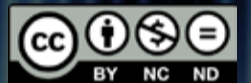


Hitos de la evaluación perceptual auditiva de la voz

¿hay evidencia?

Milestones of the auditory-perceptual evaluation of voice quality is there evidence?



Luisa Fernanda Angel **Gordillo**



VOCOLGYCENTER
todo comunica

ART Volumen 18 #2 julio - diciembre

Revista
ARETÉ

ISSN-I: 1657-2513 | e-ISSN: 2463-2252 *Fonoaudiología*

ID: 1657-2513.art.18206

Title: MILESTONES OF THE AUDITORY-PERCEPTUAL EVALUATION OF VOICE QUALITY:

Subtitle: Is there evidence?

Título: Hitos de la evaluación perceptual auditiva de la voz

Subtítulo: ¿hay evidencia?

Alt Title / Título alternativo:

[en]: Milestones of the auditory-perceptual evaluation of voice quality: is there evidence

[es]: Hitos de la evaluación perceptual auditiva de la voz: ¿Hay evidencias?

Author (s) / Autor (es):

Angel Gordillo

Keywords / Palabras Clave:

[en]: Voice quality; Auditory-perceptual evaluation; Voice perceptual evaluation; Perceptual Scales; GRBAS; CAPE-V; Reliability

[es]: Calidad de voz; Evaluación auditiva-perceptual; Evaluación perceptual de voz; Escalas Perceptuales; GRBAS; CAPE-V; Confiabilidad

Submitted: 2018-07-19

Accepted: 2018-08-19

Resumen

La evaluación perceptual auditiva es la herramienta más antigua y con mayor uso para determinar la idoneidad de la voz humana. Como resultado de una revisión minuciosa en PubMed se identifica que este enfoque ha evolucionado en lo conceptual y lo metodológico enmarcado en tres momentos: el primero caracterizado por el uso indiscriminado de etiquetas particulares, el segundo centrado en la construcción y refinamiento de escalas con confiabilidad intra e interevaluador, y a futuro, su posicionamiento como un estándar de oro que exalta los determinantes humanos y sociales de la voz. Las escalas más utilizadas en las investigaciones fueron GRBAS y CAPE-V. Adicionalmente, se argumenta la idoneidad de los vocablos más apropiados en idioma español para la denominación de “voice quality” equiparable a calidad vocal (indicador descriptivo multidimensional) o calidad vocal (valoración de severidad) según corresponda

Abstract

The auditory perceptual evaluation is the oldest and most used tool to determine the suitability of the human voice. As a result of a thorough review in PubMed it is identified that this approach has evolved in the conceptual and the methodological framed in three moments: the first characterized by the indiscriminate use of particular labels, the second focused on the construction and refinement of scales with intra- and inter-evaluator reliability, and in future, its positioning as a gold standard that exalts the human and social determinants of the voice. The scales most used in the reports were GRBAS and CAPE-V. Additionally, the suitability of the most suitable words in Spanish language is argued for the denomination of “voice quality” comparable to “calidad vocal” (multidimensional descriptive indicator) or “calidad vocal” (severity estimate) as appropriate.

Citar como:

Angel Gordillo, L. F. (2018). Hitos de la evaluación perceptual auditiva de la voz: ¿hay evidencia?. *Areté* issn-l:1657-2513, 18 (2), 65-74. Obtenido de: <https://revistas.iberoamericana.edu.co/index.php/arete/article/view/1413>

Luisa Fernanda Angel Gordillo, Msc BsH

Source | Filiación:

Universidad Nacional de Colombia

BIO:

Fonoaudióloga. o Maestría/Magister Universidad Nacional de Colombia - Sede Bogotá
Maestría en discapacidad e inclusion social

City | Ciudad:

Bogotá DC [co]

e-mail:

lfangelg@unal.edu.co

Hitos de la evaluación perceptual auditiva de la voz

¿hay evidencia?

Milestones of the auditory-perceptual evaluation of voice quality: is there evidence?

Luisa Fernanda Angel **Gordillo**

La cualidad/calidad vocal es el atributo vocal más accesible pues puede ser evaluado continuamente por el clínico, la persona misma y cualquiera que escuche la voz (Kreiman & Gerratt, 2011) Ha sido equiparada con el “timbre”, definida por los estándares ANSI/ASA (American National Standards Institute 2015) como un atributo multidimensional de la sensación auditiva que permite al oyente determinar si dos sonidos que no son idénticos, pero le son presentados como similares por tener la misma intensidad, tono, localización espacial (resonancia) y duración, son disimiles; y este atributo multidimensional frecuentemente se concreta mediante adjetivos calificativos (Ferrer, Haderlein, Maryn, Bodt, & Nöth, 2018)

La obtención de este atributo mediante el empleo de los sentidos, principalmente el auditivo, suele ser una práctica ideal para documentar los progresos del tratamiento y la eficacia del mismo, porque se orienta precisamente a realzar el acontecimiento que motivó a la persona a buscar un tratamiento. En consecuencia, la eficacia de esta práctica debe ser objeto de un razonamiento metódico para comprender la evolución conceptual y operativa de la misma a través de los tiempos y la relevancia que ha tenido como estrategia valorativa de la voz.

Si bien el propósito de esta revisión es el de reunir la evidencia y presentarla bajo esta consideración, debe en primera instancia esclarecerse el empleo de la terminología más apropiada para hablantes de español, y luego exponer los hechos claves y fundamentales dentro del contexto histórico de la evaluación perceptual auditiva.

Una revisión minuciosa pero no exhaustiva, se ejecutó exclusivamente en PubMed empleando los términos de búsqueda “auditory-perceptual evaluation” (118), “auditory-perceptual” AND “voice evaluation” (17), “GRBAS” (352), “CAPE-V” (59), “voice quality” AND “voice evaluation” (141). Se excluyeron los artículos cuya aplicación de la prueba no constituía en sí misma el objetivo de la investigación, sino que tan solo se le empleaba

como medio para determinar el comportamiento bajo una condición médica en particular o un desorden de voz. Por el contrario, se incluyeron los documentos cuyo propósito central era el de recabar, evaluar y producir información en torno a la evaluación perceptual auditiva, tales como revisiones sistemáticas, tutoriales, artículos de reflexión e investigación. No hubo ninguna restricción en cuanto al tiempo de generación de la información. Como resultado de la lectura de 30 artículos que cumplían los criterios de inclusión y exclusión, y pudieron ser ubicados, se categorizó la información en tres rangos de tiempo que mostraban similitud en los propósitos de aplicación del enfoque. Los acontecimientos puntuales y significativos que marcaron cada rango de tiempo delimitan los hitos en el desarrollo de la evaluación perceptual auditiva de la voz.

Cualidad vs. Calidad vocal: Un dilema de interpretación

En la literatura científica uno de los elementos fundamentales en la evaluación y descripción de la voz humana hace referencia a “voice quality”, (Kreiman y Gerratt, 2011) una combinación de palabras que en el idioma español puede ser traducido (o debería decirse mejor interpretado) como “cualidad” o “calidad” vocal. Para los cuidadosos de la exactitud de los términos empleados en cualquier ámbito, pero particularmente en el ámbito académico, se plantea una disyuntiva frente a interpretar este término como “calidad” vocal en lugar de “cualidad” vocal y viceversa. Sin embargo, las tendencias para preferir uno sobre otro deberían epistemológicamente corresponder con las perspectivas y orientaciones de quienes emplean dichos términos. Esta disertación tiene como propósito exponer las similitudes semánticas de ambos términos que le permitirían asertivamente ser empleados como uno u otro, así como también, las divergencias semánticas gestadas bajo perspectivas que subyacen a una u otra interpretación. Y de esta forma dirimir las controversias académicas vigentes sobre si el empleo de “cualidad vocal” se constituye en una denominación incorrecta del pasado.

Similitudes y divergencias semánticas

Los términos cualidad y calidad comparten exactamente el mismo origen tanto del latín como del griego. Ambos se derivan del latín *qualitas*, -*ātis*, y este del griego. ποιότης *poiótēs*. Según la Real Academia (de la lengua) Española (RAE, 2016) el significado de cualidad es (1) Elemento o carácter distintivo de la naturaleza de alguien o algo; (2) Cualidad positiva, especialmente de una persona; y (3) Calidad, condición o naturaleza de algo o de alguien. En este sentido, el término cualidad vocal podría emplearse tanto para describir un elemento o carácter distintivo de la naturaleza de la voz como la valoración positiva

de la calidad, condición o naturaleza de la voz.

Por su parte el término *calidad* es (1) Propiedad o conjunto de propiedades inherentes a algo, que permiten juzgar su valor (2) Estado de una persona, naturaleza, edad y demás circunstancias y condiciones que se requieren para un cargo o dignidad; y (3) Importancia o gravedad de algo. En tal sentido, el término *calidad vocal* debería emplearse para describir una propiedad o conjunto de propiedades inherentes a la voz que le permiten juzgar su valor y darle importancia o gravedad (severidad).

Es así como epistemológicamente si se están describiendo caracteres distintivos de la naturaleza de la voz (una voz proyectada, sensual, brillante, ronca, estridente, áspera, aspirada, soplada y otros similares), es acertado emplear el término *cualidad vocal* la cual por definición también involucra la calidad como parte de su definición. Por otra parte, si tan solo se desea emplear estas propiedades inherentes para juzgar su valor, sea más acertado emplear el término *calidad vocal*.

Por su parte la palabra anglosajona “*quality*” se define como (Steiner, 1997) (1) un atributo o característica en este caso equiparable a cualidad, (2) puede ser relativo a la excelencia, en este caso equiparable a calidad, (3) relativo a la naturaleza, equiparable a ambos términos: cualidad y calidad; y/o (4) en el ámbito musical a timbre y tono, con lo que “*voice quality*” se traduce como timbre de voz, tono de voz.

Sobre este particular, el concepto de cualidad vocal ideal o de una cualidad vocal que sea socialmente aceptada se ha asimilado concretamente al vocablo calidad vocal. Existen cientos de cualidades vocales que podrían ser clasificadas como ideales. Los especialistas de la voz (Profesores de canto, Otorrinolaringólogos, Fonoaudióloga(o)s) pretenden identificar un estado “ideal” para un acto comunicativo y la búsqueda de lo ideal depende de la perspectiva de cada uno. Pero en la concepción de la profesión de fonoaudiología, no sólo se intenta juzgar la voz de la persona, sino sencillamente describir su naturaleza fisiológica (aunque algunos clínicos tradicionalistas dirían describir el síntoma), sobre la cual determinar acciones para la modificación de dicho comportamiento.

En conclusión, el vocablo utilizado cualidad / calidad refleja la concepción epistemológica (*atributo descrito / valoración atribuida*) de quién la emplea. Por consiguiente, el empleo del término *cualidad vocal* podrá emplearse para la descripción de los atributos de una voz en particular como en la evaluación; pero si la intención es la de establecer una severidad o una valoración concluyente como ocurre en el diagnóstico, sería más acertado emplear el término *calidad vocal*. Adicionalmente, ha de tenerse en cuenta que la *cualidad vocal* se circunscribe a la caracterización del timbre y no se consideran las categorías tono, sonoridad y fonética; mientras que la *calidad vocal* suele incluir en su concepción también los otros atributos del sonido como lo son el tono y la sonoridad.

Evaluación perceptual auditiva de la voz

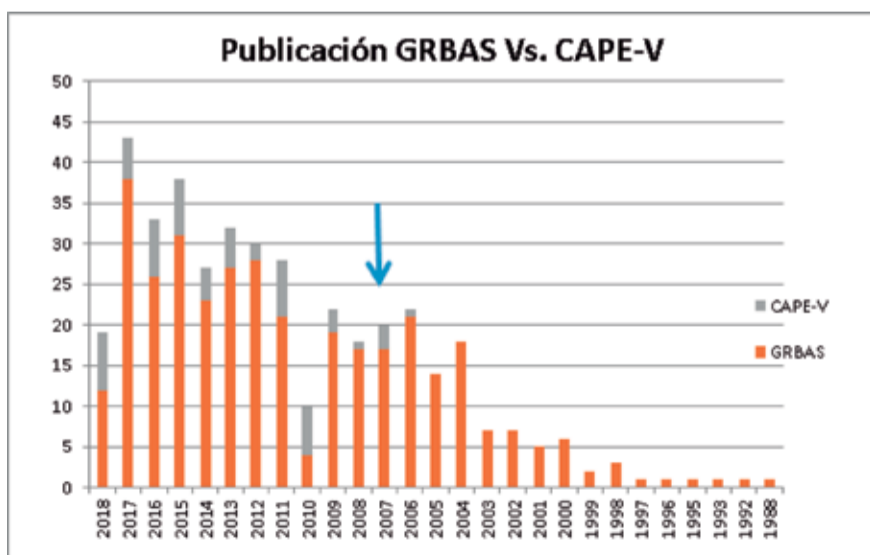
Las personas buscan tratamiento para los desórdenes de la voz porque no suenan normales y, a menudo, deciden si el tratamiento ha sido exitoso en función de si suenan mejor o no (Kreiman, Gerratt, Kempster, Berke, 1993).

. Hoy día, una evaluación clínica completa de la voz por lo general incluye la evaluación perceptual auditiva, además de la obtención de imágenes endoscópicas y la evaluación acústica y aerodinámica del funcionamiento de la voz. Conocer una voz exige la consolidación de diversos tipos de información. La voz se puede mirar desde diferentes perspectivas, una de las cuales es observarla como un “hecho físico” producto de la acción coordinada del sistema respiratorio, los pliegues vocales y el movimiento supraglótico; y como tal se puede conocer perceptualmente utilizando la vista, el oído, y el tacto.

En los orígenes de la profesión de Fonoaudiología la evaluación se centró en la información perceptual, aunque más adelante se introdujera el desarrollo de tecnologías acústicas y fisiológicas para la recolección de información sobre la voz. Tradicionalmente la voz humana ha sido descrita contemplando las características de un sonido: duración, tono, intensidad y timbre. Y para caracterizar este último se emplearon términos como voz ronca, áspera, raspada, metálica, con escape de aire/soplada, hipernasal, hiponasal, desnasal. Este enfoque subjetivo usa calificaciones perceptivas de los oyentes en una dimensión dada, generalmente en una escala predeterminada. En contraposición, en el enfoque objetivo de la evaluación del timbre empleando tecnologías acústicas, se miden algunos parámetros de la forma de onda acústica, con la creencia de que estos parámetros de alguna manera influyen en la respuesta provocada. Pero sea cual fuere el enfoque, se reconoce que la calidad vocal es multidimensional y recibe influencia de factores tales como el tono, la intensidad, la precisión articulatoria, la velocidad y ritmo del habla, y características de resonancia (Ferrand, 2012)

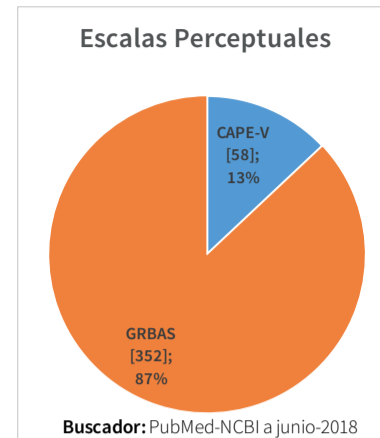
La evaluación perceptual-auditiva de la voz, tanto en el pasado como en el tiempo presente está llena de dificultades, incluida la falta de terminología estándar o de un procedimiento estándar para describir la cualidad de la voz desordenada, y la mala fiabilidad inherente a los juicios de percepción auditiva de la calidad de la voz; y sin embargo se considera el procedimiento de evaluación de la voz que más utilizan los patólogos del habla y el lenguaje/Fonoaudióloga(o)s. Con tal énfasis son dos las escalas mayormente difundidas hasta junio de 2018, en las publicaciones periódicas: GRBAS y más recientemente CAPE V (ver, gráficos 1 y 2).

Gráfico 1. Línea de tiempo de investigaciones empleando GRBAS (1988-2018) y CAPE-V (2006-2018).



Datos obtenidos de PubMed. Elaboración propia.

Gráfico 2. Comparativo de número de investigaciones empleando GRBAS (1988-2018) y CAPE-V (2006-2018).



Datos obtenidos de PubMed. Elaboración propia.

La escala GRBAS de la Sociedad Japonesa de Logopedia y Foniatría recibió su presentación internacional y reputación a través de la publicación de Hirano (Hirano, 1981) en 1981 y al parecer se convirtió en el método más utilizado. El GRBAS (y sus versiones adaptadas RBH, GRB, GRBASI, RASATI, GRBSPLxx) son la escala probablemente más compacta de todos los sistemas de calificación perceptivos y fue concebida para su uso diario por cualquier miembro del equipo de voz. Lo anterior debido principalmente a la simplicidad de la escala, con solo cinco parámetros (GRBAS: severidad-ruido-aire-debilidad-tensión) y cuatro categorías de severidad (cero/sin compromiso - tres/severo). Tal vez por este motivo, GRBAS es ampliamente utilizado a nivel internacional, a pesar de parecer demasiado insensible a pequeñas variaciones en la calidad de la voz. También ha sido objeto de críticas por la falta de instrucciones estándar para la administración y el análisis, aun cuando De (De Bodt et al,1997) afirma que la sociedad Japonesa hizo una cinta estandarizada con muestras de voces típicas representadas para la escala GRBAS. Reducciones de GRBAS son la versión GRB, (Dejonckere et al.,2001) y en los países de habla alemana, la escala RBH (Rauhigkeit, / Behauchtheit / Heiserkeit). Por otra parte, para dar cuenta de la posible intermitencia de las alteraciones o la inestabilidad, otras adicionan un componente I: para formar la escala GRBASI o la versión traducida denominada RASATI ampliamente utilizada en Brasil. Para enfatizar la relación con la Escala CAPE-V podría ser representada por el acrónimo GRBSPLxx, donde la minúscula “xx” representa los parámetros adicionales opcionales del CAPE-V.

Por su parte, la escala CAPE-V de la Asociación Americana de Patólogos de Habla y Lenguaje desarrolló por consenso un método estándar para la evaluación auditiva-perceptual en 2002, y bajo el liderazgo de Richard I. Zraick, intentó evaluar la validez y confiabilidad del CAPE-V. Su equipo encontró que las calificaciones de CAPE-V se correspondían bien con las calificaciones de las mismas voces por parte de los mismos jueces en las escalas ya existentes de GRBAS. Este hallazgo estableció la validez empírica (concurrente) del CAPE-V. El equipo de Zraick encontró una fiabilidad del evaluador ligeramente mejorada para el CAPE-V sobre las calificaciones de GRBAS. Pero esta confiabilidad aún varía mucho entre evaluadores y parámetros perceptuales (Hilman, 2013). En CAPE-V, se elimina la dimensión de debilidad (astenia) y se incluyen el tono y el volumen, lo que permite calificar las dimensiones en una escala visual analógica fuera del alcance de la cualidad vocal (timbre), además de proporcionar dos parámetros adicionales para ser especificados por el oyente si lo considera relevante: las deficiencias de resonancia o el temblor. Ambos hechos, en concepto de sus partidarios, hacen de CAPE-V una escala mejor adaptada para la disfonía en general que GRBAS, que se limita a la ronquera. CAPE-V también define un protocolo que describe los ejercicios vocales sobre los cuales deben efectuarse las calificaciones (Ferrer et al., 2018)

¿hay evidencia

La evaluación perceptual es fundamental para la evaluación de la cualidad/calidad vocal y su impacto en las habilidades de comunicación. Sin embargo, es uno de los temas más controvertidos en la evaluación de voz, debido a la mala correlación entre calificadores. La revisión de la literatura revela una amplia variedad de escalas de calificación y datos de confiabilidad que fluctúan de un estudio a otro (Kreiman et al., 1993); (De Bodt, Wuyts, Van, 1997); (Blaustein, 1993); (Bassich, Ludlow, 1986). Múltiples factores contribuyen a este fenómeno: la falta de consenso en una terminología para describir la cualidad de la voz, diferentes estrategias de percepción entre jueces (el acuerdo entre los evaluadores está influenciado por los tipos de escala, la muestra de habla y el calificador) y la importancia de la capacitación y la experiencia. A pesar de la sofisticación y la experticia se cuestiona la confiabilidad en las calificaciones perceptuales intraevaluador e interevaluadores (Kreiman y Gerratt, 2010). Para que las clasificaciones perceptuales sean significativas, los oyentes deben usar escalas coherentes: un evaluador en particular debe evaluar una muestra de voz de la misma manera cada vez que la oye. Además, para que las clasificaciones sean clínicamente útiles, el acuerdo entre diferentes evaluadores debe ser alto: cada evaluador que escuche una muestra de voz debe calificarlo de manera similar. Por lo tanto, la confiabilidad de tales juicios es un tema central en el estudio de la calidad de la voz y los desórdenes de la voz en el transcurrir de los tiempos.

A continuación, se exponen los propósitos y prácticas de investigación durante cada período de tiempo.

Hitos

Austin (1996) afirma que la recolección de datos focalizados en la cualidad/calidad vocal se puede remontar hasta los escritos de los romanos sobre oratoria, en los cuales se identifican términos familiares como áspero, claro, brillante, suave, débil, agudo, profundo, sordo, delgado, ronco y metálico. Las descripciones incluían términos relacionados con los estados emocionales y de personalidad de los hablantes (confundido, triste), y con la habilidad retórica y la articulación (articulado, distinto). Sin embargo, las similitudes con los términos empleados hoy día superan con creces las diferencias, y muchas etiquetas han estado en uso constante durante siglos. El peso de esta larga tradición da a tales prácticas un anillo de verdad que refuerza la creencia generalizada en su validez (Kreiman & Gerratt, 2011).

1930-1950

Desde 1939 cuando el estudio de (Moore, 1939) clasificó las voces de su muestra poblacional para correlacionarlas con rasgos de personalidad a partir de sus cualidades vocales (con escape de aire, nasal, áspera y metálica), se rastrea nuevamente esta práctica de clasificación de la voz con el propósito de determinar el efecto de la cualidad vocal en la comunicación, en estudios de la década de los cincuenta. Los investigadores eran conscientes que los estudios experimentales de desórdenes de la calidad vocal exigían medidas de grado de severidad que fueran confiables y válidas, pero reconocían que, aunque parecían estar bastante bien identificados y establecidos como entidades clínicas, ciertos desordenes de la cualidad vocal (por ejemplo: la aspereza, la nasalidad, el escape de aire) variaban de texto a texto. Por ejemplo, Perkins (1971) en su libro manifestó haber identificado 27 términos en nueve textos con solo dos cualidades mencionadas por todos los autores: ronquera y nasalidad; y (Gelfer, 1988) identificó más de 57 términos. Ya fuera por las etiquetas descriptivas particulares con las que se les identificaban o el número de categorías en las que se les agrupaba, la divergencia de definiciones y descripciones sobre la base de la información existente excluía la posibilidad de una clasificación final (Sherman, 1952); (Sherman, 1954); (Diehl, 1956)

Tabla 1. Artículos investigativos de la década de los sesenta y setenta centrados en la cualidad vocal. con estímulo modelo.

Autor (año)	Dimensión	Escala de calificación	Confiabilidad
Shipp & Huntington (1965)	Ronquera	8 puntos (0-7)	.17 - .69
Lively & Emanuel - 1970	Ronquera	5 puntos	.98
Kruel & Hecker - 1971	Escape de aire	Desconocido	.95*

* con estímulo modelo.
Fuente: Elaboración propia

1960 - 1990

La perspectiva de una clasificación fue la inquietud que animó a investigadores de la década de setenta y ochenta a formular y construir herramientas que sistematizaran las estrategias de percepción de los jueces/evaluadores; y a partir de allí, sus investigaciones a determinar la validez y confiabilidad de dichas herramientas.

Tabla 2. Artículos investigativos de la década entre los sesenta y los ochenta centrados en la escala perceptual auditiva.

Autor (año)	Dimensiones	Escala de calificación	Confiabilidad
Darley, Aronson, & Brown (1969)	38	7 puntos	80%
Hammarberg et al. (1980)	28	5 puntos	.93 a .97
Bassich C J, Ludlow CL. (1986)	13	7 puntos	80%

Fuente: Elaboración propia

En la década de los noventa la determinación de la validez y confiabilidad de las herramientas tomó tal preponderancia que condujo a reconocer la influencia psicoacústica del evaluador en el resultado de las mismas. Bajo esta premisa el estudio de (Kreiman, et al., 1993) reconoció que la calidad de la voz percibida se había medido utilizando una variedad de escalas:

- ➔ **Calificaciones categóricas:** implican la asignación de muestras de voz en categorías discretas y sin orden (por ejemplo, entrecortada, áspera).
- ➔ **Intervalos de apariencia igual:** requieren que los oyentes asignen un número entre uno y n a una muestra de voz, donde n es el número de puntos en la escala. (*Perfil vocal de Buffalo*)
- ➔ **Visuales analógicas (VA):** son líneas indiferenciadas, a menudo de 100 mm de largo. (*CAPE V*)
- ➔ **Estimaciones directas de la magnitud:** los oyentes asignan un número a una muestra de voz para indicar el grado en que posee la cualidad que se evalúa. (*GRBAS*)
- ➔ **Comparaciones pareadas:** los oyentes comparan dos estímulos y juzgan el grado de similitud / diferencia relativa en alguna dimensión. (*prueba de clasificación auditiva de Newcastle - NeAR*)

Y empleando dos estrategias de confiabilidad que son tanto clínica como estadísticamente importantes:

- ➔ **Consistencia:** La relación de una calificación con otra de la misma muestra es constante, aunque ella pueda diferir de oyente a oyente. Se le designa como confiabilidad *Intra evaluador*.
- ➔ **Concordancia:** dos oyentes asignan significados idénticos a cada punto de escala, por ejemplo, su idea de lo que constituye ronquera o aspereza, sus definiciones de normalidad (no ronco o no áspero) y la distancia entre los puntos intermedios de la

escala, son todos iguales. Esto solo implica que los oyentes califican las voces de forma paralela, sin implicar que los valores de la escala tengan el mismo significado. Se le designa como confiabilidad *Inter evaluador*.

Por lo tanto, saber que un conjunto de calificaciones es confiable no necesariamente permite la predicción de las calificaciones de un oyente a las de otro, a menos que también se conozcan sus niveles

Gráfico 3. Caracterización de los jueces.



relativos de clasificación (por ejemplo, A siempre califica las voces un punto por debajo de B).

En 57 estudios revisados (1951-1990). (Kreiman et al, 1993) los hallazgos sobre la caracterización de los jueces, el entrenamiento para la calificación de la escala y las escalas empleadas, se pueden resumir en los gráficos 3, 4 y 5 respectivamente.

Gráfico 4. Entrenamiento de los jueces.

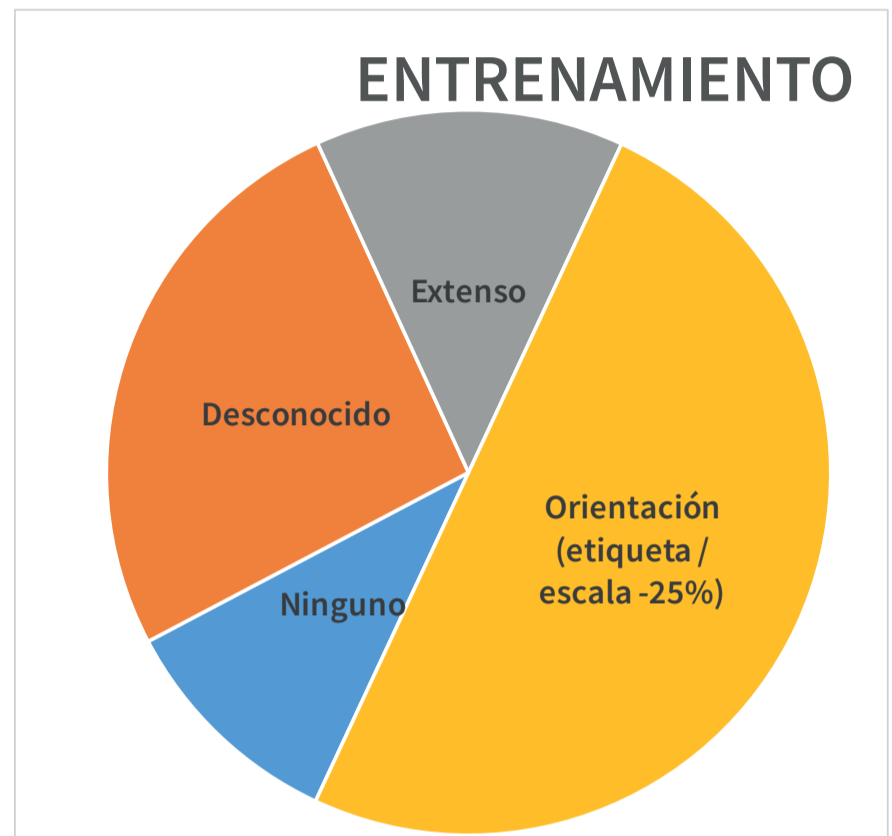
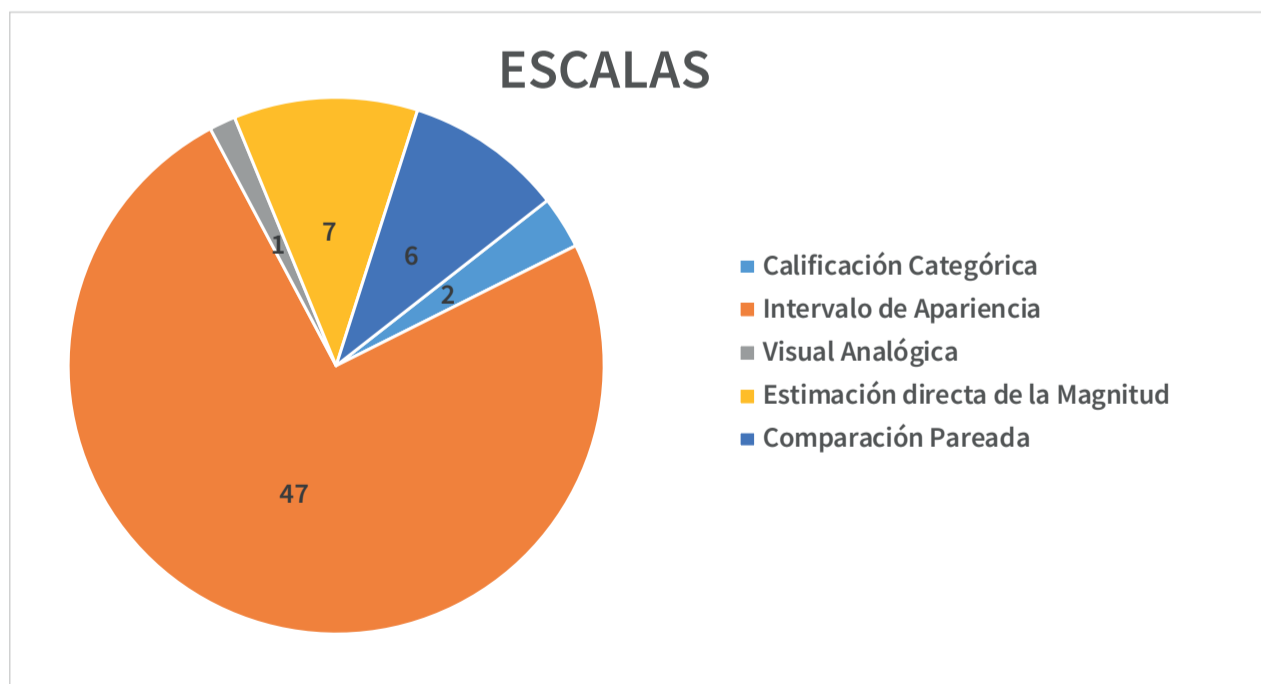


Gráfico 5. Caracterización de las escalas de evaluación perceptual auditiva



Elaboración propia a partir de los datos expuestos en la revisión de Kreiman et al. (1993).

En los estudios iniciales se habían identificado debilidades metodológicas tales como la falta de evaluación de la confiabilidad (concordancia) entre evaluadores, muy pocas repeticiones de la calificación para la estimación de la fiabilidad intraevaluador, no reportar los intervalos de confianza y el uso de estadísticas inadecuadas.

Pero con la implementación de nuevas investigaciones los resultados sugirieron que los niveles promedio de confiabilidad intra e interevaluador fueron relativamente altos, aun cuando las

calificaciones de las voces variaron ampliamente entre los evaluadores: Las clasificaciones variaron más entre voces patológicas que en voces normales y más entre leves a moderadas que en los extremos de escala; en las calificaciones de las Estimaciones directas de la magnitud las voces sonaron más ásperas la segunda vez que se evaluaron, aunque este fenómeno no se evidenció en las escalas visuales analógicas; y la presentación del orden de la tarea no tuvo un efecto significativo en el rendimiento del oyente evaluador.

En resumen, el hecho clave en la evaluación perceptual auditiva de la voz entre 1960 y 1990 fue el refinamiento gradual de la validez y confiabilidad de los procesos y herramientas empleadas para tal fin. Se afianzó la consistencia y la concordancia como estrategias de confiabilidad intraevaluador e interevaluador respectivamente.

2000-2018

A finales del siglo XX e inicios del siglo XXI se ha direccionado el énfasis en la construcción de modelos sistemáticos que emplean la capacitación y la experticia con el propósito de controlar algunas de las fuentes de variabilidad en las percepciones de la calidad vocal de los jueces. Actualmente se reconoce y destaca la contribución de los oyentes como una parte esencial de lo que después de todo es un fenómeno perceptual, en el cual el factor humano juega un papel importante en el uso de esta herramienta; y por tanto, las definiciones basadas en la fisiología no pueden acomodarse a los efectos de una medición de la voz que no es necesariamente constante entre los oyentes. (Barsties, 2015) Al respecto, (Kreiman & Gerratt, 2010) reconocen que los últimos 2.000 años han generado conciencia y descripciones de importancia sobre la voz y sus usos, pero el trabajo previo no había llevado a una comprensión teórica del “por qué” de la cualidad vocal. Como resultado, no existía evidencia psicoacústica para el desarrollo o validación de las técnicas de medición. La cualidad vocal nunca se había estudiado como parte de una teoría más amplia que abarque toda la cadena hablada, por lo que no podemos predecir cambios en el sonido vocal debido a cambios en las vibraciones de las cuerdas vocales o interpretar el cambio cuando este ocurre. Sin embargo, es posible que se esté llegando a una solución del problema a largo plazo al generar medidas confiables y válidas de la voz, que involucren precisamente su naturaleza psicoacústica.

Bajo este razonamiento, investigaciones recientes han identificado los factores que influyen en las calificaciones perceptuales de la voz. Según (Barsties y De Bodt, 2015) estos factores se pueden agrupar en tres categorías: oyente, estímulo y escala.

La primera categoría, la del oyente, se relaciona con el tamaño del grupo, factores internos intrínsecamente inestables del estándar interno de los oyentes (por ejemplo, fallas en la memoria, la atención, fatiga y errores), la restricción de la memoria auditiva para el reconocimiento, el sesgo del oyente frente al conocimiento de los antecedentes del diagnóstico médico, la experiencia del evaluador (disminuye o aumenta la confiabilidad), la formación profesional (por ejemplo, en otorrinolaringología, fonoaudiología, profesores de canto, fonetistas y estudiantes de pregrado o postgrado), la formación musical y el entrenamiento en el juicio perceptivo auditivo (por ejemplo, de oyentes nativos).

Sobre estos últimos aspectos, (Kreiman et al; 1990). mostraron que oyentes ingenuos (no entrenados) y expertos utilizan diferentes estrategias perceptuales y ello explicaría el por qué oyentes expertos muestran menos acuerdo que los oyentes ingenuos sobre la importancia de varios aspectos de la calidad de la voz, aunque sus calificaciones son algo más consistentes que la de los oyentes no entrenados (Gelfer, 1988) Por otra parte la formación profesional o los antecedentes culturales no parecen tener gran influencia en la calificación perceptual (por ejemplo, sobre la percepción de la ronquera) (Anders, 1988; Yamaguchi, 2003; Chaves, 2017.)

La segunda categoría, el estímulo, abarca aspectos como el efecto de la tarea seleccionada, mayor desacuerdo para establecer los desórdenes de voz entre leves y moderados que en las voces normales

o en los casos severos de desórdenes de voz, el uso de estímulos de anclaje (es decir, patrón de referencia) mientras se califica, y el transcurrir de las categorizaciones causadas por el contexto perceptual (por ejemplo, después de escuchar un número de voces patológicas levemente severas, la calificación para una voz patológica grave moderada se vuelve más severa mediante un cambio del estándar interno del oyente) (Kreiman,1992).

Téngase en cuenta que en los fragmentos de habla deben considerarse diferentes contextos incluyendo como mínimo la prolongación de vocales, la repetición de frases y habla conectada. En estos fragmentos la mejor correlación entre jueces empleando GRBAS, se ha encontrado para el criterio de severidad (G). La confiabilidad test-retest parece ser más confiable para ruido (R) y aire (B), y disminuye para debilidad (A) y tensión (S). Y a pesar de que las impresiones sobre A y S fueron menos consistentes, A mostró mayor variabilidad (De Bodt, 1997; Iwarsson,2017)

La tercera categoría, la escala, incluye el factor de las diferencias de confiabilidad resultantes de la escala visual-analógica u ordinal. Por una parte se destaca el abordaje de la coherencia y la concordancia cuando se presentan las escalas de percepción, prestando especial atención a las debilidades metodológicas y uso estadístico inadecuado que ponen en peligro la comparación entre escalas (Kreiman,1993; Iwarsson, 2017) Aunque una escala analógica visual parece permitir un juicio más fino sobre la calidad de la voz con una mayor libertad de juicio, el estudio de Wuyts et al. señaló que la concordancia disminuye considerablemente. Al parecer el acuerdo entre jueces es claramente mayor con la escala ordinal y las diferencias en la interpretación de los diferentes ítems de la escala también se reflejan mejor en la escala ordinal que en la escala visual analógica, lo que podría propiciar favoritismo por las escalas ordinales al reflejar con mayor precisión la realidad. Pese a lo anterior, cuando se correlacionan los datos perceptuales con las mediciones instrumentales la concordancia es mayor al usar escalas visuales-analógicas (88%) que escalas ordinales (64%) (Yu, 2002).

Por lo anteriormente expuesto, la prioridad es la de impulsar la confiabilidad involucrando estos tres factores, y con ello toma gran relevancia la capacitación al asumir la experiencia auditiva (Hirano,1981;Bassich y Ludlow, 1986 ; Anders et, al 1988; Fex,1992). En otras palabras, debe darse preponderancia al valor de la calibración auditiva de quienes ejercerán como jueces en la evaluación perceptual auditiva de la voz; así como se contempla en pruebas objetivas la calibración de los equipos de medición. Y en el proceso el factor humano-cultural, no se configura como un obstáculo o debilidad del proceso, sino todo lo contrario, como una fortaleza que da valor diagnóstico al proceso.

En este paso de calibración auditiva debe prestarse especial atención a dificultades para aislar atributos individuales de patrones complejos acústicos de voz (el factor más importante); inestabilidad de los estándares internos de los oyentes para diferentes cualidades; resolución de la escala de medición; y la magnitud del atributo que se mide. Cuando estos factores se controlaron mediante el uso de una tarea de ajuste de método, los oyentes acordaron exactamente el 96% del tiempo en sus evaluaciones perceptuales auditivas de la voz (Kreiman,2007).

Quizás por ello, las definiciones más recientes de la cualidad vocal (ANSI;2015) plantean como aspecto relevante en su definición el requisito de la ocurrencia del atributo en el sonido (objetivo) y la obtención de la sensación asociada (perceptual) en el oyente. Consecuentemente, se formula el propósito de relacionar la sensibilidad y especificidad de la herramienta, en comparación con los datos acústicos o aerodinámicos obtenidos. (Ferrer,2018) Aquí, sensible significa que la medida se

correlaciona altamente con la percepción. Específico significa que no se correlaciona con otros síntomas o aspectos perceptivos, y consistente significa que estos resultados son similares en diferentes estudios y bases de datos. Por consiguiente, la correlación de la evaluación perceptual con el dato acústico (o de cualquier otra índole instrumental como lo aerodinámico) está acrecentando el interés investigativo, y no precisamente para otorgarle credibilidad a la evaluación perceptual auditiva sino para posicionarla como una herramienta gold estándar.

En conclusión, **el juicio perceptivo auditivo es una herramienta de evaluación significativa para describir la cualidad/calidad de la voz bajo la consideración de muchos factores (oyente, estímulo y escala). Es casi imposible controlar todos estos aspectos en la práctica clínica, pero pueden jugar un papel importante para mejorar la precisión y la confiabilidad. Pese a lo anterior, el criterio de percepción auditiva sigue siendo el estándar de oro, pero su uso rutinario en este momento debe ser el soporte para establecer el grado de anormalidad en la voz (Barsties, 2015).**

Conclusiones

La evaluación perceptual auditiva de la voz es la herramienta más antigua y con mayor uso para determinar la idoneidad de la cualidad/calidad de la voz humana. Por ello, en primera instancia se hace necesario emplear apropiadamente los vocablos cualidad/calidad, señalando que durante la evaluación se explora la cualidad vocal directamente relacionada con el timbre de la voz, mientras que para el diagnóstico se juzga la calidad de la voz junto con los otros componentes del sonido como lo son el tono y la sonoridad.

En segunda instancia, reconocer las transformaciones que ha presentado la evaluación perceptual auditiva de la voz desde los escritos de los romanos sobre oratoria quienes caracterizaban la cualidad vocal, hasta mediados del siglo XX cuando los reportes de investigación sobre las prácticas de evaluación de la voz reportaron el uso de esta herramienta en sus protocolos. A partir de allí es posible identificar tres hitos, el primero motivado por el interés de caracterizar la cualidad vocal, pero sin que existiese unificación semiológica, ni metodológica sobre las cualidades vocales a identificar. El segundo hito a partir de la década de los sesenta, comprometido con el desarrollo de escalas y concreción de las cualidades vocales que impulsó la construcción de pruebas de evaluación perceptual auditiva con validez y confiabilidad, entre ellas GRBAS y CAPE-V. El tercer hito, a puertas del siglo XXI encaminó sus esfuerzos a fortalecer la confiabilidad reconociendo el factor humano de la voz y su multidimensionalidad, para entonces respaldar el valor psicoacústico de la evaluación perceptual auditiva. Por consiguiente, son propósitos vigentes y a futuro (Kreiman & Gerratt, 2011) el controlar sistemáticamente algunas de las fuentes de variabilidad en las percepciones de los jueces como son el oyente, el estímulo y la escala para obtener la mayor confiabilidad intraevaluador e interevaluadores, (America, 2016) determinar la sensibilidad y especificidad de la herramienta comparándola con herramientas instrumentales acústicas y aerodinámicas y posicionándola como una medida gold estándar, y (Ferrer, Haderlein, Maryn, Bodt, & Nöth, 2018) determinar la calidad vocal como parte de un constructo más amplio que abarque toda la cadena hablada, en un modelo que integra la producción y la percepción.

Referencias

- Acero, G., Cuervo, C., Florez, R. La voz un cambio de imagen. Documento de trabajo. Grupo Comunicando 90; 1991.
- American National Standards Institute. American National Standard bioacoustical terminology (ANSI S3.20-2015). New York, NY: Author. Disponible en: <https://asastandards.org/Terms/timbre/> Página web de ASA - Acoustical Society of America [acceso 24 Jun 2016; actualizada 24 Jun 2016]
- Austin, G. Chironomia, or a Treatise on Rhetorical Delivery. London: 1806. Ed. Mary Margaret Robb and Lester Thonssen. Carbondale, IL: Southern Illinois UP, 1966.
- Barsties B, De Bodt M. Assessment of voice quality: Current state-of-the-art. *Auris Nasus Larynx*. 2015 Jun; 42(3):183-8.
- Bassich C J, Ludlow CL. The use of perceptual methods by new clinicians for assessing voice quality. *J Speech Hear Disord* 1986; 51:125-33.
- Blaustein S, Asher B. Reliability of perceptual voice assessment. *J Commun Disord* 1983;16:157—61.
- Darley FL, Aronson AE, Brown JR. Differential diagnostic patterns of dysarthria. *J Speech Hear Res*. 1969 Jun;12(2):246-69
- De Bodt MS, Wuyts FL, Van de Heyning PH, Croux C. Test-retest study of the GRBAS scale: influence of experience and professional background on perceptual rating of voice quality. *J Voice*. 1997 Mar; 11(1):74-80.
- Dejonckere, P. H., Bradley, P., Clemente, P., Cornut, G., Crevier- Buchman, L., Friedrich, G., Woisard, V. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques: Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society. *Eur Arch Otorhinolaryngol*. 2001 Feb; 258(2):77-82.
- Diehl, C.F.; McDonald, E.T. Effect Of Voice Quality On Communication. *J Speech Hear Disord*. 1956 Jun;21(2):233-7
- Ferrand, C.T. Voice disorders scope of the theory and practice. The Allyn & Bacon communication sciences and disorders series. Pearson Education Inc. USA; 2012; p. 129-134.
- Ferrer CA, Haderlein T, Maryn Y, de Bodt MS, Nöth E. Collinearity and Sample Coverage Issues in the Objective Measurement of Vocal Quality: The Case of Roughness and Breathiness. *J Speech Lang Hear Res*. 2018 Jan 22; 61(1):1-24.
- Gelfer MP. Perceptual attributes of voice: development and use of rating scales. *J Voice*. 1988 Dec; 2(4): 320–326.
- Hammarberg B, Fritzell B, Gauffin J, Sundberg J, Wedin L. Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngol*. 1980 Nov-Dec; 90(5-6):441-51.
- Hillman, R. SIGnatures: The Quest for Voice Assessment Guidelines. The ASHA Leader, March 2013, Vol. 18 (SIG N), 54-55. Disponible en: <https://leader.pubs.asha.org/article.aspx?articleid=1784846> Página web de ASHA - American Speech-Language-Hearing Association. [acceso 20 Feb 2018; actualizada 31 May 2018]
- Hirano M. Clinical examination of voice. New York: Springer Verlag, 1981:81-4.
- Kreiman J, Geratt BR, Kempster GB, Erman A, Berke GS. Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *J Speech Hear Res* 1993; 36:21-40.
- Kreiman J, Geratt BR, Precoda K. Listener experience and perception of voice quality. *J Speech Hear Res*. 1990 Mar; 33(1):103-15.
- Kreiman J, Gerratt B. Perceptual assessment of voice quality: Past, present, and future. En: Ma EPM & Yiu EML. Handbook of voice assessments. San Diego: Plural Publishing; 2011 p. 291 - 304.
- Kreiman, J.; Gerratt, BR: Perceptual Assessment of Voice Quality: Past, Present, and Future. SIG 3 Perspectives on Voice and Voice Disorders, American Speech-Language-Hearing Association, July 2010, Vol. 20:62-67.
- Kreul EJ, Hecker MH. Descriptions of the speech of patients with cancer of the vocal folds. II. Judgments of age and voice quality. *J Acoust Soc Am*. 1971 Apr; 49(4):Suppl 2:1283+.
- Lively MA, Emanuel FW. Spectral noise levels and roughness severity ratings for normal and simulated rough vowels produced by adult females. *J Speech Hear Res*. 1970 Sep; 13(3):503-17.
- Moore, W. E. Personality traits and voice quality deficiencies. *J Speech Disord* 1939; Vol. 4: 33-36.

Hitos de la evaluación perceptual auditiva de la voz

¿hay evidencia

Perkins, W. H. Vocal function: Assessment and therapy. In L. E. Travis (Ed.), Handbook of speech pathology and audiology (pp. 481-503): New York: Appleton-Century-Crofts.

Real Academia Española. Diccionario de la lengua española. Edición del tricentenario. Disponible en: <http://dle.rae.es/?id=BRNxtDx> Página web de la Real Academia Española [acceso 16 Oct 2016; actualizada 31 May 2018]

Sherman D. and Linke E. The Influence Of Certain Vowel Types On Degree Of Harsh Voice Quality. J Speech Hear Disord. 1952 Sep;17(4):401-8.

Sherman D. The Merits Of Backward Playing Of Connected Speech In The Scaling

of Voice Quality Disorders. J Speech Hear Disord. 1954 Sep;19(3):312-21.

Shipp T, Huntington DA. Some acoustic and perceptual factors in acute-laryngitic hoarseness. J Speech Hear Disord. 1965 Nov; 30(4):350-9.

Steiner, R J. International dictionary. English / Spanish Spanish/English. Second edition. Simon & Schuster's, Inc. Webster's New World; USA; 1997; p. 604.

Yamauchi, E. J., Imaizumi, S., Maruyama, H., & Haji, T. Perceptual evaluation of pathological voice quality: A comparative analysis between the RASATI and GRBASI scales. Logoped Phoniatr Vocol. 2010 Oct; 35(3):121-8.

Con el acompañamiento de:



Calle 79 # 18-18 of. 206
Bogotá

