

**Elisei, Natalia Gabriela ; Evin, Diego Alexis**

*Evaluación perceptivo auditiva de voces degradadas y su correlación con medidas acústicas*

Revista de Psicología Vol. 8 N° 16, 2012

Este documento está disponible en la Biblioteca Digital de la Universidad Católica Argentina, repositorio institucional desarrollado por la Biblioteca Central "San Benito Abad". Su objetivo es difundir y preservar la producción intelectual de la Institución.

La Biblioteca posee la autorización del autor para su divulgación en línea.

Cómo citar el documento:

Elisei, N. G., Evin, D. A. (2012). Evaluación perceptivo auditiva de voces degradadas y su correlación con medidas acústicas [en línea], *Revista de Psicología*, 8(16). Disponible en:  
<http://bibliotecadigital.uca.edu.ar/repositorio/revistas/evaluacion-perceptivo-auditiva-voces-degradadas.pdf> [Fecha de consulta:.....]

- Stein, A. & Rosemberg, C.R. (en prensa). Redes de colaboración en situaciones de alfabetización familiar con niños pequeños. Un estudio en poblaciones urbano marginadas de Argentina. *Interdisciplinaria*.
- Strauss, A. & Corbin, J. (1991). *Basics of qualitative research. Grounded theory. Procedures and techniques*. Londres: Sage.
- Taylor, D. (1983). *Family literacy: Young children learning to read and write*. Londres, Inglaterra: Heineman.
- Taylor, D. & Strickland, D.S. (1986). *Family Storybook Reading*. Portsmouth, New Hampshire, E.E.U.U.: Heinemann.
- Teale, W. (1986). Home background and young children's literacy development. En W. Teale & E. Sulzby (Eds.), *Emergent literacy: Reading and writing* (pp. 172-206). Norwood, Nueva Jersey, E.E.U.U.: Ablex.
- Vernon-Feagans, L., Hammer, C., Miccio, A. & Manlove, E. (2003). Early language and literacy skills in low-income African American and Hispanic children. En S. Neuman & D. Dickinson (Eds.) *Handbook of early literacy research* (pp.192-210). Nueva York, E.E.U.U.: Guilford.
- Volk, D. & de Acosta, M. (2004). Mediating networks for literacy learning: The role of Puerto Rican siblings. En E. Gregory, S. Long & D. Volk (Eds.), *Many pathways to literacy* (pp. 25-39). Nueva York, E.E.U.U.: Routledge Falmer.
- Wasik, B. (2004) *Handbook of family literacy*. Nueva York, E.E.U.U.: Guilford.
- Weizman, Z.O. & Snow, C. (2001). Lexical input as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology*, 17, 265-279.

## Evaluación perceptivo auditiva de voces degradadas y su correlación con medidas acústicas

*Auditory - perceptual assessment of degraded voices and its correlation with acoustic measures*

Natalia Gabriela Elisei\*  
Diego Alexis Evin\*\*

### Resumen

El objetivo es determinar utilizando las mediciones acústicas, qué información es más relevante para el oyente al momento de categorizar el grado general de disfonía. Se eligieron 8 (4 voces femeninas y 4 voces masculinas). Cada emisión fue evaluada auditivo perceptualmente a través del ítem G de la escala GRBAS por 10 oyentes experimentados y acústicamente mediante medidas de aperiodicidad, ruido y caos. El estudio estadístico de análisis discriminante señala la importancia de GNE, Jit y Jitter\_cc y Lyapunov como parámetros predictores del grado general de disfonía. La aplicación del método k-means evidencia que existen rasgos en los parámetros acústicos empleados que permiten agrupar objetivamente las voces estudiadas con 100% de precisión para

la clase 0, 96% a la clase 2 y 79% a la clase 3. Un mayor número y variabilidad de casos se necesita a fin de verificar los resultados preliminares.

**Palabras clave:** Desórdenes vocales, Análisis acústico de voz, GRBAS, Análisis perceptual de la voz

### Abstract

The goal is to determine using acoustic measurements, which information is the most relevant to listeners at the time of categorizing the overall degree of dysphonia. Eight voice signals were chosen (4 female voices and 4 male voices). Each

\* Doctora en Ciencias de la Salud y Fonoaudióloga. Centro Interdisciplinario de Investigaciones en Psicología Matemática y Experimental Dr. Horacio J. A. Rimoldi, Consejo Nacional de Investigaciones Científicas y Técnicas (CIIPME-CONICET). Correo Electrónico: natalia.elisei@gmail.com

\*\* Doctor en Ciencias de la Computación. Laboratorio de Investigaciones Sensoriales. Instituto de Neurociencias, Hospital de Clínicas. Laboratorio LIS INIGEM. Correo Electrónico: diegoevin@gmail.com

Correo electrónico: natalia.elisei@gmail.com

Fecha recepción: 29 de febrero de 2012 – Fecha aceptación: 30 de octubre de 2012

voice was perceptually evaluated through the item G of GRBAS scale by 10 experienced listeners and acoustically by aperiodicity, noise and chaos measures. The statistical study by discriminant analysis shows the importance of GNE, Jit and Lyapunov Jitter\_cc as parameters and predictors of overall degree of dysphonia. The application of the k-means evidence there are features in the acoustic parameters that allow us to objectively group the voices studied with 100% accuracy for class 0,96% for class 2 and 79% for class 3. A greater number and variability of cases are need to verify those preliminary results.

**Key words:** Voice disorders, Acoustic voice analysis, GRBAS, Perceptual voice analysis.

Cuando se intenta detectar y caracterizar las voces patológicas en la clínica, el objetivo es documentar los cambios significativos de la voz, es decir, aquellos que no resulten despreciables ni producto del azar. Estos cambios pueden documentarse a través de la evaluación auditivo perceptual y el análisis acústico de la señal.

La popularidad de las escalas perceptivo-auditivas surge de múltiples factores, pero, sin lugar a dudas, uno de los puntos importantes es que la voz es en su naturaleza perceptual (Kreiman, Gerratt, & Berke, 1994) y que existen limitaciones en la medición instrumental de la voz. Desde este enfoque, la voz es fundamentalmente un fenómeno de percepción en respuesta a un estímulo acústico. Es por esto que la evaluación perceptivo auditiva se convierte en

una evaluación “gold standard” de este fenómeno.

Se postula que compartimos una realidad perceptiva común (Wuyts, De Bodt, & Van de Heyning, 1999). Consecuentemente, si esto fuese así, las descripciones de la percepción de la voz serían intuitivas para cualquiera de nosotros (Kreiman, Gerratt, & Kempster, 1993). Algunos de los atractivos de este enfoque subyacen en la idea de que quizás sea una manera de evaluar las características de la voz que de otro modo sería difícil medir (Kreiman et al., 1993).

Sin embargo, las voces patológicas han sido clasificadas y descritas con diferentes términos y/o escalas que denotan su impresión perceptual: ronquera, aspereza, soplo, etc. Diferentes escalas han sido introducidas para evaluar las voces patológicas y algunos especialistas han comenzado a usarlas.

La percepción puede evaluar la voz desde una perspectiva más estrecha a otra más amplia. Si se define la voz en términos acotados, la evaluación perceptual también intentará reflejarlo, y supondrá escalas de evaluación perceptual centradas en la fuente glótica, es decir, sólo en el sonido producido por las cuerdas vocales. Sin embargo, otras incluirán la influencia del tracto vocal que se extiende desde las cuerdas vocales a los labios. Perspectivas más globales y cuestionamientos más sutiles dirán que es necesario definir un modelo de percepción de la voz humana patológi-

ca o degradada y que las aproximaciones conocidas no son realmente válidas (Kreiman et al., 1993).

Entre los diversos sistemas de descripción de voces patológicas, las más comunes se concentran en el grado de “ronquera” sinónimo de disfonía (Hirano, 1981) (Nawka, Anders, Cebulla, & Zurakowski, 1997). La percepción de las anomalías en la voz podría ser calificada si se ofreciera una descripción detallada y se la independizara de la lengua o idioma. Lamentablemente, la selección y la definición de estas escalas han sido controvertidas y no siempre comunes a todos los especialistas en voz. La estandarización de estas escalas es una necesidad sentida en el área. Un requisito parecería ser la definición cuantitativa de las escalas basada en los correlatos acústicos y fisiológicos.

Varios estudios han correlacionado características acústicas y calidades perceptuales en voces patológicas (Dejonckere, Obbens, de Moor & Wienke, 1993; de Krom, 1995; Eskenazi, Childers & Hicks, 1990; Hammarberg, Fritzell, Gauffin & Sundberg, 1986; Hammarberg, Fritzell, Gauffin, Sundberg, & Wedin, 1980; Kreiman, Gerratt & Berke, 1994; Kreiman, Gerratt, Precoda & Berke, 1992; Murry, Singh & Sargent, 1977; Rammage, Pppard & Bless, 1992); o bajo condiciones fisiológicas de la glotis (Hirano, Hibi, Terasawa & Fujiu, 1986; Rammage et al., 1992). Algunas de las correlaciones asocian la aspereza con la impresión de

pulsos glóticos irregulares de una componente de ruido de baja frecuencia. Las medidas relacionadas con la perturbación de la frecuencia (jitter) serían más importantes en la evaluación de la aspereza que las relativas a la amplitud, aunque principalmente la relación armónica a ruido (HNR, harmonic to noise ratio) parece ser el principal indicador de esta característica. La voz soplada se relaciona con el ruido originado por las turbulencias creadas por una funcionalidad insuficiente de la glotis. La modalidad de la vibración de las cuerdas vocales sería ineficiente y se acompañaría con una fricción audible. Esta fricción podría deberse a la turbulencia en el flujo de aire. Los posibles indicadores de esta característica serían las medidas de perturbación de la frecuencia y los parámetros de energía espectral (como GNE, Glottal to Noise Energy Ratio). La impresión auditiva de astenia se asociaría a debilidad en la fonación espontánea, voz hipocinética o hipofuncional y por lo contrario, la impresión auditiva de excesivo esfuerzo con tensión en la fonación espontánea. Una combinación de la voz áspera y soplada generaría la voz ronca (*hoarseness*), con un flujo de aire no modulado debido a un cierre glótico ausente o incompleto, o irregularidades en la fonación concernientes a la frecuencia, amplitud o fase. Los principales predictores serían los mismos que para la aspereza y voz soplada, es decir los parámetros de perturbación de la amplitud (*shimmer*) y frecuencia (*jitter*).

Resumiendo, los resultados son usualmente ambiguos, y a veces hasta contradictorios, las co-variaciones de las medidas acústicas así como las interacciones son temas aún irresueltos.

En este trabajo se estudia el grado de acuerdo entre los oyentes para puntuar perceptualmente una muestra de voces seleccionadas y se calculan las correlaciones para un grupo de medidas acústicas de aperiodicidad (6 medidas diferentes de perturbación de la frecuencia o jitter, a saber: jitter\_cc (ANAGRAF) y jit, jitta, RAP, PPQ, DDP (PRAAT); 3 medidas de ruido (harmonics-to-noise\_ratio (ANAGRAF), HNR (PRAAT) y GNE y una medida de caos implementada *ad-hoc* para este estudio (LE). El objetivo es determinar utilizando las mediciones acústicas, qué información es más relevante para el oyente al momento de categorizar la calidad vocal.

### Materiales y método

Fueron evaluadas 194 muestras de una base de datos (Elisei, 2011). Se seleccionaron 40 muestras pasibles de representar las 5 cualidades de la GRBAS en sus 4 niveles 0 al 3. Se generaron los espectrogramas de banda estrecha para la totalidad de las muestras seleccionadas y se tipificaron según la clasificación de señales de Titze (Titze, 1995). Finalmente se eligieron 8 (4 voces femeninas y 4 voces masculinas). Esta reducción fue una decisión comprometida y sopesada entre

la necesidad de puntuación de estímulos para realizar nuestro experimento y las restricciones de fatiga y habituación que impone el sistema auditivo, sumado a la condición de iteración de cada muestra para constatar la consistencia de las respuestas. En el Anexo se detallan cada una de las señales utilizadas.

### Métodos

Un panel de 10 oyentes entrenados, residentes en el área de fonoaudiología del Hospital de Clínicas José de San Martín, puntuó el corpus. Se solicitó que utilizaran la escala perceptual que tradicionalmente usan en la clínica diaria, la GRBAS (Hirano, 1981). Se analizó sólo el ítem G: grado general de la ronquera o disfonía. Posteriormente se correlacionó un grupo de medidas acústicas con el puntaje perceptual.

Para analizar acústicamente las señales se empleó la funcionalidad del Voice Report de PRAAT "Doing phonetics by computer", de acceso público y gratuito, disponible en Internet (Boersma & Weenink, 1996) y de ANAGRAF versión V09.10 (Gurlekian, 2001) con la función de reporte desarrollado ad-hoc en el Laboratorio de Investigaciones Sensoriales. Los criterios para la selección de las medidas acústicas fueron variados. Prevaleció la tradición en medidas como jitter\_cc, harmonics-to-noise\_ratio (ANAGRAF) jit, jitta, RAP, PPQ, DDP y HNR (PRAAT); la docu-

mentada robustez en la selección del GNE (Michaelis, Froehlich, & Werner Strube, 1998) y la curiosidad sobre la influencia de LE (Giovanni, Ouaknine, & Triglia, 1999).

### Resultados y discusión

En relación al panel de jueces lo que se pudo observar es que ofrecen una respuesta psicofísica más acertada a medida que aumenta la magnitud de la degradación espectral de la señal. Esto es, a medida que el componente de ruido aumenta, aumenta el acuerdo aparente entre los jueces. Ver Figura 1.

Es importante resaltar que se esperaba mayor acuerdo o consenso en los extremos (*anchors*) del continuo y variaciones en los intermedios pero esto no ocurrió en la puntuación 0 caracterización de voces normales (G0). Una posible respuesta es que el panel estuvo compuesto por especialistas que evalúan y buscan detectar la patología vocal. La tendencia del panel fue estimar los grados de disfonía o ronquera partiendo siempre de una leve base patológica y desestimando la presencia de voces normales en la secuencia aleatoria a pesar de ser advertidos en la consigna.

Se analizó estadísticamente la contribución de varios parámetros medidos acústicamente con los perceptualmente puntuados por los oyentes, y lo que muestran los coeficientes de correlación de Spearman (Spearman's rho) es que

todos los rangos de los datos varían conjunta y significativamente ( $p < 0,01$ ) con las respuestas dadas.

Con el propósito de identificar las características (variables) que diferencian a dos o más grupos, y determinar cuántas de ellas son necesarias para lograr la mejor clasificación posible, se realizó un estudio empleando análisis discriminante. A través de esta técnica es posible buscar una o más funciones, conformadas por combinaciones lineales de las variables, que permitan separar las instancias suministradas según las identidades de clases correspondientes. Una vez obtenido el conjunto de funciones discriminantes es posible emplearlas para predecir la identidad de clase de nuevos ejemplos.

En primer lugar conviene resaltar que a partir de los análisis llevados a cabo en la etapa de estudio de correlaciones entre variables, se había observado que si bien las clases de los datos del problema en cuestión se solapaban en distintos grados, tal solapamiento no era completo, lo que sugiere que existen diferencias entre las clases, y genera expectativas respecto de los resultados de la aplicación del análisis discriminante. En primer lugar se realizó un análisis de ANOVA univariado para determinar si existía alguna variable no significativa, el resultado de dicha prueba se presenta en la Tabla 1.

Esta prueba brinda estadísticos F que permiten contrastar la hipótesis de igualdad de medias entre los grupos en

cada variable independiente. Se puede ver que si se adopta el umbral de 0,05 como límite para aceptar dicha hipótesis, en ningún caso se debería aceptar la hipótesis de medias iguales.

La Tabla 2 permite comparar de manera global la capacidad discriminante de cada función. Por ejemplo en este caso la función 1 explica el 77,7% de las diferencias existentes entre los sujetos de los grupos, mientras que la segunda función logra explicar un 21,9% de dichas diferencias y la tercera un 0,4%.

En la Tabla 3 se brinda el detalle de los coeficientes estandarizados de la función canónica discriminante. En esa tabla se puede comparar la importancia de cada variable a la hora de predecir el grupo de pertenencia perceptual. Se puede notar que la variable *GNE* es la más importante para todas las funciones discriminantes, salvo para la primera, mientras que *Jit* es importante en la primera, y segunda en importancia en la segunda y tercera funciones discriminantes.

En la Tabla se presentan los coeficientes de clasificación de Fisher que permiten obtener la función de clasificación para cada grupo. Para aplicar estos coeficientes se calcula cada una de las funciones para un sujeto, y se lo clasifica en el grupo para el cual la función obtiene el mayor valor. Se puede observar que tanto en la Tabla 3 como en la Tabla 4 no se muestra las variables *Jitta*, *RAP*, *PPQ*, *DDP* y *HNR*. Estas variables no pasaron la prueba de tolerancia. Probablemente

se deba a que la varianza de las mismas fueron completamente explicadas por las demás variables, es decir que las mismas fueron variables redundantes. En el proceso de cálculo se dispuso que el algoritmo estime las probabilidades iniciales de las clases a partir de los tamaños de cada clase de acuerdo a los datos presentados. En la Tabla se presenta los resultados de la clasificación obtenidos mediante esta técnica utilizando validación cruzada (opción 1 de N). El clasificador arrojó tasas de reconocimiento de grado de la disfonía (*G* de *GRBAS*) relativamente buenas para todos los puntajes excepto para el 1 en cuyo caso la mitad de los casos se confunden con la clase 0 y la otra mitad con la clase 2.

A partir de los resultados anteriores se evaluó la posibilidad de que los datos disponibles no estuviesen separados de manera compatible con las puntuaciones perceptuales, sino de acuerdo a otro ordenamiento. Para ello, se experimentó con un algoritmo de clustering o agrupamiento, específicamente *K-means*. A diferencia de la tarea de clasificación llevada a cabo anteriormente a través de la metodología de análisis discriminante, en la tarea de clustering, se buscan funciones que separen el conjunto de datos disponibles a partir de similitudes intrínsecas o naturales de los datos, y no a partir de etiquetas de clases conocidas a priori. Es decir, con el análisis discriminante se intentó determinar funciones que permitiesen obtener una puntuación perceptual similar a la asignada por los

profesionales empleando como observaciones las medidas acústicas, mientras que en el caso de clustering, se emplea *k-means* para tratar de detectar si existen grupos de pacientes que se puedan identificar según las similitudes de sus mediciones acústicas, sin considerar las clases perceptuales. El algoritmo *k-means* parte definiendo el número *k* de grupos en que se dividirá el conjunto de datos disponibles. El algoritmo posteriormente genera hipotéticos centros para los *k* grupos, conocidos como centros iniciales de los clústeres. Posteriormente compara la distancia de cada ejemplo disponible respecto a esos centros y los asocia al clúster con el centro más cercano. Cuando todos los ejemplos de entrada fueron asociados a un clúster se calcula el nuevo centro de cada clúster y se itera este proceso hasta que se alcanza algún criterio de finalización. Este algoritmo finalmente ofrece el valor de los centros finales, que se pueden considerar como los elementos prototípicos de cada grupo, y quedan definidos en coordenadas respecto a las variables de entrada. Finalmente se pueden comparar las características de los distintos centros en términos de sus coordenadas para proponer cuáles son las particularidades de cada grupo de datos. Como este algoritmo necesita que el usuario indique un valor de *k*, se decidió buscar igual número de grupos que de valores en la escala perceptual *GRBAS*, es decir 4. Los centros de clúster iniciales estimados por el algoritmo aparecen

en la Tabla 6. Estos son vectores cuyos valores se basan en las diez variables: *Jitter\_cc* (*ANAGRAF*), *harmonics-to-noise ratio* (*ANAGRAF*), *GNE*, *Lyapunov* (*ad-hoc*), *Jit*, *Jitta*, *RAP*, *PPQ*, *DDP* y *HNR* (*PRAAT*) estandarizadas.

En la Tabla 7 se muestran los centros finales de los conglomerados. Se puede ver que el conglomerado 1 contiene los menores valores para las medidas de perturbaciones en frecuencia y las mayores relaciones señal a ruido: *harmonics-to-noise ratio* (*ANAGRAF*). Se los podría asociar con las voces normales. El clúster 2 es parecido al 1 pero con mayores niveles de perturbación y menor *HNR*. El clúster 3 tiene los mayores niveles de perturbación (*Jit*, *Jitta*, *RAP*, *PPQ*, *DDP*) y menores *HNR* (*PRAAT*). Se los podría asociar con voces más alteradas. El grupo 4 tiene el mayor valor de *Jitter\_cc* y el más alto valor de *Lyapunov*. Ver Figura 2.

Los centros finales obtenidos a través del procedimiento anterior se pueden emplear para clasificar nuevos casos. Para ello se debe calcular la distancia de la nueva instancia respecto a cada centro y asignarle la identidad de clase correspondiente al centro más próximo. A partir de los resultados obtenidos tras la aplicación de este método, es posible aseverar que existen rasgos en los parámetros acústicos empleados que permiten agrupar objetivamente las voces estudiadas, y que este mismo procedimiento se puede aplicar para determinar la pertenencia de nuevas voces a los grupos definidos.

### Conclusiones

La voz no es un objeto tangible ni un órgano físico. La voz es un producto efímero de un sistema muy complejo de conductas que está sujeto a variaciones de momento a momento. Las mediciones impuestas al fenómeno de la producción vocal, por lo tanto, no son sólo relaciones entre entidades físicas sino aspectos del comportamiento. Demostrar una relación entre la observación de fenómenos acústicos y conductas vocales específicas generalmente es un proceso problemático. Individualmente, tanto la evaluación perceptual como la acústica han mostrado tener sus limitaciones. La tendencia es investigar si las dos pruebas se pueden correlacionar, con el propósito de desarrollar una medida que incorpore las ventajas de ambos. El análisis acústico representa sólo una parte de la evaluación de la calidad vocal, y el análisis de la percepción puede complementar y fundamentar la medida.

Los resultados mostraron una significativa correlación entre los parámetros de ruido y de percepción del grado G. Estos análisis proporcionaron información acerca de la importancia de tres parámetros como predictores: GNE que mide la proporción de energía de fuente glótica y/o por turbulencia, Jit y Jitter<sub>cc</sub> mide la aperiodicidad y Lyapunov el caos de la señal.

Los análisis efectuados en este trabajo constituyen un primer acercamiento

hacia la determinación de correlaciones entre parámetros físicos y perceptuales y sugieren que la relación entre medidas de percepción y acústicas de la voz sólo pueden ser identificadas integrando enfoques perceptuales y acústicos a la cualidad o calidad vocal. Sin embargo, resulta necesario evaluar estos resultados con un mayor número y variabilidad de casos, para poder generalizar estas relaciones.

Como trabajo futuro, se analizará la presencia de ruido aditivo a la señal de voz, dado que aparentemente esta es una de las características que mayormente relevaron los oyentes durante las evaluaciones perceptuales de este trabajo.

Es por esto que se refuerza la hipótesis de una percepción global no desagregada sino integradora que podría investigarse a través de las técnicas psicofísicas de producción de la magnitud más que de estimación.

Esto es, modelando la calidad como una función de una señal particular y un oyente específico. Por eso, el análisis por síntesis puede proveer un significado alternativo al modelar la cualidad vocal como función de ambos, voz y oyente. Se lograría así independizar al oyente de su patrón interno sugiriéndole producir por equiparación, manipulando un sintetizador, la percepción de un estímulo *target* a uno nuevo sintetizado.

### Referencias

- Boersma, P., & Weenink, D. (1996). *Praat: doing phonetics by computer*, 4.6.06. Obtenido de Computer program: <http://www.praat.org>
- de Krom, G. (1995). Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments. *J Speech Hear Res*, 38, 794-811.
- Dejonckere, P. H., Obbens, C., de Moor, G. M., & Wienke, G. H. (1993). Perceptual evaluation of dysphonia: reliability and relevance. *Folia Phoniatrica*(45), 76-83.
- Elisei, N. G. (2011). *Evaluación acústica y perceptual de la voz para la detección y caracterización de los desórdenes vocales*. Tesis doctoral, Universidad de Buenos Aires, Buenos Aires.
- Eskenazi, L., Childers, D., & Hicks, D. (1990). Acoustic Correlates of Vocal Quality. *J. Speech Hear. Res.*, 33, 298-306.
- Giovanni, A., Ouaknine, M., & Triglia, J. M. (1999). Determination of largest Lyapunov exponents of vocal signal: application to unilateral laryngeal paralysis. *J. Voice*(13), 341-354.
- Gurlekian, J. A. (2001). La percepción auditiva. En ICCE, Bustos, & Sanchez (Edits.), *La percepción auditiva: un enfoque transversal* (págs. 51-90). Madrid: CEPE.
- Hammarberg, B., Fritzell, B., Gauffin, J., & Sundberg, J. (1986). Acoustic and perceptual analysis of vocal dysfunction. *J Phonetics*(14), 533-547.
- Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J., & Wedin, L. (1980). Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngol*(90), 441-451.
- Hirano, M. (1981). *Clinical Examination of the voice*. New York: Springer Verlag.
- Hirano, M., Hibi, S., Terasawa, R., & Fujiu, M. (1986). Relationship between aerodynamic, vibratory, acoustic and psychoacoustic correlates in dysphonia. *Journal of Phonetics*, 14, 445-456.
- Kreiman, J., Gerratt, B. R., & Kempster, G. B. (1993). Perceptual Evaluation of voice quality: review, tutorial, and a framework for future research. *J Speech Hear Res*(36), 21-40.
- Kreiman, J., Gerratt, B. R., Precoda, K., & Berke, G. S. (1992). Individual differences in voice quality perception. *J Speech Hear Res*(35), 512-520.
- Kreiman, J., Gerratt, B. R., & Berke, G. S. (1994). The multidimensional nature of pathologic vocal quality. *Journal of Acoustical Society of America*, 96(3), 1291-1302.
- Michaelis, D., Froehlich, M., & Werner Strube, H. (1998). Selection and combination of acoustic features for the description of pathologic voices. *J. Acoust. Soc. Am.*, 3(103), 1628-1639.
- Murry, T., Singh, S., & Sargent, M. (1977). Multidimensional classification of abnormal voice qualities. *J. Acoust. Soc. Am.*, 61, 1630-1635.
- Nawka, T., Anders, L. C., Cebulla, M., & Zurakowski, D. (1997). The speaker's formant in male voices. *J Voice*(11), 422-428.
- Rammage, L. A., Peppard, R. C., & Bless, D. M. (1992). Aerodynamic, Laryngoscopic, and Perceptual-Acoustic Characteristics in Dysphonic Females with Posterior Glottal Chinks: A Retrospective Study. *Journal of Voice*, 6, 64-78.

Titze, I. R. (1995). *Workshop on acoustic voice analysis; Summary statement*. National Center for Voice and Speech, Denver, Colorado, Estados Unidos.  
Wuyts, F. L., De Bodt, M. S., & Van de

Heyning, P. H. (1999). Is the reliability of a visual analog scale higher than an ordinal scale? An experiment with the GRBAS scale for the perceptual evaluation of dysphonia. *J Voice*, 13, 508-517.

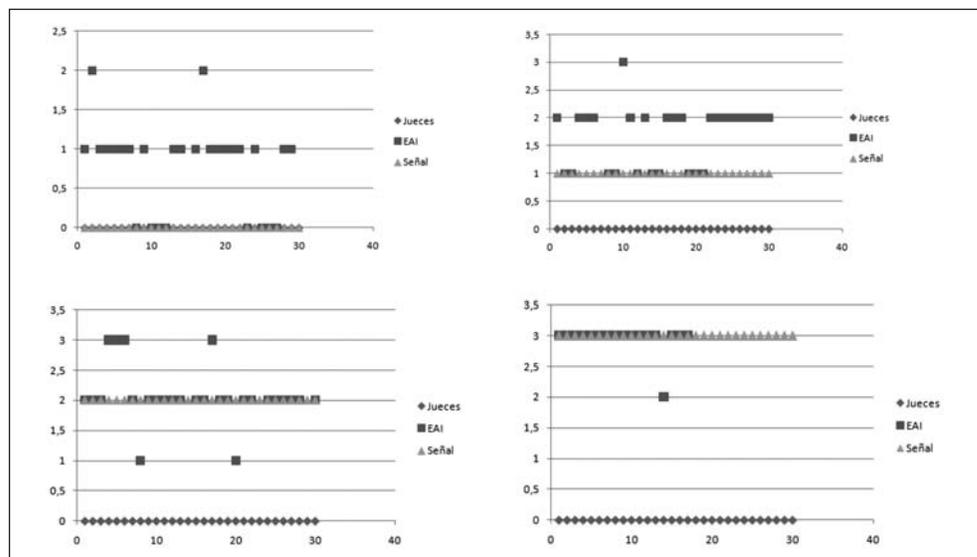


Figura 1: Las respuestas (EAI) del panel de jueces a los estímulos presentadas. La señal va aumentando su grado de alteración de 0: normal a 3: severo replicando la escala GRBAS

Tabla 1. *Análisis de ANOVA univariado*

	Wilks' Lambda	F	df1	df2	Sig.
Jitter_cc	,963	1,655	3	129	,180
Harmonics-to- Noise Ratio	,562	33,463	3	129	,000
Glottal Noise Energy	,432	56,533	3	129	,000
Lyapunov	,830	8,779	3	129	,000
Jit	,275	113,417	3	129	,000
Jitta	,275	113,367	3	129	,000
RAP	,276	112,540	3	129	,000
PPQ	,277	112,213	3	129	,000
DDP	,277	112,502	3	129	,000
HNR	,294	103,427	3	129	,000

Tabla 2. *Autovalores de la Función Discriminantes*

Función	Autovalor	% de Varianza	% Acumulado	Canonical Correlation
1	2,868	77,7	77,7	,861
2	,810	21,9	99,6	,669
3	,015	,4	100,0	,123

Tabla 3. *Coefficientes estandarizados de las funciones discriminantes canónicas*

	Función		
	1	2	3
Jitter_cc	,017	,198	,025
Harmonics to Noise_Ratio	-,368	-,982	-,373
Glottal Noise Energy	2,385	6,679	2,640
Lyapunov	1,385	4,092	,581
Jit	2,778	4,861	1,780

Tabla 4.

*Coefficientes de la función de clasificación*

	Puntaje			
	0	1	2	3
Jitter_cc	5,124	5,355	5,658	5,443
Harmonics to Noise_Ratio	,492	-1,552	-4,677	-4,307
Glottal Noise Energy	137,924	156,869	186,026	181,686
Lyapunov	38,423	44,357	52,449	51,001
Jit	97,331	110,743	131,274	135,194
(Constant)	-30,369	-36,094	-48,424	-50,796

Tabla 5.

*Resultados de la clasificación*

	Puntaje	Grupo de pertenencia pronosticado				Total	
		0	1	2	3		
Original	Count	0	7	0	0	0	7
		1	23	0	20	0	43
		2	1	0	53	1	55
		3	0	0	6	22	28
	%	0	100,0	,0	,0	,0	100,0
		1	53,5	,0	46,5	,0	100,0
		2	1,8	,0	96,4	1,8	100,0
		3	,0	,0	21,4	78,6	100,0
Cross-validated	Count	0	7	0	0	0	7
		1	23	0	20	0	43
		2	1	0	53	1	55
		3	0	0	6	22	28
	%	0	100,0	,0	,0	,0	100,0
		1	53,5	,0	46,5	,0	100,0
		2	1,8	,0	96,4	1,8	100,0
		3	,0	,0	21,4	78,6	100,0

Tabla 6.

*Centros iniciales de los conglomerados*

	Clúster			
	1	2	3	4
Jitter_cc	-1,7797	-1,2990	,24436	1,47990
Harmonics-to-noise_ratio	2,07226	,48666	-,66596	,34631
Gne	1,20094	,97350	-,91203	,37540
Lyapunov	-,77150	-,71745	-,58784	1,47904
Jit	-,66503	-,24617	2,15903	-,43530
Jitta	-,65795	-,21377	2,15595	-,41933
RAP	-,65982	-,27164	2,16113	-,42000
PPQ	-,63303	-,26628	2,16344	-,45243
DDP	-,65906	-,27222	2,16126	-,41991
HNR	1,39738	-,13960	-1,8027	,20275

Tabla 7.

*Centros de los conglomerados finales*

	Cluster			
	1	2	3	4
Jitter_cc	-1,7797	-,79742	,24436	,93341
Harmonics to Noise Ratio	2,07226	,63417	-,66596	,01745
Glottal Noise Energy	1,20094	,98009	-,91203	,33792
Lyapunov	-,77150	-,70984	-,58784	1,51739
Jit	-,66503	-,37015	2,15903	-,44002
Jitta	-,65795	-,34156	2,15595	-,46310
RAP	-,65982	-,38921	2,16113	-,42809
PPQ	-,63303	-,37735	2,16344	-,44752
DDP	-,65906	-,38936	2,16126	-,42830
HNR	1,39738	,31097	-1,8027	,07447

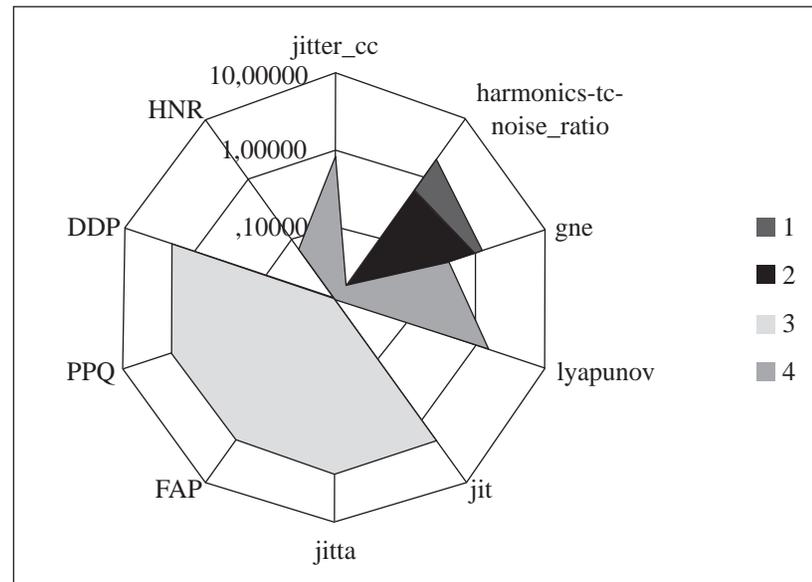
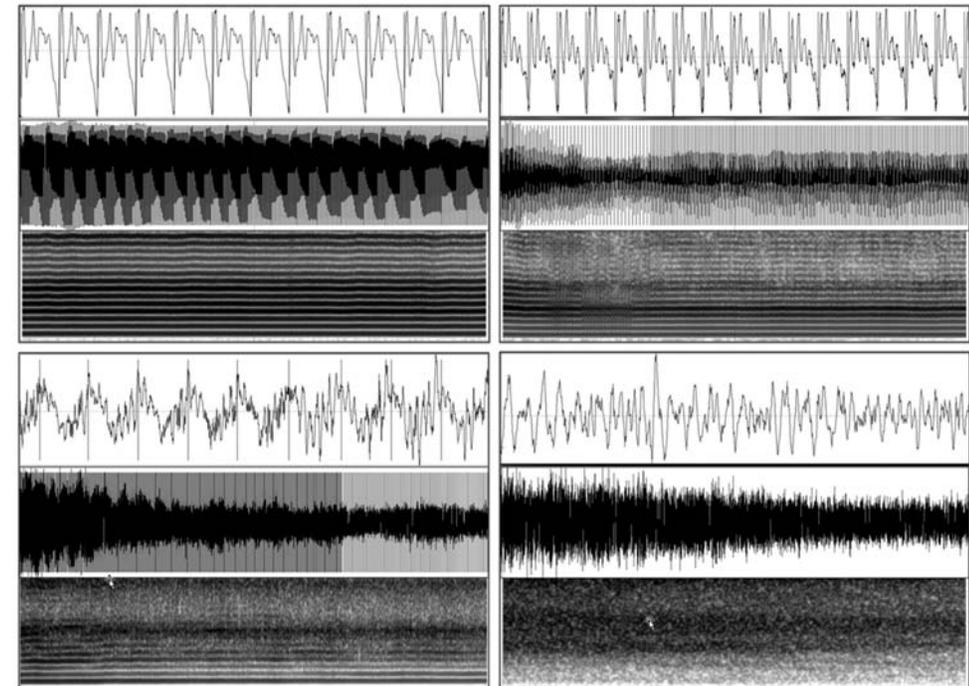


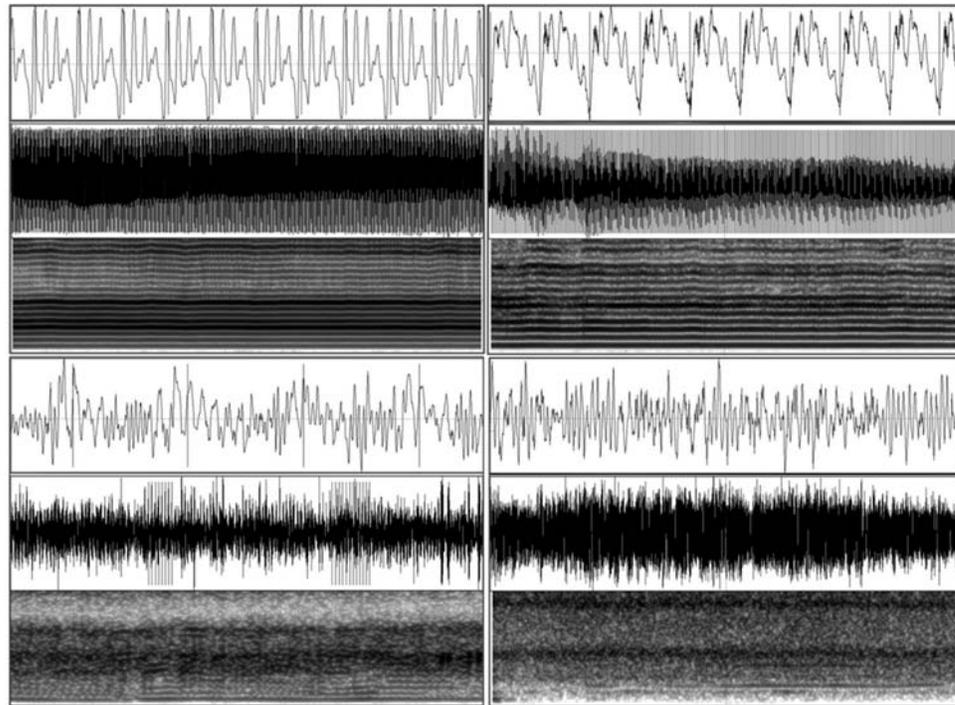
Figura 2. Representación radial del log 10 de las distancias medidas a los centros de los clúster

Anexo 1.

Estímulos utilizados en el test perceptual. Para la selección se siguió los criterios de tipificación de señales de Titze (Titze, 1995) y la modificación de (Sprecher, Olszewski, Jiang, & Zhang, 2010).



Hablantes femeninos. Tipos de señales, de izquierda a derecha: 1, 2, 3 y 4.



Hablantes masculinos. Tipos de señales, de izquierda a derecha: 1, 2, 3 y 4.

## Anexo 2.

### Glosario de medidas acústicas

Sistema	Medida acústica	Definición técnica
PRAAT	jitt	Diferencia absoluta media entre dos períodos consecutivos, dividido por el período promedio.
	jitta	Diferencia absoluta media entre dos períodos consecutivos.
	jitter_RAP	Perturbación relativa media: la diferencia absoluta media entre un período y el promedio de la misma y sus dos vecinos, dividido por el período promedio.
	jitter_PPQ	Cociente perturbación del período de cinco puntos: la diferencia absoluta media entre un período y el promedio de ella y sus cuatro vecinos más cercanos, dividido por el período promedio.
	jitter_DDP	Diferencia absoluta media entre diferencias consecutivas entre dos períodos consecutivos, dividido por el período promedio. El valor es tres veces RAP.
	HNR	Relación entre la potencia de la señal y la potencia de ruido que corrompe la señal
ANAGRAF	Jitter_cc (ANAGRAF)	Requiere coincidir con Jitter_pr para que la medición sea válida. La validez depende si la diferencia es de hasta 3 ciclos, hasta 5 ciclos o hasta 10 ciclos.
	harmonic-to-noise_ratio (ANAGRAF)	Adaptación del algoritmo de Eiji Yumoto
Michaelis	Glottal to noise Energy ratio GNE	Proporción de excitación glótica/ ruido
Evin	Lyapunov coefficients	Algoritmo de algoritmo de Wolf con las simplificaciones y parámetros sugeridos por Giovanni