

Revista de Investigación en Logopedia



revistalogopedia.uclm.es
ISSN - 2174-5218

2 (2017) 108-127

Análisis acústico de la voz: medidas temporales, espectrales y cepstrales en la voz normal con el Praat en una muestra de hablantes de español

Jonathan Delgado 1, Nieves M^a. León 2, Alejandra Jiménez 2, y Laura M. Izquierdo 1

1 CREN Salud, Tenerife, España

2 Hospital Universitario de Canarias, Tenerife, España

Resumen

El análisis acústico es una herramienta que proporciona información objetiva sobre la voz. En los últimos años, medidas del espectro medio a largo plazo (LTAS) y cepstrales, como la prominencia del pico cepstral suavizado (CPPs), han complementado a las medidas utilizadas tradicionalmente, demostrando en multitud de estudios una alta correlación con el grado de severidad de la disfonía. El objetivo de este trabajo descriptivo fue calcular, en el Praat, los valores de normalidad de medidas temporales, espectrales y cepstrales en una muestra de 50 hablantes de español (25 hombres y 25 mujeres) atendiendo a los principales factores que influyen en su fiabilidad como el tipo micrófono, el nivel de ruido ambiental, el programa de análisis y los parámetros acústicos utilizados. Se realizaron dos muestras de voz para cada sujeto: 1) una /a/ sostenida con la que se calcularon la CPPs y los parámetros de la frecuencia fundamental (F0), de ruido y de perturbación de la frecuencia y de la amplitud, y 2) una muestra de habla conectada donde se calcularon la CPPs y las pendientes del LTAS. Los resultados del análisis con la vocal sostenida muestran diferencias significativas en función del sexo en la F0, el jitter absoluto y en todos los parámetros de la perturbación de la amplitud y del ruido. En habla conectada se observan diferencias significativas entre hombres y mujeres en la pendiente espectral obtenida a partir de la línea de tendencia a través del LTAS y en la CPPs.

Palabras clave: Análisis acústico; Praat; Voz normal,

Acoustic analysis of the voice: Temporal, spectral and cepstral measurements in the normal voice with the Praat in a sample of Spanish speakers

Abstract

Acoustic analysis is a tool that provides objective information about the voice. In recent years, long-term average spectrum (LTAS) and cepstral measurements, such as the smoothed cepstral peak prominence (CPPs), have complemented the traditional measures, demonstrating in many studies a high correlation with the degree of severity of dysphonia. The objective of this descriptive study was to calculate the normal values of temporal, spectral and cepstral measures in a sample of 50 Spanish speakers (25 men and 25 women) in the Praat, taking into account the main factors that influence their reliability as the microphone type, the ambient noise level, the analysis program and the acoustic parameters used. Two voice samples were made for each subject: 1) sustained /a/ with which the CPPs were calculated and the parameters of the fundamental frequency (F0), noise and frequency and amplitude perturbations, and 2) a connected speech sample where CPPs and LTAS slopes were calculated. The results of the analysis with the sustained vowel show significant differences as a function of the sex in the F0, the jitter absolute and in all the parameters of the perturbation of the amplitude and the noise. In connected speech, significant differences were observed between men and women in the spectral slope obtained from the trend line through the LTAS and the CPPs.

Key words: Acoustic analysis; Normal voice; Praat.

Correspondencia con los autores: jonathan@cren.es

Recibido 18 Agosto 2017. Primera revisión 6 Octubre 2017. Aceptado 10 Octubre 2017.

Introducción

El análisis acústico de la voz es una herramienta que proporciona información objetiva sobre la voz de forma no invasiva. Para su interpretación se requiere de valores de normalidad de referencia y su fiabilidad está sujeta a factores como el tipo de micrófono, el ruido ambiental, el software de análisis y los parámetros acústicos utilizados.

En cuanto al tipo de micrófono se ha demostrado que éste tiene un impacto significativo en el análisis acústico de la voz (Titze y Winholtz, 1993). Para obtener grabaciones con una calidad óptima, Barsties y de Bolt (2015) recomiendan usar: 1) micrófonos de condensador, preferiblemente de diadema, 2) cardioideos, 3) con un rango de frecuencia entre 20-20000 Hz, 4) que presenten una frecuencia de respuesta plana con una desviación máxima de 2 dB, 5) un nivel de ruido equivalente menor o igual a 25 dB(A), 6) un nivel máximo de presión sonora igual o superior 126 dB SPL para un factor de distorsión no lineal del 3%, y 7) una alta sensibilidad, con un mínimo de -60 dB.

Respecto al nivel de ruido ambiental, es imprescindible realizar una medición instrumental de éste en el entorno de grabación. Las medidas de la perturbación de la frecuencia y de la amplitud son muy sensibles a este tipo de ruido, incrementándose a medida que este ruido ambiental aumenta (Ingrisano, Perry, Jepson, 1998). Delisyski, Shaw y Evans (2005) recomiendan valores de la ratio señal-ruido (SNR) por encima de los 30 dB para tener mediciones de voz con una precisión por encima del 95%.

Existen muchos programas de análisis acústico que se utilizan para la evaluación clínica de la voz como el *Pratt*, el *Anagraf*, el *Dr. Speech* o el *Multidimensional Voice Program (MDVP)*, entre otros. El Praat es un programa de distribución libre ampliamente utilizado en la clínica y aunque los resultados de los parámetros acústicos (como la frecuencia fundamental o F_0 , el jitter o el shimmer) no son comparables numéricamente con otros programas debido a diferentes algoritmos en la extracción de la F_0 , se ha comprobado que presentan la misma fiabilidad que los obtenidos en el *Anagraf* (Elisei, 2012), en el *Dr. Speech* (Núñez-Batalla, González-Márquez, Peláez-González, González-Laborda, Fernández-Fernández y Morato-Galán, 2014) o en el *MDVP* (Oğuz, Kiliç y Şafac, 2011; Maryn, Corthals, De Bolt, Cauwenberge y Deliyski, 2009). En relación a las medidas cepstrales pasa lo mismo, la prominencia del pico cepstral suavizado (CPPs) obtenida en el *Praat* presenta la misma fiabilidad que la obtenida en el programa *Speech Tool* (Maryn y Weenick, 2015) y en el programa *Analysis of Dysphonia in Speech and Voice (ADSV)* (Sauder, Bretl y Eadie, 2017) aunque estas medidas no son comparables debidos a diferencias en el procesamiento de la señal de los programas.

Otro aspecto a tener en cuenta en relación a las medidas del espectro medio a largo plazo (LTAS) y a las medidas cepstrales es la configuración de los parámetros acústicos en el programa de análisis elegido. Por ejemplo, Sauder *et al.* (2017) utilizaron para el cálculo de la CPPs la configuración que viene por defecto en el *Praat* y Delgado-Hernández, León-Gómez, Izquierdo-Arteaga y Llanos-Fumero (2017) la configuración que se utiliza en el índice acústico de calidad vocal v. 3.01 (*Acoustic Voice Quality Index, AVQI*) (Barsties y Maryn, 2016). Los resultados son similares en ambos estudios pero los valores no son comparables numéricamente. Por tanto, para poder interpretar correctamente los resultados del análisis acústico es determinante

disponer de valores de referencia de las diferentes medidas acústicas del software de análisis utilizado, así como, conocer (en el caso de las medidas de LTAS y de la CPPs) la configuración que se usó en la obtención de esos valores normativos.

Parámetros acústicos

Tradicionalmente se han utilizado medidas basadas en la F_0 , en los parámetros de perturbación de la frecuencia y la amplitud y en la ratio armónico-ruido (HNR). Estas medidas a corto plazo dependientes del tiempo tienen algunas limitaciones que hay que tener en cuenta. Por un lado, estos parámetros acústicos necesitan de la detección exacta de los ciclos de la F_0 y, por tanto, son menos fiables en las voces menos armónicas, es decir, en voces con un mayor grado de disfonía. Una solución a este inconveniente la aporta el análisis cepstral. La CPPs es una medida de la periodicidad armónica espectral (Ferrer, De Bodt, Maryn, Van de Heyning, Hernández-Díaz, 2007) y no requiere una determinación exacta de los ciclos de la F_0 como el jitter, el shimmer o la HNR, lo que la hace particularmente adecuada para la evaluación de la calidad vocal, incluso en las voces más disfónicas (Moers, Möbius, Rosanowski, Nöth, Eysholdt y Haderlein, 2012).

Otra limitación la encontramos en el tipo de muestra de voz que se utiliza en los análisis. Normalmente se usan vocales sostenidas ya que su producción es más fácil y tienen menor influencia de la articulación o de variedades dialectales (Brinca, Batista, Tavares, Gonçalves y Moreno, 2014), pero tienen el inconveniente de que no representan la voz diaria del paciente. Por ello, medidas del espectro medio a largo plazo y cepstrales han ido haciéndose un hueco en la valoración acústica de la voz ya que se pueden calcular en habla conectada. Se ha demostrado una alta correlación con la severidad de la disfonía tanto de las medidas del LTAS (Lowell, Colton, Kelley y Hahn, 2011) como de las cepstrales (Delgado-Hernández et al., 2017; Sauder et al., 2017; Lowell et al., 2011).

Por todo lo comentado anteriormente, la evaluación acústica de la voz necesita: 1) de unas condiciones óptimas de grabación, 2) de valores de normalidad de referencia con los que comparar a las voces disfónicas, y 3) de la correcta

interpretación de múltiples medidas que nos permitan tener una visión global y fiable de la voz del paciente.

El objetivo de este trabajo descriptivo es obtener valores de normalidad en las principales medidas temporales, espectrales y cepstrales que se pueden realizar en el *Praat* en una muestra de hablantes de español.

Método

Participantes

En este estudio participaron voluntariamente 50 sujetos, 25 hombres y 25 mujeres, hablantes de español, normofónicos y sin antecedentes de patología vocal con una media de edad de 37.38 años (desviación típica 13,07). Todos los sujetos pasaron una evaluación perceptiva por parte de una logopeda especialista en rehabilitación de la voz con 20 años de experiencia.

Grabación de las voces

Se obtuvieron dos muestras de voz diferentes para cada sujeto de estudio, un total de 100 muestras: Una vocal sostenida /a/ durante 4 segundos aproximadamente y la lectura de las cuatro frases fonéticamente balanceadas del test de la matriz de frases para hablantes de español (Hochmuth, Brand, Zokoll, Castro, Wardenga y Kollmeier, 2012) utilizadas en el estudio de Delgado-Hernández et al. (2017). Las voces fueron grabadas con un interfaz de audio alessis i/o2 express y un micrófono de condensador de diadema AKG C544L, cardioide, con una gama de frecuencias de 20 a 20000 Hz, un nivel de ruido equivalente de 22 dB(A), un nivel de presión sonora límite para un factor de distorsión no lineal del 3% de 130 dB SPL y una sensibilidad de -28 dB. Se utilizó el programa *Praat* v. 5.4.22 (Boersma y Weenink, 2015) con una frecuencia de muestreo de 44.1 Khz y 16 bits de resolución. Para garantizar una calidad de grabación óptima en todos los audios, se calculó *post hoc* la ratio señal-ruido (*signal-to-noise ratio* [SNR]) propuesta por Deliyski et al. (2005).

Medidas acústicas

A continuación se describe cada una de las medidas acústicas estudiadas en este trabajo:

Parámetros de la frecuencia:

- *MeanF₀*: es la media del la F₀ y nos muestra el número de veces que las cuerdas vocales se abren y cierran por segundo.

- *F₀SD (F₀ Standar Desviation)*: hace referencia a la variabilidad de la F₀.

Parámetros de la perturbación de la frecuencia:

- *Jitter (local)*: representa la variabilidad absoluta de la frecuencia periodo a periodo, dividido por la frecuencia media total de los periodos de la señal.

- *Jitter (local, absolute)*: es la diferencia absoluta media entre periodos consecutivos.

- *RAP (Relative Average Perturbation)*: corresponde a la variabilidad relativa de la frecuencia entre un periodo y la media de los dos periodos contiguos, dividido por la frecuencia media total de los periodos de la señal. Aporta por tanto, un suavizado de tres ciclos.

- *PPQ5 (Five-Point Period Perturbation Quotient)*: corresponde a la variabilidad relativa de la frecuencia entre un periodo y la media de los cuatro periodos contiguos, dividido por la frecuencia media total de los periodos de la señal. Aporta por tanto, un suavizado de cinco ciclos.

Parámetros de la perturbación de la amplitud:

- *Shimmer (local)*: Representa la variabilidad absoluta entre la amplitud de periodos consecutivos, divididos por la media total de la amplitud.

- *Shimmer (local, absolute)*: es el logaritmo en base 10 de la media absoluta de la diferencia entre la amplitud de periodos consecutivos, multiplicado por 20.

- *APQ3 (Three-Point Amplitude Perturbation Quotient)*: Corresponde a la variabilidad absoluta entre la amplitud de un periodo y la media de las amplitudes de los dos periodos contiguos, divididos por la media total de la amplitud.

- *APQ5 (Five-Point Amplitude Perturbation Quotient)*: Corresponde a la variabilidad absoluta entre la amplitud de un periodo y la media de las amplitudes de los cuatro periodos contiguos, divididos por la media total de la amplitud.

- *APQ11 (Eleven-Point Amplitude Period Perturbation Quotient)*: Representa a la variabilidad absoluta entre la amplitud de un periodo y la media de las amplitudes de los diez periodos contiguos, divididos por la media total de la amplitud.

Parámetro de ruido:

- *HNR (Mean Harmonic to Noise Ratio)*: La relación armónico-ruido es la medida que cuantifica la cantidad de ruido aditivo en la señal de la voz. Es el logaritmo en base 10 de la ratio entre la energía periódica y la energía correspondiente al ruido, multiplicado por 10.

Parámetros del espectro medio a largo plazo:

- *LTAS_{slope}*: La pendiente espectral es la diferencia entre la energía de 0 a 1000 Hz y la energía de 1000 a 10000 Hz del LTAS (figura 1).

- *LTAS_{tilt-trendline}*: Representa la diferencia entre la energía de 0 a 1000 Hz y la energía de 1000 a 10000 Hz en la línea de tendencia a través del LTAS (figura 1).

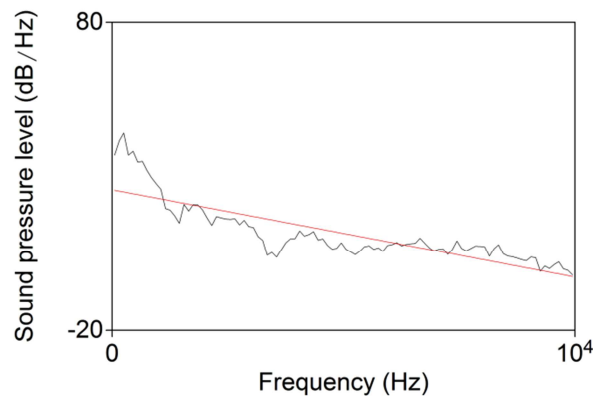


Figura 1. Representación del espectro medio a largo plazo (en negro) y de la línea de tendencia a través del LTAS (en rojo).

Parámetro del Cepstrum:

- *CPPs (Smoothed Cepstral Peak Prominence)*: La prominencia del pico cepstral suavizado representa la diferencia entre el pico cepstral más prominente, que

corresponde al primer *rahmónico* (*rahmonic*), y el punto con la misma *quefrecuencia* (*quefrecny*) sobre la línea de regresión a través del cepstrum suavizado (figura 2).

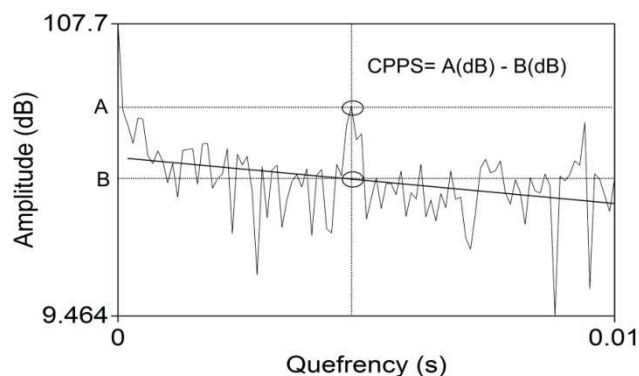


Figura 2. Representación gráfica del cepstrum y del cálculo de la CPPs (Delgado-Hernández et al., 2017). A= valor en dB del punto superior del primer *rahmónico* o pico cepstral, B= valor en dB del punto sobre la línea de regresión que tiene la misma *quefrecny* que el pico cepstral.

Análisis acústico

En la tabla 1 se enumeran los parámetros acústicos estudiados en este trabajo. Para el análisis acústico de las medidas temporales se utilizó el audio de la vocal sostenida /a/. Se tomaron los 3 segundos centrales de la producción en el espectrograma de banda ancha. Luego se calcularon con el *Praat* (Boersma y Weenink, 2015) los parámetros relacionados con la frecuencia (Mean F_0 , F_0SD), con la perturbación de la frecuencia (jitter, jitter absoluto, RAP y PPQ5), con la perturbación de la amplitud (shimmer, shimmer absoluto, APQ3, APQ5 y APQ11) y con el ruido aditivo en la señal (HNR). En el caso de las medidas espectrales, se utilizaron las grabaciones de las frases (habla conectada) y se calcularon los parámetros $LTAS_{slope}$ y $LTAS_{tilt-trendline}$. Para ello se utilizó la configuración utilizada en el AVQI v. 3.01 (Barsties y Maryn, 2016):

Configuración $LTAS_{slope}$: *Sound [frases] → analyse spectrum → to Ltas... → Bandwidth (Hz): 1 → Query → Get slope → Low band (Hz): 0, 1000; High band (Hz): 1000, 10000; Averaging method: energy.*

Configuración $LTAS_{tilt-trendline}$: *Sound [frases] → analyse spectrum → to Ltas... → Bandwidth (Hz): 1 → Compute Trend line... → Frequency range (Hz): 1, 10000 → Query*

→ *Get slope* → *Low band (Hz): 0, 1000; High band (Hz): 1000, 10000; Averaging method: energy.*

En el cálculo de la medida cepstral CPPs se utilizaron tanto las grabaciones de la vocal sostenida como las del habla conectada. Se obtuvieron, por tanto, dos valores de CPPs por sujeto, uno con la /a/ y otro con las frases. Para el cálculo de la CPPs se usó la configuración utilizada en el estudio de Delgado-Hernández et al. (2017):

Configuración CPPs: *Sound [vocal o frases] → analyse periodicity → PowerCepstrogram [Pitch floor= 60 Hz; Time step= 0.002 seg; Maximum frequency=5000 Hz; Pre-emphasis= 50 Hz] → Query → Get CPPS... [Substract tilt before smoothing= No; Time averaging window= 0.01 seg; Quefreny averaging window=0.001 seg; Peak search pitch range= 60-330 Hz; Tolerance (0-1)= 0.05; Interpolation=parabolic; Tilt line quefreny range= 0.001-0.0 seg; Line type= straight; Fit method= robust].*

Tabla 1. Parámetros acústicos estudiados.

Parámetros de la F_0	Media de la F_0
	Variabilidad de la F_0
Parámetros de perturbación de la frecuencia	Jitter
	Jitter absoluto
	RAP
	PPQ5
Parámetros de perturbación de la amplitud	Shimmer
	Shimmer absoluto
	APQ3
	APQ5
Parámetros de ruido	APQ11
	HNR
Parámetros espectrales	LTAS _{slope} (habla conectada)
	LTAS _{tilt-trendline} (habla conectada)
Parámetros cepstrales	CPPs (vocal sostenida)
	CPPs (habla conectada)

Análisis estadístico

El análisis de los resultados se realizó con el programa estadístico R v. 3.3.1 (R Core Team, Viena, Austria). Se calcularon las medias y desviaciones típicas de todos de los parámetros temporales, espectrales y cepstrales. Luego se realizaron comparaciones entre hombres y mujeres en cada medida acústica. En primer lugar se realizó la prueba Kolmogorov-Smirnov para comprobar la normalidad de la distribución de los datos. En aquellas medidas donde la distribución resultó ser normal ($p > 0.05$), se estudió la homocedasticidad de las varianzas con la prueba de Levene y se calculó el estadístico paramétrico t de student. En las medidas donde la distribución no cumplió el criterio de normalidad se calculó el estadístico no paramétrico U de Mann Whitney.

Resultados

Las grabaciones realizadas en este trabajo muestran una SNR media de 39.72 dB (DT 5.10), por tanto, tienen una precisión por encima del 95%. Para el análisis acústico se consideran óptimos los valores superiores a 30 dB (Deliyski et al. 2005). En la tabla 2 podemos observar las medias, desviaciones típicas (DT) e intervalos con ± 1.5 DT de los parámetros acústicos estudiados en este trabajo.

En la figura 3 se muestra el análisis exploratorio de los datos de la media de la F_0 , el jitter, el shimmer y la HNR. Como era de esperar se encontraron diferencias significativas entre el valor medio de F_0 entre hombres y mujeres (*Media F_0* : $U=0$, $p=0.000$). Además de una mayor F_0 , se encontró una mayor variabilidad en el grupo de las mujeres (F_0DS : $t_{(48)}=-2.65$, $p=0.01$).

Los parámetros de la perturbación de la frecuencia no muestran diferencias en cuanto al sexo (*jitter*: $t_{(48)}=1.26$, $p=0.12$; *RAP*: $U=380.5$, $p=0.018$; *PPQ5*: $U=392.5$, $p=0.12$), excepto el jitter absoluto donde se observan mayores valores en hombres así como mayor variabilidad (*jitter absoluto*: $U=554$, $p=0.000$). En cambio, los parámetros de perturbación de la amplitud muestran diferencias estadísticamente significativas en función del género, con valores más altos en el grupo de los hombres (*shimmer*: $U=479$, $p=0.001$; *shimmer absoluto*: $U=458.5$, $p=0.004$; *APQ3*: $t_{(35.14)}=2.73$, $p=0.009$; *APQ5*: $t_{(34.60)}=3.47$, $p=0.001$; *APQ11*: $U=546.5$, $p=0.000$).

Tabla 2. Medias, desviaciones típicas (DT) e intervalos con ± 1.5 DT.

Medidas acústicas	Total (n=50)		Hombres (n=25)		Mujeres (n=25)	
	Media (DT)	± 1.5 DT	Media (DT)	± 1.5 DT	Media (DT)	± 1.5 DT
Mean F ₀ (Hz)	152.47 (47.13)	[81.77-223.16]	109.99 (15.08)	[87.37-132.61]	194.94 (23.44)	[171.50-230.10]
F ₀ SD (Hz)	1.42 (0.63)	[0.47-2.36]	1.20 (0.48)	[0.48-1.92]	1.65 (0.69)	[0.61-2.68]
Jitter (%)	0.45 (0.18)	[0.18-0.72]	0.49 (0.19)	[0.20-0.77]	0.41 (0.17)	[0.15-0.66]
Jitter abs (μ s)	34.60 (21.91)	[1.73-67.46]	47.22 (23.22)	[12.39-82.05]	21.97 (10.44)	[6.31-37.67]
RAP (%)	0.26 (0.11)	[0.09-0.42]	0.27 (0.11)	[0.10-0.43]	0.24 (0.11)	[0.07-0.40]
PPQ5 (%)	0.26 (0.10)	[0.11-0.41]	0.28 (0.11)	[0.11-0.44]	0.23 (0.10)	[0.08-0.38]
Shimmer (%)	1.93 (0.97)	[0.47-3.38]	2.38 (1.10)	[0.73-4.03]	1.48 (0.54)	[0.67-2.29]
Shimmer (dB)	0.17 (0.08)	[0.05-0.29]	0.21 (0.10)	[0.06-0.36]	0.14 (0.05)	[0.06-0.21]
APQ3 (%)	0.93 (0.46)	[0.24-1.62]	1.09 (0.55)	[0.26-1.91]	0.76 (0.27)	[0.35-1.16]
APQ5 (%)	1.03 (0.48)	[0.31-1.75]	1.25 (0.55)	[0.42-2.07]	0.82 (0.26)	[0.43-1.21]
APQ11 (%)	1.49 (0.71)	[0.42-2.55]	1.90 (0.78)	[0.73-3.07]	1.08 (0.25)	[0.70-1.45]
HNR (dB)	24.29 (3.72)	[18.71-29.87]	22.94 (3.23)	[18.09-27.78]	25.63 (3.76)	[19.99-31.27]
LTAS _{slope} (dB)	-26.79 (2.99)	[(-31.27)- (-22.30)]	-27.27 (3.07)	[(-31.87)- (-22.66)]	-26.3 (2.90)	[(-30.65)- (-21.95)]
LTAS _{tilt-trendline} (dB)	-10.70 (1.02)	[(-12.23)- (-9.17)]	-11.05 (0.99)	[(-12.53)- (-9.56)]	-10.35 (0.95)	[(-11.77)- (-8.92)]
CPPS _{a/} (dB)	15.19 (2.22)	[11.86-18.52]	15.54 (2.29)	[12.10-18.97]	14.84 (2.13)	[11.64-18.03]
CPPS _{frases} (dB)	8.09 (0.83)	[6.84-9.33]	7.86 (0.76)	[6.72-9.00]	8.40 (0.77)	[7.24-9.55]

En relación al parámetro de ruido aditivo en la señal (HNR) se observan valores significativamente más altos en el grupo de las mujeres (HNR: $U=191, p=0.01$).

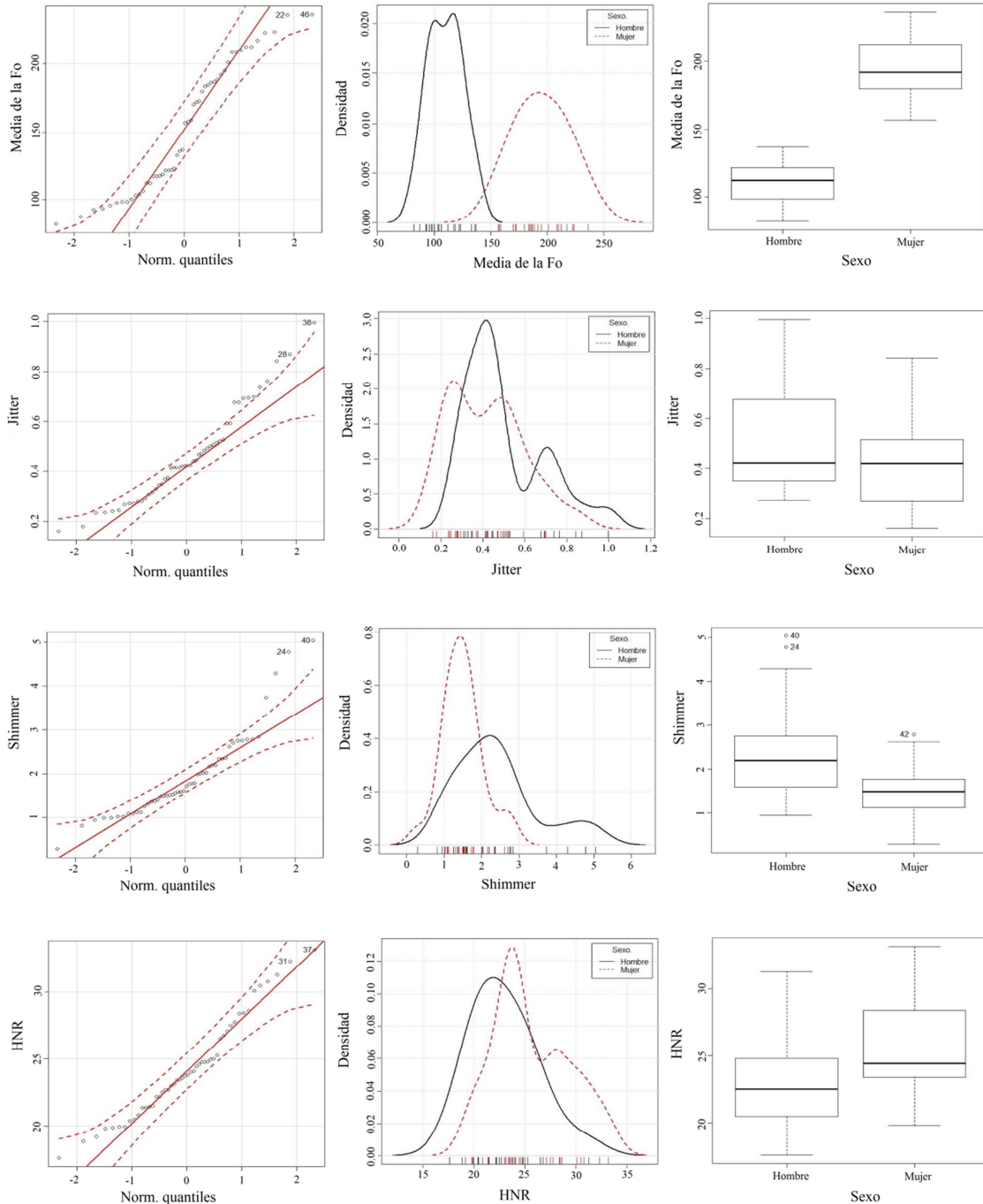


Figura 3. Análisis exploratorio de los datos: Gráficas Q-Q Normal de la muestra (columna de la izquierda) y las gráficas de densidad (columna del centro) y de cajas y bigotes (columna de la derecha) en función del sexo, de la media de la F₀ (primera fila), el jitter (segunda fila), el shimmer (tercera fila) y la HNR (última fila).

En la figura 4 se muestra el análisis exploratorio de los datos de la $LTAS_{slope}$ y la $LTAS_{tilt-trendline}$. Los resultados de los parámetros del espectro medio a largo plazo muestran que la pendiente espectral es similar en ambos sexos ($LTAS_{slope}$: $t_{(48)} = -1.12$, $p = 0.26$) mientras que la línea de tendencia de la inclinación a través del LTAS muestra mayores valores en las mujeres ($LTAS_{tilt-trendline}$: $t_{(48)} = -2.52$, $p = 0.01$)

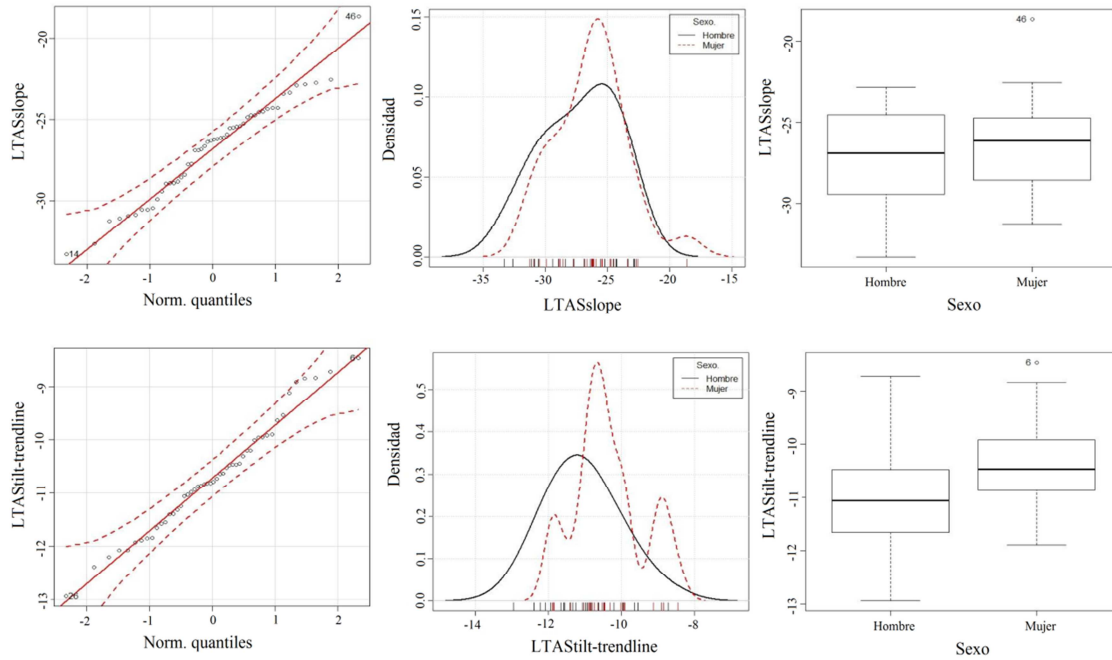


Figura 4. Análisis exploratorio de los datos: Gráficas Q-Q Normal de la muestra (columna de la izquierda) y las gráficas de densidad (columna del centro) y de cajas y bigotes (columna de la derecha) en función del sexo, de la media de la $LTAS_{slope}$ (fila de arriba) y la $LTAS_{tilt-trendline}$ (fila de abajo).

En la figura 5 se muestra el análisis exploratorio de los datos de la CPPs tanto en vocal sostenida como en habla conectada. En cuanto a la prominencia del pico cepstral suavizado, no se observan diferencias de género en la CPPs obtenida a partir de la vocal sostenida /a/ ($CPPs_{/a/}$: $t_{(48)} = 1.11$, $p = 0.26$). En relación a la CPPs obtenida en habla conectada a través de las frases, los resultados muestran valores mayores en mujeres ($CPPs_{frases}$: $t_{(48)} = -2.77$, $p = 0.007$).

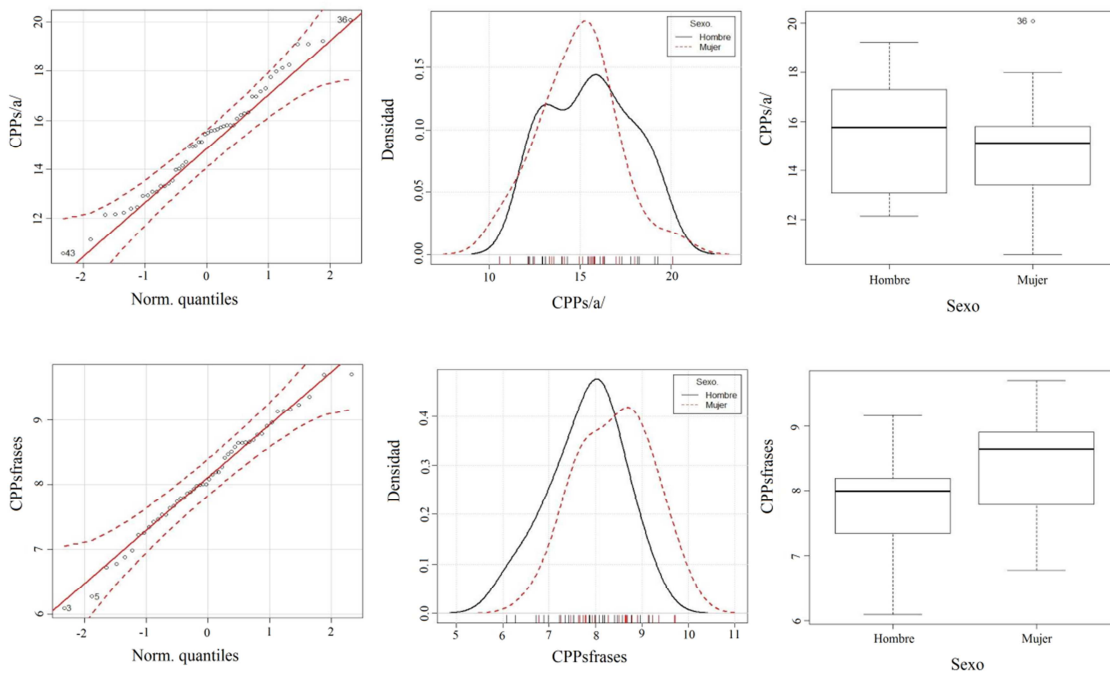


Figura 5. Análisis exploratorio de los datos: Gráficas Q-Q Normal de la muestra (columna de la izquierda) y las gráficas de densidad (columna del centro) y de cajas y bigotes (columna de la derecha) en función del sexo, de la CPPs tanto en la vocal sostenida (fila de arriba) como en las frases (fila de abajo)

Discusión

El objetivo de este trabajo descriptivo fue obtener valores de normalidad en las principales medidas temporales, espectrales y cepstrales que se pueden realizar en el *Praat* en una muestra de hablantes de español.

Los resultados muestran valores más altos de F_0 en mujeres debidas básicamente a las características fisiológicas entre hombres y mujeres. Esto se debe a que las cuerdas vocales (CCVV) en mujeres miden entre 12.5 y 17 mm y las del hombre entre 17 y 23 mm (Kaplan, 1971). Esta menor extensión provoca una mayor vibración (entre 171 y 230 ciclos por segundo según los resultados de este trabajo). En la siguiente fórmula de se puede observar como el valor de la F_0 cae a medida que aumenta la longitud de las CCVV (Núñez-Batalla, 2008):

$$F_0 = (1/2L)\sqrt{\sigma/\rho}$$

L =Longitud de las CCVV, σ = tensión longitudinal, ρ =densidad tisular

Por la acción de la musculatura de la laringe las CCVV se pueden estirar 3 o 4 mm adicionales. En la voz patológica, valores anormales de F_0 se pueden deber a un exceso de esfuerzo de la musculatura laríngea o a un aumento de la densidad de la cubierta de la CV aportando valores más bajos de F_0 de los esperados.

En relación a los parámetros de perturbación de la frecuencia se encontraron valores más altos del jitter absoluto en hombres. El jitter absoluto mide la variación ciclo a ciclo en unidades de tiempo, por lo que está influenciado por la F_0 del individuo (Cobeta y Núñez, 2013). En cambio en el jitter relativo desaparecen esas diferencias ya que se calcula la esta variación inter-ciclos en porcentaje. La voz que presenta valores anormalmente altos de jitter son percibidos como ronquera (Cobeta y Núñez, 2013).

Se obtuvieron valores de perturbación de la amplitud significativamente menores en el grupo de las mujeres al igual que los valores obtenidos en el estudio de Elisei (2012), aunque aún no está clara la relación entre estas medidas y el género (Baken y Orlikoff, 2000). El shimmer mide la perturbación de la amplitud ciclo a ciclo y está condicionada a la amplitud de la vibración de las cuerdas vocales y a la presión subglótica. Como se trata de parámetro de la intensidad se ve afectado por un soporte respiratorio inadecuado, un cierre glótico incompleto o unas cuerdas vocales poco flexibles (Cobeta y Núñez, 2013).

En relación a la ratio armónico-ruido se encontraron resultados similares a otros trabajos en nuestro idioma (Elisei, 2012) con mayor valor en mujeres. La disminución del HNR implica una mayor proporción de ruido aditivo en la voz y por tanto mayor grado de disfonía.

En la pendiente espectral no se encontraron diferencias en función del sexo. En cambio, si se observan diferencias cuando la pendiente espectral se calcula a partir de la línea de tendencia a través del LTAS observándose valores más altos en mujeres. El análisis del LTAS se correlaciona con el esfuerzo vocal (Núñez-Batalla, Corte-Santos, Señaris-González, Rodríguez-Prado y Suárez-Nieto, 2004). Un mayor esfuerzo de adducción, principalmente de las bandas ventriculares, da como resultado una voz pobre en elementos armónicos de alta frecuencia en el espectro pero rica en ruido (Dejonkere, 1983; Klatt y Klatt, 1975), y se relaciona con la voz aérea, áspera o ronca. En cambio un menor esfuerzo vocal suele provocar una pobre presión subglótica y una

mala adducción de las CCVV lo que conlleva una pobreza armónica dentro del espectro (Núñez-Batalla et al., 2004). En ambos supuestos se produce un aumento del valor (en dB) del $LTAS_{slope}$ y del $LTAS_{tilt-trendline}$ y, por tanto, una menor pendiente espectral.

La CPPs desarrollada por Hillenbrand, Cleveland y Erickson (1994) y Hillebrand y Houde (1996) presenta en nuestro estudio valores similares en ambos sexos cuando se extrae a partir de la vocal sostenida /a/. En habla conectada las mujeres presentan voces más armónicas que los hombres mostrando un mayor pico cepstral. Se ha demostrado en multitud de estudios la utilidad de este parámetro acústico para diferenciar entre voz normal y patológica con una alta correlación con las valoraciones perceptivas de la severidad de la disfonía (Delgado-Hernández et al., 2017; Sauder et al., 2017; Heman-Ackah et al., 2014; Balasubramaniam, Bhat, Fahim y Raju, 2011; Maryn, Roy, De Bodt, Van Cauwenberge y Corthals, 2009; Eadie y Baylor, 2006; Awan y Roy, 2006; Heman-Ackah et al., 2003; Heman-Ackah, Michael y Goding, 2002; Wolfe, Martin y Palmer, 2000; Dejonckere, 1998; Wolfe y Martin, 1997; Dejonckere y Wieneke, 1996; Hillebrand y Houde, 1996; Hillenbrand et al., 1994; de Krom, 1993). A medida que se incrementa la severidad de las alteraciones en la voz disminuye el valor de la CPPs, tanto en vocales sostenidas como en habla conectada (Delgado-Hernández et al., 2017; Maryn et al., 2009).

Conclusión

El análisis acústico aporta información objetiva de la voz de forma indirecta. Para su interpretación necesitamos, por un lado, de los valores de normalidad de las diferentes medidas acústicas y, por otro, de estudios de precisión diagnóstica en pacientes disfónicos de diferentes grados de severidad y etiología. En un próximo estudio se abordará la precisión diagnóstica de las medidas acústicas utilizadas en este trabajo en población española.

Para utilizar los valores de normalidad obtenidos en este estudio hay que tener en cuenta que se debe usar el programa *Praat*, un tipo de micrófono con similares características al de este trabajo, una SNR mayor de 30 dB y la misma configuración en el cálculo de la $LTAS_{slope}$, $LTAS_{tilt-trendline}$ y la CPPs.

Bibliografía

- Awan S. N., y Roy N. (2006). Toward the development of an objective index of dysphonia severity: a four-factor acoustic model. *Clinical Linguistics and Phonetics*, 20, 35–49.
- Baken, R. J., y Orlikoff, R. F. (2000). *Clinical Measurement of Speech and Voice*, 2ª ed. San Diego: Singular Publishing Group.
- Balasubramaniam, R. K., Bhat, J. S., Fahim, S. 3rd, y Raju, R. 3rd. (2011). Cepstral analysis of voice in unilateral adductor vocal fold palsy. *Journal of Voice*, 25, 326–329.
- Barsties, B., y De Bodt, M. (2015). Assessment of voice quality: current state-of-the-art. *Auris Nasus Larynx*, 42, 183-188.
- Barsties, B., y Maryn, Y. (2016). External validation of the Acoustic Voice Quality Index version 03.01 with extended representativity. *Annals of Otolaryngology, Rhinology & Laryngology*, 125, 571-583.
- Boersma P., y Weenink D. (2015). Praat: doing phonetics by computer [Computer program]. Version 5.4.22. <http://www.praat.org/>. Consultado el 6 de Febrero de 2017.
- Brinca, L. F., Batista, A. P. F., Tavares, A. I., Gonçalves, I. C., y Moreno, M. L. (2014). Use of cepstral analyses for differentiating normal from dysphonic voices: A comparative study of connected speech versus sustained vowel in European Portuguese female speakers. *Journal of Voice*, 28, 282-286.
- Cobeta, I., y Núñez, F. (2013). Análisis de la señal acústica. En I. Cobeta, F. Núñez y S. Fernández (Eds.). *Patología de la voz* (pp. 188-198). Barcelona: Marge Médica Books.
- de Krom, G. (1993). A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *Journal of Speech and Hearing Research*, 36, 254–266.
- Dejonckere, P. H. (1998). Cepstral voice analysis: link with perception and stroboscopy. *Revue Laryngologie- Otologie- Rhinologie (Bord)*, 119, 245–246.
- Dejonckere, P., y Wieneke, G. (1996). Cepstra of normal and pathological voices: correlation with acoustic, aerodynamic and perceptual data. En M. J. Ball y M. Duckworth (Eds.). *Advances in Clinical Phonetics Vol. 6.* (pp. 217-226). Amsterdam: John Benjamins.

- Dejonkere, P. H. (1983). Recognition of hoarseness by means of LTAS. *International Journal of Rehabilitation Research*, 6, 343-345.
- Delgado-Hernández, J., León-Gómez, N., Izquierdo-Arteaga, L., y Llanos-Fumero, Y. (2017). Análisis cepstral de la voz normal y patológica en adultos españoles. Medida de la Prominencia del Pico Cepstral Suavizado en vocales sostenidas versus habla conectada. *Acta Otorrinolaringológica Española*. Doi: 10.1016/j.otorri.2017.05.006.
- Deliyski, D. D., Evans, M. K., y Shaw, H. S. (2005). Influence of data acquisition environment on accuracy of acoustic voice quality measurements. *Journal of Voice*, 19, 176–186.
- Deliyski, D. D., Shaw, H. S., Evans, M. K., y Vesselinow, R. (2006). Regression tree approach to studying factors influencing acoustic voice analysis. *Folia Phoniatrica et Logopaedica*, 58, 274-288.
- Deliyski, D. D., Shaw, H. S., y Evans, M. K. (2005). Adverse effects of environmental noise on acoustic voice quality measurements. *Journal of Voice*, 19, 15-28.
- Eadie, T. L., y Baylor, C. R. (2006). The effect of perceptual training on inexperienced listeners judgments of dysphonic voice. *Journal of Voice*, 20, 527–544.
- Elisei, N. G. (2012). Análisis acústico de la voz normal y patológica utilizando dos sistemas diferentes: ANAGRAF y PRAAT. *Interdisciplinaria*, 29, 271-286.
- Ferrer, C. A., De Bodt, M., Maryn, Y., Van de Heyning, P., y Hernández-Díaz, M. E. (2007). Properties of the cepstral peak prominence and its usefulness in vocal quality measurements. En C. Manfredi (Ed.). *Models and analysis of vocal emissions for biomedical applications: 5th international workshop* (pp. 93-96). Florencia: Firenze University Press.
- Heman-Ackah, Y. D., Michael, D. D., Baroody, M. M., Ostrowski, R., Hillenbrand, J., Heuer, R. J., ... y Staloff, R. T. (2003). Cepstral peak prominence: a more reliable measure of dysphonia. *Annals of Otolaryngology, Rhinology and Laryngology*, 112, 324–333.
- Heman-Ackah, Y. D., Michael, D. D., y Goding G. S. Jr. (2002). The relationship between cepstral peak prominence and selected parameters of dysphonia. *Journal of Voice*, 16, 20–27.

- Heman-Ackah, Y. D., Sataloff, R. T., Laureyns, G., Lurie, D., Michael, D. D., Heuer, R., ... y Lyons, K. (2014). Quantifying the cepstral peak prominence, a measure of dysphonia. *Journal of Voice*, 28, 783-788.
- Hillenbrand, J., Cleveland, R. A., y Erickson, R. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research*, 37, 769-778.
- Hillenbrand, J., y Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *Journal of Speech and Hearing Research*, 39, 311-321.
- Hochmuth, S., Brand, T., Zokoll, M. A., Castro, F. Z., Wardenga, N., y Kollmeier, B. (2012). A Spanish matrix sentence test for assessing speech reception thresholds in noise. *International Journal of Audiology*, 51, 536-544.
- Ingrisano, D. R., Perry, C. K., y Jepson, K. R. (1998). Environmental noise: a threat to automatic voice analysis. *American Journal of Speech-Language Pathology*, 7, 91-96.
- Kaplan, H. M. (1971). *Anatomy and physiology of speech* (2nd ed.). Nueva York: Mcgraw-hill book company.
- Klatt, D. H., y Klatt, L. C. (1975). Analysis, synthesis, and perception on voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, 87, 820-857
- Lowell, S. Y., Colton, R. H., Kelley, R. T., y Hahn, Y. C. (2011). Spectral and cepstralbased measures during continuous speech: capacity to distinguish dysphonia and consistency within a speaker. *Journal of Voice*, 25, 223-232.
- Maryn, Y., Corthals, P., De Bodt, M., Van Cauwenberge, P., y Deliyski, D. (2009). Perturbation measures of voice: a comparative study between Multi- Dimensional Voice Program and Praat. *Folia Phoniatica et Logopaedica*, 61, 217-226.
- Maryn, Y., Roy, N., De Bodt, M., Van Cauwenberge, P., y Corthals, P. (2009). Acoustic measurement of overall voice quality: a meta-analysis. *Journal of the Acoustical Society of America*, 126, 2619-2634.
- Maryn, Y., y Weenink, D. (2015). Objective dysphonia measures in the program Praat: smoothed cepstral peak prominence and acoustic voice quality index. *Journal of Voice*, 29, 35-43.

- Moers, C., Möbius, B., Rosanowski, F., Nöth, E., Eysholdt, U., y Haderlein, T. (2012). Vowel-and text-based cepstral analysis of chronic hoarseness. *Journal of Voice*, 26, 416-424.
- Núñez-Batalla, F. (2008). Función fonatoria. *Tratado de otorrinolaringología y cirugía de cabeza y cuello, 2ª ed.* (pp. 2503-2518). En C. Suárez, C., L. M. Gil-Carcedo, J. Marco, J. E. Medina, P. Ortega y J. Trinidad (Eds.). Madrid: Ed. Médica Panamericana.
- Núñez-Batalla, F., Corte-Santos, P., Señaris-González, B., Rodríguez-Prado, N., y Suárez-Nieto, C. (2004). Evaluación espectral cuantitativa de la hipofunción vocal. *Acta Otorrinolaringológica Española*, 55, 327-333.
- Núñez-Batalla, F., González-Márquez, R., Peláez-González, M. B., González-Laborda, I., Fernández-Fernández, M., y Morato-Galán, M. (2014). Acoustic voice analysis using the Praat program: comparative study with the Dr. Speech program. *Acta Otorrinolaringológica española*, 65, 170-176.
- Oğuz, H., Kiliç, M. A., y Şafak, M. A. (2011). Comparison of results in two acoustic analysis programs: Praat and MDVP. *Turkish Journal of Medical Sciences*, 41, 835-841.
- Sauder, C., Bretl, M. y Eadie, T. (2017). Predicting Voice disorder status from smoothed measures of cepstral peak prominence using Praat and Analysis of Dysphonia in Speech and Voice (ADSV). *Journal of Voice*, 31, 557-566.
- Titze, I. R., y Winholtz, W. S. (1993). Effects of microphone type and placement on voice perturbation measurements. *Journal of Speech and Hearing Research*, 36, 1177-1190.
- Wolfe, V. I., Martin, D. P., y Palmer, C. I. (2000). Perception of dysphonic voice quality by naive listeners. *Journal of Speech, Language and Hearing Research*, 43, 697-705.
- Wolfe, V., y Martin, D. (1997). Acoustic correlates of dysphonia: type and severity. *Journal of Communication Disorders*, 30, 403-415.