 1 y PhonB tiene las categorías fricativo-postalveolar-velar.

55 Si así ocurre, el sistema pasa a la etapa 660.

En una etapa 620, se realiza una comprobación del valor de Loop: si es igual 0, el sistema pasa a la etapa 642.

60 En esa etapa, PhonA se sustituye temporalmente en la comparación con PhonB y TmpPhonA; tiene las mismas características que PhonA, excepto por el hecho de que en lugar de ser africado es oclusivo.

En una etapa 628, se realiza una comprobación para saber si TmpPhonA tiene las categorías labiodentales; si así ocurre en una etapa 636, las categorías dentales se eliminan del vector de categorías.

65 En una etapa 632, se realiza una comprobación para saber si TmpPhonA tiene la categoría postalveolar; Si así ocurre, dicha categoría se reemplaza por la categoría alveolar en una etapa 644.

En una etapa 640, se realiza una comprobación para saber si TmpPhonA tiene las categorías alveolar-palatal; si así ocurre, se elimina la categoría palatal.

ES 2 312 851 T3

En una etapa 652 PhonA se reemplaza temporalmente (hasta alcanzar la etapa 144) en comparación con PhonB por TmpPhonA; tiene las mismas características que PhonA, excepto por el hecho de que es fricativo en lugar de ser africado.

5 La etapa 656 marca la evolución hacia la comparación de la etapa 132 al comparar TmpPhonA con PhonB.

Una etapa 660 marca el retroceso a la etapa 144.

El diagrama de flujo de la figura 8 describe detalladamente la etapa 140 del diagrama de flujo de la figura 2.

10 Una etapa 700 se alcanza se alcanza si PhonA es una consonante y PhonB es una vocal o si PhonA es vocal y PhonB consonante. El fonema TmpPhonA se establece como el fonema cero.

15 En una etapa 705, se realiza una comprobación para saber si PhonA es vocal y PhonB es consonante. Si así ocurre, la próxima etapa es la etapa 780.

En una etapa 710, se realiza una comprobación para saber si PhonA es aproximante-semiconsonante.

Si no es así, el sistema pasa directamente a la etapa 780.

20 En una etapa 720, se realiza una comprobación para saber si PhonA es palatal. Si así ocurre, en una etapa 730 TmpPhonA se transforma en una vocal cerrada anterior no acentuada y la comparación de una etapa 120 se realiza entre TmpPhonA y PhonB.

25 En una etapa 740, se realiza una comprobación para saber si PhonA es bilabial-velar. Si así ocurre, en una etapa 750 TmpPhonA se transforma en un vocal redondeada-posterior-cerrada-no acentuada y la comparación de la etapa 120 (figura 2) se realiza entre TmpPhonA y PhonB.

30 En una etapa 760, se realiza una comprobación para saber si PhonA es bilabial-palatal. Si así ocurre, en una etapa 770 TmpPhonA se transforma en un vocal redondeada-cerrada-no acentuada y la comparación de la etapa 120 se lleva a cabo entre TmpPhonA y PhonB.

Una etapa 780 marca el retroceso del sistema a la etapa 144.

35 A continuación, se muestran las tablas 1 y 2, mencionadas en diversas ocasiones anteriormente.

40 Por supuesto, sin prejuicios contra los principios subyacentes de la invención, las discrepancias y las realizaciones pueden variar, significativamente, con respecto a lo que se ha descrito, solo por medio de ejemplos, sin apartarse del ámbito de la invención como se define en las reivindicaciones adjuntas.

(Tabla pasa a página siguiente)

45

50

55

60

65

ES 2 312 851 T3

TABLA 1

Distancias de los rasgos vocálicos (e)

5		CERRADA	CERRADA	MEDIOCERRADA	MEDIA	MEDIOABIERTA	ABIERTA-	ABIERTA
10			MEDIOCERRA				MEDIOABIERTA	
15	CERRADA	0	2*LStep	6*LStep	7*Lstep	8*Lstep	12*Lstep	14*LStep
20	CERRADA-		0	4*LStep	5*Lstep	6*Lstep	10*Lstep	12*Lstep
25	MEDIOCERRADA			0	1*Lstep	2*Lstep	6*Lstep	8*Lstep
30	MEDIOCERRADA				0	1*Lstep	5*Lstep	7*Lstep
35	MEDIA					0	4*Lstep	6*Lstep
40	MEDIOABIERTA						0	2*Lstep
45	ABIERTA							0
50	MEDIOABIERTA							
55	ABIERTA							0
60								
65								

Tabla 2: Valores a añadir a Score

	BILABIAL	LABIODENTAL	DENTAL	ALVEOLAR	POST ALVEOLAR	RETRO-FLEJA	PALATAL	VELAR	UVULAR	FARÍNGEA	GLOTAL
BILABIAL	+KPlace1	+KPlace2	+0	+0	+0	+0	+0	+0	+0	+0	+0
LABIODENTAL	+KPlace2	+KPlace1	+KPlace2	+0	+0	+0	+0	+0	+0	+0	+0
DENTAL	+0	+0	+KPlace1	+KPlace2	+0	+0	+0	+0	+0	+0	+0
ALVEOLAR	+0	+0	+KPlace3	+KPlace	+KPlace2	+KPlace3	+0	+0	+0	+0	+0
POST ALVEOLAR	+0	+0	+0	+KPlace3	+KPlace1	+KPlace2	+0	+0	+0	+0	+0
RETRO FLEJA	+0	+0	+0	+KPlace3	+KPlace3	+KPlace	+KPlace2	+0	+0	+0	+0
PALATAL	+0	+0	+0	+0	+KPlace3	+KPlace2	+KPlace1	+KPlace2	+0	+0	+0
VELAR	+0	+0	+0	+0	+0	+0	+0	+KPlace	+0	+0	+0
UVULAR	+0	+0	+0	+KPlace2	+0	+0	+0	+KPlace2	+KPlace1	+0	+0
FARÍNGEA	+0	+0	+0	+0	+0	+0	+0	+0	+0	+KPlace1	+0
GLOTAL	+0	+0	+0	+0	+0	+0	+0	+0	+0	+0	+KPlace1

Referencias citadas en la descripción

Esta lista de referencias citadas por el solicitante está prevista únicamente para ayudar al lector y no forma parte del documento de patente europea. Aunque se ha puesto el máximo cuidado en su realización, no se pueden excluir errores u omisiones y la OEP declina cualquier responsabilidad en este respecto.

Documentos que no son patentes citados en la descripción

• C. **TRABER** *et al.* From multilingual to polyglot speech synthesis. *Proceedings of the Eurospeech, 1999*, 835-838 [0008]

• W.N. **CAMPBELL**. Foreign-language speech synthesis. Proceedings ESCA/COCSDA ETRW on Speech Synthesis, 1998 [0010]

• Talking Foreign. Concatenative Speech Synthesis and Language Barrier. Proceedings of the Eurospeech Scandinavia, 2001, 337-340 [0010].

20

25

30

35

40

45

50

55

60

65

REIVINDICACIONES

5 1. Un procedimiento para la conversión texto a voz de un texto (T1,..., Tn) en una primera lengua incluyendo secciones en al menos una segunda lengua, **caracterizado** por el hecho de que incluye las etapas de:

- convertir (30) dichas secciones de dicha segunda lengua en conjuntos de fonemas de dicha segunda lengua,
- asignar (40; 40b) al menos parte de dichos fonemas de dicha segunda lengua a conjuntos de fonemas de
10 dicha primera lengua,
- incluyendo dichos conjuntos de fonemas de dicha primera lengua resultante de dicha asignación en el flujo de fonemas, y
- generar (50) una señal de voz a partir de dicho flujo de fonemas resultante.

15 donde dicha etapa de asignación (40) incluye las operaciones de:

- llevar a cabo pruebas de similitud entre cada dicho fonema de dicha segunda lengua estando asignada y
20 un conjunto de fonemas de asignación candidatos de dicha primera lengua, estando dichas pruebas de similitud realizadas representando dichos fonemas de dicha segunda lengua y dichos fonemas de asignación candidatos de dicha primera lengua como vectores de categorías, donde un vector representativo de categorías fonéticas de cada dicha segunda lengua está sujeto a la comparación con un conjunto de vectores de categoría representativos de la bases fonética categoría a categoría,
- asignar valores de puntuación respectivos a dichas comparaciones categoría a categoría, estando dichos
25 valores de puntuación respectivos agregados para generar puntuaciones respectivas a los resultados de dichas pruebas, y
- asignar (40b) cada dicho fonema de dicha segunda lengua a un conjunto de fonemas de asignación de dicha
30 primera lengua seleccionada de dichos fonemas de asignación candidatos como una función de dichas puntuaciones.

35 2. El procedimiento de la reivindicación 1, **caracterizado** por el hecho de que incluye la etapa de asignar (40) dicho fonema de dicha segunda lengua a un conjunto de fonemas de asignación de dicha primera lengua seleccionada de:

- un conjunto de fonemas de dicha primera lengua incluyendo tres, dos o un fonema de dicha primera lengua,
o
- un conjunto vacío, donde no se incluye ningún fonema en dicho flujo resultante para dicho fonema en dicha
40 segunda lengua.

45 3. El procedimiento de la reivindicación 2, **caracterizado** por el hecho de que dicha etapa de asignación (40) incluye las operaciones de:

- definir un valor umbral (th) para los resultados de dichas pruebas, y
- asignar a dicho conjunto vacío de fonemas de dicha primera lengua cualquier fonema de dicha segunda
50 lengua en el que cualquiera de dichas puntuaciones no consiga alcanzar dicho valor umbral.

4. El procedimiento de la reivindicación 1, **caracterizado** por el hecho de que incluye las etapas de asignar distintos pesos a dichos valores de puntuación al agregar dichos valores de puntuación respectivos para generar dichas puntuaciones.

55 5. El procedimiento de la reivindicación 1, **caracterizado** por el hecho de que incluye la operación de seleccionar dichas categorías fonéticas de entre el grupo que consiste en:

- (a) dos categorías básicas vocal y consonante;
- (b) la categoría diptongo;
- (c) las características vocálicas (por ejemplo, vocal) acentuada/no acentuada, no silábica, larga, nasal, rótica, redondeada
- (d) las categorías vocálicas anterior, central, posterior;

ES 2 312 851 T3

- (e) las categorías vocálicas cerrada, cerrada-mediocerrada, mediocerrada, mediocerrada, media, medioabierta, abierta-medioabierta, abierta;
- (f) las categorías del modo consonántico oclusivo, nasal, vibrante múltiple, vibrante simple, fricativo, fricativa lateral, aproximante, lateral, africada;
- (g) las categorías del lugar consonántico bilabial, labiodental, dental, alveolar, postalveolar, retrofleja, palatal, velar, uvular, faríngea, glotal; y
- (h) las otras categorías consonánticas sonora, larga, silábica, aspirada, oclusiva, sorda, semiconsonante.

6. El procedimiento de la reivindicación 1, **caracterizado** por el hecho de que incluye la etapa de pronunciar (50, 60) dicho flujo de fonemas resultante por medio de una voz hablante de dicha primera lengua.

7. Un sistema para la conversión texto a voz de un texto (T1,..., Tn) en una primera lengua incluyendo secciones en al menos una segunda lengua, **caracterizado** por el hecho de que incluye:

- un transcriptor grafema/morfema (30) para convertir dichas secciones en dicha segunda lengua en fonemas de dicha segunda lengua,
- un módulo de asignación (40, 40b) para asignar al menos parte de dichos fonemas de dicha segunda lengua a conjuntos de fonemas de dicha segunda lengua,
- un módulo de síntesis de voz (50) adaptado para ser suministrado con un flujo de fonemas resultante incluyendo dichos conjuntos de fonemas de dicha primera lengua resultante de dicha asignación y el flujo de fonemas de dicha primera lengua representativa de dicho texto, y para generar una señal de voz de dicho flujo resultante de fonemas,

donde dicho módulo de asignación (40) está configurado para:

- llevar a cabo pruebas de similitud entre cada dicho fonema de dicha segunda lengua estando asignada y un conjunto de fonemas de asignación candidatos de dicha primera lengua, estando dichas pruebas de similitud realizadas representando dichos fonemas de dicha segunda lengua y dichos fonemas de asignación candidatos de dicha primera lengua como vectores de categoría fonéticos, donde un vector representativo de cada dicho fonema de dicha segunda lengua está sujeto a comparaciones con un conjunto de vectores de categoría fonéticos representativos de las categorías fonéticas de dichos fonemas de asignación candidatos en dicha primera lengua, estando dicha comparación realizada sobre una base categoría a categoría,
- asignar valores de puntuación respectivos a dichas comparaciones categoría a categoría, estando dichos valores de puntuación agregados para generar puntuaciones respectivas a los resultados de dichas pruebas, y
- asignar (40b) cada dicho fonema de dicha segunda lengua a un grupo de fonemas de asignación de dicha primera lengua seleccionada de dichos fonemas de asignación candidatos como función de dichas puntuaciones.

8. El sistema de la reivindicación 7, **caracterizado** por el hecho de que dicho modulo de asignación (40) está configurado para asignar (40b) dicho fonema de dicha segunda lengua a un grupo de fonemas de asignación de dicha primera lengua seleccionada de entre:

- un conjunto de fonemas de dicha primera lengua incluyendo tres, dos o un fonema de dicha primera lengua, o
- un conjunto vacío, donde ningún fonema está incluido en dicho flujo resultante para dicho fonema en dicha segunda lengua.

9. El sistema de la reivindicación 8, **caracterizado** por el hecho de que dicho modulo de asignación (40) está configurado para:

- definir un valor umbral (Th) para los resultados de dichas pruebas, y
- asignar a dicho conjunto vacío de fonemas de dicha primera lengua cualquier fonema de dicha segunda lengua en el que cualquiera de dichas puntuaciones no alcanza dicho valor umbral.

10. El sistema de la reivindicación 7, **caracterizado** por el hecho de que dicho modulo de asignación (40) está configurado para asignar ponderaciones diferenciadas a dichos valores de puntuación al agregar dichos valores de puntuación respectivos para generar dichas puntuaciones.

ES 2 312 851 T3

11. El sistema de la reivindicación 7, **caracterizado** por el hecho de que dicho modulo de asignación (40) está configurado para operar basándose en categorías fonéticas fuera del grupo que consiste en:

- las dos categorías básicas vocal y consonante;
- la categoría diptongo;
- las características vocálicas (por ejemplo, vocal) acentuada/no acentuada, no silábica, larga, nasal, rótica, redondeada;
- las categorías vocálicas anterior, central, posterior;
- las categorías vocálicas cerrada, cerrada-mediocerrada, mediocerrada, mediocerrada, media, medioabierta, abierta-medioabierta, abierta;
- las categorías del modo consonántico oclusiva, nasal, vibrante múltiple, vibrante simple, fricativa, fricativa lateral, aproximante, lateral, africada;
- las categorías del lugar consonántico bilabial, labiodental, dental, alveolar, postalveolar, retrofleja, palatal, velar, uvular, faríngea, glotal; y
- las otras categorías consonánticas sonora, larga, silábica, aspirada, oclusiva, sorda, semiconsonante.

12. El sistema de la reivindicación 7, **caracterizado** por el hecho de que dicho módulo de síntesis de voz (50) está configurado para pronunciar (50, 60) dicho flujo resultante de fonemas por medio de una voz hablante de dicha primera lengua.

13. Un programa informático que se puede cargar en la memoria de al menos un ordenador y que incluye porciones de software que realizan las etapas del procedimiento de cualquiera de las reivindicaciones 1 a 6.

FIG. 1

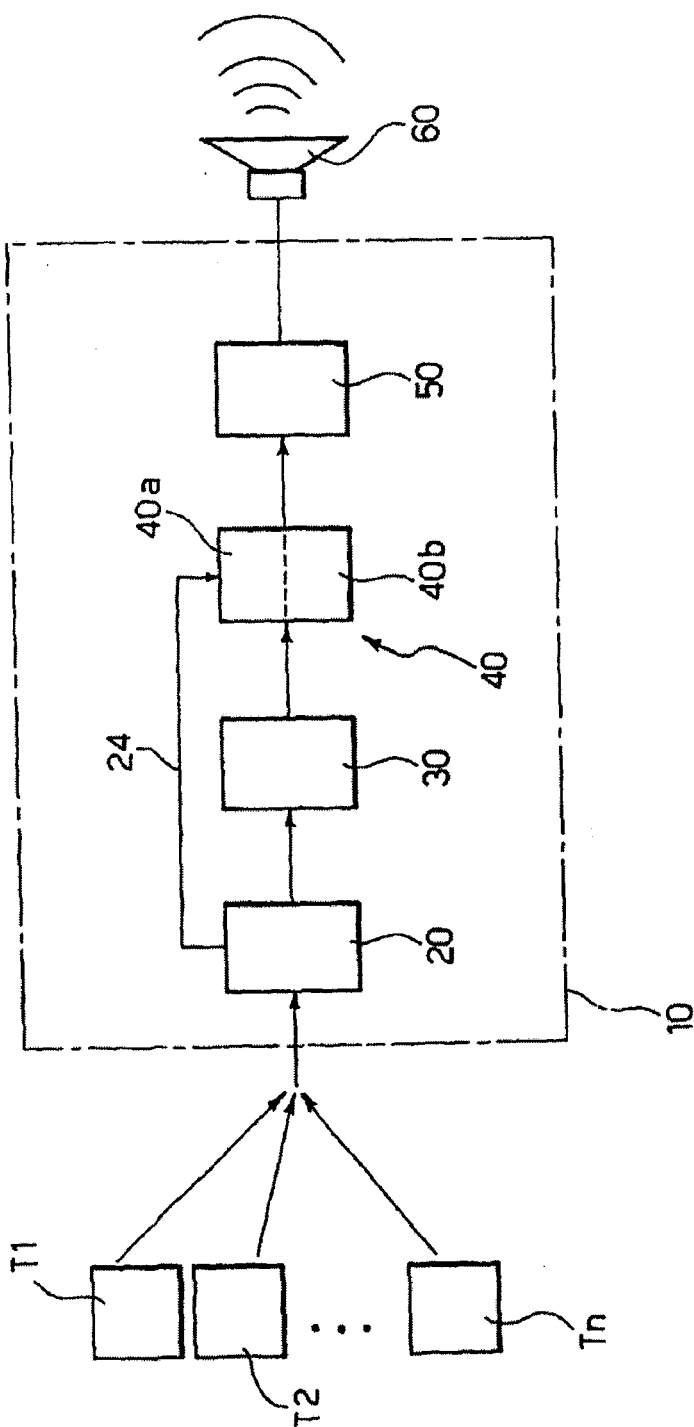


Fig. 2

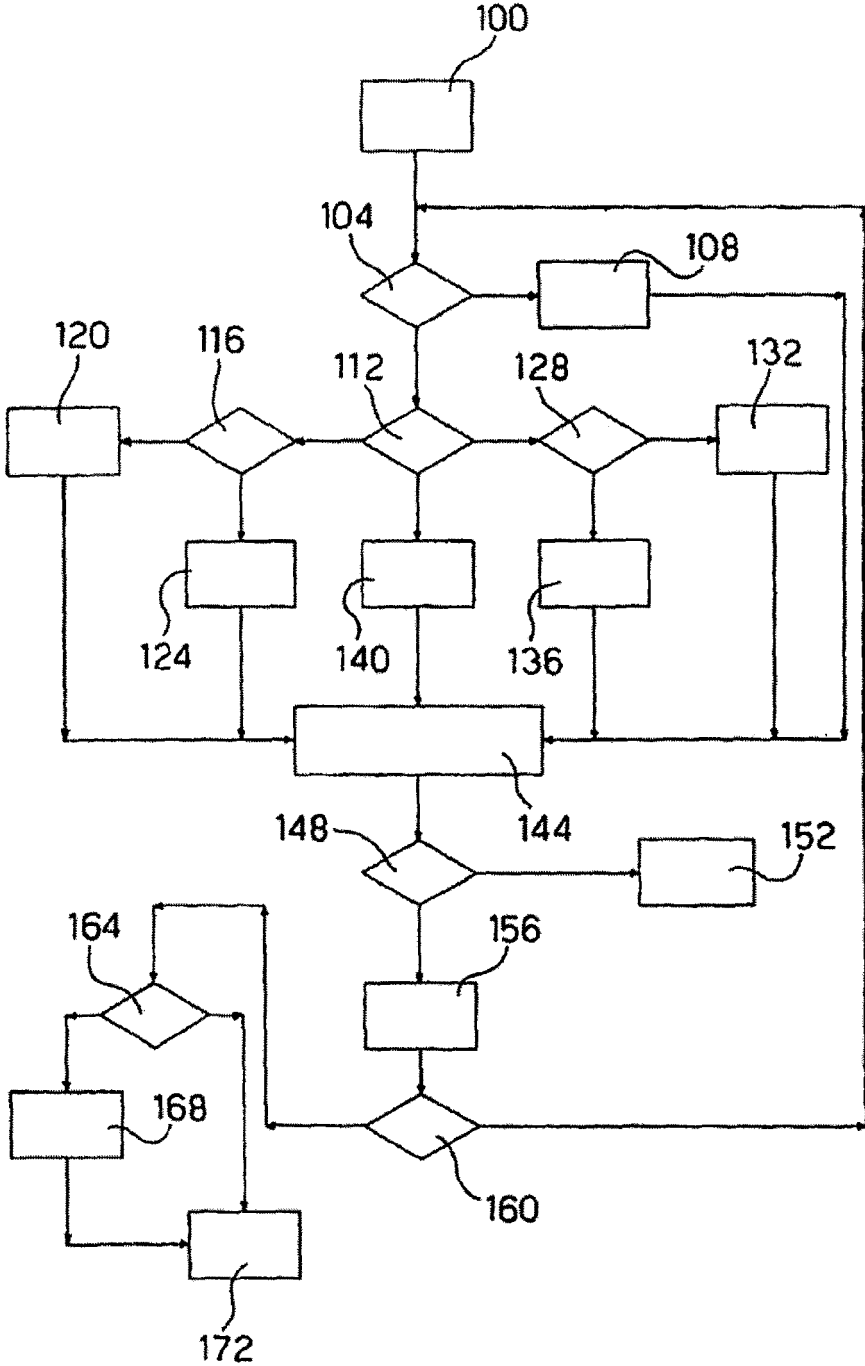


Fig-3

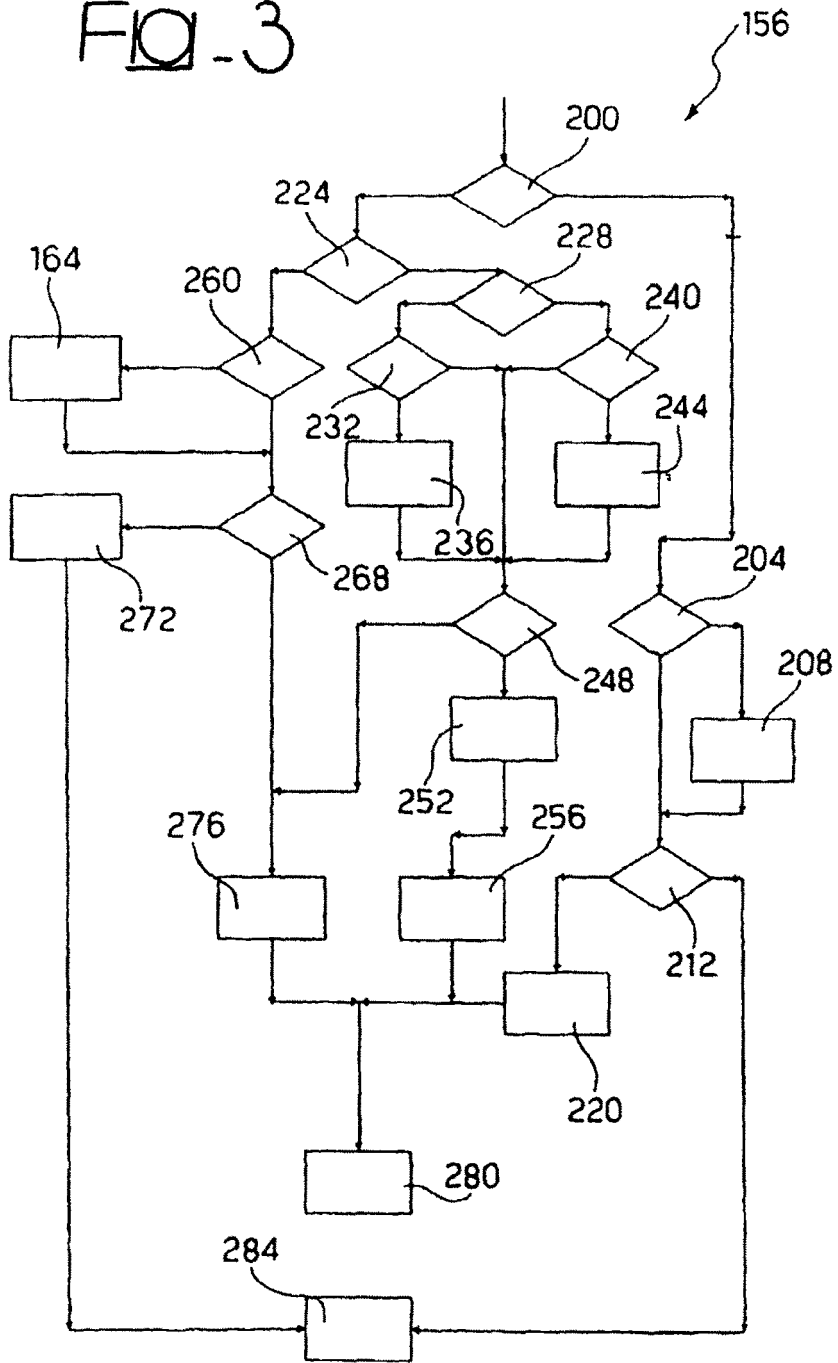


Fig. 4

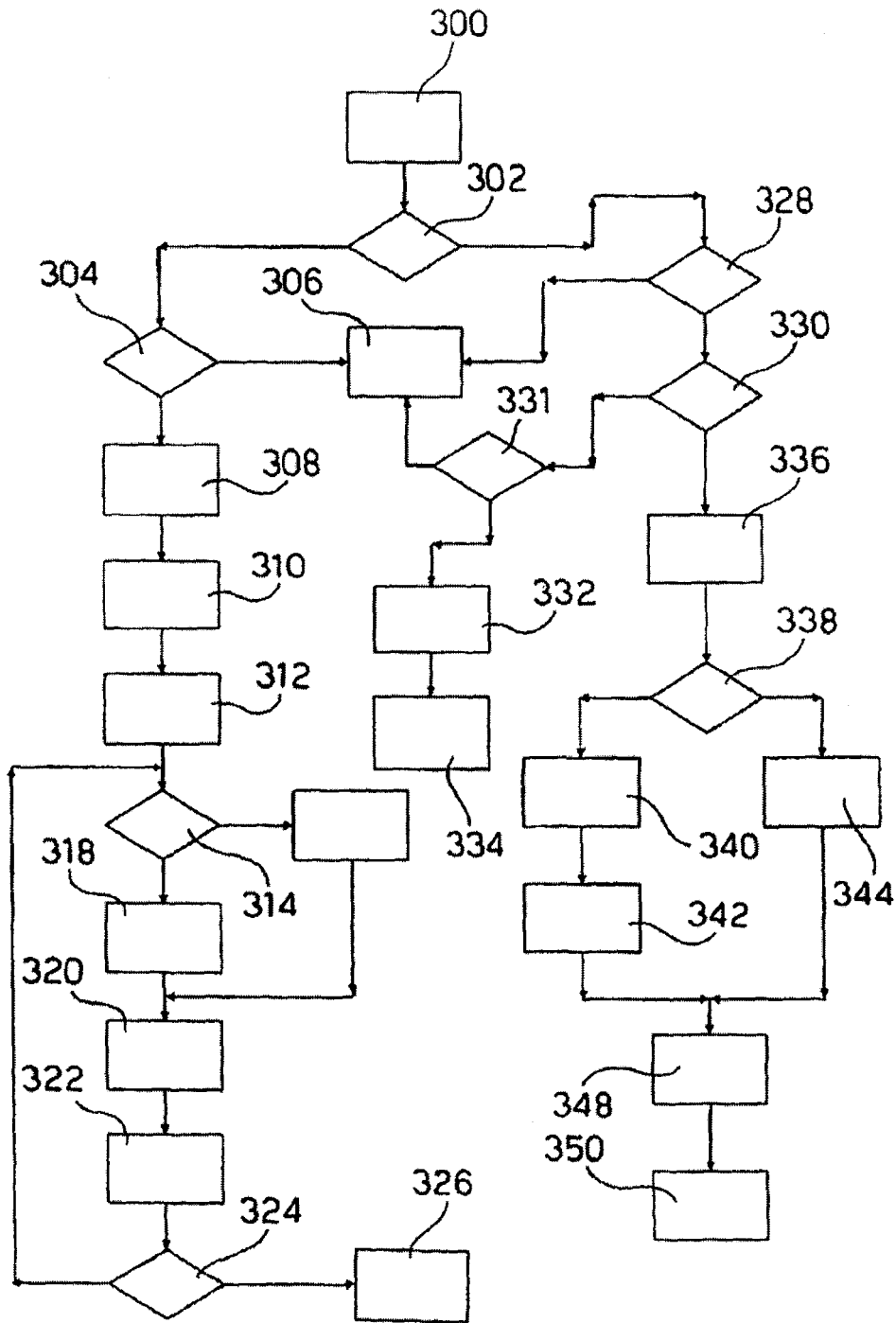


Fig. 5

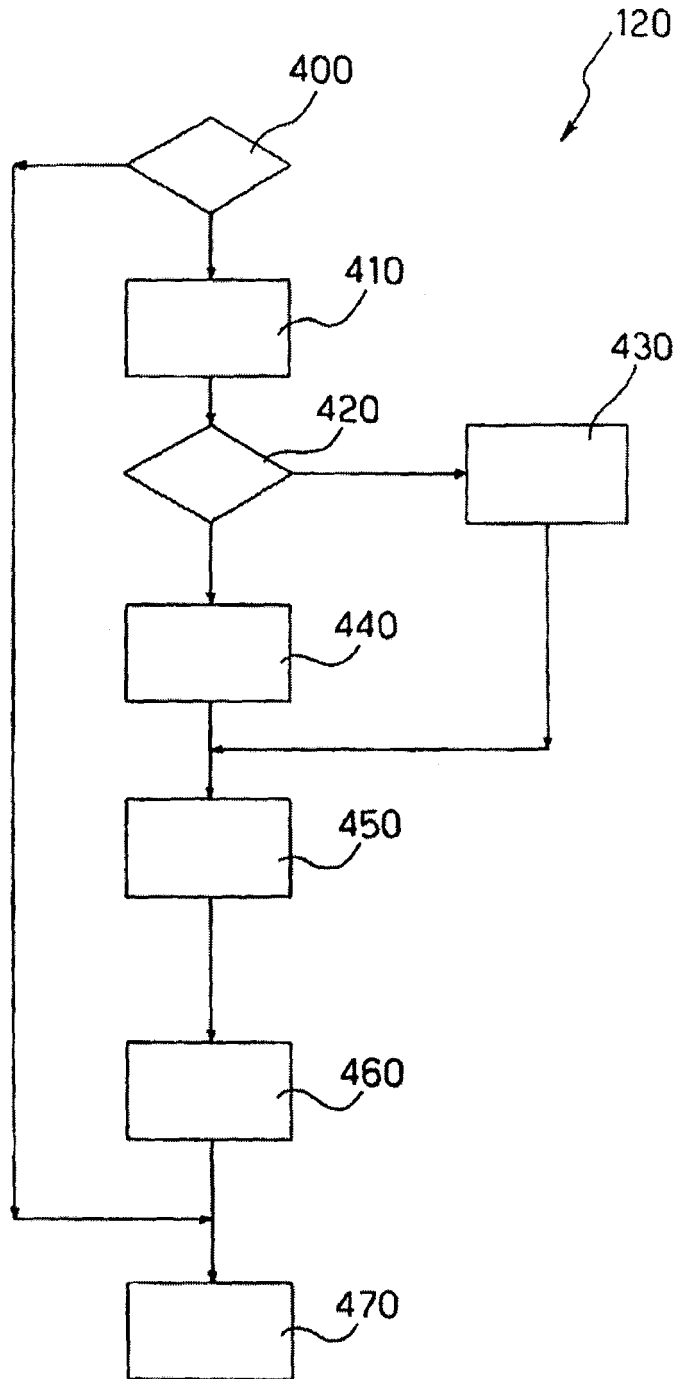


Fig. 6

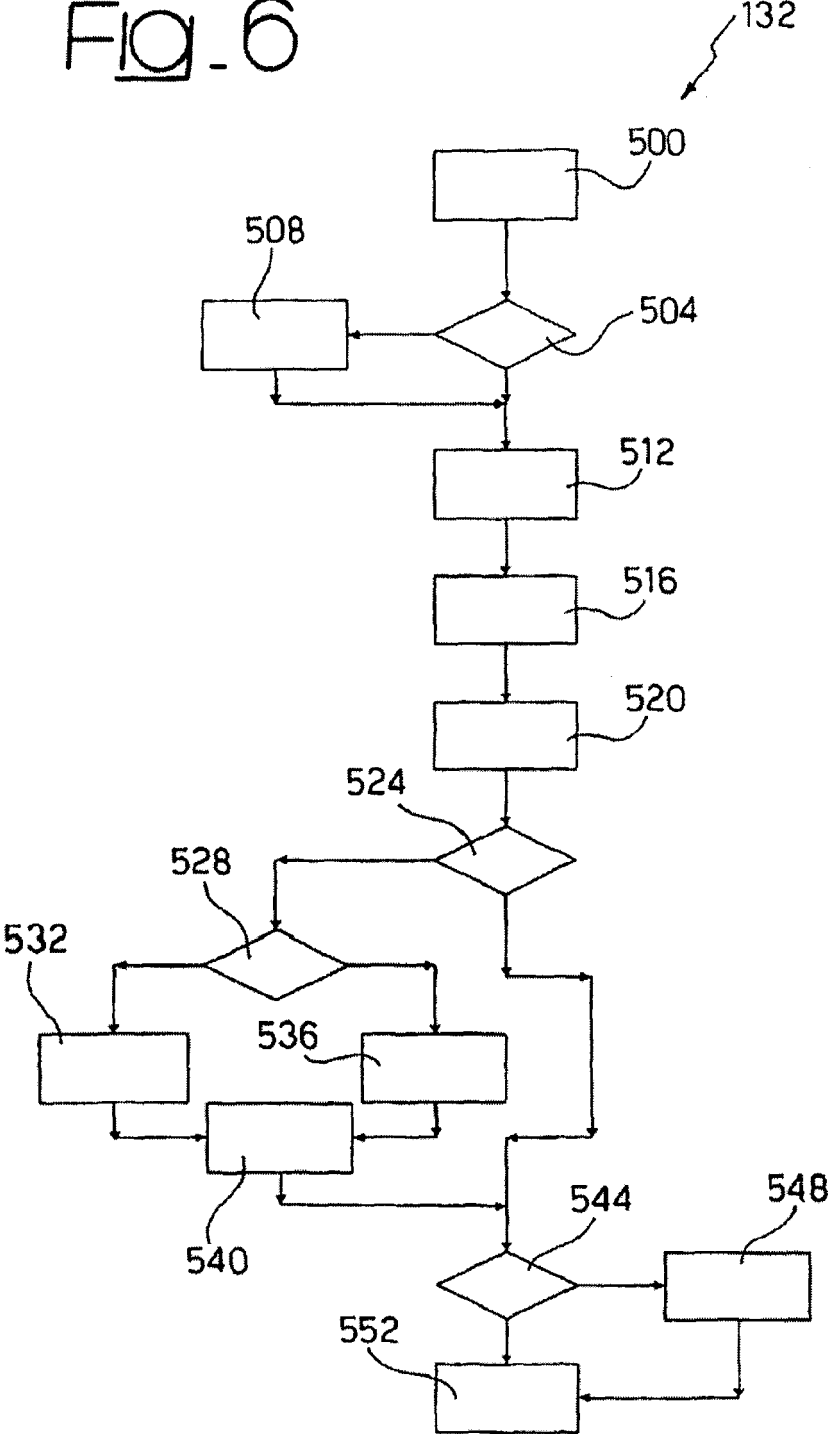


Fig. 7

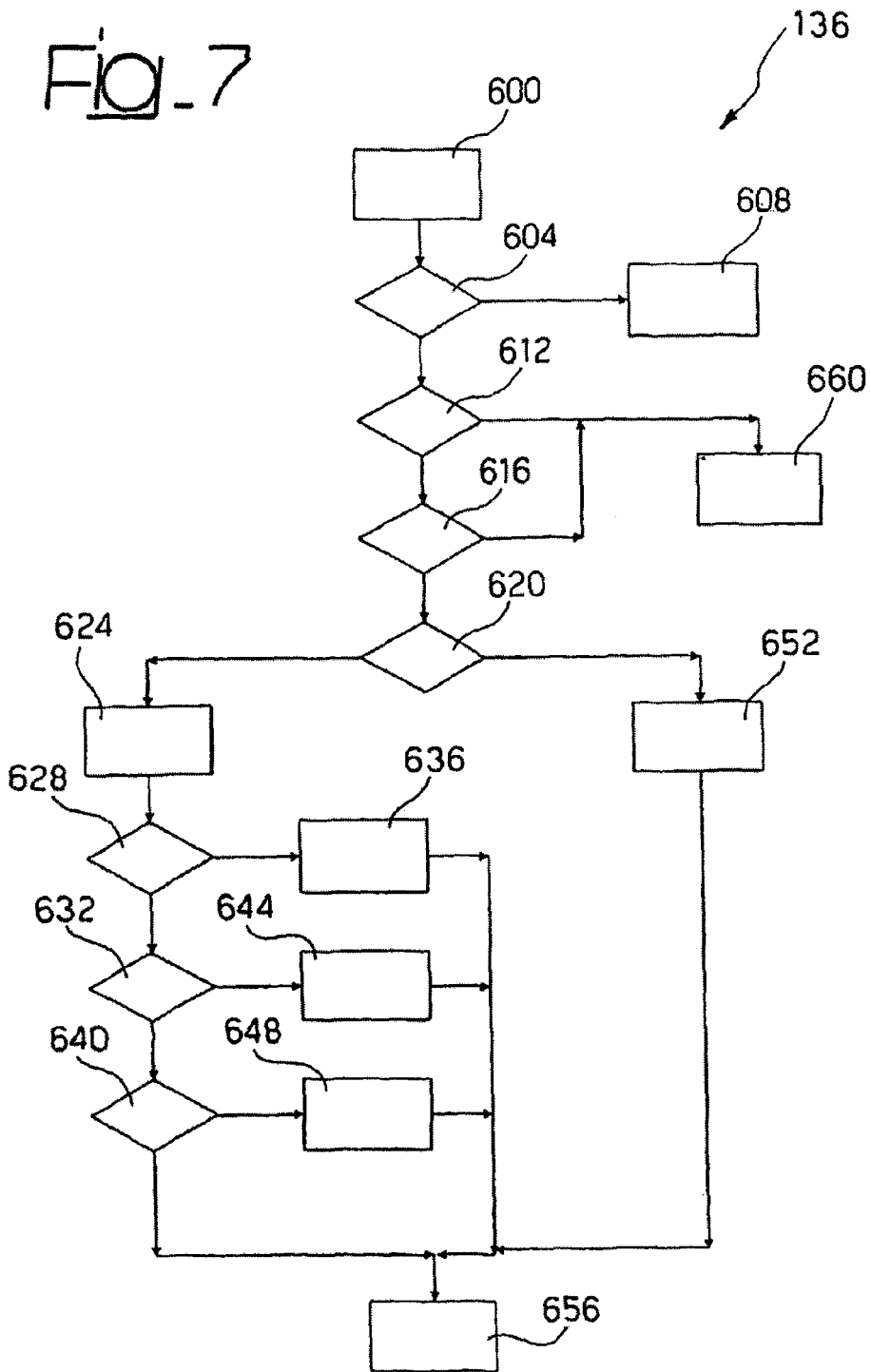


Fig. 8

