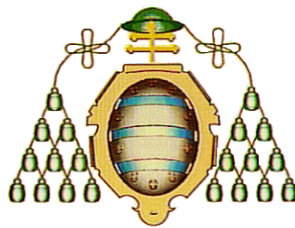


UNIVERSIDAD DE OVIEDO

Programa de Doctorado “Investigación en Cirugía y Especialidades
Médico-Quirúrgicas”



UNIVERSIDAD DE OVIEDO

TESIS DOCTORAL:

**“ANÁLISIS PERCEPTUAL DE LA DISFONÍA: ADAPTACIÓN Y
VALIDACIÓN DEL MÉTODO CAPE-V AL ESPAÑOL”**

PERCEPTUAL ANALYSIS OF DISPHONIA: ESTABLISHING VALIDITY OF THE CAPE-V

AUTORA:

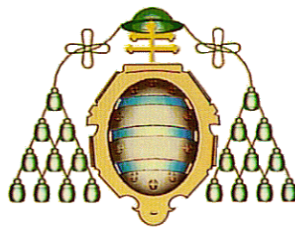
Dña. MARTA MORATO GALÁN

DIRECTOR:

Dr. D. FAUSTINO JOSÉ NUÑEZ BATALLA

UNIVERSIDAD DE OVIEDO

Programa de Doctorado “Investigación en Cirugía y Especialidades
Médico-Quirúrgicas”



UNIVERSIDAD DE OVIEDO

TESIS DOCTORAL:

**“ANÁLISIS PERCEPTUAL DE LA DISFONÍA: ADAPTACIÓN Y
VALIDACIÓN DEL MÉTODO CAPE-V AL ESPAÑOL”**

PERCEPTUAL ANALYSIS OF DISPHONIA: ESTABLISHING VALIDITY OF THE CAPE-V

AUTORA:

Dña. MARTA MORATO GALÁN

DIRECTOR:

Dr. D. FAUSTINO JOSÉ NUÑEZ BATALLA

Oviedo, Julio 2014

RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

1.- Título de la Tesis	
ANALISIS PERCEPTUAL DE LA DISFONÍA: ADAPTACIÓN Y VALIDACIÓN DEL CAPE-V AL ESPAÑOL	PERCEPTUAL ANALYSIS OF DISPHONIA: ESTABLISHING VALIDITY OF CAPE-V
2.- Autor	
MARTA MORATO GALÁN	
Programa de Doctorado: INVESTIGACIÓN EN CIRUGÍA Y ESPECIALIDADES MEDICO QUIRÚRGICAS	
Órgano responsable: DEPARTAMENTO DE CIRUGÍA Y ESPECIALIDADES MEDICO QUIRÚRGICAS	

RESUMEN (en español)

Introducción

La investigación en el campo de la evaluación perceptual de la voz se centra en la construcción de protocolos de calificación objetivos, consistentes y de difusión universal. Así, surge en una convención de especialistas en patología de la voz en EE UU en 2002 el *Consensus Auditory Perceptual Evaluation of Voice* o CAPE-V. Nuestro objetivo es adaptarlo y validarlo al idioma español para difundir su uso en este idioma. Se trata de una herramienta que se caracteriza por unas instrucciones concretas de aplicación del test y cuenta con 6 parámetros constantes: severidad general, aspereza, calidad aérea, tensión, tono e intensidad. La calificación se realiza en base a 3 tareas de fonación: vocales sostenidas, frases fonéticamente adaptadas y discurso natural. El sistema de calificación se basa en una escala de 100 puntos con apoyo analógico visual

Material y Métodos

Inicialmente se realiza la adaptación fonética de las frases al español. No consiste en una simple traducción, si no que se adapta el entorno lingüístico. Se evalúan 89 muestras vocales correspondientes a 66 individuos (30 hombres y 36 mujeres) con una edad media de 46 años (21- 80). 50 de los individuos padecen disfonía por un amplio rango de patologías vocales y 16 son sanos. Las muestras son evaluadas mediante el GRABS y el CAPE-V por dos observadores expertos. Se selecciona un subgrupo de 11 pacientes sometidos a fono-cirugía laríngea y se evalúan los resultados vocales. Se calculan a través de los estadísticos correspondientes: la validez intraobservador e interobservador del test, la validez de criterio o nivel de acuerdo entre el GRABS y el CAPE-V y la existencia de diferencias significativas entre las distintas tareas evaluadas. En el subgrupo de pacientes intervenidos se evalúa la sensibilidad de ambos test a los cambios post tratamiento.

Resultados:

La consistencia interobservador para SG: 0,935; CA: 0,894 ; ASP: 0,907. (Tarea 1)

La consistencia intraobservador para SG: 0,988; CA: 0,981; ASP: 0,979. (Tarea 1)

La validez de criterio para SG: 0,965; CA: 0,961; ASP: 0,950. (Tarea 1)
Sensibilidad del CAPE-V al cambio: para todos los parámetros se detectan cambios significativos.

Sensibilidad del GRABS al cambio: solo el Grado detecta diferencias entre las muestras antes y después del tratamiento.

Conclusión:

Tras la adaptación del CAPE-V al español, éste demuestra su validez y consistencia como test para calificación de la voz. Es una escala más flexible y completa que las existentes, que permite reflejar de forma fiable las características de la voz con la posibilidad de representar más matices que las escalas actuales. Evalúa tres tareas de fonación y su metodología estricta garantiza la objetividad de la que carecen la mayoría de las escalas previas. Los parámetros evaluados más consistentes son: la Severidad General (SG), Calidad Aérea (CA) y Aspereza (ASP). Se demuestran diferencias entre las calificaciones de cada tarea, sin embargo se mantiene el nivel de acuerdo y la consistencia del test.

En cuanto a la evaluación de los cambios, el CAPE-V es una escala con más recorrido que permite detectar diferencias sutiles en las muestras después de aplicar un tratamiento.

RESUMEN (en Inglés)

Introduction

The Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) was developed to provide a protocol and form for clinicians to use when assessing the voice quality of adults with voice disorders. This new tool take into account three vocal tasks: sustained vowels, phonetically adapted phrases and natural speech. It includes several perceptual ítems evaluated into the literature as so consistent. This study examines reliability and empirical validity of the CAPE-V when used by experienced voice clinicians judging normal and disordered voices.

Method

First of all, we made the fonetic adaptation of the protocol. It is not only a linguistic translation but also a fonetic traslation to spanish language.

We have 89 record voices: 30 men and 36 women with over 46 medium age (21-80 range). In 50 cases there is a vocal pathology from all wide range of ethiologies and severities, and 16 there are control normal cases. The validity was examined in several ways. First we compared judgements made by two experienced raters of 66 disordered and 16 normal voices, using the CAPE-V and GRABS scales. Second, we compared our main rater results to stablish intrarater reliability. After, we analysed agreement level between both mentioned scales. We made a subgroup of 11 pacients treated by surgical correction of the pathology and compare results of perceptual analysis of the voice between before and after.

Results:

Interrater Reliability Global Severity (SG): 0,935; Breathiness (CA): 0,894; Roughness (ASP): 0,907 (Task 1).

Intrarrater Reliability SG: 0,988; CA: 0,981; ASP: 0,979 (Task 1).

Agreement level between CAPE-V and GRABS is SG: 0,965; CA: 0,961; ASP: 0,950 (Task 1).

Sensitivity to Change: there are differences between presurgical evaluation and postsurgical but only with CAPE-V protocol.

Conclusions

This study provides evidence of its empirical validity, which justifies the use of the CAPE-V in clinical practice, training programs, and professional development activities.

SR. DIRECTOR DE DEPARTAMENTO DE CIRUGIA Y ESPECIALIDADES MÉDICO QUIRÚRGICAS/

SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO EN CIRUGÍA Y ESPECIALIDADES MÉDICO QUIRÚRGICAS

AGRADECIMIENTOS

En primer lugar, agradecer al Dr. Nuñez Batalla su inestimable colaboración, sin la cuál probablemente este trabajo no hubiese visto la luz. Gracias por introducirme en el estudio de la voz y por motivarme en cada desaliento del camino.

Gracias al personal de la Unidad de Hipoacusia, por su cariño y su apoyo.

Gracias a amigos y compañeros residentes que siempre han estado ahí. Especialmente a mi familia, siempre están detrás de cualquiera de mis logros y por ello, este trabajo también es suyo.

ABREVIATURAS

CAPE – V: *Consensus Auditory-Perceptual Evaluation of Voice.*

ASHA: *American Speech-Language Hearing Association.*

F0: Frecuencia fundamental.

G: *Grade* o Grado.

R: *Roughness* o Aspereza.

B: *Breathiness* o calidad aérea.

A: *Asthenia* o asthenia.

S: *Strain* o tensión.

D: Diplofonía.

PQ: Cociente de fonación.

CV: Capacidad vital.

TMF: Tiempo máximo de fonación.

HNR: Relación armónico-ruido.

TACV: traducción, adaptación cultural y validación.

GS: “*Gold standard*”.

T1: Tarea 1 (CAPE-V)

T2: Tarea 2 (CAPE-V)

T3: Tarea 3 (CAPE-V)

ICC o CCI: Índice de correlación intraclase o *Intraclass correlation index*

O1: Observador 1

O2: Observador 2

INDICE DE CONTENIDOS

Capítulo 1. Introducción	
Definición de Disfonía	6
Voz Normal	6
Clasificación de disfonías	7
Evaluación y diagnóstico de la disfonía	7
Protocolo básico de valoración de la disfonía	7
Análisis perceptual de la voz	17
Calificación auditivo perceptual de la voz	19
Sensación psicoacústica de la voz	20
Práctica actual	22
<i>Consensus Auditory-Perceptual Evaluation of Voice</i> : Consenso auditivo- perceptual de la voz (CAPE-V)	23
Adaptación cultural y validación de cuestionarios de salud	27
Primera etapa: Traducción	28
Segunda etapa: Validación del cuestionario en el idioma de destino	28
Capítulo 2. Justificación de la investigación y Objetivos	32
Capítulo 3. Material y Métodos	34
Adaptación al español del CAPE-V	35
Participantes	36
Muestras vocales	37
Procedimiento de grabación	37
Metodología de aplicación del CAPE-V	37
Evaluación auditivo-perceptual de las muestras vocales (GRABS y CAPE-V)	38
Capítulo 4. Resultados	
Datos descriptivos de la muestra	51
Fiabilidad Test- retest o validez intraobservador	60
Fiabilidad o Validez interobservador	69

Validez de criterio	82
Diferencia de tareas	100
Sensibilidad al cambio	105
Capítulo 5. Discusión	111
Conclusiones	126
Anexo A: Informe de valoración CAPE-V adaptado al español	127
Anexo B: Informe de valoración CAPE-V original	128
Bibliografía	129

Capítulo 1

INTRODUCCION

El método de evaluación perceptual de la voz denominado CAPE-V (*Consensus Auditory-Perceptual Evaluation of Voice*) es el resultado del esfuerzo de la *American Speech-Language-Hearing Association* (ASHA) para crear un protocolo clínico que pueda ser utilizado en la realización de juicios auditivo-perceptuales de la calidad vocal de una forma estandarizada. Los parámetros evaluados, las tareas vocales y los procedimientos que conforman el CAPE-V representan las recomendaciones consensuadas por expertos en la percepción humana, científicos dedicados al estudio del habla y de la voz, y fonoaudiólogos especializados en los trastornos vocales. No existe un método ideal para obtener una calificación precisa y válida de la calidad vocal, éste método deriva del más actualizado conocimiento sobre los factores multidimensionales en los que se basa la medición psicofísica de la percepción humana.

Los autores que describieron el CAPE-V esperan que sirva de soporte y estímulo para mejorar la práctica clínica en el campo de la evaluación auditivo-perceptual de la voz. Este trabajo se dirige a la adaptación de este método: CAPE-V al idioma español para después validarla clínicamente con el fin de que pueda ser aplicada a los pacientes hispanohablantes.

1. DEFINICIÓN DE DISFONÍA

Se considera que existe un trastorno de la voz cuando su calidad, tono, intensidad, o flexibilidad difiere de las voces de las demás personas del mismo sexo, edad y grupo cultural¹. Una voz que suena de forma anormal provoca que la mayoría de los adultos busquen atención médica. Aunque no tenga previamente una buena voz, el paciente se alarma por el cambio, dado que asume que estos cambios pueden indicar la presencia de una enfermedad laríngea.

1.1. Voz normal

Es más difícil definir lo que es una voz normal que cualquier otro componente del habla o del lenguaje debido a que, por naturaleza, la variedad de voces es ilimitada. Existe una variabilidad vocal muy amplia inherente a cada individuo y los requisitos que cumple una voz cuando es adecuada son muy variados y con un rango de variación muy amplio. Moore² describe la complejidad de la tarea: “es obvio que no

existe un único sonido que pueda denominarse voz normal, en su lugar existen voces de niños, voces de niñas, voces de mujeres, voces de hombres, voces de ancianos...” En cada uno de esos tipos de voces se puede reconocer la voz normal y la voz que no lo es. La localización del umbral que separa una de otra es juzgada por cada oyente según sus estándares culturales, de educación y su experiencia en la distinción de voces patológicas. Sin embargo, una vez localizada la frontera que separa lo adecuado y lo inadecuado, resulta obvio que cada observador ha adquirido los conceptos de normalidad y anormalidad. Es importante para los clínicos resaltar que las alteraciones vocales tienen una base cultural y son socialmente determinadas, lo que hace difícil determinar un umbral objetivo para calificar un timbre “normal”. Hasta el momento no existe una definición objetiva y aceptada por la comunidad científica de lo que es una voz normal.

Por tanto, solo se pueden establecer unos estándares generales que debe cumplir una voz normal³ y son los siguientes:

1. La calidad vocal debe ser agradable. Este criterio implica la presencia de cierta cualidad musical y la ausencia de ruido o atonalidad.
2. El tono debe ser adecuado. El nivel tonal debe ser apropiado para la edad y sexo de la persona que emite la voz.
3. El volumen debe ser apropiado. La voz no debe ser tan débil que no pueda ser escuchada bajo unas condiciones ordinarias donde se habla, ni debe ser tan intensa que llame la atención de forma indeseada.
4. La flexibilidad debe ser la adecuada. La variedad o flexibilidad se refiere a las variaciones en el tono y volumen que ayudan a la expresión de énfasis, significado o sutilezas que indican los sentimientos del individuo.

1.2. Clasificación de las disfonías

La alteración de la voz puede ser interpretada como un signo de enfermedad, como un síntoma de enfermedad, o como un mero trastorno de la comunicación:

La disfonía como signo de una enfermedad.

Ante un paciente cuya voz suene anormal, la principal preocupación del clínico consiste en determinar si la disfonía significa la presencia de una enfermedad. Las consideraciones estéticas o comunicativas son en ese momento secundarias, al ser

consciente que una disfonía puede ser el signo cardinal de una alteración laríngea, primera manifestación de una enfermedad grave, local o sistémica. Por tanto, la causa o causas del trastorno de la voz han de ser averiguadas si es posible. Una voz con calidad aérea que aparece de forma gradual y progresiva puede tener una significación menor desde el punto de vista estético, social y comunicativo, pero puede ser el primer signo de la aparición de un tumor cerebral o una enfermedad de Parkinson. Una vez que se conoce la causa de la disfonía y haya sido tratada, se toma en consideración la implicación comunicativa de la misma, con el fin de llevar a cabo las medidas rehabilitadoras que sean precisas.

La disfonía como síntoma de una enfermedad.

La palabra síntoma usada con propiedad, hace referencia a la queja con la que el paciente expone una sensación subjetiva, real o imaginaria. Podemos distinguir tres variaciones sobre este tema:

a. Que la voz sea percibida como anormal tanto por el médico como por el paciente con lo que se establece un acuerdo mutuo en cuanto a la necesidad de estudiar y tratar el problema.

b. Que el médico esté convencido de la necesidad de investigar y tratar un trastorno vocal, pero que el paciente no. La situación se origina tanto porque el primero sobre-estime la presunta alteración vocal, como que exista una indiferencia por parte del paciente a un problema real. En ambas, el diagnóstico y el tratamiento discurrirán en una franca o tácita resistencia, desinterés o incluso hostilidad por parte del paciente.

c. Que el paciente esté convencido de que su voz está alterada aunque el médico crea que el problema es trivial o que no existe. Este conflicto se asocia por lo general a una reacción exagerada por parte del paciente, y frecuentemente es una secuela del proceso de recuperación de una enfermedad o intervención quirúrgica laríngea. Estas reacciones son la expresión de hostilidad, perfeccionismo o de una depresión que requiere una asistencia psicológica.

La disfonía como trastorno de la comunicación.

Aunque la disfonía pueda ser el indicador de salud o enfermedad, la voz también puede ser evaluada como un instrumento de comunicación. Desde este punto de vista hay que tener en cuenta una serie de consideraciones: por una parte hay que preguntar si la voz es o no adecuada para lograr la inteligibilidad del discurso, si sus características estéticas son aceptables, y si satisface los requerimientos sociales y laborales del paciente.

Por tanto, la voz tiene una importancia personal, social y económica. Conforme se asciende en la escala socioeconómica, aumenta la importancia de poseer una voz efectiva y agradable. Con pocas excepciones, cuanto mayor dependencia social y profesional se tiene de la voz, son más devastadores los efectos de una disfonía.

En la actualidad no existe una nomenclatura estándar de los trastornos de la voz ni de la patología de las cuerdas vocales. La presentación de una fotografía de una determinada lesión vocal en un foro científico suele provocar controversia en la audiencia acerca del nombre “correcto” de la patología. Esto provoca una deficiente comunicación entre los profesionales y los pacientes dificultando la elaboración de protocolos sobre los tratamientos más adecuados o meta-análisis de los estudios publicados. Se han propuesto multitud de sistemas de clasificación y nomenclatura, por lo general con inadecuadas definiciones y clasificación fisiopatológica.

Los libros de texto tradicionales han tratado de clasificar los trastornos vocales como orgánicos o funcionales, o en relación con el cierre o falta de cierre de las cuerdas vocales. Esas categorías descriptivas carecen de referencia a consideraciones fisiopatológicas. Una clasificación más amplia debería cumplir las siguientes condiciones:

1. Ser entendida por los miembros de todas las disciplinas involucradas en los trastornos de la voz (investigadores, médicos, foniatras, maestros de canto, logopedas).
2. Informar de la presencia o ausencia de una lesión.
3. Ser utilizada consistentemente por los profesionales involucrados en el cuidado de la voz.
4. Ser objetiva.

El sistema de nomenclatura debería no precisar de equipos diagnósticos complejos y debería tener un impacto directo en el tratamiento y resultados de los trastornos de la voz. En la actualidad no existe tal sistema de clasificación, aunque existen propuestas que pueden ayudar a identificar los distintos trastornos vocales y las distintas enfermedades de las cuerdas vocales.

Una clasificación propuesta y publicada por Rosen et al., se basa en la evaluación visual y auditivo-perceptual, en los tratamientos de los distintos problemas y está cimentada en la anatomía y en la fisiología⁴. La clasificación puede ser modificada y ampliada según aparezcan nuevos conocimientos. Este sistema divide los trastornos de la voz en cuatro grandes categorías: trastornos no-orgánicos, problemas orgánicos, trastornos del movimiento (alteraciones del control neuromuscular), y enfermedades sistémicas que afectan al aparato vocal.

1. Trastornos no-orgánicos de la voz.

En estos trastornos la morfología y el movimiento de las cuerdas vocales son normales. Los trastornos que encajan en esta categoría son: la disfonía por tensión muscular, la disfonía funcional, la disfonía por conversión y la disfonía psicógena. Estas enfermedades presentan una kinesiólogía y biomecánica alteradas de la laringe en general y de las cuerdas vocales en particular. Otros trastornos pueden involucrar el apropiado control de la interacción entre el flujo aéreo subglótico con la posición de las cuerdas. Estos trastornos tienen una variedad de orígenes, desde psicológico hasta idiopático, pero el elemento común es una disfonía asociada con una anatomía y motilidad normal de las cuerdas vocales. En estos trastornos no-orgánicos de la voz el problema que subyace es la coordinación del trabajo respiratorio con la posición y tensión vocal, y se propone denominarlos “trastornos funcionales de la voz” con distintas subcategorías^{5,6}. Esta clasificación parece razonable porque la anatomía y fisiología se piensa que son normales, pero la habilidad para usar el mecanismo vocal para la producción de una voz normal no es funcional.

2. Trastornos orgánicos de la voz.

Incluye cambios patológicos tangibles de la laringe en general y de las cuerdas vocales en particular. Los trastornos orgánicos de la voz pueden ser clasificados basándose en la estructura estratificada de las cuerdas vocales descrita por Hirano: epitelio, lámina propia y músculo vocal⁷.

3. Trastornos del movimiento.

Las anomalías del control neuromuscular causan alteraciones de los movimientos laríngeos. Las enfermedades que se incluyen en esta categoría son: la parálisis vocal unilateral, la paresia vocal, el movimiento paradójico de las cuerdas vocales, la parálisis vocal bilateral, la disfonía espasmódica (abductora, adductora y mixta) y el temblor esencial de la laringe.

4. Enfermedades sistémicas que afectan al aparato vocal.

Con frecuencia, las enfermedades sistémicas pueden alterar la producción vocal al afectar al aparato fonador. Estas enfermedades incluyen el reflujo faringo-laríngeo, que es una manifestación de la enfermedad por reflujo gastro-esofágico, infecciones de la laringe y enfermedades neurológicas como la enfermedad de Parkinson.

2. EVALUACIÓN Y DIAGNÓSTICO DE LA DISFONIA

2.1. Protocolo básico de valoración de la disfonía.

La naturaleza multidimensional de la función vocal obliga a estudiarla por medio de distintos métodos. Esta diversidad en el estudio hace difícil su aplicación en la práctica clínica y entorpece la difusión del análisis acústico de la voz. Es por ello por lo que se precisa de un conjunto mínimo de características en la presentación de los resultados con el fin de permitir comparaciones y meta-análisis.

El objetivo principal del Comité de Foniatría de la *European Laryngological Society* (ELS) es el de difundir y adaptar un protocolo básico con el fin de normalizar los procedimientos de estudio de la disfonía. El propósito de este protocolo básico es el de uniformar la metodología para la valoración funcional de la patología vocal, y permitir así comparaciones relevantes en la literatura a la hora de presentar o publicar los resultados de un tratamiento vocal, o de un nuevo procedimiento de estudio.

La ELS ha propuesto un conjunto de parámetros básicos para la valoración de las disfonías, entre los que deben incluirse: (a) percepción; (b) endoscopia y videoestroboscopia; (c) parámetros acústicos; (d) aerodinámica y eficiencia; y (e)

autoevaluación por el paciente. Cada uno de estos parámetros tiene una relevancia específica aportando en conjunto una visión multidimensional.

Existe un reducido grupo de trastornos muy concretos que necesitan de un protocolo específico, con el fin de conseguir una evaluación mejor adaptada e incrementar su sensibilidad. Es el caso de disfonías espasmódicas y voces de sustitución (las no generadas por las cuerdas vocales) en su estudio es necesario incluir características como: la inteligibilidad y la fluidez. El objetivo final de la adaptación del CAPE-V como nuevo protocolo práctico de evaluación es el de alcanzar acuerdo y uniformidad en cuanto a la metodología para la valoración funcional de la patología vocal, incorporando los principios que recoge el protocolo publicado por el Comité de Foniatría de la *European Laryngological Society* (ELS).

Prerrequisito: la grabación de la señal vocal

La grabación de audio es la herramienta básica más valiosa de la valoración de la voz. Una vez obtenida una grabación de calidad, puede ser almacenada permitiendo posteriores investigaciones sobre ella. Es esencial conservar las grabaciones como archivos con el fin de poder recuperarlos con facilidad. Los sistemas de análisis acústico comercialmente disponibles graban directamente la señal, digitalizan y guardan apropiadamente en los ordenadores donde están instalados. Se recomienda configurarlos con el fin de que las grabaciones tengan una frecuencia de muestreo de 20.000 Hz. Se deben llevar a cabo de forma ideal en una habitación sonoamortiguada o en su defecto conseguir un ruido ambiente menor de 50 dB. La distancia de la boca al micrófono ha de ser constante a unos 10 cm, y éste debe colocarse en ángulo de 45 a 90° para reducir el ruido aerodinámico de la boca durante el habla.

Muestra a grabar

Una grabación estándar debe consistir en:

- /a:/ o /e:/ en un tono e intensidad cómodos, grabado en tres intentos con el fin de evaluar si existe variabilidad en la calidad vocal.
- /a:/ o /e:/ ligeramente más intensa, con el fin de evaluar posibles cambios en su calidad.

- Una frase sencilla o un corto pasaje literario estándar.

Para la selección fonética de la frase o pasaje hay que tener en cuenta que contenga una vocalización constante y esté libre de fricativas para no sesgar el cálculo de la relación armónico-ruido. Un pasaje para su lectura y grabación podría ser el siguiente:

“Platero es pequeño, peludo, suave; tan blando por fuera, que se diría todo de algodón, que no lleva huesos. Sólo los espejos de azabache de sus ojos son duros cual dos escarabajos de cristal negro”.

Fragmento de la obra *Platero y yo*, Juan Ramón Jiménez

Percepción de disfonía y parámetros clásicos

Se propone usar el término “disfonía” para designar cualquier clase de patología vocal percibida: la alteración puede referirse al tono o intensidad, así como al timbre o características rítmicas o prosódicas. El término “ronquera” se recomienda limitarlo para designar las alteraciones de la calidad vocal o timbre, y excluye el tono, intensidad y ritmo. La calificación de la ronquera se realiza a partir del habla conversacional escuchada durante la anamnesis del paciente, la severidad de la misma se cuantifica bajo el parámetro G (*Grade* o Grado) de la escala GRABS propuesta por Hirano en 1981 y califica la calidad vocal global, integrando todos los componentes alterados.

Se han identificado dos componentes principales de la ronquera: la calidad aérea o *Breathiness* (B) que es la impresión audible de la pérdida de aire turbulento a través de una glotis insuficientemente cerrada y la aspereza o *Roughness* (R) que es la impresión audible de pulsos glóticos irregulares, fluctuaciones anormales en la frecuencia fundamental o impulsos percibidos por separado (*vocal fry*), incluye la diplofonía y las roturas de voz. Cuando está presente la diplofonía debe registrarse como “D”. Estos parámetros han mostrado la suficiente fiabilidad y reproductibilidad interobservador e intraobservador en la clínica.

Los parámetros que describen el comportamiento vocal “astenia” y “tensión” (*asthenia, strain: A y S*) son considerados como menos fiables en la actualidad y han sido omitidos del protocolo básico.

Para la calificación, es frecuente en la escala GRABS el empleo de cuatro rangos de calificación (0 = normal o ausencia de alteración, 1 = ligeramente alterado, 2 =

moderadamente alterado y 3 = severamente alterado), sin embargo existen escalas con apoyo analógico visual o con recorridos más amplios (100 puntos).

Videolaringoestroboscopia.

La videolaringoestroboscopia es la principal herramienta clínica para el diagnóstico etiológico de los trastornos de la voz. Es utilizada para estudiar la calidad de la vibración vocal, por lo que también valora la efectividad de los tratamientos. La pertinencia de los parámetros estroboscópicos se basa en la combinación de fiabilidad (reproductividad inter e intraobservador), no redundancia y sentido clínico (relación con los conceptos fisiológicos). Los parámetros básicos son:

- **Cierre glótico:** gradación cuantitativa del espacio que existe entre ambas cuerdas vocales al cerrarse. Se recomienda que se especifique el tipo de defecto de cierre glótico (longitudinal, dorsal, ventral, irregular, oval o en reloj de arena).
- **Regularidad:** gradación cuantitativa del grado de irregularidad presente en la onda mucosa percibido con el estroboscopio.
- **Onda mucosa:** gradación cuantitativa de la onda mucosa, representación de la fisiología de la estructura en capas de las cuerdas.
- **Simetría:** gradación cuantitativa del movimiento “especular” de ambas cuerdas. Generalmente la asimetría está causada por limitaciones en la actividad vibratoria de una lesión (cicatriz difusa, quiste, leucoplasia, etc).

Cada parámetro estroboscópico debe puntuarse con una escala de cuatro puntos o analógico-visual. Esta exploración se puede documentar por medio de su grabación en video.

Se recomienda observar y grabar las imágenes estroboscópicas en diferentes modos de fonación (el grado del cierre glótico generalmente aumenta con el aumento de la intensidad), sin embargo, esta gradación básica se refiere a fonación en tono e intensidad cómodas. Para permitir comparaciones entre distintas exploraciones del mismo paciente, se recomienda utilizar el mismo tipo de endoscopio (rígido o flexible, si rígido: el mismo ángulo).

Aerodinámica.

El parámetro aerodinámico más simple de la fonación es el tiempo máximo de fonación (TMF) en segundos. Consiste en la prolongación de la vocal /a:/ lo más posible tras una inspiración máxima, en un tono e intensidad cómodos. Es una de las medidas clínicas de la valoración de la voz más utilizadas. Se precisa una demostración previa y se realizan tres intentos, seleccionando el más prolongado para compararlo con la norma. Para eliminar posibles sesgos debidos a una capacidad respiratoria de soporte del paciente, que compense un defecto de cierre glótico se puede recurrir a calcular la siguiente relación: (flujo aéreo promedio o Cociente de Fonación = PQ)

$$PQ = \frac{\text{Capacidad Vital}(ml)}{\text{TMF}(seg)}$$

La capacidad vital (CV) se define como el volumen de aire intercambiado en la boca entre la posición de inspiración máxima y la espiración completa. Se puede cuantificar de forma fiable usando un espirómetro de mano. La CV depende de factores antropométricos, especialmente del peso.

Acústica.

Los parámetros acústicos son medidas objetivas y no invasivas de la función vocal. Las medidas de perturbación (en periodo y amplitud: *jitter* y *shimmer*, respectivamente) así como los cálculos de la relación armónico-ruido (HNR) parecen ser las medidas muy robustas y determinan los elementos perceptuales básicos de la calidad vocal: grado, aspereza y calidad aérea. En general, pueden ser obtenidos mediante programas informáticos de forma sencilla y asequible. La limitación principal de estos sistemas es que los datos objetivos no han conseguido correlacionarse con la patología orgánica subyacente y no pueden utilizarse de forma fiable en todos los tipos de disfonía, puesto que las señales fuertemente aperiódicas no pueden analizarse. Las medidas de la perturbación, siempre que sean menores del 5% se tienen por fiables.

De esta manera, *jitter* y *shimmer* son los parámetros propuestos como medidas acústicas básicas, para analizar señales vocales obtenidas de la vocal /a:/ sostenida en intensidad y tono cómodos. El jitter es calculado como la diferencia media entre

periodos de ciclos vocales adyacentes dividida por el periodo medio. Es, por tanto, una medida derivada de la F0 (frecuencia fundamental de la voz). Para el *shimmer*, se realiza un cálculo similar sobre las amplitudes pico a pico. Obviamente, para permitir comparaciones pre y postratamiento se requieren similares técnicas y materiales.

También incluido en este apartado de medidas acústicas está el fonetograma. Descrito por Calvet en 1953, el fonetograma es la representación gráfica de la capacidad fonatoria de la laringe, lo que supone medir la intensidad o volumen en decibelios, tanto máximo como mínimo, que una persona es capaz de emitir y mantener al menos durante dos segundos en un tono determinado. Esta medida se realiza en todo el espectro de tonos que el paciente es capaz de producir, desde el más grave hasta el más agudo, pronunciando siempre una misma vocal, que puede ser la /a/, la /i/ o la /u/. También se conoce como «perfil del rango vocal». El perfil del rango vocal es un indicador sensible de la función vocal. El fonetograma está recomendado desde 1994 por el National Center for Voice and Speech de Estados Unidos, y se emplea como parte de la evaluación acústica y fonatoria de la voz patológica.

Autoevaluación por el paciente.

Esta evaluación de la voz, aunque subjetiva por definición, es de mucha importancia en la práctica clínica diaria. Es el paciente el que vive con su voz y está influenciado por los aspectos culturales y sociales que son relevantes para la voz en su medio. Esta evaluación necesita de una cuantificación cuidadosa y se precisa comparar y correlacionar con los datos de la valoración objetiva. El objetivo básico es el de diferenciar la calidad vocal alterada y la severidad de la incapacidad o minusvalía en la vida diaria social y profesional. Se puede calcular un índice de incapacidad vocal basándose en las respuestas del paciente a un conjunto de preguntas cuidadosamente elaborado: además de los aspectos ya mencionados, también investiga la posible repercusión emocional de la disfonía. Sin embargo, para el protocolo básico se recurre a una evaluación subjetiva mínima por parte del paciente sobre una doble escala analógico-visual de 10 cm: la impresión acerca de la calidad vocal y la impresión acerca de las repercusiones que la alteración vocal tiene en su vida diaria social y profesional. El extremo izquierdo de la escala o cero se relaciona con la voz normal y el extremo derecho o 10 con la voz extremadamente alterada⁸.

3. ANALISIS PERCEPTUAL DE LA VOZ

La recomendación del *National Center for Voice and Speech* (www.ncvs.org) es que la evaluación de la voz patológica debe comenzar con la realización de un análisis espectrográfico para clasificarla, con el fin de determinar los métodos de estudio idóneos en cada caso en particular.

Así, una voz con el suficiente grado de periodicidad (tipo 1) debería ser analizada por medio de los parámetros de perturbación citados previamente (jitter, shimmer y HNR). Si la voz presenta aperiodicidad, subarmónicos y roturas de voz (tipo 2) no puede ser analizada de forma fiable con los parámetros antes mencionados. Una perturbación mayor del 5% nos indica que estamos ante una voz de tipo 2, que debe ser estudiada por un método de clasificación perceptual como el GRABS y por un método visual como el espectrograma. Las voces caóticas (tipo 3) solo pueden ser estudiadas por el momento mediante métodos perceptuales, en tanto no se apliquen de forma práctica las dimensiones que estudian los fenómenos caóticos (dimensión fractal, exponente de Lyapunov, etc).

Por tanto, la evaluación perceptual es un método imprescindible en la valoración de la voz patológica, dado que es necesario aplicarla en un importante número de pacientes en los cuales las medidas de la perturbación a corto plazo no son fiables o no se pueden calcular⁹.

A pesar de que la percepción de la voz es la principal herramienta para juzgar si existe una alteración en la voz, los métodos que califican la calidad vocal nunca podrán llevarse a cabo con una perfecta validez y fiabilidad. Conceptualmente, el método ideal sería aquel que cumpla tres condiciones: 1) que distinga de una forma fiable las voces normales de las patológicas, siendo útil también para monitorizar los cambios en la calidad vocal del paciente a lo largo del tiempo, 2) que se correlacione con la fisiopatología y los parámetros acústicos objetivos y 3) que tenga claramente establecido el tipo y cuantía del entrenamiento requerido por el usuario y si se precisan patrones para el aprendizaje del método.

Los métodos de calificación perceptual de la voz tienen importantes inconvenientes: por una parte no existe una terminología estándar para describir o estadificar las

alteraciones de la calidad vocal y tampoco existe una definición estándar de lo que se considera como voz normal, por otra parte hay que tener en cuenta la falta de fiabilidad inherente de las valoraciones auditivo-perceptuales de la calidad vocal y la inevitable variabilidad en la producción vocal individual.

La mayoría del trabajo que se ha llevado a cabo en el estudio de la percepción humana de la voz se ha centrado en el tono y la intensidad, en cambio, se ha realizado menor esfuerzo en el estudio de la percepción de la calidad vocal, principalmente por su naturaleza multidimensional y por la dificultad de cuantificar este fenómeno relacionado con el timbre. Para optimizar la calificación perceptual de esta compleja característica vocal se recomienda el uso de escalas numéricas con al menos 15 sub-categorías o, alternativamente, el empleo de escalas analógico-visuales continuas y el uso de voces patrón que ayuden a localizar los extremos (normal y más severo) de la escala.

En la actualidad no existe un método que haya demostrado ser mejor que los demás en la tarea de medir la calidad vocal percibida. El conjunto de conocimientos actuales todavía es inadecuado para poder diseñar una herramienta clínica que resuelva todos los inconvenientes científicos que se presentan. En efecto, los esfuerzos para conseguirlo se han acompañado de un sinnúmero de problemas relacionados con la fiabilidad, utilidad y validez. Sin embargo, existen intentos de desarrollar protocolos con el fin de documentar las características auditivo-perceptuales de las alteraciones de la calidad vocal. La construcción de un procedimiento consistente y específico con un formato documental ha servido para dotar a estos métodos de consistencia clínica y mejora de comunicación entre los clínicos.

Los estudios llevados a cabo hasta la fecha no sostienen la sustitución de las medidas instrumentales por la valoración auditivo-perceptual y esta, para tener consideración clínica, ha de seguir un procedimiento estándar. Los dos métodos más empleados en la actualidad que proporcionan un protocolo estándar son el método GRABS y el CAPE-V. Actualmente este último solo se encuentra disponible en inglés¹⁰.

La sensación psicoacústica de la voz es la principal herramienta en la evaluación de los trastornos vocales, tanto para los pacientes cuando explican el cambio que notan en su voz, como para los clínicos a la hora de describirlos^{11,12,13,14} Los especialistas en

fonología describen las alteraciones vocales mediante sus propios juicios perceptuales, basados en su experiencia, en aspectos culturales, sociales y asociados al medio en el que trabajan. Sin embargo, las calificaciones auditivo perceptuales de la voz nunca se llevarán a cabo con perfecta validez ni fiabilidad, porque se tratan de métodos subjetivos^{15,16,17,18,19}.

Es un tema controvertido porque los evaluadores no tienen un entrenamiento específico y además existe una amplia variabilidad en la correlación inter e intra-evaluadores. Se ha demostrado que los observadores experimentados juzgan de forma más consistente que los inexpertos y usan estrategias perceptuales distintas, determinándose que 8 horas de entrenamiento en observadores previamente inexpertos son suficientes para obtener un 80% de fiabilidad interobservador⁹. Por eso existe la clara necesidad de crear un protocolo para realizar estas evaluaciones de una forma más teórica, clínicamente significativa y consistente.

3.1. Calificación Auditivo-Perceptual de la voz

Hasta ahora el juicio perceptual ha sido una tarea incierta, marcada por desacuerdos entre los especialistas en fonología y por la variabilidad en la recogida y exposición de los datos. En la actualidad no existe un método que haya demostrado ser mejor que los demás en la tarea de medir la calidad vocal percibida. El conjunto de conocimientos actuales es todavía inadecuado para poder diseñar una herramienta clínica que resuelva todos los temas científicos relevantes.

Sin embargo, en los últimos años existe un creciente interés en el desarrollo de procedimientos normalizados y, un buen ejemplo, es la reunión de consenso celebrada en Pittsburg en 2002 y organizada por la American Speech-Language-Hearing Association's Division 3 donde se sentaron las bases de la versión inicial del CAPE-V²⁰, cuya validación al idioma español es el objetivo de este estudio.

En este contexto existe acuerdo en los siguientes principios orientadores:

1. Las dimensiones perceptuales deberían reflejar un conjunto mínimo de parámetros perceptuales con significado clínico.
2. Los procedimientos y resultados deberían de ser fácilmente obtenidos.

3. Los procedimientos y resultados deberían ser aplicables a un amplio rango de patologías vocales y situaciones clínicas variadas.
4. Las puntuaciones deberían presentar una fiabilidad intra e interobservador en los ulteriores estudios de validación.
5. Se deberían seleccionar voces con el fin de considerar un futuro uso como patrones y posible entrenamiento.

Los estudios llevados a cabo hasta la fecha no sostienen la sustitución de las medidas instrumentales por la valoración auditivo-perceptual. Y paradójicamente, para tener consideración clínica, la valoración perceptual ha de seguir un procedimiento estándar como el que se propone en el consenso que describe el método CAPE-V. En esta tendencia a la normalización, otra de las tareas más importantes es consensuar el tipo de escala que debe utilizarse para optimizar las evaluaciones^{17,21,22}

3.2. Sensación Psicoacústica De La Voz

La percepción psicoacústica del sonido se basa en los mecanismos de la percepción humana y en la discriminación del sonido. Durante este proceso el sonido sufre modificaciones a su paso a través del oído medio y externo, como se manifiesta en la curva de discriminación acústica: la amplitud de frecuencias procesadas por nuestro oído es el resultado de la morfología y biomecánica coclear, y además, la integración de la información acústica se realiza a niveles superiores, en la corteza cerebral. Las investigaciones sobre los mecanismos de la percepción humana y la discriminación acústica han logrado avances en cuanto a la percepción del tono y de la intensidad, permitiendo la construcción de métodos de escala para estos fenómenos. Sin embargo, se observa que se ha invertido mucho menos esfuerzo en el estudio de la percepción de la calidad del sonido o timbre, fundamentalmente porque es un fenómeno multidimensional, complejo, y muy difícil de cuantificar objetivamente. Dos sonidos que tienen el mismo tono e intensidad pueden diferenciarse por el timbre. A modo de ejemplo, citaremos el trabajo de Zwicker et al.²³, que intenta explorar la calidad del sonido relacionada con conceptos tales como; nitidez, agrado, fluctuación, fuerza y dureza.

Autores conocidos y pioneros en psicoacústica²⁴ han investigado cómo es posible que los sistemas sensoriales sean capaces de diferenciar cambios mínimos en el umbral

del sistema neurosensorial. Examinando los métodos del análisis psicofísico han concluido que éste ha de tener en cuenta: (1) la definición de estímulo (¿Qué propiedades del estímulo son importantes para la percepción?); (2) la definición de experiencia perceptual (¿Cuáles son las características o dimensiones de la percepción y cómo se interrelacionan entre ellos?) y (3) la explicación del proceso que relaciona percepción y estímulo (Proceso sensorial, de toma de decisión, proceso cognitivo: codificación, transformación, grabación, etc).

Es importante marcar la diferencia entre métodos de escala directos e indirectos. Los métodos directos se consideran más apropiados que los indirectos para aplicaciones clínicas, ya que utilizan procedimientos de medida porcentuales o por intervalos. Para optimizar los métodos directos debe tenerse en cuenta el rango actual y la distribución del estímulo que está siendo analizado, si se utiliza un estímulo estándar, si se afecta por fenómenos secuenciales, o si el entrenamiento mejora la estratificación^{24,25}. Se establecen dos recomendaciones específicas respecto al desarrollo de un instrumento clínico en la evaluación perceptual de la calidad vocal: primero, se recomienda la utilización de rangos de escala numéricos con al menos 15 divisiones, o en su defecto, el uso de escalas gráficas o analógicas visuales; segundo, la localización de los puntos de referencia (por ejemplo; normal o muy severo) han de dejar un espacio extra al final de la escala para evitar los “efectos de fin de escala”.

Entre los métodos de evaluación auditivo-perceptual de los trastornos vocales que se usan en la actualidad, como son el Análisis del Perfil Vocal de Laver²⁶, el Abordaje para la Evaluación de la Voz de Estocolmo²⁷ y el método GRABS¹², el más extendido es éste último a pesar de importantes limitaciones que limitan su fiabilidad²⁸.

3.3. Práctica actual

Como ya se ha dicho no existe un método que haya demostrado ser mejor que los demás en la tarea de medir la calidad vocal percibida. El conjunto de conocimientos actuales todavía es inadecuado para poder diseñar una herramienta clínica que resuelva todos los temas científicos relevantes. Los estudios llevados a cabo hasta la fecha no sostienen la sustitución de las medidas instrumentales por la valoración auditivo-perceptual. Para tener consideración, sin embargo, la valoración perceptual ha de seguir un procedimiento estándar. Los dos métodos más empleados en la

actualidad que proporcionan un protocolo estándar son el método GRABS y el CAPE-V.

El método GRABS es el que en la actualidad se está empleando de forma generalizada, fue diseñado por la Sociedad Japonesa de Logopedas y Foniatras y divulgado gracias a su publicación por Hirano en su libro *Clinical Examination of the Voice*¹². Este protocolo puntúa en una escala, cinco parámetros o ítems que se recogen en el acrónimo que se usa para su denominación: el Grado (*Grade*), califica la calidad vocal global, integrando todos los componentes alterados; la Aspreza (*Roughness*), que es la impresión audible de pulsos glóticos irregulares, fluctuaciones anormales en la F0 o impulsos percibidos por separado, incluye la diplofonía y las roturas de voz; el Escape aéreo (*Breathiness*), que es la impresión audible de la pérdida de aire turbulento a través de una glotis insuficientemente cerrada e incluye cortos momentos áfonos; y la Astenia (*Asthenicity*) y la Tensión (*Strain*), que describen el comportamiento vocal y son considerados como menos fiables en la actualidad. La calificación se lleva a cabo escuchando el habla conversacional corriente durante la anamnesis, puntuando con un 0 si el parámetro se juzga como normal, con un 1 si se considera levemente afectado, con un 2 si es moderado y con un 3 si hay una severa desviación de la normalidad. Los parámetros que han mostrado una suficiente fiabilidad en los estudios de análisis de componentes principales²⁹ son el parámetro B y el R, que presentan una suficiente fiabilidad (reproducibilidad intra e inter-observador) para ser usados en la clínica. Como los parámetros de comportamiento (A y S) tienen menos fiabilidad, surgió una escala simplificada, GRB, que es similar a la escala RBH (*Rauhigkeit* para aspreza, *Behauchtheit* para escape aéreo y *Heiserkeit* para la ronquera), usada en los países germánicos³⁰.

3.4. Consensus Auditory-Perceptual Evaluation of Voice: Consenso de Evaluación Auditivo-Perceptual de la Voz (CAPE-V)

En Junio de 2002 tiene lugar la conferencia de consenso de la evaluación auditivo perceptual de la voz, organizada por la *American Speech–Language Hearing Association* (ASHA: Division 3; Patología Vocal) y la Universidad de Pittsburg.

El objetivo de esta reunión fue diseñar y promover la utilización de un protocolo estandarizado para documentar las calificaciones auditivo perceptuales de las

patologías vocales, y que finalmente se materializó en el protocolo denominado CAPE-V.

Los asistentes a esta conferencia, un grupo internacional de expertos en voz, lenguaje y percepción humana, concluyen que no existe mejor forma de evaluación para la calificación vocal que la elaboración de un protocolo estandarizado. El conocimiento actual es insuficiente para diseñar herramientas clínicas con relevancia científica en la medición psicofísica y la estratificación de los resultados, además los intentos previos han incurrido en errores de validez, utilidad y consistencia.

Estos problemas también estuvieron presentes a la hora de desarrollar el CAPE V. Sin embargo, la colaboración de los especialistas, incorporando múltiples perspectivas y documentando las bases perceptuales del procedimiento han permitido construir un protocolo de evaluación con un formato concreto y acompañado de un conjunto de procedimientos específicos a seguir. Esto último constituye la característica principal de este protocolo, ya que ningún otro antes incorpora información sobre la recogida de muestras y este hecho parece relevante, ya que mejora la comunicación y coherencia entre profesionales.

En este contexto, los autores están de acuerdo en los siguientes principios:

- Las dimensiones perceptuales deberían reflejarse en un conjunto de parámetros perceptuales que resulten significativos.
- Los procedimientos realizados y los resultados deben ser obtenidos convenientemente.
- Los procedimientos y resultados deben ser aplicables a un amplio rango de patologías vocales.
- Los juicios deben optimizarse en cuanto a consistencia entre observadores a través de estudios de validación.
- Por último, deben considerarse algunos casos para utilizarlos como ejemplos clave, para establecer límites dentro de los rangos patológicos e incluso en el entrenamiento.

Elementos Específicos del CAPE – V

El CAPE-V califica importantes atributos vocales que han sido identificados por los miembros del grupo de trabajo del consenso por ser comúnmente utilizados, de reconocida importancia y fácilmente entendidos. Los atributos son: (a) Severidad global; (b) Aspereza; (c) Voz aérea; (d) Tensión; (e) Tono; (f) intensidad. El CAPE-V exhibe cada atributo usando una línea de 100 mm para formar una escala analógico-visual, que a su vez se encuentra subdividida en 3 intervalos específicos correspondientes a las categorías: leve, medio y grave. El observador indica el grado de desviación percibida respecto de la normalidad en esta escala marcando sobre la línea horizontal. Las calificaciones se establecen por observación directa de la emisión vocal por parte del paciente y no por lo que el paciente refiere ni por otros medios. A la derecha de la escala aparecen dos letras: la C y la I para referirse a si la desviación de la normalidad del atributo a calificar es, continua o bien es intermitente.

Descripción de los atributos

Severidad global: Impresión global de desviación de la normalidad en la voz.

Aspereza: Irregularidad percibida en la fuente vocal.

Voz aérea: Escape audible de aire en la fonación.

Tensión: Percepción de excesivo esfuerzo en la vocalización (hiperfunción).

Tono: Correlato perceptual de la frecuencia fundamental de la voz. La escala documenta cuánto se desvía el paciente del tono apropiado para su edad, sexo y referentes culturales. La dirección de la desviación (alto o bajo) debe ser anotada.

Intensidad: Correlato perceptual de la intensidad de la vibración vocal.

Escalas en blanco para atributos adicionales

Los seis atributos arriba descritos han de ser considerados como los atributos mínimos que han de ser valorados en este método. Adicionalmente, se incluyen dos escalas sin etiquetar para que puedan ser registrados aquellos atributos que sobresalgan en la escucha de la voz y que no estén incluidos en los anteriores. Por ejemplo, si un individuo muestra una afonía, ésta debe ser registrada en una escala en blanco, dejando las demás sin anotar.

Grabación

El paciente ha de estar cómodamente sentado en un ambiente silencioso para proceder a la grabación de las tres tareas vocales que el paciente ha de completar: una vocal aislada, la repetición de frases y una muestra de habla espontánea. Han de cumplirse los siguientes procedimientos estándar durante la grabación: usar un micrófono de condensación situado en un azimut de 45° frente a la boca y a 4 cm de distancia. Se recomienda utilizar un ordenador con 16 bits de resolución y una frecuencia de muestreo no menor de 20 kHz.

Tareas vocales a realizar

Tarea 1: Vocales sostenidas

Se utilizan dos vocales para esta tarea. Una vocal laxa o abierta como la /a/ y una tensa como la /i/. Ha de instruirse al paciente para que produzca una vocalización cómoda, con su voz normal y prolongarla hasta que se le indique. Puede ser preciso mostrar un ejemplo. Se lleva a cabo en tres ocasiones y durante 3.5 segundos cada una de ellas. Posteriormente se repite la tarea con la vocal /i/. Esta tarea da la oportunidad de escuchar la voz del paciente sin influencia de la articulación. Estas vocales además pueden ser analizadas acústicamente.

Tarea 2: Frases

Se han diseñado seis oraciones para provocar distintos comportamientos de la laringe y signos clínicos. Las oraciones originales en idioma inglés son: (a) *The blue spot is on the key again*; (b) *How hard did he hit him*; (c) *We were away a year ago*; (d) *We eat eggs every Easter*; (e) *My mama makes lemon jam*; (f) *Peter will keep at the peak*.

La primera oración produce todos los sonidos vocálicos del idioma inglés, es una frase que es útil para estudiar las influencias coarticulatorias de las tres vocales /a, i, u/, la segunda hace hincapié en el golpe glótico suave al iniciarse las palabras con la /h/, la tercera oración es totalmente sonora y nos ofrece el contexto para juzgar la presencia de espasmos o interrupciones vocales y la habilidad que tiene el sujeto para enlazar las palabras manteniendo la vocalización, la cuarta provoca el ataque glótico duro, la quinta incorpora sonidos nasales, lo que sirve para detectar una posible hiponasalidad; y la última oración se equilibra con consonantes explosivas sordas lo

que da la oportunidad para valorar la presión intraoral y la posible hipernasalidad y emisión aérea nasal.

Se debe entregar al paciente una tarjeta por cada oración para que la lea y la pronuncie. En caso de que tenga dificultades para la lectura ha de repetir la frase que se le diga.

Tarea 3: Habla espontánea

Se debe grabar al menos 20 segundos de conversación natural utilizando preguntas estándar de entrevista tales como “Hábleme sobre su problema vocal” o “Dígame cómo percibe su voz”.

Puntuación

Antes de comenzar el proceso de calificación debemos disponer de la grabación completa de las tres tareas descritas. Si se juzga que la calidad de todas las tareas es uniforme se procede a calificar solamente una escala. Sin embargo, si se perciben discrepancias entre las distintas tareas se debe calificar cada una por separado. Se debe cubrir solamente un formulario por cada paciente, numerando las tareas si han de ser calificadas de forma distinta (1 para la vocal sostenida, 2 para las frases y 3 para la conversación natural). Una vez calificadas las escalas han de ser medidas de forma física desde la izquierda hasta el punto marcado, indicando los milímetros resultantes referidos al total (por ejemplo 70/100 mm) y también haciendo mención a las etiquetas descriptivas que correspondan al grado de desviación general, como por ejemplo; desviación media o desviación grave. Se recomienda el empleo de ambas formas de calificación. El anexo A recoge el documento recomendado para la puntuación de las tres tareas que componen el CAPE-V y el anexo B incluye el formulario original del CAPE-V en inglés.

Se recomienda que para sucesivas calificaciones del mismo paciente, se disponga de la calificación CAPE-V previa con el propósito de compararlas. Esta estrategia de comparación otorga al método una mejor consistencia interna y fiabilidad de las sucesivas puntuaciones del paciente, especialmente cuando se pretende valorar los resultados terapéuticos.

4. ADAPTACION CULTURAL Y VALIDACIÓN DE CUESTIONARIOS DE SALUD

La traducción, adaptación cultural y validación (TACV) de cuestionarios para su uso en otros idiomas es un proceso laborioso que, sin embargo, cuando se lleva a cabo de forma sistemática permite obtener una herramienta de medición equivalente a su versión original. El método CAPE-V precisa de la realización de estas tareas ya que se altera de forma importante su contenido en el proceso de adaptación a un idioma distinto al que fue descrito. Es importante seguir las recomendaciones metodológicas publicadas, puesto que si el proceso de TACV no se lleva a cabo de manera rigurosa, pueden producirse errores.

El uso de herramientas no equivalentes al cuestionario original puede producir resultados no fiables o confusos que podrían limitar el intercambio de información entre la comunidad científica^{31,32}.

El proceso de traducción y adaptación debe ir seguido de un proceso de validación en la lengua de destino, lo cual permite minimizar el sesgo de información que podría asociarse a la administración de cuestionarios en países con idiomas y culturas diferentes. Por ello, se proponen una serie de pasos a seguir durante la etapa de adaptación y validación del CAPE-V al español.

4.1. Primera etapa: Traducción

En esta etapa se traduce la herramienta partiendo de su versión original y procurando mantener la estructura del cuestionario. El objetivo es conseguir que el instrumento resultante mantenga la equivalencia semántica, idiomática y conceptual con el cuestionario original^{33,34}. Todo el cuestionario, incluyendo las instrucciones, los ítems y las opciones de respuesta, se traducirá utilizando este método, recopilando todo en un informe.

4.2. Segunda etapa: Validación del cuestionario en el idioma de destino

La correcta traducción y adaptación cultural de un cuestionario no siempre garantiza la preservación de sus propiedades psicométricas, por lo que es necesaria su validación en el idioma de destino³³.

Para que un cuestionario se considere válido, debe reunir las siguientes características: a) ser fiable y capaz de medir sin error; b) ser capaz de detectar y medir cambios, tanto entre individuos como en la respuesta de un mismo individuo a través del tiempo; c) ser sencillo, viable y aceptado por pacientes, usuarios e investigadores; d) ser adecuado para medir el fenómeno que se pretende medir, y e) reflejar la teoría subyacente en el fenómeno o concepto que se quiere medir.

Todas estas características están relacionadas con dos propiedades de los cuestionarios: la fiabilidad y la validez³³

Para validar los cuestionarios se recomienda la siguiente secuencia:

1. Fiabilidad

Es el grado en que un instrumento es capaz de medir sin errores. Mide la proporción de variación en las mediciones que es debida a la diversidad de valores que adopta la variable y no al posible error sistemático o aleatorio³³.

La fiabilidad determina la proporción de la variancia total atribuible a diferencias verdaderas entre los sujetos^{35,36}. Para demostrar la fiabilidad del protocolo CAPE-V en español debemos evaluar sus dos dimensiones: la fiabilidad intraobservador o fiabilidad test-retest y la fiabilidad interobservador.

a) Fiabilidad intraobservador o fiabilidad test-retest: este aspecto hace referencia a la consistencia del instrumento, cuando se administra el mismo método a la misma población en dos momentos diferentes. El tiempo que debe transcurrir entre la primera vez (test) y la segunda (retest) dependerá del fenómeno que se esté midiendo. No debe ser muy largo para evitar que el fenómeno observado sufra variaciones que alterarían el valor de la consistencia y tampoco debe ser demasiado corto para evitar el recuerdo de las respuestas (efecto aprendizaje)³⁷.

b) Fiabilidad inter-observador: es el grado de acuerdo que hay entre dos o más evaluadores que valoran a los mismos sujetos con el mismo instrumento³⁸. Las limitaciones principales se deben a la posibilidad de que existan acuerdos entre los observadores que no sean debidos al azar y la posibilidad de que exista un error sistemático (sesgo de información) de alguno de los evaluadores³⁷. Esta propiedad no es evaluable cuando se trata de cuestionarios auto cumplimentados, ya que es el

propio individuo quien proporciona las respuestas sin que exista interferencia de los investigadores. No es el caso del CAPE-V.

2. Validez

Es la capacidad del cuestionario para medir aquel constructo para el que ha sido diseñado^{37,39}. Puede evaluarse para todas o sólo para alguna de sus cuatro dimensiones: validez aparente o lógica, de contenido, de criterio y de constructo.

a) Validez aparente o lógica: se refiere al grado en que un cuestionario, a juicio de los expertos y de los usuarios, mide de forma lógica lo que quiere medir^{33,39}. Cuando se carece de validez aparente o lógica, los sujetos sometidos a estudio pueden no ver la relación entre las preguntas que se les formulan y el objeto para el cual han accedido a contestar. Este hecho puede provocar el rechazo de los participantes. En nuestro caso el estudio se lleva a cabo en un entorno clínico para el estudio y diagnóstico de las alteraciones de la calidad vocal, por tanto los sujetos son pacientes que acuden a la consulta del especialista en trastornos vocales.

b) Validez de contenido: los constructos suelen estar compuestos por varias dimensiones, y en el caso de la calidad vocal es un hecho. La validez de contenido es el grado en que la herramienta es capaz de medir la mayor parte de las dimensiones del constructo^{33,38,39}. Un cuestionario con alta validez de contenido es aquel que mide todas las dimensiones relacionadas con el constructo que se quiere estudiar. Su evaluación es un proceso formal y consiste en valorar si los ítems del cuestionario son una muestra representativa de aquello que se quiere medir. Se trata de una evaluación empírica, basada en juicios de diferente procedencia, como son las opiniones de los autores de la herramienta, los resultados de estudios piloto, los razonamientos realizados por los autores de una TACV y el análisis cualitativo de los comentarios realizados por los participantes durante el proceso.

En este trabajo, el constructo estudiado es el análisis perceptual de la voz, cuyas dimensiones o ítems a valorar son las recomendadas por un conjunto de expertos convocados por la ASHA y reunidos en 2002. Estos expertos enumeran una serie de ítems básicos y elaboran un protocolo pormenorizado de evaluación basado en los conocimientos científicos existentes hasta el momento e intentando subsanar los inconvenientes de otros protocolos ya validados.

c) Validez de criterio: establece la validez de un instrumento comparándola con un criterio externo, prueba de referencia o “*gold standard*”. Tiene dos dimensiones: la validez concurrente o grado en que el resultado del cuestionario concuerda con el *gold standard*, y la validez predictiva o grado en que es capaz de pronosticar un determinado resultado^{33, 38, 39}.

El “*gold standard*” (GS) debe ser un método alternativo, equivalente, independiente de los resultados del cuestionario, fiable, exacto, objetivo y ampliamente aceptado como medida válida. Cuando reúne estos requisitos, la prueba de referencia es capaz de dar un resultado siempre positivo en presencia de fenómeno a estudiar y siempre negativo en ausencia del mismo.

Siempre que se disponga de un GS, debería evaluarse la validez de criterio concurrente, siguiendo cinco pasos: 1) selección del GS; 2) selección de una muestra de sujetos representativa de la población; 3) administración del cuestionario y obtención del resultado para cada individuo; 4) evaluación de cada individuo con el GS, y 5) comparación de los resultados obtenidos con el cuestionario y el GS.

Esta secuencia es la que se lleva a cabo en el presente estudio utilizando el GRABS como GS y cuya metodología en cuanto a la selección y características de la muestra se expondrá en el apartado de “Material y Métodos”.

El análisis de la validez de criterio concurrente supone examinar la fuerza de la correlación existente entre el resultado del cuestionario y el del GS³⁷.

d) Validez de constructo: es el grado en que las mediciones que resultan de las respuestas del cuestionario pueden considerarse una medición del fenómeno estudiado^{33, 38, 39}. Su evaluación consiste en contrastar las hipótesis que se han formulado sobre el comportamiento de las puntuaciones de un instrumento en situaciones diferentes. Existen varios métodos para su evaluación, que deben realizarse cuando el fenómeno a medir es abstracto o no es posible comparar con un GS.

Capítulo 2

JUSTIFICACIÓN DE LA INVESTIGACIÓN Y OBJETIVOS

El método de evaluación auditivo perceptual de la voz conocido como CAPE-V representa la mejor forma de abordar la calificación vocal por: su planteamiento en forma de protocolo detallado, por tener en cuenta tanto la voz sin influencias de la articulación así como el habla conectada y natural y por su buena fiabilidad y rendimiento en este cometido, superando a los métodos conocidos y difundidos en la actualidad. Este protocolo está descrito y desarrollado en idioma inglés, por lo que no puede ser implementado por pacientes hispanohablantes.

Como no constan antecedentes bibliográficos sobre trabajos que pretendan adaptar y validar este consenso al idioma español, se considera justificado abordar este problema. Por tanto, el objeto del presente trabajo de investigación se centra en adaptar y validar el método auditivo perceptual de evaluación de la voz conocido como CAPE-V al idioma español.

HIPOTESIS

El protocolo que está descrito para llevar a cabo la valoración perceptual utilizando el método CAPE-V ha sido validado para un uso clínico en su versión original en inglés. Si ha sido correctamente traducido y adaptado al idioma castellano, se podrá demostrar su validez y fiabilidad por medio de los métodos estadísticos indicados para ello.

OBJETIVOS GENERALES

1. Adaptar el protocolo del Consenso para la Valoración Auditivo-Perceptual de la Voz (CAPE-V) al idioma español.
2. Validar la adaptación del CAPE-V al español para su uso clínico en el ámbito de la valoración auditivo-perceptual de la calidad vocal.

Objetivos específicos:

- a) Demostrar la consistencia o fiabilidad del CAPE-V a nivel intraobservador e interobservador.
- b) Demostrar la validez de criterio entre los parámetros del CAPE-V y los del GRABS.
- c) Analizar la relación entre las calificaciones perceptuales a través de las distintas tareas vocales.

d) Calcular la sensibilidad al cambio del CAPE-V y del GRABS.

Capítulo 3

MATERIAL Y MÉTODOS

En este apartado se expone de forma detallada el proceso llevado a cabo para la adaptación cultural y validación del método CAPE-V al español.

Se abordan los aspectos prácticos y concretos que se han explicado en el último epígrafe de la introducción detallando pormenorizadamente la metodología del estudio.

1. Adaptación al español del CAPE-V.

Para el desarrollo del presente estudio se realizó inicialmente una adaptación del CAPE-V²⁰, originariamente en inglés, al español, la cual se adjunta en el anexo B de este trabajo.

Este proceso no consistió en una mera traducción inglés-español, sino que fue necesario llevar a cabo una adaptación y construcción de las oraciones, ya que una traducción literal de las mismas al castellano no permitiría la valoración de los elementos de la calidad vocal deseados.

Así, para diseñar las nuevas oraciones se tuvo en cuenta el contexto fonético que se pretende evaluar en cada caso. Para esta adaptación se contó con el asesoramiento de una logopeda (AAM).

A continuación se presentan las seis oraciones que se proponen para la adaptación del método CAPE-V al español:

a) Oración 1. Diseñada para estudiar la influencia coarticulatoria de las distintas vocales:

- Oración original en idioma inglés: “*The blue spot is on the key again*”.
- Oración adaptada por este trabajo: “Nuria ojeó una pajarita y una blusa amarillas”.

b) Oración 2. Ofrece un contexto adecuado para valorar la producción de ataques glóticos suaves y transiciones de sorda a sonora.

- Oración original en idioma inglés: “*How hard did he hit him*”.
- Oración adaptada por este trabajo: “Marta multa mi moto más mágica”.

c) Oración 3. Produce todos los fonemas sonoros y crea un contexto adecuado para juzgar la posible existencia de espasmos/detenciones y la habilidad del paciente para enlazar una palabra con la siguiente (manteniendo la vocalización).

- Oración original en idioma inglés: “*We were away a year ago*”.
- Oración adaptada por este trabajo: “La llave brilla en la mano”.

d) Oración 4. Incluye muchas palabras que comienzan con una vocal y que pueden provocar ataques glóticos duros, dando la oportunidad para saber si éstos ocurren en el paciente.

- Oración original en idioma inglés: “*We eat eggs every Easter*”.
- Oración adaptada por este trabajo: “Irene adora hacer huevos al horno”.

e) Oración 5. Incluye numerosas consonantes nasales dando así la oportunidad para valorar la hiponasalidad y si es posible la estimulación por medio de la terapia vocal resonante.

- Oración original en idioma inglés: “*My mama makes lemon jam*”.
- Oración adaptada por este trabajo: “Mamá me mima una mano”.

f) Oración 6. Presenta una ausencia total de consonantes nasales y crea un contexto útil para evaluar la presión intraoral y si existe una hipernasalidad y emisión aérea nasal.

- Oración original en idioma inglés: “*Peter will keep at the peak*”.
- Oración adaptada por este trabajo: “Ata tu zapato a tu pata”.

2. Participantes

En este trabajo han participado un total de 66 personas: 30 hombres y 36 mujeres, con edades comprendidas entre los 21 y 80 años, con una edad media de 46 años.

Estas 66 personas se dividieron en dos grupos: el primero consta de 50 sujetos con disfonía (24 mujeres y 26 hombres) y el segundo de 16 sujetos sanos (12 mujeres, 4 hombres).

Los sujetos disfónicos que participaron en el estudio fueron reclutados de entre los pacientes que acuden al servicio de consultas externas del Servicio de Otorrinolaringología del Hospital Universitario Central de Asturias entre julio de 2013 y marzo de 2014. Todos los pacientes accedieron a participar en el estudio tras explicarles la naturaleza del mismo y otorgaron el consentimiento verbal expreso para utilizar sus voces en el mismo. En la tabla 1 se recogen los diagnósticos de este grupo. El grupo de sujetos sanos consta de voluntarios del entorno hospitalario (médicos

residentes, enfermeras del hospital y amistades) que accedieron a participar en el estudio y que no presentan ninguno de los siguientes criterios de exclusión: ser menor de edad o estar o haber estado diagnosticado de alguna patología laríngea.

PATOLOGÍA	N
Pólipo vocal	13
Neoformaciones Laríngeas	10
Edema de Reinke	9
Quistes Subepiteliales	6
Nódulos	4
Pseudoquiste Seroso	4
Leucoplasia	1
Parálisis vocal idiopática	1
Granuloma tercio posterior	1
Papilomatosis laríngea	1
TOTAL	50

Tabla 1. *Diagnósticos del grupo de pacientes que participa en el estudio.*

Dentro del grupo de pacientes disfónicos, se crea un subgrupo con 22 pacientes (12 mujeres y 10 hombres) que han sido sometidos a tratamiento quirúrgico mediante microlaringoscopia directa para extirpar la lesión tipo masa que presentaban. En la tabla 2 se exponen sus diagnósticos. Se trata de un subgrupo de pacientes homogéneo en cuanto al tipo de disfonía que presentan y al cambio esperado tras la intervención. Estos pacientes cuentan con dos grabaciones: una preoperatoria y otra posterior a la intervención (aproximadamente dos meses después la intervención). Ambas muestras se randomizan y clasifican de forma independiente.

PATOLOGÍA	N
Pólipo vocal	9
Edema de Reinke	4
Quistes Subepiteliales	6
Nódulos	2
Granuloma tercio posterior	1
TOTAL	22

Tabla 2. Diagnósticos del grupo de pacientes intervenido quirúrgicamente.

3. Muestras vocales.

Se grabaron un total de 88 muestras vocales:

- 66 correspondientes a los sujetos de los grupos: disfónico y control
- 22 muestras postoperatorias de los pacientes intervenidos quirúrgicamente.

A cada participante, tanto del grupo disfónico como del grupo control, se les pidió que realizaran las tres tareas que incluye la descripción original del protocolo CAPE-V siguiendo la metodología que se detalla. Sin embargo, las tareas se llevan a cabo en idioma español utilizando para ello las oraciones fonéticamente adaptadas que se han expuesto previamente.

En general las muestras vocales se eligen en relación al parámetro de afectación general, grado o severidad general sin tener en cuenta la relevancia del resto de los parámetros. De cualquier forma se ha cuidado la inclusión de muestras en un rango de patologías amplio y representativo de los distintos tipos de alteraciones vocales.

4. Procedimiento de grabación

El individuo realizó las tareas sentado cómodamente en un lugar con ambiente tranquilo (menos de 50dB de ruido de fondo). El evaluador hizo una grabación de audio de la actuación de la persona en las tres tareas que se le pidieron. La grabación se realizó utilizando una grabadora digital con altavoz de 300 mW modelo SONY ICD-BX700 colocada en ángulo de 45° con una distancia desde la boca del sujeto al micrófono de unos 4 centímetros. Las grabaciones se recogen en un ordenador con más de 16 bits de resolución de tarjeta de sonido y una frecuencia de muestreo de no menos de 20 KHz.

5. Metodología de aplicación del protocolo CAPE-V

A continuación se detalla la metodología llevada a cabo para la recogida de las muestras vocales tal y como describe el protocolo original del CAPE-V. El doctorando es quien realiza la recogida de las muestras garantizando la calidad de las mismas. También será el encargado de aleatorizar las muestras garantizando la realización del proceso a simple ciego.

- Tarea 1:

El evaluador comunica al individuo “Diga el sonido /e:/ manteniéndolo con su voz normal hasta que le pida que pare”. La persona efectúa esta tarea tres veces durante 3-5 s. cada una. De las tres muestras, se escoge la que tenga mejor calidad y no presente artefactos o silencios.

- Tarea 2:

El evaluador proporciona al individuo 6 tarjetas en las que aparecen las 6 frases en español adaptadas fonéticamente que constituyen esta tarea. El evaluador comunica al sujeto que lea las oraciones, respetando varios segundos de silencio entre ellas, de la forma más natural posible. En caso de imposibilidad para la lectura, el evaluador se le pidió que repitiera las frases.

- Tarea 3:

El evaluador grabó al menos 20 s. de conversación natural utilizando la pregunta: ¿Cómo percibe su voz?. Se realiza la misma pregunta a todos los individuos sean disfónicos o sanos, intentando que durante la grabación no revelen su condición ni citen el tipo de patología que padecen. De esta manera, mantenemos el estudio a doble ciego.

6. Evaluación auditivo-perceptual de las muestras vocales (GRABS y CAPE-V)

El análisis auditivo-perceptual de las muestras vocales se lleva a cabo por dos especialistas familiarizados con ambos métodos. El doctorando almacena la serie de voces a calificar en dos secuencias diferentes determinadas de forma aleatoria, garantizando la confidencialidad de las mismas.

El observador 1 calificó con el método GRABS toda la serie escuchando para ello exclusivamente la grabación de la vocal sostenida. En un segundo tiempo, unos siete días después, califica la misma serie con el método CAPE-V, escuchando ya todas las tareas grabadas. Finalmente, en una tercera ocasión realiza otra calificación de toda la serie mediante el CAPE-V. El observador 2 realiza la calificación de toda la serie con el protocolo CAPE-V.

Para la calificación, las voces fueron reproducidas por los observadores en un ambiente exento de ruido y las escucharon en los altavoces del ordenador en bloques de 20 con un intervalo de descanso entre los bloques consecutivos de 10 minutos. Aunque ambos examinadores tienen experiencia en el análisis auditivo-perceptual de la voz, recibieron dos escalas de voces patrón que fueron construidas utilizando una ayuda visual espectrográfica para voces aéreas y voces ásperas que se explica con más detalle en el apartado siguiente⁴⁰.

Tras la escucha de cada voz, el observador procedió a su calificación sin compartir información con el otro, pudiendo repetir la escucha cuantas veces considerase necesario. El estudio se realiza a simple ciego, puesto los calificadores no conocen el diagnóstico del paciente⁴¹.

GRABS: Mediante la reproducción sonora de cada muestra vocal se calificó con ayuda visual espectrográfica por el observador 1, puntuando cada uno de los siguientes ítems de 0 a 3 (0 = normal, 1 = leve, 2 = moderado, 3: severo): G (Grado), el grado global de afectación vocal. R (Roughness, Aspereza), la calidad de la voz relacionada con la impresión de pulsos glóticos irregulares, de un componente de ruido de baja frecuencia, de aspereza o vocal fry. B (Breathiness, voz aérea), la voz relacionada con el ruido originado por las turbulencias creadas por una glotis insuficiente. A (Astenia), la impresión auditiva de debilidad en la fonación espontánea, Voz hipocinética o hipofuncional. S (Strain, tensión vocal), la impresión auditiva de excesivo esfuerzo, de tensión asociada con la fonación espontánea. El hecho de emplear la ayuda visual espectrográfica se consideró importante para que esta calificación, que va a ser utilizada como “*gold standard*” para evaluar la validez de criterio de la adaptación del CAPE-V, contara con la máxima fiabilidad⁴⁰.

CAPE-V: Una vez realizadas y grabadas las tareas, tanto el observador 1 como el 2 completaron el formulario adaptado al español del CAPE-V (Anexo A) con su valoración de cada parámetro. Para ello, se escuchan las muestras correspondientes a las tres tareas vocales de cada individuo o muestra y se califican una por una, marcando los resultados en un formulario distinto para cada tarea. Se coloca una marca de verificación en cada una de las escalas visuales analógicas que corresponde a cada uno de los seis parámetros analizados teniendo en cuenta que el extremo

izquierdo de la línea indica normalidad o ausencia de cualidad y el derecho, la máxima desviación de la normalidad.

En el protocolo original se contempla la posibilidad de detectar diferencias en la ejecución de las distintas tareas vocales, que obliguen a reflejar calificaciones distintas para cada tarea del protocolo. Sin embargo, el protocolo original no recomienda la utilización sistemática de formularios distintos. Si la ejecución no es uniforme, se coloca en cada escala una marca para cada tarea etiquetando la marca con los subíndices 1, 2 y 3 relativos a las diferentes tareas. En el caso de existir discrepancias también dentro de una misma tarea, a los números anteriores se le añadieron letras para indicar los diferentes ítems dentro de cada una (1a, 1b, 2a, 2b, 2c, 2d, 2e, 2f).

Para facilitar la recogida de datos y evitar sesgos de información, en nuestro trabajo se evalúa cada tareas de forma independiente y en un formulario distinto. El objetivo es detectar, si existieran, las diferencias entre las distintas tareas vocales evaluadas, ya que es uno de los elementos exclusivos e innovadores del CAPE-V. Una vez que el evaluador completa las calificaciones, calcula con la ayuda de una regla, la distancia que existe entre el extremo izquierdo de la escala con la marca de verificación, teniendo en cuenta que la longitud total de la línea de resultados es de 100 mm. La distancia resultante se escribe, en la parte derecha de cada escala. Esta distancia puede expresarse como puntos respecto a un total de 100, que serían equivalentes a los 100 mm de la denominación en milímetros.

Además de la escala numérica, los resultados también pueden expresarse utilizando las etiquetas descriptivas (leves, moderadas o severas) que hay bajo cada una de las escalas analógicas-visuales. De esta manera, se construye una base de datos con las calificaciones cuantitativas y cualitativas referentes a cada individuo y a cada tarea vocal.

No hemos tenido en cuenta la posibilidad de reflejar características adicionales en nuestros formularios. Constituye una característica importante del test, ya que permite la valoración de voces complejas o con características poco comunes. Sin embargo, se aleja del objetivo del presente estudio.

Patrones de Referencia:

Se crearon escalas para determinados ítems del GRABS (“Breathiness” y “Roughness”) mediante patrones espectrográficos que ilustran los extremos de severidad con el fin de que los observadores comparen el espectrograma de la voz que están analizando. En un extremo se observa una voz normal para el parámetro y en el extremo contrario un ejemplo de máxima severidad. La muestra que ilustra la severidad máxima se etiquetará con un porcentaje del 90% de afectación con el fin de evitar el “efecto fin de escala”. El parámetro B se ilustra con la presencia de ruido en el espectrograma siguiendo los grados de Yanagihara⁴² y el parámetro R se ilustra con la presencia de subarmónicos⁴³. En la figura 1A se muestra la escala para el parámetro B y en la figura 1B la escala para el parámetro R (Figuras cedidas por el Dr. Nuñez Batalla incluidas en el artículo: El espectrograma de banda estrecha como ayuda para el aprendizaje del método GRABS de análisis perceptual de la disfonía).

Figura 1 A. Escala analógico-visual para el parámetro Breathiness, con tres grados de severidad, donde se pueden observar los espectrogramas de banda estrecha correspondiente.

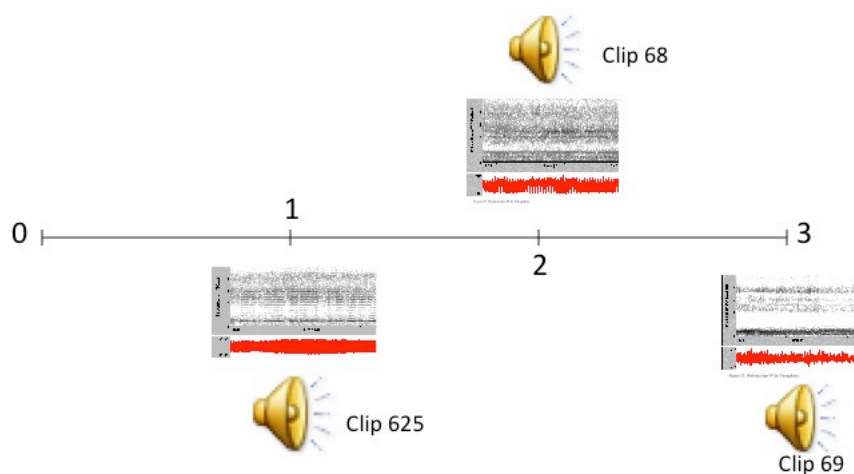
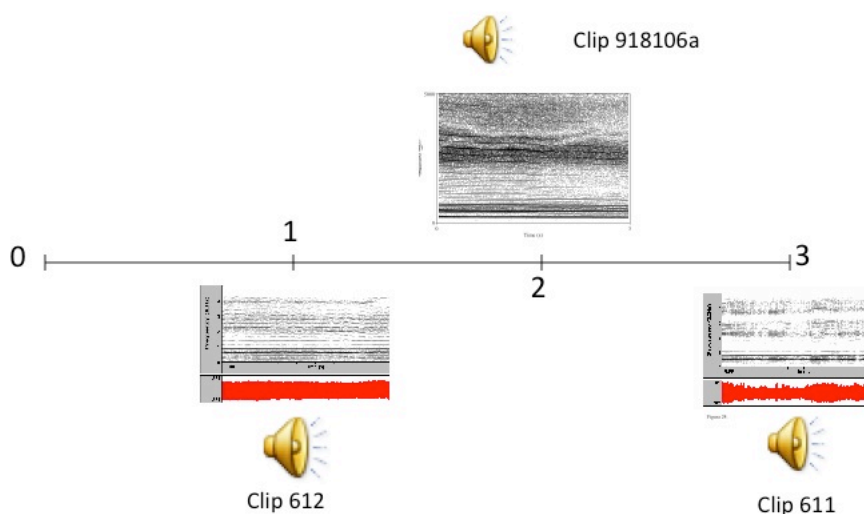


Figura 1 B. Escala analógico-visual para el parámetro Roughness, con tres grados de severidad, donde se pueden observar los espectrogramas de banda estrecha correspondientes.



7. Procedimientos y Análisis estadístico.

Se llevará a cabo el proceso de validación del protocolo CAPE-V siguiendo los pasos incluidos en el apartado 4 de la introducción: Adaptación cultural y validación de cuestionarios de salud. Una vez expuestos conceptualmente los parámetros a validar, en el siguiente apartado se especifican las consideraciones analíticas referentes a la validación de los mismos.

Para el trabajo estadístico se utiliza el programa informático SPSS 18 (PASW Statistics 18). Inicialmente se realiza un análisis estadístico básico de las calificaciones obtenidas. Al tratarse de la validación de una escala de medida, es necesario establecer puntos de corte para discriminar cada rango de patología vocal en la nueva prueba y equipararlo con la existente para el GRABS. Este hecho constituye en sí mismo un problema, puesto que la valoración de la voz depende de criterios subjetivos condicionados por el entorno social, los requerimientos vocales y profesionales de cada individuo y otros muchos factores.

El GRABS consta de una escala de cuatro valores ordinales, del 0 al 3. Sin embargo el CAPE-V consta de una escala del 0 al 100 que teóricamente podría comportarse como una variable continua. Sin embargo, se comprueba en la práctica que las calificaciones

adquieren valores discretos. Esto quiere decir, que es infrecuente que los observadores califiquen alguna muestra con valores decimales, suelen ser valores enteros y más concretamente valores cuyas unidades sean: 0 ó 5.

Clasificación por Grupos del CAPE-V: se agrupan las calificaciones según los tres intervalos en los que se divide la escala visual analógica: leves, medios y graves.

Corresponderán a leves aquellos con calificaciones que oscilan entre: 0 y 33; medios aquellos con puntuaciones entre 34 y 66 y graves las muestras con puntuaciones superiores a 66 puntos.

Clasificación por Categorías del CAPE-V: los mismos resultados se desdoblan también en 4 categorías de calificación para hacerlos comparables a la escala del GRABS que dispone de 4 rangos. Para ello, los grupos medio y grave se corresponderán con las categorías 2 y 3, respectivamente, puesto que coinciden con los mismos intervalos de calificación. La diferencia se encuentra en el grupo de leves, que al pasar a categorías se desdobra en dos grupos. El primero está formado por todas las puntuaciones 0 que constituirá por tanto la categoría 0 y el segundo grupo formado por resto de puntuaciones localizadas entre 1 y 33. (Tabla 3)

Rango GRABS	Grupo CAPE-V	Categoría CAPE-V
0	Leve	0
1		1
2	Medio	2
3	Grave	3

Tabla 3. Correlación entre GRABS y CAPE-V

1. Fiabilidad (reliability)

- Fiabilidad test- retest o consistencia intraobservador: Mide la estabilidad de las puntuaciones otorgadas por el mismo evaluador en los mismos sujetos y con el mismo método, en momentos diferentes. Para ello, el observador 1 califica mediante el CAPE-V a todos los sujetos en dos ocasiones diferentes, con un lapso de tiempo entre ellas de 2 a 6 semanas.
- Fiabilidad o consistencia interobservador: Consiste en estimar el grado de acuerdo entre dos o más evaluadores. Para ello las tareas grabadas en audio

de cada uno de los participantes del estudio serán valoradas por los dos evaluadores de forma independiente, siguiendo las instrucciones del protocolo del CAPE-V.

El estadístico utilizado para evaluar estas dos tareas es el Coeficiente de Correlación Intraclase (ICC), que permite valorar la concordancia entre dos medidas de variables cuantitativas³⁶. En la literatura existente⁴⁴, es frecuente el uso del Coeficiente de Pearson (r) como indicador de fiabilidad debido a su facilidad de obtención técnica. Este coeficiente de Pearson mide la intensidad de la asociación lineal entre dos variables pero no informa sobre el nivel de acuerdo observado. Por eso no resulta una medida adecuada para calcular la fiabilidad o consistencia, puesto que si dos instrumentos miden sistemáticamente cantidades diferentes uno del otro, la correlación puede ser perfecta ($r=1$) y sin embargo, la concordancia es nula. El índice que debe utilizarse para esta tarea es el ICC en el modelo de efectos aleatorios, esta condición permite seleccionar de forma aleatoria el observador y la muestra. Según el resultado obtenido, clasificaremos la fuerza de la concordancia como sigue:

VALOR ICC	FUERZA DE LA CONCORDANCIA
> 0,90	Muy Buena
0,90 – 0,70	Buena
0,70 – 0,50	Intermedia
0,50 – 0,30	Mediocre
<0,300	Mala

Tabla 4. Correlación valores ICC y Fuerza del nivel de acuerdo.

Estos valores nos servirán de referencia para la interpretación de los resultados obtenidos. Es necesario recordar que las calificaciones vocales son valoraciones subjetivas a las que pretendemos dar la mayor objetividad posible y hacerlas así comparables. No hay que olvidar este detalle en la interpretación de los datos, puesto que no podemos ser estrictos en la consideración de los resultados.

2. Validez

Es la capacidad de un instrumento para medir aquello para lo que ha sido diseñado³³. Como se explica en la introducción existen diferentes dimensiones para medir la validez de un instrumento: validez lógica, validez de contenido, validez de criterio y validez de constructo. La validez lógica y la validez de contenido son dimensiones de base conceptual, que en nuestro caso damos por demostradas teniendo en cuenta que este protocolo fue creado a partir de los conocimientos y la experiencia de expertos científicos en la materia.

El prototipo de validez para el análisis de pruebas diagnósticas es la validez de criterio, por eso será el que analicemos en este caso. Cuando se diseña un nuevo instrumento de medición se dispone de algún método alternativo, criterio externo o criterio de referencia del fenómeno estudiado con validez demostrada. El criterio externo debe ser una medición independiente y debe obtenerse por un método diferente, en el que no intervengan los resultados del cuestionario que se pretende validar.

En nuestro estudio, el criterio de referencia es el GRABS, por ser ampliamente utilizada y estar validada a nivel mundial. El nuevo test, el denominado CAPE-V, dispone de parámetros homólogos a los que contiene el GRABS y dos ítems complementarios que no se incluyen en la valoración del criterio de referencia. Los criterios para juzgar estos dos parámetros nuevos, se basarán en la valoración por parte de expertos. (Tabla 5)

CAPE-V	GRABS
Severidad General	Grado
Aspereza	Roughness
Calidad Aérea	Breathiness
Tensión	Strain
Tono	--
Intensidad	--

Tabla 5. Parámetros del GRABS y CAPE-V

Para el cálculo de la validez, utilizaremos por un lado medidas de correlación a través de parámetros como el coeficiente de Spearman, que es un análogo al coeficiente de correlación de Pearson (r) utilizado para la correlación de variables cuantitativas continuas sin distribución normal o con variables discretas. Siendo el GRABS una variable discreta es necesario utilizar este parámetro. Por otro lado y como se explica previamente, es necesario hallar la concordancia o nivel de acuerdo entre las mediciones de los dos test, mediante el índice de Kappa o el Coeficiente de correlación intraclase. Éste último será el que utilizemos puesto que es más adecuado para variables complejas.

3. Relación entre las distintas tareas o Validez de constructo

La evaluación de las muestras vocales con el CAPE-V se realiza a través de distintas tareas de fonación, siendo esta característica una de las más importantes que se pueden atribuir a este test. Cabe plantear la duda razonable de que evaluando tareas de fonación distintas, las calificaciones obtenidas pueden ser diferentes. Para reforzar la importancia de esta característica, que nos permite realizar una evaluación más completa y exhaustiva, es importante analizar las relaciones entre las distintas tareas evaluadas.

La metodología estadística se basará en una diferencia entre las medias de cada grupo de calificación según la tarea de fonación y el parámetro. Solo se tendrá en cuenta la significación estadística de la diferencia, puesto que para hallar la magnitud de la diferencia habría que realizar un planteamiento estadístico diferente y adaptado específicamente, probablemente con un número muestral más amplio. La cuantificación de la diferencia se aleja del principal objetivo del presente estudio por lo que no se realiza.

Por otro lado, la comparación de las tareas vocales se acompañará de datos de correlación y de nivel de acuerdo. Aunque existan diferencias en la calificación, atribuibles a la variabilidad subjetiva derivada de la calificación de tareas diferentes, los niveles de acuerdo informarán sobre la consistencia del juicio emitido. Como ya hemos mencionado previamente, el mejor indicador del nivel de acuerdo para nuestra muestra es el coeficiente de correlación intraclase o CCI.

4. Sensibilidad al cambio⁴⁵

Es el grado con que se obtienen los diferentes resultados en aplicaciones repetidas del mismo instrumento cuando se ha producido un cambio real en el estado de salud. Por lo tanto, puede definirse como la capacidad de un instrumento para detectar un cambio. La determinación del cambio dependerá: del cambio producido, y de la capacidad del instrumento de medida para detectar el cambio. En nuestro caso, aplicaremos la sensibilidad al cambio interna, que se define específicamente como la capacidad de una medida para cambiar en un lapso temporal determinado, refiriéndose a la posibilidad de detectar cualquier cambio estadísticamente significativo.

Se aplicará a un grupo de muestras homogéneo, que es el subconjunto de 22 pacientes que se someterán a tratamiento quirúrgico. Como se expuso previamente, se prevé que este subgrupo se vea afectado de una forma similar por el cambio, ya que está constituido por pacientes afectados de patología benigna de la voz. Las muestras son sometidas a calificación con el CAPE-V en dos momentos: antes y después de la cirugía. El estadístico que se utilizará será la prueba t de Student asumiendo que la hipótesis nula es la ausencia de cambio en la respuesta media de una escala en dos momentos. Este estadístico se centra exclusivamente, en la significación estadística del cambio observado pero no es un estadístico adecuado para evaluar la magnitud del cambio.

El objetivo de nuestro estudio no es la detección de la magnitud de la sensibilidad al cambio. Por este motivo, solo se valorará la diferencia entre las calificaciones medias de cada parámetro antes y después del tratamiento aplicado, haciendo hincapié en la significación estadística y no en la magnitud de los resultados que se obtengan³⁵.

Capítulo 4
RESULTADOS

DATOS DESCRIPTIVOS DE LA MUESTRA

En este apartado se exponen las calificaciones para cada uno de los parámetros estudiados de cada muestra analizada. Los resultados se exponen mediante tablas que reflejan las puntuaciones por distribuidas por decenas.

La muestra se ha seleccionando atendiendo al parámetro severidad general y por tanto éste destaca por ofrecer un rango amplio de calificaciones, representativo de la población general. Incluimos dentro de este parámetro la clasificación en grupos y en categorías. Sin embargo, para el resto de parámetros solo se incluye la tabla de calificaciones general por decenas, por considerar de escasa relevancia el estudio minucioso y exhaustivo del resto de los datos. En cada tabla, resaltará visualmente la celda de datos que contenga el mayor porcentaje evaluado en cada tarea, para resaltar visualmente el grupo de calificación con mayor frecuencia de puntuación.

Después de la exposición de cada parámetro del CAPE-V en cada una de sus tareas, se exponen los resultados de sus homólogos en el GRABS, a través de una tabla que refleja las calificaciones por categorías del 0 al 3.

Parámetro	Media Tarea 1	Media Tarea 2	Media Tarea 3
Severidad General	52,98	36,07	37,02
Aspereza	28,82	16,85	17,75
Calidad Aérea	43,99	29,16	30,51
Tensión	29,89	16,29	16,63
Tono	33,93	22,58	25,17
Intensidad	36,4	25,06	26,52

Tabla 6. Medias de todos los parámetros del CAPE-V para cada tarea.

PARÁMETRO SEVERIDAD GENERAL (CAPE-V)

Las puntuaciones se distribuyen de forma homogénea en todas las tareas sin que ninguna decena, grupo o categoría destaque sobre las demás. Se trata de una muestra balanceada en relación a este parámetro.

Severidad General	Tarea 1 (%)	Tarea 2 (%)	Tarea 3 (%)
0	16,9	33,7	32,6
10	3,4	4,5	4,5
20	7,9	7,9	14,6
25	0	2,2	0
30	2,2	5,6	3,4
40	7,9	7,9	3,4
50	6,7	7,9	6,7
60	6,7	4,5	7,9
70	16,9	6,7	6,7
75	3,4	2,2	1,1
80	9,0	5,6	5,6
90	7,9	1,1	3,4
100	11,2	10,1	10,1
Total	100	100	100

Tabla 7. Calificaciones para SG

SG Grupos	Tarea 1 (%)	Tarea 2 (%)	Tarea 3 (%)
Leve	30,3	53,9	56,2
Medio	21,3	20,2	16,9
Grave	48,3	25,8	27,0
Total	100	100	100

SG Categorías	Tarea 1 (%)	Tarea 2 (%)	Tarea 3 (%)
0	16,9	33,7	34,8
1	13,5	20,2	21,3
2	21,3	20,2	16,9
3	48,3	25,8	27,0
Total	100	100	100

Tablas 8 y 9. Distribución de datos en grupos y categorías para la SG

PARÁMETRO ASPEREZA (CAPE-V)

La mayor parte de las puntuaciones se clasifican como normales, con puntuación 0. Las tareas 2 y 3 tienden a calificar de forma más frecuente las muestras como normales.

Aspereza	% Tarea 1	% Tarea 2	% Tarea 3
0	34,8	50,6	51,7
10	4,5	6,7	7,9
20	16,9	15,7	13,5
25	0	1,1	0
30	10,1	9,0	6,7
40	3,4	2,2	1,1
50	6,7	5,6	7,9
60	6,7	2,2	4,5
70	7,9	3,4	3,4
75	1,1	1,1	0
80	3,4	2,2	2,2
90	3,4	0	1,1
100	1,1	0	0
Total	100	100	100

Tabla 10. Calificaciones totales para ASP

PARÁMETRO CALIDAD AÉREA (CAPE-V)

En términos generales la muestra incluye pocas alteraciones de la calidad aérea. Para la calificación a través del CAPE-V la muestra es bastante homogénea con un claro predominio de las calificaciones normales en las tareas 2 y 3.

CALIDAD AÉREA	% TAREA 1	% TAREA 2	% TAREA 3
0	19,1	40,4	40,4
10	3,4	7,9	3,4
20	10,1	10,1	15,7
30	10,1	6,7	5,6
40	9,0	6,7	6,7
50	11,2	6,7	3,4
60	9,0	2,2	2,2
70	9,0	4,5	6,7
75	1,1	1,1	1,1
80	4,5	2,2	1,1
90	2,2	1,1	3,4
100	11,2	10,1	10,1
Total	100	100	100

Tabla 11. Calificaciones para CA

PARÁMETRO TENSIÓN (CAPE – V)

La mayor parte de las muestras se califican como normales, a pesar de que la tarea 1 detecta menor proporción de pacientes en este rango normal en relación al porcentaje detectado en las tareas 2 y 3.

TENSIÓN	% TAREA 1	% TAREA 2	% TAREA 3
0	36	71,9	69,7
10	4,5	2,2	3,4
20	11,2	4,5	5,6
30	12,4	2,2	3,4
40	10,1	2,2	2,2
50	5,6	5,6	3,4
60	4,5	0	1,1
70	2,2	0	0
80	3,4	0	0
90	4,5	4,5	3,4
100	5,6	6,7	7,9
Total	100	100	100

Tabla 12. Calificaciones parámetro TE

PARÁMETRO TONO (CAPE -V)

La muestra escogida no se caracteriza por las alteraciones del tono, ya que la mayor parte de las puntuaciones se localizan en el rango de la normalidad.

TONO	% TAREA 1	% TAREA 2	% TAREA 3
0	33,7	55,1	57,3
10	1,1	2,2	0
20	6,7	5,6	6,7
30	14,6	10,1	10,1
40	6,7	2,2	3,4
50	11,2	6,7	6,7
60	9,0	5,6	2,2
70	4,5	2,2	3,4
80	3,4	2,2	2,2
90	4,5	3,4	3,4
100	4,5	4,5	4,5
Total	100	100	100

Tabla 13. Calificaciones parámetro T

PARÁMETRO INTENSIDAD (CAPE – V)

La muestra no se caracteriza por alteraciones de la intensidad, de esta manera todas las tareas califican la mayor parte de las muestras como normales.

INTENSIDAD	% TAREA 1	% TAREA 2	% TAREA 3
0	33,7	49,4	47,2
10	1,1	3,4	3,4
20	7,9	7,9	9,0
30	6,7	6,7	7,9
40	6,7	5,6	3,4
50	10,1	10,1	9,0
60	12,4	3,4	6,7
70	10,1	4,5	2,2
80	1,1	0	1,1
90	5,6	3,4	4,5
100	4,5	5,6	5,6
Total	100	100	100

Tabla 14. Calificaciones parámetro Int

PARÁMETRO GRADO (GRABS)

<i>Grado</i>	Frecuencia	Porcentaje
0	17	19,1
1	12	13,5
2	29	32,6
3	31	34,8
Total	89	100,0

Tabla 15. Distribución de Frecuencias para el Grado

Con la escala GRABS, en concreto para el parámetro Grado, la mayor parte de las puntuaciones se localizan en la región patológica de la tabla, en la categoría 3.

PARÁMETRO ROUGHNESS – ASPEREZA (R) (GRABS)

<i>Roughness</i>	Frecuencia	Porcentaje
0	40	44,9
1	27	30,3
2	16	18,0
3	6	6,7
Total	89	100,0

Tabla 16. Frecuencias para la calificación del parámetro *R*

La categoría 0 agrupa el mayor porcentaje de casos (44,9%) y le sigue en frecuencia la categoría 1 (30,3%) mientras que los dos últimos grupos son menos representativos. Coincidiendo con el CAPE-V, la mayor parte de las muestras se califican como normales.

PARÁMETRO “*BREATHINESS*”-VOZ AÉREA (B) (GRABS)

<i>Breathiness</i>	Frecuencia	Porcentaje
0	18	20,2
1	23	25,8
2	25	28,1
3	23	25,8
Total	89	100,0

Tabla 17. Frecuencias para la calificación del parámetro *B*

La distribución de las calificaciones con el GRABS es muy homogénea, aunque una leve diferencia destaca la categoría 2 como la más frecuente.

PARÁMETRO “*STRAIN*”-TENSIÓN DEL GRABS (S)

<i>Strain</i>	Frecuencia	Porcentaje
0	44	49,4
1	29	32,6
2	9	10,1
3	7	7,9
Total	89	100,0

Tabla 18. Calificación del parámetro *S*

La mayor parte de las muestras se clasifican como normales y se incluyen en la categoría 0 de la clasificación GRABS en cuanto a la tensión.

FIABILIDAD TEST RE TEST o VALIDEZ INTRA OBSERVADOR:

La fiabilidad “test re test” permite medir el nivel de acuerdo entre dos calificaciones realizadas por el mismo observador en dos momentos diferentes. El intervalo de tiempo entre ambas calificaciones debe ser de al menos 15 días.

La puntuación de las 89 muestras vocales a través del protocolo CAPE-V se comporta como una variable cuantitativa discreta y el coeficiente más adecuado para calcular la fiabilidad intraobservador es el Coeficiente de Correlación Intraclase (CCI) que se interpretará según se describe en Material y Métodos.

La exposición de la consistencia intraobservador de cada parámetro se realiza a través de una tabla con los siguientes datos: medias correspondientes a la primera calificación y a la segunda, medida de correlación de Pearson (r), coeficiente de correlación intraclase y su intervalo de confianza al 95% con sus correspondientes límites; superior e inferior. Tal y como se ha explicado previamente, el Pearson no determina el nivel de acuerdo entre dos calificaciones, sin embargo, está incluido en la exposición de los resultados para informar sobre el nivel de correlación de las dos calificaciones.

La denominación de las variables se hará mediante las siglas correspondientes a cada parámetro y se adjuntará un dígito numérico que indicará si se trata de la primera calificación (1) o de la segunda calificación(2).

TAREA 1:

Los coeficientes de correlación de todos los parámetros son mayores de 0,90. Esto nos permite afirmar que las calificaciones del observador principal son muy consistentes. Exponemos una tabla con los valores de correlación de Pearson y los niveles de acuerdo. Los parámetros están ordenados en relación a su nivel de acuerdo o CCI.

TAREA 1	Pearson	CCI
Severidad General	0,977	0,988
Calidad Aérea	0,963	0,981
Aspereza	0,959	0,979
Intensidad	0,944	0,971
Tono	0,930	0,963
Tensión	0,927	0,961

Tabla 19. Parámetros estadísticos Fiabilidad IntraObservador Tarea 1

VALIDEZ INTRAOBSERVADOR DE LA SEVERIDAD GENERAL

Las puntuaciones medias para la primera y la segunda clasificación son muy similares. Tal como indican valores tan elevados del Pearson y el CCI, las dos variables mantienen una correlación y un nivel de acuerdo muy fuerte.

	Puntuación		CCI	Intervalo de Confianza 95%	
	Media	Pearson (r)		Limite Inf.	Límite Sup.
SG 1	52,98				
SG 2	50,97	0,977	0,988	0,982	0,992

Tabla 20. Diferencia de Medias SG, correlación y acuerdo.

VALIDEZ INTRAOBSERVADOR DE LA ASPEREZA

La correlación entre las medias de ambas variables demuestra que las calificaciones son muy similares. El nivel de acuerdo y su estrecho intervalo de confianza ponen de relevancia la buena consistencia del parámetro Aspereza .

	Puntuación		CCI	Intervalo de Confianza al 95%	
	Media	Pearson (r)		Limite Inf.	Límite Sup.
ASP 1	28,82				
ASP 2	26,18	0,950	0,979	0,968	0,986

Tabla 21. Diferencia de Medias ASP, correlación y acuerdo.

VALIDEZ INTRAOBSERVADOR DE CALIDAD AÉREA

La calidad aérea presenta unas puntuaciones medias en ambos momentos muy similares y con alta correlación. El nivel de acuerdo es muy bueno, por lo que este parámetro demuestra elevada consistencia para la tarea 1 de CA.

	Puntuación		CCI	Intervalo de Confianza al 95%	
	Media	Pearson (r)		Límite Inf.	Límite Sup.
CA 1	43,99				
CA 2	41,80	0,963	0,981	0,971	0,988

Tabla 22. Diferencia de Medias CA, correlación y acuerdo.

VALIDEZ INTRAOBSERVADOR DE TENSIÓN

El parámetro tensión tiene un nivel de correlación y de acuerdo entre ambas variables muy elevado. Para la tarea 1, el parámetro tensión tiene una elevada consistencia.

	Puntuación	Pearson (r)	CCI	Intervalo de Confianza al	
	Media			95%	
				Límite Inf.	Límite Sup.
Te 1	29,89				
Te 2	23,26	0,927	0,961	0,941	0,974

Tabla 23. Diferencia de Medias TE, correlación y acuerdo.

VALIDEZ INTRAOBSERVADOR EN EL TONO

Las medias de las calificaciones para el momento 1 y el 2 guardan una correlación elevada y un nivel de acuerdo muy bueno. El tono en la tarea 1 tiene una consistencia elevada.

	Puntuación	Pearson (r)	CCI	Intervalo de Confianza al	
	Media			95%	
				Límite Inf.	Límite Sup.
Ton 1	33,93				
Ton 2	30,67	0,930	0,963	0,945	0,976

Tabla 24. Diferencia de Medias TON, correlación y acuerdo.

VALIDEZ INTRAOBSERVADOR DE LA INTENSIDAD

Las medias para el parámetro intensidad están muy próximas y guardan niveles de correlación elevados. El nivel de acuerdo es muy bueno, por lo que la consistencia interna del parámetro en la tarea 1 es elevada.

	Puntuación	Pearson (r)	CCI	Intervalo de Confianza al	
	Media			95%	
				Limite Sup.	Límite Inf.
Int 1	36,40				
Int 2	33,82	0,944	0,971	0,956	0,981

Tabla 25. Diferencia de Medias INT, correlación y acuerdo.

TAREA 2:

Igual que ocurría en la tarea 1, los parámetros para la tarea 2 tienen un nivel de acuerdo y correlación muy fuertes. En la siguiente tabla, se exponen ordenados según el CCI. Estos resultados traducen una validez intraobservador para la tarea 2 muy consistente.

TAREA 2	Pearson	CCI
Severidad General	0,945	0,972
Calidad Aérea	0,940	0,969
Tensión	0,939	0,968
Intensidad	0,853	0,921
Tono	0,849	0,918
Aspereza	0,809	0,895

Tabla 26. Parámetros estadísticos Fiabilidad IntraObservador Tarea 2

VALIDEZ INTRAOBSERVADOR DE SEVERIDAD GENERAL:

Las medias para la severidad general son muy similares, la correlación y el nivel de correlación son elevados por lo que para la tarea 2 este parámetro tiene una consistencia muy buena.

	Puntuación	Pearson (r)	CCI	Intervalo de Confianza al	
	Media			95%	
SG 1	36,40			Límite Inf.	Límite Sup.
SG 2	34,55	0,945	0,972	0,957	0,981

Tabla 27. Diferencia de Medias SG, correlación y acuerdo.

VALIDEZ INTRAOBSERVADOR DE LA ASPEREZA:

La correlación y el nivel de acuerdo observados para el parámetro aspereza son muy elevados.

	Puntuación	Pearson (r)	CCI	Intervalo de Confianza al	
	Media			95%	
ASP 1	17,53			Límite Inf.	Límite Sup.
ASP 2	14,21	0,809	0,895	0,839	0,931

Tabla 28. Diferencia de Medias ASP, correlación y acuerdo.

VALIDEZ INTRAOBSERVADOR DE CALIDAD AÉREA (CA):

Las medias de ambas calificaciones son similares, su correlación y su nivel de acuerdo es elevada.

	Puntuación	Pearson (r)	CCI	Intervalo de Confianza al	
	Media			95%	
CA 1	29,49			Límite Inf.	Límite Sup.
CA 2	27,53	0,940	0,969	0,953	0,980

Tabla 29. Diferencia de Medias CA, correlación y acuerdo.

VALIDEZ INTRAOBSERVADOR DE LA TENSIÓN:

Las dos variables guardan buena correlación y muy buen nivel de acuerdo. La consistencia intraobservador de la tensión es muy fuerte.

	Puntuación	Pearson (r)	CCI	Intervalo de Confianza al	
	Media			95%	
TE 1	15,06			Límite Inf.	Límite Sup.
TE 2	13,71	0,939	0,968	0,952	0,979

Tabla 30. Diferencia de Medias TE, correlación y acuerdo.

VALIDEZ INTRAOBSERVADOR DEL TONO:

Para el tono la correlación y el nivel de acuerdo son muy altos. La fiabilidad intraobservador es buena para este parámetro.

	Puntuación	Pearson (r)	CCI	Intervalo de Confianza al	
	Media			95%	
TO 1	22,92			Límite Inf.	Límite Sup.
TO 2	19,21	0,849	0,918	0,876	0,946

Tabla 31. Diferencia de Medias TO, correlación y acuerdo.

VALIDEZ INTRAOBSERVADOR PARA LA INTENSIDAD:

Al igual que los parámetros previos, la intensidad tiene unos valores muy consistentes.

Tabla 32. Diferencia de Medias INT, correlación y acuerdo.

	Puntuación	Pearson (r)	CCI	Intervalo de confianza al	
	Media			95%	
INT 1	25,39			Límite Inf.	Límite Sup.
INT 2	20,00	0,853	0,921	0,879	0,948

TAREA 3:

La tarea 3 muestra resultados similares a los de las tareas previas. Los coeficientes de correlación intraclase son muy elevados demostrando la fuerte consistencia intraobservador.

TAREA 3	Pearson	CCI
Severidad General	0,956	0,978
Calidad Aérea	0,947	0,972
Tensión	0,911	0,952
Aspereza	0,879	0,935
Intensidad	0,808	0,894
Tono	0,741	0,851

Tabla 33. Parámetros estadísticos Fiabilidad IntraObservador Tarea 2

VALIDEZ INTRAOBSERVADOR SEVERIDAD GENERAL:

Las medias de ambas variables guardan puntuaciones muy similares, el nivel de acuerdo es muy elevado. La consistencia interna de la SG en la tarea 3 es muy elevada.

	Puntuación	Pearson (r)	CCI	Intervalo de Confianza al	
	Media			95%	
				Límite Inf.	Límite Sup.
SG 1	36,01				
SG 2	32,25	0,956	0,978	0,966	0,985

Tabla 33. Diferencia de Medias SG, correlación y acuerdo.

VALIDEZ INTRAOBSERVADOR DE LA ASPEREZA:

Los valores de ambas variables tienen una correlación y un nivel de acuerdo muy elevado. La fiabilidad intraobservador para este parámetro en la tarea 3 es muy fuerte.

	Puntuación	Pearson (r)	CCI	Intervalo de Confianza al	
	Media			95%	
				Límite Inf.	Límite Sup.
ASP 1	16,63				
ASP 2	12,92	0,879	0,935	0,957	0,982

Tabla 34. Diferencia de Medias ASP, correlación y acuerdo.

VALIDEZ INTRAOBSERVADOR DE LA CALIDAD AÉREA:

El nivel de correlación es elevado tal y como demuestra el CCI. La consistencia intraobservador de la calidad aérea es muy elevada.

	Puntuación	Pearson (r)	CCI	Intervalo de Confianza al	
	Media			95%	
				Límite Inf.	Límite Sup.
CA 1	29,27				
CA 2	25,73	0,947	0,972	0,957	0,982

Tabla 35. Diferencia de Medias CA, correlación y acuerdo.

VALIDEZ INTRAOBSERVADOR DE LA TENSIÓN:

Para la tensión las medias de las calificaciones son similares, tienen una correlación alta y el nivel de acuerdo es muy elevado.

	Puntuación	Pearson (r)	CCI	Intervalo de Confianza al	
	Media			95%	
TE 1	15,96			Límite Inf.	Límite Sup.
TE 2	10,79	0,911	0,952	0,927	0,968

Tabla 36. Diferencia de Medias TE, correlación y acuerdo.

VALIDEZ INTRAOBSERVADOR TONO:

Las medias para los parámetros de ambas variables son muy similares. El coeficiente de Pearson demuestra una correlación buena y según el CCI el nivel de acuerdo es también bueno, por tanto la consistencia intraobservador del tono es alta.

	Puntuación	Pearson (r)	CCI	Intervalo de confianza al	
	Media			95%	
TO 1	22,13			Límite Inf.	Límite Sup.
TO 2	16,85	0,741	0,851	0,773	0,902

Tabla 37. Diferencia de Medias TO, correlación y acuerdo.

VALIDEZ INTRAOBSERVADOR DE LA INTENSIDAD:

La correlación de las puntuaciones para la intensidad es elevada y el nivel de acuerdo es bastante bueno, lo que traduce una consistencia intraobservador buena para este parámetro.

	Puntuación	Pearson (r)	CCI	Intervalo de Confianza al	
	Media			95 %	
INT 1	24,04			Límite Inf.	Límite Sup.
INT 2	17,42	0,808	0,894	0,838	0,930

Tabla 38. Diferencia de Medias INT, correlación y acuerdo.

FIABILIDAD O VALIDEZ INTEROBSERVADOR

Consiste en la comparación de las calificaciones de dos observadores experimentados e independientes para una subpoblación de la muestra de 50 casos. En este apartado, las calificaciones de la muestra se encuentran clasificadas en tres grupos: leves, medios y graves, teniendo en cuenta los criterios explicados en Material y Métodos

Se realiza un análisis de fiabilidad mediante el CCI, que informa del grado de acuerdo entre las muestras de cada observador. Se calcula a doble efecto aleatorio puesto que, ambas variables: los observadores y las muestras, se comportan de forma aleatoria dentro del conjunto de muestras.

Los resultados se muestran ilustrados a través de tablas de contingencia y tablas de datos. Se utilizará la abreviatura O1 para designar al observador número 1 y O2 para designar al observador número 2.

TAREA 1:

En términos generales, la correlación intraclase de los parámetros valorados en la tarea 1 está por encima de 0,70 en todos los casos, es decir, existe una buena correlación entre las calificaciones de los dos observadores.

PARÁMETRO	CCI
Severidad General	0,935
Aspereza	0,907
Calidad Aérea	0,894
Tensión	0,648
Tono	0,795
Intensidad	0,763

Tabla 39. Consistencia Interobservador para la Tarea 1

VALIDEZ INTEROBSERVADOR DE LA SEVERIDAD GENERAL.

El ICC Promedio es de 0,935 (IC 95%: 0,886 – 0,963). El nivel de máximo acuerdo se localiza en las muestras leves y graves, pero se observa menor concordancia entre observadores en los casos moderados.

		SG O2			Total
		<= 30	31 - 65	66+	
SG O1	<= 30	11	2	0	13
	31 - 65	3	11	0	14
	66+	0	7	16	23
Total		14	20	16	50

Tabla 40. Tabla de contingencia SG O1 * SG O2

	Correlación intraclase ^b	Intervalo de confianza 95%	
		Límite inferior	Límite superior
Medidas promedio	,935	,886	,963

Tabla 41. Valores CCI para SG

VALIDEZ INTEROBSERVADOR PARA LA ASPEREZA

La tabla de contingencias refleja que la mayor concordancia se obtiene en las calificaciones leves y graves, en el rango de calificaciones moderadas la concordancia entre los datos es menor. El coeficiente de correlación intraclase promedio es de 0,907 (IC95% 0,837 – 0,947).

		Asp O2			Total
		<= 30	31 - 65	66+	
Asp O1	<= 30	23	6	1	30
	31 - 65	2	5	4	11
	66+	0	0	9	9
Total		25	11	14	50

Tabla 42. Tabla de contingencia ASP O1 * ASP O2

	Correlación intraclase ^b	Intervalo de confianza 95%	
		Límite inferior	Límite superior
Medidas promedio	,907	,837	,947

Tabla 43. Valores CCI para ASP

VALIDEZ INTEROBSERVADOR DE LA CALIDAD AÉREA

La correlación entre ambos observadores es buena en las calificaciones leves y moderadas, pero no hay tanta concordancia en las calificaciones más graves, al contrario de lo que sucede previamente. El ICC mantiene un valor de 0,894 (IC 95% 0,814 – 0,940).

		CA O2			Total
		<= 30	31 - 65	66+	
CA O1	<= 30	16	3	0	19
	31 - 65	4	14	2	20
	66+	1	6	4	11
Total		21	23	6	50

Tabla 44. Tabla de contingencia CA O1* CA O2

	Correlación intraclase ^b	Intervalo de confianza 95%	
		Límite inferior	Límite superior
Medidas promedio	,894	,814	,940

Tabla 45. Valores CCI para CA

VALIDEZ INTEROBSERVADOR DE LA TENSION

El ICC de la tensión es de 0,648 (IC 95% 0,380 – 0,800) que se traduce en una relación moderada para las calificaciones de la tensión entre los calificadores, a expensas de los tres grupos.

		Te O1			Total
		<= 30	31 - 65	66+	
Te O2	<= 30	18	1	0	19
	31 - 65	13	8	0	21
	66+	5	1	4	10
Total		36	10	4	50

Tabla 46. Tabla de Contingencia Te O1* Te O2

	Correlación intraclase ^b	Intervalo de confianza 95%	
		Límite inferior	Límite superior
Medidas promedio	,648	,380	,800

Tabla 47. Valores CCI para Te

VALIDEZ INTEROBSERVADOR DEL TONO

El CCI es de 0,795 (IC 95% 0,639 – 0,884) donde la mayor concordancia aparece en el grupo de leves.

		Ton O2			Total
		<= 30	31 - 65	66+	
Ton O1	<= 30	25	4	0	29
	31 - 65	4	10	2	16
	66+	0	4	1	5
Total		29	18	3	50

Tabla 48. Tabla de Contingencia Ton O1* Ton O2

	Correlación intraclass ^b	Intervalo de confianza 95%	
		Límite inferior	Límite superior
Medidas promedio	,795	,639	,884

Tabla 49. Valores CCI para Ton

VALIDEZ INTEROBSERVADOR DE LA INTENSIDAD

El ICC es de 0,866 (IC95% 0,763 – 0,924). La mayor correlación se observa en los grupos extremos: leves y graves.

		Int O2			Total
		<= 30	31 - 65	66+	
Int O1	<= 30	21	0	0	21
	31 - 65	10	8	2	20
	66+	0	7	2	9
Total		31	15	4	50

Tabla 50. Tabla de Contingencia Int O1* Int O2

	Correlación intraclass ^b	Intervalo de confianza 95%	
		Límite inferior	Límite superior
Medidas promedio	,866	,763	,924

Tabla 51. Valores CCI para Int

TAREA 2:

El Índice de Correlación Intraclase de los parámetros valorados en la tarea 2 tiene unos resultados muy variables; entre un mínimo de 0,545 para la intensidad y un máximo de 0,879 para la Severidad General. Estos coeficientes tienen más heterogeneidad en esta tarea que en la tarea 1.

Como se puede ver en la tabla inferior, la severidad general, la aspereza y la calidad aérea tienen correlaciones fuertes que traducen un elevado nivel de acuerdo, sin embargo otros parámetros como la tensión, el tono o la intensidad tienen correlaciones interobservador menores y por lo tanto niveles de acuerdo intermedios y regulares.

En general, los grupos extremos de la calificación: leves y graves son aquellos en los que se alcanza mejor acuerdo. En la mayoría de los casos es el grupo de moderados el que introduce mayor variabilidad entre los calificadores.

Parámetro	CCI
Severidad General	0,879
Aspereza	0,775
Calidad Aérea	0,773
Tensión	0,698
Tono	0,710
Intensidad	0,545

Tabla 52. Consistencia Interobservador para la Tarea2

VALIDEZ INTEROBSERVADOR PARA LA SEVERIDAD GENERAL:

Existe muy buena correlación para todos los grupos de la muestra. El ICC se encuentra en 0,879 (IC 95% 0,786 – 0,931) un nivel elevado.

		SG O1			Total
		<= 30	31 - 65	66+	
SG O2	<= 30	20	1	0	21
	31 - 65	8	12	2	22
	66+	0	1	6	7
Total		28	14	8	50

Tabla 53. Tabla de Contingencia para SG O1* SG O2

	Correlación intraclase ^b	Intervalo de confianza 95%	
		Límite inferior	Límite superior
Medidas promedio	,879	,786	,931

Tabla 54. Valor CCI para SG

VALIDEZ INTEROBSERVADOR DE LA ASPEREZA:

El CCI es de 0,775 (IC 95% 0,603 – 0,872). La correlación entre los dos observadores es alta y tiene lugar a expensas de los grupos extremos de la muestra.

		ASP O2			Total
		<= 30	31 - 65	66+	
ASP O1	<= 30	31	6	3	40
	31 - 65	1	4	1	6
	66+	0	0	4	4
Total		32	10	8	50

Tabla 55. Tabla de Contingencia para ASP O1* ASP O2

	Correlación intraclase ^b	Intervalo de confianza 95%	
		Límite inferior	Límite superior
Medidas promedio	,775	,603	,872

Tabla 56. Valores de CCI para ASP

VALIDEZ INTEROBSERVADOR DE LA CALIDAD AÉREA

El CCI entre los dos observadores es de 0,773 (IC 95% 0,600 – 0,871).

		CA O2			Total
		<= 30	31 - 65	66+	
CA O1	<= 30	27	7	0	34
	31 - 65	3	8	0	11
	66+	1	1	3	5
Total		31	16	3	50

Tabla 57. Tabla de contingencia CA O1 * CA O2

	Correlación intraclase ^b	Intervalo de confianza 95%	
		Límite inferior	Límite superior
Medidas promedio	,773	,600	,871

Tabla 58. Valores de CCI para CA

VALIDEZ INTEROBSERVADOR DE LA TENSIÓN

El CCI para la tensión es de 0,698 (IC 95% 0,468 – 0,829).. El grupo de disfonías calificadas como moderadas es el que introduce más confusión en los datos.

		TE O2			Total
		<= 30	31 - 65	66+	
TE O1	<= 30	32	14	0	46
	31 - 65	1	1	0	2
	66+	0	0	2	2
Total		33	15	2	50

Tabla 59. Tabla de Contingencia para Te O1* Te O''

	Correlación intraclase ^b	Intervalo de confianza 95%	
		Límite inferior	Límite superior
Medidas promedio	,698^c	,468	,829

Tabla 60. Valor CCI para Te

VALIDEZ INTEROBSERVADOR DEL TONO

El CCI es de 0,710 (IC 95% 0,490 – 0,836). El grupo de disfonías leves y graves son los que mayor porcentaje de acuerdo obtienen.

		Ton O2			Total
		<= 30	31 - 65	66+	
Ton O1	<= 30	34	8	1	43
	31 - 65	2	2	0	4
	66+	0	1	2	3
Total		36	11	3	50

Tabla 61. Tabla de Contingencia para Ton O1* Ton O2

	Correlación intraclase ^b	Intervalo de confianza 95%	
		Límite inferior	Límite superior
Medidas promedio	,710	,490	,836

Tabla 62. Valores CCI para Ton

VALIDEZ INTEROBSERVADOR DE LA INTENSIDAD

El CCI es de 0,545 (IC 95% 0,198 – 0,792). La muestra contiene muchas disfonías clasificadas como leves, pero aún así la concordancia entre las calificaciones es pobre.

		INT O2			Total
		<= 30	31 - 65	66+	
INT O1	<= 30	32	2	0	34
	31 - 65	9	4	0	13
	66+	2	0	1	3
Total		43	6	1	50

Tabla 63. Tabla de Contingencia para Int O1* Int O2

	Correlación intraclase ^b	Intervalo de confianza 95%	
		Límite inferior	Límite superior
Medidas promedio	,545	,198	,742

Tabla 64. Valor CCI para Int

TAREA 3:

La correlación entre las calificaciones de los dos observadores independientes para la tarea 3 es buena y por tanto el nivel de acuerdo es elevado, puesto que en la mayoría de los parámetros está por encima del 0,700. Sin embargo la intensidad tiene un nivel de acuerdo menor.

TAREA 3	ICC
Severidad General	0,833
Aspereza	0,750
Calidad Aérea	0,769
Tensión	0,738
Tono	0,806
Intensidad	0,617

Tabla 65. Consistencia Interobservador para la Tarea 3

VALIDEZ INTEROBSERVADOR DE LA SEVERIDAD GENERAL

El CCI es de 0,833 (IC 95% 0,705 – 0,905). La tabla de contingencia muestra una buena concordancia en los grupos extremos de la muestra, el grupo intermedio introduce mucha variabilidad en el nivel de acuerdo.

		SG O2			Total
		<= 30	31 - 65	66+	
SGO1	<= 30	19	10	0	29
	31 - 65	1	7	3	11
	66+	0	4	6	10
Total		20	21	9	50

Tabla 66. Tabla de Contingencia para SG O1* SG O2

	Correlación intraclase ^b	Intervalo de confianza 95%	
		Límite inferior	Límite superior
Medidas promedio	,833	,705	,905

Tabla 67. Valor CCI para SG

VALIDEZ INTEROBSERVADOR DE LA ASPEREZA

El coeficiente de correlación intraclase es 0,750 (IC95% 0,560 – 0,858) con nivel de significación estadística. El mayor nivel de acuerdo se alcanza en el grupo de leves.

		ASP O2			Total
		<= 30	31 - 65	66+	
ASP O1	<= 30	31	6	3	40
	31 - 65	1	4	2	7
	66+	0	0	3	3
Total		32	10	8	50

Tabla 68. Tabla de Contingencia para ASP O1* ASP O2

	Correlación intraclase ^b	Intervalo de confianza 95%	
		Límite inferior	Límite superior
Medidas promedio	,750	,560	,858

Tabla 69. Valor CCI para ASP

VALIDEZ INTEROBSERVADOR DE LA CALIDAD AÉREA

El CCI es de 0,769 (IC 95% 0,593 – 0,869). El nivel de acuerdo se produce a expensas del grupo de leves, en el resto de los grupos se observa más dispersión de los datos.

		CA O2			Total
		<= 30	31 - 65	66+	
CA O1	<= 30	27	7	0	34
	31 - 65	3	7	0	10
	66+	1	2	3	6
Total		31	16	3	50

Tabla 70. Tabla de Contingencia para CA O1* CA O2

	Correlación intraclase ^b	Intervalo de confianza 95%	
		Límite inferior	Límite superior
Medidas promedio	,769	,593	,869

Tabla 71. Valor CCI para CA

VALIDEZ INTEROBSERVADOR DE LA TENSION

El coeficiente de correlación intraclase es de 0,738 (IC 95% 0,538 – 0,851). El mayor nivel de acuerdo se produce en el grupo de leves y graves.

		TEN O2			Total
		<= 30	31 - 65	66+	
TEN O1	<= 30	33	14	0	47
	31 - 65	0	1	0	1
	66+	0	0	2	2
Total		33	15	2	50

Tabla 72. Tabla de Contingencia para Te O1* Te O2

	Correlación intraclase ^b	Intervalo de confianza 95%	
		Límite inferior	Límite superior
Medidas promedio	,738	,538	,851

Tabla 73. Valor CCI para Te

VALIDEZ INTEROBSERVADOR DEL TONO

El coeficiente de correlación intraclase es de 0,806 (IC 95% 0,658 – 0,890) con significación estadística. La mayoría de las muestras son leves y en general se alcanza un buen nivel de acuerdo.

		TON O2			Total
		<= 30	31 - 65	66+	
TON O3	<= 30	34	4	1	39
	31 - 65	2	6	0	8
	66+	0	1	2	3
Total		36	11	3	50

Tabla 74. Tabla de Contingencia para Ton O1* Ton O2

	Correlación intraclase ^b	Intervalo de confianza 95%	
		Límite inferior	Límite superior
Medidas promedio	,806	,658	,890

Tabla 75. Valor CCI para Ton

VALIDEZ INTEROBSERVADOR DE LA INTENSIDAD

El coeficiente de correlación intraclase es de 0,617 (IC 95% 0,326 – 0,783). La mayor correlación se localiza en el grupo de muestras clasificadas como leves.

		INT O2			Total
		<= 30	31 - 65	66+	
INT O1	<= 30	34	2	0	36
	31 - 65	6	5	0	11
	66+	2	0	1	3
Total		42	7	1	50

Tabla 76. Tabla de Contingencia para Int O1* Int O2

	Correlación intraclase ^b	Intervalo de confianza 95%	
		Límite inferior	Límite superior
Medidas promedio	,617	,326	,783

Tabla 77. Valor CCI para Int

VALIDEZ DE CRITERIO

El prototipo de validez para el análisis de pruebas diagnósticas es la validez de criterio, que consiste en evaluar la correlación de la prueba estudiada con la prueba considerada como “gold standard” o criterio de referencia. El criterio externo debe ser una medición independiente y debe obtenerse a través de un método diferente, en el que no intervengan los resultados del cuestionario que se pretende validar.

En nuestro caso, el criterio de referencia que utilizamos es el GRABS por considerarse el protocolo de evaluación perceptual de la voz más extendido en su uso, sin embargo, existen múltiples publicaciones que recogen imperfecciones del método GRABS, asegurando que no es un método perfecto para la evaluación de las alteraciones vocales. En este apartado, evaluaremos la consistencia y el nivel de acuerdo entre las calificaciones realizadas con ambos protocolos. Se trata de uno de los epígrafes con más importancia dentro de este trabajo, puesto que nos permitirá validar el nuevo protocolo CAPE-V como prueba fiable y consistente en la evaluación de las alteraciones vocales en español.

Ante las imperfecciones detectadas en el GRABS, la relación y la consistencia entre ambos protocolos debe ser cuidadosamente interpretada. Al tratarse de calificaciones subjetivas, debemos aplicar parámetros estadísticos adecuados (no paramétricos) a riesgo de perder robustez en los cálculos. El protocolo de evaluación CAPE-V pretende subsanar estos inconvenientes al tratarse de una escala analógica visual con puntuaciones cuantitativas discretas, que nos permiten realizar calificaciones más precisas y con mayor consistencia.

Los parámetros estadísticos utilizados para este fin son: el Coeficiente de Correlación Intraclase (CCI) y el coeficiente de correlación de *Spearman*. Para medir el nivel de acuerdo, utilizamos el CCI que constituye el índice adecuado para calcular acuerdos entre variables complejas y la interpretación de la fuerza de concordancia en relación al valor numérico obtenido, se realizará de acuerdo a la tabla de datos expuesta al final del texto. El coeficiente de *Spearman* se utiliza para reflejar la relación entre dos variables, es un estadístico no paramétrico cuyo uso está indicado en este caso, al ser el GRABS una variable ordinal. Aunque no mide el nivel de acuerdo entre las variables, nos parece interesante incluir este parámetro como complemento para evaluar la relación entre las diferentes tareas del CAPE-V y el GRABS, aunque se logren distintos grados de acuerdo entre ellos, ilustraremos la correlación o no entre las calificaciones.

Para reflejar con más claridad las relaciones entre cada uno de los protocolos, se incluyen ilustraciones en forma de gráficos de distribución de cajas, que representan en abscisas la variable correspondiente a las calificaciones del GRABS y en ordenadas, las puntuaciones correspondientes al CAPE-V. Éstas últimas, se muestran agrupadas en cuatro categorías tal y como se detalla en el apartado de estadística básica para hacerlas comparables con las categorías del GRABS.

VALOR CCI	FUERZA DE LA CONCORDANCIA
> 0,90	Muy Buena
0,90 – 0,70	Buena
0,70 – 0,50	Intermedia
0,50 – 0,30	Mediocre
<0,300	Mala

Tabla 78. Valores de CCI y su fuerza de concordancia

TAREA 1 (CAPE-V)

En este apartado se ponen en relación las puntuaciones obtenidas con el GRABS y con la tarea 1 del CAPE-V, ambos correspondientes a la evaluación de vocales sostenidas.

En casi todos los casos, se obtienen índices de correlación de Spearman y nivel de CCI por encima del 90%, lo que se traduce en una concordancia muy buena para los resultados entre ambos protocolos. Para todos los parámetros de la tarea 1 los estadísticos mantienen datos muy similares que reflejan la casi perfecta relación y consistencia entre ambas escalas de medida, con datos robustos para la evaluación perceptual de las calificaciones.

Parámetro CAPE - GRABS	SPEARMAN	CCI
Severidad General – G	0,923	0,965
Aspereza – R	0,931	0,950
Calidad Aérea – B	0,924	0,961
Tensión – S	0,857	0,912

Tabla 79. Índices de correlación para los parámetros de la tarea 1

GRADO (GRABS) - SEVERIDAD GENERAL (CAPE-V):

El coeficiente de Spearman para la relación entre parámetros homólogos es de $0,891 \pm 0,030$ y el CCI es de $0,965$ (IC al 95%: $0,947 - 0,977$). Estos datos demuestran un nivel de concordancia entre ambos realmente alto y una buena correlación. Atendiendo a la tabla de contingencia, se observa que los parámetros que más concordancia demuestran son los valores extremos de la tabla.

	Grade				Total
	0	1	2	3	
0	15	0	0	0	15
SG 1	2	10	0	0	12
2	0	2	16	1	19
3	0	0	13	30	43
Total	17	12	29	31	89

Tabla 80. Tabla de contingencia entre SG y G

Índices de Correlación	Valor		
Spearman	0,891	Error Típ. Asint	Significación
		0,030	0,000
ICC (Int Confianza 95%)	0,965	Limite Inf.	Límite Sup.
		0,947	0,977

Tabla 81. Valores de los índices de correlación entre Severidad General y Grado

La gráfica de distribución de cajas que representa el parámetro Severidad General (CAPE-V) frente al Grado (GRABS) demuestra dos variables con una correlación lineal buena.

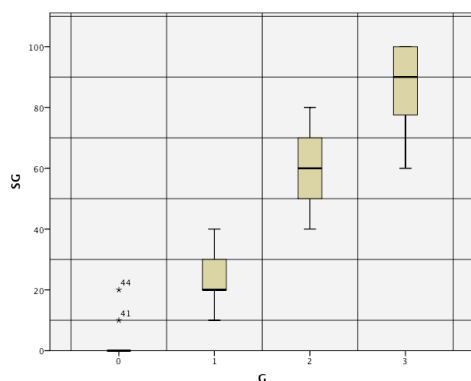


Gráfico de cajas Severidad general y Grado

ROUGHNESS (GRABS) – ASPEREZA (CAPE-V)

Se expone la tabla de contingencia con los datos correspondientes a la relación entre la Asperidad del CAPE – V (ASP) y el parámetro *Roughness* del GRABS. El coeficiente de Spearman es de $0,920 \pm 0,19$ y el CCI es de $0,950$ (IC 95%: $0,925 - 0,967$), ambos indican una buena correlación y nivel de acuerdo entre parámetros homólogos de los dos protocolos de calificación.

	<i>Roughness</i>				Total
	0	1	2	3	
ASP categoría 0	31	0	0	0	31
ASP categoría 1	6	22	0	0	28
ASP categoría 2	0	5	10	0	15
ASP categoría 3	0	1	7	7	15
Total	37	28	17	7	89

Tabla 82. Tabla de contingencias entre ASP y R

Indices de Correlación	Valor		
Spearman	0,920	Error Típ. Asint	Significación
		0,019	0,000
ICC (Int. Confianza al 95%)	0,950	Límite Inf.	Límite Sup.
		0,925	0,967

Tabla 83. Valores de los índices de correlación para Asperidad y Roughness

Para ilustrar esta relación, se acompaña el gráfico de distribución de cajas que demuestra la correlación lineal entre ambas variables.

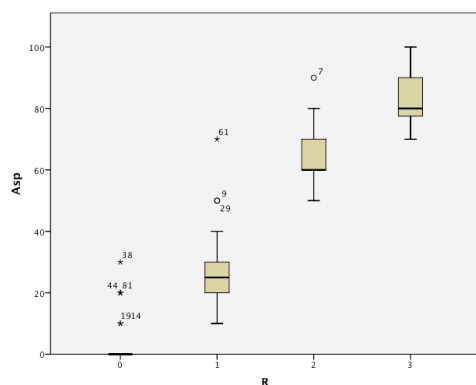


Gráfico de cajas Asperidad y R

BREATHINESS (GRABS) - CALIDAD AÉREA (CAPE-V)

Se compara la calidad aérea (CA) del CAPE-V con su homólogo en el GRABS, *Breathiness*. La correlación de *Spearman* es de $0,919 \pm 0,023$ y CCI es de $0,961$ (IC al 95%: $0,940 - 0,974$) que muestran elevada correlación y nivel de acuerdo.

El mayor porcentaje de acuerdo se alcanza en las categorías extremas de la tabla.

	<i>Breathiness</i>				Total
	0	1	2	3	
0	17	0	0	0	17
1	1	17	3	0	21
2	0	6	18	2	26
3	0	0	4	21	25
Total	18	23	25	23	89

Tabla 84. Tabla de Contingencias CA y B

Indices de Correlación	Valor		
Spearman	0,919	Error Típ. Asint.	Significación
		0,023	0,000
ICC (Int. Confianza 95%)	0,961	Límite Inf.	Límite Sup.
		0,940	0,974

Tabla 85. Valores de los índices de correlación para Calidad Aérea y *Breathiness*

Se adjunta el gráfico de distribución de cajas que muestra una progresión lineal entre las puntuaciones de ambos protocolos.

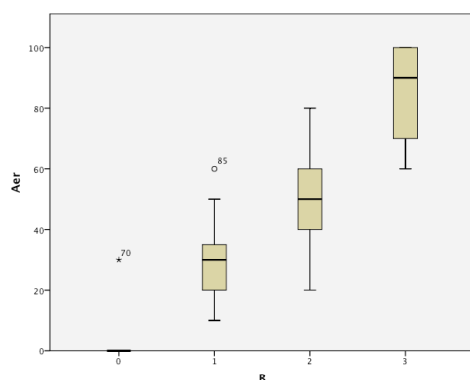


Gráfico de distribución de cajas entre la CA y B

STRAIN (GRABS) – TENSIÓN (CAPE-V)

Se compara el parámetro tensión (TE) del CAPE –V y el parámetro *Strain* del GRABS. La correlación de *Spearman* es de $0,822 \pm 0,045$ y el CCI es de $0,912$ (IC al 95%: $0,865- 0,942$). Como en los parámetros previos el nivel de correlación y acuerdo es elevado. En la tabla de contingencia se observa que la categoría 0 es la que más contribuye al nivel de acuerdo, puesto que para el resto de categorías el acuerdo es muy bajo y justifica los datos obtenidos.

	<i>Strain</i>				Total
	0	1	2	3	
0	30	2	0	0	32
1	12	13	0	0	25
2	2	14	2	0	18
3	0	0	7	7	14
Total	44	29	9	7	89

Tabla 86. Tabla de contingencias entre TE y S

Indices de Correlación	Valor		
Spearman	0,822	Error Típ. Asint	Significación
		0,045	0,000
ICC (Int Confianza 95%)	0,912	Limite Inf.	Límite Sup.
		0,865	0,942

Tabla 87. Valores de los índices de correlación para Tensión y Strain

Como para el resto de los parámetros previos, la gráfica de distribución de cajas refleja una correlación lineal entre ambos parámetros.

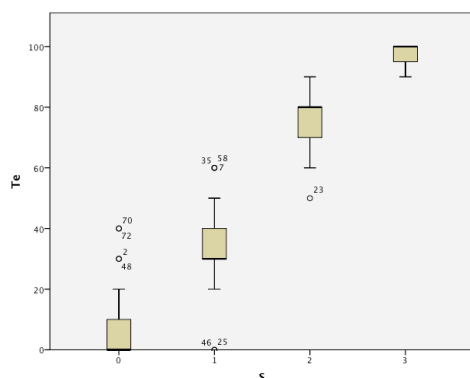


Gráfico de distribución de cajas entre T y S

TAREA 2 (CAPE-V)

En este apartado se comparan las calificaciones obtenidas para los parámetros del CAPE-V de las frases fonéticamente adaptadas correspondientes a la tarea 2 con sus homólogos del GRABS. Por lo tanto el contexto de evaluación es el siguiente, se comparan las calificaciones de dos tareas vocales distintas puesto que obtenemos el GRABS a partir de la fonación de vocales mientras que la tarea 2 del CAPE-V corresponde con el habla conectada.

En general, las relaciones obtenidas entre ambos tienen una fuerza de concordancia de intermedia a buena siendo para la severidad general su valor máximo y para la calidad aérea el valor mínimo. Sin embargo, es importante destacar que los coeficientes de Spearman indican una asociación intermedia/baja, indicando que las calificaciones perceptuales mantienen discreta relación pero tienen una buena concordancia. A efectos prácticos significa que el protocolo CAPE-V es válido y consistente, permitiendo clasificar de forma correcta a más del 70% de las muestras.

Con el protocolo CAPE-V las calificaciones de la tarea 2 tienden a ser menores que las calificaciones del GRABS y que las calificaciones obtenidas para la tarea 1, por eso el nivel de correlación es menor que el calculado para la tarea previa. Sin embargo, como decimos este hecho no condiciona un nivel de acuerdo bajo, si no que se mantiene en niveles aceptables que garantizan la consistencia de la evaluación y nos permite validar la tarea 2 del CAPE-V como prueba fiable para la calificación perceptual de la voz.

Parámetro CAPE - GRABS	SPEARMAN	CCI
Severidad General – G	0,668	0,789
Aspereza – R	0,651	0,755
Calidad Aérea – B	0,162	0,133
Tensión - S	0,563	0,819

Tabla 88. Índices de correlación para los parámetros de la tarea 2

GRADO (GRABS) – SEVERIDAD GENERAL (CAPE-V)

La correlación de *Spearman* para este parámetro es de $0,668 \pm 0,66$ y CCI es de $0,789$ (IC al 95%: $0,679- 0,862$). Existe un nivel de acuerdo y de correlación intermedio entre las calificaciones.

El número de pacientes calificados con categoría 0 ó 1 es muy voluminoso para el protocolo CAPE-V y es menos representativo con el GRABS. Esto refleja una situación particular, en la que pacientes calificados en categorías intermedias 1 ó 2 con el GRABS se evalúan o clasifican como normales o categoría 0 calificando la tarea 2 del CAPE-V y esto puede ser atribuido a las diferencias de la calificación intrínsecas a la tarea vocal evaluada.

	Grade				Total
	0	1	2	3	
0	16	6	6	2	30
1	0	3	9	6	18
2	0	3	11	4	18
3	1	0	3	19	23
Total	17	12	29	31	89

Tabla 89. Tabla de Contingencia entre G y SG

Indices de Correlación	Valor		
Spearman	0,668	Error Típ. Asint.	Significación
		0,073	0,000
CCI (Int. Confianza 95%)	0,789	Límite Inf.	Límite Sup.
		0,679	0,862

Tabla 90. Valores de los índices de correlación para G y SG

Los gráficos de dispersión muestran una distribución de las calificaciones para la Severidad General/Grado heterogéneas, con rangos de variación muy amplios y poco predecibles para los valores del CAPE-V. A pesar de que el rango de correlación sea intermedio, el gráfico refleja mejor la situación.

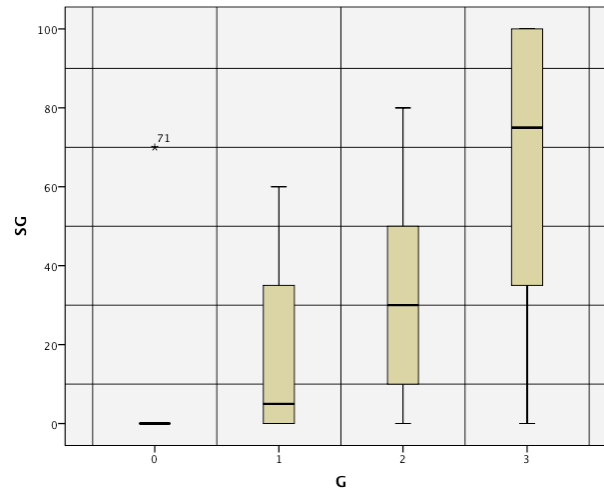


Gráfico de distribución de cajas entre G y SG

ROUGHNESS (GRABS) – ASPEREZA (CAPE-V)

El coeficiente de *Spearman* es de $0,651 \pm 0,75$ y el CCI es de 0,755 (IC al 95%: 0,627–0,839). El nivel de acuerdo y la correlación es intermedia entre los datos. Existe un grupo importante de acuerdo en la categoría 0 que como ya hemos mencionado previamente corresponde con la escasa alteración de las muestras en el parámetro aspereza. Sin embargo, el resto de las puntuaciones tiene una distribución dispersa y con escasa concordancia.

	<i>Roughness</i>				Total	
	0	1	2	3		
ASP	0	32	8	2	1	43
	1	4	16	7	3	30
	2	0	4	6	0	10
	3	1	0	2	3	6
Total		37	28	17	7	89

Tabla 91. Tabla de Contingencia ASP y R

Indices de Correlación	Valor		
Spearman	0,651	Error Típ. Asint.	Significación
		0,075	0,000
CCI (Int. Confianza 95%)	0,755	Límite Inf.	Límite Sup.
		0,627	0,839

Tabla 92. Valores de los índices de correlación para ASP y R

El gráfico de distribución de cajas representa unos resultados con mucha variabilidad, aunque exista correlación entre las puntuaciones, no es una correlación lineal ni predecible.

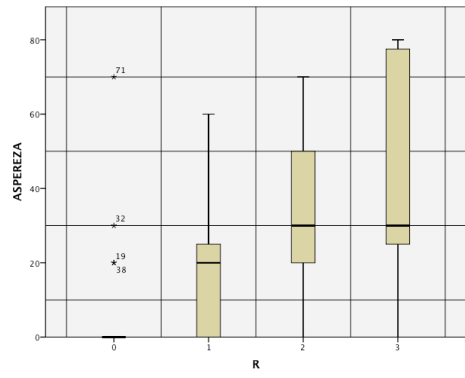


Gráfico de distribución de cajas para Asp y R

BREATHINESS (GRABS) - CALIDAD AÉREA (CAPE-V)

El nivel de correlación de *Spearman* es de $0,162 \pm 0,83$ y el CCI es de $0,133$ (IC al 95%: $-0,320 - 0,431$). Son coeficientes estadísticos de correlación y acuerdo muy bajos que revelan escasa relación entre ambos parámetros. Prueba de este hecho es la tabla de contingencias en la que no se observan grupos o categorías con aceptables niveles de acuerdo.

		<i>Breathiness</i>				Total
		0	1	2	3	
CA	0	35	0	1	0	36
	1	18	2	1	0	21
	2	10	4	0	1	15
	3	16	1	0	0	17
Total		79	7	2	1	89

Tabla 93. Tabla de Contingencia entre CA y B

Indices de Correlación	Valor		
Spearman	0,162	Error Típ. Asint.	Significación
		0,083	0,000
CCI (Int. Confianza 95%)	0,133	Límite Inf.	Límite Sup.
		-0,320	0,431

Tabla 94. Valores de los índices de correlación para CA y B

La escasa correlación del parámetro se ilustra con el gráfico. No se observa relación alguna entre los parámetros ni tampoco acuerdo entre ellos.

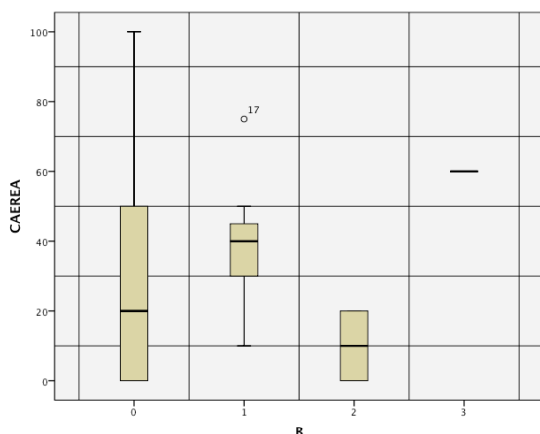


Gráfico de distribución de cajas para CA y B

STRAIN (GRABS) – TENSIÓN (CAPE-V)

En este parámetro es importante destacar que el nivel de correlación es intermedio, mientras que el nivel de acuerdo es alto. Esta diferencia debe interpretarse de forma cuidadosa, puesto que puede que la relación entre las calificaciones sea dispar pero finalmente los dos protocolos mantienen la consistencia y permiten clasificar los parámetros en la misma categoría.

	S				Total	
	0	1	2	3		
Te categorías	0	40	22	2	1	65
	1	2	5	1	0	8
	2	2	2	3	0	7
	3	0	0	3	6	9
Total	44	29	9	7	89	

Tabla 95. Tabla de Contingencias entre S y TE

Indices de Correlación	Valor		
Spearman	0,563	Error Típ. Asint.	Significación
		0,090	0,000
CCI (Int. Confianza 95%)	0,819	Límite Inf.	Límite Sup.
		0,724	0,881

Tabla 96. Valores de los índices de correlación para Tensión y Strain

El gráfico no muestra correlación entre ambas variables. Los datos de los grupos de menor calificación agrupan muchas puntuaciones 0. Los grupos de mayor calificación tienen una distribución muy variable.

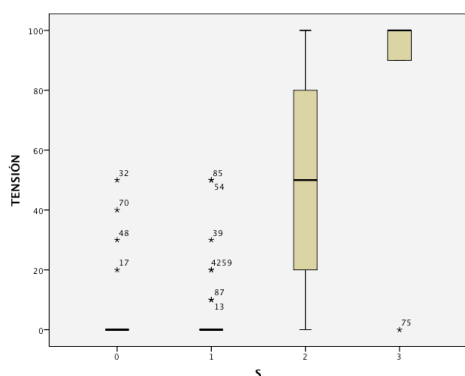


Gráfico de distribución de cajas entre Te y S

TAREA 3 (CAPE-V)

En este apartado se ponen en relación los resultados del protocolo GRABS con los resultados del CAPE-V obtenidos con la Tarea 3 correspondiente a la evaluación del habla conectada. Por lo tanto se evalúan las calificaciones obtenidas mediante la fonación de vocales sostenidas frente a las calificaciones realizadas mediante la evaluación del discurso natural.

Se obtienen resultados con un nivel de correlación y de acuerdo intermedio/altos que permiten establecer la consistencia y validez de las calificaciones a través de los distintos protocolos. El parámetro con mayor nivel de acuerdo es la severidad general mientras que el parámetro con peor resultado en el nivel de acuerdo es la calidad aérea.

Los parámetros estadísticos adquieren mejores resultados para la tarea 3 que para la tarea 2, esto puede deberse a que las alteraciones vocales durante el discurso natural se ven amortiguadas por el efecto de la articulación de la palabra y minimizan o hacen menos perceptibles estas alteraciones para el calificador. Sin embargo, el contexto lingüístico proporcionado por las frases de la tarea 2 está ideado para explorar determinadas actitudes de fonación, por lo que pueden poner de manifiesto alteraciones que pasen desapercibidas durante el discurso natural.

PARÁMETRO	Spearman	CCI
Severidad General – G	0,781	0,866
Aspereza – R	0,745	0,842
Calidad Aérea – B	0,616	0,742
Tensión – S	0,533	0,799

Tabla

101. Índices de correlación para los parámetros de la tarea 3

GRADO (GRABS) - SEVERIDAD GENERAL (CAPE-V)

El coeficiente de *Spearman* es $0,781 \pm 0,50$ y el ICC $0,866$ (IC al 95%: $0,795-0,912$).

El nivel de acuerdo y de correlación tienen valores intermedio/altos que reflejan buena concordancia de las calificaciones obtenidas a través de los distintos protocolos de evaluación.

		Grade				Total
		0	1	2	3	
SG	0	19	6	5	1	31
	1	0	4	11	4	19
	2	0	2	9	4	15
	3	0	0	4	20	24
Total		19	12	29	29	89

Tabla 102. Tabla de Contingencia entre SG y G

Indices de Correlación	Valor		
Spearman	0,781	Error Tip. Asintót.	Significación
		0,050	0,000
ICC (Int. Confianza 95%)	0,866	Límite Inf.	Límite Sup.
		0,795	0,912

Tabla 103. Valores de los índices de correlación para SG y G

Existen correlación lineal entre ambas variables, puesto que cuanto mayor es la puntuación en el CAPE-V tiende a ser mayor la calificación del GRABS. Sin embargo el intervalo de puntuaciones dentro del cual oscilan los datos en cada categoría es muy amplio y esto justifica la correlación moderada de los datos.

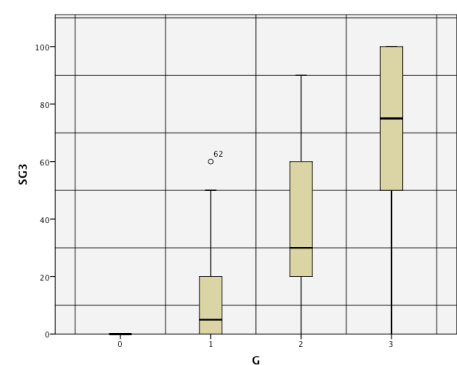


Gráfico de distribución de cajas para G y SG

ROUGHNESS (GRABS) – ASPEREZA (CAPE-V)

El coeficiente de *Spearman* es de $0,745 \pm 0,060$ y el CCI es de $0,842$ (IC al 95%: $0,759 - 0,896$). Éstos valores traducen una buena concordancia entre parámetros homólogos siendo la mayor parte de los acuerdos en la categoría 0 de ambos.

		<i>Roughness</i>				Total
		0	1	2	3	
ASP	0	37	8	2	0	47
	1	2	15	7	1	25
	2	1	3	6	2	12
	3	0	1	1	3	5
Total		40	27	16	6	89

Tabla 104. Tabla de Contingencia para R y ASP

Indices de Correlación	Valor		
Spearman	0,745	Error Tip. Asintót.	Significación
		0,060	0,000
ICC (Int. Confianza 95%)	0,842	Límite Inf.	Límite Sup.
		0,759	0,896

Tabla 105. Valores de los índices de correlación para R y ASP

Como en los casos anteriores la distribución de las calificaciones del CAPE-V son heterogéneas y tienen rangos de variación amplios.

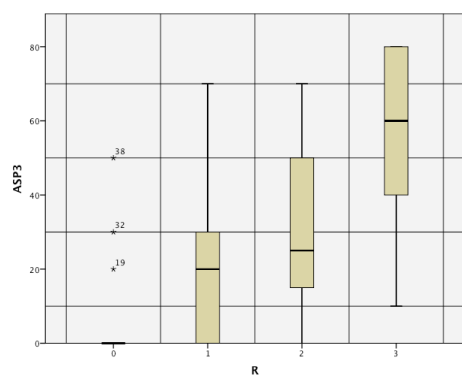


Gráfico de dispersión de cajas para R y Asp

BREATHINESS (GRABS) - CALIDAD AÉREA (CAPE-V)

El ICC es de $0,616 \pm 0,078$, mientras que el CCI es de $0,742$ (IC al 95%: $0,607 - 0,831$), son datos que reflejan una concordancia moderada. Como ocurre con otros ítems evaluados, la zona de mayor acuerdo se localiza en la categoría 0.

		<i>Breathiness</i>				Total
		0	1	2	3	
CA	0	20	11	7	1	39
	1	1	7	8	4	20
	2	0	2	7	2	11
	3	2	1	5	11	19
Total		23	21	27	18	89

Tabla 106. Tabla de contingencia para CA y B

Indices de Correlación	Valor		
Spearman	0,616	Error Tip. Asintót.	Significación
		0,078	0,000
ICC (Int. Confianza 95%)	0,742	Límite Inf.	Límite Sup.
		0,607	0,831

Tabla 107. Valores de los índices de correlación para CA y B

Como en los casos anteriores, se observa que el rango de calificaciones en las tres últimas categorías del GRABS abarcan casi todas las puntuaciones del CAPE-V, cuestión que justifica el moderado nivel de acuerdo. La categoría con mayor nivel de acuerdo es la 0.

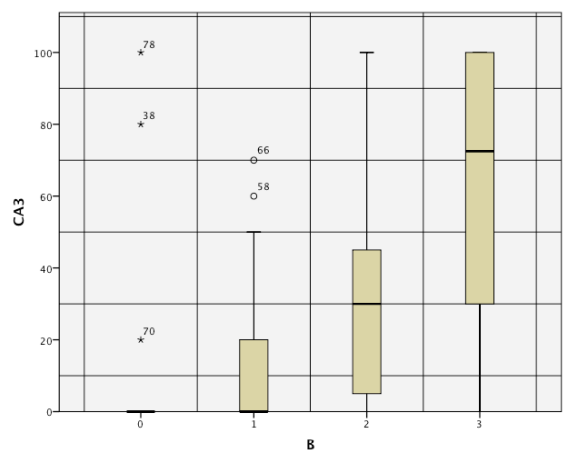


Gráfico de distribución de cajas entre B y CA

STRAIN (GRABS) – TENSIÓN (CAPE-V)

El índice de *Spearman* es bastante bajo ($0,533 \pm 0,095$) sin embargo, el acuerdo es de $0,799$ (IC al 95%: $0,694 - 0,868$). Este hecho se justifica por la influencia de la categoría 0 en los resultados, que forma la mayor parte de las calificaciones concordantes para con los dos métodos.

		S				Total
		0	1	2	3	
TE	0	39	17	3	0	59
	1	5	9	1	0	15
	2	1	2	3	0	6
	3	1	0	3	5	9
Total		46	28	10	5	89

Tabla 108. Tabla de contingencia entre S y TE

Índices de Correlación	Valor		
Spearman	0,533	Error Típ. Asint.	Significación
		0,95	0,000
ICC	0,799	Límite Inf.	Límite Sup.
		0,694	0,868

Tabla 109. Valores de los índices de correlación para S y TE

La expresión visual de los datos, traduce la concordancia de la categoría 0 y el amplio rango de variación del resto de grupos.

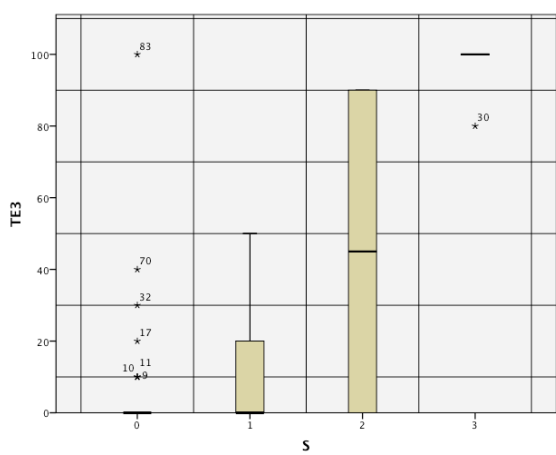


Gráfico de dispersión de cajas para Te y S

DIFERENCIA DE MEDIAS PARA LAS DISTINTAS TAREAS

La evaluación de las muestras vocales a través de distintas tareas de fonación, es una de las características más importantes y que diferencian a esta herramienta de evaluación. Cabe plantear la duda razonable de que midiendo tareas de fonación distintas las calificaciones obtenidas sigan mostrando consistencia y validez.

En el siguiente apartado realizaremos una comparación entre las medias de cada tarea para cada parámetro del test. Centraremos la comparación en la existencia o no de significación estadística entre la diferencia de medias, el nivel de correlación de los datos (medido a través del coeficiente de Pearson) y el nivel de acuerdo (obtenido a través del Coeficiente de Correlación Intraclase). No se tendrá en cuenta la magnitud de la diferencia, puesto que la muestra y el planteamiento estadístico del presente estudio no son adecuados para lograr tal objetivo.

Los resultados se exponen por cada parámetro; iniciando en primer lugar la comparación de la T1 frente a la T2 poniendo en relación las calificaciones obtenidas mediante la evaluación de vocales evaluación de frases. Después se compara la T1 con la T3, resultado de evaluar vocales sostenidas y el discurso natural. Por último, se evalúan las calificaciones obtenidas a través de las dos últimas tareas que se obtienen mediante el discurso natural y frases.

SEVERIDAD GENERAL

Se observan medias dispares con diferencias significativas entre ellas, por lo que, los datos ofrecen diferencias notables entre las dos tareas. Sin embargo, la buena correlación y el nivel de acuerdo alcanzados mantienen la consistencia de la calificación.

	Medias	Pearson	Dif. de Medias	Nivel de Signif.	CCI	Intervalo de Confianza al 95%	
						Límite Inf.	Límite Sup.
SG 1	52,98						
SG 2	36,40	0,742	16,573	0,000	0,852	0,774	0,903

Tabla 110. Diferencia medias SG T1 y T2

La comparación entre la tarea 1 y la tarea 3 refleja como en el caso anterior datos de medias significativamente diferentes. Los datos de correlación y de acuerdo son buenos, por tanto el valor de las calificaciones no se ve alterado.

	Medias	Pearson	Dif. de Medias	Nivel de Signif.	CCI	Intervalo de Confianza al 95%	
						Límite Inf.	Límite Sup.
SG 1	52,98						
SG 3	36,01	0,782	16,966	0,000	0,877	0,813	0,919

Tabla 111. Diferencia medias SG T1 y T3

Finalmente, la comparación de las dos últimas tareas no muestra diferencias significativas entre ellas y unos datos de correlación y acuerdo muy elevados.

	Medias	Pearson	Dif. de Medias	Nivel de Signif.	CCI	Intervalo de Confianza al 95%	
						Límite Inf.	Límite Sup.
SG 2	36,40						
SG 3	36,01	0,937	0,393	0,770	0,967	0,950	0,978

Tabla 112. Diferencia medias SG T2 y T3

ASPEREZA

Como en el caso anterior la comparación entre las tareas 1 frente a las tareas 2 y 3 reflejan unas medias estadísticamente diferentes, pero con correlaciones y niveles de acuerdo moderado/altos que mantienen la validez de la calificación.

	Medias	Pearson	Diferencia de medias	Nivel de signif.	CCI	Intervalo de Confianza al 95%	
						Límite Inf.	Límite Sup.
ASP 1	28,82						
ASP 2	17,53	0,609	11,292	0,000	0,743	0,609	0,831

Tabla 113. Diferencia medias ASP T1 y T2

	Medias	Pearson	Diferencia de medias	Nivel de signif.	CCI	Intervalo de Confianza al 95%	
						Límite Inf.	Límite Sup.
ASP 1	28,82						
ASP 3	16,63	0,631	12,191	0,000	0,760	0,634	0,842

Tabla 114. Diferencia medias ASP T1 y T3

Al comparar las tareas 2 y 3 los datos reflejan unas medias muy similares sin que existan diferencias significativas entre ellas. La correlación es muy elevada, así como el nivel de acuerdo entre las dos.

	Medias	Pearson	Diferencia de medias	Nivel de signif.	CCI	Intervalo de confianza al 95%	
						Límite Inf.	Límite Sup.
ASP 2	17,53						
ASP 3	16,63	0,855	0,899	0,493	0,922	0,881	0,949

Tabla 115. Diferencia medias ASP T2 y T3

CALIDAD AÉREA

La evaluación de la consistencia para las calificaciones de la calidad aérea a través de tareas de fonación distintas, revela diferencias significativas entre las medias para la tarea 1 frente a la tarea 2 y 3, mientras que los niveles de correlación y acuerdo son buenos que confirman la validez de la calificación.

	Medias	Pearson	Diferencia de medias	Nivel de signif.	CCI	Intervalo de confianza al 95%	
						Límite Inf.	Límite Sup.
CA 1	43,99						
CA 2	29,49	0,789	14,494	0,000	0,882	0,820	0,922

Tabla 116. Diferencia medias CA T1 y T2

	Medias	Pearson	Diferencia de medias	Nivel de signif.	CCI	Intervalo de confianza al 95%	
						Límite Inf.	Límite Sup.
CA 1	43,99						
CA 3	29,27	0,803	14,719	0,000	0,889	0,832	0,927

Tabla 117. Diferencia medias CA T1 y T3

Comparando las tareas 2 y 3 las medias son muy similares sin obtener diferencias significativas entre ellas. Los niveles de correlación y de acuerdo son muy fuertes, reflejando un nivel de acuerdo muy alto entre las calificaciones de ambas tareas.

	Medias	Pearson	Diferencia de medias	Nivel de signif.	CCI	Intervalo de confianza al 95%	
						Límite Inf.	Límite Sup.
CA 2	29,49						
CA 3	29,27	0,937	0,225	0,864	0,967	0,950	0,978

Tabla 118. Diferencia medias CA T2 y T3

TENSIÓN

El parámetro tensión refleja un comportamiento similar al de parámetros previos: la comparación entre las medias de T1 frente a las medias de la T2 y T3 refleja diferencias estadísticas significativas pero al mismo tiempo, guardan un nivel de correlación y de acuerdo fuertes.

	Medias	Pearson	Diferencia de medias	Nivel de signif.	CCI	Intervalo de confianza al 95%	
						Límite Inf.	Límite Sup.
TE 1	29,89						
TE 2	15,06	0,729	14,831	0,000	0,843	0,761	0,897

Tabla 119. Diferencia medias Te T1 y T2

	Medias	Pearson	Diferencia de medias	Nivel de signif.	CCI	Intervalo de confianza al 95%	
						Límite Inf.	Límite Sup.
TE 1	29,89						
TE 3	15,96	0,761	13,933	0,000	0,863	0,792	0,910

Tabla 120. Diferencia medias Te T1 y T3

Sin embargo, comparando tareas similares como son T2 y T3 se obtienen medias similares sin diferencias significativas. Los datos reflejan niveles de acuerdo muy fuerte entre ambas calificaciones.

	Medias	Pearson	Diferencia de medias	Nivel de signif.	CCI	Intervalo de confianza al 95%	
						Límite Inf.	Límite Sup.
TE 2	15,06						
TE 3	15,96	0,932	0,899	0,445	0,965	0,947	0,977

Tabla 121. Diferencia medias Te T2 y T3

SENSIBILIDAD AL CAMBIO:

Este apartado tiene como objetivo detectar las diferencias en la evaluación de 11 muestras vocales tras aplicarles tratamiento quirúrgico. Se trata de un subgrupo de 22 muestras en total que se evalúan antes y después de la aplicación de tratamiento quirúrgico mediante fonomicrocirugía laríngea. Las 11 muestras son seleccionadas por presentar patologías benigna de la laringe, en concreto lesiones exudativas del espacio de Reinke como: nódulos fibrosos o pólipos, por lo que representan una muestra homogénea de pacientes.

Se exponen los resultados para cada parámetro del CAPE-V y del GRABS. Para la **severidad general** se detectan diferencias significativas con el GRABS y con el CAPE-V sin embargo, para los parámetros **Aspereza** y **Calidad Aérea**, solo se observan diferencias significativas con el CAPE-V mientras que el GRABS no detecta diferencias. Para el parámetro **tensión** no se detectan diferencias significativas con ninguno de los dos protocolos. Los parámetros **tono** e **intensidad** no se reflejan en este apartado por varios motivos. Primero y fundamental, las muestras elegidas no se caracterizan por tener alterados estos parámetros. Segundo, no se realiza ningún tratamiento específico para realizar cambios en el tono o intensidad de la voz, por lo que no son esperables cambios entre las dos evaluaciones. Y tercero, estos parámetros no tienen homólogo en la escala GRABS y la detección del cambio de forma aislada con el CAPE-V no nos permitiría comparación posible.

En los resultados que se exponen solo vamos a tomar en cuenta la existencia o no de diferencias significativas entre las calificaciones pre y postoperatorias, no de su magnitud. Como se ha expuesto previamente, el cálculo de la magnitud de las diferencias

requiere la elección de una población de tamaño y características adecuadas, así como el diseño estadístico apropiado.

El presente estudio no tiene por objetivo evaluar la magnitud de las diferencias para el tratamiento aplicado, sino que estamos validando nuestra herramienta para la detección de cambios.

Tarea 1, 2 y 3	Diferencias pre y postoperatorias	
	<u>CAPE-V</u>	<u>GRABS</u>
Severidad General	Si	Si
Aspereza	Si	No
Calidad Aérea	Si	No
Tensión	No	No

Tabla 122. Resumen de las diferencias halladas para todas las tareas evaluadas

SENSIBILIDAD AL CAMBIO PARA LA SEVERIDAD GENERAL:

Existen diferencias significativas entre la población antes y después de la cirugía con ambos test para las tres tareas de evaluación.

TAREA 1

SG	Medias	Grado	Medias	Diferencias		Significación
Pre	68,86	Pre	2,36	SG	27,727	0,000
Post.	41,14	Post	1,68	Grado	0,682	0,004

Tabla 123. Severidad General/Grado Tarea 1

TAREA 2

SG	Medias	Grado	Medias	Diferencias		Significación
Pre	46,82	Pre	2,36	SG	31,818	0,000
Post	15,00	Post	1,68	G	0,682	0,004

Tabla 124. Severidad General/Grado Tarea 2

TAREA 3

SG	Medias	Grado	Medias	Diferenciación		Significación
Pre	50,68	Pre	2,36	SG	38,864	0,000
Post	11,82	Post	1,68	G	0,682	0,004

Tabla 125. Severidad General/Grado Tarea 3

SENSIBILIDAD AL CAMBIO EN LA ASPEREZA

Se observan diferencias significativas entre las muestras pre y postoperatorias para la aspereza con el protocolo CAPE-V. Sin embargo, las diferencias no son significativas con el protocolo GRABS.

TAREA 1

ASP	Medias	R	Medias	Diferencias	Significación
Pre	47,73	Pre	1,55	ASP 23,864	0,002
Post	23,86	Post	0,91	R 0,636	0,007

Tabla 126. Aspereza/Roughness Tarea 1

TAREA 2

ASP	Medias	R	Medias	Diferencias	Significación
Pre	33,86	Pre	1,55	ASP 25,000	0,000
Post	8,86	Post	0,91	R 0,636	0,007

Tabla 127. Aspereza/Roughness Tarea 2

TAREA 3

ASP	Medias	R	Medias	Diferencias	Significación
Pre	38,18	Pre	1,55	ASP 31,364	0,000
Post	6,82	Post	0,91	R 0,636	0,007

Tabla 128. Aspereza/Roughness Tarea 3

SENSIBILIDAD AL CAMBIO PARA LA CALIDAD AÉREA

Las diferencias entre las medias de ambos grupos adquieren significación con el parámetro CA del protocolo CAPE-V. Sin embargo, con el GRABS la diferencia entre los grupos no alcanza valores significativos.

TAREA 1

CA	Medias	B	Medias	Diferencias	Significación	
Pre	57,27	Pre	1,86	CA	23,182	0,000
Post	34,09	Post	1,41	B	0,455	0,038

Tabla 129. Calidad Aérea/Breathiness Tarea 1

TAREA 2

CA	Medias	B	Medias	Diferencias	Significación	
Pre	35,23	Pre	1,86	CA	25,227	0,000
Post	10,00	Post	1,41	B	0,455	0,038

Tabla 130. Calidad Aérea/Breathiness Tarea 2

TAREA 3

CA	Medias	B	Medias	Diferencias	Significación	
Pre	38,41	Pre	1,86	CA	31,136	0,000
Post	7,27	Post	1,41	B	0,455	0,038

Tabla 131. Calidad Aérea/Breathiness Tarea 3

SENSIBILIDAD AL CAMBIO PARA LA TENSIÓN:

Los resultados no muestran diferencias para el parámetro tensión entre la evaluación preoperatoria y la postoperatoria con ninguno de los dos protocolos.

Hay que destacar que la muestra estudiada se elige de forma aleatoria y no se caracteriza específicamente por sus alteraciones en la tensión, por tanto puede que no sea una muestra representativa para el estudio de la tensión. De hecho las medias registradas son muy bajas, lo que verifica que las muestras carecen de afectación en el parámetro tensión. Ésta puede constituir la razón fundamental por la que no se detecten cambios en este parámetro independientemente del efecto de la escala.

TAREA 1

Tensión	Media	Strain	Media	Diferencias	Significación
Pre	30,45	Pre	0,73	Tensión 12,727	0,044
Post	17,73	Post	0,41	Strain 0,318	0,090

Tabla 133. Tensión/*Strain* Tarea 1

TAREA 2

Tensión	Media	Strain	Media	Diferencias	Significación
Pre	6,36	Pre	0,73	Tensión 1,818	0,550
Post	4,55	Post	0,41	Strain 0,318	0,090

Tabla 134. Tensión/*Strain* Tarea 2

TAREA 3

Tensión	Media	Strain	Media	Diferencias	Significación
Pre	11,36	Pre	0,73	Tensión 9,545	0,029
Post	1,82	Post	0,41	Strain 0,318	0,090

Tabla 135. Tensión/*Strain* Tarea 3

Capítulo 6
DISCUSSION

Las escalas de calificación auditivo perceptual de la voz se utilizan en la clínica para evaluar la calidad vocal dentro de una valoración multidimensional de la voz. A pesar de que en la actualidad pueden realizarse análisis de la señal vocal, medidas de la presión a nivel laríngeo y exploraciones estroboscópicas completas, estos parámetros tienen limitaciones en cuanto a su correlación con el nivel de afectación clínica. Además, se ven totalmente sesgados en pacientes con determinadas alteraciones de la calidad vocal, como por ejemplo, las voces caóticas. Es en este subgrupo donde la evaluación basada en la sensación psicoacústica del observador tiene un papel más importante. Por este motivo, siempre es recomendable completar el estudio de la voz con protocolos de evaluación vocal.

Para tal fin en la práctica clínica habitual existen múltiples herramientas de calificación perceptual de la voz que se basan en las recomendaciones y los estudios de expertos que se han difundido a nivel internacional, sin que exista homogeneidad en los criterios de evaluación o en la metodología de su aplicación. Cada protocolo incluye determinados parámetros de evaluación y sistemas de calificación, siendo entre sí métodos de evaluación heterogéneos y difícilmente comparables. En general, tampoco existen pautas concretas sobre la aplicación de los protocolos de análisis vocal durante la realización de las calificaciones.

Entre los métodos de análisis perceptual de la voz más extendidos, citaremos aquellos revisados durante la reunión de la *American Speech-Language-Hearing Association* (ASHA) en 2002, celebrada con el objetivo de crear un protocolo de evaluación universal y que se utilice como *gold standard* en la calificación perceptual de la voz. De esta forma, en Suecia existe una larga tradición en el estudio de la voz a través del *Stockholm Voice Evaluation Approach (SVEA)*⁴⁶. En USA, los especialistas en voz suelen utilizar el *Buffalo Voice Profile*⁴⁷ no obstante, en Reino Unido el sistema de calificación más extendido es el *Vocal Profile Analysis* basado en las teorías de Laver⁴⁸. Sin embargo, el *GRABS*, escala descrita por Hirano en 1981 y cuyo nombre corresponde a las siglas de los parámetros que evalúa (*Grade, Roughness, Asthenia, Breathiness* y *Strain*) se recomienda en muchos países y se utiliza con el fin de

realizar calificaciones homogéneas y hacerlas comparables a la hora de publicarlas. Revisando la literatura, podríamos citar hasta 57 publicaciones que debaten aspectos relevantes a calificar en la calidad de la voz y elaboran recomendaciones a tener en cuenta para la evaluación⁴⁹.

A pesar de los esfuerzos por encontrar un test universalmente válido, existen inconvenientes importantes que han de ser tenidos en cuenta, tales como: la falta de terminología estandarizada para describir y clasificar las patologías vocales, la ausencia de una definición precisa de voz normal, y la escasa consistencia y/o validez inherente a las calificaciones vocales debido a su subjetividad.

Por este motivo, los métodos de análisis vocal utilizados pueden tener arraigo en la práctica clínica habitual pero carecen, en muchos casos de una metodología definida o referencias vocales que aseguren un entrenamiento homogéneo de los calificadores, además de contar con sistemas dispares y escasamente comparables para la evaluación. Por todo esto, los métodos perceptuales son instrumentos de medida poco robustos con limitaciones en el uso para la comparación de resultados entre distintos tratamientos vocales.

Este trabajo se dedica al estudio del *Consensus Auditory-Perceptual Evaluation of Voice* o denominado a través de sus siglas CAPE-V, que surge en 2002 en el citado contexto científico, fruto de la reunión de la División de Voz y patología vocal de la ASHA. El objetivo de los especialistas allí reunidos fue construir un sistema de calificación completo, preciso y consensuado, que permitiese reflejar de la forma más objetiva posible todos los aspectos multidimensionales de la calificación vocal y que sea aceptado y utilizado por la amplia mayoría de profesionales en la patología de la voz.

Dentro de los elementos específicos que distinguen al CAPE-V como nuevo sistema de calificación destaca una descripción detallada, a modo de instrucciones sobre la aplicación del protocolo, la recogida de datos y el sistema de calificación que se incluye como apéndice en la descripción original del protocolo²⁰. Según se demuestra, la homogeneidad de criterios en la aplicación del cuestionario mejora la consistencia entre los distintos observadores⁵⁰.

El CAPE-V estipula que deben calificarse tres tareas vocales específicas: las vocales sostenidas /i/ y /e/, la lectura de seis frases con contextos lingüísticos concretos y el discurso natural mediante la respuesta a una pregunta *standard* (Por ejemplo: “¿Cómo afecta su problema vocal a su vida cotidiana?” o “¿Cuáles son sus aficiones?”).

La evaluación mediante la fonación de vocales en un tono adecuado y confortable, permite escuchar la voz del paciente sin influencias de la articulación de palabras, además de servir para realizar el análisis acústico y obtener parámetros objetivos como: jitter, shimmer, F0... La calificación mediante frases fonéticamente adaptadas permite evaluar aspectos específicos y concretos del habla conectada y de la influencia de la co-articulación vocal en un entorno lingüístico controlado. Por este motivo, el protocolo CAPE-V debe ser fonéticamente adaptado al español y no sería correcto realizar una traducción literal del texto, puesto que se perdería el entorno lingüístico y fonético específico, requerido en la evaluación.

Nuestro trabajo comienza en la adaptación al idioma español de las frases en inglés descritas en el documento original, tarea que hasta el momento no había sido realizada, imposibilitando el uso del CAPE-V en nuestro medio. Contamos para ello, con la colaboración de una logopeda, que ha realizado el trabajo de adaptación lingüística de las frases, basándose en la descripción fonética incluida en el artículo original del CAPE-V²⁰.

La primera frase (“Lucía ojeó una pajarita y una blusa amarillas”) se utiliza para examinar la influencia co-articulatoria de las vocales /a, i, o, u/. La segunda frase (“Marta multa mi moto más mágica”) proporciona el contexto de evaluación para los ataques glóticos suaves y las transiciones de sorda a sonora. La frase tercera (“La llave brilla en la mano”) permite la calificación de fonemas sonoros, la detección de espasmos vocales y la habilidad para la conexión de las palabras. Frase cuarta (“Irene adora hacer huevos al horno”) contiene palabras que se inician en vocal y evalúan los ataques glóticos duros. La quinta frase (“Mamá me mimma una mano”) evalúa las consonantes nasales permitiendo detectar la existencia de hiponasalidad. La sexta frase (“Ata tu zapato a tu pata”) se caracteriza por la ausencia de consonantes nasales que permite evaluar la presión intraoral, la posible hipernasalidad y la emisión nasal de aire. Estas frases ponen de manifiesto situaciones fonéticas comprometidas que nos

ayudarán a detectar distintos tipos de alteraciones vocales en relación a la capacidad de resonancia del aparato fonador o la capacidad de articulación vocal, que no son valorables mediante el estudio de vocales sostenidas pero que permiten completar la calificación.

La tercera tarea constituye el discurso conversacional que en este protocolo se define como tarea específica para su evaluación. Esta tarea sustituye de un modo formal y protocolizado a la evaluación del habla natural del individuo que suele realizarse durante la anamnesis del paciente, que de forma clásica representa la primera muestra vocal con la que tiene contacto el especialista permitiendo que el paciente se exprese cómodamente sin interrupciones.

A pesar que uno de los importantes atractivos del CAPE-V, y su elemento distintivo principal, son las distintas tareas de fonación que incorpora y que deben analizarse por separado, la mayor parte de los artículos existentes publican en sus resultados puntuaciones medias sin especificar las diferencias en relación a la tarea evaluada y en otros casos sin mencionar respecto a que tarea se basa la calificación. Nuestro trabajo incluye calificaciones independientes para cada tarea, lo que consideramos que es uno de los ejes fundamentales del estudio, ya que nos permitirá evaluar las relaciones entre las distintas tareas. Puede que esta información tenga escasa relevancia a nivel clínico, pero es una característica interesante e importante a resaltar en los estudios de investigación sobre la voz.

Una vez adaptado a nuestro idioma, el siguiente paso es validar el CAPE-V como herramienta útil y fiable en la valoración perceptual de la voz, asegurándonos que refleje las características vocales relevantes de la voz humana, es decir, demostrar la validez de constructo. De esta manera, se pretende conseguir un protocolo de evaluación de la voz integral, completo y universalmente validado por los especialistas. Para conseguir este último paso es necesaria su adaptación e implantación en los países de habla hispana, que no había sido realizado hasta el momento.

El CAPE-V incorpora seis parámetros específicos para la calificación perceptual de la voz, que han sido exhaustivamente definidos y estudiados en publicaciones

internacionales relevantes sobre la voz desde hace décadas^{12,47,51,52}, por este motivo han sido seleccionados para ser incluidos en este sistema de calificación. Estos parámetros son: la severidad general, la aspereza, la calidad aérea o cantidad de aire, la tensión, el tono y la intensidad.

El CAPE-V permite flexibilizar e individualizar la calificación mediante la posibilidad de añadir otras características de interés, que se denominan *parámetros en blanco* y que no aparecen de forma tan frecuente en la población general, tales como; el temblor, los espasmos, los *falsetto* o el *vocal fry*. En nuestro estudio se han obviado estos parámetros en blanco, puesto que constituyen características adicionales que otorgan versatilidad al instrumento, pero se escapan del objetivo del presente trabajo, que es demostrar la validez empírica y de constructo de los parámetros definidos en el protocolo original del CAPE-V. Otra de las ventajas de la nueva herramienta, es el apoyo de las calificaciones en una escala híbrida que cuenta con una representación visual analógica con marcas visuales que se combinan con una escala numérica con un recorrido de 100 mm. Se utilizará para la valoración de los seis parámetros vocales mediante la colocación de una marca sobre la escala. El extremo izquierdo refleja la voz normal y el extremo derecho la afectación máxima del parámetro en cuestión. La medida desde el extremo izquierdo de la línea hasta la marca realizada representará en milímetros la calificación del parámetro, según la afectación valorada. Además, se constituyen tres intervalos visuales dentro de la escala que corresponden a diferentes grados de afectación: leve, medio y grave. Con estas características, el CAPE-V ha demostrado mejores valores de consistencia en las calificaciones realizadas y es necesario destacar que una parte importante que facilita estos resultados es la escala híbrida utilizada.

Aunque existe controversia en este aspecto, las herramientas de calificación con apoyo en escalas visuales analógicas ofrecen calificaciones más consistentes que aquellas que no lo tienen, como el GRABS por ejemplo^{49,53}.

El siguiente objetivo del estudio es demostrar la validez empírica del CAPE-V a través de la comparación de los resultados con el *gold standard*, el sistema GRABS²⁸.

El GRABS fue publicado en 1981 por Hirano y utilizado inicialmente por la *Japan Society of Logopaedics*. Posteriormente se ha extendido su uso hasta constituir una de las escalas de valoración perceptual más utilizada en todo el mundo. Consiste en una escala ordinal compacta que consta de 4 rangos de calificación (del 0 = normal al 3 = grave). No se describe un protocolo específico de entrenamiento para el análisis, ni normas de aplicación determinadas o sistemas visuales de apoyo para mejorar la consistencia interna de las calificaciones. Está basada en la calificación de cinco parámetros vocales: el grado (*Grade*), la aspereza (*Roughness*), la cantidad de aire (*Breathiness*), astenia (*Asthenia*) y tensión (*Strain*). Entre éstos, aquéllos que mejor consistencia tienen son: el grado, la cantidad de aire y la astenia^{54,55}. Por las características ordinales de las variables que incorpora, no es posible utilizar test paramétricos en los estudios de investigación sobre la voz y esto condiciona la potencia o robustez de los resultados obtenidos. Por su amplia difusión clínica se ha constituido como la principal prueba de referencia en la calificación perceptual, pero no es un sistema de calificación óptimo y tiene puntos débiles que afectan a la consistencia del estudio. Está sujeto a una importante variabilidad relacionada con: el entrenamiento de los observadores, las condiciones de evaluación de la muestra y las limitaciones propias del sistema de calificación ordinal con un recorrido corto que limita sus resultados como patrón de referencia. También hay que reconocer que hasta el momento no existe ningún método de calificación con una consistencia perfecta. (Gerrat & Kreiman, 2000; Gerrat, Kreiman, Antonazas-Barroso, & Berke, 1993)

En suma, el GRABS es una escala rígida y de escaso recorrido, con un potencial reducido para expresar la variedad de matices de la voz humana y con limitaciones en la evaluación de los cambios vocales^{53,56}. Este hecho es otro sesgo importante que penaliza al GRABS como *gold standard*.

En la revisión de estudios previos que comparan los dos protocolos, es obligado citar el estudio de Karnell en 2007⁵⁰, que estudia las muestras vocales correspondientes a 34 pacientes y las evalúa mediante el GRABS y el CAPE-V observando una buena correlación entre ambos. En congruencia con lo expuesto por Karnell, la validez interobservador que alcanza el protocolo CAPE-V en nuestro estudio es mejor que la obtenida con el GRABS, variando en un rango entre 0,83 - 0,93. Este hecho puede estar en relación con: la mejor consistencia que presentan las herramientas de

calificación que se apoyan en una escala analógica visual⁵³ y con las instrucciones sobre la forma de aplicar la herramienta de calificación, que evita sesgos sistemáticos en la aplicación del instrumento por cada observador. Por este motivo el nuevo protocolo que pretendemos validar al español constituye una herramienta de calificación perceptual de la voz robusta y consistente, diseñada para mejorar los fallos que tienen otros protocolos de calificación.

A diferencia del GRABS y otros protocolos de evaluación, el CAPE-V incorpora en la evaluación vocal rutinaria el tono y la intensidad. Estos parámetros no se incluyen como parámetros clásicos en los protocolos de valoración perceptual (por ejemplo el GRABS) y por ello la comparación de series de resultados amplios en la literatura es limitada⁵⁶. Los artículos que los incluyen, demuestran que son parámetros poco consistentes y que presentan mucha variabilidad inter- e intra-observador^{28,63}. Las alteraciones del tono y de la intensidad no son frecuentes en la población general y su calificación es controvertida porque no se han definido con exactitud los límites de la normalidad para estos parámetros. Su calificación se realiza de acuerdo al entorno cultural y al entrenamiento del observador, según se considere la voz en relación a la edad, género o contexto social del paciente. Para la evaluación de la intensidad, tiene mucha importancia la calidad de la muestra vocal analizada. Para el tono, existen situaciones puntuales y concretas en las que la alteración de este parámetro se pone de manifiesto y para los que, en la actualidad no existen sistemas de evaluación adaptados como son los trastornos de la identidad sexual. En nuestro estudio las muestras no presentan alteraciones significativas en el tono y la mayoría se clasifican como normales en relación a este parámetro. La consistencia para el tono es superior al 80% mientras que para la intensidad el acuerdo entre observadores está alrededor del 60%. La interpretación de estos resultados puede resultar controvertida pero es congruente con los resultados obtenidos en otros estudios publicados⁵⁶.

Las calificaciones de las 89 muestras vocales fueron realizadas por dos especialistas en patología vocal con una larga experiencia y trayectoria profesional en el tratamiento de la misma y en el uso del GRABS como escala de valoración perceptual. Los observadores proceden de centros hospitalarios diferentes, por lo que podemos asumir que no tienen el mismo entrenamiento para la calificación perceptual de la voz, ni tampoco están en contacto con la misma población diana. Por este

motivo, nos parece representativo elegir a estos dos calificadores para el estudio, como fiables representantes de la comunidad de expertos en fonología de nuestro medio.

En nuestro estudio se incluyen un total de 89 muestras vocales procedentes de una selección de 67 pacientes, entre ellos; 50 son pacientes disfónicos captados para el estudio desde la consulta de voz y los 17 restantes corresponden a voluntarios sanos incluidos en el estudio como muestras de voz normales. El criterio de inclusión para los voluntarios sanos fue la ausencia de patología laríngea conocida o percepción de voz disfónica por el propio sujeto. Los pacientes incluidos inicialmente en el estudio presentaban disfonías moderadas causadas por patología benigna de la laringe, de esta forma al realizar el análisis perceptual de la voz y los cálculos estadísticos, los resultados aparecen sesgados por falta de extremos patológicos y variabilidad representativa de la población general. Para corregir estos sesgos fue necesario ampliar el tamaño muestral, así como el rango de patologías vocales estudiadas para enriquecer la muestra. De esta manera, se realizó un cálculo del tamaño muestral necesario para obtener resultados robustos, teniendo en cuenta la naturaleza subjetiva de las variables estudiadas y en relación a estudios con características similares ya publicados. Entre las publicaciones consultadas una de las más destacadas es el estudio de validación del protocolo CAPE-V publicado por Zraick et al. en 2011²⁸ que incluye 74 muestras vocales (de entre las cuales 37 son disfónicas, 22 correspondientes a controles sanos, 11 muestras repetidas y 4 voces para entrenamiento) y el estudio de Karnell et al 2007⁵⁰, que compara los resultados obtenidos del CAPE-V y el GRABS para un total de 103 pacientes con disfonías orgánicas y funcionales de variada etiología. En nuestro estudio incorporamos 89 muestras procedentes de 67 pacientes, todos dieron su consentimiento expreso y son mayores de edad, con un rango de edades comprendidas entre 21 y 80 años. Se incluye un subgrupo de 22 pacientes en los que se realiza una primera evaluación vocal y tras el tratamiento quirúrgico se realiza otra calificación perceptual de la voz con el objetivo de determinar la sensibilidad al cambio del CAPE-V frente al GRABS.

Las grabaciones se realizaron en una habitación sonoamortiguada garantizando la ausencia de ruido ambiental con una grabadora digital localizada a unos diez centímetros del individuo y colocada en ángulo de 45°. En general, el desarrollo de las

grabaciones transcurrió sin incidencias, aunque los pacientes de mayor edad colaboraron con mayor dificultad fundamentalmente por problemas de visión o comprensión de la dinámica de las tareas vocales. Estos inconvenientes se subsanaron a través de la explicación clara y precisa de las tareas a realizar y la estrecha colaboración con el paciente para la lectura o repetición, en algunos casos, de las frases adaptadas para la tarea 2. Esto conlleva el consumo de más tiempo para la realización de la prueba, que en comparación con el GRABS podría constituir un inconveniente. Frente al CAPE-V una de las ventajas que dispone el GRABS para su difusión internacional es su facilidad y rapidez de aplicación en una consulta rutinaria de voz, sin consumir un tiempo excesivo, ni necesitar una metodología compleja o un esfuerzo importante por parte del paciente o del explorador⁵¹.

La muestra escogida para la realización de este estudio se realizó basándose en la severidad global de las muestras vocales con el objetivo de incorporar un amplio rango de voces en relación a este parámetro. Las calificaciones para la severidad general son aquellas que obtienen mejores resultados para todos los ítems de nuestro estudio (consistencia intra e interobservador y validez de criterio) hecho que probablemente también esté en relación con la mayor consistencia de este parámetro. La aspereza o la calidad aérea adquieren valores relativamente buenos, pero siempre algo peores que la severidad general, esto puede deberse a limitaciones del tamaño muestral en relación a estos parámetros o a su menor consistencia, ya citada en la literatura⁵⁷.

El objetivo principal de este estudio es el de comparar los resultados de calificación vocal obtenidos mediante dos herramientas clínicas utilizadas simultáneamente, el GRABS y el CAPE-V, permitiendo así validar esta última como protocolo de calificación en el idioma español. En este contexto, es necesario hacer referencia a las diferencias técnicas⁵⁰ entre ambas herramientas, que podrían afectar a su consistencia; el criterio de referencia o GRABS consta de una escala ordinal con cuatro rangos de evaluación mientras que el CAPE-V se utiliza como una variable cuantitativa discreta con apoyo en rangos visuales.

Sería lógico pensar que si las diferencias técnicas de las que hablamos no tuviesen impacto en las calificaciones, los resultados para ambos test serían similares y

consistentes. Este es un tema ampliamente discutido en la literatura, y uno de los principales sesgos atribuidos a la escala GRABS; la penalización en la consistencia de la herramienta atribuida al sistema de medida o calificación. De tal forma que, la variación en la evaluación vocal atribuible a un punto sobre una escala de cuatro rangos representaría una variabilidad de un 25%, mientras que, la variación en la calificación de un punto sobre una escala en la que existen 100 rangos de calificación representaría una variación de un 1%. Podemos decir que la escala GRABS es una escala demasiado compacta, con un corto recorrido que dificulta la expresión o evaluación correcta de las variaciones de los parámetros vocales, limitando la calificación por motivos técnicos. Este inconveniente se subsana en el CAPE-V, al disponer de un rango más amplio de evaluación y contar con el apoyo en un sistema analógico visual que aporta una doble ventaja: reflejar con más facilidad los matices atribuibles a los parámetros vocales y mayor consistencia como sistema de medida ⁵⁸.

Teniendo en cuenta estas especificaciones técnicas, es importante hablar sobre la interpretación de los índices de correlación obtenidos durante el estudio. Al tratarse de calificaciones perceptuales y subjetivas no pretendemos conseguir resultados matemáticamente exactos, por lo tanto contamos con rangos de interpretación flexibles.

Una de las importantes ventajas que incorpora el CAPE-V y que defendemos como instrumento de valoración, es la evaluación de la voz mediante tres tareas de fonación diferentes; las vocales sostenidas, el habla conectada en un entorno fonético controlado y el discurso natural. La utilización de las vocales sostenidas en la evaluación integral de la voz ha sido tradicionalmente muy utilizada en la clínica, probablemente por la facilidad y consistencia para el cálculo de las medidas acústicas objetivas y el espectrograma sin la influencia de la articulación o los inicios y finales de fonación. También existen otras razones, como la facilidad y rapidez de obtención durante la consulta médica sin consumir mucho tiempo ni requerir técnicas o recursos complejos. En nuestra experiencia durante la aplicación del protocolo la utilización de vocales sostenidas es la tarea más fácil y rápida de obtener. La evaluación a través de las frases genera conflictos en la ejecución y consume más tiempo, sobretudo en personas de avanzada edad. En algunos casos, no es posible utilizar tarjetas con las frases como propone el protocolo original, y el médico debe implicarse activamente

formulando la frase para que el paciente la repita después.

Existen además otras razones técnicas, puesto que las vocales sostenidas se generan en un entorno controlado y estable para la calificación mientras que el habla natural incorpora cambios rápidos, inicios o terminaciones vocales bruscas y fluctúa en relación a la prosodia⁵⁹. Las vocales sostenidas no se afectan por las pausas vocales, el estrés o el contexto fonético. En nuestro estudio podemos observar una penalización en los resultados atribuidos a la tarea 2 en relación a la validez de criterio que solo puede explicarse debido al entorno fonético forzado y comprometido en el que se desarrolla la calificación, puesto que los resultados en el mismo apartado para la tarea 3 obtienen mejores resultados (ICC tarea 2: 0,789 Vs ICC tarea 3: 0,866). Se trata de sutiles variaciones que sería necesario estudiar con mayor profundidad y en condiciones de estudio más controladas. En este contexto existen pocas publicaciones que comparen el discurso natural frente al habla conectada en un entorno fonético controlado. Por otro lado, la evaluación del discurso natural puede verse afectada por parámetros externos como; el temblor, la actitud o el estado de ánimo durante la exploración... Sin embargo, el habla conversacional es mucho más representativa de la voz habitual del individuo.

Uno de los artículos más relevantes en este aspecto fue publicado por Zraick et al. en 2005 y evalúa un total de 29 muestras vocales mediante vocales sostenidas, discurso natural y voz conectada, determinando que existen diferencias significativas entre la primera tarea vocal y las segundas, siendo las calificaciones más severas con las vocales sostenidas. No observa diferencias significativas entre las dos tareas relativas al habla conectada: el discurso natural y la lectura de un pasaje literario⁶⁰. Estos hallazgos se encuentran en congruencia con los estudios de Wolfe et al 1995⁶¹. Existen numerosas publicaciones que estudian la comparación entre ambas tareas de fonación sin lograr consenso en la materia⁶². A pesar de todo esto, y con el objetivo de simplificar, no es raro encontrar artículos que adaptan la metodología del CAPE-V y simplifican los resultados evitando calificar las tres tareas vocales, como el estudio de Kelchner en la población pediátrica, donde sólo utiliza la tarea 2 para la calificación vocal⁵⁶. En nuestro estudio observamos resultados muy similares a los publicados por Zraick et al.

La comparación de nuestros resultados se realiza con el *Gold Standard* actual que es

el GRABS, como mencionamos anteriormente no es un patrón de referencia óptimo, aunque es difícil conseguir niveles de consistencia y validez elevados en la calificación perceptual de la voz. Obtenemos diferencias significativas entre las calificaciones de la tarea 1 frente a las calificaciones de la tarea 2 y 3. Como es de esperar, las tareas correspondientes al habla conectada tienen un comportamiento similar; los niveles de acuerdo son superiores al 0,94 para todos los parámetros evaluados. En relación con la tarea 1 se obtienen diferencias significativas frente a las tareas 2 y 3, pero no condicionan desacuerdo con las valoraciones previas manteniendo una fuerza de concordancia por encima de 0,85. De tal modo que en relación al parámetro severidad general, observamos que las calificaciones adquieren resultados más patológicos en la tarea 1 o fonación de vocales, siendo las calificaciones correspondientes a las dos tareas siguientes de menor gravedad y similares entre sí. Los datos obtenidos para el parámetro Grado con la escala GRABS, traducen un comportamiento similar a los resultados obtenidos con la tarea 1 del CAPE-V, con tendencia a sobredimensionar la calificación, puesto que el grupo con más peso es el correspondiente al rango 3 o máximo de la clasificación.

Por otro lado, y en relación a la validez de criterio, los resultados obtenidos con el CAPE-V para la tarea 1 son similares a los obtenidos con el GRABS garantizando un nivel de acuerdo muy elevado entre los dos test (ICC T1: 0,965). Sin embargo, los resultados de la T2 y T3 reflejan una menor afectación en la gravedad de la voz en comparación con la T1 y esto disminuye el nivel de acuerdo en relación al GRABS (ICC tarea 2: 0,789; ICC tarea 3: 0,866). En este último caso, la fuerza de la concordancia es de menor magnitud que la obtenida, pero se mantiene en niveles de acuerdo intermedio/alto. Mención aparte requiere el comportamiento de la tarea 2, que refleja el peor nivel de acuerdo con el GRABS de las 3 tareas evaluadas de forma muy evidente (CCI: 0,13). Puede deberse a la penalización atribuible al contexto fonético dirigido que potencie las alteraciones vocales de las muestras y artefacte la percepción de la calidad aérea.

Separando los resultados por parámetros, es la severidad general aquél que obtiene mejores resultados (ICC: 0,78- 0,96). La aspereza, la calidad aérea y la tensión reflejan una buena correlación entre el CAPE-V y el GRABS pero con menos fuerza que la severidad general y unos rangos en el coeficiente de acuerdo más amplios.

De esta forma quedaría validado el protocolo CAPE-V al idioma español, quedando patente su concordancia con los resultados obtenidos con el actual criterio de referencia que es el GRABS. El CAPE-V aporta mayor riqueza en la calificación perceptual puesto que ofrece una evaluación integral de la calidad de la voz, enriqueciendo y completando la calificación.

La consistencia intraobservador detectada en nuestro estudio es muy elevada, en todas las tareas y para todos los parámetros. Esto refrenda la profesionalidad y buen entrenamiento de nuestro observador, puesto que alcanza correlaciones superiores al 90% en todos los casos^{28,56}.

En cuanto a la fiabilidad interobservador de nuestro estudio tiene rangos variables de los que podemos sacar varias conclusiones. Entre los observadores los parámetros más consistentes son la severidad general, la aspereza y la calidad aérea. Éstos independientemente de la tarea, adquieren valores elevados y reflejan una buena fiabilidad o consistencia. El resto: la intensidad, el tono y la tensión, toman cifras variables comprendidas en un rango de consistencia intermedio. A la luz de los datos publicados en la literatura existente son conclusiones confirmadas por otros autores: la severidad, la aspereza y la calidad aérea son parámetros de consistencia probada. La intensidad y el tono son ítems de difícil calificación debido a la dificultad de definir y cuantificar sus alteraciones⁵⁴⁶³⁶⁴⁶⁵.

Autores como Karnell⁵⁰ han publicado estudios de consistencia interobservador con resultados similares a los de nuestra serie, y sugieren que el protocolo CAPE-V tiene más sensibilidad a las pequeñas diferencias de calificación que la escala GRABS. Esta característica también se destaca en la publicación de Zraick et al, defendiendo unos mejores resultados en consistencia del CAPE-V frente al GRABS.

La variabilidad en los parámetros internos de cada calificador es tan amplia que compromete la fiabilidad de las valoraciones. Se discute si la evaluación de las muestras por parte de un número elevado de calificadores puede añadir más consistencia a los datos o restar consistencia. Según Kreiman y Gerratt¹⁷ si provienen de equipos de trabajo distintos añaden más variabilidad a las calificaciones debido a la

flexibilidad de sus parámetros internos de calificación, a su diferente entrenamiento y práctica clínica.

Por otra parte, la fiabilidad intraobservador también varía en función de la tarea a evaluar. Se comprueba en nuestro estudio que la valoración de vocales sostenidas tiene valores más consistentes que la valoración del discurso natural. Podríamos decir que esta afirmación está en relación con el mayor uso de la vocal sostenida como medio, en nuestro entorno clínico, para la valoración perceptual de la voz. Los calificadores profesionales suelen estar familiarizados con la calificación de vocales sostenidas y en algunos casos se apoyan en el estudio del espectrograma de banda estrecha, hecho que mejora la consistencia de las calificaciones⁴⁰. Aunque la discusión entre la calificación de vocales o habla conectada es antigua, en la práctica clínica la calificación del habla conectada es menos frecuente. Concretamente en nuestro caso, los observadores están menos familiarizados con la calificación del discurso natural frente a las vocales. Este hecho justifica los hallazgos referidos y por este motivo podríamos justificar que las dos últimas tareas sean ligeramente menos consistentes.

Otro inconveniente para la calificación es la dificultad en el reconocimiento aislado de los parámetros como; aspereza, cantidad de aire o tensión, que puede resultar más difícil durante el discurso natural.

El habla natural incorpora determinadas características o detalles relevantes de la percepción vocal, como terminaciones e inicios rápidos de la frase, variaciones en la frecuencia fundamental y la amplitud de la onda vocal dependiendo de la entonación y roturas vocales y silencios.

Por lo tanto, puede que la calificación de vocales sostenidas no evalúe aspectos clave de la calidad vocal. Son muchos los autores que afirman que la exclusiva calificación de vocales sostenidas, no es representativa de la valoración perceptual de la voz^{60,66}.

Aunque las medidas objetivas de calificación de la voz tienen importancia, hasta la fecha no existe consenso en su utilidad práctica como medida; tanto para la discriminación entre normal y patológico o para calificar la severidad de un trastorno vocal. Por tanto, hay autores que avalan las calificaciones perceptuales como más seguras (Rabinov, Kreiman, Gerratt y Bielamovicz). Además, las medidas de

perturbación pueden verse influidas por la entonación u otros efectos de modulación vocal.

La severidad general se comporta como el ítem más consistente. Existen diferencias significativas entre las tareas mencionadas, pero los niveles de acuerdo y de correlación adquieren valores muy elevados. Resultados similares obtiene la calidad aérea y la tensión, que se revelan como parámetros muy consistentes, con diferencias entre las distintas tareas vocales. En el caso de este último parámetro, la literatura no avala nuestra afirmación, sin embargo puede que nuestra muestra no sea específicamente representativa de las alteraciones en la tensión. El parámetro aspereza, tiene un comportamiento similar, pero los índices estadísticos no alcanzan niveles de acuerdo tan elevados. Nos sirve para afirmar que las calificaciones tienen validez, pero demuestran que la aspereza no es un parámetro con tanta consistencia y que existirá más variabilidad entre las mediciones de cada tarea.

En cuanto a la sensibilidad al cambio, solo se observan diferencias significativas en los dos protocolos para la severidad general. Probablemente se deba a la elevada consistencia del parámetro, siendo éste el que más se afecta de forma global después de un tratamiento sobre las cuerdas vocales. Para el resto, se observan diferencias significativas solo con el CAPE-V en la aspereza, calidad aérea, intensidad y tono. El GRABS no permite detectar diferencias entre dos poblaciones cuando los cambios son pequeños, solo detecta grandes diferencias en parámetros muy consistentes. Esto ocurre con la severidad general, que es el único parámetro que obtiene diferencias significativas. Por otra parte, el CAPE-V permite demostrar cambios más sutiles, ya que alcanza datos significativos para casi todos los parámetros evaluados. Es una escala más flexible y con mayor sensibilidad para los cambios en los parámetros.

Por otra parte, la aspereza y la calidad aérea son parámetros relacionados con la patología de la cuerda vocal afecta. Una lesión sobre la cuerda que provoque efecto masa imprimirá aspereza a la voz, mientras que una parálisis vocal aumentara el escape de aire durante la fonación. La comparación de la función vocal pre y postoperatoria es importante para evaluar el rendimiento terapéutico y la eficacia del tratamiento aplicado, para ello es necesario un test que nos permita detectar los cambios. Con el protocolo CAPE-V somos capaces de detectar cambios tras un

tratamiento y sin embargo, el GRABS que constituye el *gold standard* no es capaz de detectarlos. Como ya hemos mencionado se trata de una escala rígida y de corto recorrido, que no es capaz de detectar cambios sutiles.

El parámetro tensión no alcanza diferencias significativas con ninguna de las escalas, pero es importante compararlo con los resultados para la intensidad y el tono, que aunque solo se evalúan en el CAPE-V si demuestran datos significativos. A priori esta diferencia puede deberse a la existencia real de diferencias pre y postoperatorias. Sin embargo, el tipo de patologías que se incluyen en este subgrupo de 22 muestras no afectan de forma específica a la tensión, la intensidad o el tono. Por lo tanto en ninguna de ellas sería esperable una diferencia significativa en la puntuación de sus parámetros pre y postoperatorios, en contra de lo que se observa en los resultados. Habrá que tomar precaución en la forma de interpretar los datos^{20,50}.

CONCLUSIONES:

1. El CAPE-V es un protocolo adaptado al español y validado para su uso como protocolo de evaluación auditivo perceptual de la voz normal y patológica. Es una herramienta completa, flexible y sensible a los cambios con capacidad para evaluar de forma precisa el resultado de acciones terapéuticas sobre las cuerdas vocales.
2. La calificación perceptual de la voz mediante distintas tareas vocales enriquece la valoración permitiendo describir la calidad vocal con una visión amplia y muy cercana a la realidad, incorporando características y matices vocales que complementan a los parámetros objetivos y a otros test de evaluación vocal. Se detectan diferencias entre la calificación de las distintas tareas, pero no comprometen la consistencia de la valoración.
3. El sistema de calificación del protocolo CAPE-V se basa en una escala híbrida: numérica y visual. Las referencias visuales mejoran la consistencia de la calificación y el apoyo numérico con amplio recorrido (0 a 100 mm) permite objetivar la amplia riqueza de matices propios de cada muestra vocal.
4. La consistencia del protocolo CAPE-V es mayor que la detectada para el GRABS y tiene resultados más robustos. Por este motivo, la proponemos como escala fiable y precisa para realizar la evaluación perceptual de la voz.

ANEXO A: PROTOCOLO CAPE-V (ESPAÑOL)

PACIENTE:

FECHA:

Tarea 1: Vocales Sostenidas /a:/ y /e:/

Tarea 2: Frases

Nuria ojeó una pajarita y una blusa amarillas

Marta multa mi moto más mágica

La llave brilla en la mano

Irene adora hacer huevos al horno

Mama me mimma una mano

Ata tu zapato a tu pata

Tarea 3: Habla natural ¿Cómo percibe su voz?/¿Cuáles son sus aficiones?

SEVERIDAD GENERAL



ASPEREZA



CALIDAD AÉREA



TENSIÓN



TONO



INTENSIDAD



Epígrafes a utilizar: T1 (tarea 1); T2 (tarea 2); T3 (tarea 3)

C (consistente) I (intermitente)

REALIZADO POR:

ANEXO B: PROTOCOLO ORIGINAL DEL CAPE-V

Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)

Name: _____

Date: _____

The following parameters of voice quality will be rated upon completion of the following tasks:

1. Sustained vowels, /a/ and /i/ for 3-5 seconds duration each.
2. Sentence production:
 - a. The blue spot is on the key again.
 - b. How hard did he hit him?
 - c. We were away a year ago.
 - d. We eat eggs every Easter.
 - e. My mama makes lemon muffins.
 - f. Peter will keep at the peak.
3. Spontaneous speech in response to: "Tell me about your voice problem." or "Tell me how your voice is functioning."

Legend: C = Consistent I = Intermittent
 MI = Mildly Deviant
 MO = Moderately Deviant
 SE = Severely Deviant

				<u>SCORE</u>
Overall Severity	_____	_____	_____	C I <u> /100</u>
	MI	MO	SE	
Roughness	_____	_____	_____	C I <u> /100</u>
	MI	MO	SE	
Breathiness	_____	_____	_____	C I <u> /100</u>
	MI	MO	SE	
Strain	_____	_____	_____	C I <u> /100</u>
	MI	MO	SE	
Pitch	(Indicate the nature of the abnormality): _____			C I <u> /100</u>
	_____	_____	_____	
	MI	MO	SE	
Loudness	(Indicate the nature of the abnormality): _____			C I <u> /100</u>
	_____	_____	_____	
	MI	MO	SE	
_____	_____	_____	_____	C I <u> /100</u>
	MI	MO	SE	
_____	_____	_____	_____	C I <u> /100</u>
	MI	MO	SE	

COMMENTS ABOUT RESONANCE: NORMAL OTHER (Provide description): _____

ADDITIONAL FEATURES (for example, diplophonia, fry, falsetto, asthenia, aphonia, pitch instability, tremor, wet/gurgly, or other relevant terms): _____

Clinician: _____

-
- ¹ Aronson AE. Clinical Voice Disorders. An interdisciplinary approach. New York: Thieme Inc;1990.
- ² Moore JP. Organic voice disorders. New Jersey: Prentice-Hall Englewood Cliffs ;1971.
- ³ Johnson W, Brown SF, Curtis JF, Edney CW, Keaster J. Speech handicapped school children. New York: Harper & Brothers; 1965.
- ⁴ Rosen CA, Murry T. Nomenclature of voice disorders and vocal pathology. *Otolaryngol Clin N Am.* 2000; 33: 1035-45.
- ⁵ Koufman JA, Blalock PD. Classification and approach to patients with functional voice disorders. *Ann Otol Rhinol Laryngol.*1982; 91: 372-377.
- ⁶ Morrison MD, Rammage LA. Muscle misuse voice disorders: Description and classification. *Acta Otolaryngol (Stockh).*1993; 113: 428-434
- ⁷ Hirano M. Estructure and vibratory pattern of the vocal folds. In: Sawashima N, Cooper FS (eds): Dynamic aspects of speech production. Tokyo, University of Tokyo Press, 1977, pp13-27.
- ⁸ Dejonckere PH. A basic protocol for functional assessment of voice pathology. *Eur Arch Otolaryngol.* 2000; 258: 77-82.
- ⁹ Núñez Batalla F, Corte Santos P, Sequeiros Santiago G, Señaris González B, Suárez Nieto C. Evaluación Perceptual de la Disfonía: Correlación con los Parámetros Acústicos y Fiabilidad. *Acta Otorrinolaringol Esp.* 2004; 55: 282-287.
- ¹⁰ Nuñez F, Mate MA. Historia Clínica y valoración subjetiva de la voz (calidad de vida en relación con la voz). En: Patología de la voz. Barcelona: Marge Medica Books; 2013. p111-118.
- ¹¹ Carding, P. N., Wilson, J. A., MacKenzie, K., & Deary, I. J. Measuring voice outcomes: State of the science review. *J Laryngol Otol.* 2009; 123: 823–829.
- ¹² Hirano M. Psycho-acoustic evaluation of voice. In: Arnold GE, Winckel F, Wyke BD (eds) Disorders of Human Communication 5. Clinical Examination of Voice. Vienna, Austria: Springer Verlag; 1981: 81-84.
- ¹³ Oates J, Russell A. Learning voice analysis using an interactive multi-media package: Development and primary evaluation. *Journal of Voice* 1998; 12: 500-512.
- ¹⁴ Wilson DK. Voice problems of children. 3rd ed. Baltimore MD: Williams and Wilkins; 1987.
- ¹⁵ Kreiman J, Gerratt B, Ito M. When and why listeners disagree in voice quality assessment tasks. *J Acoust Soc Am.* 2007;122:2354-64.
- ¹⁶ Kreiman J, Gerratt B. Perception of aperiodicity in pathological voice. *J Acoust Soc Am.* 2005;117:2201-11.
- ¹⁷ Kreiman J, Gerratt B. Sources of listener disagreement in voice quality assessment. *J Acoust Soc Am.* 2000;108(4):1867-76.
- ¹⁸ Kreiman & Gerratt. Validity of rating scale measures of voice quality. *J Acoust Soc Am.* 1998 Sep;104:1598-608.
- ¹⁹ Bielamowicz S, Kreiman J, Gerratt BR, Dauer MS, Berke GS. Comparison of voice analysis systems for perturbation measurement. *J Speech Hear Res.* 1996;39(1):126-34.
- ²⁰ Kempster GB, Gerratt BR, Verdolini Abbott K, Barkmeier-Kramer J, Hillman RE. Consensus Auditory Perceptual Evaluation of Voice: Development of a standardized clinical protocol. *Am J Speech Lang Pathol.* 2009; 18: 124-132.
- ²¹ Callan D, Kent R, Roy N, Tasko S. The use of self-organizing maps for the classification of voice disorders. In: Kent R and Ball M (Eds.), Voice quality measurement. San Diego CA: Singular Publishing Group;2000 p. 103-116.

-
- ²² Gerratt B, Kreiman J. Measuring vocal quality with speech synthesis. *J Acoust Soc Am*. 2001; 110: 2560-2566.
- ²³ Zwicker E, Fastl H, Frater H. Psychoacoustics: facts and models. New York: Springer Verlag; 1999.
- ²⁴ Gescheider GA, Marks LE. Psychophysical scaling. In: HE Pashler y SS Stevens (Eds.), *Steven`s Handbook of Experimental Psychology*, 3rd. Edition, Indianapolis: John Wiley and Sons; 2002. p. 91-138
- ²⁵ Marks L, Algom D. Psychophysical scaling. In: MH Birnbaum (Ed), *Measurement, judgment, and decision making*. San Diego CA: Academic Press; 1998.
- ²⁶ Carding P, Carlson E, Epstein R, Mathieson L, Shewell C. Formal perceptual evaluation of voice quality in the United Kingdom. *Logoped Phoniatr Vocol*. 2000;25(3):133-8.
- ²⁷ Hammamberg B. Pathological voice qualities. Perceptual and acoustic characteristics of a set of Swedish “reference” voices. *Bull Audiophonology*.1992; 8: 39-52.
- ²⁸ Zraick RI, Kempster GB, Connor NP, Thibeault S, Klaben BK, Bursac Z, Glace LE. Establishing validity of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). *Ame J Speech-Language Pathol* 2011; 20: 14-22.
- ²⁹ Dejonckere PH, Lebacq J. Acoustic, perceptual, aerodynamic and anatomical correlations in voice pathology. *J Otorhinolaryngol Relat Spec*. 1996; 58: 326-332.
- ³⁰ Nawka T, Anders LC, Wendler J. Die auditive Beurteilung heiserer Stimmen nach dem RBH-System. *Sprache, Stimmes Gehör* 1994; 18: 130-133.
- ³¹ Alexandre NMC, Guirardello Ede B. Cultural adaptation of instruments utilized in occupational health. *Rev Panam Salud Publica* 2002;11:109-111.
- ³² García de Yébenes MJ, Rodríguez-Salvanés F, Carmona-Ortells L. Validación de cuestionarios. *Reumatol Clin* 2009;5:171-177.
- ³³ Beaton DE, Bombardier C, Guillemin F, Bosi-Ferraz M. Guidelines for the process of cross-cultural adaptation of self-reports measures. *Spine*. 2000;25:3186-3191.
- ³⁴ Carvajal A, Centeno C, Watson R, Martínez M, Rubiales AS. How is an instrument for measuring health to be validated? *An Sist Sanit Navar*. 2011;34:63-72.
- ³⁵ Argimon-Pallas JM, Jimenez-Villa J. *Métodos de investigación clínica y epidemiológica*. 3a. ed. Madrid: Elsevier España; 2004.
- ³⁶ Müller R, Büttner P. A critical discussion of intraclass correlation coefficients. *Stat Med* 1994;13:2465-2476.
- ³⁷ Ramada-Rodilla José María, Serra-Pujadas Consol, Delclós-Clanchet George L. Adaptación cultural y validación de cuestionarios de salud: revisión y recomendaciones metodológicas. *Salud pública Méx*. 2013;55(1): 57-66. Disponible en: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0036-36342013000100009&lng=es.
- ³⁸ Mookink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, *et al*. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63:737-745.
- ³⁹ Aday LA, Cornelius LJ. *Designing and conducting health surveys: a comprehensive guide*. 3rd ed. San Francisco: Jossey-Bass Publisher; 2006.
- ⁴⁰ Núñez-Batalla F, Díaz-Molina JP, García-López I, Moreno-Méndez A, Costales-Marcos M, Moreno-Galindo C, Martínez-Cambor P. El espectrograma de banda estrecha como ayuda para el aprendizaje del método GRABS de análisis perceptual de la disfonía. *Acta Otorrinolaringol Esp*. 2012; 63: 173-179.

-
- ⁴¹ Eadie T, Sroka A, Wright DR, Merati A. Does knowledge of medical diagnosis bias auditory-perceptual judgments of dysphonia? *J Voice*. 2011;25: 420–429.
- ⁴² Yanagihara N. Significance of harmonic changes and noise components in hoarseness. *J Speech Hear Res*. 1967;10(3):531-41.
- ⁴³ Cavalli L, Hirson A. Diplophonia reappraised. *J Voice*. 1999;13:542-556.
- ⁴⁴ Prieto. L, Lamarca R, Casado A. La evaluación de la fiabilidad en las observaciones clínicas: el coeficiente de correlación intraclase. *Med Clin (Barc)*. 1998 7;110(4):142-5.
- ⁴⁵ García Yébenes Prous, M. J. G., Rodríguez Salvanés, F., & Carmona Ortells, L. Sensibilidad al cambio de las medidas de desenlace. *Reumatología Clínica*. 2008. 4(6), 240-247.
- ⁴⁶ Hammarberg B. Pathological voice qualities. Perceptual and acoustic characteristics of a set of Swedish “reference” voices. *Bull Audiophonology*. 1992;8:32-52
- ⁴⁷ Wilson DK. Voice problems of children. Baltimore: Williams & Wilkins; 1987
- ⁴⁸ Carding P, Carlson E, Epstein R, Mathieson L, Shewell C. Formal perceptual evaluation of voice quality in the United Kingdom. *Logoped Phoniat Vocol*. 2000;25:133-138
- ⁴⁹ Kreiman J, Gerratt B, Kempster GB, Erman A, Berke Gs. Perceptual evaluation of voice quality: review, tutorial, and a framework for research. *J Speech Hear Res*. 1993;36:21-40
- ⁵⁰ Karnell M, Melton S, Childes J, Coleman T, Dailey S, & Hoffman H. Reliability of clinician-based (GRBAS and CAPE-V) and patient-based (V-RQOL and IPVI) documentation of voice disorders. *Journal of voice*. 2007; 21:576–590
- ⁵¹ DeBodt MS, Wuyts FL, Van de Heyning PH, Lambrechts L. The perceptual evaluation of voice disorders. *Acta Oto Rhino Laryngologica*. 1996;50:283-291
- ⁵² Fairbanks G. Voice and articulation drillbook. Philadelphia: Harper & Row; 1960
- ⁵³ Wuyts FL, De Bodt MS, Van de Heyning PH. 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*. 1999;13:508–517
- ⁵⁴ DeBodt MS, Wuyts FL, Van de Heyning PH, Croux C. Test-retest study of the GRBAS scale: influence of experience and profesional background on perceptual rating of voice quality. *J Voice*. 1997; 11:74-80.
- ⁵⁵ Dejonckere PH, Remacle M, Fresnl-Elbaz E, Woisard V, Crevier Buchman L, Millet B. Differentiated perceptual evaluation of pathological voice quality: reliability and correlation with acoustic measurements. *Rev Laryngol Otol Rhinol*. 1996;117:219-224.
- ⁵⁶ Kelchner L N, Brehm S, Weinrich B, Middendorf J, deAlarcon A, Levin L, and Elluru R. Perceptual evaluation of severe pediatric voice disorders: rater reliability using the consensus auditory perceptual evaluation of voice. *J Voice*, 2010;24:441-449.
- ⁵⁷ Solomon NP, Helou LB, Stojadinovic A. Clinical versus laboratory ratings of voice using the CAPE-V. *J Voice*. 2011;25: 7-14.
- ⁵⁸ Yiu EM, Murdoch B, Hird K, Lau P. Perception of synthesized voice quality in connected speech by Cantonese speakers. *J Acoust Soc Am*. 2002;112:1091-1101.
- ⁵⁹ Awan SN, Roy N. Toward the development of an objective index of dysphonia severity: a four-factor acoustic model. *Clin Linguist Phon*. 2006;20:35–49.

-
- ⁶⁰ Zraick R, Wendel D, Smith-Olinde L. The effect of speaking task on perceptual judgment of the severity of dysphonic voice. *J Voice*. 2005; 19:574-581.
- ⁶¹ Wolfe VI, Cornell R, Fitch J. Sentence/vowel correlation in the evaluation of dysphonia. *J Voice*. 1995; 9:297-303.
- ⁶² de Krom G. Consistency and reliability of voice quality ratings for different types of speech fragments. *J Speech Hear Res*. 1994;37:985-1000.
- ⁶³ Dejonckere PH, Obbens C, de Moor GM, Wieneke GH. Perceptual evaluation of dysphonia: reliability and relevance. *Folia Phoniatr (Basel)*. 1993;45:76-83.
- ⁶⁴ Moers C, Mobius B, Rosanowski F, Noth E, Eysholdt U, Haderlein T. Vowel- and text-based cepstral analysis of chronic hoarseness. *J Voice*. 2012;26:416-424.
- ⁶⁵ Choi SH, Zhang Y, Jiang JJ, Bless DM, Welham NV. Nonlinear dynamic- based analysis of severe dysphonia in patients with vocal fold scar and sulcus vocalis. *J Voice*. 2012;26:566-576.
- ⁶⁶ Maryn Y, Corthals P, Cauwenberge PV, Roy N, de Bodt M. Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. *J Voice*. 2010; 24(5); 540-555.