(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 284 932**

(51) Int. Cl.:
G10L 11/04 (2006.01)
H03M 7/30 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Número de solicitud europea: **02774815 .1**
(86) Fecha de presentación : **08.11.2002**
(87) Número de publicación de la solicitud: **1444685**
(87) Fecha de publicación de la solicitud: **11.08.2004**

(54) Título: **Método para comprimir datos de diccionario.**

(30) Prioridad: **12.11.2001 FI 20012193**

(45) Fecha de publicación de la mención BOPI:
16.11.2007

(45) Fecha de la publicación del folleto de la patente:
16.11.2007

(73) Titular/es: **Nokia Corporation**
Keilalahdentie 4
02150 Espoo, FI

(72) Inventor/es: **Tian, Jilei**

(74) Agente: **Curell Suñol, Marcelino**

DESCRIPCIÓN

Método para comprimir datos de diccionario.

Antecedentes de la invención

La presente invención se refiere al reconocimiento de voz independiente del hablante, y de forma más precisa a la compresión de un diccionario de pronunciación.

Durante los últimos años se han desarrollado diferentes aplicaciones de reconocimiento de voz, por ejemplo, para interfaces de usuario en automóviles y terminales móviles, tales como teléfonos móviles, dispositivos PDA y ordenadores portátiles. Entre los métodos conocidos para los terminales móviles se incluyen métodos para llamar a una persona específica pronunciando en voz alta su nombre en el micrófono del terminal móvil y estableciendo una llamada con el número según el nombre pronunciado por el usuario. No obstante, los métodos dependientes del hablante actuales requieren habitualmente un entrenamiento del sistema de reconocimiento de voz de manera que reconozca la pronunciación correspondiente a cada nombre. El reconocimiento de voz independiente del hablante mejora la usabilidad de una interfaz de usuario controlada por voz, ya que se puede omitir la etapa de entrenamiento. En la selección del nombre independiente del hablante, la pronunciación de los nombres se puede almacenar de antemano, y el nombre pronunciado por el usuario se puede identificar con la pronunciación predefinida, por ejemplo, una secuencia de fonemas. Aunque en muchos idiomas la pronunciación de muchas palabras se puede representar mediante reglas, o incluso modelos, estas reglas o modelos todavía no pueden generar correctamente la pronunciación de algunas palabras. No obstante, en muchos idiomas, la pronunciación no se puede representar mediante reglas generales de pronunciación, sino que cada palabra tiene una pronunciación específica. En estos idiomas, el reconocimiento de la voz se basa en el uso de los denominados diccionarios de pronunciación en los cuales, en una estructura de tipo lista, se almacenan una forma escrita de cada palabra del idioma y la representación fonética de su pronunciación.

En los teléfonos móviles, el tamaño de la memoria está limitado frecuentemente debido a razones de costes y de tamaño del hardware. Esta situación impone también limitaciones sobre las aplicaciones de reconocimiento de voz. En un dispositivo capaz de contener múltiples idiomas de la interfaz de usuario, la solución del reconocimiento de voz independiente del hablante con frecuencia usa diccionarios de pronunciación. Debido a que un diccionario de pronunciación tiene un tamaño habitualmente grande, por ejemplo, 37 KB para dos mil nombres, es necesario comprimir el mismo para almacenarlo. En términos generales, la mayoría de los métodos de compresión de texto se incluyen en dos clases: basados en diccionarios y basados en estadística. Existen varias implementaciones diferentes en la compresión basada en diccionarios, por ejemplo, el LZ77/78 y el LZW (Lempel-Ziv-Welch). Combinando un método estadístico, por ejemplo, la codificación aritmética, con técnicas de modelado potentes, se puede lograr un rendimiento mejor que con solamente métodos basados en diccionarios. No obstante, el problema del método basado en la estadística es que requiere una memoria de trabajo (memoria intermedia) grande durante el proceso de descompresión. Por esta razón, esta solución no es adecuada para ser usada en dispositivos electrónicos portátiles pequeños tales como terminales móviles.

El documento US-A-5 930 754 da a conocer un método para procesar un diccionario de pronunciación con vistas a su compresión. El léxico de pronunciación consta de ortografías emparejadas con representaciones fonéticas correspondientes. La secuencia de letras se alinea con su secuencia correspondiente de fonos. Con las secuencias alineadas se entrena una red neuronal.

Aunque los métodos de compresión existentes son en general satisfactorios, la compresión de los diccionarios de pronunciación no es suficientemente eficaz para los dispositivos portátiles.

Breve descripción de la invención

El objetivo de la invención es proporcionar un método de compresión más eficaz para comprimir un diccionario de pronunciación. El objetivo de la invención se alcanza mediante un método, dispositivos electrónicos, un sistema y productos de programa de ordenador caracterizados por los aspectos que se dan a conocer en las reivindicaciones independientes. En las reivindicaciones subordinadas se exponen las formas de realización preferidas de la invención.

Según un primer aspecto de la invención tal como el reivindicado en la reivindicación independiente 1, el diccionario de pronunciación se preprocesa antes de la compresión. El preprocesado se puede usar junto con cualquier método para comprimir un diccionario. En el preprocesado, cada entrada del diccionario de pronunciación se alinea usando un algoritmo estadístico. Durante la alineación, una secuencia de unidades de carácter y una secuencia de unidades de fonema se modifican de manera que presenten el mismo número de unidades en las secuencias. A continuación, las secuencias alineadas de unidades de carácter y unidades de fonema se entrelazan de manera que cada unidad de fonema se inserta en una posición predeterminada con respecto a la unidad de carácter correspondiente.

Típicamente, una secuencia de unidades de carácter es una secuencia de texto que contiene letras. Dependiendo del idioma, el conjunto alfabético se puede ampliar de manera que incluya más letras o símbolos que el alfabeto inglés convencional.

Una secuencia de unidades de fonema representa la pronunciación de la palabra y habitualmente contiene letras y símbolos, por ejemplo, "@", "A:", "{" en la notación SAMPA (Alfabeto Fonético para Métodos de Evaluación de la Voz). El alfabeto fonético también puede contener caracteres no imprimibles. Como un fonema se puede representar con más de una letra o símbolo, los fonemas se separan por un carácter de espacio en blanco.

De acuerdo con un segundo aspecto de la invención tal como el reivindicado en la reivindicación independiente 8, un dispositivo electrónico está configurado para convertir en una secuencia de unidades de fonema una entrada de cadena de texto. En la memoria del dispositivo se almacena un diccionario de pronunciación comprimido y preprocesado que comprende entradas, comprendiendo las entradas un primer conjunto de unidades que comprende unidades de carácter y un segundo conjunto de unidades que comprende unidades de fonema, en el que las unidades del primer conjunto y las unidades del segundo conjunto se alinean y entrelazan insertando cada unidad de fonema en una posición predeterminada con respecto a la unidad de carácter correspondiente. Se halla una entrada coincidente para la entrada de cadena de texto a partir del diccionario de pronunciación preprocesado usando las unidades del primer conjunto de unidades de la entrada de las posiciones predeterminadas. A partir de la entrada coincidente se seleccionan y concatenan en una secuencia de unidades de fonema unidades del segundo conjunto de unidades. De la secuencia de unidades de fonema se eliminan también los espacios vacíos.

Según un tercer aspecto de la invención tal como el reivindicado en la reivindicación independiente 11, un dispositivo electrónico está configurado para convertir en una secuencia de unidades de carácter una entrada de información de voz. En la memoria del dispositivo se almacena un diccionario de pronunciación comprimido y preprocesado que comprende entradas, comprendiendo las entradas un primer conjunto de unidades que comprende unidades de carácter y un segundo conjunto de unidades que comprende unidades de fonema, en el que las unidades del primer conjunto y las unidades del segundo conjunto se alinean y entrelazan insertando cada unidad de fonema en una posición predeterminada con respecto a la unidad de carácter correspondiente. Los modelos de pronunciación para la representación fonémica de cada entrada bien se almacenan en la memoria junto con el diccionario de pronunciación o bien se crean durante el proceso. Se halla una entrada coincidente para la información de voz comparando la información de voz con los modelos de pronunciación y seleccionando la entrada que presente una mayor correspondencia. A partir de la entrada coincidente, se seleccionan y concatenan en una secuencia de unidades de carácter unidades del primer conjunto de unidades. Finalmente, de la secuencia de unidades de carácter se eliminan los espacios vacíos.

Una de las ventajas de la invención es que con el preprocesado descrito, se reduce la entropía (H) del diccionario. Según la teoría de la información, un índice de entropía (H) bajo indica que se puede alcanzar una compresión más eficaz, ya que el índice de entropía determina el límite inferior para la compresión (la relación de compresión con la mejor compresión sin pérdidas posibles). Esta situación permite una mejor compresión, y el requisito de memoria es menor. Además, el diccionario de pronunciación resulta relativamente sencillo y rápido de aplicar para el reconocimiento de voz.

En una de las formas de realización de la invención, el algoritmo HMM-Viterbi está adaptado para ser usado para la alineación. El algoritmo HMM-Viterbi garantiza que la alineación se realiza de una manera óptima en el sentido estadístico, y por lo tanto minimiza la entropía remanente de la entrada del diccionario. Además, una de las ventajas al usar el algoritmo HMM-Viterbi en la alineación es que se puede alcanzar una alineación más óptima en el sentido estadístico.

En otra de las formas de realización de la invención, al preprocesado se le añade una etapa de establecimiento de correspondencias. El establecimiento de correspondencias se puede realizar bien antes o bien después de la alineación. En esta etapa, se establece una correspondencia de cada unidad de fonema con un símbolo y en lugar de representar las unidades de fonema mediante múltiples caracteres, se usa un único símbolo para indicar las unidades de fonema. Mediante el uso de la técnica de establecimiento de correspondencias, se pueden eliminar de la entrada los caracteres de espacio en blanco, e incluso sigue siendo posible la decodificación de la secuencia entrelazada. La eliminación de caracteres de espacio en blanco mejora adicionalmente la relación de compresión. Adicionalmente, una de las ventajas del establecimiento de correspondencias es que el método se puede adaptar a múltiples idiomas, o incluso se puede usar una gran tabla de correspondencias para todos los idiomas en el dispositivo.

Breve descripción de los dibujos

A continuación se describirá más detalladamente la invención por medio de formas de realización preferidas y haciendo referencia a los dibujos adjuntos, en los cuales:

la Figura 1 es un diagrama de bloques que ilustra un dispositivo de procesado de datos, el cual soporta el preprocesado y la compresión del diccionario de pronunciación según una de las formas de realización preferidas de la invención;

la Figura 2 es un diagrama de flujo de un método según una de las formas de realización preferidas de la invención;

la Figura 3 ilustra el uso del algoritmo HMM para la alineación del diccionario de pronunciación;

la Figura 4 muestra las etapas de preprocesado para una entrada del diccionario;

la Figura 5 es un diagrama de bloques que ilustra un dispositivo electrónico, el cual hace uso del diccionario de pronunciación preprocesado;

la Figura 6 es un diagrama de flujo que ilustra el uso del diccionario de pronunciación preprocesado cuando una cadena de textos se convierte en un modelo de pronunciación según una de las formas de realización preferidas de la invención; y

la Figura 7 es un diagrama de flujo que ilustra el uso del diccionario de pronunciación preprocesado cuando información de voz se convierte en una secuencia de unidades de texto según una de las formas de realización preferidas de la invención;

Descripción detallada de la invención

La Figura 1 ilustra un dispositivo de procesado de datos (TE) únicamente en relación con las partes relevantes para una forma de realización preferida de la invención. El dispositivo de procesado de datos (TE) puede ser, por ejemplo, un ordenador personal (PC) o un terminal móvil. La unidad de procesado de datos (TE) comprende medios I/O (I/O), una unidad de procesado central (CPU) y una memoria (MEM). La memoria (MEM) comprende una parte de memoria de solo lectura ROM y una parte reescribible, tal como una memoria de acceso aleatorio RAM y una memoria FLASH. La información usada para comunicarse con diferentes partes externas, por ejemplo, un CD-rom, otros dispositivos y el usuario, se transmite a través de los medios I/O (I/O) hacia/desde la unidad de procesado central (CPU). La unidad de procesado central (CPU) proporciona un bloque de preprocesado (PRE) y un bloque de compresión (COM). La funcionalidad de estos bloques se implementa típicamente ejecutando un código de software en un procesador, aunque también se puede implementar con una solución de hardware (por ejemplo, un ASIC) o en forma de una combinación de las dos. El bloque de preprocesado (PRE) proporciona las etapas de preprocesado de una forma de realización preferida ilustrada de forma detallada en la Figura 2. El bloque de compresión (COM) proporciona la compresión del diccionario de pronunciación, para lo cual se pueden usar varios métodos de compresión diferentes, por ejemplo, el LZ77, el LZW o una codificación aritmética. El preprocesado se puede combinar con cualquiera de los otros métodos de compresión para mejorar la eficacia de la compresión.

El diccionario de pronunciación que es necesario preprocesar y comprimir se almacena en la memoria (MEM). El diccionario también se puede descargar desde un dispositivo de memoria externo, por ejemplo, desde un CD-ROM o una red, usando los medios I/O (I/O). El diccionario de pronunciación comprende entradas que, a su vez, incluyen cada una de ellas una palabra en una secuencia de unidades de carácter (secuencia de texto) y en una secuencia de unidades de fonema (secuencia de fonemas). La secuencia de unidades de fonema representa la pronunciación de la secuencia de unidades de carácter. La representación de las unidades de fonema depende del sistema de notación fonética usado. Se pueden usar varios sistemas de notación fonética diferentes, por ejemplo, el SAMPA y el IPA. El SAMPA (Alfabeto Fonético para Métodos de Evaluación de la Voz) es un alfabeto fonético legible por máquina. La Asociación Fonética Internacional proporciona una norma de notación, el Alfabeto Fonético Internacional (IPA), para la representación fonética de numerosos idiomas. Se podría usar por ejemplo una entrada de diccionario que haga uso del sistema de notación fonética SAMPA:

Secuencia de Texto	Secuencia de Fonemas	Entrada
Father	F A: D @	Father f A: D @

La entropía, indicada mediante una *H*, es un atributo básico, el cual caracteriza el contenido de datos de la señal. Es posible hallar la forma más corta de presentar una señal (de comprimirla) sin perder ningún dato. La longitud de la representación más corta viene indicada por la entropía de la señal. En lugar de calcular el valor de entropía exacto de forma individual para cada señal, Shannon ha establecido un método para realizar una estimación del mismo (ver, por ejemplo, A Mathematical Theory of Communication, The Bell System Technical Journal, de C.E.Shannon, Vol. 27, págs. 379 a 423, 623 a 656, julio, octubre, 1948). Este planteamiento se describirá brevemente a continuación.

Sea $P(l_j|l_i)$ la probabilidad condicional de que el carácter en curso sea la letra j-ésima del alfabeto, dado que el carácter anterior es la letra i-ésima, y $P(l_i)$ la probabilidad de que el carácter anterior sea la letra i-ésima del alfabeto. El índice de entropía H_2 de los estadísticos de segundo orden es

$$H_2 = -\sum_{i=1}^{m} P(l_i) \cdot \sum_{j=1}^{m} P(l_j \mid l_i) \cdot \log_2 P(l_j \mid l_i) \qquad (1)$$

El índice de entropía H en un caso general viene dado por

$$H = \lim_{n \to \infty} -\frac{1}{n} \sum p(B_n) \cdot \log_2 p(B_n) \qquad (2)$$

en la que B_n representa los primeros caracteres. Es prácticamente imposible calcular el índice de entropía según la anterior ecuación (2). Usando este método de predicción de la ecuación (1), es posible realizar una estimación según la cual el índice de entropía de un texto inglés de 27 caracteres es aproximadamente 2,3 bits/carácter.

Para mejorar la compresión de un diccionario de pronunciación, se usa el preprocesado del texto para reducir su entropía.

La Figura 2 ilustra un método según una de las formas de realización preferidas de la invención. El método se centra en el preprocesado del diccionario de pronunciación para reducir el índice de entropía (H).

Cada entrada está alineada (200), es decir, las secuencias de texto y fonemas se modifican de manera que haya tantas unidades de fonema en la secuencia de fonemas como unidades de carácter en la secuencia de texto. Por ejemplo, en inglés, una letra se puede corresponder con cero, uno, o dos fonemas. La alineación se obtiene insertando epsilones (valores nulos) grafémicas o fonémicas entre las letras de la cadena de texto, o entre los fonemas de la secuencia de fonemas. Se puede evitar el uso de epsilones grafémicas introduciendo una lista corta de seudofonemas que se obtienen concatenando dos fonemas de los cuales se sabe que se corresponden con una única letra, por ejemplo, "x->k s". Para alinear las entradas, se debe definir el conjunto de fonemas permitidos para cada letra. La lista de fonemas incluye los seudofonemas correspondientes a la letra y la posible épsilon fonémica. El principio general consiste en insertar un valor nulo grafémico (definido como épsilon) en la secuencia de texto y/o un valor nulo fonémico (denominado también épsilon) en la secuencia de fonemas cuando sea necesario. A continuación se presenta la palabra usada anteriormente como ejemplo después de la alineación.

Secuencia de Texto	Secuencia de Fonemas	Entrada Alineada
father	f A: D @	father f A: D ε ε @

En este caso, la palabra "father" tiene 6 unidades y después de la alineación hay 6 fonemas en la secuencia de fonemas: "f A: D ε ε @". La alineación se puede realizar de varias maneras diferentes. Según una de las formas de realización de la invención, la alineación se realiza con el algoritmo HMM-Viterbi. El principio de funcionamiento de la alineación se ilustra y describe más detalladamente en la Figura 3.

Después de la alineación (200), se establece preferentemente una correspondencia (202) de cada fonema usado en el sistema de notación fonética con un único símbolo, por ejemplo, un código ASCII de bytes. No obstante, el establecimiento de las correspondencias no es necesario para obtener las ventajas de la invención, aunque puede mejorarlas adicionalmente. El establecimiento de las correspondencias se puede representar, por ejemplo, en una tabla de correspondencias. A continuación se presenta un ejemplo de cómo se podrían establecer las correspondencias de los fonemas de la palabra usada como ejemplo:

Símbolo de Fonemas	Número ASCII	Símbolo ASCII
f	0x66	f
A:	0x41	A
D	0x44	D
@	0x40	@
ε	0x5F	-

Al representar cada fonema con un símbolo, los dos caracteres que representan una unidad de fonema se pueden sustituir por solamente un símbolo ASCII de 8 bytes. Como consecuencia, el ejemplo queda:

Secuencia de Fonemas	Secuencia con Correspondencia Establecida (números ASCII)	Secuencia con Correspondencia Establecida (símbolos)
f A: D ε ε @	0x66 0x41 0x44 0x5F 0x5F 0x40	f A D_ _ @

Después de representar los fonemas con un símbolo, se pueden eliminar los espacios entre las unidades. También se puede eliminar el espacio entre la secuencia de texto y la secuencia de fonemas con la correspondencia establecida y alineada ya que existe el mismo número de unidades en ambas secuencias y es evidente qué caracteres pertenecen al texto y cuáles a la representación fonética.

Entrada alineada y con correspondencia establecida
fatherfAD_@

El establecimiento de correspondencias de las unidades de fonema con símbolos individuales (202) es una etapa importante para el entrelazado ya que se pueden evitar los caracteres de espacios en blanco. El establecimiento de correspondencias también mejora adicionalmente el propio resultado final, ya que los caracteres individuales ocupan menos espacio en comparación con, por ejemplo, las combinaciones de dos caracteres, y se incrementa la correlación con el carácter de texto correspondiente. El orden de la alineación (200) y el establecimiento de correspondencias (202) no afectan al resultado final, el establecimiento de correspondencias (202) también se puede llevar a cabo antes que la alineación.

La tabla de correspondencias depende únicamente del método de notación fonética usado en el diccionario de pronunciación. La misma se puede implementar de manera que sea independiente del idioma para que no sean necesarios diferentes sistemas o implementaciones en relación con los diferentes dialectos o idiomas. Si se usara una pluralidad de diccionarios de pronunciación en diferentes métodos de notación fonética, existiría la necesidad de tablas de correspondencia dependientes para cada método de notación fonética.

Después de la alineación (200) y del establecimiento de correspondencias (202), las entradas se entrelazan (204). Como el patrón carácter ->fonema tiene una probabilidad mayor (una entropía menor) que el patrón de letras consecutivo, especialmente si la alineación se ha llevado a cabo de forma óptima, se incrementa la redundancia. Esto se puede realizar insertando fonemas de pronunciación entre las letras de la palabra para formar una única palabra. En otras palabras, las unidades de fonema se insertan a continuación de las unidades de carácter correspondientes. Después de la alineación (200), la secuencia de texto y la secuencia de fonemas tienen el mismo número de símbolos y resulta sencillo hallar el par carácter-fonema. Por ejemplo:

Secuencia de Texto	**Secuencia de Fonemas**	**Entrada Entrelazada**
father	*FAD*_ _ @	f*f*a*A*t*D*h_e_r@

en la que los símbolos en cursiva y en negrita significan fonemas de pronunciación. Resulta evidente a partir del ejemplo que la composición y la descomposición de una entrada entre los formatos original y nuevo están definidas de una forma única, ya que la secuencia de texto y la secuencia de fonemas, que están entrelazadas, contienen el mismo número de unidades.

Después del preprocesado, se puede llevar a cabo la compresión (206) del diccionario de fonemas preprocesado.

La Figura 3 ilustra el HMM de grafemas para alinear las representaciones de texto y fonética de una entrada.

El Modelo de Markov Oculto (HMM) es un método estadístico bien conocido y ampliamente usado que se ha aplicado, por ejemplo, en el reconocimiento de voz. A estos modelos se les hace referencia también como fuentes de Markov o funciones probabilísticas de la cadena de Markov. La suposición subyacente en el HMM es que una señal se puede caracterizar satisfactoriamente como un proceso aleatorio paramétrico, y que los parámetros de un proceso estocástico se pueden determinar/estimar de una manera precisa y bien definida. Los HMM se pueden clasificar en modelos discretos y modelos continuos según si los acontecimientos observables asignados a cada estado son discretos, tales como palabras de código, o si los mismos son continuos. En cualquiera de los casos, la observación es probabilística. El modelo en el proceso estocástico subyacente no es observable de forma directa (está oculto) sino que puede verse únicamente a través de otro conjunto de procesos estocásticos que producen la secuencia de observaciones. El HMM está compuesto por estados ocultos con transición entre los estados. La representación matemática incluye tres elementos: la probabilidad de transición entre estados entre los estados en cuestión, la probabilidad de observación de cada estado y la distribución inicial de los estados. Dados un HMM y una observación, se usa el algoritmo de Viterbi para proporcionar la alineación de los estados de observación siguiendo el mejor camino.

En la presente invención se admite que el HMM se puede usar para resolver el problema de la alineación óptima de una secuencia observada con los estados del Modelo Oculto de Markov. Además, se puede usar el algoritmo de Viterbi en relación con el HMM para hallar la alineación óptima. Se puede hallar más información sobre los Modelos Ocultos de Markov y sus aplicaciones, por ejemplo, en el libro “Speech Recognition System Design and Implementation Issues”, págs. 322 a 342.

En primer lugar, para un par determinado letra-fonema, las penalizaciones p(f/l) se inicializan con cero si el fonema f se puede hallar en la lista de los fonemas permitidos de la letra l, en cualquier otro caso las mismas se inicializan con unos valores positivos elevados. Con los valores iniciales de las penalizaciones, el diccionario se alinea en dos etapas. En la primera etapa, se generan todas las alineaciones posibles para cada entrada del diccionario. Sobre la base de todas las entradas alineadas, a continuación se vuelven a calificar los valores de las penalizaciones. En la segunda etapa, se halla únicamente la mejor alineación individual para cada entrada.

Para cada entrada, se halla la alineación óptima con el algoritmo de Viterbi sobre el HMM de grafemas. El HMM de grafemas tiene una entrada (ES), una salida (EXS) y estados de letras (S1, S2 y S3). Las letras de las cuales se puede establecer una correspondencia con seudofonemas se gestionan de manera que presentan un estado de duración (EPS). Los estados 1 a 3 (S1, S2, S3) son los estados que se corresponden con las letras de la palabra. El estado 2

(S2) se corresponde con una letra que puede producir un seudofonema. Se permiten saltos desde todos los estados anteriores al estado en curso para soportar las epsilones fonémicas.

Cada estado y el estado de duración tienen un testigo que contiene una penalización acumulada (en forma de una suma de probabilidades logarítmicas) de alineación de la secuencia de fonemas con respecto al HMM de grafemas y las secuencias de estados que se corresponden con la puntuación acumulada. La secuencia de fonemas se alinea con respecto a las letras pasando a través de la secuencia de fonemas desde el comienzo hasta el final, un fonema cada vez. Para hallar la alineación de Viterbi entre las letras y los fonemas, se lleva a cabo un paso de testigo. Cuando los testigos pasan de un estado a otro, los mismos recogen la penalización de cada estado. El paso del testigo también puede conllevar la división de testigos y la combinación o selección de testigos para entrar en el siguiente estado. Para todos los estados del HMM se halla el testigo que al final presenta la penalización acumulada más baja. Sobre la base de la secuencia de estados del testigo, se puede determinar la alineación entre las letras de la palabra y los fonemas.

La alineación funciona correctamente para la mayoría de las entradas, aunque existen algunas entradas especiales que no se pueden alinear. En tales casos, se aplica otra alineación sencilla: se añaden epsilones grafémicas o fonémicas al final de las secuencias de letras o fonemas.

La Figura 4 ilustra más detalladamente el preprocesado de la entrada usado como ejemplo según una de las formas de realización preferidas de la invención.

La entrada original (400) tiene las dos partes, una secuencia de texto “father” y una secuencia de fonemas “f A: D @”. Estas dos secuencias se separan con un carácter de espacio en blanco y también se separan con caracteres de espacio en blanco las unidades de fonema.

En la alineación (402), las epsilones fonémicas y grafémicas se añaden de manera que haya el mismo número de unidades en ambas secuencias. En la palabra del ejemplo son necesarios dos epsilones fonémicas y el resultado de la secuencia de fonemas es “f A: D ε ε @”.

El establecimiento de la correspondencia (404) de las unidades de fonema con una representación de un símbolo hace que cambie solamente la secuencia de fonemas. Después del establecimiento de correspondencias, la secuencia de fonemas de la palabra del ejemplo es “f A D __ @”.

Cuando se establece una correspondencia (404) de la entrada, es posible eliminar los caracteres de espacio en blanco (406). Como consecuencia, se produce una cadena “fatherfAD__ @”.

La última etapa es el entrelazado (408) y la entrada del ejemplo es “ffaAtDh_e_r@”. A continuación, la entrada se puede procesar adicionalmente, por ejemplo, se puede comprimir.

Todas estas etapas se describen más detalladamente en la figura 2.

El método de preprocesado descrito anteriormente, incluyendo también el establecimiento de correspondencias (202), se probó experimentalmente. El experimento se llevó a cabo usando el Diccionario de Pronunciación de la Carnegie Mellon University, el cual es un diccionario de pronunciación para el inglés norteamericano que contiene más de 100.000 palabras y sus transcripciones. En el experimento, el rendimiento se evaluó en primer lugar usando métodos de compresión típicos basados en diccionarios, el LZ77 y el LZW, y un método de compresión basado en estadísticas, la compresión aritmética de 2º orden. A continuación se realizaron pruebas del rendimiento con el método de preprocesado junto con los métodos de compresión (LZ77, LZW y aritmético). En la Tabla 1, los resultados, proporcionados en kilobytes, muestran que el rendimiento del método de preprocesado es mejor en todos los caso. En general, el mismo se puede usar con cualquier algoritmo de compresión.

TABLA 1

Comparación del rendimiento de la compresión, sometida a prueba usando el diccionario de pronunciación de inglés de la CMU. Los resultados se encuentran en kilobytes

Método	Antes de la compresión	Compresión sin preprocesado	Compresión con preprocesado	Mejora
LZ77	2580	1181	940	20,4%
LZW	2580	1315	822	37,5%
Aritmético	2580	899	501	44,3%

Tal como se puede observar a partir de la Tabla 1, el preprocesado mejora la compresión con todos los métodos de compresión. Combinado con el método de compresión LZ77, el preprocesado mejoró la compresión en más de un 20%. La mejora es todavía mayor cuando el preprocesado se combina con el método LZW o con el método aritmético, proporcionando una mejor compresión en aproximadamente un 40%.

Debería entenderse que la invención se puede aplicar a cualquier diccionario de aplicación general que se use en el reconocimiento de voz y la síntesis de voz o en todas las aplicaciones en las que sea necesario almacenar un diccionario de pronunciación con un uso eficaz de la memoria. También es posible aplicar la invención a la compresión de otras listas cualesquiera que comprendan grupos de entradas de texto con una correlación elevada al nivel de los caracteres, por ejemplo, diccionarios comunes que presenten todas las formas de una palabra y programas de corrección ortográfica.

La Figura 5 ilustra un dispositivo electrónico (ED) únicamente en las partes relevantes para una forma de realización preferida de la invención. El dispositivo electrónico (ED) puede ser, por ejemplo, un dispositivo PDA, un terminal móvil, un ordenador personal (PC) o incluso cualquier dispositivo accesorio destinado a ser usado con los mismos, por ejemplo, unos auriculares inteligentes o un dispositivo de control remoto. El dispositivo electrónico (ED) comprende medios I/O (IO), una unidad de procesado central (PRO) y una memoria (ME). La memoria (ME) comprende una parte de memoria de solo lectura ROM y una parte reescribible, tal como una memoria de acceso aleatorio RAM y una memoria FLASH. La información usada para comunicarse con diferentes partes externas, por ejemplo, con la red, otros dispositivos o el usuario, se transmite a través de los medios I/O (IO) hacia/desde la unidad de procesado central (PRO). La interfaz de usuario, tal como un micrófono o un teclado que permita alimentar una secuencia de caracteres hacia el dispositivo, forma parte por lo tanto de los medios I/O (IO). Se puede descargar un diccionario de pronunciación preprocesado desde el dispositivo de procesado de datos (TE) hacia el dispositivo electrónico (ED) a través de los medios I/O (IO), por ejemplo, en forma de una descarga desde la red. A continuación, el diccionario se almacena en la memoria (ME) para su uso posterior.

Las etapas mostradas en las Figuras 6 y 7 se pueden implementar con un código de programa de ordenador ejecutado en la unidad de procesado central (PRO) del dispositivo electrónico (ED). El programa de ordenador se puede cargar en la unidad de procesado central (PRO) a través de los medios I/O (IO). La implementación también se puede realizar con una solución de hardware (por ejemplo, un ASIC) o con una combinación de las dos opciones mencionadas. Según una de las formas de realización preferidas, el diccionario de fonemas almacenado en la memoria (ME) del dispositivo (ED) se preprocesa tal como se describe en la Figura 2.

En la Figura 6 la unidad de procesado central (PRO) del dispositivo electrónico (ED) recibe una entrada de cadena de texto que es necesario convertir en un modelo de pronunciación. La cadena de texto introducida puede ser por ejemplo un nombre que el usuario ha añadido, usando los medios I/O (IO), a una base de datos de contacto del dispositivo electrónico (ED). En primer lugar, es necesario hallar (600) una entrada coincidente de entre el diccionario de pronunciación preprocesado que está almacenado en la memoria (ME). El hallazgo de la entrada coincidente se basa en la comparación de la cadena de texto introducida con las unidades de carácter de las entradas. Como las entradas están entrelazadas, una cadena de entradas es una combinación de unidades de carácter y de fonema. Si el entrelazado se realiza según la forma de realización preferida descrita en la Figura 2, cuando se compara la cadena introducida con la entrada, solamente se usa una de cada dos unidades. Las unidades de carácter de la entrada se pueden hallar seleccionando las unidades impares, comenzando desde la primera. La comparación se realiza con la cadena de caracteres original de la entrada, y por lo tanto se ignoran los espacios vacíos, por ejemplo, las epsilones grafémicas. Existen varios métodos y algoritmos para hallar la entrada coincidente, los cuales son conocidos de por sí para los expertos, y no es necesario describir los mismos en el presente documento, ya que no forman parte de la invención. Cuando las unidades de carácter coinciden exactamente con las unidades de la cadena de texto introducida, se ha hallado la entrada coincidente. No obstante, debería entenderse que en algunas aplicaciones podría resultar ventajoso usar alternativamente un algoritmo que no encuentre las coincidencias exactas, por ejemplo, uno que haga uso de los denominados comodines.

Cuando se ha hallado la entrada coincidente, se seleccionan (602) las unidades de fonema de la entrada. Debido al entrelazado (realizado según la forma de realización preferida descrita en la Figura 2), se usa una de cada dos unidades de la cadena de entrada. Para determinar las unidades de fonema, la selección comienza a partir de la segunda unidad. A continuación, las unidades seleccionadas se pueden concatenar para crear la secuencia de unidades fonémicas.

Como las entradas están alineadas, la secuencia de unidades de fonema puede incluir espacios vacíos, por ejemplo, epsilones fonémicas. Los espacios vacíos se eliminan para crear una secuencia que conste únicamente de fonemas (604).

Si el preprocesado del diccionario de fonemas incluyera también un establecimiento de correspondencias, es necesario un establecimiento de correspondencias inversas (606). El establecimiento de correspondencias inversas se puede llevar a cabo usando una tabla de correspondencias similar a la usada durante el preprocesado, aunque en un orden inverso. Esta etapa hace que cambie el primer método de representación, por ejemplo, representación de un carácter, de las unidades fonémicas al segundo método de representación, por ejemplo, el SAMPA, que se usa en el sistema.

Cuando se crea la secuencia de unidades de fonema, la misma típicamente se procesa adicionalmente, por ejemplo, se crea un modelo de pronunciación de la secuencia. Según una de las formas de realización, se crea un modelo de pronunciación para cada fonema usando, por ejemplo, el algoritmo HMM. Los modelos de pronunciación de los fonemas se almacenan en la memoria (ME). Para crear un modelo de pronunciación de una entrada, de la memoria (608) se recupera un modelo de pronunciación para cada fonema de la secuencia de fonemas. A continuación, estos modelos de fonema se concatenan (610) y se crea el modelo de pronunciación para la secuencia de fonemas.

La conversión de una entrada de cadena de texto en un modelo de pronunciación tal como se ha descrito anteriormente también se puede distribuir entre dos dispositivos electrónicos. Por ejemplo, el diccionario preprocesado se almacena en el primer dispositivo electrónico, por ejemplo, en la red, en la que se realiza el descubrimiento de una entrada coincidente (600). A continuación, la entrada coincidente se distribuye al segundo dispositivo electrónico, por ejemplo, un terminal móvil, en el que se realiza el resto del proceso (etapas 602 a 610).

La Figura 7 ilustra una forma de realización preferida en la que una información de voz se convierte a una secuencia de unidades de carácter en un dispositivo electrónico (ED) que utiliza un diccionario de pronunciación preprocesado. La unidad de procesado central (PRO) del dispositivo electrónico (ED) recibe una entrada de información de voz a través de los medios I/O (IO). Esta información de voz requiere ser convertida a una secuencia de unidades de carácter para su uso posterior, por ejemplo, para mostrarla en forma de texto en la pantalla o para compararla con una cadena de texto de una orden de voz predeterminada de un dispositivo controlado por voz.

El descubrimiento de una entrada coincidente (702) se basa en la comparación de la información de la voz de entrada con los modelos de pronunciación de cada entrada en el diccionario de pronunciación. De este modo, antes de la comparación, se modela (700) la pronunciación de cada entrada. Según una de las formas de realización preferidas, los modelos se crean en el dispositivo electrónico (ED). El diccionario de fonemas ya está entrelazado y alineado, y por lo tanto el modelado se puede realizar tal como se describe en la Figura 6, siguiendo las etapas 602 a 610. Cuando el modelado se realiza en el dispositivo electrónico (ED), se incrementa la necesidad de la capacidad de procesado y de la memoria de trabajo. En cambio, el consumo de memoria para almacenar el diccionario de pronunciación se puede mantener a un nivel bajo.

De acuerdo con una segunda forma de realización preferida, los modelos se crean antes que el preprocesado del diccionario de pronunciación en el dispositivo de procesado de datos (TE). El modelado se puede realizar tal como se describe en la Figura 6, siguiendo las etapas 608 y 610. Como el modelado se realiza antes que el preprocesado y el diccionario no está todavía entrelazado, alineado o con correspondencias establecidas, no son necesarias las etapas 602 a 606. A continuación, el modelo de pronunciación se almacena en la memoria (MEM) junto con la entrada. Cuando el diccionario se transfiere al dispositivo electrónico (ED), también se transfieren los modelos. En esta solución, son necesarias una capacidad de procesado y una memoria de trabajo menores para convertir la información de voz a una secuencia de texto. En cambio, se incrementa el consumo de memoria de la memoria de almacenamiento (ME).

El descubrimiento de una entrada coincidente (702) se realiza usando la información de la voz de entrada y los modelos de pronunciación de las entradas almacenadas en la memoria (ME). La información de voz se compara con cada entrada y se calcula una probabilidad del nivel de coincidencia de la información de la voz de entrada con el modelo de pronunciación de cada entrada. Después de calcular las probabilidades, se puede hallar la entrada coincidente seleccionando la entrada que presente la mayor probabilidad.

A continuación, a partir de la entrada coincidente se seleccionan las unidades de carácter (704). Debido al entrelazado, realizado tal como se describe en la Figura 2, se usa una de cada dos unidades de la cadena de entrada. La selección debe comenzar a partir de la primera unidad para obtener las unidades de carácter. A continuación, estas unidades seleccionadas se pueden concatenar para formar una secuencia de unidades grafémicas.

Debido a la alineación, la secuencia de las unidades grafémicas puede incluir espacios vacíos, por ejemplo, epsilones grafémicas. Para crear una secuencia que presente solamente grafemas, se eliminan los espacios vacíos (706). Como consecuencia, se obtiene una cadena de texto que se puede usar posteriormente en el sistema.

Un dispositivo electrónico, por ejemplo, un teléfono móvil con una interfaz de usuario para automóviles, dispone de un reconocimiento de voz independiente del hablante para órdenes de voz. Cada orden de voz es una entrada del diccionario de pronunciación. El usuario desea realizar una llamada telefónica mientras está conduciendo. Cuando el reconocimiento de voz está activo el usuario pronuncia “LLAMADA”. El teléfono recibe la orden de voz con un micrófono y transmite la información de voz a través de los medios I/O hacia la unidad de procesado central. La unidad de procesado central convierte la entrada de voz en una secuencia de texto tal como se describe en la Figura 7. La secuencia de texto se transmite a través de los medios I/O hacia la pantalla para proporcionar al usuario una indicación retroalimentada sobre lo que está haciendo el dispositivo. Además del texto en la pantalla, el dispositivo también proporciona una retroalimentación de audio. El modelo de pronunciación de la entrada coincidente, el cual se creó como parte del proceso de conversión de voz-a-texto, se transfiere a través de los medios I/O hacia el altavoz. A continuación, el teléfono realiza una llamada telefónica al número que ha seleccionado el usuario.

REIVINDICACIONES

1. Método para preprocesar un diccionario de pronunciación con vistas a su compresión en un dispositivo de procesado de datos, comprendiendo el diccionario de pronunciación por lo menos una entrada, comprendiendo la entrada una secuencia de unidades de carácter y una secuencia de unidades de fonema, **caracterizado** porque el método comprende las siguientes etapas:

- alinear (200) dicha secuencia de unidades de carácter y dicha secuencia de unidades de fonema usando un algoritmo estadístico; y

- entrelazar (204) dicha secuencia alineada de unidades de carácter y dicha secuencia alineada de unidades de fonema insertando cada unidad de fonema en una posición predeterminada con respecto a la unidad de carácter correspondiente.

2. Método según la reivindicación 1, **caracterizado** porque dicho algoritmo estadístico utiliza un algoritmo HMM-Viterbi.

3. Método según la reivindicación 1, **caracterizado** porque dichas unidades de fonema se sitúan junto a unidades de carácter correspondientes.

4. Método según cualquiera de las reivindicaciones anteriores, **caracterizado** porque dicha secuencia alineada de unidades de carácter y dicha secuencia alineada de unidades de fonema se realizan de manera que incluyan el mismo número de unidades insertando epsilones grafémicas en dicha secuencia de unidades de carácter y/o epsilones fonémicas en dicha secuencia de unidades de fonema.

5. Método según cualquiera de las reivindicaciones anteriores, **caracterizado** porque dichas unidades de carácter son letras o caracteres de espacio en blanco.

6. Método según cualquiera de las reivindicaciones anteriores, **caracterizado** porque dichas unidades de fonema son letras o caracteres de espacio en blanco que representan un único fonema o una épsilon fonémica y dicha una unidad se indica mediante por lo menos un carácter.

7. Método según la reivindicación 1, **caracterizado** porque el método comprende la siguiente etapa: se establece una correspondencia (202) de cada unidad de fonema con un símbolo.

8. Dispositivo electrónico configurado para convertir una entrada de cadena de texto en una secuencia de unidades de fonema, **caracterizado** porque comprende:

- unos medios para almacenar un diccionario de pronunciación comprimido y preprocesado que comprende entradas, comprendiendo las entradas un primer conjunto de unidades que comprende unidades de carácter y un segundo conjunto de unidades que comprende unidades de fonema, en el que las unidades del primer conjunto y las unidades del segundo conjunto se alinean (200) y se entrelazan (204) insertando cada unidad de fonema en una posición predeterminada con respecto a la unidad de carácter correspondiente;

- unos medios para hallar una entrada coincidente para dicha entrada de cadena de texto a partir de dicho diccionario de pronunciación preprocesado usando dicho primer conjunto de unidades de la entrada de entre posiciones predeterminadas;

- unos medios para seleccionar, de entre dicha entrada coincidente, unidades de dicho segundo conjunto de unidades de posiciones predeterminadas y concatenarlas en una secuencia de unidades de fonema; y

- unos medios para eliminar espacios vacíos de dicha secuencia de unidades de fonema.

9. Dispositivo electrónico según la reivindicación 8, **caracterizado** porque dicho dispositivo electrónico es un terminal móvil en un sistema de comunicaciones móviles.

10. Dispositivo electrónico según la reivindicación 8, **caracterizado** porque comprende además unos medios para establecer una correspondencia de cada unidad de fonema desde un primer método de representación fonémica a un segundo método de representación fonémica.

11. Dispositivo electrónico configurado para convertir una entrada de información de voz en una secuencia de unidades de carácter, **caracterizado** porque comprende:

- unos medios para almacenar un diccionario de pronunciación comprimido y preprocesado que comprende entradas, comprendiendo las entradas un primer conjunto de unidades que comprende unidades de carácter y un segundo conjunto de unidades que comprende unidades de fonema, en el que las unidades del primer conjunto y las unidades

del segundo conjunto se alinean y se entrelazan insertando cada unidad de fonema en una posición predeterminada con respecto a la unidad de carácter correspondiente;

- unos medios para almacenar o crear modelos de pronunciación de la representación fonémica de cada entrada;

- unos medios para hallar una entrada coincidente para dicha información de voz comparando dicha información de voz con dichos modelos de pronunciación y seleccionando la entrada que presente la mayor correspondencia;

- unos medios para seleccionar, de entre dicha entrada coincidente, unidades de dicho primer conjunto de unidades de posiciones predeterminadas y concatenarlas en una secuencia de unidades de carácter; y

- unos medios para eliminar espacios vacíos de dicha secuencia de unidades de carácter.

12. Sistema que comprende un primer dispositivo electrónico y un segundo dispositivo electrónico dispuestos en una conexión de comunicación mutua, estando configurado el sistema para convertir una entrada de cadena de texto en una secuencia de unidades de fonema, **caracterizado** porque:

- dicho primer dispositivo electrónico comprende unos medios para almacenar un diccionario de pronunciación comprimido y preprocesado que comprende entradas, en el que las entradas se alinean y se entrelazan insertando cada unidad de fonema en una posición predeterminada con respecto a la unidad de carácter correspondiente, comprendiendo las entradas un primer conjunto de unidades que comprende unidades de carácter y un segundo conjunto de unidades que comprende unidades de fonema;

- dicho primer dispositivo electrónico comprende unos medios para hallar una entrada coincidente para dicha entrada de cadena de texto a partir de dicho diccionario de pronunciación preprocesado usando dicho primer conjunto de unidades de la entrada;

- dicho primer dispositivo electrónico comprende unos medios para transmitir dicha entrada coincidente hacia el segundo dispositivo electrónico;

- dicho segundo dispositivo electrónico comprende unos medios para recibir dicha entrada coincidente desde el primer dispositivo electrónico;

- dicho segundo dispositivo electrónico comprende unos medios para seleccionar, de entre dicha entrada coincidente, unidades de dicho segundo conjunto de unidades y concatenarlas en una secuencia de unidades de fonema; y

- dicho segundo dispositivo electrónico comprende unos medios para eliminar espacios vacíos de dicha secuencia de unidades de fonema.

13. Producto de programa de ordenador cargable en la memoria de un dispositivo de procesado de datos, **caracterizado** porque comprende un código que se puede ejecutar en el dispositivo de procesado de datos consiguiendo que el dispositivo de procesado de datos:

- recupere de la memoria un diccionario de pronunciación que comprende por lo menos una entrada, comprendiendo la entrada una secuencia de unidades de carácter y una secuencia de unidades de fonema;

- alinee (200) dicha secuencia de unidades de carácter y dicha secuencia de unidades de fonema usando un algoritmo estadístico; y

- entrelace (204) dicha secuencia alineada de unidades de carácter y dicha secuencia alineada de unidades de fonema insertando cada unidad de fonema en una posición predeterminada con respecto a la unidad de carácter correspondiente.

14. Producto de programa de ordenador que se puede cargar en la memoria de un dispositivo electrónico, **caracterizado** porque comprende un código que se puede ejecutar en el dispositivo electrónico consiguiendo que el dispositivo electrónico:

- recupere de la memoria un diccionario de pronunciación preprocesado que comprende entradas, comprendiendo las entradas un primer conjunto de unidades que comprende unidades de carácter y un segundo conjunto de unidades que comprende unidades de fonema, en el que el primer conjunto de las unidades y el segundo conjunto de las unidades se alinean y se entrelazan insertando cada unidad de fonema en una posición predeterminada con respecto a la unidad de carácter correspondiente;

- halle (600) una entrada coincidente, a partir de dicho diccionario de pronunciación preprocesado, para una entrada de cadena de texto usando dicho primer conjunto de unidades de la entrada de entre las posiciones predeterminadas e ignorando espacios vacíos;

- seleccione (602), de entre dicha entrada coincidente, las unidades de dicho segundo conjunto de unidades de las posiciones predeterminadas y las concatene en una secuencia de unidades de fonema; y

- elimine (604) espacios vacíos de dicha secuencia de unidades de fonema.

15. Producto de programa de ordenador que se puede cargar en la memoria de un dispositivo electrónico, **caracterizado** porque comprende un código que se puede ejecutar en el dispositivo electrónico consiguiendo que el dispositivo electrónico:

- recupere de la memoria un diccionario de pronunciación preprocesado que comprende entradas, comprendiendo las entradas un primer conjunto de unidades que comprende unidades de carácter y un segundo conjunto de unidades que comprende unidades de fonema, en el que el primer conjunto de unidades y el segundo conjunto de unidades se alinean y se entrelazan insertando cada unidad de fonema en una posición predeterminada con respecto a la unidad de carácter correspondiente;

- almacene o cree (700) modelos de pronunciación de la representación fonémica de cada entrada;

- halle (702) una entrada coincidente para dicha información de voz comparando dicha información de voz con dichos modelos de pronunciación y seleccionando la entrada que presente la mayor correspondencia;

- seleccione (704), de entre dicha entrada coincidente, las unidades de dicho primer conjunto de unidades de las posiciones predeterminadas y las concatene en una secuencia de unidades de carácter; y

- elimine (706) espacios vacíos de dicha secuencia de unidades de carácter.

16. Dispositivo de procesado de datos que comprende memoria para almacenar un diccionario de pronunciación que comprende por lo menos una entrada, comprendiendo la entrada una secuencia de unidades de carácter y una secuencia de unidades de fonema, **caracterizado** porque el dispositivo comprende:

- unos medios para recuperar de la memoria un diccionario de pronunciación que comprende por lo menos una entrada, comprendiendo la entrada una secuencia de unidades de carácter y una secuencia de unidades de fonema;

- unos medios para alinear dicha secuencia de unidades de carácter y dicha secuencia de unidades de fonema usando un algoritmo estadístico; y

- unos medios para entrelazar dicha secuencia alineada de unidades de carácter y dicha secuencia alineada de unidades de fonema insertando cada unidad de fonema en una posición predeterminada con respecto a la unidad de carácter correspondiente.

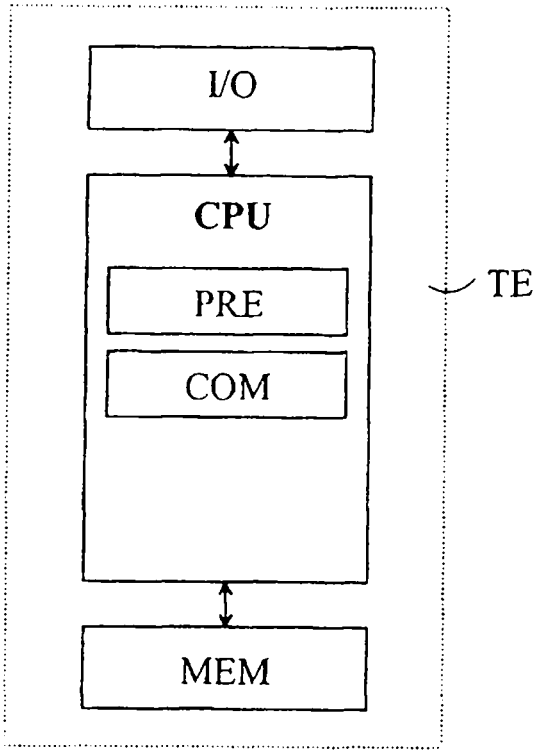

Fig. 1

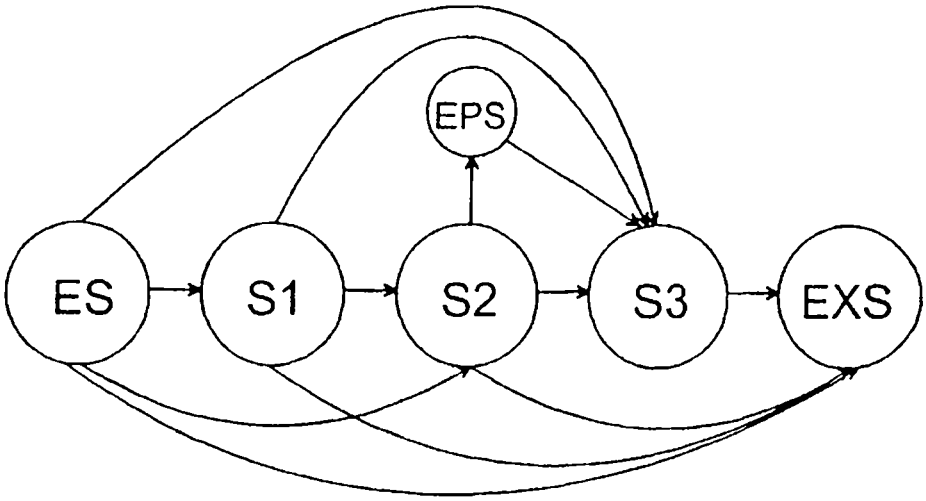

Fig. 3

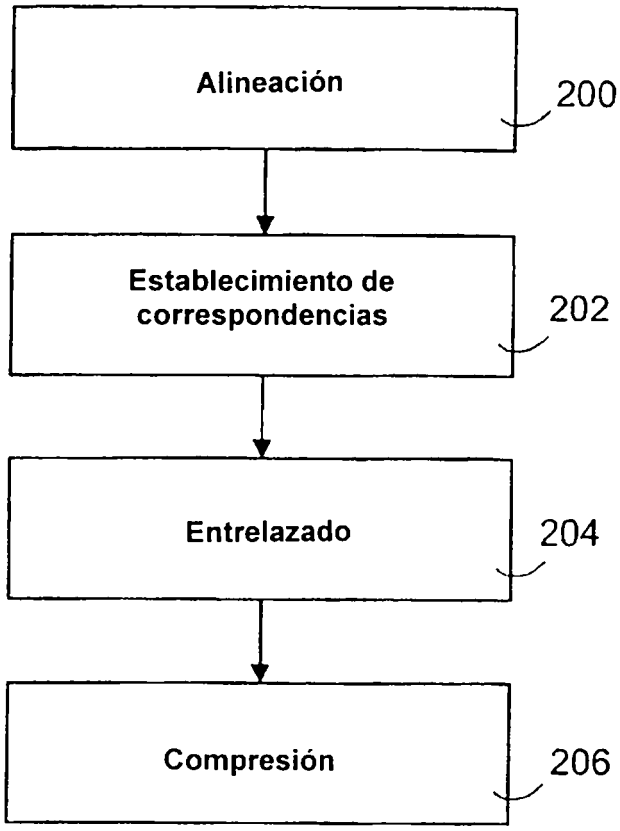
Alineación
200
Establecimiento de correspondencias
202
Entrelazado
204
Compresión
206

Fig. 2

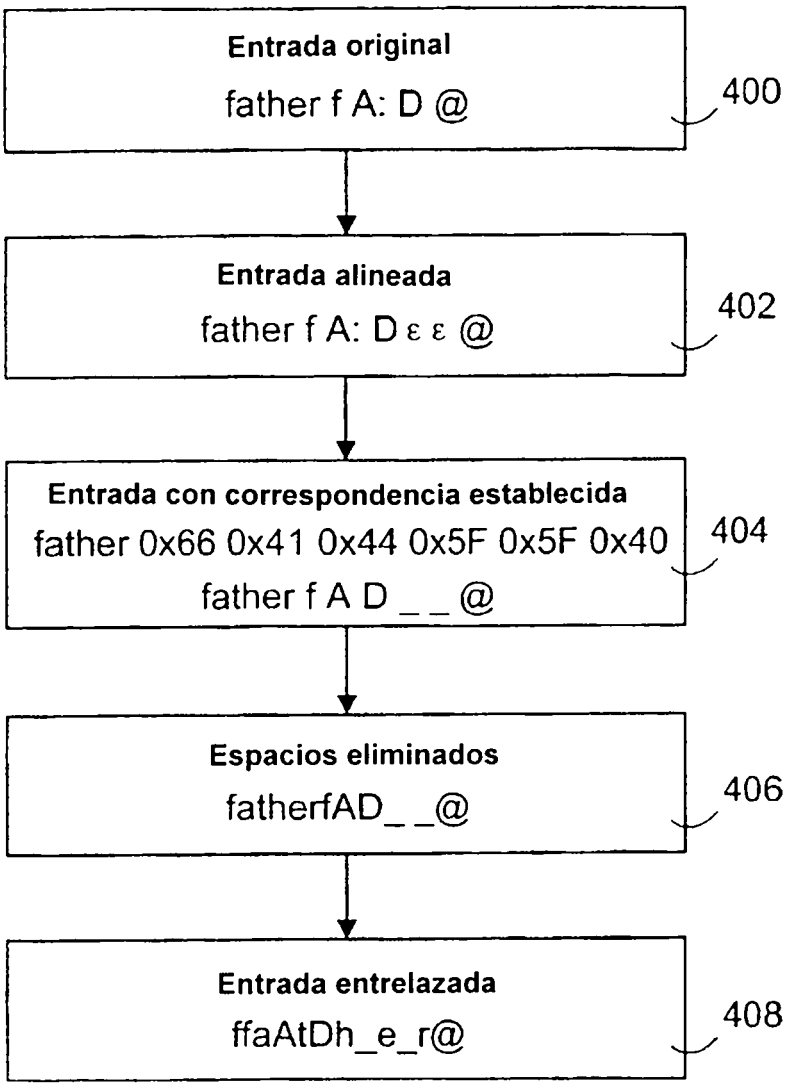
Entrada original
father f A: D @
400
Entrada alineada
father f A: D ε ε @
402
Entrada con correspondencia establecida
father 0x66 0x41 0x44 0x5F 0x5F 0x40
father f A D _ _ @
404
Espacios eliminados
fatherfAD_ _@
406
Entrada entrelazada
ffaAtDh_e_r@
408

Fig. 4

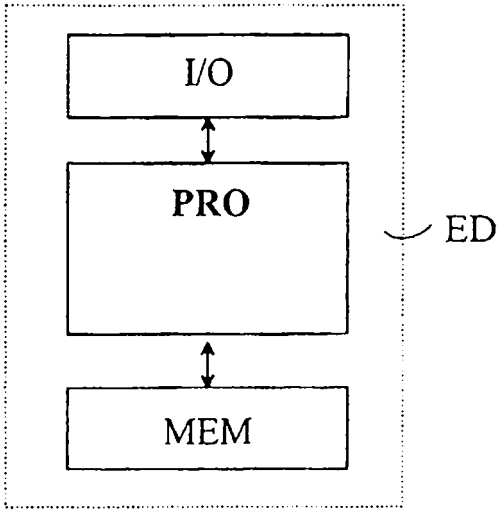

Fig. 5

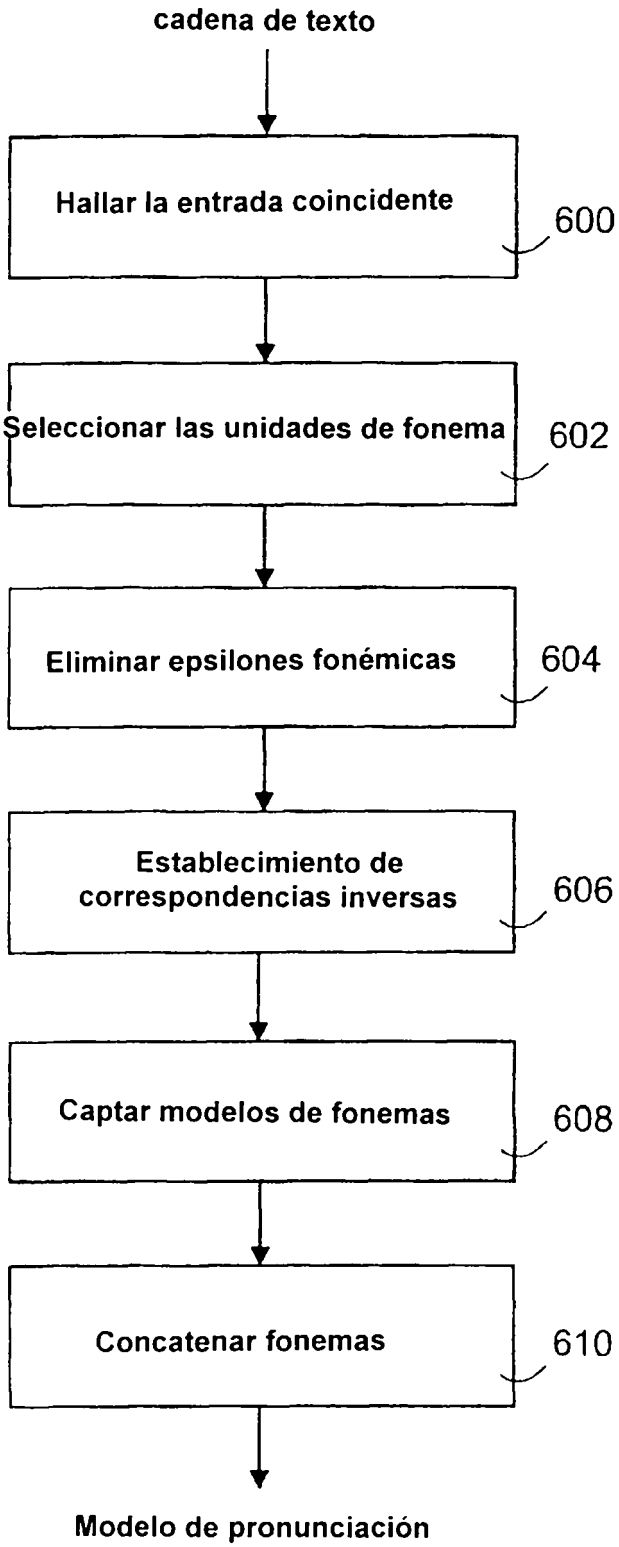
cadena de texto
Hallar la entrada coincidente
600
Seleccionar las unidades de fonema
602
Eliminar epsilones fonémicas
604
Establecimiento de correspondencias inversas
606
Captar modelos de fonemas
608
Concatenar fonemas
610
Modelo de pronunciación

Fig. 6

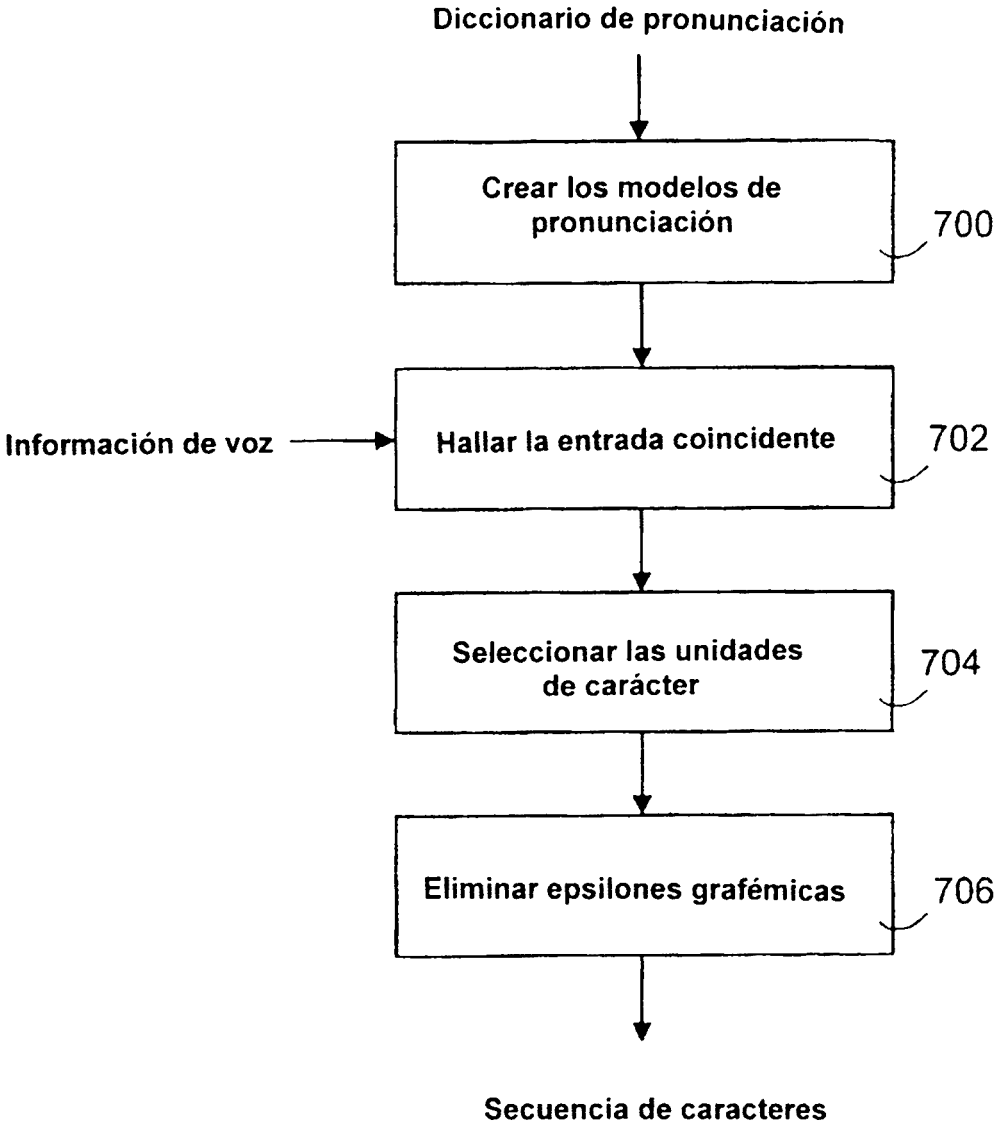
Diccionario de pronunciación
Crear los modelos de pronunciación
700
Información de voz
Hallar la entrada coincidente
702
Seleccionar las unidades de carácter
704
Eliminar epsilones grafémicas
706
Secuencia de caracteres

Fig. 7