

APLICACIÓN DEL ANÁLISIS ROC A LA EVALUACIÓN DE LA VOZ HUMANA: OBTENCIÓN DE EVALUADORES FIABLES

M. Pelegrina*, J.A. Adrián* y M.J. Rodríguez-Parra**
Universidad de Málaga* y Universidad de Granada**

Resumen

Este trabajo evalúa un conjunto de ítems que hace referencia a la calidad de la voz humana. Se aplican dos modelos: La evaluación de las características de la voz mediante la característica operativa del receptor (o análisis ROC) y la obtención de la fiabilidad de “jueces” (evaluadores múltiples). Como material objeto de la evaluación hemos utilizado un cuestionario para el diagnóstico de la voz (Casado y Adrián, 2002) que incorpora variables físicas, fisiológicas, psicológicas y lingüísticas. Estas variables (y sus correspondientes interacciones) se encuentran incluidas en los siguientes constructos teóricos: *Fatiga vocal, tensión muscular, extensión vocal (o escala de la voz), higiene vocal, ansiedad (y estrés), antecedentes de riesgo, procesos anteriores ante una disfonía, factores subjetivos asociados y factores hormonales.*

Los datos obtenidos nos permiten obtener una muestra sistemática de evaluadores fiables y un material (o conjunto de ítems) que permitirán a su vez diagnosticar la calidad de la voz, prevenir y evitar daños y controlar la evolución de los pacientes. Una vez que hemos obtenido un modelo válido y fiable, sería conveniente aplicarlo para evaluar modelos artificiales de voz, *mutatis mutandis*. De hecho, dependiendo de los objetivos que pretendamos, esta metodología podría ser usada para llevar a cabo aproximaciones más claras a un modelo de funcionamiento de voz, formas o estilos de comunicación o cambios de actitudes del oyente...

Abstract

This work evaluates a set of items that studies the quality of human voice. Two models have been applied: the evaluation of voice characteristics using the Receiver Operating Characteristic analysis (ROC) and the reliability amongst expert and non-expert judges (multiples raters). As material assessment, we used a questionnaire which considers physical, physiological, psychological, and linguistic variables in the voice diagnosis (Casado y Adrián, 2002). These variables (and their corresponding interactions) are including in the following theoretical constructs: *vocal fatigue, muscular tension, vocal extension (vocal scale), vocal hygiene, anxiety (and stress), background of risk, previous processes of dysphonia, subjective associated factors and hormonal influences.*

The data provide us a systematic sample of reliable judge criteria and a material (a set of items) that will allow specialists to obtain a better diagnosis and perception of the quality of a voice, preventing damages and control the patient improvement. Once we have obtained a valid and reliable model, it could be worth to apply it for artificial models of voice; *mutatis mutandis*. In fact, depending on the objectives that we want to get, this methodology could be also use to build a clearer model of voice functioning, forms or styles of communication, or changes of listener attitudes....

Introducción

Seleccionar una muestra de expertos que evalúen la calidad de la voz humana es una tarea relativamente compleja. En efecto, la voz humana incorpora variables físicas, fisiológico-anatómicas, psicológicas y las propias del lenguaje: fonéticas, morfológicas, sintácticas y pragmáticas. Un experto en voz no es un foniatra, ni un psicólogo, ni un especialista en medicina, ni siquiera un lingüista; pero es todo ello a la vez. En la actualidad, por ejemplo, se publican trabajos tan diversos como aquellos que estudian la voz como indicador en el diagnóstico de la esquizofrenia (Benson, 2002), o bien investigaciones que relacionan la voz con los procesos cognitivos y emotivos (Cook, 2002), y finalmente, otro ejemplo puede ser el estudio de la voz en periodos determinados del desarrollo y en actividades específicas (Kovacic y Buanovac, 2002).

No obstante, el objetivo de esta investigación consiste en clasificar los contenidos que hacen referencia a la calidad de la voz humana, así como a la obtención de una muestra de expertos que evalúen en qué medida las variables de los constructos citados en el resumen de este trabajo responden a dicha calidad.

Así, una vez seleccionada una muestra de expertos, pretendemos también determinar los mejores ítems para evaluar la voz. Con este objetivo partimos de un cuestionario inicial de 64 ítems (Adrián, Casado y González, 2001) que evaluamos con el fin de seleccionar la muestra más reducida posible de ítems que nos permita evaluar correctamente la voz. Lo que pretendemos es conseguir un número mínimo de ítems necesarios que cumplan con los principios de exhaustividad y exclusividad. Es decir, deben estar todos los ítems necesarios y cada uno de ellos debe representar variables independientes entre sí.

Otro objetivo consiste en formar técnicos expertos evaluadores de voz. Esta especialidad podría corresponder a titulados en logopedia, lingüística, psicología, etc. La evaluación de cuestionarios relacionados con la voz permite mejorar éstos, pero también conseguir evaluadores entrenados. La realización de esta actividad y la consecución de este objetivo es una actividad didáctica dentro de las especialidades citadas, especialmente en logopedia. Aun así, nos encontramos con que son muchas las medidas y datos que se obtienen en una evaluación funcional de la voz, y muchos los protocolos de evaluación propuestos y utilizados por los profesionales (p.e. Perelló, 1996; Bonet, 1997, Dejonckere, 2000, Dejonckere et al., 2001). Un reto importante que se plantean los especialistas hoy es conseguir acuerdos sobre un protocolo de exploración básico que aporte uniformidad metodológica al amplio y heterogéneo espectro de sistemas de medida utilizados en el diagnóstico de la voz.

Finalmente, pretendemos poner a prueba la eficiencia de los modelos de evaluación propuestos por Johnson y Albert (1999); Pelegrina, Ruiz y Wallace (2000), y Swets y Pickett (1982). Una integración de los anteriores trabajos permite obtener datos empíricos basados en la intervención de evaluadores múltiples (o *multiple raters*), la posible representación y estimación los datos obtenidos y la realización de juicios diagnósticos y/o evaluativos. Esta metodología es aplicable en ámbitos muy diversos, aunque en la práctica se ha centrado en el diagnóstico industrial, psicológico y médico. Finalmente, mediante este trabajo pretendemos extrapolar las citadas técnicas para la evaluación y diagnóstico de la voz artificial.

Así, los modelos empíricos obtenidos mediante evaluadores múltiples permiten iniciar un análisis de los datos obtenidos y la mejora progresiva del cuestionario. Es decir, una muestra de evaluadores realiza una primera valoración de 64 ítems. Ello permitirá la posible reducción de ítems y las mejoras de su calidad. Este proceso se puede valorar mediante estadística bayesiana o mediante cadenas de Markov, aunque tradicionalmente en psicología se han aplicado las técnicas de tests-retests (o diseños de medidas repetidas).

Para centrar los objetivos anteriores, partimos de una adaptación de la ecuación propuesta por Jhonson y Albert (1999) aplicada a la evaluación que permite predecir las funciones que obtendremos mediante los datos y la metodología que utilizaremos en las respuestas:

$$g(\{t_{ie}\}, \mathbf{Z}, \gamma, \sigma_e^2, a, b | \text{datos})$$

$$\alpha \prod_{e=1}^e \prod_{i \in T_i} \left\{ \frac{1}{\sigma_e} \exp \left[-\frac{(t_{ie} - Z_i)^2}{2\sigma_e^2} \right] I(\gamma_{e,y_{ie}} t_{ie} \leq \gamma_{e,y_{ij}}) \right\} \quad (1)$$

donde g es la distribución posterior conjunta, t_{ie} representa el conjunto de rasgos latentes percibido por los evaluadores (o jueces), $\mathbf{Z}$ es la matriz de densidad de probabilidades, γ es el punto de corte (o criterio) de los evaluadores (o jueces), σ^2 es la varianza de los evaluadores (o jueces) y a y b son posibles variables a comparar.

La fórmula anterior tiene dos partes fundamentales: una distribución exponencial referida a la distribución de datos obtenida después de la evaluación, y un procedimiento de evaluación mediante *rating scale*, como veremos seguidamente. La distribución exponencial representa y estima los datos obtenidos. Sin embargo, el procedimiento para la obtención de respuestas se desarrolla a partir de la teoría de la detección de señales (TDS) (Green y Swets, 1966) y de la elaboración de curvas características operativas del receptor (ROC) (Egan, 1975). Ver tabla 1.

Tabla 1. Respuestas basadas en la TDS para obtener curvas ROC

Estímulos	Respuestas	
	Sí	No
Señal	Aciertos 5, 4, 3, 2, 1	Rechazos incorrectos 5, 4, 3, 2, 1
Ruido	Falsas alarmas 5, 4, 3, 2, 1	Rechazos correctos 5, 4, 3, 2, 1

Nota: La tabla 1 muestra un procedimiento para evaluar señales y ruidos con las posibles respuestas, junto con una escala de estimación (o *rating scale*), con la que nos referimos a la respuesta dada ante la señal. El 5 indica que la respuesta es “muy segura”, 4 “segura”, 3 regular, 2 “insegura”, 1 “muy insegura”.

No obstante, para evaluar los ítems objeto de nuestra investigación asumimos que no hay “ruido” en dichos estímulos, ya que han sido elaborados por expertos en diagnóstico de la voz. Es decir el cuestionario propuesto se refiere a variables que asumimos afectan a la calidad de la voz y a su diagnóstico, por lo que la tabla 1 queda reducida a la tabla 2.

Tabla 2. Respuestas basadas en la TDS para obtener curvas ROC

Estímulos	Respuestas	
	Sí	No
Señal	Aciertos 5, 4, 3, 2, 1	Rechazos incorrectos 5, 4, 3, 2, 1

Nota: La tabla 2 muestra las respuestas posibles ante la presencia de un estímulo “señal”.

Si observamos la tabla 2, vemos que los resultados de la evaluación permiten obtener 10 puntos en una escala. En efecto, las respuestas a la señal (ítem) se valoran mediante “Sí” o “No” con lo cual los puntos de la escala van desde No 5 a Sí 5, representando una escala ordinal: No 5 equivale a un valor nulo de la señal (o ítem) y Sí 5 al valor máximo. Esto permite plantearnos una hipótesis de evaluación de los ítems según la cual los expertos ofrecerán más respuestas Sí a la señal que los no expertos. Una vez obtenidos los datos se calcula el índice de fiabilidad entre expertos.

Método

Participantes: Como grupo no experto participaron 28 estudiantes (mujeres y hombres) de primer curso de la licenciatura de psicología. Como grupo de expertos participaron 28 estudiantes de último curso de la diplomatura de logopedia. En el retest participaron otros 28 estudiantes de logopedia del mismo curso citado.

Material: Se valoró el Protocolo “Teatinos” de evaluación y diagnóstico funcional de la voz (Casado y Adrián, 2002).

Resultados y discusión

Los resultados permitieron una primera valoración sobre la diferencia entre las varianzas obtenidas en las respuestas de evaluadores expertos y no expertos. En este sentido y mediante la prueba de Levene se obtuvieron diferencias muy significativas desde el punto de vista estadístico: $F=309, p<0.000$. La magnitud del efecto fue de 0.71. Según estos datos las respuestas de los no expertos presentó una variabilidad muy alta.

En segundo lugar obtuvimos la correlación entre evaluadores. Así, la correlación entre las dos mitades del grupo de 28 “expertos” fue: $r=0,913$, $p<0,000$. No se dio correlación estadísticamente significativa en el grupo de los no expertos, sino que hubo una varianza amplia entre ellos como hemos indicado anteriormente.

Dados los resultados anteriores, elaboramos una réplica entre expertos con objeto de confirmar o descartar los resultados iniciales. Ésta consistió en obtener los datos de otros 28 evaluadores “expertos”. La correlación entre éstos y los expertos del primer grupo fue de $r=0,995$, $p<0,000$, y el coeficiente alfa de fiabilidad fue 0,997. Estos resultados confirman la alta correlación y fiabilidad entre expertos. Según estos datos, los alumnos de logopedia evalúan los ítems relacionados con el diagnóstico y calidad de la voz de una forma altamente fiable consistente.

Sin embargo, habíamos propuesto en la introducción una hipótesis predictiva sobre la distribución exponencial de los resultados empíricos. Así, observamos en la figura 1 el ajuste de las puntuaciones directas obtenidas mediante la función exponencial. En los evaluadores no expertos el resultado fue de $r^2 = 0.86$, mientras que en los evaluadores expertos $r^2 = 0.72$.

Insertar aproximadamente aquí la figura 1

Además, a partir de las frecuencias empíricas analizadas y representadas mediante la figura 1, se han obtenido las proporciones acumuladas (curva ROC), ver figura 2. La transformación obtenida de los datos empíricos ofrece para el grupo de no expertos: $r^2 = 0.96$ y para el grupo de expertos: $r^2 = 0.89$.

Insertar aproximadamente aquí la figura 2

Los resultados anteriores han permitido cubrir los objetivos planteados al inicio de esta investigación. Los analizamos siguiendo el orden secuencial de la aplicación empírica. En primer lugar la transformación de un modelo de evaluación basado en la TDS y las curvas ROC ha resultado útil para seleccionar una muestra sistemática de evaluadores expertos y fiables en el ámbito de la voz. En segundo lugar las distribuciones obtenidas mediante el ajuste exponencial representan una predicción del modelo de acuerdo con Johnson y Albert (1999). Dicha predicción se refiere a la actuación de los evaluadores múltiples y a su ajuste mediante curvas ROC, de acuerdo con la citada función exponencial.

Finalmente es nuestra intención proponer los modelos utilizados en este trabajo para que puedan ser extrapolados en la evaluación de la calidad de la señal en el habla artificial. Esto permitirá modular los logros de un determinado sistema en el ámbito de la voz y en el contexto pragmático del lenguaje.

Referencias

- Adrián, J.A., Casado, J.C., González, M. (2001) Protocolo preliminar “Teatinos” de evaluación y diagnóstico funcional de la voz. *Escritos de Psicología*; 5, 81-88.
- Benson, R. J. (2002). The influence of bias against the schizophrenia diagnostic label in social skills assessment. *Dissertation Abstracts International*, 62(12-B), 5953.
- Bonet, M. (1997) Disfonías. En: M. Puyuelo (Dir.), *Casos clínicos en logopedia*. Barcelona: Masson.
- Casado, J.C. y Adrián, J.A. (2002) *La evaluación clínica de la voz. Fundamentos médicos y logopédicos*. Archidona(Málaga): Aljibe
- Cook, N. (2002). Tone of voice and mind: the connections between intonation, emotion, cognition and consciousness. *Advances in Consciousness research*, 47.
- Dejonckere, P.H. (2000) Perceptual and laboratory assessment of dysphonia *Voice Disorders of Phonosurgery*, 33(4):731-749.
- Dejonckere, P.H., Bradley P., Clemente, P., Cornut, C., Crevier-Buchman, L., Friedrich, G., Van De Heyning, A., Remacle, M. and Woisard, V. (2001) A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. *European Archives of Otorhinolaryngology*, 258, 77-82.
- Egan, J. P. (1975). *Signal detection theory and ROC analysis*. New York: Academic press
- Green, D. M., y Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley.
- Johnson, V. E. & Alvert, J. M. (1999). *Ordinary data modeling*. New York: Springer-Verlag.
- Kovacic, G., Buanovac, A. (2002). Acoustic Characteristics of adolescent actors’ and non-actors voices. *Folia Phoniatrica et Logopedica*, 54 (3), 125-132.
- Pelegrina, M., Ruiz, y Wallace, A. (2000). Índices de discriminación y análisis ROC en el contexto de los modelos lineales generalizados. *Metodología de las Ciencias del Comportamiento*, 2, (1), 63-83.
- Perelló, J. (1996) *Evaluación de la voz, lenguaje y audición*. Barcelona: Lebón.
- Swets, J. A. & Pickett, R. M. (1982). *Evaluation of diagnostic systems. Methods from signal detection theory*. New York: Academic Press.

Anexo

Orden en la valoración de los ítems. Hemos incorporado del 1 al 30. El primer número representa el orden de los ítems hasta 30, el segundo la frecuencia empírica de elecciones, y el tercero el número de orden del ítem en el cuestionario original.

- 1.- 395 (6) Siente alivio tras esto (después de carraspear y aclararse la voz).
- 2.- 251 (49) ¿Padece o ha padecido problemas vocales de nódulos, pólipos, alguna enfermedad o traumatismo de las cuerdas vocales?
- 3.- 250 (12) ¿Siente que hace esfuerzos al hablar?
- 4.- 238 (58) ¿Ha seguido anteriormente algún otro tratamiento de rehabilitación vocal?: quirúrgico/ farmacológico/logopédico.
- 5.- 237 (28) ¿Fuma usted?.

- 6.- 235 (50) En su actividad profesional o vida diaria debe usar o usó en el pasado continuamente su voz (por ejemplo hablar mucho por teléfono)? ¿Debe o debió en el pasado permanecer en ambientes muy ruidosos hablando a un volumen muy alto forzando su voz
- 7.- 234 (3) ¿Cuándo ha empezado a notar molestias al hablar? (dolor, “gallos”, ronquera, fatiga, etc.) (se refiere al proceso disfónico actual, no a posibles procesos anteriores).
- 8.- 233 (48) ¿Tiene, o ha tenido, problemas asmáticos?
- 9.- 230 (60) *En caso de que respondiera qfirmativamente a la pregunta 47: ¿Estuvo en algún momento tomando medicamentos para la afección de O.R.L. que padeció?*
- 10.- 230 (8) ¿Tiene dolor de garganta, picor o sensación de cuerpo extraño?
- 11.- 226 (43) ¿Padece o ha padecido alguna enfermedad de O.R.L.? En caso afirmativo: ¿Cuál es y cuándo?
- 12.- 224 (38) ¿ Se considera una persona nerviosa, ansiosa o que se irrita fácilmente?
- 13.- 222 (2) En caso de tener molestias ¿dónde se localizan generalmente?
- 14.- 221 (13) Si habla algún tiempo seguido ¿nota que se fatiga y que a su voz le cuesta sonar y se cansa?
- 15.- 219 (24) ¿Padece o ha padecido alguna enfermedad pulmonar ? En caso afirmativo ¿Cuál es y cuándo?
- 16.- 216 (6) ¿Con frecuencia carraspea o tiene que aclararse la voz?.
- 17.- 215 (32) ¿En sus salidas nocturnas nota que tiene que gritar para hacerse entender y que rápidamente pierde la fuerza de su voz?.
- 18.- 214 (61) ¿Qué es lo que cree que le alivia más rápidamente su fatiga vocal, ronquera, dolor, etc. (disfonía)?:
- El uso de fármacos o remedios caseros: Pastillas para la garganta, inhalaciones, gárgaras de miel y limón, etc.
 - Dormir y descansar adecuadamente.
 - No hablar durante un tiempo.
 - Acudir al logopeda (en caso de haberlo hecho alguna vez).
 - Otros.
- 19.- 213 (29) ¿Toma bebidas alcohólicas?
- 20.- 207 (44) ¿Ha sufrido o sufre alguna enfermedad?. En caso afirmativo ¿Cuál es y cuándo?
- 21.- 207 (53) ¿Son regulares sus ciclos menstruales? (en su caso)
- 22.- 206 (47) ¿Tiene o ha tenido problemas alérgicos o asmáticos?
- 23.- 204 (37) ¿Se siente tenso físicamente?
- 24.- 204 (4) ¿En qué momento del día siente más dificultades o nota peor su voz?
- 25.- 203 (14) Si habla en voz alta ¿su voz enronquece?
- 26.- 203 (11) Nota habitualmente ronquera al hablar?
- 27.- 202 (21) ¿Siente molestias o tensión en la zona de la garganta al hablar?
- 28.- 202 (64) ¿Se ha sometido anteriormente a algún tipo de terapia física o psicológica?
- 29.- 202 (54) ¿Presenta desarreglos hormonales con frecuencia?
- 30.- 201 (26) En caso afirmativo ¿De qué forma? (en referencia a si tiene frecuentemente sequedad de garganta).

Figuras y pié de las figuras:

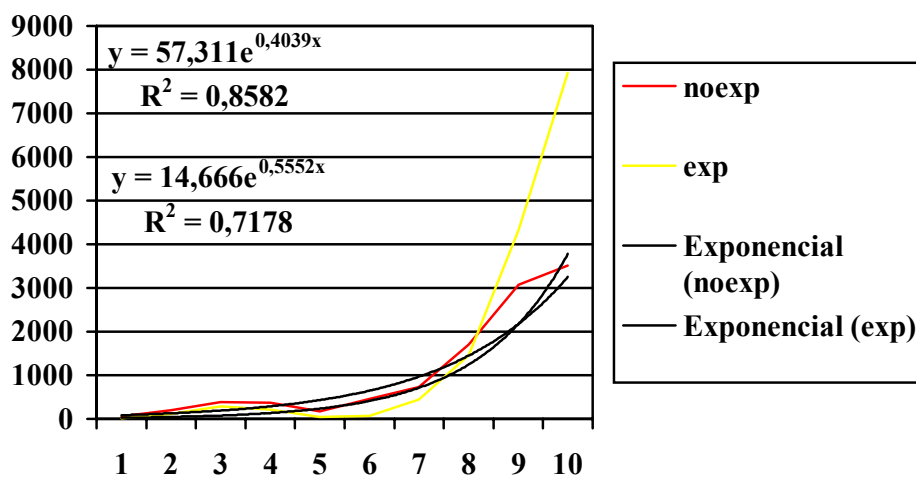


Figura 1: Representación obtenida (en color) y estimada (en negro) de las frecuencias obtenidas en la valoración de un cuestionario de voz mediante un grupo de no expertos y un grupo de expertos.

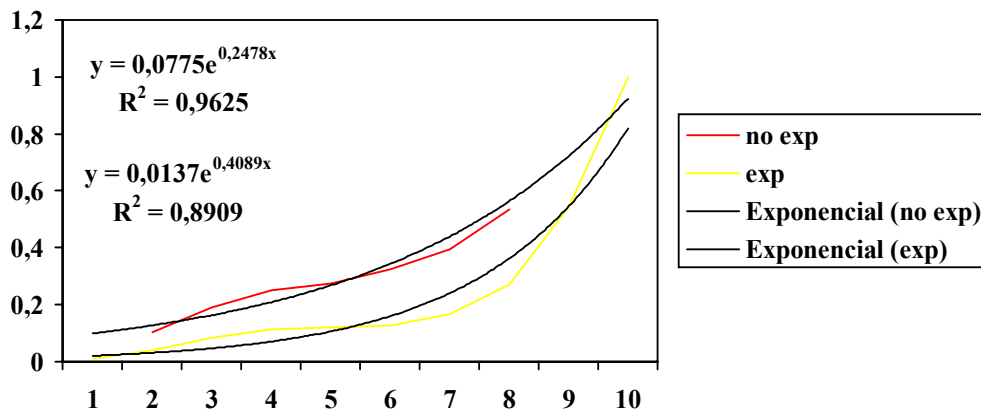


Figura 2. Representaciones obtenidas y estimadas de las curvas ROC correspondientes a evaluadores no expertos y expertos.