

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/318441405>

# Application of the Acoustic Voice Quality Index for Objective Measurement of Dysphonia Severity

Article in *Acta Otorrinolaringologica (English Edition)* · July 2017

DOI: 10.1016/j.otoeng.2017.06.007

---

CITATIONS

5

---

READS

117

5 authors, including:



**Faustino Nuñez-Batalla**

Hospital Universitario Central de Asturias

147 PUBLICATIONS 1,145 CITATIONS

SEE PROFILE



**Jose L. Llorente**

Hospital Central de Asturias

359 PUBLICATIONS 3,306 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



NEWBORN HEARING SCREENING [View project](#)



Voice disorders [View project](#)



ARTÍCULO ORIGINAL

## Aplicación del índice acústico de calidad vocal para la cuantificación objetiva de la severidad de la disfonía



Faustino Núñez-Batalla<sup>a,\*</sup>, Estefanía Díaz-Fresno<sup>b</sup>, Andrea Álvarez-Fernández<sup>b</sup>, Gabriela Muñoz Cordero<sup>a</sup> y José Luis Llorente Pendás<sup>a</sup>

<sup>a</sup> Servicio de Otorrinolaringología, Hospital Universitario Central de Asturias, Oviedo, España

<sup>b</sup> Facultad de Psicología y Logopedia, Universidad de Oviedo, Oviedo, España

Recibido el 27 de octubre de 2016; aceptado el 8 de noviembre de 2016

Disponible en Internet el 15 de febrero de 2017

### PALABRAS CLAVE

Prominencia del pico cepstral suavizado;  
Índice acústico de calidad vocal;  
Praat.

### Resumen

**Introducción:** En las últimas décadas se han estudiado muchos parámetros acústicos con el fin de cuantificar la disfonía, sin embargo las medidas actuales podrían no ser sensibles a la calidad vocal percibida. En un metaanálisis donde se evaluó la relación entre la calidad global percibida de la voz se identificaron numerosas medidas acústicas que no dependen de la extracción del periodo fundamental tales como las derivadas del cepstrum y que pueden ser usadas tanto en vocales sostenidas como en habla conectada. Un método específico recientemente diseñado para cuantificar la severidad global de la disfonía es el índice acústico de calidad vocal (AVQI) que es un constructo multivariante que combina múltiples marcadores acústicos para generar un único número que se correlaciona razonablemente con la calidad global vocal.

**Métodos:** Esta investigación se basó en un archivo de voces grabadas procedentes de dos grupos de individuos: 60 personas sanas vocalmente y 58 participantes con trastornos vocales. Se grabaron vocales sostenidas y una muestral de habla conectada para su análisis y así obtener los seis parámetros incluidos en el AVQI utilizando el programa Praat. El análisis estadístico de los resultados se llevó a cabo con el programa SPSS para Windows versión 12.0.

**Resultados:** Correspondencia entre el valor del AVQI y el grado de severidad global de la disfonía: correlación  $r = 0,68$  ( $p < 0,000$ ), y entre el AVQI y el parámetro B del GRABS:  $r = 0,68$  ( $p < 0,000$ ).

**Conclusiones:** Los resultados obtenidos muestran una buena correlación con la calidad vocal percibida. En conclusión, el AVQI es una medida multivariante, accesible, factible y razonablemente válida para medir clínicamente la severidad global de la severidad de la disfonía.

© 2017 Elsevier España, S.L.U. y Sociedad Española de Otorrinolaringología y Cirugía de Cabeza y Cuello. Todos los derechos reservados.

\* Autor para correspondencia.

Correo electrónico: [fnunezb@telefonica.net](mailto:fnunezb@telefonica.net) (F. Núñez-Batalla).

**KEYWORDS**

Smoothed cepstral peak prominence;  
Acoustic voice quality index;  
Praat.

## Application of the acoustic voice quality index for objective measurement of dysphonia severity

**Abstract**

**Introduction:** Over the past several decades, many acoustic parameters have been studied as sensitive to and to measure dysphonia. However, current acoustic measures might not be sensitive measures of perceived voice quality. A meta-analysis which evaluated the relationship between perceived overall voice quality and several acoustic-phonetic correlates, identified measures that do not rely on the extraction of the fundamental period, such the measures derived from the cepstrum, and that can be used in sustained vowel as well as continuous speech samples. A specific and recently developed method to quantify the severity of overall dysphonia is the acoustic voice quality index (AVQI) that is a multivariate construct that combines multiple acoustic markers to yield a single number that correlates reasonably with overall vocal quality.

**Methods:** This research is based on one pool of voice recordings collected in two sets of subjects: 60 vocally normal and 58 voice disordered participants. A sustained vowel and a sample of connected speech were recorded and analyzed to obtain the six parameters included in the AVQI using the program Praat. Statistical analysis was completed using SPSS for Windows, version 12.0.

**Results:** Correlation between perception of overall voice quality and AVQI: A significant difference exists ( $t(95) = 9.5$ ;  $p < .000$ ) between normal and dysphonic voices.

**Conclusions:** The findings of this study demonstrate the clinical feasibility of the AVQI as a measure of dysphonia severity.

© 2017 Elsevier España, S.L.U. and Sociedad Española de Otorrinolaringología y Cirugía de Cabeza y Cuello. All rights reserved.

**Introducción**

La valoración objetiva de la calidad vocal es una parte esencial del proceso de valoración de la disfonía en una consulta de voz. Tanto en el contexto clínico como en el de investigación existen dos formas de describir la severidad percibida de la misma. La primera, genérica y global, se refiere a la sensación compleja del grado percibido de la disfonía, y se conoce por la «G» (de «grado»), «severidad del trastorno de la voz», «severidad de la disfonía» y «severidad global». La segunda, consiste en calificar dimensiones perceptuales específicas y singulares de la calidad vocal, de las cuales, las mejor conocidas son la aspereza (*roughness*) y la calidad aérea de la voz (*breathiness*).

Sin embargo, existen evidencias científicas sobre la dificultad de lograr calificaciones fiables dado que dependen de estándares internos del examinador, de su sensibilidad para una determinada dimensión, de su fatiga, atención, exposición a variadas voces patológicas y su entrenamiento en la calificación perceptual de la calidad vocal<sup>1,2</sup>. Dado que la valoración perceptual presenta sesgos subjetivos y otras posibles inexactitudes, muchos investigadores han tratado de correlacionar los resultados de las medidas fonético-acústicas con las calificaciones de la calidad vocal y severidad de la disfonía. Este hecho ha sido tremendamente favorecido por la aparición de los sistemas de grabación digital, la disponibilidad de algoritmos de análisis automatizados y que los parámetros acústicos se obtengan de forma no invasiva, a la vez que nos brindan de forma sencilla una cuantificación de la mejoría obtenida durante el proceso terapéutico. Todo ello originó un considerable interés

en la cuantificación clínica de la calidad vocal mediante técnicas de análisis acústico. Sin embargo, la profusión de parámetros, la complejidad de su determinación y la no muy clara asociación clínica con el trastorno vocal hizo que cayera en desuso. Además, los parámetros comúnmente usados derivan de análisis que se basan en la detección de la periodicidad. Esto reduce la fiabilidad de los mismos en el análisis de aquellas voces que presentan una aperiodicidad significativa<sup>3</sup>.

Se pueden encontrar más de 100 algoritmos acústicos y un gran número de programas informáticos que calculan distintos parámetros vocales. El hecho de que las correlaciones entre la calidad global de la voz y los parámetros acústicos varíen sustancialmente, han puesto en entredicho la validez y la utilidad práctica de esos determinantes acústicos. A este respecto, se estudió mediante el coeficiente de correlación la relación entre las medidas acústicas y los juicios perceptuales del examinador de la severidad de una disfonía en un estudio metaanalítico de la multitud de parámetros descritos para ser utilizados en el análisis acústico de la voz. Maryn et al.<sup>4</sup> calcularon la relación entre dichos parámetros y las valoraciones perceptuales globales de la calidad vocal y se organizaron los parámetros de acuerdo a una jerarquía basada en los resultados del análisis estadístico. Los parámetros que mostraron un coeficiente de correlación con la severidad de la disfonía más alto estaban relacionados con el pico cepstral, tanto en vocales sostenidas como en habla conectada. Derivado de este estudio se desarrolló un constructo multivariante denominado índice acústico de calidad vocal (*acoustic voice quality index* [AVQI]).

**Tabla 1** Diagnósticos del grupo de pacientes que se incluyen en el estudio

Patología	n
Pólipos	15
Malformación	12
Edema	11
Quiste subepitelial	8
Nódulos vocales	4
Pseudoquiste seroso	4
Leucoplasia	1
Parálisis vocal	1
Granuloma	1
Papilomatosis laríngea	1
Total	58

En este trabajo se estudia el AVQI, que descrito por Maryn et al.<sup>5</sup>, pretende cuantificar la severidad global de la disfonía teniendo en cuenta los hallazgos del metaanálisis arriba mencionado. El índice se propone aglutinar en un dato numérico único una serie de parámetros acústicos relevantes, donde el mayor peso recae sobre el análisis cepstral de la señal vocal y puede suponer un cambio en el paradigma del análisis acústico de la disfonía.

## Material y métodos

### Sujetos

Se estudió un grupo de 58 pacientes con disfonía que acudieron a la consulta de voz para su valoración. En la [tabla 1](#) se resumen los diagnósticos finales a los que se llegó. Seis muestras vocales no pudieron ser finalmente analizadas por defectos técnicos en la grabación. Como «voces control» se realizó la grabación de voluntarios sanos hasta completar 60 casos, con edades entre los 20 y los 60 años. Se considera como individuo sano aquel que asegura no presentar síntomas o molestias en la voz, ni antecedentes de intervenciones quirúrgicas sobre las cuerdas vocales.

### Grabación de las voces

A todos los participantes se les pidió que emitieran una vocal /e/ sostenida y estable durante 3 segundos y leer las frases del método CAPE-V de evaluación auditivo-perceptual de la voz en español<sup>6</sup>. Ambas muestras vocales fueron grabadas en una habitación silenciosa (nivel de ruido ambiente menor de 50 dB) utilizando con una grabadora *digital IC Recorder* de SONY (ICD PX820), de la que obtenemos las grabaciones en formato mp3, compatible con el programa Praat. La frecuencia de muestreo fue de 44,1 kHz y la resolución de 16 bits. Para realizar la grabación, se colocó la grabadora a unos 8 cm de distancia de la boca del sujeto, manteniéndose este de pie, con los brazos y los hombros relajados. Una vez colocados, se le pidió tomar aire, y a continuación emitir la vocal /e/ de manera continua, con un tono y una frecuencia normal y con una duración de entre 3 y 5 segundos.

## Análisis perceptual de las voces

El análisis auditivo-perceptual de las muestras vocales se llevó a cabo por dos especialistas familiarizados con el método. La serie de voces a calificar se almacenaron en dos secuencias diferentes, determinadas de forma aleatoria. El observador 1 calificó con el método GRABS toda la serie escuchando para ello exclusivamente la grabación de la vocal sostenida. Siete días después calificó la misma serie con el método CAPE-V, escuchando ya todas las tareas grabadas. El observador 2 llevó a cabo las mismas puntuaciones. Para la calificación, las voces fueron reproducidas por los observadores en un ambiente exento de ruido y las escucharon en los altavoces del ordenador o auriculares en bloques de 20, con un intervalo de descanso entre los bloques consecutivos de 10 min. Tras la escucha de cada voz, el observador procedió a su calificación sin compartir información con el otro, pudiendo repetir la escucha cuantas veces considerara necesario. Tampoco conocían el diagnóstico del paciente, ni que se encontraban voces normales entre las de los pacientes, puesto que podría comprometer la fiabilidad de las calificaciones.

## Análisis acústico

Tras la grabación vocal, se comenzó a realizar el análisis acústico a través de un programa informático denominado *Praat*, el cual es un programa de libre distribución, multiplataforma, de código abierto y gratuito. Diseñado principalmente para realizar investigaciones en el campo de la fonética fue desarrollado en la Universidad de Ámsterdam por Paul Boersma y David Weenink a partir del año 1992. El *Praat*<sup>7,8</sup> es un software que presenta muchos propósitos, permitiendo el análisis acústico, síntesis articulatoria, procesamiento estadístico de datos, edición y manipulación de señales de audio, etc. Incluso permite que el usuario pueda crear sus propias rutinas y añadirlas a los menús del programa.

A continuación, se analizan las grabaciones con el programa Praat. En la ventana de *Praat objects* también se puede pedir al programa que ejecute el análisis del espectro, donde encontraremos el «espectro promediado a largo plazo» (LTAS) y dentro de él obtener los valores del «*slope*» y el «*tilt*». También se puede pedir el análisis de la periodicidad, donde se encuentra el *powerCepstogram*, y luego obtener la cifra de «prominencia del pico cepstral suavizado» (CPPS). En este paso se obtienen los siguientes parámetros que están incluidos en el cálculo del AVQI: 1. Pendiente espectral general (*general spectral slope, slope*): diferencia entre la energía en el rango 0-1 kHz y la energía en el rango 1-10 kHz del LTAS. 2. Inclínación de la línea de tendencia espectral (*spectral trend-line inclination, Tilt*): diferencia entre la energía en el rango 0-1 kHz y la energía en el rango 1-10 kHz de la línea de tendencia a lo largo del LTAS. 3. Prominencia del pico cepstral suavizado (*smoothed cepstral peak prominence, [CPPS]*): distancia entre el pico del primer *rahmónico* y el punto de *quefrecny* equivalente en la línea de regresión a través del cepstrum suavizado.

Además de poder ver el espectrograma de la voz que se está analizando, con el *Praat* también se pueden obtener datos cuantitativos de los diferentes parámetros acústicos

de la voz: de entre todos los parámetros que nos ofrece el análisis del *Praat*, se recogen los que se incluyen en la ecuación de regresión que calcula el AVQI: 1. *Shimmer Local*: diferencia media absoluta entre las amplitudes de periodos sucesivos dividida por la amplitud promedio. 2. *Shimmer Local dB*: especificada en el programa como el logaritmo de base 10 de las diferencias entre las amplitudes de los sucesivos periodos multiplicados por 20. 3. Relación armónico ruido media (*mean harmonics to noise ratio [HNR]*): logaritmo de base 10 del ratio entre la energía periódica y la energía del ruido multiplicado por 10.

Una vez obtenidos los parámetros antes mencionados, ya se puede determinar el AVQI, que se calcula aplicando una fórmula de regresión (obtenida mediante un método de análisis de regresión lineal múltiple) descrita por Maryn et al.<sup>5</sup>, en la que se introducen los resultados de los marcadores acústicos que se obtienen de la forma antes descrita:

$$\text{AVQI} = 9,072 - 0,245 \times \text{CPPS} - 0,161 \times \text{HNR} - 0,470 \times \text{SL} \\ + 6,158 \times \text{SLdB} - 0,071 \times \text{Slope} - 0,170 \times \text{Tilt}$$

### Análisis estadístico

Los resultados obtenidos en el programa Praat y las puntuaciones del análisis perceptual de las voces se introdujeron en una hoja de cálculo de Microsoft Office Excel 2007 (Microsoft Office Excel 2007 para Windows. Microsoft Corporation. Redmond, WA, EE. UU.). Los datos recogidos y almacenados en la base de datos se transfirieron al programa de análisis estadístico IBM SPSS 20 (*Statistical Package for the Social Sciences*). Para el análisis de los resultados, y en concreto, para establecer el grado de correlación entre los parámetros hallados con los grados de severidad de la disfonía se empleó el coeficiente de correlación de Pearson, que es una medida de la relación lineal entre dos variables aleatorias cuantitativas, que establece el grado de relación de dos variables siempre y cuando ambas sean cuantitativas. El valor del índice de correlación varía en el intervalo [-1,1]. Si  $r = 1$ , existe una correlación positiva perfecta, dependencia total entre las dos variables denominada relación directa por cuanto si una aumenta la otra también lo hace en proporción constante. Si  $0 < r < 1$ , existe una correlación positiva. Si  $r = 0$ , no existe relación lineal. Si  $-1 < r < 0$ , existe una correlación negativa. Si  $r = -1$ , existe una correlación negativa perfecta, que demuestra una dependencia total entre las variables en forma de una relación inversa. Aunque no existe un criterio firme o consenso universal para evaluar la magnitud de los coeficientes de correlación<sup>9</sup>, se escogió la cifra de corte de  $r = 0,60$  para distinguir entre un fuerte o un moderado grado de correlación entre los parámetros analizados<sup>10</sup>. Para la determinación de las diferencias entre dos medias muestrales y para la construcción del intervalo de confianza para la diferencia entre las medias de dos poblaciones se empleó la «t de student» para muestras pareadas con el fin de comparar los resultados pre- y postoperatorios y la «t de student» para muestras independientes con el fin de comparar el resultado del AVQI obtenido en los controles sanos con los pacientes, AVQI pre- y postoperatorio y AVQI obtenido del análisis de vocal sostenida versus frase.

**Tabla 2** Estadística descriptiva de los resultados del AVQI

	AVQI Casos	AVQI Controles
N	52	60
Media	7,70	6,23
Mediana	7,76	6,50
Desviación estándar	0,90	0,70
Rango	4,10	3,60
Mínimo	5,70	4,10
Máximo	9,80	7,70

### Resultados

A. Estadística descriptiva. En la [tabla 2](#) se exponen los resultados de la estadística descriptiva tanto de las voces sanas, como de las patológicas. Existe una diferencia significativa ( $t[95] = 9,5$ ;  $p < 0,000$ ) entre ambos grupos, lo que demuestra que el índice AVQI es capaz de distinguir una voz normal de otra patológica.

B. Correspondencia entre el valor del AVQI y el grado de severidad global de la disfonía ([tabla 3](#)). El valor de la correlación entre el AVQI y el grado global severidad de la disfonía encontrado fue:  $r = 0,68$  ( $p < 0,000$ ), y entre el AVQI y el parámetro B del GRABS:  $r = 0,68$  ( $p < 0,000$ ).

C. Correspondencia entre el valor del AVQI y el espectrograma de banda estrecha. El valor de la correlación entre el AVQI y el grado de Yanagihara del espectrograma de banda estrecha hallado fue:  $r = 0,76$  ( $p < 0,000$ ).

D. Análisis del habla conectada. Los resultados de la comparación entre el índice obtenido mediante el estudio de una vocal sostenida y mediante una frase («Irene adora hacer huevos al horno») demuestran que existe una diferencia estadísticamente significativa entre ambas muestras, al exhibir una distinta severidad objetivada mediante el AVQI, siendo mayor en el caso de las frases  $t(29) = -12,3$ ;  $p < 0,000$ . En el AVQI obtenido de las vocales sostenidas se encontró una media de 7,3 con una desviación estándar de 1,07 y un rango entre 5,3 y 9,8. En el AVQI obtenido de las frases se encontró una media de 9,7 con una desviación estándar de 0,70 y un rango entre 8,5 y 11,6.

E. Análisis longitudinal del AVQI. Se realizó una comparación entre las puntuaciones del AVQI pre- y postoperatorio en un grupo de 20 pacientes que fueron intervenidos fonocirúrgicamente (9 pólipos, 4 edemas de Reinke, 6 quistes subepiteliales, uno nódulos vocales). En el AVQI obtenido de las voces preoperatorias se encontró una media de 7,8 con una desviación estándar de 0,84 y un rango entre 6,0 y 9,6. En el AVQI obtenido de las voces postoperatorias se encontró una media de 7,05 con una desviación estándar de 1,12 y un rango entre 5,3 y 9,3. De la comparación entre el AVQI pre- y el postoperatorio se obtiene que existe una diferencia significativa, con menores valores en el postoperatorio:  $t(19) = 2,47$ ;  $p < 0,023$ .

### Discusión

Los métodos de análisis acústico tienen una larga historia en la valoración clínica de la voz. Además de las medidas de la frecuencia fundamental y de la intensidad vocal para objetivar el tono y la intensidad vocal, se han propuesto



**Tabla 3** Correspondencia entre el valor del AVQI y el grado de severidad global de la disfonía

	Correlaciones			
	AVQI	G	B	Yanagihara
AVQI				
Correlación de Pearson	1	0,684	0,684	0,767
Sig. (2-tailed)		0,000	0,000	0,000
N	112	112	112	112

numerosos parámetros para objetivar el tipo y severidad de la disfonía<sup>11</sup>. El análisis objetivo de la voz es particularmente atractivo debido a que es barato, no es invasivo y es fácil de ejecutar, permite obtener un resultado numérico, con lo que se puede hacer un seguimiento de los resultados del tratamiento, y permite compartir esta información entre los distintos profesionales implicados. Sin embargo, su uso en la clínica diaria no se corresponde con las expectativas en cuanto a su generalización, especialmente tras la aparición de programas informáticos al alcance de todos, hecho que se explica por la falta de correlación entre los parámetros y la severidad de la disfonía percibida, por una parte, y por la multitud de parámetros descritos en diferentes programas, por otra.

Uno de los puntos débiles de las medidas de la perturbación y del ruido (jitter, shimmer, NNE, HNR) es que dependen de un análisis temporal para separar la señal vocal en periodos de tono discretos. Aun en vocales prolongadas pueden producirse, de forma ocasional, errores de análisis de la duración precisa de cada periodo; incluso, de forma más frecuente, la señal vocal presenta elementos significativos de ruido que alteran la periodicidad. Este problema hace imposible estudiar mediante estos parámetros las voces moderadas o severamente disfónicas, lo cual supone una gran limitación de los parámetros arriba indicados. Por otro lado, la necesidad de contar con un tono estable hace que estas medidas de perturbación sean inapropiadas para estudiar muestras de habla conectada, donde la frecuencia cambia constantemente. Las medidas de ruido precisan una posición vocal constante, donde los armónicos puedan ser definidos, así como una frecuencia fundamental del habla estable. De esta forma, con el fin de obtener unas medidas más exactas del ruido en las voces, se precisa de un abordaje analítico que no sea dependiente del tiempo, como es el análisis espectral a largo plazo y el cepstrum.

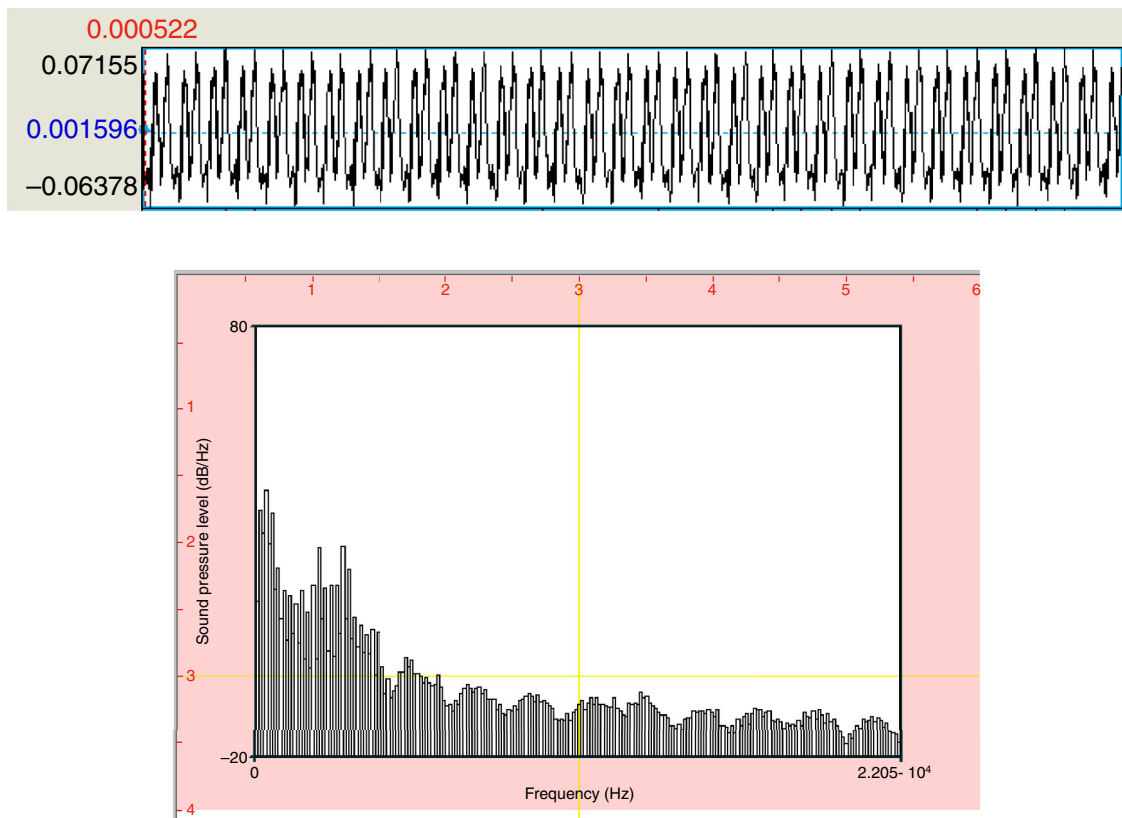
El análisis espectral a largo plazo tiene el potencial de medir el espectro del sonido de una muestra moderadamente prolongada. En este tipo de análisis, se aplica una transformada discreta de Fourier para calcular un espectro promediado a largo plazo integrado sobre toda una muestra de fonación y convierte la señal hablada del dominio temporal al dominio frecuencial<sup>12</sup> (fig. 1).

La transformada de Fourier es un espectro logarítmico que presenta la energía en frecuencias armónicamente relacionadas por medio de la separación de los componentes de amplitud y frecuencia de una onda compleja representada en tiempo y amplitud<sup>13</sup>. Utilizando la transformada de Fourier como base, el análisis cepstral es la transformada de Fourier de la transformada de Fourier descrita

anteriormente. El cepstrum es un espectro magnitud versus «quefrecy» (tiempo), medido en dB y milisegundos respectivamente. Creando un espectro logarítmico de un espectro logarítmico previo, el análisis cepstral puede mostrar una estructura armónica bien definida con un robusto componente de frecuencia fundamental y un ruido reducido tanto en vocales sostenidas, como en muestras de habla conectada producidas por un sujeto normal (fig. 2). De la relación entre la «quefrecy» y la magnitud cepstral se obtiene una línea de regresión lineal muy útil para normalizar la amplitud global de la señal (fig. 3). Las señales periódicas se asocian a los picos cepstrales más prominentes o de mayor amplitud comparados con la línea de regresión, mientras que una señal aperiódica se relaciona con una menor amplitud del pico cepstral en relación con dicha línea de regresión<sup>14</sup>. De esta forma, la cantidad de ruido o energía en una muestra de habla conectada puede ser cuantificada mediante la medida de la distancia entre el pico cepstral más prominente y la línea de regresión. Esta medida se denomina «prominencia del pico cepstral», o CPP (fig. 3).

En conjunto, las medidas cepstrales y las del «espectro promediado a largo plazo» tienen menos posibilidad de error debido a que la muestra se integra en toda su duración, a diferencia con las medidas dependientes del tiempo que miden la diferencia de ciclo a ciclo y que se basan en una prolongada colección de periodos tonales individuales. Así, los algoritmos espectrales/cepstrales teóricamente presentan un gran potencial para el estudio del habla conectada, al contrario de lo que ocurre en esta circunstancia con las medidas basadas en las diferencias ciclo a ciclo (jitter, shimmer, HNR).

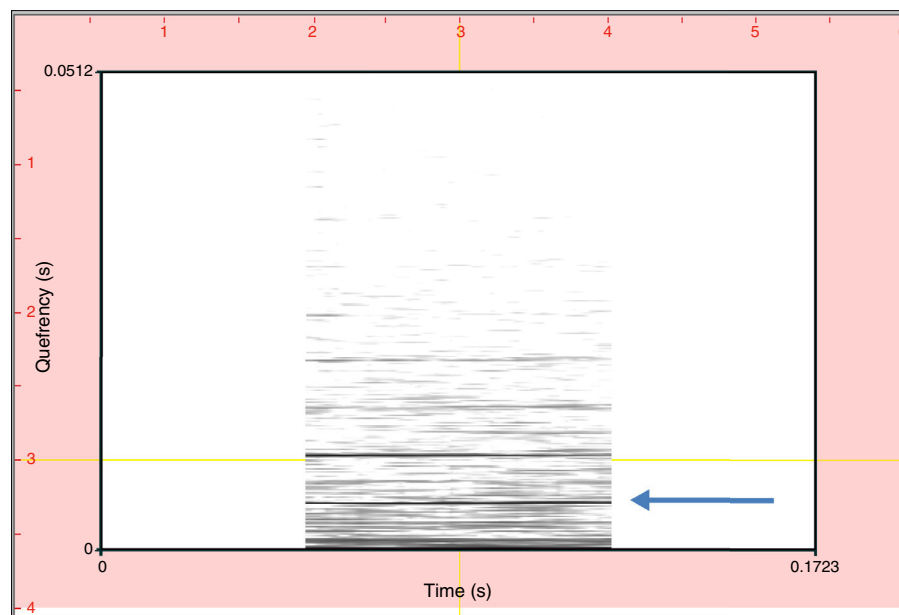
Lo más importante es averiguar si cualquiera de esas medidas acústicas es capaz de ofrecer parámetros objetivos clínicos útiles en el diagnóstico y tratamiento de los trastornos vocales. Recientemente se ha diseñado un método de análisis acústico de la voz que es capaz de cuantificar la severidad global de una disfonía: AVQI. Este índice se desarrolló para cuantificar la disfonía tanto en habla conectada como en vocales sostenidas. Las vocales sostenidas inducen una acción estable en los niveles subglótico, glótico y supraglótico, al contrario del habla conectada que se caracteriza por variaciones espectrales y temporales causadas por los inicios y ceses de la producción de vocales, fonemas sordos, contexto fonético, modulaciones prosódicas en la frecuencia fundamental e intensidad, tempo del habla, entre otros fenómenos. Dado que el comportamiento vocal difiere considerablemente entre estas dos tareas de fonación y habla, la percepción del tipo y severidad de la disfonía puede ser distinta según el tipo de tarea vocal o habla observada. Aunque



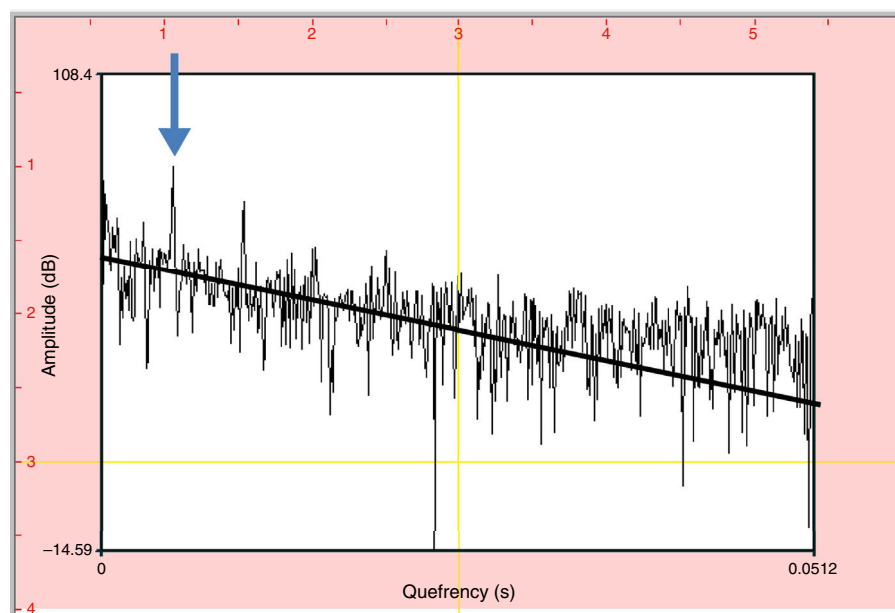
**Figura 1** Espectro promediado a largo plazo obtenido de la conversión de la señal del dominio temporal al dominio frecuencial mediante la transformada discreta de Fourier.

dos estudios no encuentran una diferencia significativa en la percepción auditiva de la severidad entre vocales sostenidas y habla conectada<sup>15,16</sup>, otros estudios sí muestran diferencias en la percepción auditiva de la disfonía entre vocales y

habla conectada<sup>17,6</sup> al encontrar que los observadores califican la disfonía, especialmente su calidad aérea, de forma más severa en las vocales sostenidas. Estos hallazgos apoyan que es preciso grabar ambas tareas de fonación y habla si



**Figura 2** Power cepstrogram obtenido de aplicar la transformada de Fourier sobre la anterior representada en la figura 1. La flecha señala en primer harmónico.



**Figura 3** CPPS obtenido desde el Power cepstrum representado en la figura 2. La línea recta es la línea de regresión lineal, la flecha señala el pico cepstral más prominente, o primer *rahmónico*.

se pretende realizar un análisis perceptual e instrumental de una forma correcta. En este trabajo se obtuvieron distintos resultados cuando se analizan voces sostenidas o habla conectada, hecho que refuerza la opinión de que es preciso estudiar ambas muestras para obtener resultados que reflejen la situación real de la calidad vocal de una persona, no obstante, es deseable profundizar en el estudio del habla conectada mediante el AVQI, aunque existan evidencias que apuntan a una mayor fiabilidad cuando se analiza el habla conectada que cuando se analiza una vocal sostenida<sup>3</sup>. Dado que en la actualidad está cambiando el paradigma del análisis acústico de la disfonía, se recomienda abandonar las medidas basadas en los parámetros tradicionales para pasar a los parámetros derivados de la métrica cepstral como es el CPP (parámetro más importante del AVQI), que permiten el análisis del habla.

El AVQI es un constructo multivariante que combina múltiples marcadores acústicos para dar como resultado un único número que se correlaciona razonablemente con la severidad global de la disfonía. Este abordaje multiparamétrico está motivado por la naturaleza multidimensional de la calidad vocal y por el hecho de que no se relaciona con una única variable física o un único determinante psicoacústico.

Dicho abordaje multiparamétrico contrasta con el anterior concepto de intentar correlacionar un único parámetro acústico con la calidad vocal tal y como se correlaciona la frecuencia fundamental con el tono o la intensidad con el volumen de la voz. La calidad vocal percibida lleva todas las dimensiones perceptuales del contorno del espectro y sus cambios en el tiempo, por lo que es posible que tanto las medidas en el dominio tiempo como en el dominio frecuencia son las máximas responsables de producir la percepción de la calidad vocal<sup>18</sup>. Además, los métodos de correlación bivalente utilizados para estimar la relación proporcional entre la calificación auditivo-perceptual y un único marcador acústico tienen un insuficiente peso<sup>4</sup>. Este hecho ha

determinado que se empleen métodos estadísticos multivariantes para combinar las ventajas de las múltiples medidas disponibles con el objeto de aumentar la validez del análisis objetivo-instrumental de la calidad vocal y así ser capaces de ganar precisión en la discriminación entre los diferentes niveles de la severidad de la disfonía<sup>19</sup>.

Para construir un modelo estadístico que represente la mejor combinación de los parámetros que mejor predicen el grado de severidad de una disfonía solamente seis fueron elegidos de entre 13 medidas tras aplicar métodos estadísticos de análisis de la regresión lineal. Como resultado se obtuvo una ecuación de regresión que incluye marcadores acústicos en el dominio temporal, en el dominio frecuencial y en el dominio de la *quefrecy*, con lo que es una representación multidimensional de la severidad de la disfonía. Los parámetros que recoge la ecuación denominada AVQI son: la CPPS, la HNR, el shimmer local, el shimmer local en dB, la pendiente del espectro promediado a largo plazo (LTAS) y la inclinación de la línea de tendencia a través del LTAS<sup>5</sup>. El parámetro más importante de este modelo AVQI, es la versión suavizada de la prominencia del pico cepstral (CPPS). Esta medida representa la distancia entre el primer pico *rahmónico* y el punto con la misma *quefrecy* en la línea de regresión a través del cepstrum suavizado. Este marcador acústico se fundamenta en que cuanto más periódica es una señal vocal, mejor configuración armónica tiene definida en el espectro, y consecuentemente, se obtiene un pico cepstral más prominente. Desde su introducción en el campo de la cuantificación de la calidad vocal por Hillenbrand<sup>20</sup> se ha comprobado que es una medida fiable y válida de la calidad vocal global, especialmente de la calidad aérea, en numerosos estudios<sup>21</sup>. Este parámetro aporta valiosa información acerca del grado en el cual puede ser detectada una estructura armónica bien definida en la señal vocal, sin basarse en la detección de la periodicidad. El CPPS predice mejor que el jitter y el shimmer la disfonía<sup>22</sup>.



Adicionalmente, Maryn et al.<sup>4</sup> realizaron un metaanálisis sobre los coeficientes de correlación entre las calificaciones auditivo-perceptuales del grado global de severidad de la disfonía y 69 medidas acústicas obtenidas de vocales sostenidas y 26 obtenidas de habla conectada. Solamente cuatro marcadores acústicos en vocales sostenidas y tres en habla conectada satisficieron los criterios metaanalíticos para ser considerados como las medidas más prometedoras del análisis acústico de la voz. La prominencia del pico cepstral suavizada (CPPS), sin embargo, fue la única medida acústica que obtuvo la suficiente validez tanto en vocal sostenida, como en habla conectada, por lo que se considera como una medida superior de la severidad de la disfonía. En combinación con los otros cinco parámetros acústicos del AVQI, el CPPS aporta una sólida base para el estudio objetivo y cuantitativo de la severidad de la disfonía. Dado que el AVQI se desarrolló con la ayuda de los paquetes informáticos gratuitos que se pueden bajar libremente de la web (programa *Praat*) está al alcance de la mayoría de los profesionales.

De los resultados del presente trabajo se demuestra que el AVQI discrimina las voces patológicas de las normales tomando como referencia la calidad global percibida por el examinador, de esta forma, el índice se puede utilizar para documentar de forma longitudinal los resultados de los tratamientos vocales. También se observa buena correlación entre el resultado numérico del índice y la clasificación de Yanagihara del espectrograma de banda estrecha, lo que apoya, junto con la calificación perceptual global, su validez. En esta serie existen distintos resultados cuando se analizan voces sostenidas o habla conectada, este hecho refuerza la opinión de que es preciso estudiar ambas muestras para obtener resultados que reflejen la situación real de la calidad vocal de una persona, no obstante es deseable profundizar en el estudio del habla conectada mediante el AVQI.

En conclusión, el AVQI es una medida multivariante, accesible, factible y razonablemente válida para medir clínicamente la severidad global de la severidad de la disfonía en muestras de vocales sostenidas y habla conectada.

## Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

## Bibliografía

- Eadie TL, Baylor CR. The effect of perceptual training on inexperienced listeners' judgments of dysphonic voice. *J. Voice*. 2006;20:527-44.
- Núñez-Batalla F, Díaz-Molina JP, García-López I, Moreno-Méndez A, Costales-Marcos M, Moreno-Galindo C, et al. El espectrograma de banda estrecha como ayuda para el aprendizaje del método GRABS de análisis perceptual de la disfonía. *Acta Otorrinolaringol Esp*. 2012;63:173-9.
- Halberstam B. Acoustic and perceptual parameters relating to connected speech are more reliable measures of hoarseness than parameters relating to sustained vowels. *ORL*. 2004;66:70-3.
- Maryn Y, de Bodt M, van Cauwenberge PB, Corthals P. Acoustic measurement of overall voice quality: A meta-analysis. *J Acoust Soc Am*. 2009;126, 2619-1634.
- Maryn Y, Weenink D. Objective dysphonia measures in the program Praat: smoothed cepstral peak prominence and Acoustic Voice Quality Index. *Journal of Voice*. 2015;29:35-43.
- Núñez-Batalla F, Morato-Galán M, García-López I, Avila-Menéndez A. Validation of the Spanish adaptation of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). *Acta Otorrinolaringol Esp*. 2015;66:249-57.
- Boersma P, Weenink D. Praat Manual. Amsterdam: University of Amsterdam, Phonetic Science Department. 2006. Disponible en: <http://www.fon.hum.uva.nl/praat/manualsByOthers.html>.
- Boersma P, Weenink D. Praat: doing phonetics by computer [Computer program]. Version 6.0.07 [consultado 30 Nov 2015]. Disponible en: <http://www.praat.org/>
- Frey LR, Botan CH, Friedman PG, Krepes GL. Investigating communication: An introduction to research methods. Englewood Cliffs, NJ: PrenticeHall; 1991.
- Franzblau AN. A primer of statistics for non-statisticians. New York: Harcourt, Brace & Company; 1958.
- Buder EH. Acoustic analysis of voice quality: a tabulation of algorithms 1902-1990. En: Kent RD, Ball MJ, editores. Voice quality measurement. San Diego, CA: Singular Publishing Group; 2000. p. 119-244.
- Baken RJ, Orlikoff RF. Clinical measurement of speech and voice. 2. nd ed San Diego, CA: Singular Publishing Group; 2000.
- Hillenbrand J, Houde RA. Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *J Speech Lang Hear Res*. 1996;39:311-21.
- Awan SN, Roy N, Jette ME, Meltzner GS, Hillman RE. Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: comparisons with auditory perceptual judgments from the CAPE-V. *Clin Linguist Phon*. 2010;24:742-58.
- de Krom G. Consistency and reliability of voice quality ratings for different types of speech fragments. *J Speech Hear Res*. 1994;37:985-1000.
- Revis J, Giovanni A, Wuyts F, Triglia J. Comparison of different voice samples for perceptual analysis. *Folia Phoniatr Logop*. 1999;51:108-16.
- Maryn Y, Roy N. Sustained vowels and running speech in the auditory-perceptual evaluation of dysphonia severity. *J Soc Bras Fonoaudiol*. 2012;24:107-12.
- Kreiman J, Gerratt BR. Validity of rating scale measures of voice quality. *J Acoust Soc Am*. 1998;104:1598-608.
- Awan SN, Roy N. Toward the development of an objective index of dysphonia severity: a four-factor acoustic model. *Clin Linguist Phon*. 2006;20:35-49.
- Hillenbrand J, Cleveland RA, Erickson R. Acoustic correlates of breathy vocal quality. *J Speech Hear Res*. 1994;37:769-78.
- Samlan RA, Story BH, Bunton K. Relation of perceived breathiness to laryngeal kinematics and acoustic measures based on computational modeling. *J Speech Lang Hear Res*. 2013;56:1209-23.
- Heman-Ackah YD, Heuer RJ, Michael DD, Ostrowski R, Horman M, Baroody MM, et al. Cepstral peak prominence: A more reliable measure of dysphonia. *Ann Otol Rhinol Laryngol*. 2003;112:324-33.