(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 728 871**

(51) Int. CI.:

G10L 15/02 (2006.01)
G10L 25/75 (2013.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **17.12.2014 PCT/IB2014/067023**

(87) Fecha y número de publicación internacional: **25.06.2015 WO15092711**

(96) Fecha de presentación y número de la solicitud europea: **17.12.2014 E 14821858 (9)**

(97) Fecha y número de publicación de la concesión europea: **03.04.2019 EP 3084757**

(54) Título: **Método y aparato para reconocimiento de voz automático**

(30) Prioridad:

18.12.2013 GB 201322377

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
29.10.2019

(73) Titular/es:

OXFORD UNIVERSITY INNOVATION LIMITED (100.0%)
Buxton Court, 3 West Way, Botley
Oxford OX2 0JB, GB

(72) Inventor/es:

LAHIRI, ADITI;
REETZ, HENNING y
ROBERTS, PHILIP

(74) Agente/Representante:

ARIAS SANZ, Juan

DESCRIPCIÓN

Método y aparato para reconocimiento de voz automático

Esta invención se refiere a un método de reconocimiento de voz automático y un aparato que puede hacerse funcionar para llevar a cabo el método.

Antecedentes de la invención

El reconocimiento de voz automático es una tecnología ampliamente utilizada, cuyas aplicaciones incluyen programas de dictado, programas de menú del emisor de llamada para sistemas telefónicos, y 'asistentes' por respuesta de voz en teléfonos móviles.

Un problema con tales sistemas es la carga computacional requerida para moverse del discurso codificado digitalmente para identificar las palabras reales articuladas. Los sistemas comerciales se basan en sistemas de emparejamiento por plantillas y estadísticos, en los que un espectro acústico particular y sus cambios durante periodos de tiempo se emparejan con un conjunto conocido de espectros o características espectrales. En estos sistemas, se usan los modelos de Hidden Markov y otros algoritmos de búsqueda de patrones con fines generales. El sistema se entrena en base a ejemplos de discurso real, y realiza su mejor estimación sobre qué información procedente de cualquier señal dada es relevante para la tarea de reconocimiento. La desventaja con tales sistemas es que requieren una gran cantidad de procesamiento, para emparejar espectros con valor extremadamente informativo. Por consiguiente, los programas de dictado deben entrenarse para trabajar de manera eficaz con la voz de un usuario particular. Cuando no es posible proporcionar un funcionamiento robusto, tal como en sistemas de menú de emisor de llamada, solo pueden identificarse un intervalo relativamente limitado de respuestas posibles. Incluso en ese caso, los sistemas de reconocimiento de voz convencionales pueden fallar al reconocer correctamente discurso con un fuerte acento regional o nacional, o cuando el locutor tiene una dificultad en el habla.

Se ha propuesto un enfoque alternativo, basándose en teorías lingüísticas, en el que las características fonológicas individuales se identifican dentro de la señal acústica (véase, por ejemplo, Lahiri, Aditi & Reetz, Henning, 2002. 'Underspecified recognition'. En Carlos Gussenhoven & Natasha Warner (eds.), Laboratory Phonology 7, 637-676, Berlín: Mouton of Gruyter; y Reetz, Henning: 'Underspecified Phonological Features for Lexical Access'. En Phonus 5, Institute of Phonetics (instituto de fonética), Universidad de Saarland, 161-173). Este enfoque se basa en el hecho de que los sonidos hablados específicos aparecen en el espectro acústico de maneras identificables, de modo que una sección de discurso hablado puede usarse para identificar una secuencia de características. Sin embargo, este enfoque no se ha implementado de manera eficaz a día de hoy.

Sumario de la invención

Según un primer aspecto de la invención se proporciona un método según la reivindicación 1.

Los parámetros acústicos de la señal de voz dentro de cada ventana de tiempo pueden comprender uno o más del valor de amplitud cuadrática media, la frecuencia fundamental de la señal de voz (F0), la frecuencia de uno o más formantes, F1, F2, F3 en la señal de voz y un espectro de la señal de voz.

Cada ventana de tiempo puede ser de 20 ms.

Cuando se calcula un espectro de la señal de voz, el método puede comprender además determinar un valor de inclinación global calculando la pendiente de una línea de regresión sobre la totalidad del espectro, un primer valor de inclinación calculando la pendiente de una línea de regresión sobre un primer intervalo de frecuencia, y un segundo valor de inclinación calculando la pendiente de una línea de regresión sobre un segundo intervalo de frecuencia.

El primer intervalo de frecuencia puede oscilar entre 300 Hz y 1500 Hz y el segundo intervalo de frecuencia puede oscilar entre 1500 Hz y 5000 Hz.

El método puede comprender determinar que la característica [CONSONÁNTICA] está activa si el valor de inclinación global >-0 dB/Hz.

El método puede comprender determinar que la característica [SONORA] está activa si el valor de inclinación global >0,015 dB/Hz.

El método puede comprender determinar que la característica [BAJA] está activa si la característica [SONORA] está activa y la frecuencia del primer formante F1 es >600 Hz.

El método puede comprender determinar que la característica [ALTA] está activa si la característica [SONORA] está activa y la frecuencia del primer formante F1 es <450 Hz.

El método puede comprender determinar que la característica [CORONAL] está activa si la característica [SONORA] está activa y la frecuencia del segundo formante F2 se encuentra en el intervalo 2000 Hz < F2 < 2750 Hz.

El método puede comprender determinar que la característica [DORSAL] está activa si la característica [SONORA] está activa y la frecuencia del segundo formante F2 < 1000 Hz.

El método puede comprender determinar que la característica [RTR] está activa si la característica [SONORA] está activa y la frecuencia del tercer formante F3 es < 4000 Hz y la frecuencia del primer formante F1 es > 400 Hz o la frecuencia del segundo formante F2 se encuentra en el intervalo 1300 Hz < F2 < 1500 Hz.

El método puede comprender determinar que la característica [OBSTRUYENTE] está activa si el valor de inclinación global es < -0,015 db/Hz.

El método puede comprender determinar que la característica [ALTA] está activa si la característica [OBSTRUYENTE] está activa y el primer valor de inclinación menos el segundo valor de inclinación es mayor que 2.

El método puede comprender determinar que una característica de pausa está activa si el valor de amplitud cuadrática media está por debajo de un umbral.

El método puede comprender determinar que la característica [PLOSIVA] está activa si la característica [OBSTRUYENTE] está activa y está precedida por una característica de pausa.

El método puede comprender determinar que la característica [NASAL] está activa si la frecuencia del primer formante F1 se encuentra en el intervalo 100 Hz < F1 <400 Hz, el ancho de banda del primer formante es menor de 300 Hz, la amplitud del primer formante F1 es superior a 70 dB y la amplitud del segundo formante F2 es menor de 90 dB.

El método puede comprender determinar que la característica [LABIAL] está activa si la frecuencia del segundo formante F2 se encuentra en el intervalo 700 Hz < F2 < 1200 Hz.

El método puede comprender determinar que la característica [ESTRIDENTE] está activa si la inclinación global es >0,1 dB/Hz.

El método puede comprender determinar que la característica [VOCÁLICA] está activa si la frecuencia del primer formante F1 se encuentra en el intervalo 200 Hz < F1 < 1000 Hz y el ancho de banda del primer formante es menor de 500 Hz.

El método puede comprender generar una secuencia de características fonológicas determinando las características fonológicas activas en cada ventana de tiempo y emitiendo las características de voz en orden cronológico.

Una secuencia de puntos de tiempo que no se encuentran en una zona estable y que se encuentran entre zonas estables pueden determinarse para formar zonas no estables.

La longitud de zona estable mínima puede ser 30 ms.

La puntuación de inestabilidad para un punto de tiempo puede verse aumentada para cada característica presente en el punto de tiempo anterior pero no presente en el punto de tiempo.

La puntuación de inestabilidad para un punto de tiempo puede verse aumentada para cada característica presente en el punto de tiempo, pero no presente en el punto de tiempo anterior.

La puntuación de inestabilidad para un punto de tiempo puede verse aumentada cuando el punto de tiempo y el punto de tiempo anterior comprenden características que forman pares exclusivos de manera mutua.

El método puede comprender determinar una penalización relativamente grande para una característica en la zona estable si una característica incompatible está presente en el segmento fonológico emparejado de la entrada léxica.

El método puede comprender determinar una penalización relativamente pequeña para una característica en la zona estable cuando una característica en la zona estable no está presente en el segmento fonológico emparejado de la entrada léxica o viceversa.

No pueden determinarse penalizaciones para una característica en la zona estable si la misma característica está presente en el segmento fonológico emparejado de la entrada léxica.

La penalización puede depender de la fracción de la zona estable en la que la característica está activa.

Una penalización de zona no estable puede determinarse para cada característica en una zona no estable dependiendo de las características presentes en el segmento fonológico emparejado de la entrada léxica alineadas con la zona estable en cada lado de la zona no estable.

Si un segmento fonológico de la entrada léxica se encuentra entre los segmentos fonológicos emparejados de la entrada léxica alineados con la zona estable en cada lado de la zona no estable, puede determinarse una penalización de zona no estable para cada característica en la zona no estable en comparación con ese segmento fonológico.

El método puede comprender comparar los segmentos fonológicos secuenciales con una pluralidad de entradas léxicas e identificar una palabra de la entrada léxica con la puntuación de emparejamiento más baja.

El método puede comprender solo comparar los segmentos fonológicos secuenciales con una entrada léxica si el número de segmentos fonológicos en la entrada léxica está dentro de un intervalo limitado del número de zonas en los segmentos fonológicos secuenciales.

Según un segundo aspecto de la invención se proporciona un aparato que puede hacerse funcionar para realizar un método según una cualquiera de las reivindicaciones anteriores.

El aparato puede comprender una base de datos que almacena un diccionario, comprendiendo el diccionario una pluralidad de entradas léxicas, comprendiendo cada entrada léxica una descripción de una palabra en cuanto a segmentos fonológicos.

El aparato puede hacerse funcionar para recibir una señal que comprende una señal de voz.

Breve descripción de los dibujos

Una realización de la invención se describe a modo de ejemplo solo con referencia a los dibujos adjuntos, en los que;

La figura 1 es una trayectoria de flujo que muestra el método completo,

la figura 2 es una ilustración esquemática de un aparato para realizar el método,

la figura 3 es un diagrama de flujo que muestra la etapa de análisis acústico de la figura 1 en más detalle,

la figura 4 es un ejemplo de un espectro acústico calculado durante la etapa de análisis acústico de la figura 1,

la figura 5 es un ejemplo de un espectro acústico suavizado que muestra la ubicación de los formantes,

la figura 6 es un diagrama de flujo que muestra la etapa de determinación de características de la figura 1 en más detalle,

la figura 7 es un diagrama de flujo que muestra la etapa de segmentación de la figura 1 en más detalle,

la figura 8 es un diagrama de flujo que muestra la etapa de emparejamiento de la figura 1 en más detalle,

las figuras 9a a 9c son gráficas que muestran penalizaciones determinadas en el método de la figura 8,

la figura 10 muestra un espectro de señal de voz durante el tiempo y un flujo de características clasificadas por zonas correspondiente tras las características de determinación y segmentación,

la figura 11 muestra la señal de voz de la figura 10 durante el procedimiento de emparejamiento con características fonológicas emparejadas con zonas estables, y

la figura 12 muestra la señal de voz de la figura 10 con características fonológicas finales emparejadas con todas las zonas.

Descripción detallada de las realizaciones preferidas

Ahora, haciendo referencia específica a los dibujos en detalle, se señala que las particularidades mostradas son a modo de ejemplo y con fines de discusión ilustrativa de las realizaciones preferidas de la presente invención únicamente, y se presentan para proporcionar lo que se considera que es la descripción más útil y fácilmente comprensible de los principios y aspectos conceptuales de la invención. A este respecto, no se pretende mostrar detalles estructurales de la invención en más detalle de lo necesario para una comprensión fundamental de la

invención, haciendo la descripción tomada junto con los dibujos que se evidente para los expertos en la técnica cómo pueden llevarse a la práctica las diversas formas de la invención.

Antes de explicar al menos una realización de la invención en detalle, ha de comprenderse que la invención no se limita en su aplicación a los detalles de construcción y la disposición de los componentes expuestos en la siguiente descripción o ilustrados en los dibujos. La invención puede aplicarse a otras realizaciones o llevarse a la práctica o llevarse a cabo de diversas maneras. Asimismo, ha de comprenderse que la fraseología y terminología empleadas en el presente documento presentan fines de descripción y no deben considerarse limitativas.

Se resume un método de reconocimiento de voz automático en la figura 1, y las etapas individuales se amplían y comentan en más detalle a continuación. En la figura 1, en la etapa 10, se recibe una señal que comprende una señal de voz. La señal puede encontrarse en cualquier formato adecuado, y puede someterse a digitalización o puede comprender una señal análoga que se somete a digitalización posteriormente antes de procesarla tal como se comenta a continuación. En la etapa 11, se realiza el análisis acústico, para identificar diversos parámetros acústicos de la señal de voz y su espectro. En la etapa 12, los parámetros acústicos se usan para identificar características fonológicas en la señal de voz. La secuencia de tiempo de características fonológicas se separa en segmentos correspondientes a sonidos de voz individuales en la etapa 13. En la etapa 14, la secuencia de segmentos se compara con un diccionario 15 para identificar una palabra emparejada, y el emparejamiento resultante se emite en 16.

En la figura 2, en 20, se muestra de manera general un sistema para llevar a cabo el método. El sistema puede ser un ordenador personal, una parte de un sistema mayor tal como un servidor, o implantarse en un dispositivo móvil tal como un teléfono, a modo de ejemplo. Un aparato generalmente mostrado en 21 incluye un elemento de entrada/salida 22 para recibir la señal de voz. Un almacén de datos 23 soporta el diccionario 15, y un aparato de procesamiento mostrado generalmente en 24 puede hacerse funcionar para ejecutar el programa 25 para llevar a cabo la etapa mostrada en la figura 1, y pasar la salida a conexión de salida 26, que puede ser, por ejemplo, otro sistema, otro programa en el propio sistema u otro módulo o componente del programa 25.

Tal como se ilustra en la figura 3, los parámetros acústicos de una señal de voz recibida se determinan de la siguiente manera. La señal de voz se recibe en la etapa 30 y se divide en una secuencia de ventanas, teniendo en este ejemplo una anchura de 20ms, desviándose el inicio de cada ventana sucesiva por 1ms. Por consiguiente, la primera ventana va de 0ms a 20ms, la segunda de 1ms a 21ms, la tercera de 2ms a 22ms y así sucesivamente. La anchura y desvío de ventana pueden elegirse para tener valores diferentes si se desea, por ejemplo, dependiendo de la capacidad de procesamiento disponible, calidad de señal u otra cosa.

La primera ventana se selecciona en la etapa 31 y se calculan algunos o todos de los siguientes parámetros acústicos:

a) la amplitud cuadrática media de las muestras en la ventana, que es una medición de cómo de alta es la señal en la ventana (etapa 32).

b) se calcula el espectro de la señal en la ventana. De manera conveniente, esto puede realizarse usando una transformada rápida de Fourier (FFT) (etapa 33).

c) se calcula la frecuencia fundamental de la forma de onda en la ventana. Esto puede derivarse del espectro, o puede calcularse usando un método de autocorrelación. La frecuencia fundamental es una medición de la frecuencia con la que vibran las cuerdas vocales, lo que determina el paso global del discurso codificado (etapa 34).

d) la frecuencia, ancho de banda y amplitud de los primeros cuatro formantes, en la etapa 35. Los formantes son propiedades resonantes principales de las cavidades oral y nasal, tal como se expresa en la señal de voz. Cuando las cuerdas vocales están vibrando, habitualmente es posible detectar cuatro formantes denominados F1, F2, F3 y F4 en la señal de voz. En el presente ejemplo tal como se comenta a continuación, para el reconocimiento de Palabras en inglés solo se usan los primeros 3 formantes, pero pueden ser necesarios todos los cuatro formantes para otras lenguas. Los formantes se extraen del espectro mediante codificación predictiva lineal de manera conocida. La figura 4 es un ejemplo de un espectro en el que se han identificado los formantes. Los formantes F1 a F4 se indican mediante las bandas 40 a 43 respectivamente.

e) un valor de inclinación global que es la pendiente de una línea de regresión mínima cuadrática por la totalidad del espectro calculado en la etapa (b) anterior, tal como se muestra en 44 en la figura 5.

f) un primer valor de inclinación que es la pendiente de una línea de regresión sobre una primera parte del espectro, tal como se muestra en 45 en la figura 5. En este ejemplo, la primera parte del espectro se extiende desde 300Hz hasta 1500Hz.

g) un segundo valor de inclinación que es la pendiente de una línea de regresión sobre una segunda parte del espectro, tal como se muestra en 46 en la figura 5. El valor de inclinación global, el primer valor de inclinación y el

segundo valor de inclinación se calculan en la etapa 36.

Dependiendo de la lengua con la que va a usarse el método, pueden medirse otros parámetros acústicos según sea apropiado.

En este ejemplo, la segunda parte del espectro se extiende desde 1500Hz hasta 5000Hz. Un ejemplo de un espectro se muestra en la figura 4. La línea de regresión mínima cuadrática sobre la totalidad del espectro puede observarse en 44, la línea de regresión sobre la primera parte del espectro se muestra en 45 y la línea de regresión sobre la segunda parte del espectro se muestra en 46.

Tal como se muestra en las etapas 37 y 38, las etapas 31 a 36 se repiten para cada ventana. Cuando se han calculado los parámetros acústicos para todas las ventanas, un flujo de parámetros acústicos se emite en 39. Por consiguiente, tras la etapa 11 de la figura 1 será evidente que el contenido de información de la señal de voz se ha reducido a un flujo de formante que comprende valores para un conjunto limitado de parámetros acústicos, en puntos de tiempo específicos.

En la etapa 12 de la figura 1, ilustrados en la figura 6, los parámetros acústicos se usan para determinar la presencia de características fonológicas en cada punto de tiempo. De manera conocida, las características fonológicas son características correspondientes a características articulatorias y acústicas particulares del discurso. En la etapa 50, se recibe el flujo de parámetros acústicos, y en la etapa 51 los parámetros acústicos correspondientes a la primera ventana se usan para identificar la presencia de características fonológicas activas usando una pluralidad de criterios de prueba.

En este ejemplo, los siguientes criterios de prueba se usan para determinar qué características están activas;

a) la característica [CONSONÁNTICA] está activa si el valor de inclinación global > 0 dB/Hz.

b) la característica [SONORA] está activa si el valor de inclinación global > -0,015 dB/Hz.

c) la característica [OBSTRUYENTE] está activa si el valor de inclinación global es > -0,015 dB/Hz.

d) la característica [BAJA] está activa si la característica [SONORA] está activa y la frecuencia del primer formante F1 es >600 Hz.

e) la característica [ALTA] está activa si la característica [SONORA] está activa y la frecuencia del primer formante F1 es <450 Hz.

f) la característica [CORONAL] está activa si la característica [SONORA] está activa y la frecuencia del segundo formante F2 se encuentra en el intervalo 2000 Hz < F2 < 2750 Hz.

g) la característica [DORSAL] está activa si la característica [SONORA] está activa y la frecuencia del segundo formante F2 < 1000 Hz

h) la característica [RTR] está activa si la característica [SONORA] está activa y la frecuencia del tercer formante F3 es < 4000 Hz y o bien la frecuencia del primer formante F1 es > 400 Hz o bien la frecuencia del segundo formante F2 se encuentra en el intervalo 1300 Hz < F2 < 1500 Hz.

i) la característica [PLOSIVA] está activa si la característica [OBSTRUYENTE] está activa y está precedida por una característica de pausa.

Una característica de pausa está activa si el valor de amplitud cuadrático medio está por debajo de un umbral.

j) la característica [NASAL] está activa si la frecuencia del primer formante F1 se encuentra en el intervalo 100 Hz < F1 <400 Hz, el ancho de banda del primer formante es menor de 300 Hz, la amplitud del primer formante F1 es superior a 70 dB y la amplitud del segundo formante F2 es menor de 90 dB.

k) la característica [LABIAL] está activa si la frecuencia del segundo formante F2 se encuentra en el intervalo 700 Hz < F2 < 1200 Hz.

l) la característica [ESTRIDENTE] está activa si la inclinación global es > 0,1 db/Hz.

m) la característica [VOCÁLICA] está activa si la frecuencia del primer formante F1 se encuentra en el intervalo 200 Hz < F1 < 1000 Hz y el ancho de banda del primer formante es menor de 500 Hz.

En el presente ejemplo, el método y aparato son para usarse en el reconocimiento de las lenguas inglesa o alemana habladas, y de modo que se identifiquen el conjunto de características enumeradas anteriormente. Para otras

lenguas, puede ser necesario identificar otras características, tales como [CONTINUANTE], [RADICAL], [RÓTICA], [GLOTIS EXTENDIDA], [LATERAL] o [ATR], y algunas de las características enumeradas anteriormente pueden no ser necesarias. Para otras lenguas, como parte de la identificación de otras características, puede ser necesario medir diferentes parámetros acústicos en las etapas 32 a 36. Por ejemplo, para lenguas tonales y contornos de entonación, puede ser necesario medir la frecuencia fundamental F0 para acentos tonales tales como [H*] o [L*].

Además, los criterios de prueba pueden variar según sea apropiado, y la frecuencia y valores de inclinación dados anteriormente son ejemplos empíricos. De manera ideal, los valores deben seleccionarse para abarcar la cantidad de intervalo que sea posible dentro del que se espera que varíe la frecuencia o inclinación de formante.

Tal como se muestra mediante las etapas 52 y 53, los criterios de prueba se aplican a cada punto de tiempo en el flujo de parámetros acústicos. Cuando se ha sometido a prueba cada punto de tiempo, la secuencia resultante se denomina flujo de característica, y se emite en la etapa 54. Será evidente que la señal de voz se ha reducido, por tanto, a una sencilla enumeración de la presencia o ausencia de un conjunto limitado de características fonológicas en cada punto de tiempo específico.

Aunque el método descrito en el presente documento es, generalmente, secuencial, porque una etapa completa se realiza y almacena temporalmente antes de pasar los resultados a la siguiente etapa, será evidente que el método puede ser un procedimiento continuo, de manera que cada punto de tiempo en el flujo de parámetros acústicos puede someterse a prueba y la parte resultante del flujo de características pasarse a la etapa de segmentación tal como se describió anteriormente, y así sucesivamente a través del método completo.

En la etapa 13 de la figura 1, el flujo de características se clasifica en zonas. De manera ideal, cada zona correspondería a un sonido fonético específico. En la práctica, sin embargo, las zonas no son claramente específicas, ya que una zona sucesiva puede contener las mismas características, y las características pueden no activarse o desactivarse simultáneamente en un límite de zona. Por tanto, el método descrito en el presente documento se basa en la identificación de zonas estables dentro del flujo de característica, dentro del que todos los puntos de tiempo pueden observarse con seguridad como que pertenecen al mismo segmento.

Para identificar zonas estables en el flujo de característica, se llevan a cabo las etapas tal como se muestra en la figura 7. Comenzando con el flujo de características en la etapa 60, para el primer punto de tiempo en el flujo de características, se calcula una puntuación de inestabilidad en la etapa 61, comparando las características extraídas en ese punto de tiempo con aquellas en los puntos de tiempo anteriores, de vuelta a un número configurable de milisegundos (50 por defecto). Tal como se muestra mediante las etapas 62, 63, se calcula una puntuación de inestabilidad para cada ventana.

La puntuación de inestabilidad comienza en cero y aumenta de la siguiente manera. Para cada punto de tiempo anterior, la puntuación aumenta:

i) en 1 para cada característica que está presente en el punto de tiempo anterior, pero no en el punto de tiempo cuya puntuación se está calculando.

ii) en un valor configurable para cada característica presente en el punto de tiempo actual, pero no en el punto anterior. Este valor se establece en 5 por defecto, porque una nueva característica que aparece en el punto de tiempo actual tiene más probabilidades de indicar un nuevo segmento que un segmento que ha aparecido antes que no se ha mostrado

iii) en un valor configurable diferente para cada caso en donde los puntos de tiempo anterior y actual contienen en conjunto ambas características de uno o más de los siguientes pares exclusivos de manera mutua: [CONSONÁNTICA/VOCÁLICA], [OBSTRUYENTE/ SONORA], [ESTRIDENTE/NASAL] y [ALTA/BAJA]. Dado que estas características son exclusivas entre sí, la conmutación de una de estas características a otra se considera un indicio importante de un nuevo segmento. En el presente ejemplo este valor se establece en un valor alto, 25.

Tras haber calculado un valor de inestabilidad para cada punto de tiempo, la secuencia resultante de valores de inestabilidad, ilustrada en la etapa 64, se denomina el contorno de inestabilidad. En la etapa 65, las zonas estables se identifican usando el contorno de inestabilidad. Los puntos de secuencia de tiempo que tienen una longitud mayor que una longitud de zona estable mínima y una puntuación de inestabilidad menor que un umbral de inestabilidad se determinan para formar una zona estable. La longitud de zona estable mínima se selecciona, preferiblemente, para ser aproximadamente la del segmento en estado preparado más corto posible, por ejemplo, una vocal o una consonante nasal con una onda glotal sostenida, o el ruido sostenido de una fricativa. En el presente ejemplo, la longitud de zona estable mínima es de 30ms y el umbral de inestabilidad es 400. Cuando se determina una secuencia de puntos de tiempo para definir una zona estable, se consideran que las características dentro de la zona pertenecen al mismo segmento fonológico. Al comienzo y final de cada palabra, y entre zonas estables se encuentran zonas no estables. Características dentro de zonas no estables no se descartan; en su lugar, decidir con qué segmento fonológico deberían alinearse se considera parte del procedimiento de búsqueda léxica.

Por consiguiente, tras completar la etapa de clasificación de zona 13 en la etapa 66, el flujo de características de la etapa 12 se ha dividido en zonas no estables y estables, ahora denominado flujo de características clasificadas por zonas. No se descarta ninguna información, ya que las zonas no estables pueden representar segmentos fonológicos válidos, y el procedimiento de decidir si están presentes segmentos fonológicos válidos o no se realiza como parte de la etapa 14, la etapa de emparejamiento léxico.

La etapa de emparejamiento de léxico 14 avanza asignando una puntuación de emparejamiento a posibles entradas léxicas que pueden avenirse al flujo de características segmentadas de la etapa 13, y seleccionando la entrada léxica con la mejor puntuación de emparejamiento como la palabra emparejada. La etapa de emparejamiento léxico se comenta en más detalle con referencia a la figura 8.

En la etapa 70, los flujos de características clasificadas por zonas se reciben y se selecciona una entrada léxica en la etapa 71. En la etapa de emparejamiento, solo aquellas entradas léxicas con varios segmentos fonológicos en la entrada léxica dentro de un intervalo limitado del número de zonas en los segmentos fonológicos secuenciales se someten a prueba, y esto se comprueba en la etapa 72. En el presente método, el intervalo va desde el número de zonas estables menos un parámetro de intervalo hasta el número de zonas, ya sea estable o no estable, más el parámetro de intervalo. En este ejemplo el parámetro de intervalo es 2. Los flujos de características clasificadas por zonas tienen cuatro zonas estables y cuatro zonas no estables, de modo que solo aquellas entradas léxicas con de 2 a 10 zonas se someten a prueba.

El parámetro de intervalo puede variar, o ser diferente en los dos extremos del intervalo. Puede concebirse que cada entrada léxica tenga un recuento de segmentos fonológicos almacenados en el diccionario 15, y solo aquellas entradas léxicas con un recuento de segmentos fonológicos en el intervalo limitado se comparan con el flujo de características clasificadas por zonas.

Si la longitud de entrada léxica está dentro del intervalo aceptable en la etapa 72, entonces en la etapa 73, se comparan el flujo de características clasificado por zonas con la entrada léxica, las zonas estables en el flujo de características clasificadas por zonas se emparejan con sus mejores emparejamientos de segmento fonológico en la entrada léxica. Puede usarse cualquier algoritmo de emparejamiento óptimo adecuado, en este ejemplo, el algoritmo de Needleman-Wunsch.

Entonces se calculan las puntuaciones de penalización en la etapa 74 comparando las características en las zonas estables con las características en los segmentos de entrada léxica emparejados. Se compara cada característica en cada zona.

Se determina una penalización de incompatibilidad relativamente grande para una característica en la zona estable si una característica incompatible está presente en el segmento fonológico emparejado de la entrada léxica. Las características incompatibles en el flujo de características clasificadas por zonas y la entrada léxica se muestran en la tabla a continuación.

Característica de flujo de características clasificadas por zonas	Característica de entrada léxica incompatible
[CONS]	[VOC]
[VOC]	[CONS]
[BAJA]	[ALTA]
[ALTA]	[BAJA]
[NASAL]	[ESTRIDENTE]
[ESTRIDENTE]	[NASAL]
[CORONAL]	[CONSONÁNTICA, LABIAL]
[CORONAL]	[DORSAL]
[DORSAL]	[CONSONÁNTICA, LABIAL]
[LABIAL]	[CONSONÁNTICA, DORSAL]
[PLOSIVA]	[NASAL]

Se determina una penalización de no incompatibilidad relativamente pequeña para una característica en una zona estable en donde una característica en la zona estable no está presente en el segmento fonológico emparejado de la entrada léxica o viceversa. Se determina una penalización de emparejamiento para una característica en la zona estable si la misma característica está presente en el segmento fonológico emparejado de la entrada léxica. Preferiblemente, la penalización es dependiente de la fracción de la zona estable en la que la característica está activa.

En este ejemplo, la penalización se calcula de la siguiente manera. Para una característica que se extiende a través de una z proporcional de una zona estable;

a) si la característica de flujo de características clasificadas por zonas está presente en el segmento de entrada

léxica, la penalización se aporta mediante l + z(1 - l) en donde l es la penalización de emparejamiento,

b) si la característica de flujo de características clasificadas por zonas no está presente en el segmento de entrada léxica, la penalización se aporta mediante 1 - z(1 - n), en donde n es la no penalización de incompatibilidad, y

c) si la característica de flujo de características clasificadas por zonas es incompatible con una característica de entrada léxica en el segmento de entrada léxica emparejado, la penalización se aporta mediante 1 - z(1 - m), en donde m es la penalización de incompatibilidad.

Los valores se seleccionan, preferiblemente, de modo que m es mucho menor que n, como par de características incompatibles entre sí indica mucho más una entrada léxica errónea. En el presente ejemplo, n=0,95 y m=0, mientras que l=n. Si una característica está presente en la entrada léxica pero no en el segmento de flujo de características clasificadas por zonas, la penalización se calcula como en el caso (a) para z=0, es decir, la penalización para una característica que está presente en el segmento de entrada léxica pero no en el segmento de flujo de características clasificadas por zonas es la misma que una característica que está presente en el segmento de flujo de características clasificadas por zonas pero no en el segmento de entrada léxica. La variación de las penalizaciones calculadas para z a través del intervalo 0 y 1 se muestra en las figuras 9a a 9c, respectivamente, correspondiente a los casos (a) a (c) anteriores. Las penalizaciones se encuentran en el intervalo 0 y 1, en donde 1 indica un emparejamiento perfecto y 0 representa características completamente incompatibles.

Para una zona inestable, las características en la zona no estable se comparan con las características del segmento léxico alineado con la zona estable inmediatamente a la izquierda de la zona no estable, el segmento léxico alineado con la zona estable inmediatamente a la derecha de la zona no estable, y se compara con cualquiera de los segmentos léxicos que se encuentran entre los dos segmentos ya tenidos en consideración. Las puntuaciones para cada comparación se calculan usando los criterios (a) a (c) anteriores, y se selecciona la menor puntuación. Será evidente que este procedimiento es ventajoso porque no se descarta información potencialmente útil. Una zona no estable puede emparejarse con zonas adyacentes, y considerarse simplemente como una continuación de un segmento estable, o puede emparejarse con un segmento de una entrada léxica que todavía no se ha identificado como emparejado con una zona estable.

Finalmente, se penalizan las incompatibilidades de longitud entre el flujo de características clasificadas por zonas y la entrada léxica. Para cada zona estable extra en el flujo de características clasificadas por zonas en exceso del número de segmentos en la entrada léxica, se añade penalización. Las penalizaciones también se determinan para cada segmento de la entrada léxica con el que no se ha emparejado ninguna característica del flujo de características clasificadas por zona.

El logaritmo de cada penalización se calcula, y en la etapa 75 los valores de logaritmo añadidos para proporcionar una puntuación de penalización final. La puntuación de penalización final es, esencialmente, una medición de la calidad del emparejamiento entre el flujo de características clasificadas por zonas y la entrada léxica en comparación con un emparejamiento perfecto teórico. En la etapa 76, si la entrada léxica no es la entrada final que va a comprobarse, entonces se repiten las etapas de procedimiento 71 a 75.

Tras haber comprobado todas las entradas léxicas apropiadas, entonces en la etapa 77 la entrada léxica con la puntuación más alta se selecciona como el mejor emparejamiento, y se emite en la etapa 78.

El método se ilustra con referencia a un ejemplo particular en las figuras 10 a 12. La palabra hablada es ‘swordfish’, cuyos segmentos se representan usando el alfabeto fonético internacional como /sɔːdfɪʃ/. En cuanto a las características fonológicas, las características de los segmentos de la entrada léxica son de la siguiente manera;

/s/ [CONS, OBS, ESTRID, CONT]

/ɔː/ [VOC, SON, DOR, LAB]

/d/ [CONS, OBS, VOI]

/f/ [CONS, OBS, ESTRID, CONT, LAB]

/ɪ/ [VOC, SON, ALTA, RTR]

/ʃ/ [CONS, OBS, CONT, ALTA, ESTRID]

En la gráfica superior 80 de la figura 10, el espectro evolvente de la señal de voz se muestra, representado en la gráfica como intervalos de tiempo de 20ms. El intervalo de frecuencia se encuentra entre 0 y 5500Hz y las bandas más oscuras son frecuencias de mayor amplitud.

La gráfica inferior 81 de la figura 10 muestra la identificación de características específicas en la señal de voz, tal

como se lleva a cabo en la etapa 12. La presencia de una característica se muestra mediante una línea horizontal, correspondiente a la característica particular enumerada en el eje a mano derecha. Será evidente que no se identifiquen todas las características correspondientes a cada segmento léxico, en particular [VOC] y [ESTRID].

También representado gráficamente en la gráfica inferior de la figura 10 se encuentra un contorno de inestabilidad 82 calculado tal como se describe en la etapa 13. Las zonas estables se muestran mediante bandas grises 83a, 83b, 83c, 83d y las zonas no estables se muestran en bandas blancas 84a, 84b, 84c, 84d. En las zonas no estables 84a y 84c, el valor de inestabilidad muestra un pico puntiagudo y entonces desciende uniformemente, lo que representa un cambio abrupto y entonces una estabilidad relativa. Por el contrario, la zona no estable 84b muestra varios picos, lo que sugiere varios segmentos posibles dentro de la zona.

La banda inferior 85 de la figura 10 muestra la duración real de cada segmento en la señal de voz para comparación.

En la figura 11, el flujo de características clasificadas por zonas se compara con la entrada léxica para 'swordfish'. El algoritmo de mejor emparejamiento empareja /s/, /ɔ/, /ɪ/ y /J/ con las cuatro zonas estables 83a, 83b, 83c, 83d, dejando, obviamente, /df/ sin emparejar. En las zonas estables, será evidente que las características en el flujo de características clasificadas por zonas se extienden por la mayor parte de la zona y corresponden a las características del segmento de entrada léxica correspondiente. A partir del cálculo de penalización (c) anterior, se evaluará una penalización relativamente baja para las zonas estables.

En este ejemplo, se observará que la vocal identificada es /ɔ/ sin la marca de longitud. El método descrito en el presente documento solo usa las características definitorias para identificar un segmento sin referencia a la longitud del segmento. Puede concebirse que la longitud de una zona, con respecto a la longitud promedio de una zona, puede usarse como señal fonética para ayudar a identificar un segmento. Por ejemplo, una zona larga puede indicar un segmento correspondiente a una vocal larga tal como /ɔ:/. Una zona estable excepcionalmente larga también puede identificarse como que corresponde a segmentos idénticos sucesivos.

La figura 12 ilustra la etapa de emparejado de las zonas no estables 84a, 84b, 84c, 84d. Los emparejamientos de zona estable establecidos se muestran mediante flechas de puntos. Las zonas no estables 84a y 84c se comparan con las características de los segmentos de entrada léxica emparejados con las zonas estables adyacentes. En cada caso, las características corresponden a las características de la zona estable a la derecha y generarán una puntuación de penalización baja, lo que sugiere que la zona no estable y la zona estable son parte del mismo segmento.

En el caso de la zona no estable 84b, las características en la zona se comparan con los segmentos de entrada léxica emparejados con zonas estables 83b y 83c tal como se muestra mediante las flechas continuas, y también los segmentos de entrada léxica no emparejados que se encuentran entre los segmentos emparejados con las dos zonas estables. La puntuación de penalización más baja se asocia con el emparejamiento con el segmento /d/.

No se evalúa ninguna penalización para el número de zonas, ya que existen menos zonas estables en el flujo de características segmentado que en la entrada léxica, pero se determinará una penalización para el segmento /f/ ya que no se emparejó con una zona, al haber recibido una peor puntuación que /d/ cuando se empareja con la zona no estable 84b. La identificación final de zonas y segmentos se ilustra en la figura 12.

El método descrito en el presente documento es ventajoso de muchas maneras. Se ha encontrado, inesperadamente, que es robusto incluso cuando el emisor de llamada tiene un impedimento en el habla, un acento regional o es un hablante no nativo de la lengua que se reconoce. El método es ligero a nivel informático, al necesitar solo una fracción de los datos espectrales y acústicos para reconocer una palabra en comparación con métodos de reconocimiento de voz conocidos. El propio diccionario puede ser compacto, al necesitar solo clasificar cada posible palabra o entrada léxica en cuanto a una secuencia corta de pequeños conjuntos de características fonológicas, y, tal como se describió anteriormente, ni siquiera es esencial para identificar cada posible característica fonológica asociada con la lengua objetivo. Será evidente que el método se adapta de manera sencilla a otras lenguas, mediante el ajuste de las características fonológicas identificadas en la etapa 12 y proporcionando un diccionario para esa lengua.

Aunque anteriormente se describen etapas de método particulares, será evidente que muchas de las etapas pueden realizarse en diferente orden, de manera simultánea o de otro modo, según se requiera por una implementación, programa o sistema particular.

En la descripción anterior, una realización es un ejemplo o implementación de la invención. Los diversos aspectos de "una realización", "la realización" o "algunas realizaciones" no necesariamente se refieren en su totalidad a las mismas realizaciones.

Aunque pueden describirse diversos elementos de la invención en el contexto de una única realización, los elementos también pueden proporcionarse en cualquier combinación adecuada que se encuentre dentro del alcance de las reivindicaciones. Por el contrario, aunque la invención puede describirse en el presente documento en el

contexto de realizaciones independientes por motivos de claridad, la invención también puede implementarse en una única realización.

Además, ha de comprenderse que la invención puede llevarse a cabo o llevarse a la práctica de diversas maneras y que la invención puede implementarse en realizaciones diferentes a las señaladas en la descripción anterior.

Los significados de términos técnicos y científicos usados en el presente documento han de comprenderse como lo haría normalmente un experto habitual en la técnica a la que pertenece la invención, a menos que se defina lo contrario.

REIVINDICACIONES

1. Método de reconocimiento de voz automático, comprendiendo el método las etapas de:

recibir una señal de voz (10, 30),

dividir la señal de voz en ventanas de tiempo,

para cada ventana de tiempo;

determinar parámetros acústicos (31 a 36) de la señal de voz dentro de esa ventana, e identificar características fonológicas (51) a partir de los parámetros acústicos, de manera que se genera una secuencia de características fonológicas para la señal de voz,

separar la secuencia de características fonológicas en una secuencia de zonas (13) determinando una puntuación de inestabilidad para cada punto de tiempo en la secuencia de características fonológicas; determinándose la puntuación de inestabilidad comparando las características extraídas en un punto de tiempo con aquellas en puntos de tiempo que preceden el punto de tiempo, de vuelta a un número configurable de milisegundos; comparar las puntuaciones de inestabilidad con un umbral de inestabilidad y una longitud de zona estable mínima para identificar zonas estables y no estables, en el que las zonas no estables se encuentran entre zonas estables (65); y comparar la secuencia de zonas con una entrada léxica que comprende una secuencia de segmentos fonológicos en un diccionario almacenado para identificar una o más palabras en la señal de voz (14, 15, 16); en el que para una entrada léxica que comprende una descripción de una palabra en cuanto a segmentos fonológicos; emparejar las zonas estables con los segmentos fonológicos de la entrada léxica, y para cada zona estable, determinar una penalización para cada característica fonológica dependiendo de las características fonológicas presentes en el segmento fonológico emparejado de la entrada léxica; y

para cada característica fonológica en una zona no estable, determinar una penalización de zona no estable, dependiendo de las características fonológicas presentes en los segmentos fonológicos emparejados de la entrada léxica alineadas con la zona estable en cada lado de la zona no estable, en donde la penalización de zona no estable más baja se selecciona para contribuir a la puntuación de emparejamiento; y

calcular una puntuación de emparejamiento a partir de las penalizaciones determinadas.

2. Método según la reivindicación 1 en el que, si un segmento fonológico de la entrada léxica se encuentra entre los segmentos fonológicos emparejados de la entrada léxica alineada con la zona estable en cada lado de la zona no estable, se determina una penalización de zona no estable para cada característica fonológica en la zona no estable en comparación con ese segmento fonológico.

3. Método según una cualquiera de las reivindicaciones anteriores, en el que cada ventana de tiempo es de 20 ms.

4. Método según una cualquiera de las reivindicaciones anteriores, en el que los parámetros acústicos de la señal de voz dentro de cada ventana de tiempo comprenden uno o más de;

la amplitud cuadrática media,

la frecuencia fundamental de la señal de voz;

la frecuencia de uno o más formantes F1, F2, F3 en la señal de voz;

un espectro de la señal de voz.

5. Método según la reivindicación 4, en el que se calcula un espectro de la señal de voz, comprendiendo además el método determinar;

un valor de inclinación global calculando la pendiente de una línea de regresión sobre la totalidad del espectro,

un primer valor de inclinación calculando la pendiente de una línea de regresión sobre un primer intervalo de frecuencia, y

un segundo valor de inclinación calculando la pendiente de una línea de regresión sobre un segundo intervalo de frecuencia;

opcionalmente en el que el primer intervalo de frecuencia es desde 300 Hz hasta 1500 Hz y el segundo intervalo de frecuencia es desde 1500 Hz hasta 5000 Hz.

6. Método según una cualquiera de las reivindicaciones anteriores, que comprende generar una secuencia de características fonológicas determinando las características de voz activas en cada ventana de tiempo y emitir las características de voz en orden cronológico.

7. Método según una cualquiera de las reivindicaciones anteriores, en el que la etapa de separar la secuencia de características fonológicas en una secuencia de zonas comprende, además;

comparar las puntuaciones de inestabilidad con un umbral de inestabilidad y una longitud de zona estable mínima,

en el que se determinan una secuencia de puntos de tiempo que tienen una longitud mayor que la longitud de zona estable mínima y una puntuación de inestabilidad menor que el umbral de inestabilidad para formar una zona estable, de manera que se considera que las características que se encuentran dentro de la zona estable forman parte del mismo segmento fonológico.

8. Método según la reivindicación 7, en el que la longitud de zona estable mínima es de 30 ms.

9. Método según la reivindicación 7 o la reivindicación 8, en el que la puntuación de inestabilidad para un punto de tiempo se ve aumentada en uno o más de los siguientes:

para cada característica fonológica presente en el punto de tiempo anterior pero no presente en el punto de tiempo

para cada característica fonológica presente en el punto de tiempo, pero no presente en el punto de tiempo anterior; y

donde el punto de tiempo y el punto de tiempo anterior comprenden características fonológicas que forman pares exclusivos de manera mutua.

10. Método según una cualquiera de las reivindicaciones 7 a 9, en el que comparar la secuencia de zonas con entradas léxicas en un diccionario almacenado para identificar una o más palabras en la señal de voz comprende las etapas de;

para una entrada léxica que comprende una descripción de una palabra en cuanto a segmentos fonológicos; emparejar las zonas estables con los segmentos fonológicos de la entrada léxica,

para cada zona estable, determinar una penalización para cada característica fonológica dependiendo de las características fonológicas presentes en el segmento fonológico emparejado de la entrada léxica; y

calcular una puntuación de emparejamiento a partir de las penalizaciones determinadas.

11. Método según la reivindicación 10, en el que no se determina ninguna penalización para una característica fonológica en la zona estable si la misma característica fonológica está presente en el segmento fonológico emparejado de la entrada léxica.

12. Método según la reivindicación 10 o la reivindicación 11, en el que la penalización es dependiente de la fracción de la zona estable en la que está activa la característica fonológica.

13. Método según una cualquiera de las reivindicaciones 10 a 12, que comprende comparar la secuencia clasificada por zonas con una pluralidad de entradas léxicas e identificar una palabra de la entrada léxica con la puntuación de emparejamiento más baja y opcionalmente solo comparar los segmentos fonológicos secuenciales con una entrada léxica si el número de segmentos fonológicos en la entrada léxica está dentro de un intervalo limitado del número de zonas en los segmentos fonológicos secuenciales.

14. Aparato que incluye una base de datos que almacena un diccionario (15), comprendiendo el diccionario una pluralidad de entradas léxicas, comprendiendo cada entrada léxica una descripción de una palabra en cuanto a segmentos fonológicos, estando el aparato adaptado para realizar un método según una cualquiera de las reivindicaciones anteriores.

15. Medio legible por ordenador que contiene instrucciones que cuando se leen por un ordenador provocan que el ordenador proporcione el método según cualquiera de las reivindicaciones 1 a 13.

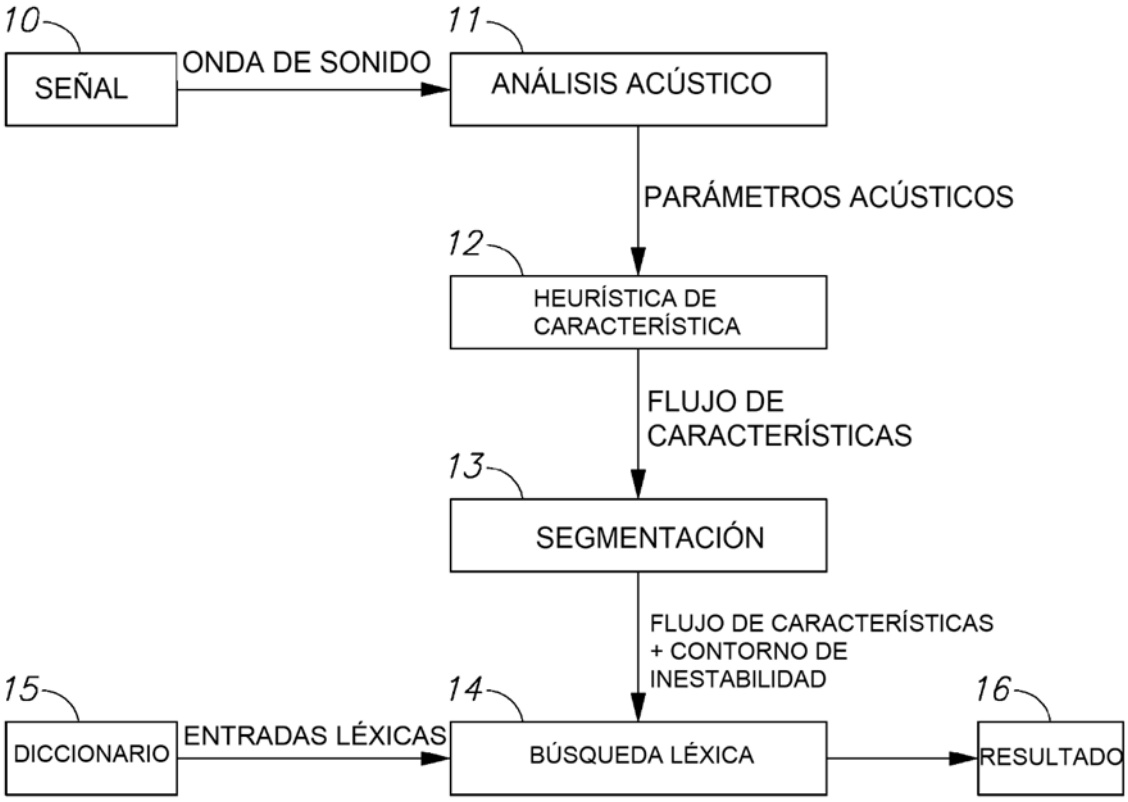
10
SEÑAL
ONDA DE SONIDO
11
ANÁLISIS ACÚSTICO
PARÁMETROS ACÚSTICOS
12
HEURÍSTICA DE CARACTERÍSTICA
FLUJO DE CARACTERÍSTICAS
13
SEGMENTACIÓN
FLUJO DE CARACTERÍSTICAS + CONTORNO DE INESTABILIDAD
15
DICCIONARIO
ENTRADAS LÉXICAS
14
BÚSQUEDA LÉXICA
16
RESULTADO

FIG.1

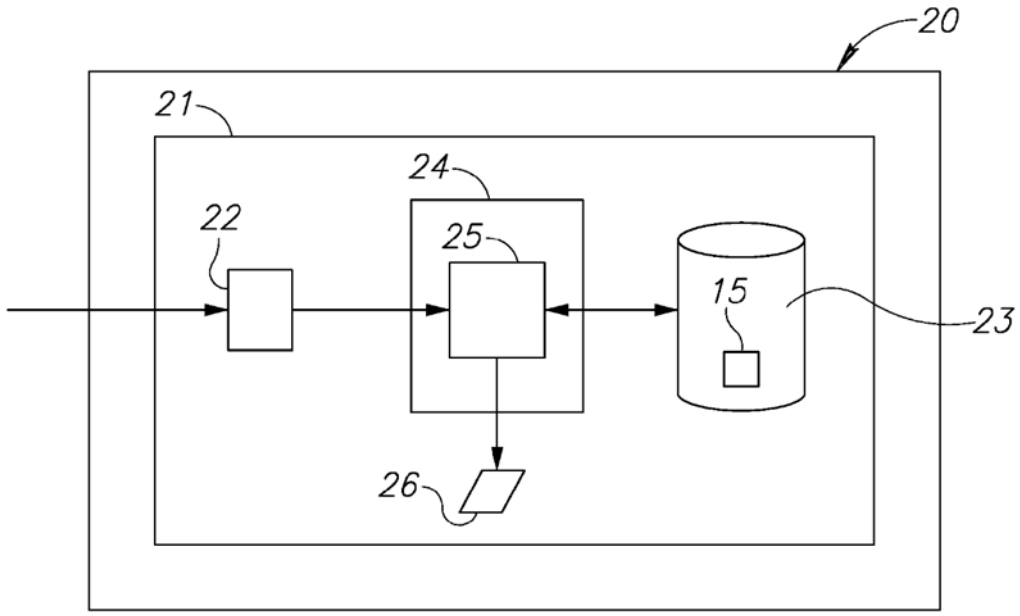
20
21
24
22
25
15
23
26

FIG.2

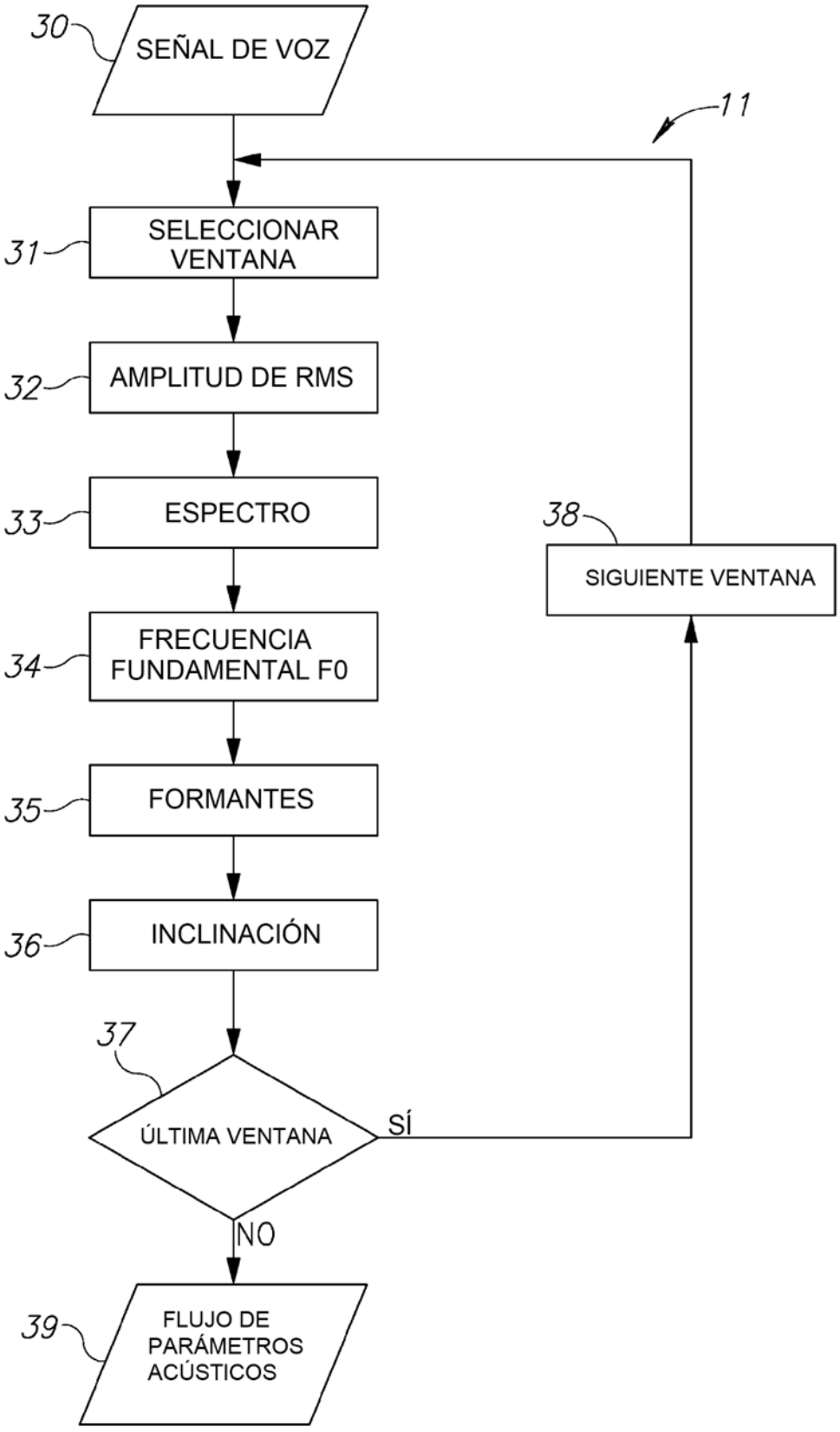
30
SEÑAL DE VOZ
11
31
SELECCIONAR VENTANA
32
AMPLITUD DE RMS
33
ESPECTRO
34
FRECUENCIA FUNDAMENTAL F0
35
FORMANTES
36
INCLINACIÓN
37
ÚLTIMA VENTANA
SÍ
NO
38
SIGUIENTE VENTANA
39
FLUJO DE PARÁMETROS ACÚSTICOS

FIG.3

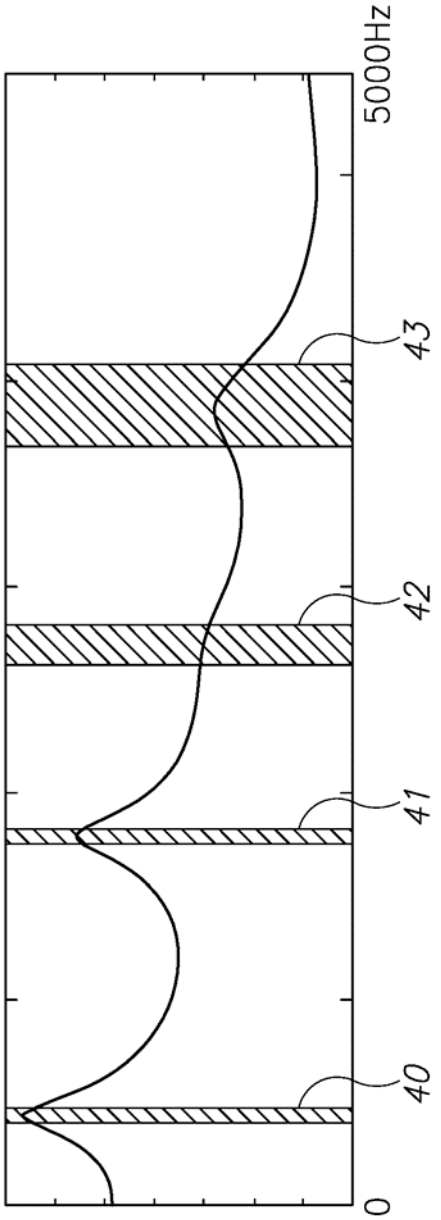

FIG.4

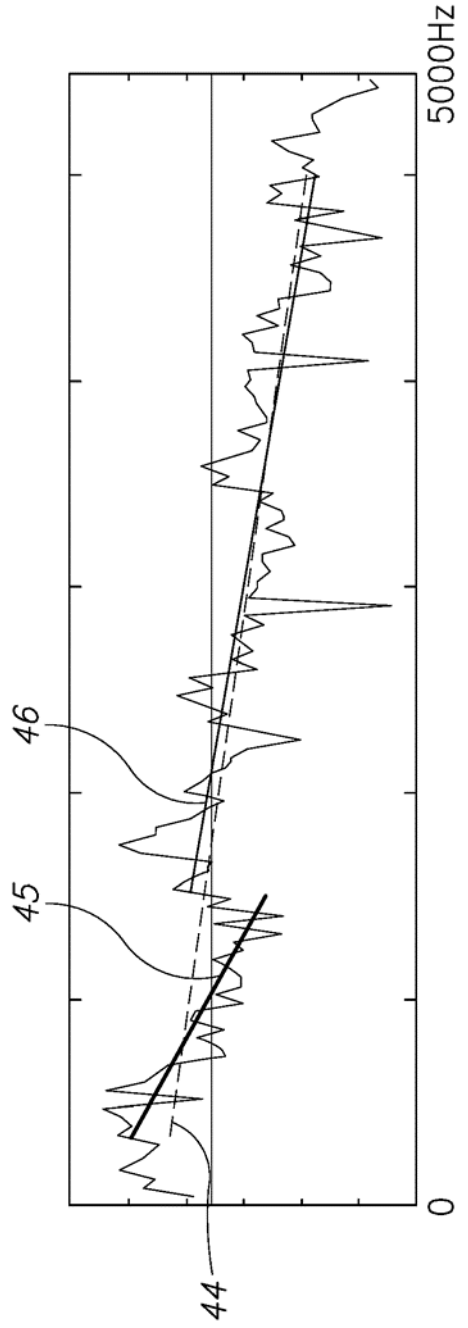

FIG.5

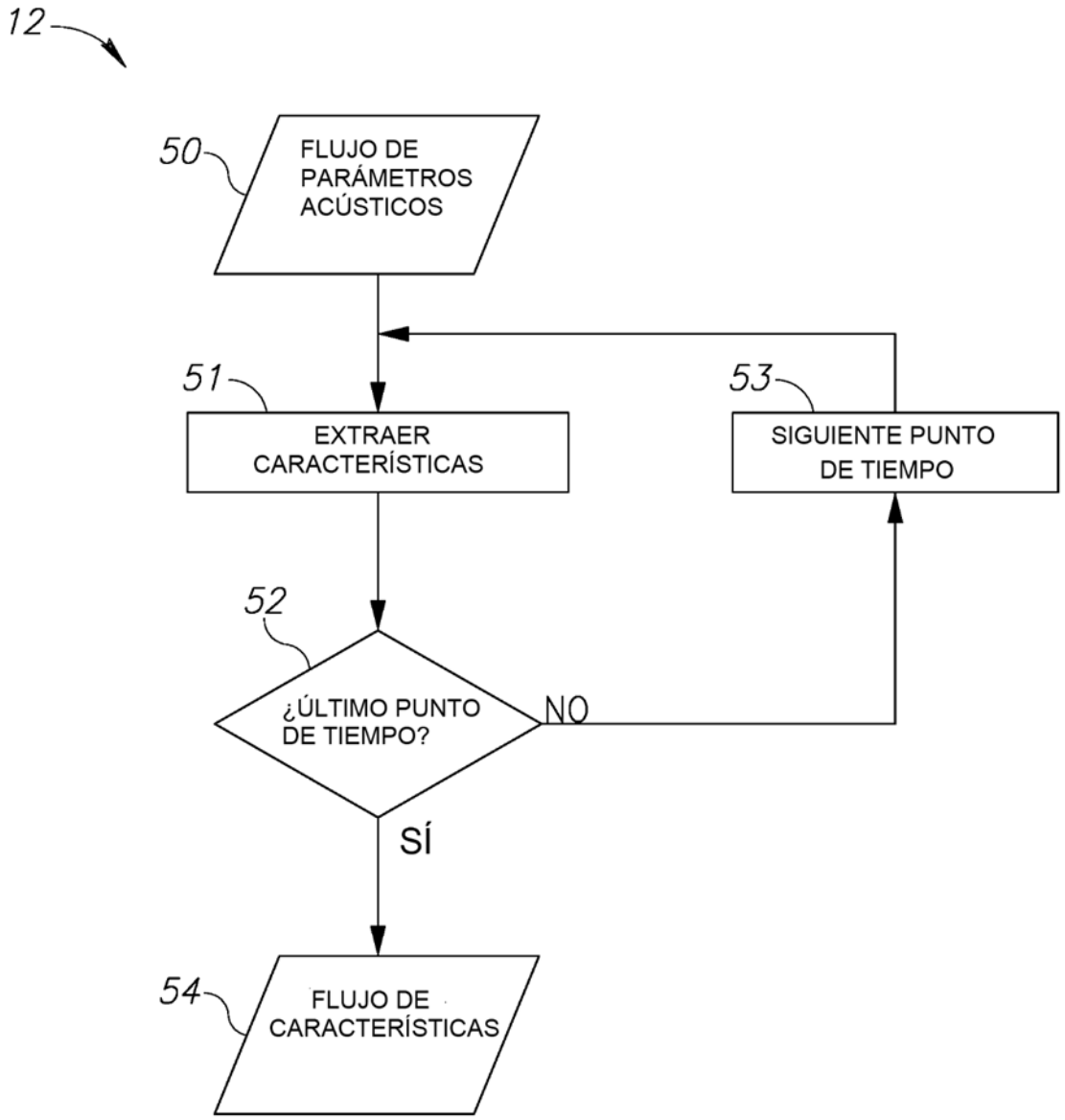
12
50
FLUJO DE PARÁMETROS ACÚSTICOS
51
EXTRAER CARACTERÍSTICAS
53
SIGUIENTE PUNTO DE TIEMPO
52
¿ÚLTIMO PUNTO DE TIEMPO?
NO
SÍ
54
FLUJO DE CARACTERÍSTICAS

FIG.6

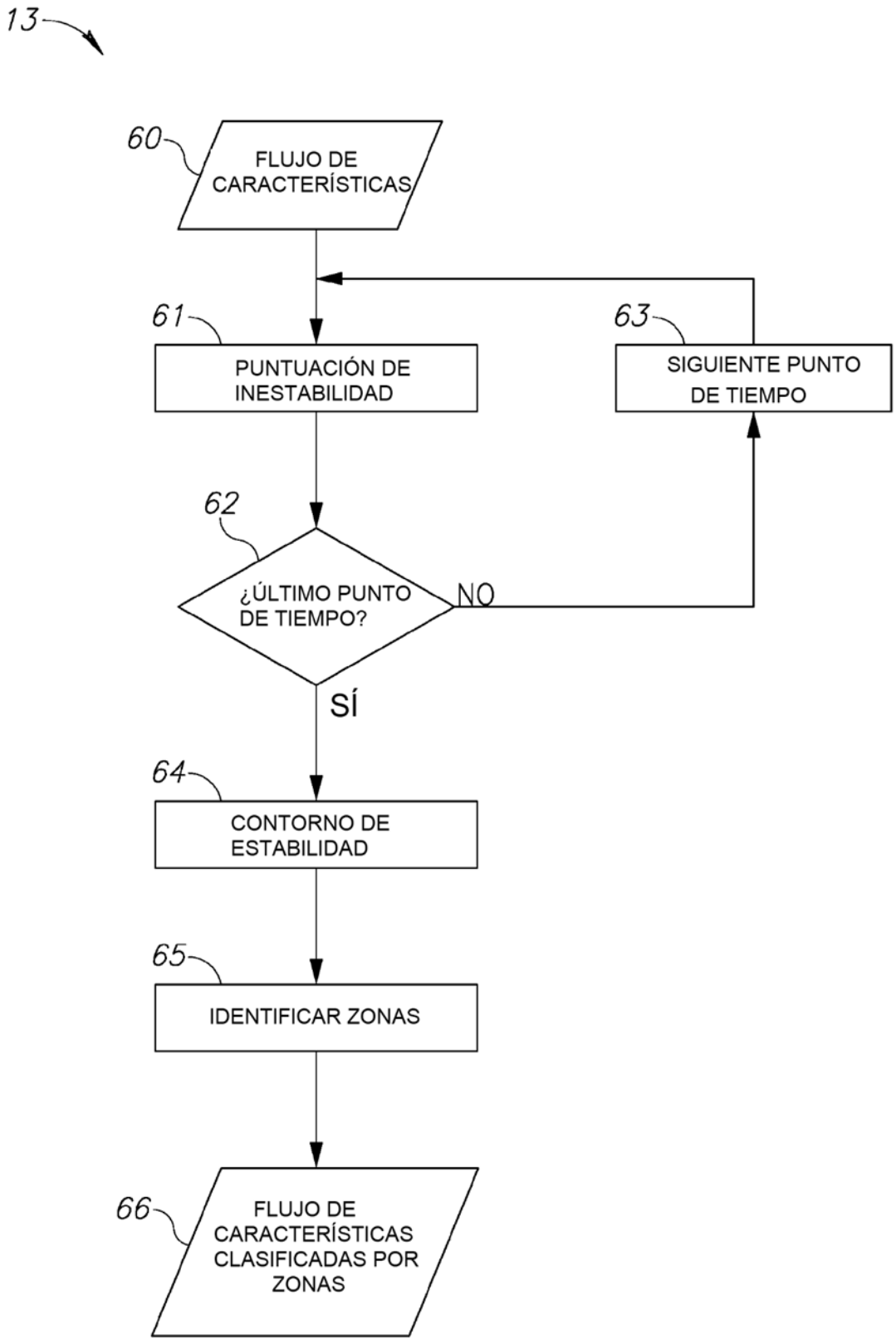
13
60
FLUJO DE CARACTERÍSTICAS
61
PUNTUACIÓN DE INESTABILIDAD
63
SIGUIENTE PUNTO DE TIEMPO
62
¿ÚLTIMO PUNTO DE TIEMPO?
NO
SÍ
64
CONTORNO DE ESTABILIDAD
65
IDENTIFICAR ZONAS
66
FLUJO DE CARACTERÍSTICAS CLASIFICADAS POR ZONAS

FIG.7

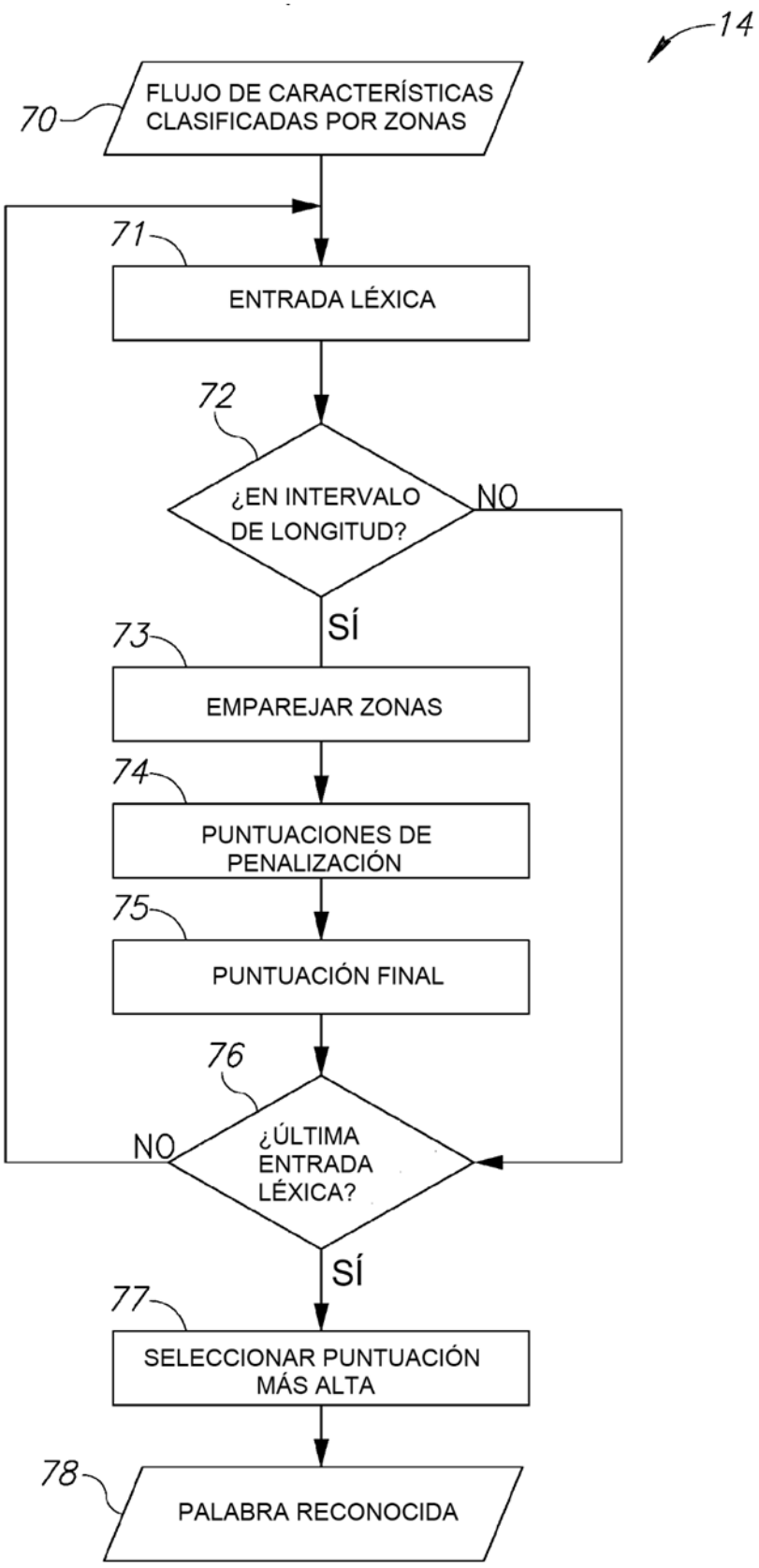
14
70
FLUJO DE CARACTERÍSTICAS CLASIFICADAS POR ZONAS
71
ENTRADA LÉXICA
72
¿EN INTERVALO DE LONGITUD?
NO
SÍ
73
EMPAREJAR ZONAS
74
PUNTUACIONES DE PENALIZACIÓN
75
PUNTUACIÓN FINAL
76
¿ÚLTIMA ENTRADA LÉXICA?
NO
SÍ
77
SELECCIONAR PUNTUACIÓN MÁS ALTA
78
PALABRA RECONOCIDA

FIG.8

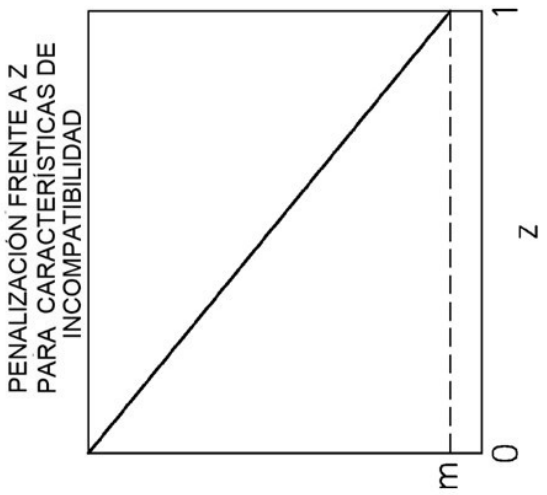

FIG.9C

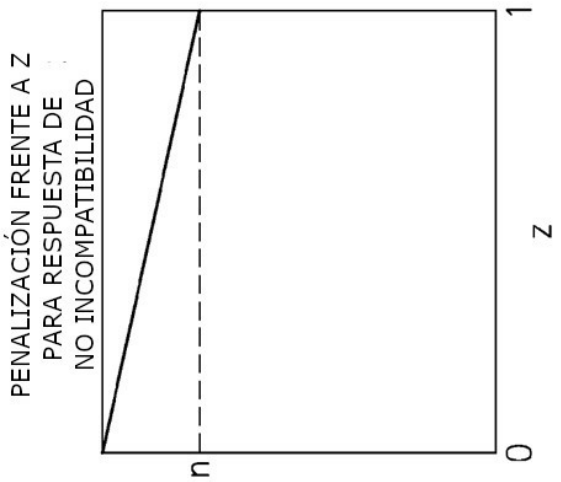

FIG.9B

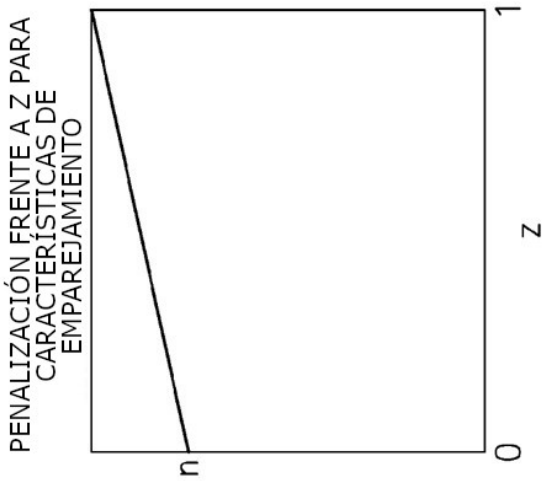

FIG.9A

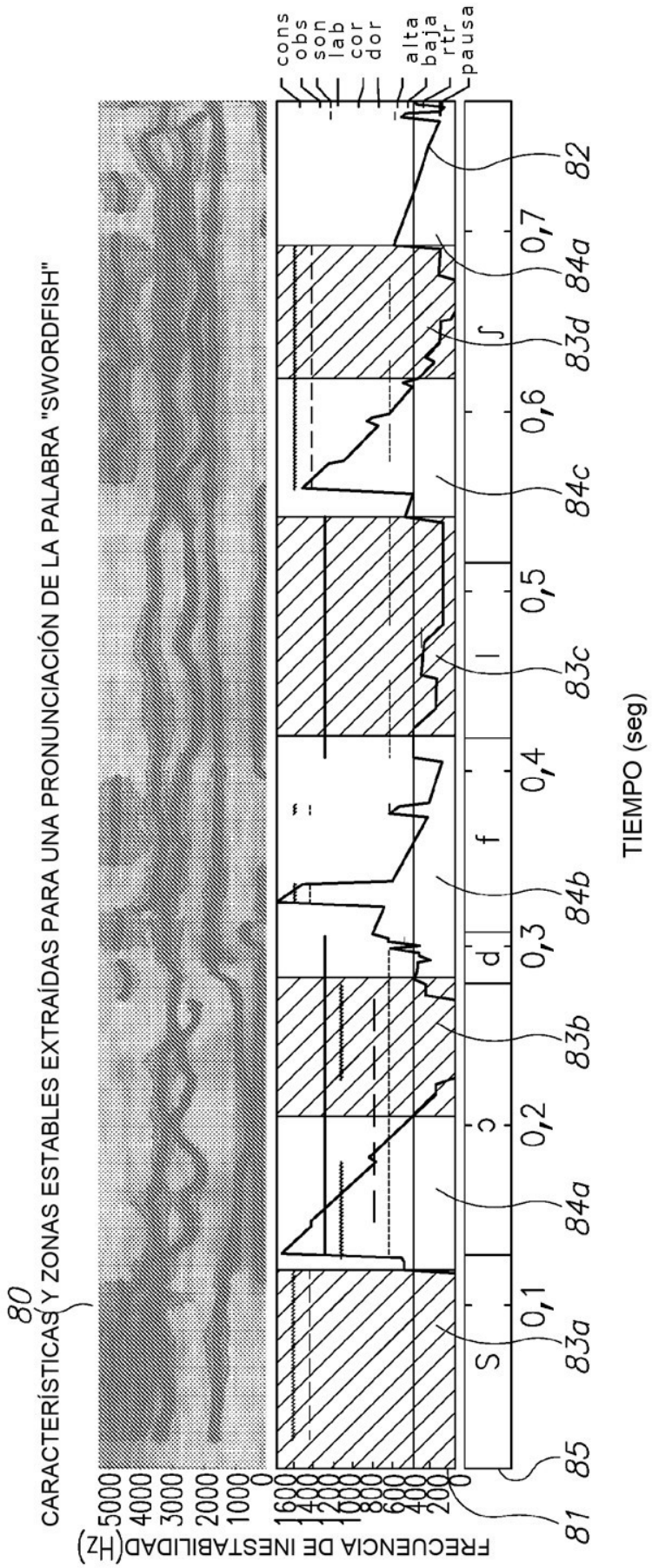
80
CARACTERÍSTICAS Y ZONAS ESTABLES EXTRAÍDAS PARA UNA PRONUNCIACIÓN DE LA PALABRA "SWORDFISH"
FRECUENCIA DE INESTABILIDAD (Hz)
5000
4000
3000
2000
1000
0
1600
1400
1200
1000
800
600
400
200
0
cons
obs
son
lab
cor
dor
alta
baja
rtr
pausa
S
c
d
f
l
ʃ
0,1
0,2
0,3
0,4
0,5
0,6
0,7
TIEMPO (seg)
81
82
83a
83b
83c
83d
84a
84b
84c
84a
85

FIG.10

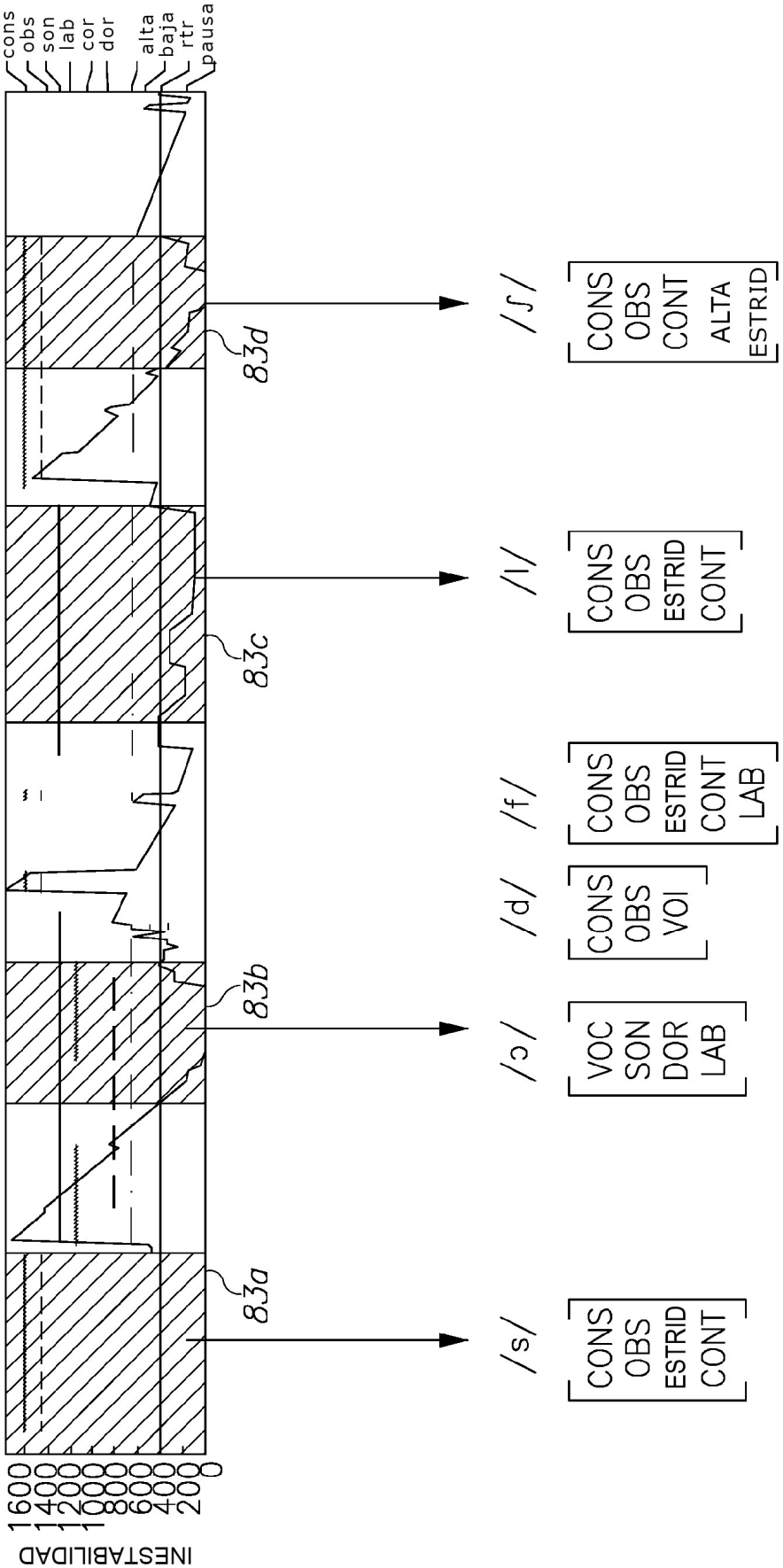
cons
obs
son
lab
cor
dor
alta
baja
rtr
pausa
INESTABILIDAD
1600
1400
1200
1000
800
600
400
200
83a
83b
83c
83d
/s/
CONS
OBS
ESTRID
CONT
/ɔ/
VOC
SON
DOR
LAB
/d/
CONS
OBS
VOI
/f/
CONS
OBS
ESTRID
CONT
LAB
/l/
CONS
OBS
ESTRID
CONT
/ʃ/
CONS
OBS
CONT
ALTA
ESTRID

FIG.11

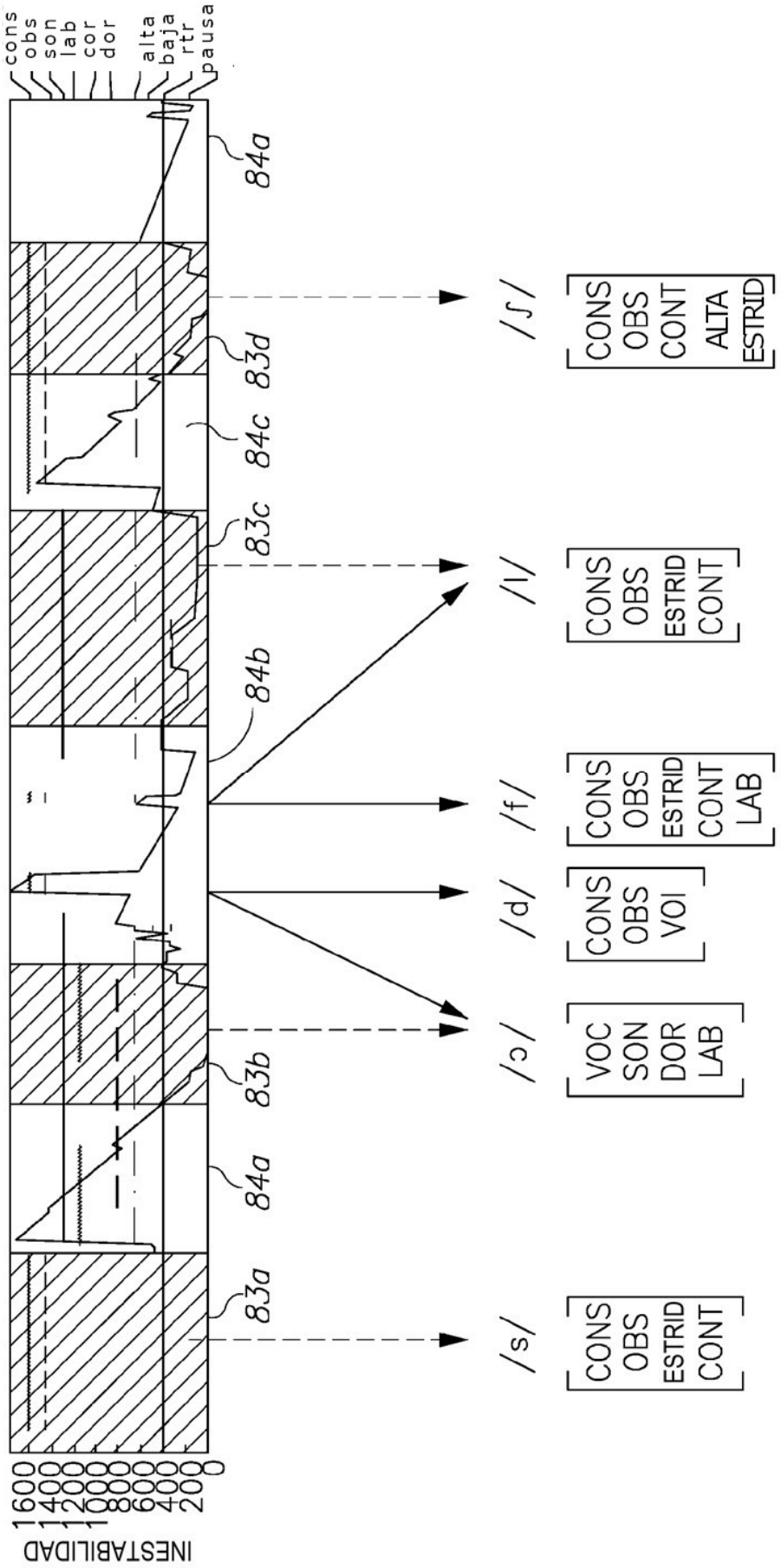
cons
obs
son
lab
cor
dor
alta
baja
rtr
pausa
84a
83d
84c
83c
84b
83b
84a
83a
/s/
/ɔ/
/d/
/f/
/l/
/ʃ/
CONS OBS ESTRID CONT
VOC SON DOR LAB
CONS OBS VOI
CONS OBS ESTRID CONT LAB
CONS OBS ESTRID CONT
CONS OBS CONT ALTA ESTRID
1600
1400
1200
1000
800
600
400
200
INESTABILIDAD

FIG.12