

CUANTIFICACIÓN DE LA CALIDAD DE LA VOZ PARA SU EVALUACIÓN CLÍNICA POR MEDIO DEL ANÁLISIS ACÚSTICO

Jesús B. Alonso¹, Miguel A. Ferrer¹, José de León², Carlos M. Travieso¹

¹ Dpto. de Señales y Comunicaciones, Universidad de Las Palmas de Gran Canaria
Campus de Tafira, 35017 Las Palmas de Gran Canaria, SPAIN
email: jalonso@dsc.ulpgc.es

² Servicio de Otorrinolaringología, General Hospital de Gran Canaria “Dr.Negrín”
Las Palmas de Gran Canaria, SPAIN

RESUMEN

En lo que respecta a la evaluación del sistema fonador, las técnicas basadas en métodos de inspección visual directa son las técnicas que permiten obtener la mejor valoración del estado del sistema fonador de un locutor, sin embargo, estas técnicas presenta ciertos inconvenientes como el hecho de tratarse de técnicas invasivas, con la correspondiente incomodidad para el locutor que se preste a la evaluación, o el hecho de que este tipo de pruebas han de ser realizadas por profesionales del ámbito médico especializado.

En este trabajo se propone un procedimiento para evaluar la calidad de la voz de un locutor a partir de una grabación de audio. Este procedimiento, consiste en evaluar cuatro características de la señal de voz: la estabilidad, la riqueza espectral, la presencia de ruido y su comportamiento no lineal. Además, se han identificado las medidas de calidad de la voz que mejor cuantifican a cada uno de estas características de la voz. Finalmente, el trabajo concluye presentando una metodología del análisis acústico para la evaluación clínica de la voz.

1. INTRODUCCIÓN

Medir la calidad de la voz de un locutor a partir de una grabación consiste en cuantificar diferentes características acústicas de la voz, las cuales permitan compararla respecto a ciertos patrones de referencia. En la bibliografía actual no existe una única medida que documente todas las características acústicas de la voz.

Una medida de la calidad acústica de la voz basada en una valoración auditiva es intrínsecamente difícil de establecer como referencia comparativa entre distintas voces y distintos evaluadores. En la bibliografía actual se han realizado varios intentos de objetivar la medida de la calidad de voz por medio de medidas clínicas multidimensionales basadas en métodos auditivos. Existen ejemplos bien conocidos como son: la escala

GRABS de Japón [1] y su formulario extendido, desarrollado y aplicado en Europa [2], o el utilizado en Suecia donde se ha trabajado en otro juego de descriptores clínicos de la voz referentes a la percepción [3], o como en el que se introduce un juego de características fonéticas las cuales intentan agregar información de la excitación del tracto vocal en la medida de la calidad de la voz [4]. Estos procedimientos de medida de la calidad de la voz pretenden alcanzar una medida objetiva a partir de una valoración subjetiva.

Existen diferentes trabajos en los que se proponen medidas objetivas de la calidad de la voz obtenidas a partir de una grabación de audio [6][7][8][9][10][11][12]. Una característica común de la mayoría de estos trabajos es la de utilizar, para el cálculo de medidas de la calidad de la voz, una grabación de un sonido sonoro sostenido, principalmente una vocal. Esto es debido a que durante la fonación de este tipo de sonidos el sistema de producción de voz pone en funcionamiento gran parte de sus mecanismos (flujo glótico de aire constante, vibración de las cuerdas vocales de forma continuada, ...) permitiendo captar cualquier anomalía de estos mecanismos. En estos trabajos, los distintos autores presentan diferentes conjuntos de medidas con la finalidad de cuantificar objetivamente la calidad de voz. Todos ellos tienen en común la necesidad de realizar diferentes medidas de la voz con la finalidad de recoger en el proceso de medida los distintos aspectos de las características acústicas de la voz.

2. LA CALIDAD DE LA VOZ DESDE DIFERENTES DOMINIOS

Una grabación de voz aporta diferentes características respecto a la calidad de la voz del locutor. Esta señal de voz registrada es posible representarla en diferentes dominios, lo cual resulta interesante ya que cada dominio en el que se puede representar la señal de voz,

manifiesta de forma preferencial alguna de las características de la voz del locutor. Los principales dominios estudiados en el procesado de señales de voz están centrados en los siguientes dominios:

- *Dominio Temporal*
- *Dominio Espectral*
- *Dominio Cepstral*
- *Dominio del Modelo Inverso*

Sin duda alguna gran cantidad de los trabajos sobre procesado digital de señales de voz existentes en la bibliografía actual centran sus estudios en estos dominios. Sin embargo, nuevas líneas de trabajo presentan nuevos dominios derivados de los anteriores.

De igual forma resulta de interés estudiar nuevos dominios que cada vez toman mayor importancia en el procesado de señales: los *dominios que manifiestan el comportamiento no lineal de la señal de voz*.

A continuación se describen las manifestaciones más relevantes de cada uno de los diferentes dominios.

2.1. Dominio Temporal

La envolvente de señal de voz de un locutor con voz de alta calidad resulta más regular que la envolvente de la señal de voz del locutor con voz de baja calidad. Este hecho resulta más significativo en intervalos de tiempo muy pequeños. Los fenómenos más característicos que permiten diferenciar entre voz de alta calidad y voz de baja calidad son:

- La energía de la señal contenida en un intervalo de tiempo pequeño, varía mucho de un intervalo al siguiente en las señales de voz de baja calidad respecto a las señales de voz de alta calidad.
- Otra característica que presenta las voces de baja calidad es la existencia de intervalos carentes de periodicidad durante fonaciones sonoras sostenidas.

2.2. Dominio Espectral

Las muestras de voz de un sonido sonoro sostenido producidas por un locutor con voz de baja calidad presentan las siguientes características diferenciales frente a las voces de alta calidad:

- Menor regularidad de la envolvente del espectro, principalmente en las bajas frecuencias.
- Mayor porcentaje de energía en las bajas frecuencias respecto a la energía total.
- Presencia de bloques de energía en las altas frecuencias, los cuales son debido a la presencia de ruido glótico.
- Gran variación del espectro de una trama con respecto a las tramas contiguas.

Una característica de las voces de alta calidad es la energía del espectro concentrada alrededor de determinados formantes, principalmente el primer y el tercer formante, mientras que las voces de baja calidad

se caracterizan por presentar gran componente de ruido alrededor de los formantes.

Las voces de alta calidad presentan gran cantidad de riqueza espectral. Por el contrario, las voces de baja calidad presentan muy poca componente armónica, estando ésta concentrada en las componentes de muy baja frecuencia. En cualquier caso, la cantidad de riqueza espectral es una peculiaridad de la voz de un locutor, sin embargo, la variación de la riqueza espectral a lo largo del tiempo durante la fonación de un sonido sonoro sostenido si es indicativo de una voz de baja calidad.

Otra característica típica de las voces de baja calidad es la de presentar, durante la fonación de un sonido sonoro sostenido, variaciones en el ritmo de vibraciones de las cuerdas vocales, lo cual es equivalente a decir que existe variaciones de la frecuencia de *pitch*.

2.3. Dominio Cepstral

Por medio de una representación cepstral es posible identificar características que permiten valorar la calidad de la voz (la envolvente del espectro, la riqueza espectral, identificar las componentes armónicas y las componentes de ruido, etc). En el caso de una fonación de un sonido sonoro sostenido, realizando una representación cepstral de diferentes tramas con duración equivalente a tres periodos de *pitch*, pueden identificarse diferentes características.

La riqueza espectral de una muestra de voz puede cuantificarse por medio de la amplitud y anchura de la componente cepstral correspondiente al *pitch*. La existencia de un pico de una amplitud considerable indica la presencia notable de energía en dicha componente armónica, siendo ésta una característica de las voces de alta calidad. También, una anchura reducida del pico cepstral, correspondiente al *pitch*, indica la alta estabilidad de la frecuencia de *pitch* para tres periodos consecutivos, también siendo ésta una característica de las voces de alta calidad. Características como la amplitud y la estrechez del pico cepstral correspondiente al segundo armónico son, al igual que ocurre con el pico del *pitch*, aspectos que diferencian a las muestras de voz de alta y baja calidad. La presencia de un pico cepstral correspondiente al primer armónico el cual resulte más estrecho que el pico *cepstral* correspondiente al segundo armónico resulta una característica propia de las voces de alta calidad.

La presencia de ruido glótico en la señal de voz puede estimarse por medio de relaciones entre diferentes regiones de los cepstrum: la componente armónica (las componentes cepstrales correspondientes al *pitch* y sus armónicos), y la componente de ruido (el resto de las componentes cepstrales).

2.4. Dominio del Modelo Inverso

En este dominio se estima, a partir de la señal de voz, la forma de onda del pulso de aire que se produce debido a las cuerdas vocales durante la fonación de un sonido

sonoro, es decir, la forma de onda de la señal residual (forma de onda del flujo glótico). Esta estimación de la forma de onda se obtiene mediante un filtro inverso sobre la señal de voz, donde se elimina el efecto del tracto vocal y el efecto de la radiación en los labios.

Las diferentes manifestaciones de la calidad de la voz en la señal glótica se cuantifican por medio de los valores de amplitudes, de los instantes de inicio de la apertura, de máxima apertura de las cuerdas vocales, de inicio del cierre, etc, y diferentes relaciones entre distintos instantes del ciclo glótico: *open quotient*, *speed quotient*, *closing quotient*, etc.

2.5. Dominio No Lineal

Los principales sistemas comerciales que permiten evaluar objetivamente la calidad de la voz de un locutor a partir de una grabación (Dr Speech (*Tiger Elemetric*),SSVA (*System for Sigle Voice Analysis*), MDVP (*Multi-Dimensional Voice Program*), EVA (*Evaluation Vocal Assiste*), CSL (*Computerized Speech Laboratory*) PRAAT, VISHACSRE (*Computerized Speech Research Environment*), MEDIVOZ, etc) no evalúan las características no lineales presentes en la señal de voz.

El modelo más popular de caracterización del sistema de producción de voz es un sistema tiempo-variante, basado en teorías acústicas lineales, consistente en un modelo de fuentes y filtros (*source/filter model*). La existencia de variaciones en la amplitud espectral de la señal de voz, en la frecuencia fundamental, para una excitación estable, son indicios para presuponer un comportamiento no lineal.

Ha sido identificada [13] la presencia de una frecuencia fundamental (f_s) y de un subarmónico ($f_s/2$) en la señal de voz, donde el efecto de los subarmónicos se manifiesta o bien en una modulación de amplitud o bien en una modulación de frecuencia, existiendo la posibilidad de que se produzcan los dos fenómenos. Estudios realizados [14] indican que un 31% de las muestras de voz de locutores patológicos presentan subarmónicos. Si embargo, también se han identificado la existencia de subarmónicos en las voces de alta calidad [15], donde el fenómeno se estima que lo presenta el 10.5% de la población, siendo esta presencia de subarmónicos no necesariamente indicativa de anomalías.

Existen dos posibles teorías para justificar la presencia de subarmónicos:

- La teoría de Titze [16]: la producción de los subarmónicos se debe a asimetrías mecánicas o geométricas entre las cuerdas vocales.
- La teoría de Svec [17]: la frecuencia subarmónica se debe a la combinación de dos modos vibratorios (bifonación: la presencia de dos frecuencias principales) cuyas frecuencias tienen la relación 3:2.

Sin embargo, ambas teorías son las mismas según [18], donde los autores indican que la presencia de bifonación puede ser inducida por asimetría entre cuerdas derecha-izquierda o por desincronización en la

vibración anterior-posterior, siendo estas asimetrías o desincronizaciones en las fases de la onda mucosa el origen de los subarmónicos [19], lo que se asocia a significativas diferencias en las masas o propiedades viscoelásticas entre las cuerdas vocales, siendo este comportamiento susceptible de ser modelado por modelos de fonación no lineales. Esta idea es compartida por [20], que propone la necesidad del uso de modelos no lineales para explicar el efecto de la viscosidad de la mucosa de las cuerdas vocales en aspectos como su vibración (la mucosa en la superficie de las cuerdas vocales genera tensión superficial y causa adhesión).

Desde otra perspectiva, el modelo tradicional del tracto vocal asume que la propagación de la onda sonora se realiza por medio de una propagación plana. Sin embargo, las medidas de presión sonora y las medidas de variación de volumen se adaptan de forma más realista a un modelo no lineal de fluido de dinámicos, debiéndose esto a que las cavidades existentes entre las cuerdas vocales y las falsas cuerdas vocales pueden provocar un torbellino, incluso pudiéndose producir torbellinos periódicos, el cual excita al tracto vocal durante la fase de cierre de las cuerdas vocales.

A partir de estudios empíricos, se ha estudiado la dimensión fractal, por parte de diferentes autores, coincidiendo entre ellos en la baja dimensionalidad de la señal de voz de alta calidad. En [5] se afirma que la cantidad de alinealidades en el sistema vocal es una importante determinación de la presencia de fonación anormal, también indica que la dimensionalidad del espacio de fases, utilizado para caracterizar el *attractor*, puede estar relacionada con la cantidad de masa de las cuerdas puestas en juego.

3. FENÓMENOS FÍSICOS IMPLICADOS EN LA CALIDAD DE LA VOZ

En el apartado anterior se ha realizado un recorrido por los principales dominios de representación en los que se puede estudiar la señal de voz, describiendo las diferentes manifestaciones, propias de cada dominio, que permiten valorar la calidad de la voz. Cada una de las diferentes manifestaciones mencionadas caracteriza un fenómeno físico que interviene en la generación de la voz y sin embargo el mismo fenómeno físico puede manifestarse en diferentes dominios de representación de la señal de voz. En este trabajo se ha identificado el conjunto de fenómenos físicos que son necesarios describir para realizar una correcta documentación de la calidad de voz de un locutor. Los cuatro fenómenos físicos identificados son los siguientes:

- **Estabilidad de la voz:** es la capacidad de un locutor de producir, durante la producción de un sonido sonoro sostenido, un flujo de aire con una intensidad constante para excitar las cuerdas vocales. Este fenómeno físico se cuantifica a partir de medidas de estabilidad de la señal de voz.

- **Riqueza espectral:** se define como la capacidad de generar, durante la fonación de un sonido sostenido, un movimiento periódico de las cuerdas vocales y de producir una excitación sonora del tracto vocal con gran cantidad de componentes espectrales. Este fenómeno físico se cuantifica por medio del cálculo de la estabilidad de la frecuencia de *pitch* y por la presencia de armónicos con energía elevada en las diferentes bandas de frecuencias.
- **Presencia de ruido:** se define como la presencia de ruido glótico en la voz, durante la fonación de un sonido sonoro sostenido, debido a carencias en la fase de cierre de las cuerdas vocales. Este fenómeno físico se cuantifica por medio de la presencia de ruido no estacionario en la voz.
- **Comportamiento no lineal:** este fenómeno físico es ocasionado por un funcionamiento anómalo de las cuerdas vocales, debido a irregularidades en las masas que interviene en la fase de cierre de cuerdas, la existencia de asimetrías en el movimiento de dichas cuerdas y factores relacionados con la mucosa que recubre a las cuerdas. Estos fenómenos son cuantificables por medio de la identificación de comportamientos no lineales en la señal de voz.

Una voz de calidad anormal presenta al menos uno de los valores correspondientes a la cuantificación de los cuatro fenómenos físicos fuera de los rangos de normalidad. Este procedimiento de cuantificación de la calidad de la voz permite identificar calidades de voz anómalas de diferente origen, es decir, la voz de un locutor con baja calidad debida a un estado incipiente de una patología orgánica presentará probablemente la alteración de la “Estabilidad de la voz” mientras que otro locutor con un estado incipiente de una patología funcional presentará valores anormales de “Presencia de ruido”. A medida que la calidad de la voz de un locutor empeora, aumentará el número de fenómenos físicos y la desviación de los valores de la cuantificación de dichos fenómenos respecto a los valores de normalidad.

Estos cuatro diferentes fenómenos físicos son cuantificables en diferentes dominios de representación de la señal de voz, existiendo diferentes medidas objetivas de la calidad de la voz que cuantifican con mayor o menor precisión un mismo fenómeno físico.

4. ESTRATEGIA PARA SELECCIONAR LAS MEDIDAS DE CALIDAD DE LA VOZ

Hasta ahora en la bibliografía actual nos encontramos diferentes escenarios en el estudio de las medidas de calidad de la voz. Por un lado, nos encontramos los resultados obtenidos a partir de bases de datos de voces de diferentes calidades donde el número de muestras es reducido, en los que se presentan los resultados obtenidos para el estudio de una determinada medida o de un subconjunto de medidas. Por otro lado existen estudios donde, a partir de una base de datos de voces de diferentes calidades con un número significativo de

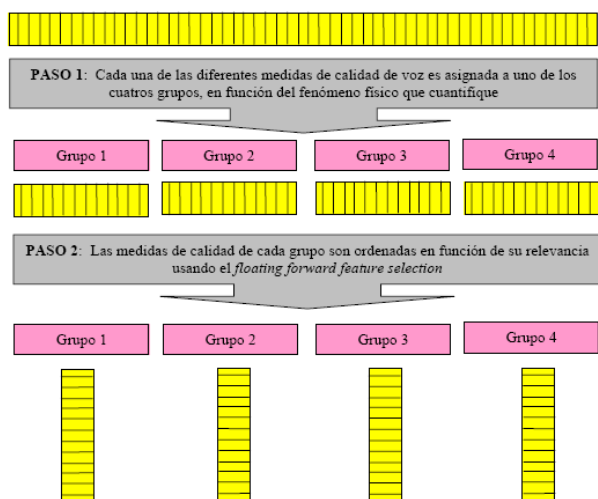
muestras, se realiza el estudio de conjuntos de medias e incluso se aplican técnicas de selección de características con el fin de cuantificar la relevancia de las diferentes medidas de calidad de la voz. Existen diferentes aspectos en la metodología de estos estudios que no permiten realizar una generalización de los resultados obtenidos.

Debido principalmente a la inexistencia de una gran base de datos, resulta necesario demostrar la independencia de los resultados de la cuantificación de la relevancia de las medidas de calidad de la voz con la base de datos.

Otro aspecto a considerar, es la gran cantidad de estudios en los que no se tiene en cuenta en el proceso de selección de medidas de la calidad de la voz todo el conocimiento existente en el ámbito clínico respecto a la naturaleza de los trastornos del sistema fonador que producen una alteración de la calidad de la voz. En este sentido, apreciamos la necesidad de disponer de medidas de calidad de la voz que cuantifiquen cada uno de los cuatro fenómenos físicos identificados con el fin de poder documentar la gran variedad de tipos de alteraciones de la calidad de la voz existentes, ya que cualquier voz de baja calidad presenta al menos la alteración de una de las cuatro fenómenos físicos identificados.

En este trabajo se propone clasificar las diferentes medidas de calidad de voz en los cuatro grupos atendiendo al fenómeno físico que cuantifica. Una vez identificadas las diferentes medidas de calidad de voz pertenecientes a cada una de los diferentes grupos de características es necesario estudiar la relevancia de las medidas para cada grupo. El método de selección de características que se propone, para estudiar la relevancia de las medidas de cada grupo, es el algoritmo *floating forward feature selection*, ya que permite obtener ordenadas las medidas de la calidad de la voz en función de su relevancia y tiene un coste computacional razonable.

En concreto, el método propuesto consiste en asignar las diferentes medidas de calidad de la voz en uno de los cuatro grupos de medidas que hemos identificado. Posteriormente, las medidas de cada grupo se ordenan en función de su relevancia por medio del método *floating forward feature selection*. Este método proporciona, para cada grupo de medidas, un conjunto óptimo de características, y es por ello que el proceso de evaluación de cada uno de los diferentes grupos de medidas se repite k iteraciones, y posteriormente se ordenan las características en función del número de ocasiones en las que para las diferentes k iteraciones estuvieron seleccionadas dentro del subconjunto óptimo.



Para la evaluación del sistema propuesto para identificar al subconjunto óptimo de medidas de calidad de la voz, se han utilizado dos bases de datos totalmente independientes. Con una de ellas, se realizó la selección de las medidas y su posterior evaluación, y con la otra base de datos sólo se realizó la evaluación, con el fin de comparar los resultados obtenidos en diferentes sistemas de clasificación.

5. DATOS

Con el fin de evaluar el sistema de selección de medidas de calidad de la voz, se utilizó una colección de voces de locutores los cuales presentan diferentes patologías en el sistema fonador y voces de locutores sanos. Dentro de las muestras de voces de locutores patológicos se dispuso de muestras con distintos grados de afección consiguiendo disponer de un variado rango de calidades de voz. Para el estudio se dispuso de dos bases de datos.

5.1. Base de Datos 1

La grabación de voz ha sido realizada en un centro hospitalario, en una habitación en la que se ha tenido en cuenta los niveles de contaminación acústica, intentándose obtener un compromiso entre muy bajos niveles de ruido, y niveles de ruido realistas, ya que si no fuese así resultaría imposible aplicar a las condiciones normales de trabajo cualquier resultado obtenido a partir de la base de datos.

El contenido de las grabaciones corresponde a la fonación de la vocal “a” castellana de forma sostenida y no susurrada. Este formato consiste en audio digital, con una frecuencia de muestreo de 22050 KHz y con una resolución de 16 bits por muestra. Se utiliza una tarjeta de sonido convencional y un micrófono común, con ancho de banda lineal superior a 11 KHz.

En la siguiente tabla se presenta los datos pormenorizados por sexos y grados de patologías.

Grupo de Muestras de Voces		Hombres	Mujeres	Total
Muestras de voces sanas		43	44	87
Muestras de voces patológicas	Disfonía leve	7	19	26
	Disfonía moderada	22	14	36
	Disfonía severa	18	13	31
				93

5.2. Base de Datos 2

Esta base de datos es la *Voice Disorders Database Model 4337* desarrollada por el *Massachusetts Eye and Ear Infirmary Voice and Speech Lab*.

Contiene 1400 muestras de voz de aproximadamente 700 sujetos de habla inglesa. Consiste en la fonación sostenida de la vocal “ah”, de pacientes con una amplia variedad de patologías del sistema fonador, al igual que 53 muestras de voces sanas. Para su elaboración se usó un micrófono capacitivo en una cámara acústicamente aislada. La señal fue registrada con un grabador DAT con una frecuencia de muestreo de 44.1 KHz, utilizándose un preamplificador calibrado.

6. RESULTADOS

Se han identificado las principales medidas de calidad de la voz presentes en la bibliografía actual, posteriormente han sido clasificadas en uno de los cuatro grupos de medidas, y tras aplicar la metodología propuesta, las medidas de cada uno de los cuatro grupos han sido evaluadas por separado [6], utilizando un diferentes sistemas de clasificación (redes neuronales probabilísticas (PNN), redes neuronales *feedforward* (NN) con una capa oculta, clasificadores basados en funciones de base radial (RBF), máquinas de vectores soporte (SVM) y clasificadores basados en los K vecinos más cercanos (KNN)). La evaluación se realizó de forma incremental, añadiendo de forma individual medidas de calidad de la voz en orden, en función de su relevancia. Se observó que la tasa de éxito en los diferentes clasificadores no se incrementaba respecto a los resultados obtenidos utilizando la medida de calidad con mayor relevancia, a medida que se añadían más medidas de calidad de la voz en el conjunto de características de entrada de los diferentes clasificadores.

Las medidas que mayor relevancia presentaron para los diferentes grupos [6] fueron:

Fenómeno Físico	Medida de Calidad de la Voz
Estabilidad de la voz	Amplitud Perturbation Quotient (APQ)
Riqueza Espectral	Media Valor Cepstral del primer pico de los cepstrum derivados
Presencia de Ruido	Media Relación Espectral Subarmónico - armónico de al señal residual
Comportamiento no Lineal	Media Área inferior del biespectro integrado

Los resultados obtenidos para los diferentes clasificadores son los siguientes:

Clasificador	Tasa de Éxito Base de datos 1	Tasa de Éxito Base de datos 2
PNN	80.95%	78.21%
NN	88.21%	84,36%
RBF	86.94%	84.53%
SVM	84.34%	85.31%
KNN	84.66%	85.38%

Por lo tanto, además de resultar suficiente, con una medida de calidad de la voz para cada grupo de medidas, se observa que los resultados obtenidos con la base de datos 1 (utilizada para la selección) y los resultados obtenidos con la base de datos 2 (utilizada sólo para la evaluación) resultan muy similares.

Se ha estudiado además los rangos de normalidad de las diferentes medidas, estimándose en un intervalo de confianza del 95% y asociando el valor de la medida al fenómeno físico que cuantifica, obteniendo los siguientes rangos:

Estabilidad de la Voz	0.000.....0.136
Riqueza Espectral	0.919.....1.759
Presencia de Ruido	0.189.....0.561
Comportamiento no Lineal	0.530.....0.754

7. CONCLUSIONES

En este trabajo se han identificado los diferentes fenómenos físicos que caracterizan la calidad de la voz y que necesitan ser cuantificados para una correcta documentación de la calidad de la voz, introduciendo de forma novedosa la cuantificación del comportamiento no lineal de la señal de voz.

Además, se propone una estrategia de selección de medidas, que combina la experiencia clínica con los métodos ciegos de selección de características, y que identifican las medidas de calidad de la voz que mejor permiten documentar la calidad de la voz de un locutor.

Finalmente, se propone un protocolo de evaluación de la calidad de la voz por medio de 4 medidas objetivas obtenidas a partir de una grabación de voz, que permiten asociar el valor de la medida con la cuantificación de un determinado comportamiento fisiológico del sistema fonador.

8. BIBLIOGRAFÍA

[1] Hirano, M., Clinical Examination of Voice. New York, Springer-Verlag, 1981
 [2] Dejonckere, P. H. Remale, M. Fresnel-Elbaz, E. Woisard, V. Crevier-Buchman, L. Millet, B., "Differentiated perceptual evaluation of pathological voice quality: reliability and correlations with acoustic measurements". Revue de Laryngologie Otolgie Rhinologie, vol. 117, no.2, pp. 219-224, (Bordeaux) 1996.
 [3] Hammarberg, B. y Gauffin, J., Perceptual and acoustic characteristics of quality differences in pathological voices as

related to physiological aspects., in O. Fujimura & M.Hirano (eds.), Vocal Fold Physiology, pp 283-303, 1995
 [4] Laver, J., The Gift of Speech, Edinburgh University Press,1991
 [5] Robert F. Orlikoff, R. J. Baken, "Curing Diagnosis: Improving the Taxonomy of Phonatory Dysfunction", Sixth Conference on Advances in Quantitative Laryngology. Hamburg, Germany, April 2003
 [6] Jesús B. Alonso, Metodología del Análisis Acústico para la Evaluación Clínica de la Voz, Tesis Doctoral, Universidad de Las Palmas de Gran Canaria, 2006
 [7] Boyanov, B; Hadjitodorov, S. "Acoustic analysis of pathological voices. A voice analysis system for the screening of laryngeal diseases". IEEE Engineering in Medicine and Biology Magazine, vol.16,no.4, pp.74 – 82, 1997.
 [8] Hansen, J.H.L.; Gavidia-Ceballos, L.; Kaiser, J.F. "A nonlinear operator-based speech feature analysis method with application to vocal fold pathology assessment". IEEE Transactions on Biomedical Engineering, vol.45, no. 3, pp.300-313, March 1998.
 [9] Stefan Hadjitodorov, Petar Mitev, "A computer system for acoustic analysis of pathological voices and laryngeal diseases screening", Medical Engineering Physics, vol. 24, pp. 419-429, 2002.
 [10] Michaelis D.; Frohlich M.; Strube H. W. "Selection and combination of acoustic features for the description of pathologic voices". Acoustical Society of America. vol.103, no.3, pp.1628-1640, 1998.
 [11] Boyanov B., Doskov D., Mitev P., Hadjitodorov S., Teston B. , "New cepstral parameters for description of pathologic voice", Comptes Rendus de L'Academie Bulgare des Sciences (Ann. of Bulgarian Academy of Sciences), vol.53, no.3, pp.41-44, 2000
 [12] Godino-Llorente, J.I.; Aguilera-Navarro, S.; Gomez-Vilda, P. "Non supervised neural net applied to the detection of voice impairment". Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '00. vol.6, pp.3594-3597, 2000.
 [13] Xuejing Sun and Yi Xu, "Perceived Pitch of Synthesized Voice with Alternate Cycles", Journal of Voice, vol.16, no. 4, pp. 443-459, 1995
 [14] Nuñez Batalla F, Suarez Nieto, C., Muñoz Pinto, C., Baragaño Rio, L., Álvarez Zapico, M.J., Martínez Ferreras, A., "Estudio Espectrográfico de la disfonía: Subarmónicos", Acta Otorrinolaring. Esp., vol.52,no. 1, pp.52-56, 2000.
 [15] C. Michael Haben, Karen Kost, George Papagiannis, "Lateral Phase Mucosal Asymmetries in the Clinical Voice Laboratory", Journal of Voice, vol.17,no. 1, pp.3-11,2003.
 [16] TITze, IR., Principles of Voice Production, Englewood Cliffs, NJ: Prentice-Hall, Inc, 1994 .
 [17] Svec JG, Schutte HK, Miller DG." A Subharmonic vibratory pattern in normal vocal folds", Journal of Speech and Hearing Research, vol.39, no.1, pp.135-143,1996.
 [18] Jürgen Neubauer, Patrick Eysholdt, Ulrich Eysholdt, Hanspeter Herzel, "Spatio-temporal analysis of irregular vocal fold oscillations: Biphonation due to desynchronization of spatial models", J. Acoustical Society of America, vol.110, no. 6, pp.3179-3192, December 2001.
 [19] C. Michael Haben, Karen Kost, George Papagiannis, "Lateral Phase Mucosal Asymmetries in the Clinical Voice Laboratory", Journal of Voice, vol.17,no. 1, pp.3-11,2003.
 [20] Stephane Ayache, Maurice Ouaknine, Philippe Dejonkere, Pierre Prindere, Antoine Giovanni, "Experimental Study of the Effects od Surface Mucus Viscosity on the Glottic Cycle", Journal of Voice, vol.18, no.1, 2004.