

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 364 401**

21 Número de solicitud: 201131069

51 Int. Cl.:

G10L 11/00 (2006.01)

G10L 15/00 (2006.01)

G10L 17/00 (2006.01)

G10L 19/04 (2006.01)

12

PATENTE DE INVENCION CON EXAMEN PREVIO

B2

22 Fecha de presentación: **27.06.2011**

43 Fecha de publicación de la solicitud: **01.09.2011**

Fecha de la concesión: **13.12.2011**

45 Fecha de anuncio de la concesión: **23.12.2011**

45 Fecha de publicación del folleto de la patente:
23.12.2011

73 Titular/es:
**UNIVERSIDAD POLITÉCNICA DE MADRID
C/ RAMIRO DE MAEZTU, 7
28040 MADRID, ES**

72 Inventor/es:
**GÓMEZ VILDA, PEDRO;
RODELLAR BIARGE, VICTORIA;
NIETO LLUIS, VICTOR;
ÁLVAREZ MARQUINA, AGUSTÍN y
MARTÍNEZ OLALLA, RAFAEL**

74 Agente: **Carvajal y Urquijo, Isabel**

54 Título: **MÉTODO Y SISTEMA PARA LA ESTIMACIÓN DE PARÁMETROS FISIOLÓGICOS DE LA FONACIÓN.**

57 Resumen:

La invención consiste en un método y sistema de cómputo para el registro y análisis de la voz, que permite calcular una serie de parámetros de la fonación. Estos transportan información relevante sobre influencias causadas por trastornos orgánicos (que afectan a la fisiología de la laringe) o neurológicos (que afectan a los centros cerebrales del habla). Asimismo se consideran parte esencial de la invención los procedimientos clasificatorios que permiten obtener estimaciones de la disfunción presente y de asignación de personalidad. La utilidad de la invención se enmarca, en la posibilidad de aplicar la estimación de disfunción en los centros médicos de asistencia primaria para el cribado de pacientes a los centros de atención especializada, simplificando los protocolos de exploración, ahorrando costes, y reduciendo listas de espera. También es aplicable esta metodología en la detección de la personalidad del locutor por la voz, permitiendo garantizar el acceso a instalaciones o servicios.

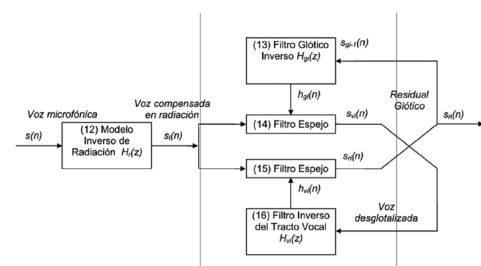


Fig. 2

DESCRIPCIÓN

Método y sistema para la estimación de parámetros fisiológicos de la fonación.

5 Sector técnico

La invención se enmarca en el sector de las tecnologías de la información y las comunicaciones con aplicación en biomedicina y la seguridad, y más concretamente en el campo de la detección y la gradación de la patología orgánica de la voz por medio de la clasificación de parámetros obtenidos de la onda glótica de la voz, y en la gestión de acceso seguro por voz.

Antecedentes de la invención

La medida de la calidad de voz como procedimiento para el diagnóstico y gradación de la patología orgánica de la voz ha conocido un auge importante en la última década. Fruto del mismo son un conjunto de aplicaciones informáticas que a partir de la voz generan índices de medida de la calidad de la misma como variantes del *jitter* (perturbación del período de fonación a lo largo del tiempo), del *shimmer* (perturbación temporal de la amplitud de la fonación ciclo a ciclo), de la relación señal/ruido (entre la parte periódica y la no periódica de un segmento de voz), del índice *glottal/noise* (proporción entre la energía de la onda glótica respecto al ruido residual presente en la voz), y de parámetros temporales que reflejan los procesos de apertura y cierre de los pliegues vocales durante una fonación, como son los ciclos de *recuperación*, *cierre*, *apertura* y *cerramiento*. Los procesos de estimación paramétrica se suelen llevar a cabo sobre la voz medida en el punto de captura de la misma, generalmente un micrófono de propósito general, para ser digitalizada y posteriormente procesada. Los procesos habituales son la extracción en el dominio espectral o temporal. Entre los primeros se determina su densidad espectral de potencia, y a partir de la misma los parámetros mel-cepstrum, sus primeras y segundas diferencias. También se miden por procedimientos relacionados la relación de energía armónico/ruido. La parametrización temporal parte de la reconstrucción de la fuente glótica, sobre la que se mide la duración del ciclo de fonación (duración en tiempo entre dos cierres consecutivos de los pliegues vocales), a partir de la cual se derivan los instantes de recuperación, apertura y cierre, y en base a los mismos se determinan las relaciones *glottal/noise* y las pendientes del pulso glótico.

La metodología básica que soporta la invención es la estimación precisa de la onda glótica, entendida ésta como el correlato de presión sonora que se desarrolla en la glotis durante la fonación. Los trabajos en este terreno se inician hacia principio de los años noventa, con las contribuciones de Paavo Alku y sus colegas sobre la inversión de la trama de voz para la reconstrucción del pulso glótico (Alku, P., "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering"), *Speech Communication*, vol. 11, 1992, pp. 109-118. Por lo que respecta a la estimación combinada de la onda glótica y el tracto vocal se pueden citar asimismo los trabajos de P. Murphy y su equipo (Akande, O., and Murphy, P. J. "Estimation of the vocal tract transfer function for voiced speech with application to glottal wave analysis", *Speech Communication*, Vol. 46, 2005, pp. 15-36). Por lo que hace referencia en la solicitud a la aplicación de la parametrización de la onda glótica en descripción biométrica del locutor cabe citar los trabajos clásicos de Reynolds y su grupo (Reynolds, D. A., Quatieri, T. F., Dunn, R. B., "Speaker Verification Using Adapted Gaussian Mixture Models", *Digital Signal Processing*, Vol. 10, 2000, pp. 19-41). Una buena revisión de este tipo de tecnologías de verificación identitaria por la voz se puede encontrar en el trabajo clásico de Bimbot *et. al.* (Bimbot, F., Bonastre, J. F., Fredouille, C., Gravier, G., Magrin, I., Meignier, S., Merlin, T., Ortega, J., Petrovska, D., Reynolds, A., "A Tutorial on Text-Independent Speaker Verification", *EURASIP J. on App. Sig. Proc.*, Vol. 4, 2004, pp. 430-451).

En el ámbito de la detección de la disfonía existen soluciones basadas en el análisis no diferenciado del habla, como son los sistemas:

- CSL (Computerized Speech Lab), MDVP (MultiDimensional Voice Program) y APM (Ambulatory Phonation Monitor) de Kay Elemetrics (KayPENTAX), URL: <http://www.kayelemetrics.com> (20.4.2011)
- MEDIVOZ y WPCVox de TGH ENDOSCOPIA, URL: <http://www.tghendoscopia.com/> (20.4.2011)
- Dr. Speech, URL: <http://www.drspeech.com/> (20.4.2011)
- SESANE (Software Environment for Speech Analysis and Evaluation) de SQLab, URL: <http://www.sqlab.fr/sesaneUK.htm> (20.4.2011)
- LingWaves de WEVOSYS, URL: <http://www.wevosys.com/> (20.4.2011)
- Speech Studio, de Laryngograph, URL: <http://www.laryngograph.com/> (20.4.2011)
- WaveView Software, de Glottal Enterprises, URL: <http://www.glottal.com/>
- Otras aplicaciones remotamente relacionadas con el procesado de la voz, ya que han sido desarrolladas para el tratamiento del habla son Wavesurfer (URL: <http://www.speech.kth.se/wavesurfer/>), y Praat (URL: www.praat.org).

Estas soluciones abordan el estudio de la voz por medio de análisis acústicos clásicos que tecnológicamente suponen un estado del arte anterior al propuesto en la presente solicitud. La tecnología que se pretende patentar contempla la voz como resultado de dos procesos: producción de la excitación laríngea u onda glótica en los pliegues vocales, y filtrado articulatorio por el tracto vocal, que es el recinto acústico formado por las cavidades faríngea, nasal y oral.

El segundo de los procesos es muy variable incluso para un mismo locutor, ya que está influido por el mensaje, y es más fácilmente impostable. El primero de los procesos, la producción, es menos variable para un mismo locutor, difícilmente impostable, y arrastra información del estado neurológico, emocional y fisiológico del aparato productor. Estos rasgos dotan de gran ventaja a la propuesta de la solicitud en el ámbito biométrico en general, y en particular en los ámbitos clínico y forense. Para ello se utilizan procedimientos de tratamiento de señal y reconocimiento de patrones que son la base de la solicitud que se propone.

En el ámbito de la detección de la patología neuromotora del habla se tiene conocimiento de un sistema Kiosk, para detección precoz del mal de Alzheimer, si bien tecnológicamente no es un producto avanzado, ya que solo se utiliza para registro, sin introducir análisis acústico.

En el ámbito de la identificación y verificación del locutor y en la asignación de personalidad existen los sistemas:

- VocalPassword y similares, de Persay, URL: <http://www.persay.com/> (20.4.2011)
- Verifier, de Authentify, URL: <http://www.authentify.com/> (20.4.2011)
- ASIS, KIVOX, BS y BatVox, de Agnitio, URL: <http://www.agnitio.es/> (20.4.2011)
- SecuriVox de SpeechSentinel, URL: <http://www.speechsentinel.co.uk/> (20.4.2011)
- BioVox de DTEC, URL: <http://www.dtec-bio.es/> (20.4.2011)

La característica diferencial de la solución propuesta respecto a todos estos sistemas se basa en el uso de segmentos del habla que son exclusivamente fonados, es decir, que cursan con producción de voz, en la reconstrucción de un correlato de fonación que está íntimamente ligado a la biomecánica de la fonación, y en la parametrización de dicho correlato en los ámbitos biomecánico y biométrico mediante la reconstrucción del sistema productor más aproximado al modelo laríngeo capaz de generar dicho correlato. Este proceso de aproximación al modelado del sistema fonador es más introspectivo que cualquier otro procedimiento de análisis existente, porque modela las estructuras biomecánicas responsables últimas de la producción de la voz que son los pliegues o cuerdas vocales. En la revisión del estado del arte, y en las presentaciones de tecnologías del habla a las que se ha asistido en el ámbito nacional e internacional, así como en la revisión de publicaciones especializadas y bases de datos de patentes no se ha encontrado ninguna referencia que siga la misma orientación que la propuesta. Otra de las ventajas competitivas de la presente propuesta es que se puede personalizar fácilmente para ofrecer diferentes soluciones, desde el profesional autónomo clínico o forense, a los servicios de asistencia primaria o especializada, a los cuerpos y fuerzas de seguridad, al ámbito de la seguridad privada, al acceso seguro a instalaciones físicas y servicios informáticos, a la gestión de Customer Service Optimization, etc.

Por lo que hace referencia a patentes que traten temas afines, tras la consulta de las bases de datos correspondientes se pueden citar las siguientes, y su relación con la solicitud:

- European Patent Application EP 2 124 223 A1:

METHOD AND SYSTEM FOR DIAGNOSING PATHOLOGICAL PHENOMENON USING A VOICE SIGNAL

El objetivo de la patente referida pretende ser la detección de patologías de tipo psicoacústico o de desbalance bioquímico que puedan ser establecidas mediante el análisis del habla del paciente, como entre las primeras la dislexia, desorden de déficit de atención, desorden de déficit de atención por hiperactividad, autismo, Parkinson, Alzheimer, déficit de percepción sensorial, problemas de escucha, depresión, control motor y letargia, y entre las segundas la condición cardiopulmonar, la diabetes juvenil, la deficiencia de dopamina y serotonina, el exceso de norepinefrina, testosterona, serotonina y acetilcolina o su regulación, las patologías de zona sacra y genital, y los problemas de sistema inmunológico. En cuanto a materiales, la citada patente propugna el uso de la voz (voice) o del habla (speech) sin hacer distinción entre ambos conceptos, si bien se desprende que se refiere al habla del paciente cuando dice “wherein the speech has a finite duration and each time period separating the respective plurality of sample intensity values is essentially evenly distributed within the duration of the speech” (claim 12, col. 13, l. 21-26). En cuanto a la metodología, de lo manifestado en la descripción técnica se desprende que se propone el análisis espectral de la señal de habla del paciente cuando dice “For the purposes of describing and claiming the present invention, the term “crater feature” is intended to refer to a shape (on a graph of frequency vs. intensity) which manifests a sharp drop at a first frequency continued by a relatively low level along approximately 50 Hz or more and then a relatively steep rise at a second frequency”. (FIELD OF THE INVENTION, col. 1, l. 18-24). No se muestra en esta patente un proceso de clasificación de patrones o de modelado de una base de datos de sujetos normales y/o patológicos que haga referencia a la detección de cada patología, por lo que se asume que dicho proceso de detección se basa en el marcado de unos rasgos, sin que exista un mecanismo de medición y validación del procedimiento. La propuesta en la presente solicitud tiene como ventajas frente a la patente referida el contemplar la detección de la patología orgánica del aparato fonador

incluyendo alteraciones de los pliegues vocales, como pólipos, nódulos, edemas, carcinomas de pliegue, parálisis de pliegue, etc. y de perturbaciones producidas por deterioro de los centros superiores o neuromotores que afectan exclusivamente a la laringe. Además la propuesta de la presente solicitud propugna el uso de la voz como marca biométrica para acceso seguro y cotejo forense. Todos estos objetivos son radicalmente diferenciadores de los reclamados en la patente referida, añadiendo valor en su capacidad de detección, su robustez y su precisión. En cuanto a materiales, la solicitud propugna la utilización de los correlatos de vibración de los pliegues vocales, como la excitación glótica, la onda mucosa o el residual glótico, que deben ser extraídos por filtrado inverso de los segmentos fonados de la voz, y preferentemente de vocales sostenidas. En cuanto a la metodología utilizada en la presente solicitud, a partir de la onda glótica obtenida por filtrado inverso de los segmentos fonados (inversión de la señal hablada) se calculan parámetros cepstrales, singularidades de la densidad espectral de potencia, y parámetros biomecánicos que se extraen para cada ciclo de fonación en segmentos de unos 200 msg. de fonación, lo cual supone muestras tomadas a razón de unas 100-200 por segundo. También se utilizan parámetros temporales en la presente solicitud obtenidos para cada ciclo de fonación mediante transformadas wavelet. En la solicitud propuesta se plantean diferentes modelos de clasificación de patrones, de normado de pacientes y de análisis y validación estadística de resultados. También se muestran procedimientos para la gradación de la disfonía, del acceso seguro y del cotejo forense. La nueva solución propuesta en la presente solicitud mejora lo expuesto en la patente referida en la obtención de onda glótica precisos en lugar de voz completa, en la estimación de un conjunto de parámetros con alta semántica funcional en disfonía, y en la robustez y precisión de las estimaciones frente a la variabilidad intra-locutor, mejorando la capacidad de separación inter-locutor. Estas innovaciones hacen que las diferencias entre lo manifestado en dicha patente y la presente solicitud sean sustanciales en lo que hace referencia a objetivos, materiales y metodología utilizados, siendo la nueva solución propuesta claramente más ventajosa en cuanto a ofrecer más parámetros con mejor semántica, menor variabilidad intra-locutor, y mayor precisión y robustez.

- US Patent Number US005400434A

VOICE SOURCE FOR SYNTHETIC SPEECH SYSTEM

El objetivo de la patente referida pretende ser el uso de la señal de excitación glótica previamente extraída a partir de un conjunto de locutores de referencia, para ser reproducida en sistemas de síntesis texto-voz. La metodología utiliza filtrado inverso para generar la onda glótica, sin especificar el tipo de filtrado inverso a utilizar. El material utilizado es habla registrada a partir de una base de locutores, de la cual se extrae el pulso glótico. La presente solicitud utiliza un tipo de filtrado inverso basado en celosías modelo-espejo, que son una innovación en sí mismas. Estos filtros celosía son estándares en la estimación conjunta de procesos. La nueva solución propuesta en la presente solicitud mejora lo expuesto en la patente referida en la obtención de onda glótica en lugar de voz completa, en la estimación de un conjunto de parámetros con alta semántica funcional en disfonía, y en la robustez y precisión de las estimaciones frente a la variabilidad intra-locutor, mejorando la capacidad de separación inter-locutor. Estas innovaciones hacen que las diferencias entre lo manifestado en dicha patente y la presente solicitud sean sustanciales en lo que hace referencia a objetivos, materiales y metodología utilizados, siendo la nueva solución propuesta claramente más ventajosa en cuanto a ofrecer más parámetros con mejor semántica, menor variabilidad intra-locutor, y mayor precisión y robustez.

- US Patent Number US005577160A

SPEECH ANALYSIS APPARATUS FOR EXTRACTING GLOTTAL SOURCE PARAMETERS AND FORMANT PARAMETERS

El objetivo de la patente referida pretende ser la reconstrucción de la fuente glótica junto con la función de transferencia del tracto vocal, combinando algoritmos de análisis basados en predicción lineal. Para ello utilizan conocimiento de dominio público, fundamentalmente publicado por P. Alku y otros (OTHER PUBLICATIONS). La patente se basa en la integración de diferentes métodos de análisis espectral de dominio público, siguiendo estructuras de conexión de dichos métodos, que permiten a los autores producir estimaciones de determinados parámetros de la fuente glótica (SOURCE PARAMETER EXTRACTING MEANS: Fundamental Frequency F0, Amplitude of waveform B, Open Quotient OQ, SK, C, D), así como del tracto vocal (FIRST TO SIXTH FORMANT: F1-6). Los parámetros así extraídos se combinan en un modelo espectral de la derivada del flujo glótico, y en un modelo del tracto vocal, que juntos definen un modelo completo del habla (fonada) que se desee (Fig. 27 y Fig. 28 del documento referido). La metodología que se utiliza es el filtrado LPC (Linear Predictive Coding) clásico mediante predictores transversales, la estimación de los polos y ceros de los mismos, y su uso en la eliminación de la influencia del tracto vocal en pasos escalonados siguiendo el modelo AIF (Adaptive Inverse Filtering) de P. Alku (de dominio público) para generar la fuente glótica, y mediante Transformada Rápida de Fourier (de dominio público) generar dos modelos en el dominio de la frecuencia, de fuente glótica y de transferencia de tracto vocal, que combinados e invertidos al dominio del tiempo den una descripción del habla (fonada) analizada y sintetizada. El procedimiento consiste en controlar cuidadosamente el número de formantes detectados en el modelado de la fuente glótica y del tracto vocal para evitar la intrusión de componentes del tracto vocal en la fuente glótica durante el modelado LPC de ésta. Se pretende conseguir estimaciones más precisas de ambas componentes respecto al estado del arte. Para ello se propone utilizar un sistema denominado AbS (Analysis by Synthesis) para modelar la fuente glótica, en lugar del clásico LPC, que se reserva para el modelado del tracto vocal. El proceso de modelado de la fuente glótica por AbS pasa por eliminar el primer

formante de la voz según una pluralidad de candidatos, generando una pluralidad de fuentes glóticas al eliminar los diferentes candidatos a primer formante. Estas fuentes se combinan con la estimación del tracto vocal para sintetizar voz, que se compara con la original y permite seleccionar el candidato más adecuado. La originalidad de esta metodología está en la detección y eliminación selectiva de los formantes del tracto vocal (el primero, y los superiores de forma diferenciada) para sintetizar un prototipo de fuente glótica que mejor se ajuste al perfil reducido de la voz por eliminación selectiva de formantes. Para ello se utilizan las estimaciones de los parámetros F0, OQ, SK, C y D, así como los formantes F1-6. Por su parte, la presente solicitud propugna la utilización del modelo AIF con la originalidad de realizar las estimaciones cruzadas de onda glótica y tracto vocal mediante filtros LPC implementados mediante celosías modelo-espejo según se muestra en las Figuras 2 y 5, controlando de forma empírica los órdenes (número de etapas) de dichas celosías. De esta forma, la solución propuesta en la presente solicitud es respetuosa con los patrones biométricos y biomecánicos que aparecen en la fuente glótica, y que no son respetados por la patente citada, por lo que mejoran sustancialmente la capacidad de caracterización de aquella. La nueva solución propuesta en la presente solicitud mejora lo expuesto en la patente referida en la obtención de onda glótica precisos en lugar de voz completa, en la estimación de un conjunto de parámetros con alta semántica funcional en disfonía, y en la robustez y precisión de las estimaciones frente a la variabilidad intra-locutor, mejorando la capacidad de separación inter-locutor. Estas innovaciones hacen que las diferencias entre lo manifestado en dicha patente y la presente solicitud sean sustanciales en lo que hace referencia a objetivos, materiales y metodología utilizados, siendo la nueva solución propuesta claramente más ventajosa en cuanto a ofrecer más parámetros con mejor semántica, menor variabilidad intra-locutor, y mayor precisión y robustez.

• US Patent Number US007398213B1

METHOD AND SYSTEM FOR DIAGNOSING PATHOLOGICAL PHENOMENON USING A VOICE SIGNAL

Se trata de la extensión como patente en Estados Unidos de la ya comentada patente europea European Patent Application EP 2 124 223 A1, por lo que son de aplicación a ésta las mismas consideraciones que a aquella, sin más añadidos.

• US Patent Number US 20050171774A1

FEATURES AND TECHNIQUES FOR SPEAKER AUTHENTICATION

El objetivo de la patente referida parece ser la utilización de parámetros extraídos de la fuente glótica, de los formantes, de las características temporales y de la frecuencia fundamental del habla como elementos para el reconocimiento de locutores. En cuanto a los parámetros de la fuente glótica, se citan los siguientes: la amplitud de pico, la amplitud RMS (valor eficaz), la tasa de cruces por cero, la función de autocorrelación, la longitud del arco, los coeficientes de Fourier, la trayectoria en el plano complejo de la Transformada Discreta de Fourier, la tasa de caída con la frecuencia (spectral tilt), las relaciones de amplitud y fase de los primeros armónicos, el grado de aire en voz (ruido de aspiración, coeficiente OQ (open quotient) alto, la componente de ruido, sus cruces por cero y energía, el resultado de su análisis de Fourier, el jitter y shimmer, la relación entre diferentes coeficientes de correlación de dicha señal respecto al primero, la información de fase entre fuentes glóticas distintas normalizadas. Los parámetros de formantes son los nueve primeros y sus respectivos anchos de banda. También se añade el perfil del tracto vocal y la nasalidad. En cuanto a la metodología para establecer las comparaciones se presenta una arquitectura con un sistema de extracción próximo al locutor de donde se toman los correlatos acústicos para el análisis, que se transmiten por una red de comunicaciones a un servidor remoto donde se procede a su verificación contra una base de locutores previamente elaborada, devolviéndose la decisión de autenticación al sistema próximo (Figura 1). La descripción de los métodos a utilizar no es muy precisa. Tampoco se especifica el procedimiento de autenticación de locutor que se comenta. Los puntos de relación con la presente solicitud se centran en la utilización de parámetros derivados de la onda glótica, si bien son totalmente distintos en su concepción (primeros armónicos, jitter y shimmer, sin especificar cuáles de las diferentes parametrizaciones existentes en la literatura de dominio público se proponen), trayectorias en el plano z, cruces por cero, todos ellos muy alejados de la presente solicitud (basada en parámetros de distorsión, cepstrales, singularidades de la densidad espectral de potencia de la onda glótica, parámetros biomecánicos, parámetros temporales de eficiencia glótica, que poseen una semántica claramente superior a los utilizados en la patente comentada). La nueva solución propuesta en la presente solicitud mejora lo expuesto en la patente referida en la obtención de onda glótica precisos en lugar de voz completa, en la estimación de un conjunto de parámetros con alta semántica funcional en disfonía, y en la robustez y precisión de las estimaciones frente a la variabilidad intra-locutor, mejorando la capacidad de separación inter-locutor. Estas innovaciones hacen que las diferencias entre lo manifestado en dicha patente y la presente solicitud sean sustanciales en lo que hace referencia a objetivos, materiales y metodología utilizados, siendo la nueva solución propuesta claramente más ventajosa en cuanto a ofrecer más parámetros con mejor semántica, menor variabilidad intra-locutor, y mayor precisión y robustez.

- International Publication Number WO 2010/031437 A1

METHOD AND SYSTEM OF VOICE CONVERSION

5 El objetivo de la patente referida es la conversión de la voz de un locutor (objeto) en la voz de otro locutor (objetivo), mediante el modelado de la fuente glótica y el tracto vocal en cada ciclo glótico, incluyendo la intensidad de la excitación, un conjunto de parámetros de la fuente glótica y los coeficientes del filtro todo-polo que modelan el tracto vocal. La metodología utilizada es la extracción de la fuente glótica y el tracto vocal por medio del modelo de estimación conjunta de Lu & Smith para obtener un conjunto de parámetros de onda glótica y de modelo de tracto vocal, ajustando la fuente glótica obtenida por filtrado inverso contra un modelo Rosenberg-Klatt por medio de optimización no lineal restringida. Por este procedimiento se determina un vector de características de la onda glótica compuesto por la intensidad de la excitación (Ee), los parámetros temporales de máximo flujo (Tp), de fase abierta (Te), de ajuste de retorno (Ta) y de fin de recuperación (Tc), junto con la energía de ruido de aspiración (ANE). Se genera una base de datos con diferentes locutores estimando estos parámetros. A continuación se presentan los resultados de la síntesis cruzada de voz mediante estimaciones objetivas y subjetivas. La relación con la presente solicitud se centra en los procedimientos de extracción de onda glótica, si bien la estimación conjunta de onda glótica y tracto vocal se lleva a cabo en ambos casos por procedimientos bien distintos: la estimación conjunta por Lu&Smith utilizando optimización no lineal de parámetros en el caso de la patente, y por celosías adaptativas modelo-espejo en el caso de la presente solicitud. La nueva solución propuesta en la presente solicitud mejora lo expuesto en la patente referida en la obtención de onda glótica precisos en lugar de voz completa, en la estimación de un conjunto de parámetros con alta semántica funcional en disfonía, y en la robustez y precisión de las estimaciones frente a la variabilidad intra-locutor, mejorando la capacidad de separación inter-locutor. Estas innovaciones hacen que las diferencias entre lo manifestado en dicha patente y la presente solicitud sean sustanciales en lo que hace referencia a objetivos, materiales y metodología utilizados, siendo la nueva solución propuesta claramente más ventajosa en cuanto a ofrecer más parámetros con mejor semántica, menor variabilidad intra-locutor, y mayor precisión y robustez.

- US Patent Number US006195632B1

30 EXTRACTING FORMANT-BASED SOURCE-FILTER DATA FOR CODING AND SYNTHESIS EMPLOYING COST FUNCTION AND INVERSE FILTERING

El objetivo de la patente referida es la estimación de los formantes de la voz mediante la minimización de una función de coste definida sobre el residual glótico que se denomina "arc-length" (longitud de arco). La metodología opera como sigue: la señal de voz se somete a filtrado inverso para evaluar el residual glótico, sobre el que se evalúa una función de coste. Sobre ella se lleva a cabo un proceso de optimización que permite detectar los parámetros de ajuste del filtro inverso, y reconstruir la voz sintetizada para comprobar su calidad. Los parámetros a ajustar son los polos del filtro inverso y sus anchos de banda, mientras que la medida de calidad se basa en la fijación de una serie de puntos de referencia en el correlato glótico, y en el cálculo de la tensión del arco resultante entre cada par de puntos, que se acumula como una medida cuadrática. Los parámetros del filtro inverso se modifican adecuadamente para que la medida de tensión de arco resultante se vaya minimizando iterativamente. De esta forma se consigue definir un correlato glótico y un filtro inverso "óptimos" en el sentido dicho. La relación con la presente solicitud se centra en los procedimientos de extracción de onda glótica, si bien la estimación de la onda glótica y tracto vocal se lleva a cabo en ambos casos por procedimientos bien distintos: la estimación de la fuente y el filtro por medio de la optimización de la función de "tensión de arco" en el caso de la patente que se referencia, y por celosías adaptativas modelo-espejo ilustradas en las Figuras 2 y 5 en el caso de la presente solicitud. Esta nueva solución propuesta en la presente solicitud mejora lo expuesto en la patente referida en la obtención de onda glótica precisos en lugar de voz completa, en la estimación de un conjunto de parámetros con alta semántica funcional en disfonía, y en la robustez y precisión de las estimaciones frente a la variabilidad intra-locutor, mejorando la capacidad de separación inter-locutor. Estas innovaciones hacen que las diferencias entre lo manifestado en dicha patente y la presente solicitud sean sustanciales en lo que hace referencia a objetivos, materiales y metodología utilizados, siendo la nueva solución propuesta claramente más ventajosa en cuanto a ofrecer más parámetros con mejor semántica, menor variabilidad intra-locutor, y mayor precisión y robustez.

55 Descripción de la invención

Introducción

Las limitaciones identificadas en el estado actual de la técnica en el área en que se pretende actuar son las siguientes:

- La influencia del tracto vocal en la fonación es fuertemente enmascaradora de la actividad dinámica de los pliegues vocales, y dificulta enormemente la estimación del estado fisiológico de la misma a partir del registro de la voz.
- Las estimaciones del estado fisiológico de los pliegues vocales a partir del análisis acústico de la voz se centran en el uso de parámetros de distorsión que no tienen una semántica bien definida y unívoca respecto al problema que modelan.

- La personalización del locutor a partir del habla incorpora mucha información articulatoria, dependiente del texto, que genera una alta variabilidad intra-locutor dificultando las tareas de la identificación robusta.

La presente invención resuelve los problemas y las limitaciones anteriores mediante las siguientes actuaciones:

- Se utilizan parámetros derivados de la onda glótica para determinar la actividad dinámica de los pliegues vocales, mediante la reconstrucción de dicha señal por filtrado inverso de la señal de voz. Los nuevos parámetros son estimaciones de la envolvente del espectro en frecuencia de la onda glótica así reconstruida.
- Se realizan estimaciones de los parámetros biomecánicos de los pliegues vocales mediante la adaptación de un modelo biomecánico resonante que reconstruya el comportamiento en frecuencia en una banda dada de la envolvente espectral de la onda glótica. Los parámetros biomecánicos son estimados mediante los valores del modelo biomecánico citado mediante inversión del sistema dinámico de los mismos. Estos nuevos parámetros determinan comportamientos normales y anormales de los pliegues vocales en fonación de forma mucho más directa.
- Se elimina la influencia del tracto vocal en la fonación mediante filtrado inverso, lo que reduce la variabilidad intra-locutor producida por la articulatoria. Ello mejora las tasas de discriminación inter-locutor al separar mejor las clases de locutores modelados.

Para ello se propone un sistema de captura de la señal de voz y un conjunto de métodos algorítmicos diseñados para la extracción de parámetros relevantes a partir de la onda glótica, y para su clasificación de acuerdo con una población de control normal, que permiten la determinación de la presencia de disfonía, del grado de la misma, y de la identidad del locutor.

Breve descripción de los dibujos

Figura 1. Diagrama general del sistema que describe la invención. Consiste en una toma de sonido (1-2), seguida de una codificación digital (3) bajo control de un circuito lógico programable (4) que realiza las funciones de control (5), almacenamiento (6), estimación de parámetros y clasificación de los mismos (7, 8, 9) y presentación (10) o puesta a disposición de otro sistema (11) para su interpretación por un experto.

Figura 2. Proceso para la estimación de la onda glótica por inversión del modelo de producción de la voz. Se elimina el efecto de radiación (12), se modela el pulso glótico (13), se elimina su influencia (14), se modela el tracto vocal (16) y se elimina asimismo su influencia (15) cruzándose las estimaciones con influencia eliminada (glótica o vocal) para realizar sucesivos refinamientos de los modelados. Se genera una señal residual glótica como resultado sin influencia sustancial del tracto vocal.

Figura 3. Proceso de estimación de los parámetros de referencia que se utilizarán en las realizaciones preferentes (clínica, de identificación del locutor, y forense). Mediante dos integraciones sucesivas (17) y (18) se estima la fuente glótica y el flujo glótico. La fuente glótica se utiliza para estimar los parámetros temporales del ciclo glótico (19), los parámetros de distorsión (20), y la onda acústica promedio (21). A partir de esta señal se calcula la densidad espectral de potencia (22) del correlato de onda mucosa que permite estimar los parámetros biométricos de la fonación (23) así como los parámetros biomecánicos de la cubierta del pliegue vocal (25). A partir de la onda acústica promedio se pueden calcular los parámetros biomecánicos del cuerpo del pliegue vocal (24).

Figura 4. Filtro cancelador del efecto de radiación. Está construido como una celosía de correlación parcial que es el inicio de una cadena de filtros de modelado y cancelación cruzada, que se denominan modelo-espejo.

Figura 5. Filtros celosía modelo espejo para estimación conjunta. Etapa de un filtro de modelado y cancelación que muestra el flujo a modelar (33-33') y el flujo modelado (37-37') así como el procedimiento de estimación (34) y recorrelación cruzada (35, 38) generándose los flujos que se traspasan a la etapa siguiente (36-36') y (39-39').

Figura 6. Ejemplo de reconstrucción de la fuente glótica resultante de (17-18) para voz masculina. La fuente (onda) glótica está dada por (40) mientras que el flujo glótico es (41). En (42) pueden verse sucesivos ciclos de la fuente glótica.

Figura 7. Ejemplo de reconstrucción de la fuente glótica resultante de (17-18) para voz femenina. La fuente (onda) glótica está dada por (43) mientras que el flujo glótico es (44). En (45) pueden verse sucesivos ciclos de la fuente glótica.

Figura 8. Perfil (envolvente) de la densidad espectral de potencia de la fuente glótica. Se observa una función de caída promedio inversamente proporcional a la frecuencia, salpicada por picos (46) y valles (47) sucesivos que constituyen las singularidades de dicho perfil, y que debidamente estimados en valor y posición normalizada dan lugar a los parámetros biométricos de dicha fuente glótica.

Figura 9. Ajuste de la densidad espectral de potencia de la onda acústica promedio (48) mediante una función de segundo orden (49) cuyos parámetros de ajuste se convierten en estimadores de la biomecánica del cuerpo del pliegue vocal.

Figura 10. Ajuste de la densidad espectral de potencia de la onda mucosa (50) mediante una función de segundo orden (51) cuyos parámetros de ajuste se convierten en estimadores de la biomecánica de la cubierta del pliegue vocal.

Figura 11. Ejemplo de realización preferida para la aplicación del método y sistema propuesto para la detección y gradación de la disfonía. La señal de voz (55) se captura y almacena (52, 54, 53), y sobre la misma se extrae la fuente glótica (56). Se estiman los parámetros biométricos y biomecánicos (57) y se seleccionan los más relevantes a la aplicación deseada (58). Con suficiente cantidad de individuos normofónicos se genera un modelo de locutor de referencia (58) que se utiliza en contraste mediante parámetros clasificatorios basados en diferentes algoritmos, como modelos de mezclas de gaussianas (60: GMM) no excluyendo otros modelos posibles. Se genera un índice de grado disfónico (61) que puede utilizarse para evaluar el nivel de disfonía (62: 0-1 normal, 1-2 leve, 2-3 moderado, >3 grave) y para decidir (63) posibles actuaciones consecuentes, como inspección en servicio de especialidad, etc.

Figura 12. Ejemplo de realización preferida para la aplicación del método y sistema propuesto para la identificación y verificación del locutor. La señal de voz (67) se captura y almacena (64, 66, 65), y sobre la misma se extrae la fuente glótica (70). Se estiman los parámetros cepstrales (71) a partir del espectro convencional (68), los cepstrales (72) derivados del espectro LPC (69) y los biométricos y biomecánicos (73) de la fuente glótica (70) con los que se compone un supervector (74) que se utiliza en detección de la secuencia temporal (75) y del grado de semejanza (76) con respecto a un modelo de locutor previamente producido (77). Las puntuaciones de similitud en la identificación de un locutor contra otros (78) se utilizan para generar unos índices de identidad y certeza (79) que informan sobre la personalidad estimada y la confiabilidad de la estimación. Con estas informaciones se pueden tomar decisiones sobre aceptación, rechazo o evaluación de personalidad por métodos alternativos (80).

Figura 13. Ejemplo de realización preferida para la aplicación del método y sistema propuesto para la verificación y cotejo forense del locutor. La señal de voz (84) se captura y almacena (81, 83, 82), y sobre la misma se extrae la fuente glótica (85). Se estiman los parámetros biométricos y biomecánicos (86) y se seleccionan los más relevantes a la aplicación deseada (87). Con suficiente cantidad de individuos normofónicos se genera un modelo universal de locutor (88) que se utiliza en contraste mediante parámetros clasificatorios basados en diferentes algoritmos, como modelos de mezclas de gaussianas (89: GMM) no excluyendo otros posibles. Se genera un índice de verosimilitud (61: LR) que puede utilizarse para evaluar la evidencia del cotejo (91: <0 rechazo, 0-1 evidencia débil, 1-2 evidencia fuerte, 2-3 evidencia muy fuerte) y para decidir (92) posibles actuaciones consecuentes.

Realización preferente de la invención

A continuación con carácter no limitativo, se ilustra con referencia a las figuras un modo de realización de la invención.

Para describir el sistema propuesto se parte del diagrama de bloques que se muestra en la Figura 1, que ilustra la plataforma de registro de voz, cómputo de parámetros y presentación de los mismos en una interfaz de usuario de tipo portable. En la misma, la señal de voz obtenida desde un micrófono convencional (1) o desde un teléfono (2) es capturada y codificada a digital (3) y almacenada en una memoria temporal (6) en un circuito de control constituido por una FPGA (4: Field-Programmable Gate Array) que incorpora además la programación adecuada para controlar la captura de sonido (5) y el resto de los procesos de tratamiento de señal e identificación de patrones (7: estimador de onda glótica, 8: estimador de los parámetros de referencia, 9: clasificador de patrones). Los resultados son almacenados asimismo en memoria temporal (6) desde donde o bien se ofrecen por una pantalla tipo i-pod al usuario (10) o se ponen a disposición de otros sistemas informáticos mediante una interfaz USB (11).

Por su parte el método propuesto consiste en una serie de procesos para la estimación de los parámetros de clasificación que se hallan programados dentro de la FPGA de la Figura 1. Estos procesos son los siguientes:

- Proceso para estimación de la onda glótica. Se describe en la Figura 2. Consiste en un filtro inverso (12) que compensa el efecto de radiación desde los labios al micrófono, y de un par de filtros modelo-espejo que modelan el comportamiento de la onda glótica y del tracto vocal y los eliminan sobre la señal de voz. El sistema Filtro Glótico Inverso (13) modela la onda glótica, y su Filtro Espejo (14) elimina la influencia de la misma en la voz, produciendo una voz desglotalizada. El sistema Filtro Inverso del Tracto Vocal (16) modela las resonancias de las cavidades faríngea, vocal y nasal, y su Filtro Espejo (15) elimina las mismas en la señal de voz, produciendo un residual glótico.
- Proceso para la estimación de los parámetros de referencia. Se describe en la Figura 3. El residual glótico producido en el proceso anterior se integra en un Filtro Integrador (17) y produce la Fuente Glótica. La integración de esta señal en un nuevo Filtro Integrador (18) produce el Flujo Glótico. Por su parte la Fuente Glótica se utiliza para la Estimación de sus Parámetros Temporales (19), y de sus Parámetros de Distorsión (20). Adicionalmente sobre ella estima la Onda Acústica Promedio en un Detector de Onda Acústica Promedio (21), que produce asimismo un Correlato de Onda Mucosa. A partir del Correlato de Onda Mucosa se estima la Densidad Espectral de Potencia de su Envolvente en (22), sobre la cual se estiman los Parámetros Biométricos de la Onda

Mucosa (23). También a partir del Correlato de Onda Mucosa se estiman los Parámetros Biomecánicos de la Cubierta del Pliegue Vocal (25). De la Onda Acústica Promedio se estiman los Parámetros Biomecánicos del Cuerpo del Pliegue Vocal (24).

En los procedimientos de reconstrucción del residual de onda glótica de la Figura 2 juegan un papel relevante los filtros inversos propuestos, que son parte importante de la presente solicitud porque mejoran la robustez de los procesos de estimación empleados en el estado del arte, y son más precisos que los convencionales de tipo autorregresivo. A continuación se pasa a describir cada uno de los bloques de las Figuras 2 y 3 en mayor detalle:

Bloque (12) En la Figura 4 se muestra la estructura del filtro cancelador del efecto de radiación (32) equivalente al bloque (12) de la Figura 2, el cual se implementa mediante una celosía adaptativa de modo que la señal de voz (26) es dividida en dos ramas (27) y (28) y alimenta a un estimador PARCOR (29) que se utiliza para cancelar la correlación cruzada en (30) entre (27) y (28) desfasada, generando la señal libre del efecto de radiación en (31).

Bloques (13)-(14) Se construyen como los filtros modelo-espejo, se describen en la Figura 5. En ella el filtro de modelado recibe las señales de propagación directa (33) y reversa (33') procedentes de la señal glótica sobre las que estima un coeficiente de correlación parcial (34) que se usa en (35) para eliminar dicha correlación dando lugar a dos nuevas señales de propagación directa (36) y reversa (36'). El mismo coeficiente de correlación parcial (34) se aplica sobre los flujos de propagación directa (37) y reversa (37') procedentes de la señal de voz en el filtro espejo para eliminar dicha correlación cruzada (38) dando lugar a dos nuevos flujos de propagación directa (39) y reversa (39'). Estos filtros de celosías emparejadas se pueden encadenar hasta constituir un sistema de orden adecuado al tipo de modelado requerido. En el sistema (13)-(14) suele ser suficiente encadenar de una a tres de estas etapas.

Bloques (15)-(16) Se construyen como los filtros modelo-espejo que se describen en la Figura 5. En ella el filtro de modelado recibe las señales de propagación directa (33) y reversa (33') procedentes de la voz desglotalizada sobre las que estima un coeficiente de correlación parcial (34) que se usa en (35) para eliminar dicha correlación dando lugar a dos nuevas señales de propagación directa (36) y reversa (36'). El mismo coeficiente de correlación parcial (34) se aplica sobre los flujos de propagación directa (37) y reversa (37') procedentes de la señal de voz en el filtro espejo para eliminar dicha correlación cruzada (38) dando lugar a dos nuevos flujos de propagación directa (39) y reversa (39'). Estos filtros de celosías emparejadas se pueden encadenar hasta constituir un sistema de orden adecuado al tipo de modelado requerido. En el sistema (15)-(16) puede ser necesario encadenar varias decenas de estas etapas, dependiendo de la precisión que se desee obtener en las estimaciones, y de la frecuencia de muestreo de las señales.

Bloques (17)-(18) Se construyen como integradores simples de la señal que reciben a su entrada, para lo que se pueden utilizar diferentes algoritmos con o sin fuga para garantizar la compensación de la deriva continua.

Bloque (19) Estimador de Parámetros Temporales. Se diseña para que se puedan determinar adecuadamente los parámetros temporales de relevancia en la señal glótica que se muestra en la Figura 6 (voz masculina) y 7 (voz femenina). Los parámetros de base temporal del ciclo de fonación están referidos a los instantes singulares del Ciclo de Fonación tal como se ilustra en las citadas figuras, que son:

- Instante de Inicio del Cierre Glótico. Es el punto en que la Fuente Glótica alcanza su valor mínimo en forma de espiga puntiaguda negativa (punto inicial de las Figuras 6 y 7, (40 y 43), equivalente a los puntos marcados con un asterisco /*/ en la plantilla inmediatamente inferior a la referida en dicha figura), resultante de la depresión que produce en la zona supraglótica la interrupción brusca del flujo, mientras que la columna de aire presente en el Tracto Vocal sigue su movimiento de salida dado su comportamiento inercial. Se toma como origen del ciclo glótico, en $t=0$.
- Instante de Recuperación $t=Tr$. Cuando se cierra bruscamente un conducto por el que circula un fluido, se produce una retroacción expresada como un pico negativo de presión, ya que se produce un vacío por la incapacidad del fluido en movimiento para detenerse instantáneamente por poseer una masa inerte. Ello induce un retroceso de parte del fluido, que compensa la caída de presión al cabo de un breve tiempo. El instante de recuperación es el punto temporal en que el retroceso parcial de la columna de aire presente en el Tracto Vocal equilibra de nuevo la presión supraglótica a la de reposo (atmosférica, o de referencia).
- Instante de Inicio de la Apertura $t=To$. Es el punto en que las cuerdas vocales vuelven a abrirse incipientemente.

- Instante de máxima sobrepresión $t=T_m$. Es el punto en que se alcanza el máximo de presión supraglótica.
- Instante de Inicio del Cierre $t=T_c$. Es el punto en que se alcanza la máxima apertura o luz entre los pliegues vocales, correspondiente con el máximo flujo si se pudiese descartar la influencia del Tracto Vocal, a partir del cual la luz entre pliegues (apertura) empieza a decrecer.
- Instante Final del Ciclo Glótico $t=T_g$. Es el punto en el que se alcanza el mínimo de presión supraglótica, y que coincide con el inicio de un nuevo ciclo.

Sobre la fuente glótica (señal en grueso) se estiman cuatro tiempos de referencia en la parametrización temporal: el instante de retorno (T_r), el instante de apertura (T_o), el instante de máxima amplitud (T_m) y el instante de inicio del cierre (T_c). La duración completa del ciclo glótico viene dada por T_g . La parametrización temporal se basa en la estimación de dos señales a partir de la fuente glótica: la onda acústica promedio $s_s(n)$, y el correlato de la onda mucosa $s_w(n)$, según se ilustra en el Proceso (21). Sobre el flujo glótico (señal en fino) se estima un tiempo de referencia, coincide con el máximo de dicha onda (T_M).

Bloque (20) Estimador de Parámetros de Distorsión. Se estima una serie de parámetros de distorsión, que son el *jitter*, el *shimmer de amplitud*, el *shimmer de área*, la *agudeza del cierre*, y la relación *cubierta/cuerpo*.

Bloque (21) La onda acústica promedio es una señal de carácter semisinusoidal que tiene la duración de un arco glótico para voz masculina o femenina, (40) o (43), con una amplitud que minimiza la diferencia entre su área y la de la fuente glótica. Su estimación se realiza para cada ciclo de fonación, y en forma sincrónica con el principio y fin de éste, definido desde mínimo a mínimo de la fuente glótica como se muestra en las plantillas (42) y (45) de las Figuras 6 y 7.

Bloque (22) Estimador de la Densidad Espectral de Potencia de la Envolvente de la Onda Glótica. La envolvente de la densidad espectral de potencia del correlato glótico de que se trate (onda glótica, onda mucosa) se estima como el módulo de la Transformada de Fourier de un ciclo de la citada onda, y su aspecto es similar al descrito en la Figura 8.

Bloque (23) Estimador de los Parámetros Biométricos de la Onda Mucosa. El comportamiento promedio de la envolvente de la onda glótica o de la onda mucosa es de caída inversamente proporcional a la frecuencia, mostrando ciertas singularidades en forma de picos (46) y valles (47) alternados. La estimación precisa de estos picos y valles constituye el conjunto de parámetros biométricos de la envolvente de la densidad espectral de potencia de la onda referida.

Bloque (24) Estimador de la Biomecánica del Cuerpo del Pliegue Vocal. Se basa en el ajuste de una función de transferencia de segundo orden (49) sobre la densidad espectral de la onda acústica promedio (48) como se muestra en la Figura 9 en un intervalo de baja frecuencia. Los parámetros de la función de transferencia de ajuste constituyen los Parámetros Biomecánicos del Cuerpo del Pliegue Vocal.

Bloque (25) Estimador de la Biomecánica de la Cubierta del Pliegue Vocal. Se basa en el ajuste de una función de transferencia de segundo orden (51) sobre la densidad espectral del correlato de onda mucosa (50) como se muestra en la Figura 10 en un intervalo de baja frecuencia. Los parámetros de la función de transferencia de ajuste constituyen los Parámetros Biomecánicos de la Cubierta del Pliegue Vocal.

Descripción detallada de los procesos principales realizados en los bloques

Proceso (12). Modelo Inverso de Radiación $H_f(z)$

Se lleva a cabo mediante una Celosía de Error de Predicción de primer orden como la que se muestra en la Figura 4, que opera como un filtro FIR (Finite Impulse Response) según la recursión, donde n hace referencia al índice temporal discreto:

$$f_k(n) = f_{k-1}(n) + c_{k-1}b_{k-1}(n-1) \quad eq.1$$

de modo que cuando $k=1$ y $c_0=-r_f$ (primer coeficiente de reflexión) supuesto que:

$$f_0(n) = b_0(n) = s(n) \quad eq.2$$

la celosía se comporta como un diferenciador de primer orden:

$$s_l(n) = f_l(n) = s(n) - r_f s(n-1) \quad eq.3$$

con función de transferencia dada por:

$$H_l(z) = R^{-1}(z) = 1 - r_f z^{-1} \quad eq.4$$

que cancela el polo de primer orden introducido por los efectos de la radiación en los labios.

Proceso (13). Filtro Glótico Inverso $H_{gi}(z)$

Según se muestra en la Figura 2 el filtro glótico inverso tiene por función construir un modelo de inversión espectral de la señal a su entrada, que es el residual de onda glótica a modelar $s_{ri}(n)$. En el modelado se obtienen unos coeficientes de pivotaje $\{c_k\} \rightarrow \{h_{gi}\}$ que inyectados en otra celosía semejante (filtro espejo) permiten eliminar la huella espectral de la señal modelada sobre otra señal diferente, en este caso la señal de voz compensada en radiación $s_l(n)$, dando lugar a la voz desglotalizada $s_{vi}(n)$. La estimación de los coeficientes de pivotaje se puede hacer en modo bloque o en modo adaptativo. Ambos mecanismos son utilizados por el método propuesto en la solicitud.

Proceso (14). Filtro Espejo de desglotalización

Según se muestra en la Figura 2 la función de un filtro espejo es la de realizar la deconvolución de la señal a su entrada respecto a un conjunto de parámetros $\{c_k\} \rightarrow \{h_{gi}\}$ estimados en un modelo nodriza que obtiene los mismos y los inyecta en el filtro espejo. Los dos filtros, de modelado inverso y su espejo, constituyen un estimador conjunto de procesos, y su implementación mediante celosías puede verse en la Figura 5, que muestra una etapa de este tipo de estructuras. La replicación de K de estas etapas permite construir el estimador conjunto de orden K . En este caso la señal a procesar es la voz compensada en radiación ($s_l(n)$ reducida al punto labial), dando lugar a la señal de voz desglotalizada $s_{vi}(n)$.

Proceso (15). Filtro Espejo de eliminación de la influencia espectral del Tracto Vocal

Según la Figura 2 los parámetros del filtro de modelado inverso del tracto vocal $\{c_k\} \rightarrow \{h_{vi}\}$ (16) inyectados en el correspondiente filtro espejo (15) eliminan de la señal de voz compensada en radiación $s_l(n)$ la influencia de los procesos articulatorios y dejan como resultado el residual glótico $s_{ri}(n)$.

Proceso (16). Filtro Inverso del Tracto Vocal $H_{vi}(z)$

Refiriéndose asimismo a la Figura 2, el funcionamiento de este tipo de sistema es similar al descrito en (13), aunque en este caso la señal que se modela es la señal de voz desglotalizada $s_{vi}(n)$. De este modo se deriva un conjunto de coeficientes de pivotaje $\{c_k\} \rightarrow \{h_{vi}\}$ que modelan en forma inversa el comportamiento en frecuencia del tracto vocal.

Proceso (17). Filtro Integrador estimador de la fuente glótica

Según la Figura 3 se genera la fuente glótica $s_{gi}(n)$ a partir del residual glótico $s_{ri}(n)$ por integración simple mediante la expresión:

$$s_{gi}(n) = s_{gi}(n-1) + r_l s_{ri}(n) \quad eq.5$$

donde r_l es un coeficiente de control de derivas cuyo fin es evitar la acumulación de estos efectos no deseados.

Proceso (18). Filtro Integrador estimador del flujo glótico

Según la Figura 3 se genera el flujo glótico $u_{gi}(n)$ a partir de la fuente glótica $s_{gi}(n)$ por integración simple mediante la expresión:

$$u_{gi}(n) = u_{gi}(n-1) + r_i s_{gi}(n) \quad eq.6$$

donde r_i es el correspondiente coeficiente de control de derivas.

Proceso (19). Estimador de los parámetros temporales del ciclo glótico

Un ejemplo de ciclo de la fuente glótica $s_{gi}(n)$ se puede ver en las Figuras 6 (masculina) y 7 (femenina). Los parámetros de base temporal del ciclo de fonación están referidos a los instantes singulares del Ciclo de Fonación tal como se ilustra en la Figura 6, que son:

- Instante de Inicio del Cierre Glótico. Es el punto en que la Fuente Glótica alcanza su valor mínimo en forma de espiga puntiaguda negativa (punto inicial de la Figura 6, superior, equivalente a los puntos marcados con un asterisco /*/ en (42) o (45), resultante de la depresión que produce en la zona supraglótica la interrupción brusca del flujo, mientras que la columna de aire presente en el Tracto Vocal sigue su movimiento de salida dado su comportamiento inercial. Se toma como origen del ciclo glótico, en $t=0$.
- Instante de Recuperación $t=T_r$. Es el punto en que el retroceso parcial de la columna de aire presente en el Tracto Vocal equilibra de nuevo la presión supraglótica a la de reposo (atmosférica, o de referencia).
- Instante de Inicio de la Apertura $t=T_o$. Es el punto en que las cuerdas vocales vuelven a abrirse incipientemente.
- Instante de máxima sobrepresión $t=T_m$. Es el punto en que se alcanza el máximo de presión supraglótica.
- Instante de Inicio del Cierre $t=T_c$. Es el punto en que se alcanza la máxima apertura o luz entre los pliegues vocales, correspondiente con el máximo flujo si se pudiese descartar la influencia del Tracto Vocal, a partir del cual la luz entre pliegues (apertura) empieza a decrecer.
- Instante Final del Ciclo Glótico $t=T_g$. Es el punto en el que se alcanza el mínimo de presión supraglótica, y que coincide con el inicio de un nuevo ciclo.

Sobre la fuente glótica (señal en grueso) se estiman cuatro tiempos de referencia en la parametrización temporal: el instante de retorno (T_r), el instante de apertura (T_o), el instante de máxima amplitud (T_m) y el instante de inicio del cierre (T_c). La duración completa del ciclo glótico viene dada por T_g . La parametrización temporal se basa en la estimación de dos señales a partir de la fuente glótica: la onda acústica promedio $s_s(n)$, y el correlato de la onda mucosa $s_w(n)$. Sobre el flujo glótico (señal en fino) se estima un tiempo de referencia, coincide con el máximo de dicha onda (T_M). De acuerdo con las anteriores definiciones establecidas en el Bloque (19), la estimación de cada uno de los tiempos de referencia se ajusta a los siguientes procedimientos:

$$T_r = \arg \left[\max_{0 \leq n < n_o} \{s_{mk}(n)\} \right] \quad eq.7$$

$$T_o = \arg \left[\min_{0 \leq n < N_k} \{s_{wk}(n)\} \right] \quad eq.8$$

$$T_m = \arg \left[\max_{n_o \leq n < N_k} \{s_{gk}(n)\} \right] \quad eq.9$$

$$T_c = \arg \left[s_{gk}(n_o \leq n < N_k) = s_{gk} \left(\arg \left[\min_{0 \leq n < N_k} \{s_{wk}(n)\} \right] \right) \right] \quad eq.10$$

$$T_M = \arg \left[\max \{u_{gk}(n)\} \right] \quad eq.11$$

Asimismo se estiman los siguientes parámetros de base temporal, que se detallan a continuación:

- OQ , coeficiente de apertura, que mide la duración relativa del intervalo en que la glotis está abierta con respecto a la duración del ciclo glótico T_g .
- SQ , coeficiente de velocidad, que mide la relación entre las dos partes del ciclo de apertura, antes y después del punto de amplitud máxima positiva.
- ClQ , coeficiente de cierre, que mide la relación entre la segunda mitad del ciclo de apertura, desde el punto de amplitud máxima positiva hasta el instante de cierre, y la duración del ciclo glótico T_g .
- RQ , coeficiente de retorno, que mide la relación entre el intervalo de retorno y la duración del ciclo glótico T_g .

- NAQ , coeficiente de amplitud normalizada, que mide la relación entre el valor máximo del flujo glótico (curva en línea fina) y el área del cuadrante inferior de la onda gótica por debajo de T_o .
- ArQ , coeficiente de amplitud relativa del instante de retorno respecto a la amplitud máxima.
- AoQ , coeficiente de amplitud relativa del instante de apertura respecto a la amplitud máxima.

Los parámetros anteriores se estiman de la siguiente manera:

$$OQ = \frac{T_g - T_o}{T_g} \quad eq.12$$

$$SQ = \frac{T_m - T_o}{T_g - T_m} \quad eq.13$$

$$CIQ = \frac{T_g - T_m}{T_g} \quad eq.14$$

$$RQ = \frac{T_r}{T_g} \quad eq.15$$

$$CMQ = \frac{T_M}{T_g}$$

$$NAQ = \frac{\max\{u_{gk}(n_o \leq n < N_k)\}}{s_{gk}(T_o)T_g} \quad eq.16$$

$$ArQ = \frac{s_{gk}(T_r)}{s_{gk}(T_m)} \quad eq.17$$

$$AoQ = \frac{s_{gk}(T_o)}{s_{gk}(T_m)} \quad eq.18$$

$$AMQ = \frac{s_{gk}(T_M)}{s_{gk}(T_m)} \quad eq.19$$

Finalmente se añade un grupo adicional de parámetros que mide la eficacia de la fonación, como relación entre la eficiencia de la inyección aérea y la deficiencia del cierre glótico (escapes debidos a cierre defectuoso) y que se definen como:

- ODQ , coeficiente de defecto en apertura, parámetro que estima la merma del flujo promedio en la fase de apertura glótica (desde T_o hasta T_g), debida a la presencia de cierre prematuro, o deficiencia en inyección.
- CDQ , coeficiente de defecto de cierre, parámetro que estima el flujo promedio en la fase de cierre glótico (desde 0 hasta T_o), debida a la presencia de apertura prematura, o deficiencia por escape.
- GEQ - eficiencia glótica, parámetro que estima el complementario de la deficiencia por inyección más la deficiencia por escape, como factor de mérito en la fonación.

Los parámetros anteriores se estiman de la siguiente manera:

$$W_s(\sigma, \delta) = \frac{1}{\sqrt{\sigma}} \int_{-\infty}^{\infty} \psi_g\left(\frac{t-\delta}{\sigma}\right) s_{gk}(t) dt \quad eq.20$$

$$L_c = \sum_{\delta < \delta_o} |W_s(\sigma > \sigma_1, \delta)|^2;$$

$$L_o = \sum_{\delta \geq \delta_o} |W_s(\sigma > \sigma_1, \delta)|^2; \quad eq.21$$

$$L_r = \sum_{\delta} |W_s(\sigma > \sigma_1, \delta)|^2$$

$$ODQ = \gamma_c = \frac{L_c}{L_r}; \quad eq.22$$

$$CDQ = \gamma_o = \frac{L_o}{L_r}$$

$$GEQ = 1 - \gamma_c - \gamma_o \quad eq.23$$

Donde s_{gk} es la onda glótica, ψ_g es una ondícula de tipo Gausiano, escalable en el parámetro σ y desplazable en el parámetro δ . El cuadro de asignación de Parámetros de Base Temporal del Ciclo de Fonación queda como sigue:

$$\begin{aligned} p_{47k} &= OQ; & p_{48k} &= SQ; & p_{49k} &= CIQ; \\ p_{50k} &= RQ; & p_{51k} &= CMQ; & p_{52k} &= NAQ; \\ p_{53k} &= ArQ; & p_{54k} &= AoQ; & p_{55k} &= AMQ; \\ p_{56k} &= ODQ; & p_{57k} &= CDQ; & p_{58k} &= GEQ; \end{aligned} \quad eq.24$$

Proceso (20). Estimador de los parámetros de distorsión de la fuente glótica

A partir de la fuente glótica evaluada en el ciclo de fonación k -ésimo:

$$s_{gk}(n) = s_{gi}(n); \quad n_{k-1} \leq n < n_k; \quad eq.25$$

donde n_{k-1} y n_k son los límites temporales superiores de los ciclos glóticos $k-1$ -ésimo y k -ésimo respectivamente con tamaños dados por:

$$N_{k-1} = n_{k-1} - n_{k-2}; \quad N_k = n_k - n_{k-1} \quad eq.26$$

Se estima una serie de parámetros de distorsión, que son el *jitter*, el *shimmer de amplitud*, el *shimmer de área*, la *agudeza del cierre*, y la *relación cubierta/cuerpo*, que se definen a continuación. El *jitter* se estima como:

$$p_{2k} = 2 \frac{\left| \frac{I}{N_k} - \frac{I}{N_{k-1}} \right|}{\frac{I}{N_k} + \frac{I}{N_{k-1}}} \quad eq.27$$

El *shimmer de amplitud* se estima asimismo como:

$$p_{3k} = 2 \frac{|s_{gmk} - s_{gmk-1}|}{s_{gmk} + s_{gmk-1}} \quad eq.28$$

donde s_{gmk} es el máximo valor en amplitud de pico a pico que toma la fuente glótica dentro del ciclo k -ésimo.

El *shimmer de área* se estima asimismo como:

$$p_{4k} = 2 \frac{|S_{gk} - S_{gk-1}|}{S_{gk} + S_{gk-1}} \quad eq.29$$

donde S_{gk} es el área cerrada por la amplitud pico a pico de la fuente glótica en el ciclo k -ésimo:

$$S_{gk} = \frac{1}{N_k} \sum_{i=n_{k-1}}^{n_k} s_{gk}(n-i) \quad eq.30$$

La *agudeza del cierre* se define para un punto de cierre dado en $n=n_k$:

$$p_{5k} = \frac{s_{gi}(n_k) - (s_{gi}(n_k - n_w) + s_{gi}(n_k + n_w)) / 2}{2n_w + 1} \quad eq.31$$

donde $2n_w+1$ es el tamaño de una ventana temporal alrededor del punto de cierre.

La *relación cubierta/cuerpo* se estima como:

$$p_{6k} = \frac{\sum_{n \in N_k} (s_{wk}(n))^2}{\sum_{n \in N_k} (s_{sk}(n))^2} \quad eq.32$$

donde $s_{sk}(n)$ y $s_{wk}(n)$ son la onda acústica promedio y el correlato de onda mucosa, respectivamente.

Proceso (21). Detector de la Onda Acústica Promedio

La onda acústica promedio es una señal de carácter semi-sinusoidal que tiene la duración de un arco glótico, y cuya amplitud cumple un determinado criterio, como se comenta a continuación. Su estimación se realiza para cada ciclo de fonación, y en forma síncrona con el principio y fin de éste, definido desde mínimo a mínimo de la fuente glótica (clipping) como se muestra en (42) o (45), de modo que siendo k el índice del ciclo de fonación de que se trate tendremos las siguientes definiciones:

$$s_{wk}(n) = s_{gk}(n) - s_{sk}(n) \quad eq.33$$

$s_{sk}(n)$ y $s_{mk}(n)$ son la onda acústica promedio y el correlato de onda mucosa. La onda acústica promedio es una semi-sinusoide de semiperíodo igual a la duración del ciclo de fonación T_{ck} :

$$s_{gk}(n) = s_{0k} \sin(\omega_k n \tau); \quad n \in N_k \quad eq.34$$

siendo la pulsación correspondiente:

$$\omega_k = \frac{2\pi}{T_{ck}} \quad eq.35$$

La amplitud de la semi-sinusoide representativa de la onda acústica promedio se evalúa minimizando la energía del correlato de la onda mucosa:

$$L_k = \sum_{n \in N_k} (s_{wk}(n))^2 = \sum_{n \in N_k} (s_{gk}(n) - s_{sk}(n))^2 \quad eq.36$$

respecto a dicha amplitud:

$$\frac{\partial L_k}{\partial s_{0k}} = 0 \Rightarrow s_{0k} = \frac{\sum_{n \in N_k} s_{gk} \text{sen}(\omega_k n \tau)}{\sum_{n \in N_k} \text{sen}^2(\omega_k n \tau)} \quad eq.37$$

Consecuentemente la derivada del correlato de la onda mucosa se puede estimar como:

$$s_{dk}(n) = \frac{s_{wk}(n) - s_{wk}(n - l)}{\tau} \quad eq.38$$

si se utiliza la regla rectangular por la izquierda.

Proceso (22). Estimador de la envolvente de la Densidad Espectral de Potencia del Correlato de la Onda Mucosa

La envolvente de la densidad espectral de potencia del correlato de onda mucosa se define como el módulo de la Transformada de Fourier de un ciclo de onda mucosa, esto es:

$$S_{wk}(m) = \left| \sum_{n \in N_k} s_{wk}(n) e^{j \frac{2\pi mn}{N_k}} \right| \quad eq.39$$

La Figura 8 muestra un ejemplo de dicha estimación, con los principales parámetros biométricos derivados de las singularidades de la envolvente.

Proceso (23). Estimador de los parámetros biométricos de la onda glótica: parámetros cepstrales y singularidades de la envolvente de la densidad espectral de potencia

La definición del conjunto de parámetros biométricos de la onda glótica incluye tres tipos de parámetros. El primer juego de parámetros resulta de evaluar los coeficientes cepstrales del correlato de onda mucosa ciclo a ciclo, según la definición que sigue:

$$C_{wk}(q) = \left| \sum_{m \in W_k} \log(S_{wk}(m)) e^{j \frac{2\pi mq}{W_k}} \right| \quad eq.40$$

donde W_k es el tamaño de la ventana definida en el dominio de la frecuencia sobre la densidad espectral de potencia del correlato de onda mucosa, y q es el índice de selección del parámetro cepstral correspondiente. La asignación paramétrica resultante es la siguiente:

$$\begin{aligned}
 p_{7k} &= C_{wk}(1); & p_{8k} &= C_{wk}(2); & p_{9k} &= C_{wk}(3); \\
 p_{10k} &= C_{wk}(4); & p_{11k} &= C_{wk}(5); & p_{12k} &= C_{wk}(6); \\
 p_{13k} &= C_{wk}(7); & p_{14k} &= C_{wk}(8); & p_{15k} &= C_{wk}(9); \\
 p_{16k} &= C_{wk}(10); & p_{17k} &= C_{wk}(11); & p_{18k} &= C_{wk}(12); \\
 p_{19k} &= C_{wk}(13); & p_{20k} &= C_{wk}(14);
 \end{aligned}
 \tag{eq.41}$$

El segundo juego de parámetros resulta de evaluar los valores de las singularidades del perfil de la densidad espectral de potencia del correlato de onda mucosa según se define en la Figura 8, que se expresan como:

$$\begin{aligned}
 p_{21k} &= \max \{20 \log_{10} [S_{wk}(m)]\}; \\
 p_{22k} &= \min \{20 \log_{10} [S_{wk}(M_1 < m < M_2)]\} - p_{21k}; \\
 p_{23k} &= \max \{20 \log_{10} [S_{wk}(m > M_1)]\} - p_{21k}; \\
 p_{24k} &= \min \{20 \log_{10} [S_{wk}(M_2 < m < M_3)]\} - p_{21k}; \\
 p_{25k} &= \max \{20 \log_{10} [S_{wk}(m > M_2)]\} - p_{21k}; \\
 p_{26k} &= 20 \log_{10} [S_{wk}(m = N_k/2)] - p_{21k}; \\
 p_{27k} &= \arg \max \{20 \log_{10} [S_{wk}(m)]\}; \\
 p_{28k} &= \frac{\arg \min \{20 \log_{10} [S_{wk}(M_1 < m < M_2)]\}}{p_{27k}}; \\
 p_{29k} &= \frac{\arg \max \{20 \log_{10} [S_{wk}(m > M_1)]\}}{p_{27k}}; \\
 p_{30k} &= \frac{\arg \min \{20 \log_{10} [S_{wk}(M_2 < m < M_3)]\}}{p_{27k}}; \\
 p_{31k} &= \frac{\arg \max \{20 \log_{10} [S_{wk}(m > M_2)]\}}{p_{27k}}; \\
 p_{32k} &= \frac{N_k}{2p_{27k}};
 \end{aligned}
 \tag{eq.42}$$

donde M_1 , M_2 y M_3 son los argumentos de los tres primeros máximos de la densidad espectral de potencia del correlato de onda mucosa expresada en decibelios.

El tercer juego de parámetros incluye las relaciones de aspecto de los dos primeros mínimos del perfil de la densidad espectral de potencia del correlato de la onda mucosa, definidos como:

$$\begin{aligned}
 p_{33k} &= \frac{2S_{wkdB}(m_1) - S_{wkdB}(M_1) - S_{wkdB}(M_2)}{2(M_2 - M_1)}; \\
 p_{34k} &= \frac{2S_{wkdB}(m_2) - S_{wkdB}(M_2) - S_{wkdB}(M_3)}{2(M_3 - M_2)}
 \end{aligned}
 \tag{eq.43}$$

donde S_{wkdB} es la densidad espectral de potencia referida en decibelios.

Proceso (24). Estimador de los parámetros biomecánicos del cuerpo del pliegue vocal

Se pueden obtener estimaciones fiables de los valores relativos de las masas y tensiones elásticas de los pliegues vocales a partir de la densidad espectral de potencia de la onda acústica promedio:

$$S_{sk}(m) = \left| \sum_{n \in N_k} s_{sk}(n) e^{j \frac{2\pi mn}{N_k}} \right| \quad eq.44$$

La técnica de estimación se basa en el ajuste adaptativo de la densidad espectral de potencia de la onda acústica promedio contra la función de transferencia del modelo de pliegue vocal de una masa. La hipótesis de trabajo se basa en la suposición de que la onda acústica promedio está determinada por las componentes dinámicas del pliegue, y por lo tanto, su densidad espectral de potencia está directamente relacionada con el módulo al cuadrado de la admitancia del modelo electromecánico de una masa dada por:

$$G_b(\omega) = |Y_b|^2 = \left| \frac{V_x(\omega)}{F_x(\omega)} \right|^2 = \left[(\omega M_b - \omega^{-1} K_b)^2 + R_b^2 \right]^{-1} \quad eq.45$$

$$S_{dk}(m) \approx G_b(\omega = m\Omega);$$

donde M_b , K_b y R_b son respectivamente los parámetros asociados a la masa dinámica, la elasticidad y las pérdidas del modelo de una masa cuando se toma solo el cuerpo de la misma en consideración. La estimación robusta de los parámetros del modelo se basa en la determinación de dos puntos sobre la densidad espectral de potencia de la componente dinámica tales como $\{G_{b1}, \omega_1\}$ y $\{G_{b2}, \omega_2\}$. Los Parámetros Biomecánicos de la Fuente Glótica se estiman aproximando la Densidad Espectral de Potencia de la Fuente Glótica mediante la Función de Transferencia de un sistema RLC serie cuyos elementos de circuito $-M_b$, K_b y R_b - son seleccionados por los procedimientos que se describen a continuación.

a. Estimación del parámetro de Pérdidas

El parámetro de pérdidas del cuerpo se estima como

$$R_b = \frac{F_{b0}}{\sqrt{G_r}} \quad eq.46$$

donde G_r es el valor del módulo al cuadrado de la admitancia de entrada dada por eq. 45 a la frecuencia de resonancia ω_r determinada por el primer máximo de la densidad espectral de potencia de la fuente glótica.

b. Estimación del parámetro de Masa

La masa dinámica equivalente del cuerpo de la cuerda puede estimarse como:

$$M_b = \frac{\omega_2}{\omega_2^2 - \omega_1^2} \sqrt{\frac{G_{b1} - G_{b2}}{G_{b1} G_{b2}}} \quad eq.47$$

La selección de los puntos de referencia más adecuados $\{T_{b1}, \omega_1\}$ y $\{T_{b2}, \omega_2\}$ está fuertemente relacionada con la robustez del método de estimación.

c. Estimación del parámetro de Elasticidad

Una vez determinado el parámetro de masa dinámica, el parámetro de rigidez elástica K_b puede obtenerse de la determinación precisa del pico máximo $\{T_r, \omega_r\}$, como:

$$K_b = M_b \omega_r^2 \quad eq.48$$

d. Desbalance de los parámetros biométricos

Los pliegues vocales son asimétricos desde un punto de vista anatómico, fisiológico y biomecánico, tanto en individuos normofónicos (aquellos calificados como libres de disfunción por especialistas en foniatría o logopedia previa inspección y trazado de su historial) como disfónicos (aquellos a quienes se ha diagnosticado una determinada alteración de la fonación como resultado de causa orgánica o funcional), aunque posiblemente en mayor medida en ciertas disfonías que en otras. Esta asimetría es observable como un desbalance de los parámetros biomecánicos estimados para ciclos de fonación vecinos. Este desbalance será más grande en casos donde la patología fisiológica del pliegue vocal esté presente, sobre todo si afecta de forma diferencial a ambos pliegues, como en el caso de quistes o pólipos unilaterales, por ejemplo. El desbalance de la vibración de los pliegues vocales debe corresponderse con un desbalance en las estimaciones de los parámetros biomecánicos de un individuo dado cuando se comparen ciclo a ciclo. Se acepta generalmente que la presencia de desbalance es un correlato de la patología del pliegue vocal, y que este desbalance se halla estimado en parámetros de distorsión como el *jitter* y el *shimmer*. El desbalance entre ciclos de fonación vecinos puede apreciarse en (42) y (49), ya que aunque procedentes de individuos supuestamente normofónicos, se constata la diferencia en amplitud entre ciclo y ciclo, y si bien menos perceptible, también en su duración. Por todo lo expuesto, resulta de sumo interés recoger las variaciones inter-ciclo de las estimaciones de los parámetros biomecánicos mediante medidas de desbalance de masa, tensión y pérdidas obtenidas para cada ciclo (μ_b : Desbalance de la Masa del Cuerpo; σ_b : Desbalance de las Pérdidas del Cuerpo; γ_b : Desbalance de la Tensión del Cuerpo), las cuales se pueden definir como:

$$\begin{aligned}\mu_{bk} &= (\hat{M}_{bk} - \hat{M}_{bk-1}) / (\hat{M}_{bk} + \hat{M}_{bk-1}) \\ \rho_{bk} &= (\hat{R}_{bk} - \hat{R}_{bk-1}) / (\hat{R}_{bk} + \hat{R}_{bk-1}) \\ \gamma_{bk} &= (\hat{K}_{bk} - \hat{K}_{bk-1}) / (\hat{K}_{bk} + \hat{K}_{bk-1})\end{aligned}\quad eq.49$$

donde $1 \leq k \leq K$ es el índice del ciclo de fonación, y $\hat{M}_{bk}$, $\hat{R}_{bk}$, y $\hat{K}_{bk}$ son los estimados de la masa, las pérdidas y la tensión para el ciclo k-ésimo de una muestra de voz procedente de un individuo dado. Dado que el parámetro de interelasticidad $K_{bl,r}$ no se incluye en la lista de parámetros biomecánicos habitualmente, si se considera que los pliegues son totalmente simétricos, bastará con calcular tres parámetros por pliegue (masa, elasticidad y pérdidas del cuerpo, e ídem de cubierta), y sus tres desbalances, hasta completar seis parámetros biomecánicos de cuerpo de pliegue vocal. El cuadro de asignación de parámetros queda como sigue:

$$\begin{aligned}p_{35k} &= M_{bk}; & p_{36k} &= R_{bk}; & p_{37k} &= K_{bk}; \\ p_{38k} &= \mu_{bk}; & p_{39k} &= \rho_{bk}; & p_{40k} &= \gamma_{bk};\end{aligned}\quad eq.50$$

Proceso (25). Estimador de los parámetros biomecánicos de la cubierta del pliegue vocal

Similarmente al caso de los parámetros biomecánicos del cuerpo del pliegue vocal se pueden estimar los parámetros biomecánicos de su cubierta a partir de la densidad espectral de potencia del correlato de la onda mucosa, calculado según eq. 39, en forma totalmente similar a lo descrito para (24) en referencia a los parámetros del cuerpo del pliegue, mediante el ajuste de la Función de Transferencia de un sistema de segundo orden cuyos elementos de circuito $-M_c$, K_c y R_c son seleccionados por el mismo procedimiento que en (24). Para los parámetros biomecánicos de la cubierta del pliegue vocal se utilizan derivaciones similares a partir del correlato de la onda mucosa, ya que la influencia del cuerpo de la cuerda se ha eliminado al separar la onda acústica promedio de la fuente glótica, reduciendo el problema al modelo de una sola masa, lo que facilita la aplicación de la misma metodología. Por lo que respecta a los parámetros de desbalance (μ_c : Desbalance de la Masa de la Cubierta; σ_c : Desbalance de las Pérdidas de la Cubierta; γ_c : Desbalance de la Rigidez de la Cubierta), su estimación es asimismo idéntica. La asignación de parámetros resultante es la siguiente:

$$\begin{aligned}p_{41k} &= M_{ck}; & p_{42k} &= R_{ck}; & p_{43k} &= K_{ck}; \\ p_{44k} &= \mu_{ck}; & p_{45k} &= \rho_{ck}; & p_{46k} &= \gamma_{ck};\end{aligned}\quad eq.51$$

Descripción breve de la utilidad de los diferentes parámetros

La utilidad de los diferentes parámetros evaluados mediante el proceso de estimación referido en la Figura 3 es la siguiente, sin excluir otras posibles:

Parámetro p_{1k} . Es el período del ciclo glótico, inversa de la frecuencia fundamental. Sirve, con otros, para distinguir voz masculina de femenina.

Parámetro p_{2k} . Es el *jitter* dado en la eq. 27. Sirve, junto con otros, para detectar inestabilidad en la fonación, y ayuda a caracterizar la disfonía (aplicación en detección y gradación de disfonía).

Parámetro p_{3k} . Es el *shimmer* de amplitud dado en la eq. 28. Sirve, junto con otros, para detectar inestabilidad en la fonación, y ayuda a caracterizar la disfonía (aplicación en detección y gradación de disfonía).

Parámetro p_{4k} . Es el *shimmer* de área dado en la eq. 29. Sirve, junto con otros, para detectar inestabilidad en la fonación, y ayuda a caracterizar la disfonía (aplicación en detección y gradación de disfonía).

Parámetro p_{5k} . Es la agudeza del cierre glótico dada en la eq. 31. Sirve, junto con otros, para detectar emocionalidad en la fonación, y ayuda a caracterizar la disfonía (aplicación en detección y gradación de disfonía).

Parámetro p_{6k} . Es la relación onda mucosa/onda glótica dada en la eq. 32. Sirve, junto con otros, para detectar posibles alteraciones neurológicas en un locutor, y ayuda a caracterizar la disfonía (aplicación en detección y gradación de disfonía).

Parámetros cepstrales p_{7k} - p_{20k} . Forman parte de la firma biométrica del locutor en forma compacta según la eq. 41, y junto con otros ayudan en la identificación y verificación del locutor, tanto en aplicaciones de acceso seguro como en cotejo forense.

Parámetros de perfil espectral p_{21k} - p_{34k} . Forman parte de la firma biométrica del locutor según las eq. 42 y eq. 43, así como del comportamiento normofónico o disfónico de éste, y junto con otros ayudan en la identificación y verificación del mismo (aplicaciones de acceso seguro o cotejo forense), y en la determinación de la presencia de disfonía de origen orgánico (aplicación en detección y gradación de disfonía).

Parámetros biomecánicos p_{35k} - p_{46k} . Constituyen un conjunto robusto de descriptores del funcionamiento mecánico de la glotis según las eq. 50 y eq. 51, y junto con otros ayudan en la determinación de las posibles causas de una disfonía, y en la cuantificación del grado de la misma (aplicación en detección y gradación de disfonía).

Parámetros de base temporal p_{47k} - p_{58k} . Constituyen un descriptor robusto de los instantes de interés del ciclo glótico (cierre, retorno, apertura) según la eq. 24, y junto con otros sirven en la caracterización de la disfonía (aplicación en detección y gradación de disfonía).

Ejemplos de posibles realizaciones de la invención

Modo de realización 1

Sistema para la parametrización de los correlatos de la onda glótica y su uso clínico y forense y para estudios avanzados de la voz

Se integra el método de parametrización completo sobre una plataforma similar a la de la Figura 1 sin que se excluya asimismo una plataforma informática de propósito general, que permita registrar un segmento de habla de duración arbitraria, sobre la que el experto (usuario) puede ubicar segmentos vocálicos por inspección simple, sobre el cual se extraen los parámetros que el usuario seleccione en configuración. La interfaz permite analizar la trama deseada, y comparar sus resultados contra otra trama cualquiera previamente analizada, contra un Modelo Normofónico de Locutor, para la obtención de Detección y Gradación de Disfonía, o contra un Modelo Universal de Locutor para la obtención de Cotejo Forense. Los resultados pueden verse por ventanas en pantalla, presentarse en ventanas individuales, imprimirse como figuras .pdf, y almacenarse en hoja de cálculo Excel®.

Modo de realización 2

Sistema para la monitorización y evaluación de la eficiencia de fonación por un servicio especializado de Otorrinolaringología

Se integra el método de parametrización parcial sobre una plataforma similar a la de la Figura 1 sin que se excluya asimismo una plataforma informática de propósito general, que registre un segmento de vocal /a/ de 0,2 seg. sobre el cual extraigan los parámetros p_{1k} - p_{58k} y represente los mismos sobre una interfaz de usuario incluyendo los intervalos de normalidad para dichos parámetros, con objeto de que el facultativo valore la calidad de la fonación.

Modo de realización 3

Aplicación para cribado de pacientes en centros de asistencia primaria

Se integra el método de parametrización sobre una plataforma similar a la de la Figura 1 sin que se excluya asimismo una plataforma informática de propósito general, o en dispositivo portable tipo teléfono móvil, pda o ipod, con micrófono simple, que registra un segmento de voz, y realiza varias parametrizaciones sobre segmentos contiguos en el centro de la trama capturada, representando éstos en modo semáforo según el contraste de los mismos sobre una población normada (ver Figura 11) en una interfaz de usuario reducida para que el médico de asistencia primaria pueda determinar la conveniencia o no de remitir al paciente a los servicios de especialidad. Esta circunstancia se

complementa mediante la generación de un documento electrónico en formato .pdf, una copia del cual se remite al servicio de especialidad ORL, y otra copia se proporciona al paciente.

Aplicación industrial

5

Aplicación a la detección y gradación de disfonía para el desarrollo de un modelo de asistencia primaria en la patología de la voz

Se enmarca en el contexto de la relación entre los Centros Médicos de Asistencia Primaria y los Servicios de Atención Especializada en Otorrinolaringología. La detección y gradación de la disfonía se puede llevar a cabo en una interfaz muy sencilla similar a la descrita en la Figura 1 (10), siguiendo el procedimiento de análisis que se fija en la Figura 11. Los parámetros estimados a partir de la fuente glótica para una población normofónica previamente evaluada por los Servicios Especializados ORL y almacenados en una Base de Datos con Modelos de Locutores se utilizan para construir un Modelo de Locutor Normofónico (59) para hombres, y otro para mujeres, en un margen de edad entre 18-60 años. Una toma de voz (52, 54) realizada con la interfaz de la Figura 1 (11) es automáticamente contrastada contra el Modelo de Locutor Normofónico (59), obteniéndose un análisis de contraste de un conjunto de parámetros contra la estadística del Modelo de Locutor Normofónico (61). Si los parámetros evaluados para el sujeto bajo inspección se despliegan fuera del ámbito de normalidad, se colorea un semáforo por cada parámetro (63). Con esta información visual el Médico de Atención Primaria puede tomar la decisión de derivar o no al paciente a los Servicios de Asistencia Especializada para su inspección y tratamiento. Esta función se denomina “cribado de pacientes” y está orientada a aumentar la eficacia de los Servicios Especializados evitando inspecciones innecesarias, ahorrando costes y tiempo al personal especializado.

25

Aplicación a la identificación y verificación del locutor para acceso seguro a sistemas e instalaciones

Esta aplicación permite conceder o denegar privilegios de acceso a personas por medio de su firma de voz, por medio de una interfaz como la que se muestra en la Figura 1 (10), siguiendo la descripción metodológica que se aporta en la Figura 12. En la misma a partir de la señal de voz (64, 66) se extraen parámetros del espectro de Fourier (FFT), y de predicción lineal (LPC) que se utilizan para detectar el mensaje generado por el locutor (p. ej. su nombre o un pin), así como para detectar su firma biométrica (68-73). Ésta se mezcla con la firma bio-metro-mecánica obtenida de la fuente glótica para dar lugar a un supervector (74), cuyo secuenciado se analiza para derivar el mensaje impreso en el mismo (75 Parsing HMM) y se contrasta en biometría contra la Base de Datos con Modelos de Locutores (76 Clustering GMM). Los valores del análisis se combinan para dar una puntuación (78 Fusión de Scores) que se utiliza para determinar la identidad del locutor de entre un conjunto cerrado (77), y otorgar una certeza a dicha identidad (79). En función de estos parámetros se decide conceder o denegar el acceso (80 Aceptación, Rechazo) o solicitar una nueva evaluación por voz o por otra biometría multimodal (Evaluación Alternativa).

Aplicación a la verificación y cotejo forenses para la evaluación de evidencia

Esta aplicación se basa en la interfaz (10) de la Figura 1, según la cual se puede analizar una voz pregrabada (82), o grabar una nueva (81, 83), la cual puede ser contrastada contra un modelo universal de locutor (88) previamente generado (Modelo Universal de Locutor) según la metodología expuesta en la Figura 13. El resultado de la evaluación (90 Scores LR) se contrasta contra una escala (91 Evaluación de Evidencia) que para valores por debajo de 0 favorece la Hipótesis de la Defensa (de que no hay evidencia plausible que permita vincular la voz procesada -dubitada- contra otra voz previamente grabada cuya identidad es conocida -indubitada- con el aval de la semejanza o disparidad de las mismas respecto al Modelo de Locutor Universal de que disponga la base de datos). Entre 0 y 2 la evidencia es débil (ED) o fuerte (EF), si bien no ratifica suficientemente la Hipótesis de la Acusación (de que hay evidencia plausible de vinculación entre debitada e indubitada). En estos casos se aplica el principio de in dubio pro reo. Finalmente si la evidencia está por encima de 2 se considera que la evidencia es muy fuerte (EMF) a favor de la Hipótesis de la Acusación.

Aplicación a la detección del temblor en voz para la detección precoz de la patología neurológica, y para la carga emocional del locutor

Esta aplicación se basa en la misma plataforma que la descrita en la Figura 11 para la aplicación de detección y gradación de la disfonía, siguiendo las mismas pautas metodológicas. La diferencia fundamental es que para generar las Bases de Datos de Modelos de Locutores (59), y para el contraste de la voz de un paciente contra las mismas, sólo se tienen en cuenta ciertos parámetros del conjunto completo, como son p_{k5} , p_{k6} , p_{k7} y p_{k8} junto con p_{k37} y p_{k43} , ya que se ha podido determinar que estos parámetros, y no otros, presentan una alta correlación con el deterioro neurológico del locutor y la alteración de su estado emocional según estudios previos llevados a cabo por los proponentes de la presente solicitud. Los parámetros a configurar para una aplicación de este método a Centros de Asistencia Primaria, similares a la descrita en la Figura 1 serán los citados, siendo el tipo de evaluación similar al de la Figura 11. Las Bases de Datos de Modelos de Locutores se habrán desarrollado con voz de individuos libres de patología neurológica u orgánica, previamente seleccionados en un Servicio de Especialidad ORL/Neurología. El procedimiento de evaluación y decisión será por lo demás el mismo que para la disfonía de origen orgánico ya comentada. El valor del análisis será también el de evaluar el grado de deterioro neurológico (inexistente, leve, moderado o grave) con objeto de derivar o no al paciente a los Servicios Especializados.

REIVINDICACIONES

5 1. Método para la estimación de parámetros fisiológicos de la fonación a partir de una señal de voz **caracterizado** por que comprende:

- compensar la radiación de los labios en la señal de voz mediante la cancelación del polo de primer orden generado por dicha radiación en el espectro de la señal de voz (12),
- 10 - efectuar un filtrado inverso (13, 14, 15, 16) de al menos un segmento fonado en un ciclo de fonación para una señal de voz compensada, donde dicho filtrado inverso comprende a su vez:
 - modelar la inversión espectral de la señal de voz compensada (13, 14) para extraer la señal de voz desglo-
 - 15 - modelar la inversión espectral del tracto vocal para extraer la onda glótica sin influencia sustancial del tracto vocal y para obtener los correlatos de vibración de los pliegues vocales (15, 16).

20 2. Método de acuerdo con la reivindicación 1, **caracterizado** por que la etapa de compensar la radiación de los labios comprende además:

- modelar la inversión espectral (13) del tracto vocal mediante una pluralidad de filtros de celosías adaptativas y encadenables entre sí, dichos filtros configurados para dividir la señal de voz en dos señales (27, 28) sobre las que se
- 25 calcula la correlación cruzada (29, 30) entre las dos señales divididas desfasadas (27, 28) cancelando el efecto de la radiación en los labios y generando una señal libre de radiación (31).

30 3. Método de acuerdo con la reivindicación 1 ó 2, **caracterizado** por que la etapa de modelar la inversión espectral de una señal (14, 15) comprende además:

- implementar una pluralidad de filtros modelo-espejo (13, 14) configurados para estimar la correlación parcial (34) y eliminar dicha correlación parcial (38) de la señal glótica (37, 37') debida al tracto vocal generando nuevas
- 35 señales glóticas (39, 39') sin influencia sustancial del tracto vocal.

4. Método de acuerdo con una cualquiera de las reivindicaciones anteriores, **caracterizado** por que comprende

40 5. Método de acuerdo con la reivindicación 4, que comprende estimar al menos uno de los siguientes parámetros temporales (19) sobre la onda glótica $s_{qi}(n)$:

- inicio del ciclo glótico;
- 45 - instante de recuperación T_r ;
- instante de inicio de apertura de la cuerdas vocales T_o ;
- instante de máxima presión supraglótica T_m ;
- 50 - instante de inicio del cierre de los pliegues vocales T_c ;
- instante final del ciclo glótico con mínima presión supraglótica T_g ;
- 55 - coeficientes ODQ, CDQ y GEQ.

60 6. Método de acuerdo con la reivindicación 4 ó 5, que comprende estimar al menos uno de los siguientes parámetros de distorsión (20) sobre la onda glótica $s_{qi}(n)$:

- jitter,
- shimmer de amplitud,
- 65 - shimmer de área,
- agudeza del cierre,

- relación cubierta/cuerpo.

7. Método de acuerdo con una cualquiera de las reivindicaciones 4 a 6, **caracterizado** por que comprende estimar la onda acústica promedio $v_{qi}(n)$ para estimar al menos uno de los siguientes conjuntos de parámetros biométricos (22, 23) sobre la onda glótica $s_{qi}(n)$ mediante la detección de onda acústica promedio (21):

- densidad espectral de potencia del correlato de onda mucosa,

- coeficientes cepstrales del correlato glótico de onda mucosa,

- singularidades de la envolvente de la densidad espectral de potencia del correlato glótico de onda mucosa.

8. Método de acuerdo con una cualquiera de las reivindicaciones 4 a 7, **caracterizado** por que comprende estimar la onda acústica promedio $v_{qi}(n)$ para al menos uno de los siguientes parámetros biomecánicos (24, 25) sobre la onda glótica $s_{qi}(n)$ mediante la detección de onda acústica promedio (21):

- parámetro de pérdidas,

- parámetro de masa dinámica equivalente del cuerpo de la cuerda,

- parámetro de elasticidad,

- desbalances entre ciclos de fonación respecto de:

- la masa dinámica del cuerpo,

- las pérdidas del cuerpo,

- la elasticidad del cuerpo.

9. Sistema para la estimación de parámetros fisiológicos de la fonación a partir de una señal de voz que comprende:

- medios configurados para compensar la radiación de los labios en la señal de voz mediante la cancelación del polo de primer orden generado por dicha radiación en el espectro de la señal de voz,

- medios configurados para efectuar un filtrado inverso (13,14,15,16) de al menos un segmento fonado en un ciclo de fonación para una señal de voz compensada, donde dicho filtrado inverso comprende a su vez:

- medios configurados para modelar la inversión espectral de la señal de voz compensada (14) para extraer la señal de voz desglotalizada,

- medios configurados para modelar la inversión espectral del tracto vocal para extraer la onda glótica sin influencia del tracto vocal y para obtener los correlatos de vibración de los pliegues vocales (15, 16).

10. Sistema de acuerdo con la reivindicación 9, **caracterizado** por que los medios configurados para compensar la radiación de los labios comprenden además:

- medios configurados para modelar la inversión espectral (13) del tracto vocal comprendiendo a su vez una pluralidad de filtros de celosías adaptativas y encadenables entre sí, donde dichos filtros están configurados para dividir la señal de voz en dos señales (27, 28) sobre las que se calcula la correlación cruzada (29, 30) entre las dos señales divididas desfasadas (27, 28) cancelando el efecto de la radiación en los labios y generando una señal libre de radiación (31).

11. Sistema de acuerdo con la reivindicación 9 ó 10, *caracterizado* por que los medios configurados para modelar la inversión espectral de una señal (14, 15) comprenden además:

- una pluralidad de filtros modelo-espejo (13, 14) configurados para estimar la correlación parcial debida al tracto vocal (34) y eliminar dicha correlación parcial de la señal glótica (38).

12. Sistema de acuerdo con una cualquiera de las reivindicaciones 9 a 11, **caracterizado** por que las estimaciones se realizan sobre al menos un Modelo de Locutor Normofónico y son almacenadas en unos medios de almacenamiento para ser comparadas con las estimaciones de un locutor cualquiera para determinar la presencia y grado de la disfonía de acuerdo con la desviación existente entre ambas estimaciones.

13. Sistema de acuerdo con una cualquiera de las reivindicaciones 9 a 11, **caracterizado** por que las estimaciones de un locutor cualquiera se almacenan en unos medios de almacenamiento para identificar unívocamente a dicho locutor.

5

10

15

20

25

30

35

40

45

50

55

60

65

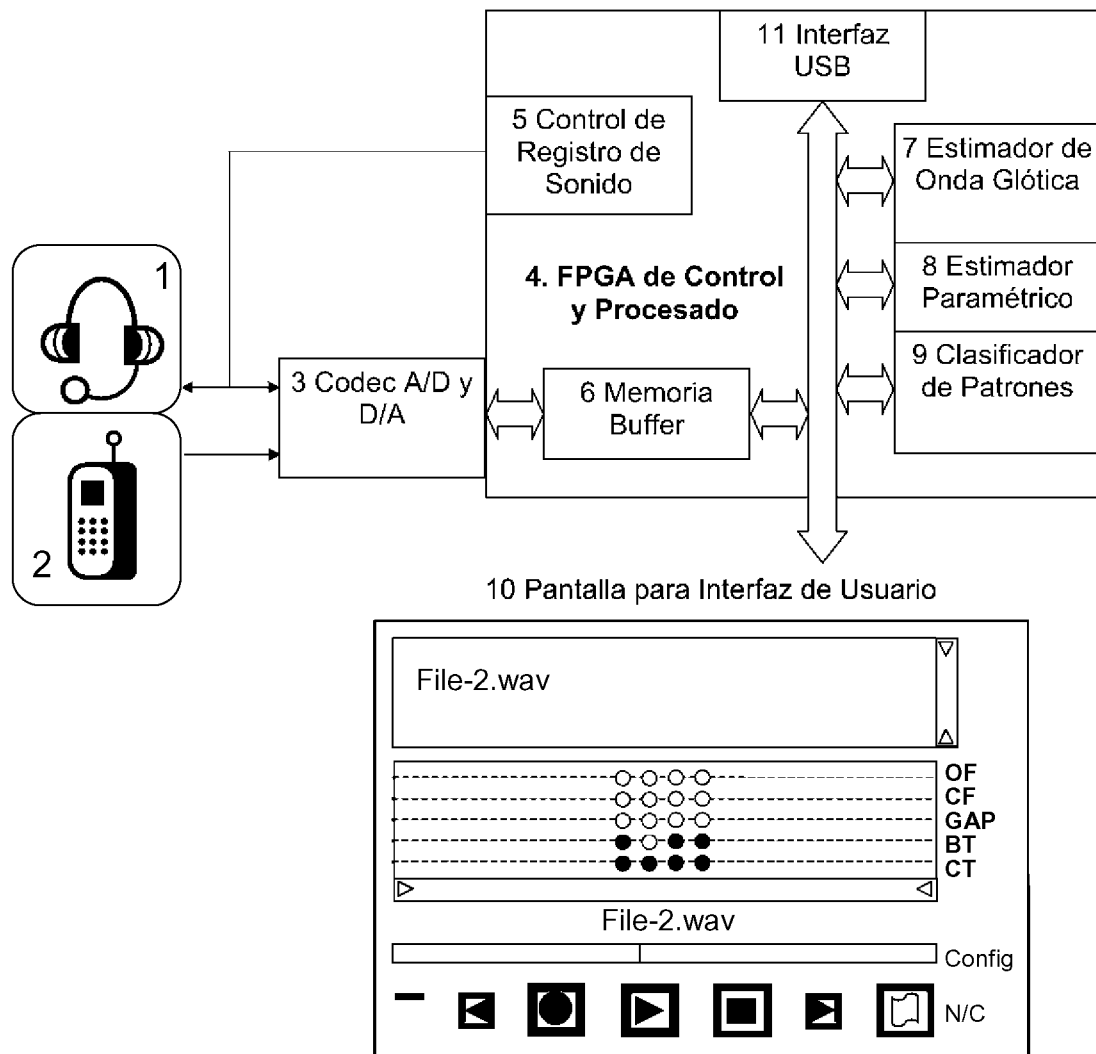


Fig. 1

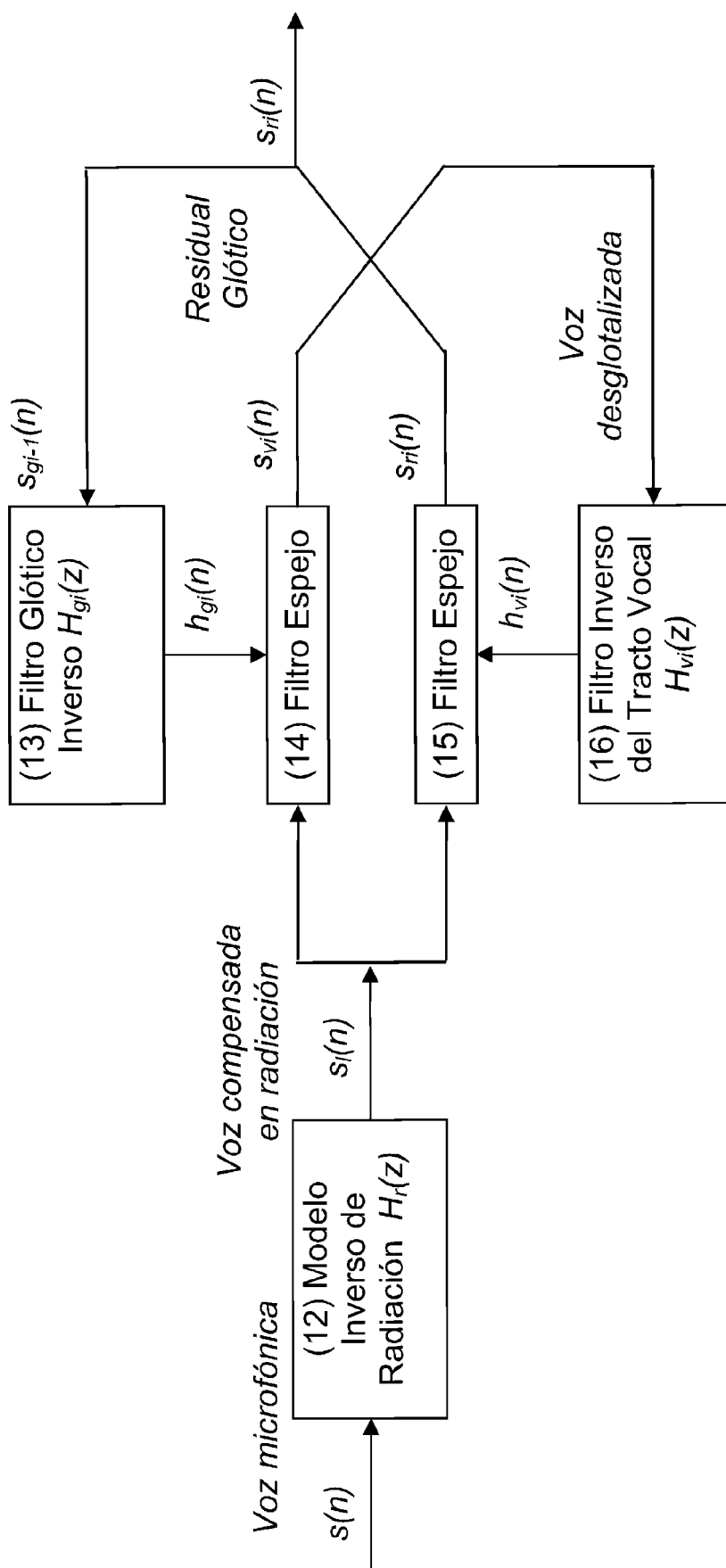


Fig. 2

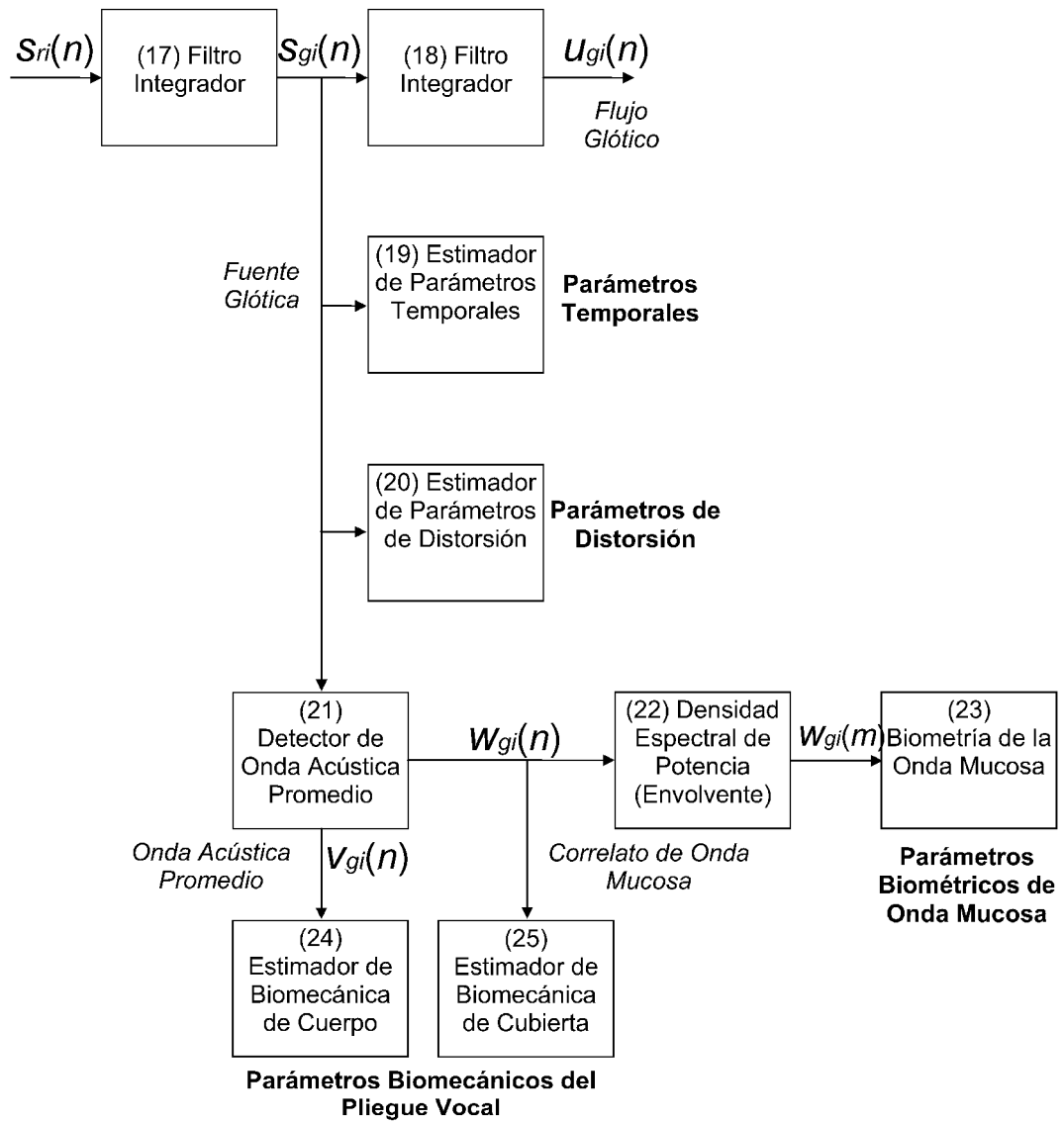


Fig. 3

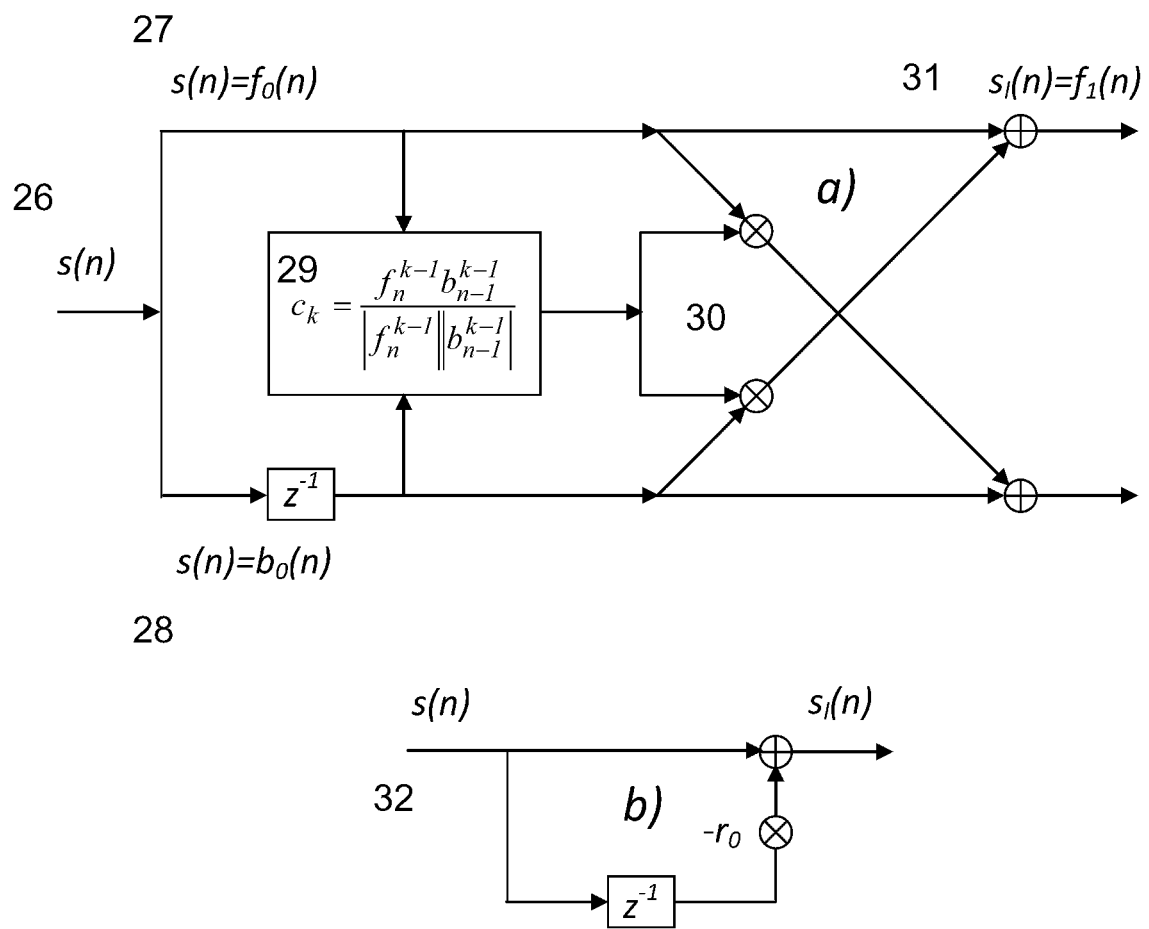


Fig. 4

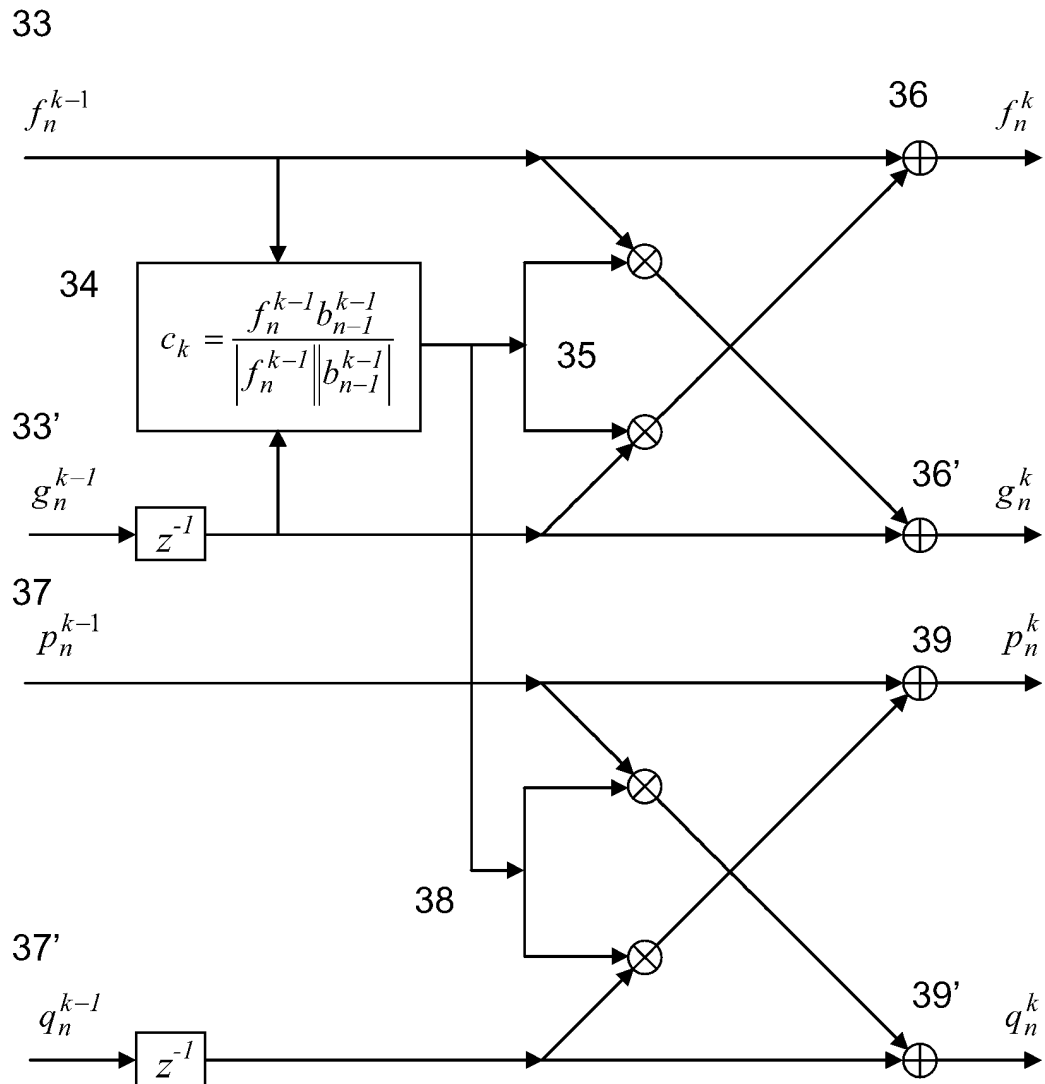


Fig. 5

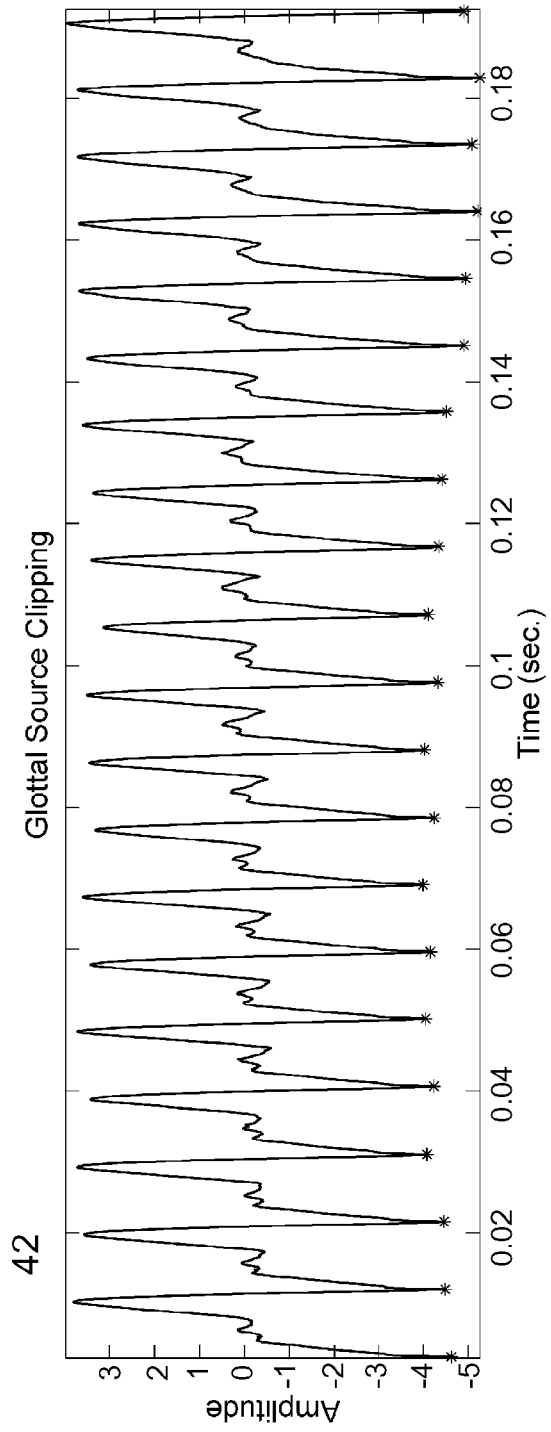
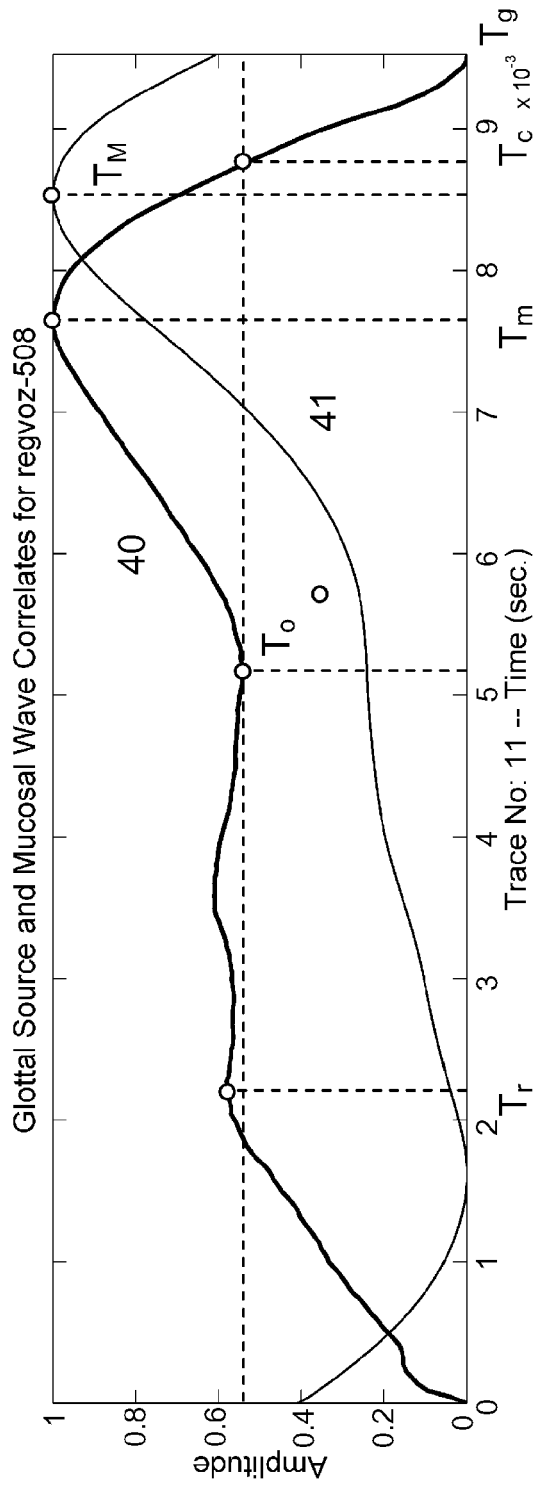


Fig. 6

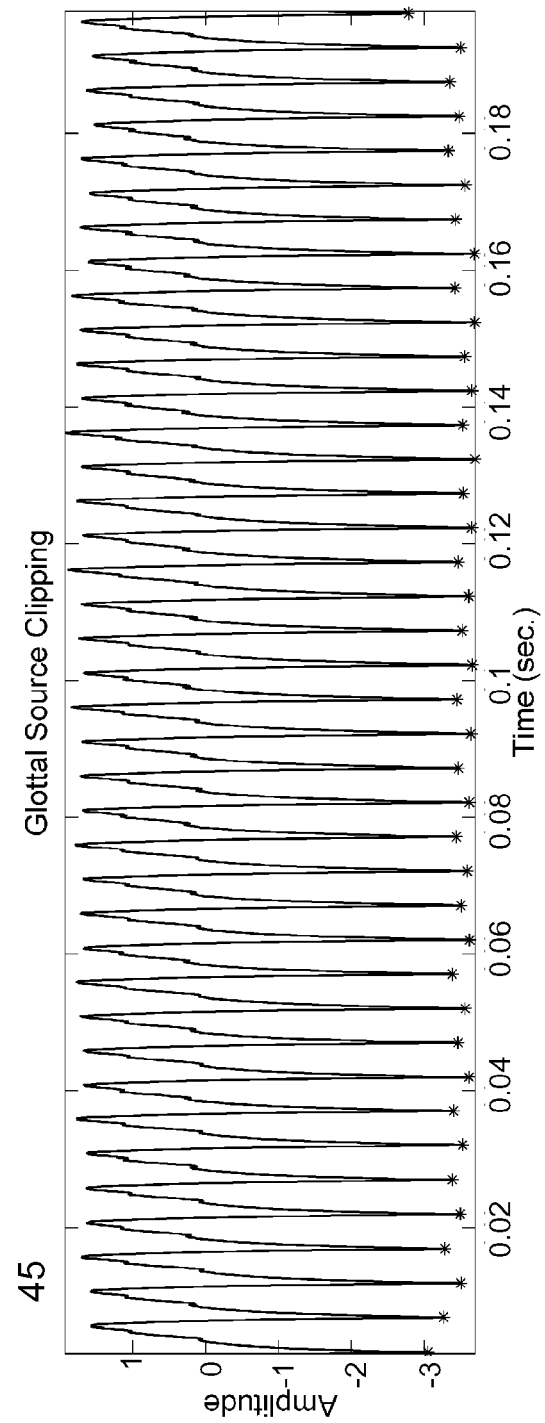
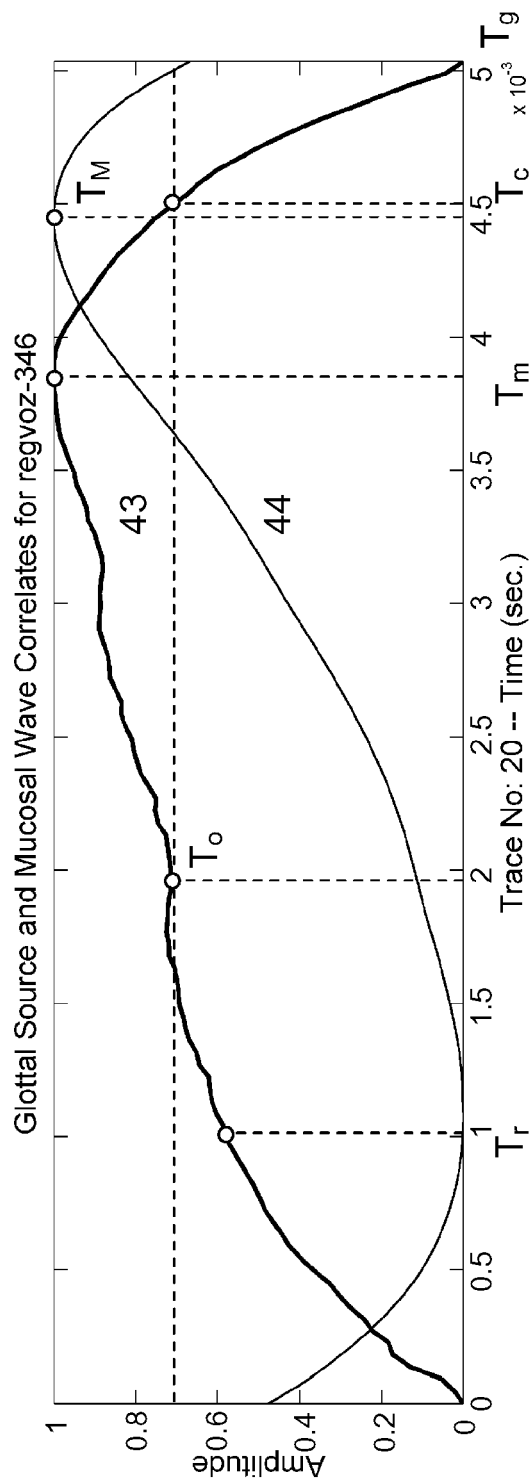


Fig. 7

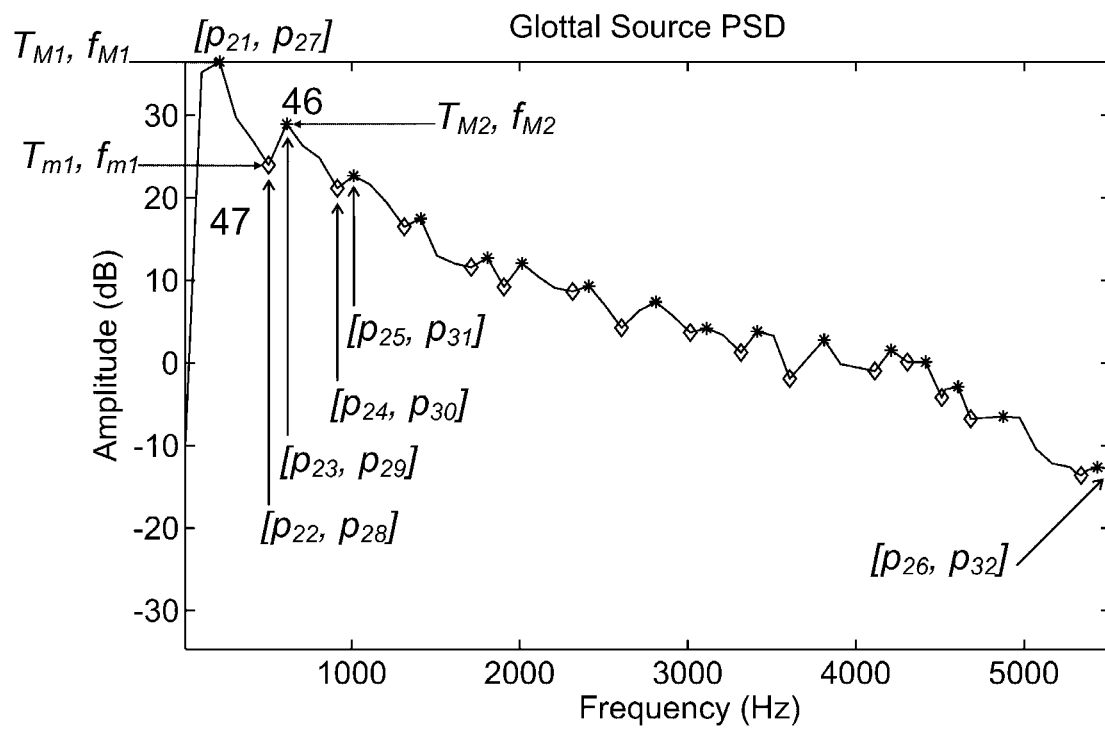


Fig. 8

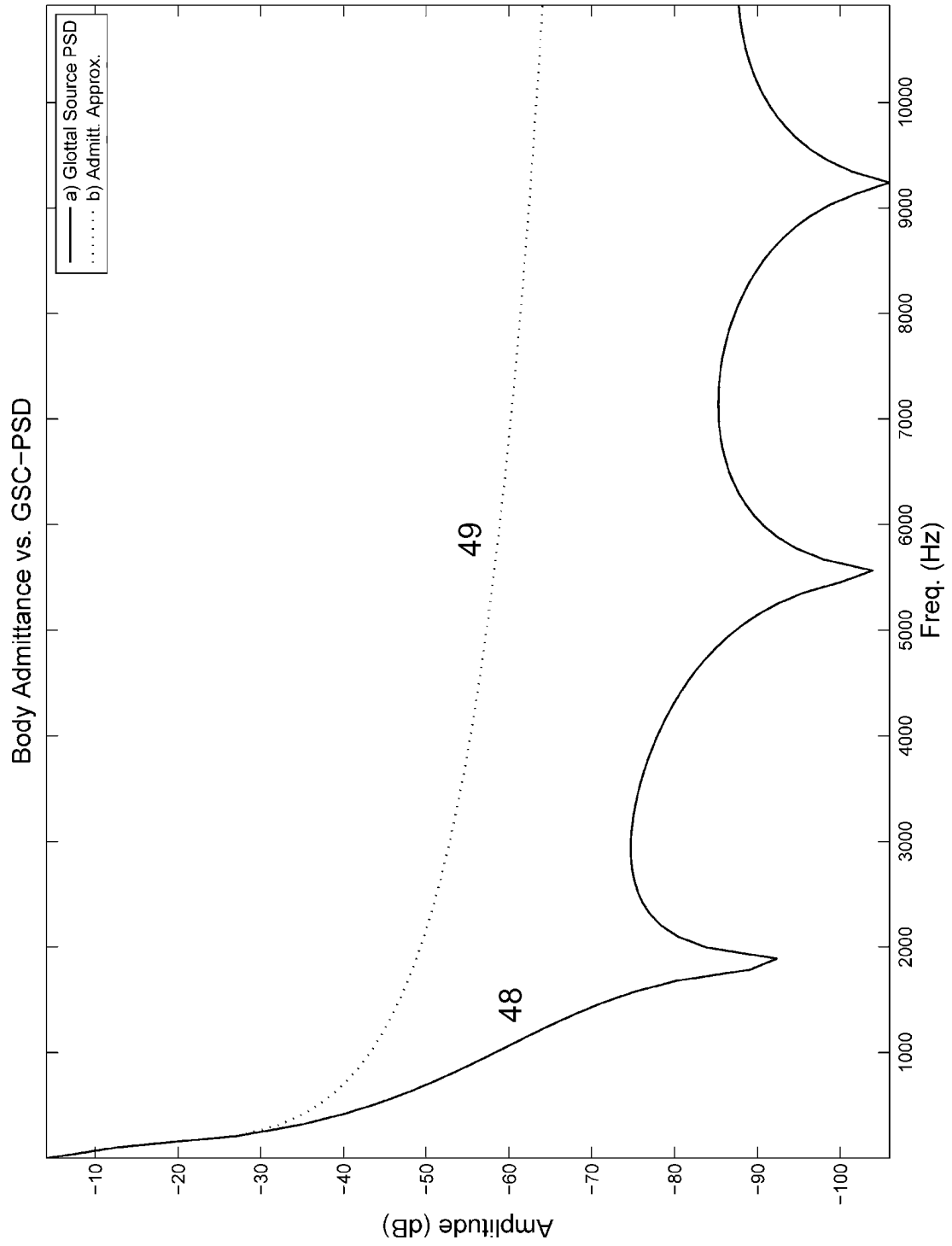


Fig. 9

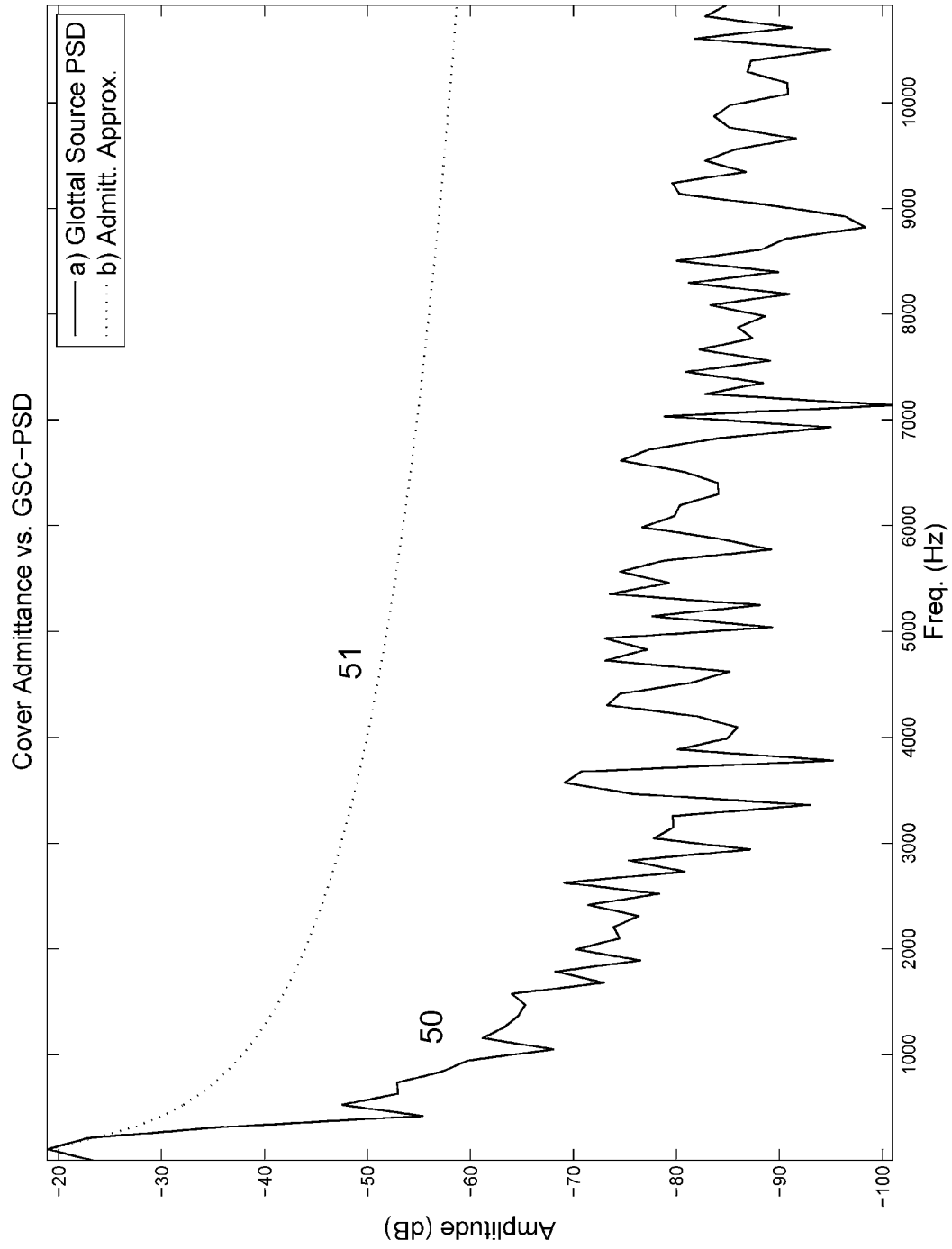


Fig. 10

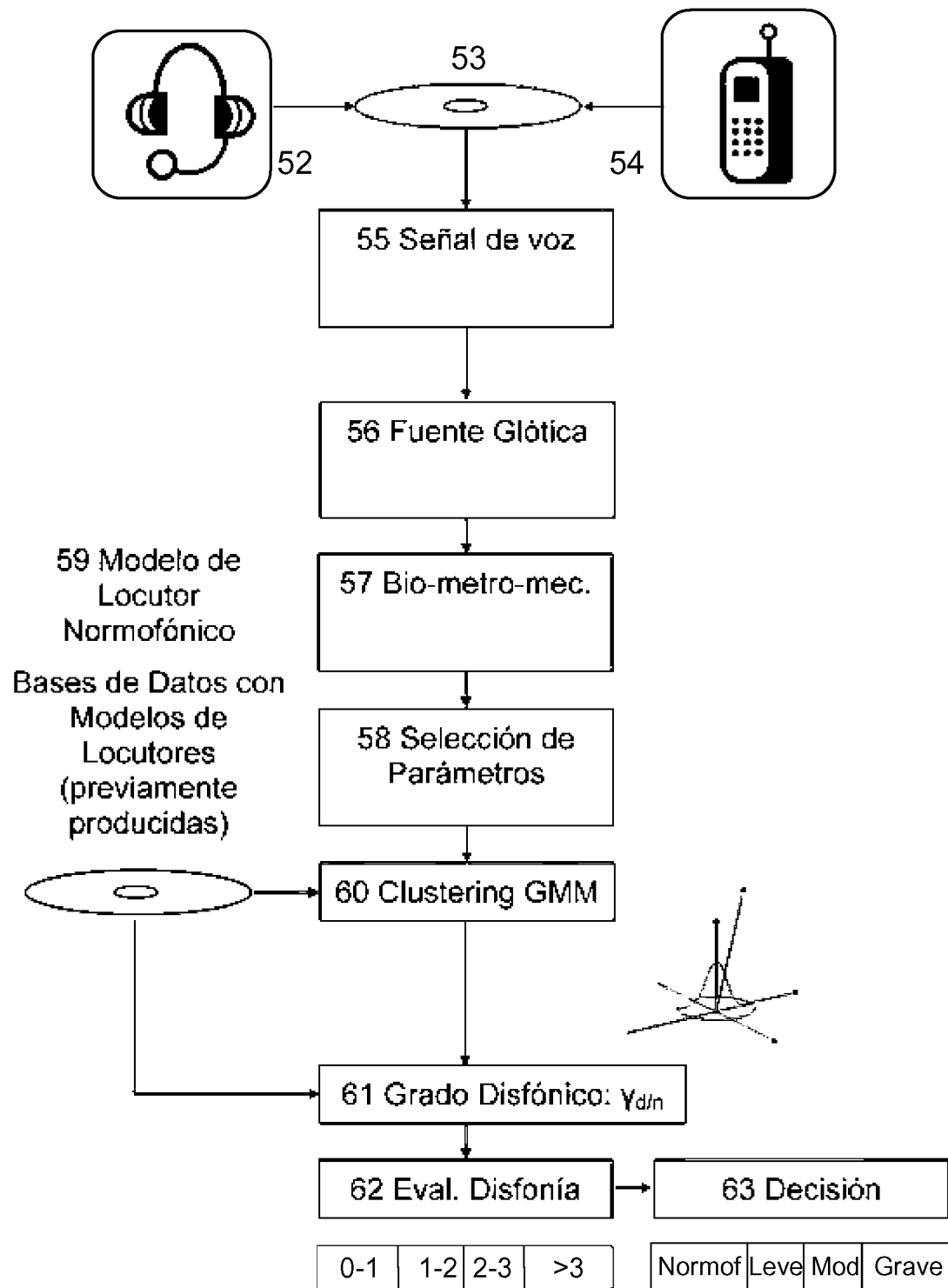


Fig. 11

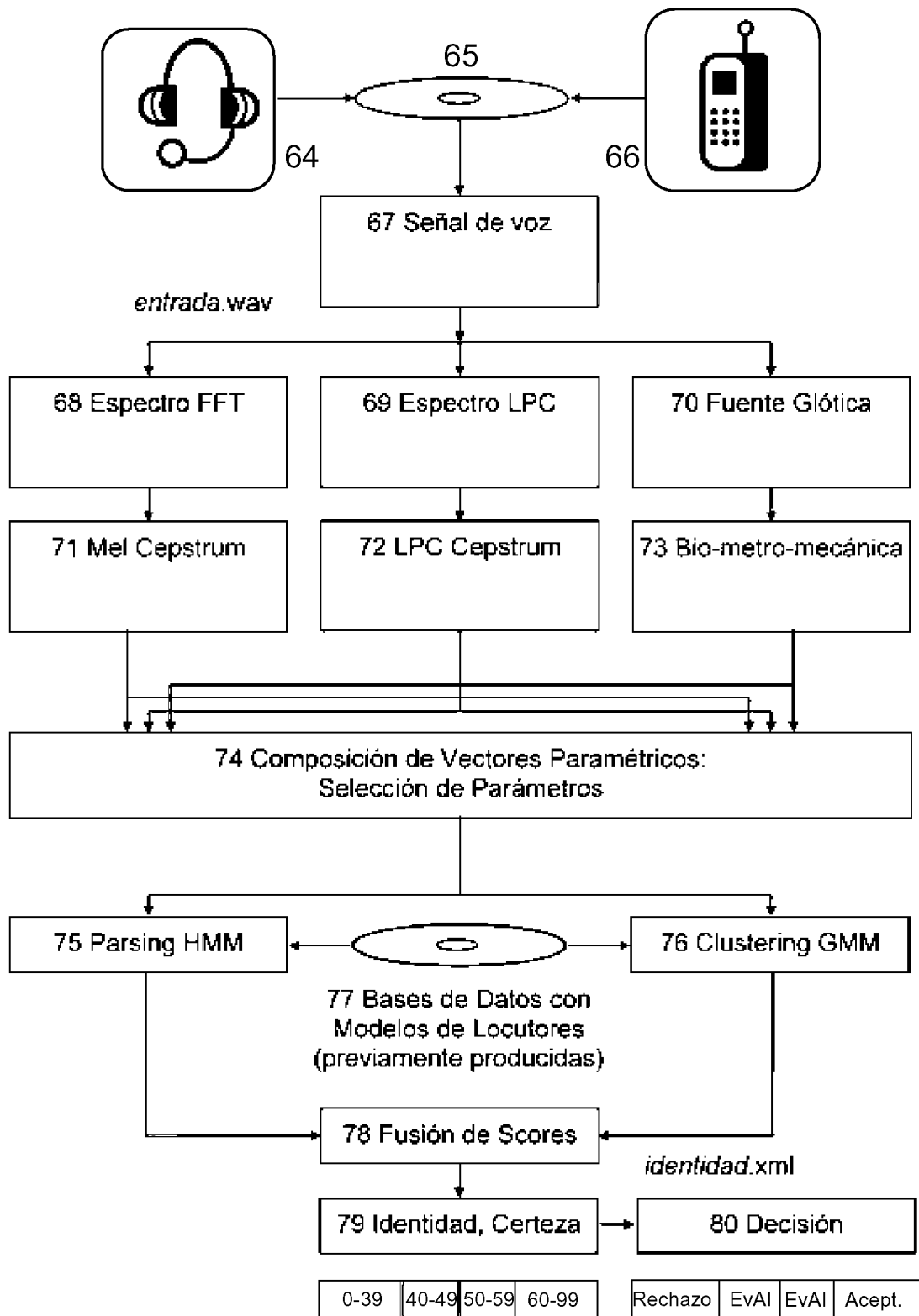


Fig. 12

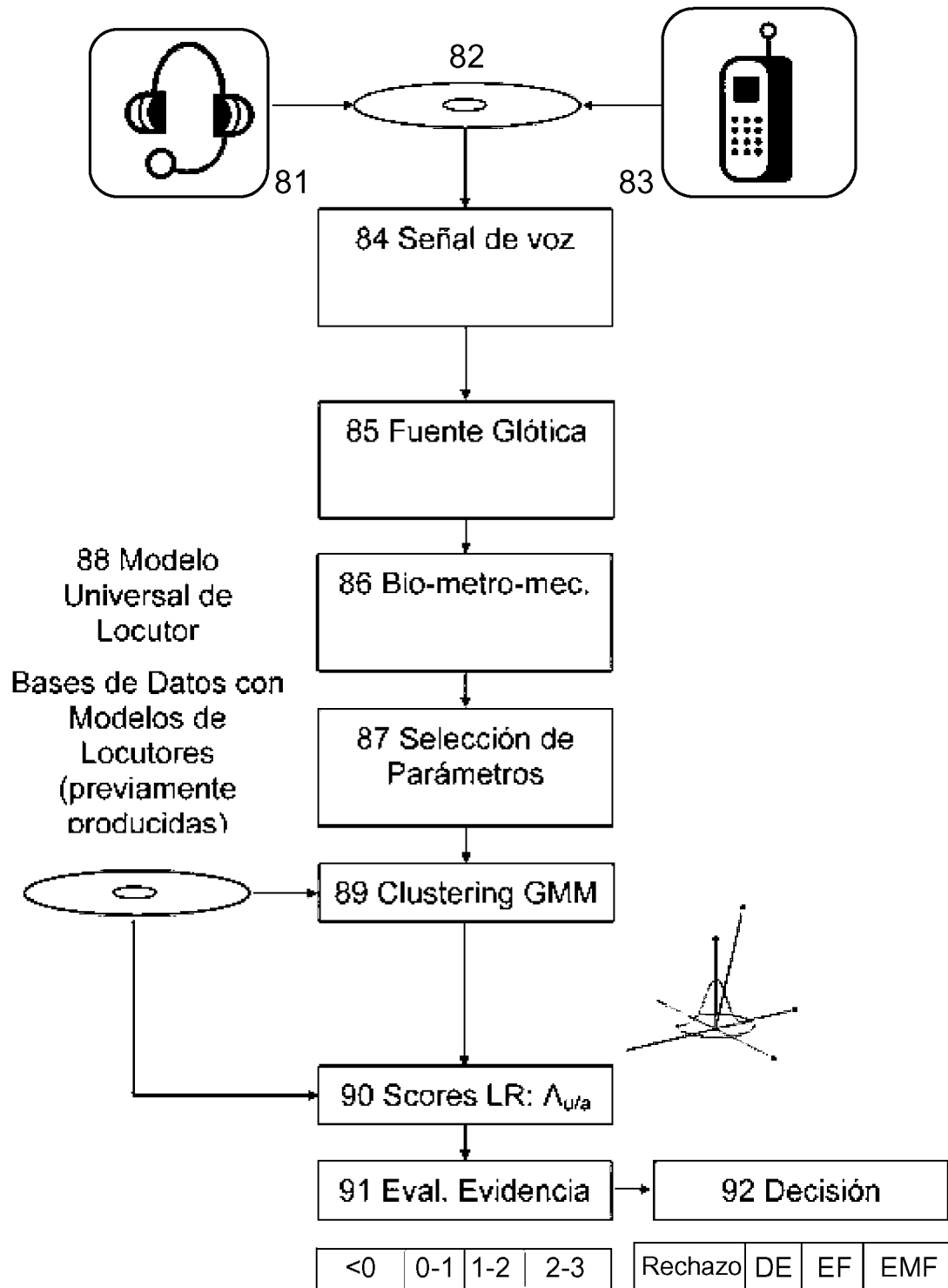


Fig. 13



OFICINA ESPAÑOLA
DE PATENTES Y MARCAS

ESPAÑA

②① N.º solicitud: 201131069

②② Fecha de presentación de la solicitud: 27.06.2011

③② Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TÉCNICA

⑤① Int. Cl.: Ver Hoja Adicional

DOCUMENTOS RELEVANTES

Categoría	Documentos citados	Reivindicaciones afectadas
A	WO 2010031437 A1 (CT DE TECNOLOGIAS DE INTERACCI et al.) 25.03.2010	1,9
A	US 5577160 A (HOSOM JOHN-PAUL et al.) 19.11.1996	1,9
A	US 7398213 B1 (LEVANON YORAM et al.) 08.07.2008	1,9
A	US 2005171774 A1 (APPLEBAUM TED H et al.) 04.08.2005	1,9

Categoría de los documentos citados

X: de particular relevancia

Y: de particular relevancia combinado con otro/s de la misma categoría

A: refleja el estado de la técnica

O: referido a divulgación no escrita

P: publicado entre la fecha de prioridad y la de presentación de la solicitud

E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado

☒ para todas las reivindicaciones

☐ para las reivindicaciones nº:

Fecha de realización del informe
19.08.2011

Examinador
M. González Vasserot

Página
1/4

CLASIFICACIÓN OBJETO DE LA SOLICITUD

G10L11/00 (2006.01)**G10L15/00** (2006.01)**G10L17/00** (2006.01)**G10L19/04** (2006.01)

Documentación mínima buscada (sistema de clasificación seguido de los símbolos de clasificación)

G10L

Bases de datos electrónicas consultadas durante la búsqueda (nombre de la base de datos y, si es posible, términos de búsqueda utilizados)

INVENES, EPODOC, WPI

Fecha de Realización de la Opinión Escrita: 19.08.2011

Declaración

Novedad (Art. 6.1 LP 11/1986)	Reivindicaciones 1-13	SI
	Reivindicaciones	NO
Actividad inventiva (Art. 8.1 LP11/1986)	Reivindicaciones 1-13	SI
	Reivindicaciones	NO

Se considera que la solicitud cumple con el requisito de aplicación industrial. Este requisito fue evaluado durante la fase de examen formal y técnico de la solicitud (Artículo 31.2 Ley 11/1986).

Base de la Opinión.-

La presente opinión se ha realizado sobre la base de la solicitud de patente tal y como se publica.

1. Documentos considerados.-

A continuación se relacionan los documentos pertenecientes al estado de la técnica tomados en consideración para la realización de esta opinión.

Documento	Número Publicación o Identificación	Fecha Publicación
D01	WO 2010031437 A1 (CT DE TECNOLOGIAS DE INTERACCI et al.)	25.03.2010
D02	US 5577160 A (HOSOM JOHN-PAUL et al.)	19.11.1996
D03	US 7398213 B1 (LEVANON YORAM et al.)	08.07.2008
D04	US 2005171774 A1 (APPLEBAUM TED H et al.)	04.08.2005

2. Declaración motivada según los artículos 29.6 y 29.7 del Reglamento de ejecución de la Ley 11/1986, de 20 de marzo, de Patentes sobre la novedad y la actividad inventiva; citas y explicaciones en apoyo de esta declaración

Los documentos citados solo muestran el estado general de la técnica, y no se consideran de particular relevancia. Así, la invención reivindicada se considera que cumple los requisitos de novedad, actividad inventiva y aplicación industrial.

1.- El objeto de la presente solicitud de patente consiste en un método y un sistema para la estimación de parámetros fisiológicos de la fonación que se enmarca en el sector de las tecnologías de la información y las comunicaciones con aplicación en biomedicina y seguridad y más concretamente en el campo de la detección y la gradación de la patología orgánica de la voz por medio de la clasificación de los parámetros obtenidos de la onda glótica de la voz y en la gestión de acceso seguro por voz.

Se propone un sistema de captura de señal de voz y un conjunto de métodos y algorítmicos diseñados para la extracción de parámetros relevantes a partir de la onda glótica y para su clasificación que permiten la determinación de la presencia de disfonía, del grado de la misma y de la identidad del locutor.

2.- El problema planteado por el solicitante es registrar y analizar la voz para calcular una serie de parámetros de la fonación que transportan información relevante sobre influencias causadas por trastornos orgánicos (que afectan a la fisiología de la laringe) o neurológicos (que afectan a los centros cerebrales del habla).

También es parte esencial los procedimientos clasificatorios que permiten obtener estimación de la disfunción presente y asignación de la personalidad.

La invención resuelve los siguientes problemas:

- La influencia del tracto vocal en la fonación es fuertemente enmascaradora de la actividad dinámica de los pliegues vocales, y dificulta enormemente la estimación del estado fisiológico de la misma a partir del registro de la voz.
- Las estimaciones del estado fisiológico de los pliegues vocales a partir del análisis acústico de la voz se centran en el uso de parámetros de distorsión que no tienen una semántica bien definida y unívoca respecto al problema que modelan.
- La personalización del locutor a partir del habla incorpora mucha información articulatoria, dependiente del texto, que genera una alta variabilidad intra-locutor dificultando las tareas de la identificación robusta.

Y lo hace mediante las siguientes actuaciones:

- Se utilizan parámetros derivados de la onda glótica para determinar la actividad dinámica de los pliegues vocales, mediante la reconstrucción de dicha señal por filtrado inverso de la señal de voz. Los nuevos parámetros son estimaciones de la envolvente del espectro en frecuencia de la onda glótica así reconstruida.
- Se realizan estimaciones de los parámetros biomecánicos de los pliegues vocales mediante la adaptación de un modelo biomecánico resonante que reconstruya el comportamiento en frecuencia en una banda dada de la envolvente espectral de la onda glótica. Los parámetros biomecánicos son estimados mediante los valores del modelo biomecánico citado mediante inversión del sistema dinámico de los mismos. Estos nuevos parámetros determinan comportamientos normales y anormales de los pliegues vocales en fonación de forma mucho más directa.
- Se elimina la influencia del tracto vocal en la fonación mediante filtrado inverso, lo que reduce la variabilidad intra-locutor producida por la articulatoria. Ello mejora las tasas de discriminación inter-locutor al separar mejor las clases de locutores modelados.

El documento D1 puede considerarse como el representante del estado de la técnica más cercano ya que en este documento confluyen la mayoría de las características técnicas reivindicadas.

Análisis de las reivindicaciones independientes 1,9

D1 se diferencia del documento de solicitud de patente en que no se modela la inversión espectral del tracto vocal para extraer la onda glótica sin influencia sustancial del tracto vocal ni para obtener los correlatos de vibración de los pliegues vocales.

Por tanto no incluye la etapa c) del método propuesto, tampoco emplea medios configurados, a los que se refiere en la reivindicación 9 para efectuar la etapa C) de la reivindicación 1 de la solicitud.

La reivindicación 1 y 9 son nuevas (Art. 6.1 LP 11/1986) y tienen actividad inventiva (Art. 8.1 LP11/1986).

Análisis del resto de los documentos

De este modo, ni el documento D1, ni ninguno del resto de los documentos citados en el Informe del Estado de la Técnica, tomados solos o en combinación, revelan la invención en estudio tal y como es definida en las reivindicaciones independientes, de modo que los documentos citados solo muestran el estado general de la técnica, y no se consideran de particular relevancia. Además, en los documentos citados no hay sugerencias que dirijan al experto en la materia a una combinación que pudiera hacer evidente la invención definida por estas reivindicaciones y no se considera obvio para una persona experta en la materia aplicar las características incluidas en los documentos citados y llegar a la invención como se revela en la misma.