

Revista Española de Lingüística

Órgano de la Sociedad Española de Lingüística

RSEL

45|1

Enero-Junio
2015

Edita
SeL

REVISTA ESPAÑOLA DE LINGÜÍSTICA
(RSEL)
45/1

Edita

SeL

REVISTA ESPAÑOLA DE LINGÜÍSTICA (RSEL)

ISSN: 0210-1874 · eISSN: 2254-8769

Depósito Legal: M-24.769-1971

DIRECTOR DE HONOR: D. Francisco Rodríguez Adrados (RAE y RAH).

DIRECTOR: Juan Antonio Álvarez-Pedrosa Núñez (UCM).

SECRETARIO: Luis Unceta Gómez (UAM).

CONSEJO DE REDACCIÓN: Montserrat Benítez (CSIC), José Antonio Berenguer (CSIC), Joaquín Garrido (UCM), Juana Gil Fernández (CSIC), Salvador Gutiérrez Ordóñez (U. León y RAE), Antonio Hidalgo (U. Valencia), Patricia Infante (CSIC), Manuel Leonetti (U. Alcalá), Eugenio Luján (UCM), Victoria Marrero (UNED), Ventura Salazar (U. Jaén), Esperanza Torrego (UAM).

CONSEJO ASESOR: Alberto Bernabé (UCM), Margarita Cantarero (SEL), Ramón Cerdá (UB), Victoria Escandell (UNED), Marina Fernández Lagunilla (UAM), José Manuel González Calvo (U. Extremadura), Emma Martinell (UB), Juan Carlos Moreno Cabrera (UAM), Gregorio Salvador (RAE), José Carlos de Torres (SEL), Jesús de la Villa (UAM).

A partir del número 38 (2008) la Revista Española de Lingüística ha recuperado el formato de dos fascículos al año, con periodicidad semestral. Los trabajos enviados para su publicación han de dirigirse al Secretario de la revista. Deberán ser originales e inéditos y ajustarse a las normas que aparecen en el número 38/2, así como en la página web de la Sociedad Española de Lingüística. Todos los trabajos son sometidos al dictamen de al menos dos evaluadores designados por el Consejo de Redacción, mediante informes de carácter confidencial.

Los derechos de publicación y difusión, bajo cualquier forma, son propiedad de RSEL. Todo texto publicado en la revista obliga a sus autores a no cederlo a terceros, sin autorización previa de la revista, quien sí queda autorizada a comercializarlo, debiendo entregar, en este caso, el 50% de los beneficios obtenidos a sus autores.

REDACCIÓN: Sociedad Española de Lingüística, Centro de Ciencias Humanas y Sociales del CSIC, c/ Albasanz, 26 - 28, 28037 Madrid. Correo electrónico: secretarioRSEL@gmail.com. <<http://www.sel.edu.es/>>

DISEÑO Y COMPOSICIÓN: Carmen Chíncoa & Carlos Curiá (produccionRSEL@gmail.com)

SERVICIOS DE INFORMACIÓN: Los contenidos de la RSEL son recogidos sistemáticamente en *Bibliographie Linguistique / Linguistic Bibliography*, *CINDOC – Base de datos Sumarios ISOC, CSA – Linguistic and Language Behavior Abstracts, Dialnet, Francis, Modern Language Association (MLA) Bibliography*.

ÍNDICE 45/1 (2015)

INTRODUCCIÓN

Nuevas aportaciones al estudio de la percepción del habla7
JUANA GIL - EUGENIA SAN SEGUNDO

ARTÍCULOS

Implicaciones perceptivas de la variación: la fricativa labiodental25
BEATRIZ BLECUA FALGUERAS - ASSUMPCIÓ ROST BAGUDANCH

La percepción de la cualidad de voz y los estereotipos vocales45
MARIANELA FERNÁNDEZ TRINIDAD

La prosodia audiovisual de la ironía verbal: un estudio de caso73
SANTIAGO GONZÁLEZ FUENTE

¿Son distintos el creak y la voz creaky?: Estudio perceptivo preliminar105
PATRICIA INFANTE

La percepción del habla en ruido: un reto para la lingüística129
y para la evaluación audiológica (estudio experimental)
VICTORIA MARRERO-AGUIAR

Percepción de los enunciados interrogativos de duda y de enfado153
sin apoyo visual en alumnos chinos de ELE
SÍLVIA PLANAS-MORALES

Percepción y evaluación de la pronunciación del español como L2175
ENRIQUE SANTAMARÍA BUSTO

ARTÍCULOS

¿SON DISTINTOS EL *CREAK* Y LA VOZ *CREAKY*? ESTUDIO PERCEPTIVO PRELIMINAR*

PATRICIA INFANTE

*Consejo Superior de Investigaciones Científicas
Universidad Complutense de Madrid*

RESUMEN

Tanto en el ámbito referido al análisis de la cualidad de voz en general como en el de las patologías vocales, es habitual encontrar trabajos en los que el objeto de estudio son dos tipos de fonación que, en palabras de Catford 1964, producen la misma impresión que «un palo al recorrer una verja». Las etiquetas que se les han aplicado son muy variadas; además, algunos autores categorizan estos dos tipos de fonación como uno solo y únicamente unos pocos investigadores realizan una distinción clara entre ambos conceptos. Esto demuestra que se trata de conceptos difíciles de perfilar y merecedores, por consiguiente, de un análisis en profundidad. Por ese motivo se ha diseñado un experimento para determinar si es posible distinguir perceptivamente los dos tipos de fonación, voz *creaky* y *creak*, o si, por el contrario, no resulta posible distinguirlos y, por tanto, deberían categorizarse bajo un mismo nombre. Los informantes que han actuado como jueces –agrupados en nativos/no nativos de español, entrenados/no entrenados en el Análisis del Perfil Vocal y expertos/no expertos en cualidad de voz *creak*(y)– han demostrado ser capaces de categorizar los distintos estímulos con un elevado grado de acuerdo y las diferencias entre las respuestas de los distintos grupos resultan significativas en este estudio preliminar.

Palabras clave: *creak*, voz *creaky*; *Vocal Profile Analysis*; percepción; fonación; cualidad de voz.

ABSTRACT

It is common to find papers in the field of voice quality and speech pathologies in which the object of study are two types of phonation that are defined as «a stick running along a railing» (Catford 1964). These types of phonation have received different names. Some

* Me gustaría expresar mi agradecimiento a la Dra. Juana Gil por sus siempre pertinentes indicaciones, al Dr. Joaquín Romero, al Dr. Antonio del Palacio, a José María Lahoz, a Marianela Fernández-Trinidad y a José Manuel Rojo, por la inestimable ayuda que me ha prestado con el análisis de los datos, así como a todos los informantes y jueces que han evaluado las muestras de habla. Este artículo forma parte de la producción resultante del proyecto CIVIL (Cualidad Individual de la Voz e Identificador de Locutor) financiado por el Ministerio de Economía y Competitividad (Plan Nacional I+D, Ref. FFI2010-21690-CO2-01).

researchers categorize these two types of phonation as one phenomenon while a small group distinguish between two concepts, «*creak*» and «*creaky voice*», which complicates its study. Defining these types of phonation is a difficult task and in depth studies are needed. A perceptual experiment has been designed to investigate if it is possible to distinguish *creaky voice* from *creak* or not, in which case they should be categorized under the same name. The judges were grouped according to the following parameters: native/non-native Spanish speakers, VPA trained / not trained and *creaky voice* experts/non-experts. Results showed that judges were able to categorize different stimuli with a high level of agreement. All differences investigated between the groups of judges were significant in this preliminary study.

Keywords: creak, creaky voice; Vocal Profile Analysis; perception; phonation; voice quality.

RECIBIDO: 26/05/2014

APROBADO: 20/01/2015

1. INTRODUCCIÓN

Este trabajo se enmarca en el estudio de la cualidad de voz. Se conoce como cualidad de voz la impresión perceptiva que producen en un oyente los ajustes laríngeos y supralaríngeos¹ de un hablante. En función de las modificaciones que un hablante realice en ellos, su voz se percibirá de una u otra manera. En el presente estudio, nos centraremos en la impresión perceptiva que producen los ajustes laríngeos.

Ladefoged 1971 entendía los diferentes ajustes fonatorios como situados en un *continuum* que mostraba las distintas configuraciones que podían adoptar los pliegues vocales desde una disposición en la cual se encuentran completamente unidos o aducidos (ataque glotal), hasta una configuración en la que se produce la abertura completa de la glotis (emisión de un sonido sordo). La gradación que él proponía se ordenaba, pues, en ataque glotal, *creaky*, modal, *breathy* y sordo, en función del grado de cierre que presentaran los pliegues vocales. Otros investigadores, como Laver, establecieron una distinción de acuerdo con los hercios a los que el hablante fonaba; así podía clasificarse la voz en tres modos de fonación o registros

1. Se entiende por ajustes supralaríngeos a las modificaciones que un hablante realiza en la cavidad oral, como el adelantamiento del ápice de la lengua, la abertura mandibular, etc. Algunos autores (sirvan como ejemplo Hollien 1974; Kreiman, Gerratt, Precoda y Berke 1992; Gobl y Ní Chasaide 2003; o Gordon y Ladefoged 2001) equiparan la cualidad de voz a la fonación. En el presente artículo se entiende por cualidad de voz el conjunto constituido por la fonación y los ajustes supralaríngeos, con independencia de que el experimento perceptivo esté centrado únicamente en los tipos de fonación.

vocales²: la voz modal³, caracterizada por una frecuencia determinada; el *falsetto*, con una frecuencia más alta; y el *creak*, registro que se caracteriza por un descenso de la frecuencia fundamental. Este estudio se centra en el modo de fonación *creak* y en la voz *creaky*, resultado de un ajuste laríngeo⁴ que se produce conjuntamente con la voz modal.

2. CREAK FRENTE A CREAKY

Uno de los factores que posiblemente más haya contribuido a dificultar el estudio de la cualidad de voz sea la abundancia de terminología existente, tal y como señalan Gerratt y Kreiman 2001:

This variability in terminology, combined with the scattering of literature across various academic disciplines, has resulted in substantial confusion about the phenomena being described. Terms that are derived from different domains (for example, acoustic vs. physiological descriptions) may often characterize the same vocal phenomenon, and a single term may be applied to rather different kinds of phonation. Further, many terms imply a particular physiologic process or perceptual reality that has not withstood or even necessarily received experimental scrutiny. As a result, readers often have significant difficulty determining precisely what vocal phenomenon an author is referring to. Thus, the study of non-modal phonation has stalled in a disorderly, descriptive stage, and the important question of which kinds of phonation are the same and which are different remains largely unaddressed.

Una buena parte de los investigadores, entre los que cabe mencionar a Gordon y Ladefoged 2001 o Laver 1980, coincide al considerar el *creak* un modo de fonación no modal, puesto que los pliegues vocales no vibran

2. En función de la bibliografía que se consulte se pueden encontrar referencias a *type of phonation* (habitualmente traducido al español como «modo de fonación») o a *vocal register* 'registro vocal'. Algunos fonetistas (Laver 1980; Hewlett y Beck 2006; Esling 1978; Esling y Wong 1983; Esling y Harris 2005; Edmonson y Esling 2006, entre otros muchos mencionables) emplean «modo de fonación», aunque otros investigadores como Hollien 1974 o Titze 2000 prefieren hablar de «registro vocal». En este trabajo se emplean ambas locuciones como sinónimos para hacer referencia a las agrupaciones hechas para caracterizar la fonación atendiendo a los hertzios, en primer lugar, y, cuando es posible, también a la configuración fisiológica.

Se empleará «tipo de fonación» como convención para hacer mención de los conceptos de *creaky* y *creak*, no como sinónimo de «modo de fonación». Como se podrá observar a lo largo del artículo, uno de los mayores problemas a los que nos enfrentamos en el ámbito de la cualidad de voz es el escaso acuerdo que existe en cuanto a terminología.

3. La voz modal se caracteriza por presentar una vibración regular de los pliegues vocales y una tensión longitudinal, medial y aductiva moderada. La mayoría de los hablantes, no obstante, suelen presentar en su fonación habitual una vibración de los pliegues vocales ligeramente irregular o con escape de aire, por este motivo es importante distinguir la voz modal de la voz habitual de un hablante.

4. Se emplea «ajuste laríngeo» para hacer referencia de manera general a cualquier tipo de modificación que pueda producirse en la laringe. La combinación del *falsetto*, la voz modal o el *creak* con algún otro tipo de fonación se denomina «ajuste» (Laver 1980; Pérez Sanz 2001).

de manera regular y no se produce en ningún momento el cierre completo de la glotis.

El *creak* se ha considerado el registro del habla más grave que los humanos podemos producir. Sus frecuencias son siempre inferiores a los 100 Hz: 10,9-52,1 Hz (McGlone 1967), 34,4 Hz (McGlone 1967), 7-78 Hz (Hollien y Michel 1968 y Gerratt y Kreiman 2001), 31,2-66,6 Hz (Hollien y Wendhal 1968), 34,6 Hz (Laver 1980), 22-71 Hz (Henton y Bladon 1988), 25-50 Hz (Marasek 1997), por debajo de 70 Hz (Titze 2000), 50-90 Hz (Hewlett y Beck 2006).

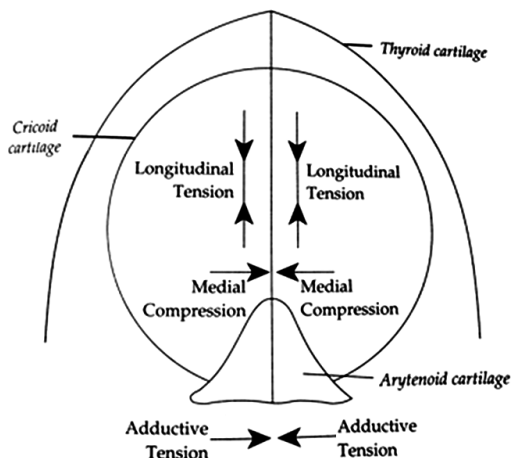


FIGURA 1. Esquema basado en la propuesta de Laver 1980 y tomado de Ní Chasaide y Gobl 1997, p. 444.

Los programas de análisis acústico habitualmente empleados por los fonetistas a menudo no muestran la frecuencia fundamental (F_0) de las grabaciones que presentan este tipo de voz. Esto bien podría deberse a que la F_0 sea extremadamente baja y el algoritmo del programa no sea capaz de calcularla, o bien a que la onda acústica que se produce en estos casos sea una onda supraperiódica (esto es, no estrictamente periódica ni aperiódica). Gerratt y Kreiman 2001, p. 369 categorizan el *creak* como un tipo de fonación supraperiódica e indican que en este tipo de voz parece existir más de un patrón que se repite en la onda acústica, lo que dificulta identificar la frecuencia fundamental, ya sea desde un punto de vista perceptivo o ya sea desde un punto de vista acústico. Hewlett y Beck 2006 y Harris y Esling 2005 simplifican y clarifican este concepto, y definen el *creak* como un registro que carece de frecuencia fundamental.

Determinar la configuración fisiológica que presentan los pliegues vocales en este tipo de fonación también resulta complicado y la descripción no es coincidente en todos los autores. Así, Laver 1980 y Hewlett y Beck

2006 indican que en el *creak* existe una baja tensión longitudinal y una alta tensión medial y aductiva (véase la Figura 1). Algunos autores (Catford 1964 y Gobl y Ní Chasaide 2003) señalan que solo la parte ligamental de los pliegues vocales vibra mientras que los cartílagos aritenoides están juntos, y que las bandas ventriculares pueden entrar en contacto ocasional con los propios pliegues vocales (Hollien, Moore, Wendahl y Michel 1966; Allen y Hollien 1973; Hollien 1974). Esling y Harris 2005, por su parte, consideran que todas estas descripciones implican la acción del esfínter ariepiglótico y, por tanto, una de las características más destacadas de este modo de fonación sería la acción de los pliegues ariepiglóticos (véase la Figura 2).

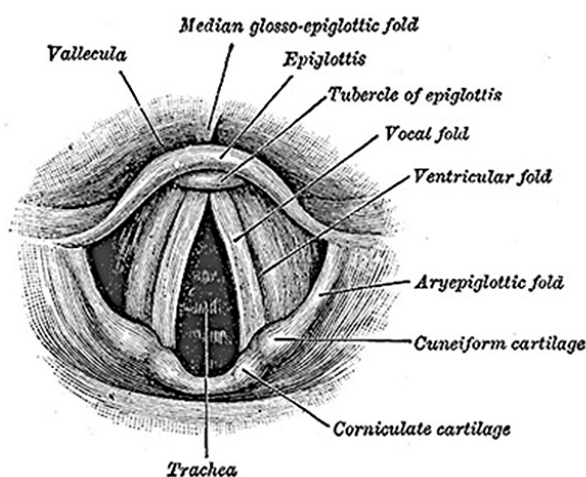


FIGURA 2. Disposición de los pliegues ventriculares y ariepiglóticos (Gray 2000).

En cualquier caso, en el *creak* los pliegues vocales están laxos y compactos, debido a la baja tensión longitudinal, y vibran de manera lenta e irregular (Eckert y Laver 1994; Hewlett; Beck 2006; Marasek 1997 entre otros), existe muy poca presión subglótica y, por tanto, el volumen de aire que pasa a través de los pliegues vocales es muy reducido (McGlone 1967; Murry 1970; McGone y Shipp 1971; Monsen y Engebretson 1977) de modo que estos no llegan a entrar en vibración y gran parte de la energía destinada a generar vibración es absorbida por las paredes del tracto vocal. Esto produce unos ciclos glóticos en los que existe una fase de abertura con muy poca amplitud y una fase de cierre muy prolongada (Moore y von Leden 1958; Wendahl, Moore y Hollien 1963; Hollien, Girard y Coleman 1977).

Por lo que respecta a lo que se conoce como voz *creaky*, Laver 1980 indicaba que durante la fonación modal era posible realizar algunos ajustes laríngeos (escape de aire, vibración irregular) para producir un «modo de fonación compuesto». La voz *creaky* sería uno de esos modos de fonación

compuestos, puesto que es una combinación entre la voz modal y el *creak* (Laver 1994; Hewlett y Beck 2006), es decir, toma de la voz modal la presencia de frecuencia fundamental y del *creak*, la configuración fisiológica. Se mueve en un rango de frecuencias más bajo que el de la voz modal de cada hablante y la onda acústica que presenta es menos irregular que la del puro *creak*. Sobre este ajuste articulatorio contamos con pocos datos, pero resulta interesante la observación que hace Laver 1975, p. 251, pues según él, la impresión perceptiva del *creaky* es diferente en función del modo de fonación o ajuste laríngeo con el que se combine: «the auditory impression of *creak* component in compound phonations is [...] different [...]. The *creaky* component in high-pitched falsetto sounds different from that in whispery *creak*⁵».

La configuración fisiológica que adoptan los pliegues vocales durante la emisión del *creak* y la voz *creaky* no termina de ser del todo clara y aún son necesarios más estudios, como señalan Esling y Harris 2005. No obstante, estos autores, p. 373, consideran que –pese a que el *creak* y la voz *creaky* produzcan una impresión perceptiva ligeramente distinta– su configuración laríngea es idéntica aunque con distinto grado de constricción del esfínter ariepiglótico⁶ como se muestra en el Esquema 1:

	Esfínter ariepiglótico
voz <i>creaky</i>	- constricción
<i>creak</i>	+ constricción

ESQUEMA 1. Tipo de fonación en función del grado de constricción del esfínter ariepiglótico.

La principal diferencia entre el *creak* y la voz *creaky* estriba en que se produzca un pequeño grupo de pulsos aperiódicos (*creak*) o que se produzcan varios grupos seguidos (voz *creaky*) (Esling y Harris 2005, p. 372). En el primer caso, la onda acústica no mantendría ningún tipo de patrón mientras que en el segundo caso existiría cierta regularidad. Aun así, todavía son necesarios más estudios acústicos para poder esclarecer las diferencias que presentan estos dos tipos de fonación⁷.

5. Para explicar la diferencia perceptiva Laver alude a dos tipos de fonación compuesta en los que el *creaky* se combina con otros ajustes laríngeos o modos de fonación. Para más detalles sobre estos tipos de fonación compuesta, véase Laver 1994, p. 198.

6. Para observar más imágenes de una laringe sin constricción frente a una con constricción véase Moisik y Esling 2007, p. 376.

7. Véase nota 2.

3. PROBLEMAS TERMINOLÓGICOS

El *creak* a menudo se ha descrito de manera metafórica: Catford 1964 lo define como semejante al «[sonido producido por] un palo al recorrer una verja»; a Henton y Bladon 1988 les recuerda al sonido del maíz al estallar; Hewlett y Beck 2006 lo comparan con una carraca; y otros autores lo han asimilado con la crepitación que se produce al quemar sal en el fuego (Pérez Sanz 2011). La voz *creaky*, aunque quizá presente menor «fuerza» que el *creak*, también podría corresponderse con estas definiciones. Esto ha provocado que en la bibliografía haya autores que se refieran a estos tipos de fonación de manera indistinta, que otros los categoricen bajo un mismo nombre sin realizar distinciones y que tan solo algunos diferencien entre ambos tipos de fonación, todo lo cual complica su estudio.

Uno de los factores que posiblemente más ha contribuido a dificultar el estudio del *creak* y de la voz *creaky* es el hecho de que la impresión perceptiva de ambos sea muy similar pero no idéntica debido a la frecuencia con que se emiten. Las etiquetas que ambos fenómenos han recibido en la bibliografía han sido muy variadas: *creak*, *creaky voice*, *low pitched noise*, *lax creaky voice*, *laryngealization*, *glottal fry*, *glottal stop*, *stiff voice*, *pulse register* o *pulse phonation*. Así, algunos autores como Epstein 2002 han empleado los términos *laryngealized*, *creaky voice* y *glottalization* como sinónimos; Lindqvist-Gauffin 1969 utiliza los nombres *glottal stop*, *laryngealization*, *creaky voice* y *creak* indistintamente, mientras que otros investigadores, como por ejemplo, Laver 1980, distinguen entre voz *creaky* y *creak*. Por ese motivo, el mayor interés reside ahora en comprobar si es posible discriminar perceptivamente estos tipos de fonación o, por el contrario, debido a su similitud, deberían categorizarse bajo un mismo nombre en el caso de que los oyentes no fueran capaces de distinguirlos. Dicho en otras palabras, ¿pueden el *creaky* la voz *creaky* distinguirse desde el punto de vista perceptivo?

En el presente estudio se analiza una serie de grabaciones que perceptivamente se asemejan a la impresión auditiva a la que aludían Catford 1964 y Henton y Bladon 1988, y se comprueba si un grupo de jueces entrenados es capaz de realizar una discriminación satisfactoria entre el *creak* y la voz *creaky*.

4. EVALUACIÓN DE LA CUALIDAD DE VOZ

Existen diversos métodos para evaluar la calidad de voz desde el punto de vista perceptivo⁸. Laver 1974 señaló la dificultad de abordar un análisis

8. En fonética clínica con frecuencia se han empleado otros métodos de evaluación, se incluyen aquí los que recopilan Gil y San Segundo 2014 p. 162-163: la escala GRBAS («Grade, Roughness, Breathiness, Asthenia, Strain» Isshiki y Takeuchi 1970; Hirano 1981, 1989), la escala

científico exhaustivo si las etiquetas que se aplicaban al habla eran vagas, ambiguas e «impresionistas», y si no existía una correspondencia entre las etiquetas y las instrucciones articulatorias, de manera que el lector fuera capaz de conseguir el sonido descrito.

Por este motivo, Laver, Wirs, Mackenzie y Hiller 1981 diseñaron el Análisis del Perfil Vocal (*Vocal Profile Analysis* o *VPA*), un método de evaluación basado en la fisiología de la producción y empleado en diversas disciplinas como la fonética, la logopedia, la fonoaudiología o el entrenamiento de la voz. Algunos investigadores consideran que no es posible realizar un análisis perceptivo en el que exista coherencia interjueces; sin embargo, Schewell 1998 prueba la validez del VPA como herramienta para la investigación en fonética y en el estudio de las patologías del habla. En la misma línea, Gerratt y Kreiman 2001 –si bien muestran su escepticismo acerca de las evaluaciones perceptivas– mantienen que el *creak* es suficientemente prominente como para ser evaluado desde este punto de vista⁹.

4.1. El Vocal Profile Analysis

La cualidad de voz –dependiente, entre otras cuestiones, de los ajustes laríngeos– es un rasgo de largo alcance, para cuya evaluación es necesario considerar la totalidad de la muestra de habla. Por este motivo, en el protocolo del Análisis del Perfil Vocal (VPA) se recogen todos los rasgos que caracterizan el habla de una persona y que, por tanto, resultan relevantes, con independencia de que estos se encuentren presentes en su habla de manera constante o intermitente¹⁰. El VPA se divide en tres secciones: una primera dedicada a los ajustes del tracto vocal, otra en la que se recoge la tensión general del tracto vocal, y una tercera centrada en los ajustes laríngeos¹¹. Mientras que para la evaluación de otros ajustes es muy recomendable que la persona entrenada conozca la lengua que evalúa para poder determinar si una característica se desvía o no de lo esperado, la evaluación de los tipos de fonación resulta más sencilla cuanto más atención se preste al sonido laríngeo emitido y menos se atienda al resto de ajustes supralaríngeos, que son los que en definitiva hacen que se articulen los sonidos y se configure el mensaje.

RBH («Roughness, Breathiness, Hoarseness» Wendler, Rauhut y Krüger 1986), Bufalo III *Voice Screening Profile* (Wilson 1987), Stockholm Voice Evaluation Approach (Hammarberg y Gauffin 1995) o el Consensus Auditory Perceptual Evaluation of Voice (CAPE-V) (Kempster, Gerratt, Verdolini Abbot, Barkmeier-Kramer y Hillman 2009).

9. Sin embargo, según estos investigadores el *creak* en combinación con la voz modal presenta más dificultades de evaluación.

10. En aquellos casos en los que la aparición de un rasgo es intermitente pero relevante, el VPA incluye la notación «i».

11. Para consultar más información actualizada sobre el Análisis del Perfil Vocal y acceder a la plantilla completa empleada en el protocolo perceptivo, véase Mackenzie Beck 2005 y 2007.

C. PHONATION FEATURES										
	SETTING	Present		Scalar Degree						
		Neutral	Non-neutral	Moderate		Extreme				
				1	2	3	4	5	6	
10. Voicing type	Voice									
	Falsetto									
	Creak									
	Creaky									
11. Laryngeal friction	Whisper									
	Whispery									
12. Laryngeal irregularity	Harsh									
	Tremor									

FIGURA 3. Extracto de una plantilla de VPA que muestra las casillas para evaluar los distintos tipos de fonación y los ajustes laríngeos.

Prescindir del contenido léxico no siempre resulta sencillo, por lo que quizá presentar a los jueces estímulos de una lengua que no conozcan facilite la tarea. Para la evaluación, el juez entrenado parte de la escucha de una grabación y se ocupa de desglosar la impresión holística de una voz en sus diferentes características. La plantilla del VPA está diseñada de modo que el evaluador pueda realizar un primer análisis muy somero en el que indica si un determinado fenómeno se encuentra presente y, a continuación, detalla el grado en que dicho se muestra fenómeno en una escala de 1 a 6, siendo 1 el mínimo y 6 el máximo.

Para la elaboración del experimento perceptivo de este trabajo se ha tomado como modelo evaluador el esquema del VPA y se ha realizado una ligera adaptación *ad hoc* para evaluar la fonación *creak* (*y*), objeto del presente estudio. En la Figura 3 se puede observar la parte del esquema del VPA dedicada a las características fonatorias que ha servido de base para diseñar el experimento.

De acuerdo con el esquema del VPA, el primer paso para la evaluación de la fonación consiste en decidir si existe una fonación neutra (voz modal) o no. La mayoría de los hablantes realiza pequeños ajustes que confieren a su voz una particularidad propia de cada persona y los evaluadores rara vez señalan la casilla «neutral» en la plantilla del protocolo, sino que habitualmente marcan la celda «no neutral».

5. OBJETIVOS E HIPÓTESIS

Una vez analizada la problemática existente en cuanto a la clasificación del *creak* y de la voz *creaky*, parece interesante esclarecer si un grupo de jueces es capaz de discriminar perceptivamente estas cualidades de manera significativa.

Las hipótesis que se pretendían corroborar con el experimento que se describe en este estudio son las siguientes:

- Distinguir la voz *creaky* del *creak* es una tarea extremadamente compleja, tal y como indican investigadores como Laver 1980; Henton y Bladon 1988; Gerratt y Kreiman 2001 o Kreiman y Sidtis 2013, entre otros.
- Los jueces que no tengan el español como L1 realizarán mejor la tarea de discriminación que los españoles, puesto que única y exclusivamente se fijarán en la fonación y obviarán el significado del enunciado.
- El grado de acuerdo entre los jueces al evaluar las categorías será mayor en los jueces que hayan sido entrenados en el VPA y en los expertos que en el resto de jueces. Se espera que las diferencias entre estos grupos sean significativas.
- Los jueces serán capaces de categorizar los estímulos con un elevado grado de coherencia individual en voz no *creak(y)*, *creak* y *creaky*.
- Esta coherencia se verá incrementada notablemente cuanto mayor sea el conocimiento del juez de la cualidad de voz *creak(y)*¹².

6. MÉTODO

6.1. *Adaptación ad hoc del Vocal Profile Analysis*

En previsión de la complejidad que puede entrañar en algunas ocasiones el etiquetado de las voces (véase la Sección 3) y para simplificar el experimento del presente estudio, se ha decidido incluir en el test perceptivo una categoría denominada «no *creak(y)*», en la que se incluyen todas las desviaciones de la configuración neutra que no sean voz *creaky* o *creak*, (recuérdese que denominamos *creak* al registro más grave del habla y *creaky* a la combinación entre el *creak* y la voz modal) cada una de las cuales cuenta con su propia categoría. Si la muestra de habla se juzgaba como «no *creak(y)*» o «*creak*» se pedía al juez que señalara la opción pertinente y procediera a escuchar el siguiente estímulo. En caso de marcar «*creaky*voice» –ajuste que lleva implícita la presencia de frecuencia fundamental– era necesario que jerarquizara su grado en moderado (1-3) o extremo (4-6)¹³.

12. Se emplea la notación *creak(y)* para agrupar aquellas voces que Catford 1964; Henton y Bladon 1988 o Hewlett y Beck 2006 definían metafóricamente, de este modo se evita etiquetar los estímulos dentro del *creak* o de la voz *creaky*. Bajo el nombre de «no *creak(y)*» se incluyen aquellos estímulos que no presentan ninguno de estos tipos de fonación pero que pueden no coincidir con la voz modal canónica.

13. Es importante señalar que, aunque inicialmente se les presentó a los jueces la gradación de la voz *creaky* en una escala de 1 a 6, en el análisis de los datos se ha estimado oportuno simplificar esta jerarquización en moderado (1-3) y extremo (4-6), puesto que existían demasiadas categorías para los casos con los que se contaba en este estudio preliminar.

6.2. Selección de estímulos

Resulta bastante complejo elicitarse muestras válidas de *creak* o voz *creaky*, especialmente si se quiere evitar que estén acompañadas de otros ajustes como puede ser la tensión laríngea¹⁴. Además, Vieira 1997 señala que los resultados en la evaluación perceptiva de la calidad de voz han demostrado no ser satisfactorios para vocales, por lo que, aunque los estímulos perdieran algo de homogeneidad, se decidió grabar a cinco hombres¹⁵, hablantes de español centropeninsular en conversación semiespontánea y elegir treinta enunciados al azar (seis por hablante) delimitados por pausas que rondaran los 40 segundos, tiempo suficiente para evaluar la calidad de voz (Beck 2005). Cada estímulo se presentó dos veces y el orden fue siempre aleatorio.

Posteriormente, un especialista en Otorrinolaringología del Hospital Universitario de La Paz estudió las voces empleadas para comprobar que no fueran patológicas, no presentaran astenia, tensión ni «rugosidad» en los pliegues vocales. Sin embargo, lo que sí se apreció es que los locutores tendían a forzar una emisión más grave de la que la longitud y el grosor de sus pliegues vocales normalmente permitían.

6.3. Jueces

Para corroborar las hipótesis era necesario encontrar jueces que cumplieran diferentes requisitos, todos ellos bastante específicos, por lo que se intentó constituir un grupo que permitiera observar su comportamiento atendiendo a distintos criterios: su LI, si habían sido entrenados en el protocolo del VPA (Laver, Wirs, Mackenzie y Hiller 1981; Beck 2007) y su experiencia en el estudio de la calidad de voz y en *creak*(y). Puesto que encontrar evaluadores con estas características tan concretas resulta complejo, en este primer experimento el grupo de jueces lo formaron ocho fonetistas. En la Tabla 1 se presentan las características de cada uno de los jueces.

Los jueces recibieron un pequeño entrenamiento previo al experimento en el que se les presentaron algunas muestras de voz sin *creak*(y) frente a otras con *creak* extremo. De este modo se les permitía familiarizarse con los tipos de fonación que se encontrarían más adelante y se les mostraban casos con presencia extrema de *creak*(y) frente a otros sin ella. Esto se

14. En la grabación del corpus de CIVIL (San Segundo, Alves, Fernández Trinidad 2013) se comprobó que era frecuente que algunos informantes al intentar producir voz *creaky* incluyeran un alto grado de tensión laríngea (*harsh voice*).

15. Los estímulos se grabaron en la cabina insonorizada del Laboratorio de Fonética del CSIC. Se empleó un micrófono inalámbrico AKG C444L y la señal se recogió con el programa Adobe Audition.

diseñó para facilitar a los evaluadores la jerarquización de los estímulos que habrían de llevar a cabo posteriormente y a la vez proporcionarles un entrenamiento que confería homogeneidad al grupo.

	Juez KK	Juez JL	Juez EL	Juez PI	Juez MF	Juez JR	Juez ZR	Juez RP
Entrenado en VPA	No	No	Sí	Sí	No	No	Sí	No
Español L1	No	Sí	No	Sí	Sí	Sí	No	Sí
Experto en <i>creak(y)</i>	Sí	No	No	Sí	No	No	No	No

Tabla 1. Esquema que detalla las características de los jueces.

6.4. Diseño del experimento

Se diseñó un experimento perceptivo de respuesta forzada MFC con el programa Praat (Boersma y Weenink 2014). Los jueces comenzaron con el entrenamiento y posteriormente realizaron el test perceptivo. En él, se les presentaba un sonido y se les pedía que lo categorizaran en «no *creak(y)*», «*creaky*» o «*creak*».

De acuerdo con los presupuestos del VPA, y tal y como se ha mencionado anteriormente, si el audio era clasificado dentro del grupo «*creaky*», era necesario también jerarquizar, de 1 (apenas audible) a 6 (grado extremo), el grado con que se percibía el estímulo. En caso de que este se clasificara en otro grupo distinto del «*creaky*», el juez simplemente debía continuar con el experimento. Los jueces podían escuchar los estímulos todas las veces que necesitaran. En la Figura 4 se muestra la interfaz que se les presentó a los jueces en el experimento.

Please decide which category best describes this sound.
When judged as creaky: rate between 1 (just noticeable) and 6 (extreme degree).
Otherwise: press NA and next button

no creak(y)

creaky

creak

1

2

3

4

5

6

NA

Play again

FIGURA 4. Muestra visual del experimento.

7. RESULTADOS Y DISCUSIÓN

En este apartado se muestran los resultados obtenidos de la evaluación global de los audios que realizaron los fonetistas, se evalúa cómo de parecidos eran los resultados de cada juez entre su primera y segunda ronda de respuestas, y el grado de acuerdo existente entre los distintos grupos de jueces en relación con cada audio.

La primera hipótesis mantenía que distinguir la voz *creaky* del *creak* era problemático para todos los fonetistas, por lo que se decidió comprobar si esto era cierto tomando todas las respuestas de los jueces y extrayendo la moda, que mostraba la categoría globalmente preferida para cada estímulo.

Para esto se realizó una ordenación de los sonidos en categorías de carácter ordinal atendiendo a dos criterios: la frecuencia y el grado de cualidad *creak* (*y*) que presentaba un estímulo (véase Figura 5). La categoría «*creak*» es la que presenta menor frecuencia y más cualidad *creak* (*y*), mientras que los grupos «*creaky*» y «no *creak* (*y*)» muestran frecuencias más altas con distinta presencia de cualidad *creak* (*y*): moderada en el caso del «*creaky*» y sin ella en la categoría «no *creak* (*y*)».

Para medir el grado de desacuerdo que existía entre los jueces para cada sonido, se tomó la diferencia absoluta entre la respuesta de cada juez y el valor modal de dicho sonido. Si hubieran respondido al azar, el grado de desacuerdo sería de 12; sin embargo, en este caso, las respuestas oscilan entre 3,60 y 5,60, por lo que aunque fueran difíciles de categorizar, podemos afirmar que los jueces fueron plenamente capaces de categorizarlos por encima del nivel de azar.

+ hercios		
-intensidad		
cualidad		no creaky
creak(y)		creaky 1-3
		creaky 4-6
+ hercios		creak
-intensidad		
cualidad		
creak(y)		

FIGURA 5. Gradación de no *creak* (*y*) a *creak*.

La Figura 6 representa la media del grado de desacuerdo que existe para las distintas etiquetas. Resulta importante señalar en este punto que, de todos los posibles ajustes a largo plazo que un hablante puede realizar, los tipos de fonación suelen ser uno de los aspectos de la cualidad de voz que mejores índices de reconocimiento muestra (Beck 2007);

sin embargo, según se sostiene en el presente trabajo, algunos tipos de fonación pueden resultar más difíciles de identificar y evaluar que otros, como ocurre aquí. A pesar de que inicialmente se esperaba que los jueces encontrarán más problemas para distinguir la voz *creaky* del *creak*, en los resultados se observa que los jueces son capaces de evaluar los extremos del eje de abscisas de la Figura 6, no *creaky* y *creak*, –y especialmente este último– con un grado de desacuerdo inferior al esperado por azar, ya que aquí tan solo es de 4; sin embargo, encuentran complicado jerarquizar la presencia de *creaky*. Esto puede explicarse si tenemos en cuenta que todos los jueces eran fonetistas y podían distinguir con relativa facilidad qué estímulos se caracterizaban por la ausencia de F0 (*creak*). En cambio, evaluar qué grabaciones presentaban voz *creaky*, y en qué grado, ha resultado más complejo porque su presencia puede no ser siempre tan clara ni fácil de graduar.

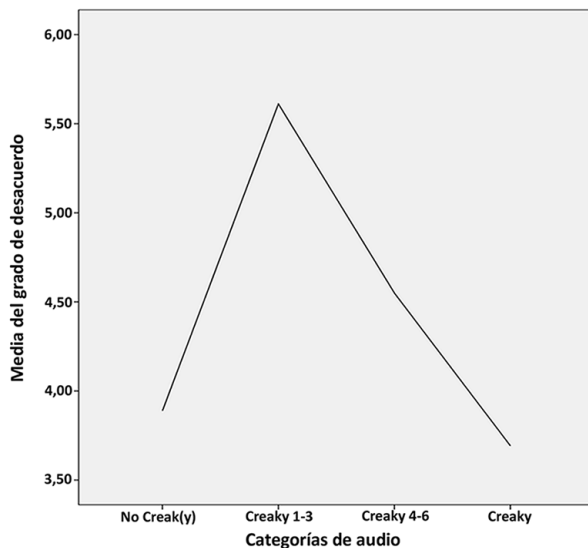


FIGURA 6. Media del grado de desacuerdo del total de jueces para las categorías «no *creak* (y)», «*creaky* 1-3», «*creaky* 4-6» y «*creak*».

Puesto que a cada juez se le presentó cada uno de los estímulos dos veces, sería natural encontrar que entre la primera ronda de respuestas y la segunda existieran ligeras variaciones, especialmente si era la primera vez que los evaluadores se enfrentaba a una tarea de discriminación de este tipo.

Es cierto también, que en algunos casos las posibles fluctuaciones entre los estímulos han sido menores y no comportan grandes diferencias, pero en otras ocasiones la confusión entre una categoría y otra muestra que el

estímulo presentado le resulta al juez difícil de categorizar. Por este motivo se ha decidido presentar el coeficiente del Rho de Spearman, puesto era necesario evaluar que los resultados de cada juez entre su primera y segunda ronda fueran similares y apuntaran en una misma dirección, es decir, que si se asignaba un valor alto en la primera ronda, también lo fuera en la segunda.

Debido a la naturaleza ordinal que se le ha otorgado aquí a las categorías puede emplearse el coeficiente de correlación Rho de Spearman. Se han obtenido unos valores medios de 0,76, que pueden considerarse muy satisfactorios si tenemos en cuenta que habitualmente el umbral de aceptación se fija en 0,40; es decir, que en líneas generales la mayoría de los jueces presentan una tendencia muy similar y en una misma dirección entre su primera y segunda ronda de respuestas. De hecho, uno de los jueces llega a presentar una correlación de 0,93 sobre un máximo de 1; esto es, sus respuestas a la primera y a la segunda escucha son prácticamente idénticas.

Para que estas conclusiones resultaran más fiables se ha decidido complementar estos datos con un test de kappa de Cohen con el fin de comprobar que los resultados obtenidos estaban por encima del nivel de azar. Esto efectivamente se verificó, ya que la media total es de 0,60 con un nivel de significatividad $<0,001$. Así pues, puede decirse que los jueces presentan resultados muy parecidos en ambas rondas de respuestas.

Una vez realizados estos primeros análisis, se elaboraron unas tablas que muestran si el grado de desacuerdo entre los jueces se veía disminuido al agruparlos atendiendo al hecho de que hubieran tenido o no formación en VPA, de que fueran nativos de español o fueran no nativos y de que presentaran o carecieran de experiencia en el estudio de la cualidad de voz *creak* (y) (Tablas 2 y 3). Téngase en cuenta que, para que las agrupaciones fueran homogéneas y los resultados definitivos, sería necesario un mayor tamaño muestral, pero dadas las dificultades que ha supuesto encontrar jueces con unas características tan específicas no ha sido posible alcanzarlo por el momento. Los resultados que se muestran, por tanto, son tan solo preliminares con respecto a cada uno de los grupos. De todos modos, no por ello dejan de tener relevancia como paso previo, puesto que es necesario obtener unos primeros resultados para, posteriormente, poder ampliar la muestra y seleccionar los jueces que presenten el perfil más adecuado para este tipo de estudio.

Las Tablas 2 y 3 muestran los 60 audios presentados y el grado de desacuerdo encontrado entre los distintos grupos de jueces en cada uno de los órdenes de repetición. Se ha decidido sombrear en color gris claro aquellos casos en los que el grado de desacuerdo era 0 y aquellos en los que el máximo desacuerdo entre todos los jueces que integraban el grupo era de un grado. La única excepción la conforma el grupo de expertos en cualidad *creak* (y), que tan solo está integrado por dos jueces y en el que

el criterio ha sido más estricto y solo se han considerado aquellos casos en los que existe un acuerdo total.

Además, para asegurar que las respuestas eran fiables, en el recuento final únicamente se han tenido en cuenta aquellos casos en los que los jueces no diferían en más de una categoría ni en la primera ronda de respuestas ni en la segunda. Estos casos están señalados en las Tablas 2 y 3 en color gris oscuro.

Primeramente se decidió analizar los resultados de un grupo de jueces que desconocía la lengua en la que estaban grabados los estímulos, pues se fijarían solo en la forma de la muestra de habla y no en el contenido y, por consiguiente, se esperaba que el reconocimiento de los distintos tipos de fonación se viera incrementado. Como se puede observar en las Tablas 2 y 3, sucede así, ya que los hablantes no nativos de español han conseguido un acuerdo mayor en la primera y segunda ronda que los hablantes de español: trece audios de acuerdo (tres con acuerdo total y diez con acuerdo parcial) frente a uno de acuerdo parcial¹⁶, conforme a lo esperado en la hipótesis 2. Si tenemos en cuenta que los jueces no nativos pertenecían a la escuela británica, famosa por el entrenamiento perceptivo de sus fonetistas, y que los representantes de la tradición española no comparten en general esa preparación, es posible que pueda explicarse por qué el grupo de hablantes no nativos de español ha obtenido unos resultados mucho mejores.

A continuación, se evaluó la actuación del grupo de jueces entrenado en el Vocal Profile Analysis. Estos fonetistas habían sido instruidos para ser capaces de cumplimentar el protocolo del VPA¹⁷ y, puesto que la tarea que se les solicitó era muy similar, se esperaba que sus resultados fueran sustancialmente mejores que los de los jueces que no habían sido entrenados en este protocolo, pero también que los de los evaluadores agrupados en nativos/no nativos de español. De acuerdo con lo previsto en la hipótesis 2, el entrenamiento en VPA comporta, en efecto, un descenso en el grado de desacuerdo en la primera y segunda ronda, como puede observarse en las Tablas 2 y 3. Así, el grupo que había sido entrenado muestra dieciséis audios con un bajo grado de desacuerdo (en 6 de los cuales existía un acuerdo total y en diez un acuerdo parcial), mientras que el grupo que no había sido entrenado tan solo coincidió en cuatro estímulos, en uno de los cuales consiguieron un acuerdo completo y en tres de ellos un acuerdo parcial.

16. Debe tenerse en cuenta que todos los recuentos de datos en los que se muestran los aciertos entre primera y segunda vuelta se realizan sobre el total de audios existentes (30), es decir, sin contar la repetición.

17. Según se muestra en los artículos especializados sobre el VPA (Beck 2005, p. 303), los jueces que han realizado este curso completo deben ser capaces de mostrar una elevada coherencia intrajuez e interjueces.

Estímulo			entrenados	no entrenados	no		expertos en	no expertos
			en VPA	en VPA	españoles	españoles	<i>creak</i> (<i>y</i>)	en <i>creak</i> (<i>y</i>)
			Media	Media	Media	Media	Media	Media
A12	Repetición	1	,33	,20	0,00	,50	,50	,17
		2	1,00	2,30	,33	1,70	2,00	1,87
A15	Repetición	1	,33	,30	,33	,30	,50	,30
		2	,33	1,20	,33	1,50	0,00	1,37
A19	Repetición	1	0,00	,80	0,00	1,20	0,00	,97
		2	,33	1,20	,33	1,50	0,00	1,37
A21	Repetición	1	1,33	,20	,33	1,70	4,50	,30
		2	3,00	1,80	0,00	2,70	4,50	1,50
A22	Repetición	1	0,00	,20	0,00	,20	0,00	,17
		2	0,00	,70	,33	1,30	,50	1,07
A25	Repetición	1	1,00	,20	,33	,50	,50	,27
		2	1,00	1,80	,33	1,70	,50	1,87
F01	Repetición	1	0,00	,20	0,00	,50	0,00	,40
		2	1,00	,20	1,00	,20	0,00	,57
F03	Repetición	1	3,00	2,70	3,00	2,70	0,00	1,50
		2	3,00	2,70	3,00	2,70	0,00	1,50
F04	Repetición	1	,33	1,30	1,33	1,00	,50	,57
		2	0,00	,50	0,00	,70	0,00	,57
F06	Repetición	1	1,00	,80	2,33	,70	,50	,57
		2	,33	,20	,33	,50	0,00	,17
F07	Repetición	1	0,00	1,30	,33	1,00	,50	,97
		2	,33	,80	,33	1,20	0,00	,97
F08	Repetición	1	2,33	1,70	2,33	2,70	0,00	2,27
		2	1,33	1,20	1,33	1,80	0,00	1,77
J11	Repetición	1	0,00	,20	,33	,20	,50	,17
		2	0,00	,50	,33	,30	,50	,27
J12	Repetición	1	1,33	1,50	2,33	2,00	4,50	1,77
		2	1,33	1,50	2,33	2,00	4,50	1,77
J14	Repetición	1	1,33	2,30	2,33	1,80	4,50	1,77
		2	2,33	2,00	2,33	2,00	0,00	1,07

TABLA 2. Medias del grado de desacuerdo para las distintas agrupaciones de jueces.

Estímulo			entrenados en VPA	no entrenados en VPA	no españoles	españoles	expertos en <i>creak</i> (y)	no expertos en <i>creak</i> (y)
			Media	Media	Media	Media	Media	Media
J15	Repetición	1	1,33	,70	2,33	,70	,50	,80
		2	1,33	1,30	2,33	,80	,50	,97
J17	Repetición	1	0,00	2,70	3,00	2,70	4,50	2,70
		2	0,00	2,70	3,00	2,70	4,50	2,40
J19	Repetición	1	0,00	,80	,33	1,30	,50	1,10
		2	,33	,20	,33	,80	,50	,70
N10	Repetición	1	,33	,20	,33	,50	0,00	,17
		2	,33	,20	,33	,50	0,00	,17
N11	Repetición	1	,33	,30	,33	,70	0,00	,40
		2	,33	,20	,33	,50	0,00	,40
N24	Repetición	1	,33	,30	,33	,80	0,00	,57
		2	,33	,20	,33	,50	0,00	,40
N4	Repetición	1	,33	1,70	,33	1,80	,50	1,87
		2	0,00	1,80	,33	1,20	,50	1,07
N8	Repetición	1	3,00	2,70	3,00	1,70	4,50	2,17
		2	,33	1,70	,33	1,50	0,00	1,20
N9	Repetición	1	0,00	,20	0,00	,50	0,00	,40
		2	0,00	0,00	0,00	,20	0,00	,17
V1	Repetición	1	0,00	,70	,33	1,30	,50	1,07
		2	0,00	,30	,33	1,50	,50	1,20
V2	Repetición	1	3,00	2,70	3,00	1,80	0,00	2,40
		2	3,00	2,70	3,00	1,80	0,00	2,40
V28	Repetición	1	3,00	0,00	3,00	1,80	0,00	2,40
		2	0,00	1,80	0,00	2,70	0,00	2,40
V31	Repetición	1	,33	,30	,33	,30	0,00	,30
		2	,33	,20	,33	,20	0,00	0,00
V32	Repetición	1	3,00	2,70	3,00	1,80	0,00	2,70
		2	2,33	2,30	2,33	2,70	0,00	2,27
V7	Repetición	1	0,00	0,00	0,00	,20	0,00	,17
		2	0,00	0,00	0,00	,20	0,00	,17

TABLA 3. Medias del grado de desacuerdo para las distintas agrupaciones de jueces.

En la misma línea, también puede observarse que cuanto mayor sea la formación que los jueces poseen en el campo de la cualidad de voz *creak(y)*, menor es el grado de desacuerdo: en el caso del grupo de expertos, once audios con un grado de acuerdo total frente a dos con acuerdo parcial, en el de los no expertos.

Como se mencionaba anteriormente, es interesante analizar los resultados atendiendo a las cuatro clasificaciones hechas: en función de si los jueces son o no fonetistas, en función de su entrenamiento en VPA, en función de si el español es o no su lengua materna y en función de su experiencia en el estudio de la cualidad *creak(y)*. Según el test de igualdad de varianzas de Lebenne, que asigna a cada grupo un valor, se ha medido el grado de agrupación global que existe en las respuestas por audio y con respecto a cada grupo.

La varianza muestra diferencias entre los distintos grupos: entre los no nativos de español y los nativos el nivel de significatividad es $<0,001$, entre los entrenados en VPA y los no entrenados es $<0,005$, y entre expertos y no expertos es $<0,001$. Por lo tanto, según lo esperado, las diferencias entre todos los grupos son significativas. Aunque efectivamente pudimos observar en las Tablas 2 y 3 que los grupos que presentan más experiencia en cualidad de voz obtuvieron unos mejores resultados que el resto, al realizar el test de igualdad de varianzas se puede comprobar que el nivel de significatividad que se obtiene para este grupo es menor que para el resto. Esto puede deberse a que uno de los jueces que integra este grupo explicó que le había resultado difícil evaluar aquellos estímulos en los que se presentaba en un mismo audio con fragmentos que incluían un grado muy extremo de cualidad *creak(y)* y otros cuya cualidad era no *creak(y)*, combinación que con frecuencia se encuentra en el habla. Ciertamente, esto explicaría estas diferencias y si revisamos los resultados que este juez obtuvo tanto en el test de kappa como en el coeficiente de correlación Rho de Spearman, sus resultados están por debajo de los de la mayoría de jueces. En próximos estudios se tratará de presentar estímulos más breves, pero también más homogéneos, para subsanar esta posible dificultad de evaluación.

Dado que un grupo estaba constituido por dos expertos frente a seis no expertos, parecía interesante poder comprobar hasta qué punto los dos expertos en cualidad de voz *creak(y)* habían logrado coincidir en sus evaluaciones. Por esa razón, se estimó oportuno realizar un test de kappa entre los expertos, que muestra que sus resultados coinciden en un 0,38 por encima del nivel de azar con una significatividad de $<0,001$. De acuerdo con la clasificación de Landis y Koch, se trata de un valor bajo, aunque no despreciable, puesto que a partir de 0,40 podría considerarse bueno; por ello, está próximo al umbral que habitualmente se marca como aceptable. Es decir, estos dos jueces lograron un acuerdo total en 23 de los 60 audios en, al menos, una ronda de respuestas y un acuerdo total en once audios,

tanto en la primera como en la segunda ronda, lo que es mucho más de lo que se había conseguido en el resto de los grupos. Estos resultados son muy interesantes ya que, claramente, no se deben al azar y muestran que, si bien es difícil categorizar los audios, estos jueces fueron capaces de hacerlo y de coincidir totalmente con respecto a un tercio de ellos.

Se muestra como ejemplo la Tabla 4, que ofrece los resultados del test de kappa de uno de los jueces expertos. Este juez ha categorizado prácticamente igual los estímulos en la primera y segunda ronda y solo son cuatro los audios en los que sus respuestas no han sido parecidas. En todo caso, según se ilustra en la tabla, las diferencias solo se han dado entre categorías colindantes, lo que sin duda indica que la respuesta en estos casos tampoco era aleatoria.

Juez experto 1		Segunda ronda				Total
		No <i>creak(y)</i>	<i>Creaky</i> 1-3	<i>Creaky</i> 4-6	<i>Creak</i>	
Primera ronda	No <i>creak(y)</i>	6	0	0	0	6
	<i>Creaky</i> 1-3	2	12	0	0	14
	<i>Creaky</i> 4-6	0	1	4	0	5
	<i>Creak</i>	0	0	1	4	5
Total		8	13	5	4	30

TABLA 4. Comparación de resultados entre la primera y la segunda ronda de respuestas de uno de los jueces expertos en calidad de voz.

Tal y como se preveía en las hipótesis, todo parece indicar que incluso a los jueces entrenados en el VPA y en calidad de voz les resultó difícil caracterizar algunos de los estímulos —especialmente los correspondientes a dos de los hablantes— y distinguir entre las categorías «*creaky*» y «no *creak(y)*», prueba de que la tarea resultaba difícil y de que la discriminación perceptiva entre categorías no es tan clara como parece que se ha presupuesto en la bibliografía tradicional, en la que se presentan algunas caracterizaciones de los distintos tipos de voz a partir de las intuiciones de los propios investigadores. No obstante, también es cierto que los jueces fueron capaces de categorizar los estímulos con una sistematicidad razonable, lo que parece indicar que, en efecto, es posible distinguir la voz *creaky* del *creak* desde el punto de vista perceptivo.

Si bien los resultados presentados aquí por el momento son preliminares, a la luz de lo expuesto puede deducirse que la situación óptima para evaluar la calidad *creak(y)* sería contar con jueces que no solo fueran fonetistas, sino también expertos en este tipo concreto de fonación, puesto

que ellos son los que han conseguido un índice de desacuerdo interno más bajo. En el orden de preferencia, estarían seguidos por los jueces entrenados en VPA y por los fonetistas que no tuvieran el español como L1.

Los resultados, en cualquier caso, serían más fiables si los grupos pudieran ser más amplios y estar constituidos por el mismo número de jueces, objetivo que se perseguirá en el futuro pese a la dificultad que entraña encontrar evaluadores con características tan específicas.

8. CONCLUSIONES

El objetivo principal del presente artículo ha consistido en comprobar si los distintos jueces eran capaces de categorizar los estímulos percibidos en *creak*, *creaky* y no *creak(y)*, entendiéndose por *creak(y)* una macrocategoría que engloba tanto el *creak* como el mucho más ambiguo *creaky*. Los jueces fueron capaces de evaluar el *creaky* y el no *creak(y)* razonablemente bien y, sin embargo, encontraron problemas para categorizar el *creaky*.

De acuerdo con los estudios que muestran qué parámetros se identifican mejor en el protocolo del Vocal Profile Analysis (Beck 2007; Camargo y Madureira 2008; Infante y Fernández-Trinidad 2014), el apartado que más índice de aciertos muestra es la evaluación de los tipos de fonación. Sin embargo, a la vista de los resultados presentados aquí, parece también manifiesto que ciertos tipos de voz resultan extremadamente difíciles de evaluar, como ocurre en este caso con la voz *creaky*, que suscita un elevado grado de desacuerdo entre los jueces. En el experimento realizado esto puede deberse a que la categorización de la voz *creaky* implicaba también jerarquizar el grado en el que este tipo de fonación se presentaba, lo que suponía una complicación añadida a la tarea. Algunos estudios recientes, como los de Kreiman y Gerratt 2000 o Kreiman; Gerratt e Ito 2007, muestran que los oyentes pueden presentar dificultades para aislar e identificar perceptivamente las características presentes en una voz y quizá esto sea lo que ocurre en algunas ocasiones con la evaluación perceptiva de la calidad *creak(y)*, especialmente con aquellos casos dentro de esta categoría que se encuentran entre la voz modal y el *creak*.

En cualquier caso, se ha podido corroborar que los jueces han sido capaces de categorizar los distintos estímulos con un elevado grado de sistematicidad en no *creak(y)*, *creak* y *creaky*. Los resultados entre unas agrupaciones de jueces y otras presentaron diferencias muy notables. Así, los grupos que menor grado de desacuerdo mostraron fueron el grupo de expertos en calidad de voz *creak(y)*, el grupo entrenado en el VPA y el grupo de jueces no hispanohablantes. Sin embargo, para que estos resultados puedan considerarse definitivos, resulta esencial poder realizar un experimento perceptivo a mayor escala y con un mayor número de jueces para, a continuación, cotejar los resultados con un análisis acústico que permita determinar qué características definen y distinguen la voz *creaky*

del *creak*. Para ello, se elegirán estímulos que se puedan comparar acústicamente y que, a ser posible, tengan el mismo grado de cualidad *creak* (*y*) de manera que la evaluación sea más sencilla. También sería conveniente aplicar la teoría de detección de señales al experimento, de manera que pueda verificarse que los jueces son capaces de mantener la atención a lo largo de su desarrollo y que sus respuestas no están sesgadas.

9. REFERENCIAS BIBLIOGRÁFICAS

- ADOBE SYSTEM (2014): *Adobe Audition* [programa de ordenador], versión 3.0.
- BECK, J. (2007): *Vocal profile analysis scheme: A user's manual*, Edimburgo, Queen Margaret University.
- BOERSMA, P. y WEENINK, D. (2014): *Praat: doing phonetics by computer*, [programa de ordenador], versión 5.3.64 <<http://www.praat.org/>>
- CAMARGO, Z. y MADUREIRA, S. (2008): «Voice quality analysis from a phonetic perspective: Voice Profile Analysis for Brazilian Portuguese (BP-VPAS)», *Proceedings of the fourth conference on speech prosody (Speech prosody 2008)*, Brasil, pp. 57-60.
- CATFORD, J.C. (1964): «Phonation types: The classification of some laryngeal components of speech production», en Abercrombie, D. Fry, D.B., MacCarthy, P.A.D., Scott, N.C. y Trim, J.M.L. (Eds.), *In honour of Daniel Jones*, Londres, Longmans, pp. 26-37.
- EDMONSON, J.A. y ESLING, J.H. (2006): «The valves of the throat and their functioning in tone, vocal register and stress: Laryngoscopic case studies», *Phonology* 23, 2, pp. 157-191.
- ECKERT, H. y LAVER, J. (1994): *Menschen und ihre Stimmen*, Weinheim, Beltz Psychologie Verlags Union.
- EPSTEIN, M. (2002): *Voice quality and prosody in English*, Tesis Doctoral, Los Ángeles, University of California.
- ESLING, J.H. (1978): *Voice quality in Edinburgh. A sociophonetic study*, Tesis Doctoral, Edimburgo, Universidad de Edimburgo.
- , y HARRIS, J. (2005): «States of the glottis: An articulatory phonetic model based on laryngoscopic observations», en Hardcastle, W. y Beck, J. (eds.), *A figure of speech: A Festschrift for John Laver*, Nueva Jersey, Lawrence Erlbaum Associates, pp. 247-383.
- , y WONG, R.F. (1983): «Voice quality setting and the teaching of pronunciation», *TESOL Quarterly* 17, 1, pp. 89-95.
- GERRATT, B. y KREIMAN, J. (2001): «Toward a taxonomy of nonmodal phonation», *Journal of Phonetics* 29, pp. 365-381.
- GIL, J. y SAN SEGUNDO, E. (2014): «La cualidad de voz en fonética judicial», en Garayzábal, E., Jiménez, M. y Reigosa, M. (eds.), *Lingüística forense. La lingüística en el ámbito legal y judicial*, Madrid, Euphonia Ediciones.
- GOBL, C. y NI CHASAIDE, A. (2003): «The role of voice quality in communicating emotion, mood and attitude», *Speech Communication* 40, pp. 189-212.
- GORDON, M. y LADEFOGED, P. (2001): «Phonation types: a cross linguistic overview», *Journal of Phonetics* 29, 383-406.

- GRAY, H. (2000): *Anatomy of the human body*, Filadelfia, Lea & Febiger.
- HENTON, C. y BLADON, R. (1988): «Creak as a sociophonetic marker», en Hyman, L.M. y Li, C.N. (eds.), *Language, speech and mind: Studies in honor of Victoria A. Fromkin*, Londres, Routledge, pp. 3-29.
- HEWLETT, N. y BECK, J. (2006): *An Introduction to the Science of Phonetics*, Nueva Jersey, Lawrence Erlbaum Associates.
- HOLLIEN, H. (1974): «On vocal registers», *Journal of Phonetics* 2, pp. 125-143.
- , (2000): «The concept of ideal voice quality», en Kent, R.D. y Ball, M.J., *Voice Quality Measurement*, San Diego, Singular Publishing Group.
- , GIRARD, G.T. y COLEMAN, R.F. (1977): «Vocal fold vibratory patterns of pulse register phonation», *Folia Phoniátrica et Logopédica* 29, pp. 200-205.
- , y MICHEL, J.F. (1968): «Vocal fry as a phonational register», *Journal of Speech and Hearing Research* 11, pp. 600-604.
- , y WENDAHL R.W. (1968): «Perceptual study of vocal fry», *Journal of the Acoustical Society of America* 43, pp. 506-509.
- INFANTE, P. y FERNÁNDEZ-TRINIDAD, M. (2014): «¿Cómo trabajar con el Vocal Profile Analysis en español?», presentado en el *XI Congreso internacional de lingüística general*, Pamplona.
- KNEIMAN J. y GERRATT, B. (2000): «Measuring vocal quality», en Kent, R. D. y Ball, M. J., *Voice Quality Measurements*, San Diego, Singular Publishing, pp. 73-102.
- KREIMAN, J., GERRATT, B., PRECODA, K. y BERKE, G.S. (1992): «Individual differences in voice quality perception», *Journal of Speech and Hearing Research* 35, pp. 512-520.
- , e ITO, M. (2007): «When and why listeners disagree on voice quality assessment tasks», *Journal of the Acoustical Society of America* 122, pp. 2354-2364.
- , y SIDTIS, D. (2013): *Foundations of voice studies: An interdisciplinary approach to voice production and perception*, Reino Unido, Wiley-Blackwell.
- LAVER, J. (1974): «Labels for voices», *Journal of the International Phonetic Association* 4, pp. 62-75.
- , (1975): *Individual features in voice quality*, Tesis Doctoral, Edimburgo, Universidad de Edimburgo.
- , (1980): *The Phonetic Description of Voice Quality*, Cambridge, Cambridge University Press.
- , (1994): *Principles of phonetics*, Cambridge, Cambridge University Press.
- , WIRZ, S., MACKENZIE, J. y HILLER, S.M. (1981): «A perceptual protocol for the analysis of vocal profiles», *Work in Progress* 14, pp. 139-155.
- LINDQVIST-GAUFFIN, J. (1969): «Quarterly progress and status report laryngeal mechanisms in speech», *Speech Transmission Laboratory-Quarterly Progress and Status Report (STL-QPSR)* 10, 2-3, pp. 26-32.
- MACKENZIE BECK, J. (2005): «Perceptual analysis of voice quality: The place of Vocal Profile Analysis», en Hardcastle, W.J. y Mackenzie Beck, J. (eds.), *A figure of speech. A festschrift for John Laver*, Londres, Lawrence Erlbaum Associates.
- MARASEK, K. (1997): *EGG & Voice Quality*, Stuttgart University [archivo en línea], <<http://www.ims.uni-stuttgart.de/phonetik/EGG/>> [enero 2014].
- MOISIK, S. y ESLING, J. (2007): «3D Auditory-articulatory modeling of the laryngeal constrictor mechanism», *International congress of phonetic sciences XVI*, pp. 373-376.

- MOORE, P. y VON LEDEN, H. (1958): «Dinamic vibrations of the vibratory pattern in the normal larynx», *Folia Phoniatica* 10, pp. 205-238.
- NÍ CHASAIDE, A. y GOBL, C. (1997): «Voice source variation», en Hardcastle, W. y Laver, J. (eds.), *The handbook of phonetic sciences*, Londres, Blackwell.
- PÉREZ SANZ, C. (2011): *Ajustes laríngeos y estilos de fonación en habla y televisión*, Tesis doctoral, Madrid, Universidad Complutense de Madrid - Instituto Universitario José Ortega y Gasset.
- SAN SEGUNDO, E., ALVES, H. y FERNANDEZ TRINIDAD, M. (2013): «CIVIL Corpus: voice quality for forensic speaker comparison», *Procedia-Social and Behavioral Sciences Corpus Resources for Descriptive and Applied Studies. Current Challenges and Future Directions: Selected Papers from the 5th International Conference on Corpus Linguistics (CILC2013)* 95, pp. 587-593.
- SHEWELL, C. (1998): «The effect of perceptual training on ability to use the vocal profile analysis scheme», *International Journal of Language & Communication Disorders* 33, 1, pp. 322-326.
- TITZE, I. (2000): *Principles of voice production*, 2^a edición, Iowa City, National Center for Voice and Speech.
- VIEIRA, M. (1997): *Automated measures of dysphonias and the phonatory effects of asymmetries in the posterior larynx*, Tesis Doctoral, Edimburgo, Universidad de Edimburgo.
- WENDAHL, R. MOORE, P. y HOLLIEN, H. (1963): «Comments on vocal fry», *Folia Phoniatica* 15, pp. 251-255.

Edita
SeL

