TESIS DOCTORAL

2018

CARACTERIZACIÓN DEL *FALSETTO* Y SUS CONSECUENCIAS PARA LA DISCRIMIINACIÓN DE VOCES

MARIANELA FERNÁNDEZ TRINIDAD

PROGRAMA DE DOCTORADO EN FILOLOGÍA. ESTUDIOS LINGÜÍSTICOS Y LITERARIOS: TEORÍA Y APLICACIONES

DIRECTORA: DRA. JUANA GIL FERNÁNDEZ

Agradecimientos

Por fin, llega el momento de los agradecimientos y, por suerte, son muchas las personas a las que quiero y tengo que dar las gracias.

A mi directora de tesis, Juana Gil Fernández, un importantísimo referente académico y personal para mí desde que la conocí en 2009 cuando viajé a Madrid para realizar el Posgrado de *Estudios Fónicos* (CSIC-UIMP). Con ella tuve el privilegio de formarme y de trabajar en el Laboratorio de Fonética durante 5 años. Las actividades que allí desarrollamos con todo el equipo (investigación, docencia y peritajes) han sido, sin duda, esenciales para el desarrollo de esta tesis doctoral. Muchas gracias por todos esos años de aprendizaje, de trabajo y de amistad y, por supuesto, por la ayuda inestimable para la realización de esta tesis.

Quiero agradecer también muy especialmente a Pedro Gómez-Vilda quien, generosamente y de forma totalmente desinteresada, me proporcionó su programa de análisis de la voz (*BioMetroSoft*®), me explicó y me orientó pacientemente tanto en su manejo como en la interpretación de los datos obtenidos a partir de él. Sin su ayuda, esta tesis tampoco hubiera sido posible.

Un agradecimiento enorme a mi amigo y colega José María Lahoz, que me ha ayudado con la estadística asesorándome con infinita paciencia, y ha compartido conmigo muchas horas de discusión sobre varios aspectos de esta investigación, absolutamente enriquecedores, además de animarme siempre.

A los profesores del Máster Oficial en Estudios Fónicos (CSIC-UIMP) les agradezco por la excelente formación que me han brindado. Entre ellos, destaco muy especialmente quienes colaboraron con la tesis:

A Norma Antoñanzas y a Heriberto Avelino y por sus cursos sobre fonación y la onda glotal.

A Joaquim Llisterri y a María Machuca, de quienes siempre aprendo fonética.

A Antonio Manzanero, que me ayudó con la *Teoría de Detección de Señales* y me renovó el interés por continuar con la tesis, en uno de esos momentos de desánimo.

A Victoria Marrero y a Ana María Fernández Planas, que me contagiaron su entusiasmo por la fonética perceptiva (y por la fonética clínica).

A Carolina Pérez Sanz, quien consiguió captar mi atención sobre la voz y el funcionamiento de la laringe. Muchas gracias por sus estupendas clases y largas charlas sobre la cualidad de la voz.

A Sandra Schwab, que me orientó con aspectos metodológicos en fonética perceptiva.

Agradezco asimismo a las instituciones que, de un modo u otro, hicieron posible este trabajo:

Al *Consejo Superior de Investigación Científicas* (CSIC) que, en dos oportunidades, me concedió una beca para mis estudios de Máster y Doctorado. En el CSIC, también doy las gracias a la Unidad de Estadística, en especial a su responsable, José Manuel Rojo.

A la *Universidad Nacional de Educación a Distancia* (UNED), y en particular, a Salvatore Bartolotta y a Ángela Romera, coordinador y secretaria del Programa de Doctorado en Filología, respectivamente, quienes me guiaron en todo momento en las gestiones burocráticas.

A la *Università degli Studi Federico II*, y muy especialmente, a la profesora Francesca Dovetto, a Anna Chiara Pagliaro, Daria D'Alessandro y Francesca Carbone, que me recibieron tan amablemente en Nápoles para realizar los estudios perceptivos.

A la *Universidad de la República* (Uruguay), en particular al *Departamento de Teoría del Lenguaje y Lingüística General* y a mis amigas Virginia Bertolotti, Serrana Caviglia, Magdalena Coll, Sylvia Costa, Mariela Grassi y Marisa Malcuori.

A mis colegas y amigos de la Universidad Complutense de Madrid (UCM) y de la Universidad Nacional de Educación a Distancia (UNED) porque me animaron para que terminara el Doctorado.

A Patricia Infante, que me prestó a sus estudiantes de Logopedia de la Universidad Complutense de Madrid para las pruebas perceptivas.

A todos los estudiantes de ambas universidades por realizar los test perceptivos. También, a los 6 informantes napolitanos que vinieron a grabar al Laboratorio de Fonética del CSIC.

A Chiara Celata, Pier Marco Bertinetto, y a todo el magnífico equipo del *Laboratorio di Linguistica* de la *Scuola Normale Superiore* (SNS) de Pisa, donde realicé una provechosa estancia de investigación.

A Alba Aguete y Elisa Fernández-Rei, por su ayuda con la utilización de FOLERPA.

A muchos otros amigos y colegas de la profesión que, de una forma o de otra, colaboraron conmigo. En especial agradezco a Nuria Martínez García, Elena Battaner, Jordi Cicres, Eva Estebas, Mark Gibson, Jennifer González, Paolo Roseano y Assumpció Rost.

En lo estrictamente personal, a mi familia (Montevideo-Genova-Casale), en particular a mis padres y abuelos.

Por supuesto, un agradecimiento infinito a Maura y a Daniele porque me apoyan siempre, en todo y ante todo.

Gracias a todos ellos he podido llegar hoy hasta aquí y espero que me sigan acompañando, como hasta ahora, en este nuevo comienzo.

Índice de contenidos

Índice de Tablas

Índice de Figuras

Lista de abreviaturas

A	Aciertos
ACP	Análisis por Componentes Principales
APRES	*Aural Perception on Reverse Speech*
C	Consonante
CCHS	Centro de Ciencias Humanas y Sociales
CIVIL	Cualidad Individual de la Voz e Identificación del Locutor
COR	*Curva Característica de Operación del Receptor*
CPs	Componentes Principales
CP1	Primer Componente Principal
CP2	Segundo Componente Principal
CP3	Tercer Componente Principal
CP4	Cuarto Componente Principal
CSIC	Consejo Superior de Investigaciones Científicas
F	*Falsetto*
F1, F2…F6	*Falsetto* del Locutor 1, *Falsetto* del Locutor 2… *Falsetto* del Locutor 6
FA	Falsas Alarmas
FOLERPA	*Ferramenta On-Line para ExpeRimentación PerceptivA*
FSR	*Forensic Speaker Recognition*
GRBAS	*Grade, Roughness, Breathiness, Asthenia, Strain*
IAFPA	*International Association for Forensic Phonetics and Acoustics*
ILG	Instituto da Lingua Galega
L	Locutor
logTR	Logaritmo de los tiempos de reacción
M	Voz modal
M1, M2…M6	Modal del Locutor 1, Modal del Locutor 2… Modal del Locutor 6
MAE	*Mucosal Average Energy*
MFCC	*Mel-frequency Cepstral Coefficients*
MFDR	*Maximum Flow Declination Rate*
MW	*Mucosal Wave*

MWC	*Mucosal Wave Correlate*
NHR	*Noise-to-Harmonics Ratio*
NIST	*National Institute of Standards and Technology*
n.s.	No significativo
O	Omisiones
PSD	*Power Spectral Density*
R	Ruido
RC	Rechazos Correctos
ROC	*Receiver Operating Characteristic*
S	Señal
SNS	*Scuola Normale Superiore*
TDS	Teoría de Detección de Señales
TR	Tiempos de reacción
UCM	Universidad Complutense de Madrid
UdelaR	(Universidad de la República)
UIMP	Universidad Internacional Menéndez Pelayo
UNED	Universidad Nacional de Educación a Distancia
UNINA	*Universitá Degli Studi di Napoli Federico II*
USC	Universidad de Santiago de Compostela
V	Vocal
VPA	*Vocal Profile Analysis*
WAV	*Waveform Audio File Format*

Resumen

Caracterización del *falsetto* y sus consecuencias para la discriminación de voces

Esta investigación persigue una doble finalidad: por un lado, pretende profundizar en el conocimiento del registro de *falsetto*, realizando un estudio detallado de los aspectos acústicos que lo caracterizan y lo diferencian de la voz modal, así como de los posibles correlatos articulatorios a los que apuntan dichos rasgos acústicos.

De otra parte, se propone evaluar, en el plano perceptivo, el daño provocado por el *falsetto* en la discriminación de voces e indagar en los posibles efectos que puede tener conocer o no la lengua del locutor comparando el español y el italiano, lenguas que no se sirven lingüísticamente de los mecanismos de fonación para significar.

Estudios previos han mostrado que el incremento de la frecuencia fundamental (f_0) o el cambio de registro hacia el *falsetto* comprometen o impiden el reconocimiento de una voz familiar o desconocida y que se trata, además, de un procedimiento bastante utilizado por delincuentes que quieren confundir y, en última instancia, ocultar su identidad. Este procedimiento de transformación de la voz, relativamente fácil de llevar a cabo y muy efectivo, despista notablemente a oyentes humanos y a reconocedores automáticos de voz. En ello radica su especial interés aplicado al ámbito de la fonética judicial.

Antes de valorar en qué medida el *falsetto* afecta el reconocimiento de voces se establecen primeramente qué rasgos vocales modifican los locutores por ser intrínsecos al cambio de registro y cuáles no se modifican, bien porque no son indispensables para conseguir un *falsetto*, bien porque los locutores no son capaces de disimular o, mejor dicho, se resisten a este tipo de disimulo. En esta tesis se intentan aislar aquellos rasgos vocales necesarios para el cambio de registro, es decir, los rasgos que puedan ser definitorios del *falsetto*. Como efecto de lo anterior, quedarán también mejor acotados aquellos que no se hayan modificado por las razones que fueren, aunque no constituye el objetivo principal de esta tesis resolver si, entre los rasgos invariables, hay algunos que pudieran ser indicativos de la individualidad de la voz.

La metodología seguida en la presente investigación es la propia de la fonética experimental. Para alcanzar el primer objetivo descrito, se grabaron a 6 locutores

masculinos, jóvenes universitarios y hablantes nativos de italiano (de Nápoles), que pronunciaron las mismas frases con su voz habitual y en *falsetto*.

A partir de las muestras de voz obtenidas se extrajeron, con un programa informático (*BioMet®Phon* de *BioMetroSoft®*), medidas relativas al comportamiento fonatorio de todos los locutores en ambos registros. Luego, mediante varios estudios estadísticos se pudieron finalmente delimitar los dos grupos de rasgos vocales que interesaba aislar. Por un lado, se encontraron aquellos parámetros que modificaron todos los locutores y que, por tanto, podrían considerarse intrínsecos al cambio de registro. Se trata de parámetros centrados en la f_0 y en el comportamiento biomecánico de los pliegues tendente a aumentarla o disminuirla (tensión, masa, irregularidades en la vibración y gastos de energía de la cubierta del pliegue, y tensión y masa del cuerpo). Todos estos rasgos que recogen información tonal aumentaron su valor en el *falsetto,* a excepción de la masa del cuerpo puesta en vibración, que disminuyó. En cuanto al comportamiento de la frecuencia fundamental f_0 en el cambio de registro se observó que cuando pasaron de voz modal a *falsetto,* los locutores analizados aumentaron su f_0 de promedio 1,5 octavas y que la desviación típica media fue mucho mayor en el *falsetto* que en la voz modal.

De otra parte, se ha podido también confinar un conjunto de rasgos que presentaron un comportamiento fluctuante, unas veces cambiaron entre registros y otras veces no, dependiendo de cada locutor. Si entre este último grupo pudiera haber alguno indicativo de la individualidad de la voz es un aspecto que no se analiza en esta investigación pero que por el interés que presenta podría dirimirse en estudios futuros.

Se ha señalado previamente en la bibliografía que los cambios de registro pueden desencadenar cambios en la velocidad de habla de los locutores. Asimismo, se ha demostrado en varios trabajos que el factor temporal en general es una clave perceptiva relevante en el reconocimiento de hablantes pues contribuye en buena medida a la caracterización individual del habla.

Del análisis del *tempo* se observó, al igual que en algunos estudios previos, que el *falsetto* provoca concomitantemente otras modificaciones en el habla, como el cambio de la velocidad de elocución, aunque la tendencia observada en esta investigación no es clara: al comparar un registro fonatorio y otro, se comprueba que en la mitad de los locutores la velocidad se acelera, mientras que en la otra mitad disminuye.

El segundo objetivo de esta tesis fue tratar de dilucidar si los parámetros que se alteran (y los que no) en el cambio de registro modal-*falsetto* son perceptivamente relevantes para los oyentes, esto es, constituyen una clave perceptiva para la discriminación de voces y el reconocimiento de locutores. Para comprobar esto último, para evaluar hasta qué punto los oyentes pueden ser capaces de reconocer a la misma persona hablando con su voz normal y con su voz disimulada en *falsetto* se elaboró un test perceptivo de discriminación igual-diferente, el cual se aplicó a 140 oyentes jóvenes universitarios. La creación de la prueba perceptiva así como su ejecución se llevaron a cabo mediante la herramienta FOLERPA, *Ferramenta On-Line para ExpeRimentaciónPerceptivA*.

Los resultados, interpretados en el marco de la *Teoría de Detección de Señales*, demostraron que el *falsetto* dificulta pero no impide la discriminación correcta de las voces. Los oyentes demostraron que son capaces, bajo ciertas circunstancias, de reconocer voces en tareas de discriminación igual–diferente. Aunque la discriminación fue muy débil, los oyentes pudieron reconocer a un mismo locutor por encima del nivel del azar cuando habló en voz modal y en *falsetto*. También se observó que los oyentes aplicaron un criterio liberal en sus respuestas, es decir, mostraron una tendencia a responder que las voces que oían pertenecían al mismo locutor.

Por último, interesaba también evaluar la posible influencia de la lengua en la discriminación de voces. Los principales estudios dedicados a evaluar el efecto del idioma en la percepción y el reconocimiento auditivo de la individualidad del hablante indican, tomados en su conjunto y a pesar de la divergencia metodológica de los experimentos, que los oyentes tienen la capacidad de atender a características de la voz que parecen ser independientes de la lengua para reconocer rasgos individualizadores del locutor. Algunos de los resultados a los que se llega en esta tesis dan apoyo adicional a este supuesto y parecen probar que las propiedades no lingüísticas de la cualidad de voz, en concreto, las derivadas del comportamiento laríngeo y algunos aspectos temporales del habla, pueden desempeñar un papel importante en el reconocimiento de locutores, y que, cuando esto ocurre, conocer la lengua del hablante no supone una ventaja decisiva.

Los experimentos perceptivos llevados a cabo en esta tesis consistieron en dos test. En el primer experimento perceptivo participaron 60 jueces españoles, universitarios, hablantes nativos de español y sin conocimientos de italiano. El segundo test de discriminación se realizó con 80 oyentes nativos de italiano (de Nápoles),

universitarios. Ninguno de los 140 oyentes estaba acostumbrado a evaluar voces y no presentaban problemas de audición. Los jueces no conocían a ninguno de los informantes grabados.

Los resultados se analizaron en conjunto y separando cada grupo de oyentes. Como se ha dicho, los oyentes españoles e italianos reconocieron las voces por encima del nivel del azar aunque con un nivel de discriminación bajo en los dos casos. Ambos grupos de oyentes aplicaron un criterio liberal (tendencia a responder que las voces pertenecían al mismo locutor) aunque entre los españoles esta tendencia fue más acusada. La única diferencia estadísticamente significativa al evaluar el efecto de la lengua en la discriminación de voces fue que los españoles cometieron más Falsas Alarmas.

Se calcularon los tiempos de reacción por grupo de oyentes. A partir del análisis realizado en esta tesis, se observó un efecto del idioma en los tiempos de respuesta: los italianos emplearon tiempos superiores a los de los españoles en todos los tipos de respuesta, es decir, tardaron significativamente más en tomar sus decisiones y ofrecer sus juicios.

Por último, esta investigación corrobora también un interesante hallazgo sobre los tiempos de reacción descubierto antes en el ámbito del reconocimiento facial. En esta tesis se constata una relación entre el tiempo de reacción y el tipo de respuesta, y se observa que los casos de acierto (entendiendo por ello Aciertos y Rechazos Correctos) presentan tiempos de respuesta menores mientras que los fallos (Falsas Alarmas y Omisiones) se caracterizan por tiempos de respuesta mayores. Todos los oyentes, independientemente de su lengua e independientemente de si estaban evaluando pares de igual locutor (AA) o de locutores distintos (AB), tardaron significativamente más cuando fallaron que cuando acertaron. Coincidencias como esta animan a seguir confrontando los estudios realizados sobre modalidades perceptivas diferentes.

Aunque la presente investigación se propuso alcanzar algunos resultados que permitieran avanzar en algo en el conocimiento básico del registro de *falsetto* y en la percepción humana de la cualidad de la voz, son indudables sus implicaciones prácticas. Muchos de sus hallazgos, en concreto aquellos que se relacionan directamente con el disimulo mediante *falsetto* y con la influencia de la lengua del locutor, permiten realizar una transferencia de conocimiento al ámbito de la fonética forense o judicial (por ejemplo, casos de cotejos de voces con fines judiciales y realización de ruedas de reconocimiento).

Abstract

Characterization of the *falsetto* register and its consequences for voice discrimination

This research has two aims: on the one hand, it intends to delve into the knowledge of the *falsetto* register by means of a detailed study of the acoustic features that characterize it and set it apart from modal voice, as well as into the possible articulatory correlates to which the acoustic features point.

On the other hand, it aims at evaluating, on the perceptual level, the damage that *falsetto* can cause on voice discrimination, and also at investigating the possible effects that the knowledge—or lack thereof—of the speaker's language can have, which is done by means of a comparison between Spanish and Italian, two languages that do not use phonation modes to convey meaning.

Previous studies have shown that the rise in fundamental frequency (f_0) or the shift in register to the *falsetto* can compromise and in some cases make impossible the recognition of any voice, whether familiar or unfamiliar; they have also shown that it is a procedure quite used by criminals who wish to confound and eventually conceal their identity. This vocal-quality transformation, which is relatively easy to achieve and very effective, is notably misleading for human listeners and automatic speech recognizers. Therein lies its particular interest to the field of forensic phonetics.

Before assessing to what extent the *falsetto* register has an effect on voice recognition, it is necessary to first establish which vocal features are modified by the speakers because they are intrinsic to the shift in register and which ones are not subject to modifications, due to the fact that either they are not indispensable in order to achieve *falsetto* or speakers are not able to disguise their voices—better said, because the vocal features resist being subject to disguise. The present research intends to isolate those vocal features that are necessary to achieve the shift in register, that is, those that are the defining features of *falsetto*. As a consequence, features that remain unmodified—for whatever the reason—will also be identified, even though the main purpose is not finding out if some among the invariable ones could indicate voice individuality.

The methodological approach is that of experimental phonetics. In order to achieve the first goal indicated above, six male native speakers of Italian (from Naples), all of whom had completed university education, were recorded. They uttered the same sentence in modal voice and in *falsetto*.

Several measurements of phonatory behavior for all speakers and both registers were taken by means of *BioMetroSoft*®'s computer program *BioMet®Phon*. Subsequently, statistical analyses made possible the identification of the two groups of vocal features characteristic of modal and *falsetto* registers. One group comprises those parameters that all speakers modified and that could thus be considered intrinsic to the register shift. These parameters are associated to f_0 and the biomechanical behavior of the vocal folds, which shows a tendency towards rising or lowering f_0 (tension, mass, irregular vibration and energy losses in the vocal fold cover, and tension and mass of the body). All these features, which are related to tonal information, had higher values for *falsetto*—with the exception of body mass in vibration, for which values were lower. Regarding the shift from modal voice to *falsetto*, the speakers' f_0 went up an average of 1.5 octaves and had a mean standard deviation significantly higher in *falsetto* than in modal voice.

The other group comprises features that exhibited a fluctuating behavior in the sense that they sometimes changed from one register to the other and other times they did not, depending on each speaker. As it has already been mentioned, whether in this group there could be features that might be indicative of voice individuality is not an aspect considered in this research, but that due to their inherent interest could be settled in future studies.

It has been noted in the literature that shifts in register can trigger changes in the speech rate of the speakers. Likewise, it has been shown in several works that the temporal cue, in a broad sense, is a relevant perceptual cue for speaker recognition because it contributes greatly to the characterization of the speaker's individuality.

Through the temporal analysis of speech it was observed—as it has already in some previous studies—that the *falsetto* register bears other concomitant modifications in speech, such as a change in speech rate, although the tendency noticed in the present research is not clear: when comparing the two registers, it becomes evident that for half of the speakers speech rate increases, whereas for the other half it decreases.

The second aim of this research was trying to elucidate whether the parameters—the ones that change as well as those that do not—in the shift between modal and *falsetto* registers are perceptually relevant for the listeners, that is, whether they constitute a perceptual cue in voice discrimination and speaker recognition. In order to evaluate to what extent listeners are able to recognize the same speaker when he uses his normal voice and his disguised voice by means of *falsetto*, a perceptual same–different discrimination task was performed by 140 young university listeners. The task was devised and also run using the tool FOLERPA (*Ferramenta On-Line para ExpeRimentación PerceptivA*).

The results, interpreted within the framework of *Signal Detection Theory*, showed that *falsetto* makes the correct discrimination of voices difficult but not impossible. Under certain circumstances, listeners proved to be capable of recognizing voices in same–different discrimination tasks. Although the discrimination rate was low, judges were able to recognize a given speaker above chance level when they use modal voice and *falsetto*. Additionally, listeners applied a liberal criterion in their answers, that is, they showed a tendency towards answering that the voices that they heard belonged to the same speaker.

Another important aspect evaluated was the possible influence that the language spoken could have on voice discrimination. The main studies dedicated to the evaluation of the effect of language knowledge in the perception and auditory recognition of the speaker's individuality—taken as a whole and in spite of the methodological differences of their experiments—indicate that listeners have the ability to pay attention to voice characteristics that seem to be language independent in order to recognize idiosyncratic features of the speaker. Some of the results obtained in this dissertation provide further support for this assumption and seem to prove that the nonlinguistic properties of voice quality, specifically those derived from laryngeal behavior and some temporal aspects of speech, can play an important role in speaker recognition, and that, when this takes place, knowing the speaker's language does not signify any advantages or almost none.

Two perceptual experiments were carried out to this end. In the first experiment 60 native speakers of Spanish participated, all university students without any knowledge of Italian. The second discrimination experiment was carried out with 80 native speakers of Italian (from Naples), also university students. None of the

140 judges was used to judging voices or had hearing problems. The judges did not know any of the recorded speakers.

The results were analyzed both as a whole and by listener group. As stated above, Spanish and Italian listeners recognized all voices above chance level, although the discrimination rate was low in both cases. Both groups applied a liberal criterion (specifically, they were inclined towards responding that the voices belonged to the same speaker), although it was more pronounced in the case of the Spanish judges. The only statistically significant difference when evaluating the effect of language knowledge in voice discrimination was that Spanish judges produced more false alarms.

Reaction times were measured for each listener group; the analyses conducted showed a language-knowledge effect on the reaction times: Italians made use of longer periods of time than those of Spaniards for all answer types, that is to say, it took them longer to reach a decision.

Lastly, this research also corroborates an interesting finding about reaction times that has already been found in the field of facial recognition. In this dissertation a relation between reaction times and answer types is confirmed: correct responses (meaning hits and correct rejections) exhibit shorter reaction times, whereas errors (i.e., false alarms and misses) are characterized by longer reaction times. All listeners, regardless of language and of whether they were evaluating pairs corresponding to the same speaker (AA) or different ones (AB), took longer to respond when they gave an incorrect response than when they gave a correct one. Coincidences such as these encourage the comparison of studies on different perceptual modalities.

Although the present research set out to obtain some results that allowed to make some progress in the basic knowledge of the *falsetto* register and in the human perception of voice quality, its practical implications are unquestionable. Many of its findings, specifically those related directly to voice disguise through *falsetto* and to the influence of the speaker's language, allow a knowledge transfer to the field of forensic phonetics (for example, cases of forensic voice comparisons or voice lineups).

Introducción

Mis estudios de Doctorado comenzaron en 2012 cuando me incorporé, gracias a la concesión de una beca de estudios[1], al equipo de investigación del Laboratorio de Fonética del Consejo Superior de Investigaciones Científicas que entonces dirigía la Dra. Juana Gil Fernández.

Desde el primer momento, me interesé por el estudio de la cualidad de la voz, un ámbito poco estudiado en español y que, sin embargo, constituía entonces una de la principales líneas de trabajo de dicho laboratorio.

El proyecto inicial de esta tesis se centraba esencialmente en estudiar cómo los cambios voluntarios en la cualidad de la voz evocaban en los oyentes distintas asociaciones de significado, también llamadas fonosimbólicas, y de qué manera la publicidad se servía de ellas para persuadir con sus mensajes verbales. El principal resultado de este comienzo fue el trabajo "La percepción de la cualidad de la voz y los estereotipos vocales", publicado en un número monográfico sobre percepción en la *Revista de la Sociedad Española de Lingüística*[2].

Al tiempo que continuaba indagando en las asociaciones semánticas que la cualidad de voz, y en concreto, ciertos modos de fonación, provocaban en los oyentes, empezaba también a integrarme cada vez más en las investigaciones que se desarrollaban en el Laboratorio a través de proyectos, presentaciones en congresos y publicaciones[3]. Fue así que, tras mi incorporación oficial al proyecto de investigación *CIVIL: Cualidad Individual de la Voz e Identificación del Locutor* (MICINN, FFI10-

[1] Beca JAE Pre-DOC del Consejo Superior de Investigaciones Científicas para desarrollar estudios de Doctorado en el Programa Oficial "Estudios Fónicos" (Consejo Superior de Investigaciones Científicas y Universidad Internacional Menéndez Pelayo). Convocatoria BOE 03/02/2011, resolución BOE 03/06/2011.

[2] Fernández Trinidad, M. (2015). La percepción de la cualidad de voz y los estereotipos vocales. *Revista Española de Lingüística*, *45*(1), 45-72.

[3] Entre todas, destaco las siguientes:

Alves, H., Fernández Trinidad, M., Gil Fernández, J., Infante, P., Pérez Sanz, C. y San Segundo, E. (2012). "Disguised voices: A perceptual experiment". *3rd European Conference of the International Association of Forensic Linguistics*, Universidad de Oporto, Portugal.

San Segundo, E., Alves, H., y Fernández Trinidad, M. (2013). CIVIL corpus: voice quality for speaker forensic comparison. *Procedia-Social and Behavioral Sciences*, *95*, 587-593.

Fernández Trinidad, M., Infante, P., Lahoz-Bengoechea, J. M. y Alves, H. (2013). "*Falsetto* as a disguise method in male voices", *31 Congreso Internacional de la Asociación Española de Lingüística Aplicada*. Universidad de La Laguna, Abril 2013.

Gil Fernández, J., Fernández Trinidad, M., Infante, P. y Lahoz-Bengoechea, J. M. (2017). "Obtaining speech samples for research and expertise in forensic phonetics". En: Orletti, F. y Mariottini, L. (Eds.) *Theories, Practices, Instruments of Forensic Linguistics* (pp. 27-50). Cambridge: Cambridge Scholars Publishing.

21690), mi curiosidad por aprender se detuvo durante los siguientes años en el estudio de los mecanismos de disimulo de la voz a través de diversos registros de fonación, en concreto, en el estudio del disimulo mediante el *falsetto*.

En 2014, a raíz de una presentación del equipo en el *Congreso Internacional de Fonética Experimental* en Valencia,[4] me propuse estudiar en profundidad el perjuicio del *falsetto* y los efectos de la lengua en la percepción de la cualidad de la voz y el reconocimiento del locutor comparando el español y el italiano, lenguas que no utilizan lingüísticamente los modos de fonación. Fue así que en la primavera de 2015 comencé a confeccionar el corpus de esta tesis (ver Capítulo 3). Un primer análisis, tanto de los aspectos de la producción de la voz como de las consecuencias perceptivas del disimulo para la discriminación de locutores, se realizó con oyentes hispanohablantes y sus resultados se publicarán próximamente[5]. La investigación, que entonces comenzaba, continuaría tres años más. Durante ese tiempo, avancé principalmente en el estudio de la producción de la voz y, a nivel perceptivo, completé el estudio añadiendo 80 jueces italianos para evaluar el efecto de la lengua en la discriminación y reconocimiento de voces. Los resultados finales de todos estos estudios se presentan ahora en la tesis doctoral.

La obra se organiza en 7 capítulos. Los dos primeros ofrecen el marco teórico general basado en la bibliografía más relevante relativa al objeto de estudio. El Capítulo 1 se centra en la producción de la voz y en los registros vocales, prestando especial atención al *falsetto*. El Capítulo 2 trata la percepción de la cualidad de la voz. En el tercer capítulo se expone el planteamiento experimental desarrollado tanto para el estudio de la producción de la voz y del habla como para el diseño de los test perceptivos de discriminación de voces. Los Capítulos 4 y 5 exponen y analizan los resultados obtenidos de ambos análisis, producción y percepción. Dada la complejidad de los mismos, al hilo de la presentación de los datos se ofrece ya una primera explicación en aras de facilitar su interpretación posterior.

En el Capítulo 6 se discuten, a la luz del estado del conocimiento actual, los resultados obtenidos en su conjunto (producción y percepción), se presentan las

[4] Fernández Trinidad, M.; Gil Fernández, J.; Infante Ríos, P., y Lahoz-Bengoechea, J. M. (2014). "El *falsetto* como procedimiento de disimulo de la voz: rasgos alterables y rasgos permanentes", comunicación presentada en el *VI Congreso Internacional de Fonética Experimental*, Valencia 5-7 de noviembre de 2014.

[5] Fernández Trinidad, M. y Rojo, J.M. (2018). Perceptual cues for individual voice quality. En: Gil Fernández J. y Gibson, M. (Eds). *Romance Phonetics and Phonology*, Oxford: Oxford University Press (en prensa).

principales conclusiones y aportaciones que se extraen de la tesis así como las futuras líneas de investigación que se abren a partir de ella.

La obra se cierra con un capítulo dedicado a la transferencia de conocimiento (Capítulo 7) en el que se retoman sintéticamente las conclusiones expuestas en el anterior capítulo y se comentan las principales aplicaciones de los resultados de este estudio al ámbito de la fonética judicial.

Capítulo 1
Producción de la voz y registros vocales

1. Producción de la voz y registros vocales

En este primer capítulo se realiza una breve descripción de cómo se produce la voz humana (§1.1) y se definen los tres registros vocales básicos (§1.2), caracterizándolos en el plano fisiológico, acústico y perceptivo (§1.3 y §1.4). Por último, y de manera muy somera, se comentan algunos métodos y técnicas que permiten analizar indirectamente el comportamiento laríngeo (§1.5).

1.1. La voz: su producción

En el sistema de producción de la voz suelen distinguirse tres partes: el *aparato respiratorio*, el *oscilador* y el *aparato resonador*. Un esquema del sistema de producción de la voz puede observarse en la Figura 1.1.

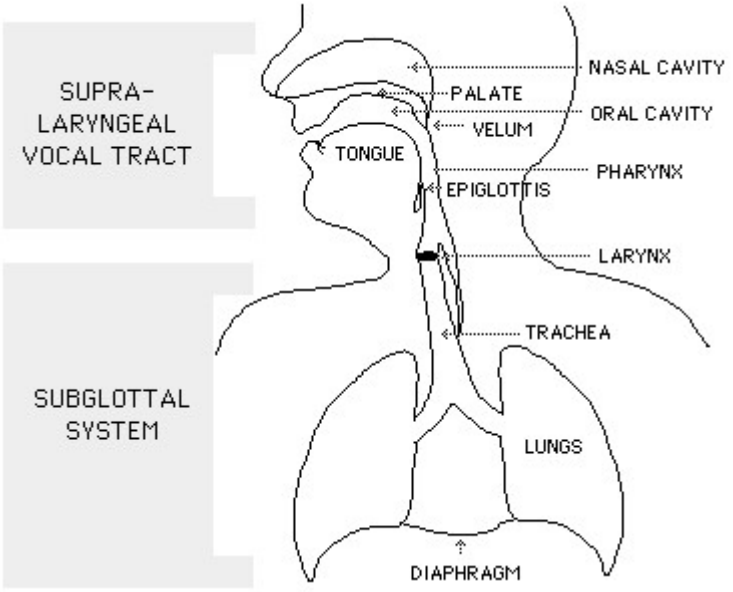


Figura 1.1. Esquema de los principales componentes que intervienen en el sistema de producción de la voz. Imagen tomada de http://www.haskins.yale.edu/featured/heads/mmsp/intro.html

El aparato respiratorio es de fundamental importancia para poder generar voz. Para iniciar su producción, el diafragma se eleva y comprime el aire en los pulmones

provocando la expulsión de la corriente aérea hacia la laringe y, en consecuencia, la activación de esta como oscilador acústico.

La producción de la voz o fonación se entiende como un proceso de transducción de energía aerodinámica, generada por el aparato respiratorio, en energía acústica, radiada hacia los labios. El flujo de aire procedente de los pulmones llega al oscilador y presiona los pliegues vocales, que están cerrados —si estuvieran abiertos no se produciría sonido— haciéndolos vibrar, de modo que la energía que conlleva ese flujo de aire pasa de ser de naturaleza aerodinámica a ser de naturaleza acústica. Esta conversión de energía, por tanto, se produce en la laringe, mediante la vibración de los pliegues vocales, que controlan el flujo de aire que pasa de las cavidades infraglóticas a las supraglóticas (véanse los trabajos de Núñez, 2013a, Titze 2000 [1994], 2006, por ejemplo).

El oscilador lo conforman los pliegues vocales que, como se explicará en seguida, no son cuerpos homogéneos o uniformes. Durante la espiración las presiones de las cavidades subglóticas y las supraglóticas son desiguales. Esta diferencia empuja lateralmente a los pliegues vocales forzando su abertura desde la parte inferior hacia la superior hasta conseguir su total separación. Cuando los pliegues se separan completamente consiguen equilibrarse las presiones sub- y supraglóticas y los pliegues vuelven a cerrarse, completando así un ciclo glótico. La frecuencia de vibración resultante de la sucesión de ciclos como el descrito es lo que se denomina frecuencia fundamental (f_0), cuyo valor viene determinado por la presión subglótica y las características físicas de los pliegues vocales (Stevens, 1998). La frecuencia de vibración de los pliegues (f_0) oscila, por lo general, entre los 100-130 Hz para los hombres y los 200-250 Hz para las mujeres, durante la fonación normal. Así lo explica, por ejemplo, Núñez (2013a):

> [P]ara iniciar la voz, las cuerdas vocales deben aproximarse para formar un canal estrecho o ligeramente cerrado que separa la subglotis de la supraglotis. Una vez que la glotis está cerrada o casi cerrada, comienza la espiración de aire desde los pulmones, con lo que aumenta la presión entre las cuerdas y se produce un empuje en contra de su elasticidad. Cuando la presión del aire es lo bastante alta como para poder separar los tejidos de las cuerdas (estando los aritenoides unidos), el aire fluye a través de la apertura glótica generada. La diferencia entre la presión subglótica y la supraglótica (atmosférica) produce una presión positiva que insufla aire desde la tráquea hacia la superficie medial de las

> cuerdas vocales. En cuanto el flujo aéreo pasa a través del estrechamiento del conducto que determina la glotis, la velocidad de sus moléculas aumenta, determinando una reducción de la presión transglótica que produce una presión negativa. Una vez que el aire fluye por la ahora abierta glotis, varias fuerzas se combinan para cerrarla de nuevo. […] Este ciclo de vibración se denomina «ciclo glótico». Los ciclos vibratorios suceden con una frecuencia media de 110 por segundo en la voz masculina y de 200 por segundo en la femenina (p. 61).

El resonador lo constituyen parte de la laringe, la faringe y las cavidades oral y nasal. Se trata de un filtro ajustable que modificará, según sus resonancias, el sonido original generado en la laringe. Es decir, los pliegues vocales generan un sonido vocal primario, el cual viene posteriormente transformado (filtrado) por las cavidades de resonancia del tracto vocal o resonador y esto da lugar al sonido que se irradia desde la boca. La forma y el tamaño del filtro se modificarán dependiendo de los cambios articulatorios que se produzcan en el tracto vocal supraglótico. Por eso se considera que los resonadores no generan la energía sino que responden a la energía que reciben (Núñez, 2013a, p. 68). En función de cómo sea el filtro, el sonido primario experimentará una serie de cambios por la resonancia: algunos de los conjuntos de armónicos que se generaron primero en la laringe (fuente) se verán amplificados —los denominados *formantes*—, mientras que otros, en cambio, resultarán atenuados (véase la Figura 1.2).

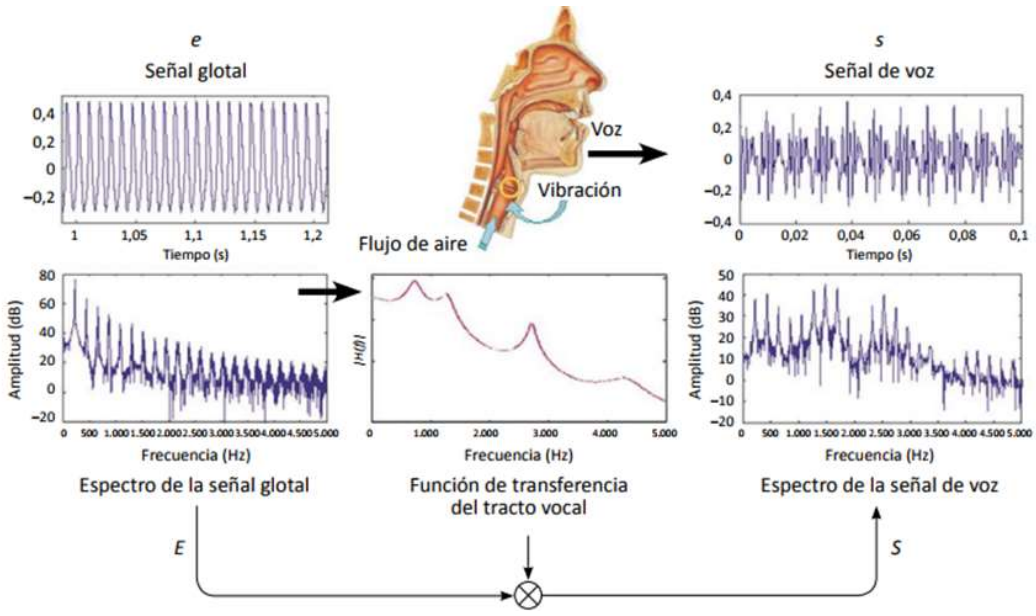


Figura 1.2. El modelo de la fuente y el filtro en la producción de la voz. Imagen tomada de Godino y Gómez-Vilda (2013, p. 97)

La voz puede definirse y analizarse de forma amplia o acotada. Cuando se aplica la definición más restrictiva de *voz*, esta se concibe como el sonido resultante del funcionamiento de la laringe, de la acción de los pliegues vocales. Según esta definición más acotada, voz y fuente laríngea (ing. *laryngeal source*) se consideran términos equivalentes. Asimismo, la voz puede entenderse en un sentido más amplio, equivalente a habla y, en tal caso, no solo se atiende al funcionamiento de los pliegues vocales sino que también interesan los efectos de las resonancias de las cavidades supraglóticas (Kreiman y Sidtis 2013 [2011], pp. 5-6).

1.2. Los registros vocales: definición

Las diferentes configuraciones fisiológicas que puede adoptar la laringe generan un resultado acústico específico, asociado a su vez, desde el punto de vista perceptivo, con una cualidad de voz determinada (Pérez Sanz, 2011, p. 50). Todos los autores coinciden en que tanto en el habla como en el canto existen registros. Sin embargo, no existe acuerdo ni en su definición, ni en su número ni en su denominación (véase Henrich, 2006). En cualquier caso, Hollien (1974) define un registro vocal como:

> a series or range of consecutively phonated frequencies which can be produced with nearly identical vocal quality and that ordinarily there should be little or no overlap in fundamental frequency (f0) between adjacent registers. Furthermore, I maintain that a voice register is a totally laryngeal event and, before the existence of a particular register can be established, it must be operationally defined: (1) perceptually, (2) acoustically, (3) physiologically, and (4) aerodynamically (pp. 125-126).

El concepto de registro se vincula con la frecuencia de fonación porque existe una relación directa entre la f_0 y cada uno de ellos. Los registros vocales básicos —para autores como Titze (2000 [1994]), entre otros— son tres, del extremo más agudo al más grave: *falsetto, modal voice* y *creak* (o *pulse register* o *vocal fry),* mencionados en la bibliografía en español con las denominaciones *falsete*, *voz modal*, *crepitación*, *voz pulsada*, *frito vocal*, entre otras que pueden encontrase también (véase más adelante

§1.4.2). Cada uno de estos registros se caracteriza por presentar valores de f_0 altos, medios y bajos, respectivamente (Colton, 1969; Colton y Hollien, 1973; Hollien, 1974; Hollien y Michel, 1968; Laver, 1980).

Titze (2000 [1994]) define los registros como "...perceptually distinct regions of vocal quality that can be maintained over some ranges of pitch and loudness" (p. 282). Como se explica claramente en Pérez Sanz (2011), para un mismo registro vocal el mecanismo laríngeo básico funciona del mismo modo dentro de un determinado rango de frecuencias. Luego, para alcanzar un determinado rango frecuencial y de intensidad no se puede seguir con la misma configuración laríngea y la voz pasa necesariamente al siguiente registro, lo que origina su cambio cualitativo (pp. 50-51).

El acuerdo no es total entre los estudiosos de la fonación respecto de los límites frecuenciales precisos que corresponden a cada registro, y distintos autores proponen diferentes rangos de frecuencias para señalar el pasaje de una 'zona' a otra. Los registros se solapan en ciertos rangos de frecuencias. El tránsito de un registro al siguiente puede ser voluntario o involuntario y puede darse de forma más o menos abrupta dependiendo del entrenamiento vocal. Por lo general, los hablantes o los cantantes no entrenados experimentan un quiebre (*break*) en la producción de la voz al moverse de un registro al siguiente, mientras que en las voces entrenadas este tránsito es siempre voluntario y por lo común mucho más suave (véanse Henrich, 2006 y Titze, 2000 [1994], pp. 281-301, para una completa discusión acerca de las transiciones entre registros).

1.3. Registros vocales: parámetros de descripción

Para describir y caracterizar los diferentes registros de forma objetiva se suele recurrir a dos grupos de parámetros: (1) parámetros fisiológicos de la actividad glótica, el comportamiento de los pliegues vocales, y (2) parámetros espectrales de la acción glotal.

1.3.1. Parámetros fisiológicos de la actividad glótica

Como fue antes mencionado (§1.2), los registros dependen del funcionamiento glótico y no de la configuración del tracto vocal en su totalidad. Según Laver (1980), los principales parámetros de la actividad glótica que conviene considerar para caracterizar los tres registros mencionados son la fuerza de cierre o tensión aductora de los pliegues vocales —ing. *Adductive Tension* (AT) —, el grado de compresión en la línea media—*Medial Compression*— (MC) y el grado de tensión longitudinal—*Longitudinal Tension*— (LT). Véanse Figura 1.3 y Tabla 1.1.

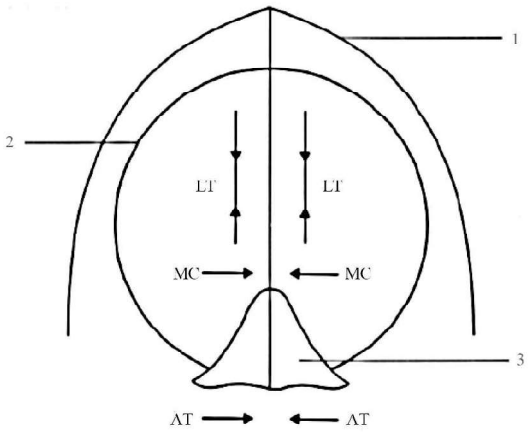


Figura 1.3. Esquema de los parámetros laríngeos, tomada de Laver (1980, p. 109). Tensión Aductora (AT), Compresión Media (MC) y Tensión Longitudinal (LT). 1. Cartílago Tiroides, 2. Cartílago Cricoides, 3. Cartílago Aritenoides.

Tabla 1.1.*Configuración básica de los pliegues vocales en los tres registros. Cuadro adaptado de San Segundo, Alves y Fernández Trinidad (2013, p. 588).*

Registros	Tensión Aductora (AT)	Compresión Media (MC)	Tensión Longitudinal (LT)
Falsetto	menor	menor	mayor
Modal	moderada	moderada	moderada
Creak	mayor	mayor	menor

Además de la configuración específica que puedan adoptar en cada caso, que se verá al presentar la caracterización de cada registro (§1.4), es necesario conocer la composición estructural de los pliegues, así como considerar la masa y la tensión implicados en la vibración durante los diversos mecanismos de fonación (véase por

ejemplo, Roubeau, 1993). Desde un punto de vista histológico, los pliegues vocales están formados por varias capas de tejidos diferentes. Desde lo más externo a lo más interno: *epitelio de la mucosa*, *lámina propia* (compuesta a su vez por tres capas) y *músculo vocal*, como se observa en la Figura 1.4.

La mucosa del pliegue vocal está formada por un epitelio que le proporciona una apariencia de "brillo blanquecino". La mucosa tiene la propiedad de ondular, es decir, puede moverse y recuperar su posición de inicio (Sañudo, Maranillo y León, 2013, p. 31). Desde una perspectiva mecánica, "el epitelio debe contemplarse como un fino estuche con la función de mantener la forma de la cuerda vocal" (Núñez, 2013a, p. 57).

La lámina propia se divide en tres capas o estratos (superficial, medio y profundo) atendiendo a su composición histológica. Esta diferenciación es importante para describir su funcionamiento durante la vibración. En palabras de Núñez (2013a, p. 57), el estrato más superficial, comparable a una "masa de gelatina suave", es muy flexible desde el punto de vista mecánico. La capa intermedia, formada fundamentalmente por fibras elásticas, es menos flexible que el estrato más superficial y se comporta como un "mazo de tiras de goma elástica". Luego, el estrato más profundo de la lámina propia, formado principalmente por fibras de colágeno, resulta ser menos flexible y su comportamiento se compara con un "mazo de hilos de algodón". Por último, se encuentra el *músculo vocalis* o músculo Tiroaritenoideo (TA), que constituye el cuerpo principal del pliegue y su rigidez se modifica en función de la propia contracción muscular.

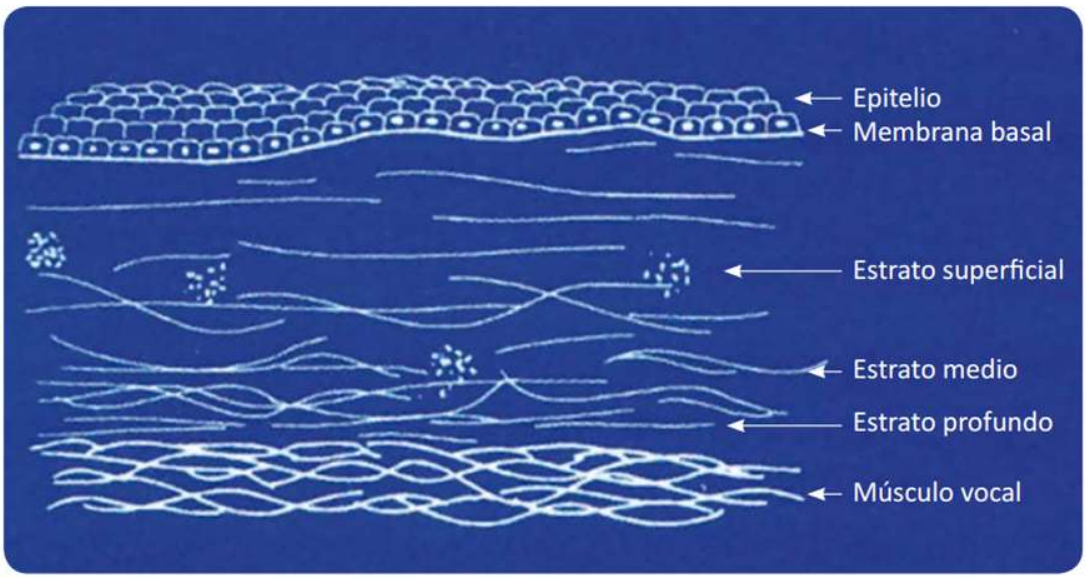


Figura 1.4. Representación de las diferentes capas o estratos del pliegue vocal vistas con un microscopio óptico. Imagen extraída de Cervera y Núñez (2013, p. 51).

Siguiendo a Titze (1994), el comportamiento biomecánico de los pliegues vocales puede afectar a dos partes diferentes, denominadas *cuerpo* y *cubierta*. Como puede observarse en la Figura 1.5, el cuerpo del pliegue vocal estaría conformado por el músculo Tiroaritenoideo –llamado también, *músculo vocalis*– y la capa profunda de la *lámina propia*, mientras que el epitelio y las capas superficial e intermedia de la lámina propia constituyen la cubierta (Titze, 2000 [1994], p. 17). Mecánicamente, los pliegues vocales se describen atendiendo al comportamiento vibrátil de estas dos partes o estructuras, el cuerpo y la cubierta. Cuando los pliegues vocales empiezan a moverse, puede darse un desfase entre el movimiento del cuerpo y el de la cubierta, lo cual será importante, como se verá, para los registros de fonación (véanse§ 1.4.1, 1.4.2 y 1.4.3).

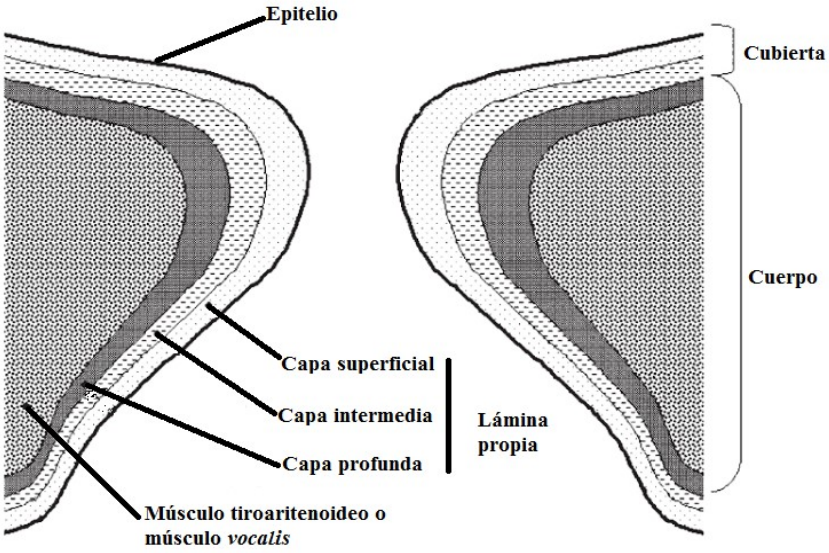


Figura 1.5. Dibujo esquemático de la estructura en capas de los pliegues vocales. Imagen adaptada de Jiang (2008, p. 18).

Como expone claramente Núñez (2013a), la masa y la tensión implicadas en la vibración de los pliegues resultan ser importantísimos para determinar la frecuencia de fonación, según la teoría cuerpo-cubierta:

> La cubierta es flexible, elástica y no contráctil, mientras que el cuerpo es más rígido y tiene propiedades contráctiles activas que permiten ajustar la rigidez y concentrar la masa. La tensión global de las cuerdas depende del acoplamiento de la cubierta al cuerpo, que varía su rigidez en función de la contracción muscular. Durante la contracción aislada del músculo tiroaritenoideo, el cuerpo de la cuerda aumenta su rigidez por el acortamiento del músculo, mientras que la cubierta se torna más laxa y flexible. Esta diferencia de tensión entre ambas capas de la cuerda, la combinación del estiramiento longitudinal y la contracción de la masa muscular, determina la amplitud de la onda mucosa. El modelo cuerpo-cubierta es útil para explicar la interacción del músculo cricotiroideo, principal control del tono, y de las contracciones del músculo tiroaritenoideo en la regulación de la frecuencia fundamental (Núñez, 2013a, p. 61).

El hecho de que los pliegues estén acortados o elongados, laxos o tensos, y que vibren con más o menos masa, es decir, que se acerquen más gruesos o más delgados, determinará que los valores de la f_0 sean más bajos o altos, respectivamente. Estas propiedades de los pliegues vienen determinadas principalmente por el comportamiento de dos músculos, el Tiroaritenoideo y el Cricoaritenoideo (Titze, 2000 [1994], pp. 214 y 217). Si el Tiroaritenoideo (TA) o músculo *vocalis* se contrae, los pliegues se acortan y aunque este músculo esté en tensión, las capas externas están laxas y vibran con más masa. Al contrario, cuando el Tiroaritenoideo está sin actividad, relajado, los pliegues se elongan y la tensión la soportan las capas externas, que están más rígidas y finas. De otra parte, el Cricotiroideo (CT) puede estar activo o sin actividad, esto es, puede tensar o no los pliegues.

Titze (2000 [1994], p. 214) explica que según cómo se dé el comportamiento conjunto o simultáneo de estos dos músculos la f_0 podrá variar en una dirección u otra[6]:

a) Cuando el Tiroaritenoideo está relajado pero se activa el Cricotiroideo, se da un aumento importante de la f_0 porque los pliegues están estirados y finos y además tensos o rígidos debido a la acción del Cricotiroideo. Al relajarse el músculo *vocalis* las capas más externas del pliegue quedan más delgadas. Como consecuencia de todo esto, la f_0 aumenta de forma importante.

[6] Puede encontrarse también una clara explicación sobre esto en Pérez Sanz (2011, p. 23).

b) Cuando ocurre la situación contraria, es decir, cuando el Tiroaritenoideo está activo y se contrae, los pliegues se acortan, engrosando el pliegue vocal y concentrando, por tanto, más masa en vibración, y, aunque el músculo *vocalis* esté en tensión por estar activo, las capas más externas oscilan con relativa laxitud. Como además el Cricotiroideo permanece sin actividad, el resultado es un descenso de la frecuencia de vibración, la oscilación se vuelve más lenta y el tono resultante es más grave.

c) Si el Tiroaritenoideo y el Cricotiroideo están ambos en actividad o contraídos se genera tensión pero no se modifica la longitud del músculo, puesto que la acción del Tiroaritenoideo se opone a la del Cricotiroideo (recuérdese que la activación del primero supone un acortamiento de los pliegues mientras la acción del segundo un alargamiento), por tanto, el efecto de ambos se neutraliza y la frecuencia fundamental aumenta solo ligeramente.

Lo dicho anteriormente puede esquematizarse de la siguiente forma (Tabla 1.2).

Tabla 1.2. *Esquema del comportamiento de los músculos Cricotiroideo y Tiroaritenoideo y su efecto en la configuración de los pliegues vocales y en el comportamiento de la* f_0.

	TA activo	TA sin actividad
CT activo	mucha tensión general aumento moderado de la f_0	pliegues elongados, rígidos y finos aumento importante de la f_0
CT sin actividad	pliegues acortados, laxos y gruesos descenso de la f_0	

El *falsetto*, como se verá en §1.4.3., respondería a la primera configuración fisiológica descrita antes en a). De hecho, en rehabilitación vocal se les suele pedir a los pacientes que padecen disfonías hipertónicas que realicen una emisión en *falsetto* para relajar el Tiroaritenoideo y tensar el Cricotiroideo, como explica Coll (2013, p. 465).

1.3.2. Parámetros espectrales de la acción glotal

La vibración de los pliegues vocales produce una señal compleja periódica compuesta por tonos puros o armónicos, que son múltiplos exactos de la frecuencia fundamental y que aparecen espaciados a intervalos regulares. En un espectro, la frecuencia que presenta mayor amplitud es la del primer armónico de la voz, es decir, la de la frecuencia fundamental —representada como f_0 o H1 *(primer armónico).* A partir de este, los armónicos sucesivos (H2, H3… H15…) van perdiendo intensidad a medida que aumenta la frecuencia hasta desaparecer por completo (véase por ejemplo, Núñez, 2013a para más detalles y la Figura 1.2)

Como se verá un poco más adelante (§1.4), las medidas espectrales que pueden extraerse de fuente glotal a partir del sonido emitido por un locutor suelen proporcionar información útil sobre los patrones de vibración de los pliegues vocales y en consecuencia sobre la cualidad de voz. Muchos trabajos, como los de Ní Chasaide y Gobl (1997); Kreiman, Gerratt, Garellek, Samlan, y Zhang (2014), entre los más recientes, muestran que diferentes tipos de voz pueden distinguirse atendiendo a la pendiente o inclinación espectral general —ing. *general spectral tilt or slope*— y a la relación entre la intensidad de la frecuencia fundamental y la de los demás armónicos del espectro. Lo que es más importante aún, estas diferencias suelen ser perceptivamente relevantes en la distinción de la cualidad de la voz, lo que las vuelve aún más útiles para comprobar los cambios en la fonación (véase por ejemplo, Gobl y Ní Chasaide, 2003; Holmes, 1973; Kreiman, Gerrat y Antoñanzas-Barroso, 2007; Kreiman *et al*., 2014). Por este motivo, se emplean frecuentemente en los estudios fonéticos para inferir el comportamiento de la laringe (véase por ejemplo, Gobl y Ní Chasaide, 2003; Holmes, 1973; Sundberg, Titze y Scherer, 1993, entre muchos otros). No obstante, y tal como se ha señalado ya en la bibliografía especializada, las mediciones obtenidas a través de estos análisis no siempre son indicios definitivos del comportamiento laríngeo (para una completa discusión sobre este punto véase por ejemplo, Ní Chasaide y Gobl, 1997).

A continuación, se ofrece una caracterización fisiológica y acústica de los tres registros básicos, considerando los parámetros recientemente mencionados y también las impresiones perceptivas que se han recogido de la bibliografía.

1.4. Caracterización fisiológica, acústica y perceptiva de los registros

1.4.1. Voz modal

El registro modal es considerado un tipo de fonación neutro, y es, en las voces masculinas no patológicas, el modo de fonación más habitual y espontáneo. Las mujeres, sin embargo, no siempre producen este tipo de fonación y es bastante frecuente encontrar en las voces femeninas otros comportamientos o hábitos de fonación, como la utilización de la *breathy voice*[7] (Hanson, 1997), o incluso del *falsetto* (Sulter y Albers, 1996). El término 'modal' fue sugerido por Hollien (1974) para evitar el de 'normal', puesto que esto implicaría de algún modo considerar por oposición los mecanismos no modales como 'anormales'.

Los valores de f_0 no son considerados ni muy altos (como en el *falsetto*), ni muy bajos (como en el registro *creak*). Se ubican en torno a los 125 Hz para los hombres, 250 Hz para las mujeres, y 350 Hz para los niños (Cobeta y Núñez, 2013, p. 193). Los datos son mucho menos específicos cuando se describe su intensidad característica y se dice que se distribuye en un amplio rango (Hollien, 1974; Laver, 1980). Por lo general, presenta valores bajos de perturbación de la frecuencia — en inglés, *jitter*— y de la amplitud —en inglés, *shimmer* — (Davis, 1979; Hirano, Yoshida y Tateishi, 1985; Monsen y Engebretson, 1977; entre muchos otros).

Las descripciones sobre el mecanismo de voz modal (por ejemplo, Hollien, 1974; Laver, 1980) señalan que los pliegues vocales presentan un grado moderado de tensión longitudinal que aumenta a medida que se incrementa también la frecuencia fundamental, un grado moderado de tensión aductiva, sin excesiva presión en la línea media. La vibración es periódica y la ondulación, amplia.

Titze explica (2000 [1994], p. 291) que, durante la fonación modal, el músculo Tiroaritenoideo (TA) se contrae, lo cual provoca que los pliegues se acorten y genera un engrosamiento del pliegue vocal, dejando más laxas las capas externas. Existe un desfase entre el contacto de los bordes superior e inferior de los pliegues vocales porque los bordes inferiores entran primero en contacto, y los bordes superiores lo hacen

[7] *Breathy voice* refiere a un tipo de fonación caracterizado fisiológicamente por presentar mucho escape de aire por el espacio glótico como consecuencia de una aducción o cierre incompleto de los pliegues vocales. Se han propuesto varias traducciones al español, como por ejemplo, "voz soplada", "voz aérea" o "voz de hálito", (véanse Gil, 2007, p. 217, y Poyatos, 1993, 1994 y 2002).

después. Los pliegues contactan, por tanto, en toda su longitud (desde el borde inferior hasta el superior) y este cierre es prolongado, lo cual se traduce acústica y perceptivamente en una voz rica en armónicos y sin ruido turbulento audible (Pérez Sanz, 2011, p. 3). Para más detalles pueden consultarse, Gunter (2003); Hollien (1974); Laver (1980); Pérez Sanz (2011); Rothenberg (1973); Titze (1994); van den Berg (1968), entre otros. El declive espectral es de 12 dB por octavas, es decir, cada duplicación de la frecuencia fundamental produce una pérdida o caída en la amplitud de 12 dB.

1.4.2. Creak

El registro más grave de los mencionados se conoce en inglés como *creak*, *pulse register* o *vocal fry,* términos traducidos en español como 'crepitación', 'voz pulsada', 'frito vocal', 'voz quebrada', 'voz rota', entre otras muchas que pueden encontrarse en la bibliografía especializada[8]. Su descripción y análisis ha generado grandes discusiones, principalmente cuando se intenta desligarlo de otros mecanismos muy similares como la voz crepitante (ing. *creaky voice*) o la laringalización (ing. *laringealization*), por ejemplo (véase Infante, 2015, p. 111 y Lirio, 2016, p.136). Trabajos recientes[9] han contribuido muy especialmente a aclarar muchas cuestiones al respecto. Sin embargo, aquí solo se caracterizará el *creak* en tanto en cuanto se opone a la voz modal y al *falsetto*, únicos dos registros que se consideran en esta tesis.

El *creak* es un tipo de fonación asociado con valores bajos de frecuencia e intensidad y escasa presión subglótica, y presenta valores altos de perturbación de la frecuencia o *jitter* (Hollien, 1974; Laver, 1980; Monsen y Engebretson, 1977; entre otros). La masa del cuerpo implicada en la vibración es mayor a la de la cubierta, los pliegues se acercan cortos y distendidos para la vibración (Hirano, 1982; Hollien, 1974; Pérez Sanz, inédito; Gómez-Vilda y Pérez Sanz, inédito, entre otros). Además, de que la vibración entre el cuerpo y la cubierta suele estar desfasada, la fase de cierre suele ser muy larga, lo que implica valores bajos de cociente de abertura (Hollien, Girard y

[8] Las traducciones que se han propuesto han sido muchas y variadas (véase Infante, 2015; Pérez Sanz, 2011; Poyatos 1993, 1994). Por este motivo, y para evitar confusiones, se ha considerado pertinente utilizar siempre uno de los términos ingleses, en concreto, *creak*.

[9] Por ejemplo, Keating, P., Garellek, M., y Kreiman, J. (2015). Acoustic properties of different kinds of creaky voice. En *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow; Keating, P., y Garellek, M. (2015). Acoustic analysis of creaky voice. Póster presentado en una sesión especial sobre voz crepitante en el *Encontro Anual da Linguistic Society of America em Portland* (OR). En el ámbito hispano, los estudios de Heriberto Avelino son, sin duda, trabajos indispensables.

Coleman, 1977 y Whitehead, Metz y Whitehead, 1984). El declive espectral para este registro es menor que el de la voz modal, de 6 dB/octava.

En esta configuración los pliegues están longitudinalmente laxos y esta falta de tensión longitudinal hace que, en la mayoría de los casos, aunque no siempre sucede así, la vibración pueda no ser periódica[10], que se produzcan pulsos irregulares de muy baja amplitud (Infante, 2015, p. 109; Pérez Sanz, 2011, p. 54). Para más detalles, el lector interesado puede consultar Allen y Hollien (1973), Hollien (1974), Hollien y Colton (1969); Monsen y Engebretson (1977); Whitehed *et al*., (1984) entre otros.

Se da en frecuencias muy bajas tanto en hombres como en mujeres (Hollien y Michel, 1968), aunque los valores y rangos presentan variaciones según los autores (véase Infante, 2015). Hollien (1974) sitúa la f_0 media del registro para voces masculinas por debajo de los 70 Hz; Laver (1980) la ubica entre los 35 y los 52 Hz. Titze (2000 [1994], p. 283) explica que cuando la f_0 es demasiado baja (< 70 Hz), la rapidez de respuesta del sistema auditivo permite detectar tramos silenciosos[11] entre las explosiones de energía acústica de cada ciclo sucesivo y la frecuencia se percibe a "impulsos" (véase también Pérez Sanz, 2011: pp. 51-52). Esto justificaría que auditivamente se lo haya descrito como una "rough-sounding voice" (Childers y Lee, 1991; Wendahl, Moore y Hollien, 1963), un sonido similar al de "a rapid series of taps, like a stick being run along a railing" (Catford, 1964) o al de "popping of corn" (Henton y Bladon, 1987), o "food cooking in a hot frying pan" (Ishi, Sakakibara, Ishiguro y Hagita, 2008). Este registro es mucho más frecuente en el habla que en el canto y en varias lenguas se utiliza con finalidades lingüísticas (Gordon, y Ladefoged, 2001) o extralingüísticas y paralingüísticas (Esling, 1978; Laver, 1980, 1994; Ní Chasaide y Gobl, 1997; Ogden, 2001 y 2004; Poyatos 2002, entre muchos trabajos).

1.4.3. Falsetto

El *falsetto* es utilizado por hombres y mujeres en el habla y en el canto. Es el registro más agudo de los tres registros vocales mencionados, generalmente supone el aumento de la f_0 en al menos una octava respecto de la voz modal. Se ha descrito

[10]Algunos estudios sin embargo sostienen que "The glottal vibration pattern is characterized by short pulses, which can be periodic and single, periodic and multiple (double, triple), or aperiodic (single and multiple), (Henrich, 2006, p. 10).
[11]Se considera que hay un silencio cuando la onda ha perdido un 99% de la amplitud inicial del pulso (Titze, 2000 [1994]).

(Childers y Lee, 1991) que el rango en el *falsetto* puede variar desde 275 hasta 634 Hz, en voces masculinas, y moverse entre los 495 y los 1131 Hz, en las femeninas (Colton, 1969; Hollien, 1974). Roubeau, Henrich y Castellengo (2009) señalan que el pasaje de voz modal a *falsetto* en voces masculinas (durante un *glissando*[12]) puede iniciarse a partir de los 238 Hz, pero, como se ha dicho antes, no todos los autores coinciden. Sin embargo, el registro siempre se sitúa entre una altura medio-alta o alta del rango frecuencial. Por lo común, presenta niveles bajos o medio-bajos de intensidad[13] y perturbaciones tanto de *jitter* como de *shimmer* (Colton, 1973 y Colton y Hollien, 1973; Hollien, 1974; Monsen y Engebretson, 1977; Laver, 1980). No parece cumplir funciones lingüísticas en las lenguas (Poyatos, 2002, p. 33). Es posible que en el habla pueda aparecer de forma más o menos puntual como variable estilística en situaciones comunicativas muy concretas, o incluso también para que un hablante manifieste simbólicamente una identidad sexual o personal (véase Podesva, 2007, pp. 496-498). Como se verá en el siguiente capítulo (§ 2.4 y § 2.4.1.), los delincuentes recurren con cierta frecuencia al *falsetto* para disimular su voz, para despistar sobre su identidad.

Desde el punto de vista de la producción presenta lógicamente diferentes características fisiológicas respecto de los otros dos registros. Como explica Titze (1994), a diferencia de lo que ocurre con la voz modal, en la que toda la superficie del pliegue vocal se pone en movimiento, la vibración en el *falsetto* tiende a afectar solamente a la cubierta, no al cuerpo y, por ello, la amplitud de vibración se ve reducida. El contacto es muy superficial, incluso a veces no existe siquiera contacto, por lo que el número de armónicos que se genera suele ser escaso. En comparación con la voz modal, en el *falsetto* se reduce la masa de los pliegues vocales en vibración al tiempo que aumentan su tensión y rigidez, fundamentalmente en las capas más externas de los pliegues vocales (Hirano, 1982; Pérez Sanz y Gómez-Vilda, inédito; Titze, 2000 [1994]). La tensión se incrementa, explica Titze, cuando la parte que vibra es más estrecha, puesto que la fuerza se concentra en una superficie menor. Como consecuencia, la vibración es muy rápida, y este incremento en la frecuencia de vibración redunda en un aumento del tono (Hollien, 1974; Laver, 1980, Titze, 1994). Titze (2000 [1994], p. 291) explica que una estrategia de aumento de la f_0, que conlleva un salto cualitativo y, por tanto, el paso hacia el registro más agudo —el *falsetto*—, es

[12] En canto, se denomina *glissando* a la sucesión rápida y continua de una nota a la siguiente.
[13] Siempre considerando la voz hablada, porque naturalmente en la voz cantada se puede producir *falsetto* con intensidades muy altas.

aumentar la tensión del CT (Cricotiroideo) y mantener inactivo el TA (Tiroaritenoideo), y es, precisamente, lo que parece ocurrir en el *falsetto*. Recuérdese que, como ha sido antes señalado, en este registro el Tiroaritenoideono está inactivo, y casi toda la tensión provocada por la actividad del Cricotiroideo queda soportada por la cubierta. Una vez en el ámbito del propio registro del *falsetto*, cualquier aumento de la f_0 implica mayor actividad del músculo Cricotiroideo, que aplica más tensión tanto en la cubierta como en el cuerpo, que no está activo (véase Titze, 2000 [1994], p. 214 y pp. 289-291).

En el *falsetto*, y a diferencia de lo que ocurre en el registro modal, la fase de abertura es más larga y la de cierre más breve, incluso en ocasiones los pliegues no llegan a cerrarse del todo (Henrich, d'Alessandro, Castellengo y Doval, 2005; Hollien, 1974; Kitzing, 1982; Monsen y Engebretson, 1977; Roubeaut *et al.*, 2009; Titze, 1994). Los valores del cociente de abertura son, pues, siempre más altos respecto de la voz modal (Henrich *et al*., 2005; Roubeau *et al*., 2009). El cociente de abertura se ha señalado como un buen indicador para distinguir la voz modal del *falsetto*, y todos los autores están de acuerdo en que es siempre mayor en el *falsetto* (véanse por ejemplo, Childers y Lee, 1991; Henrich *et al.*, 2005; Roubeaut *et al.*, 2009). En este mecanismo de fonación, los aritenoides tienden a separarse y a dejar un continuo escape de aire en la parte posterior de la glotis, lo cual se traduce acústicamente en un fuerte declive espectral, de 18 dB/octava, una pérdida de energía en la voz y esa cualidad aflautada (*piping voice*) y en ocasiones *breathy* que caracteriza auditivamente al *falsetto* (véanse Colton y Hollien, 1973; Childers y Lee, 1991; Hollien, 1974; Laver, 1980; Poyatos, 2002, p. 34 y Titze, 1994).

1.5. Métodos indirectos para evaluar el comportamiento laríngeo

1.5.1. El análisis electroglotográfico

El análisis electroglotográfico (EGG) es una técnica clásica, no invasiva, que permite la exploración indirecta del comportamiento laríngeo. La señal que se obtiene es la onda laríngea exclusivamente. Para más detalles pueden consultarse Childers y Krishnamurthy (1985), Childers, Hicks, Moore y Alsaka (1986), Childers, Hicks, Moore, Ezkenazi y Lalwani (1990), Titze, (1990). Dicha señal se obtiene mediante la colocación de dos electrodos que se disponen a ambos lados del cuello, a la altura de la nuez o bocado de Adán, sujetos por un elástico regulable. Una explicación clara de esta

técnica se ofrece en Pérez Sanz (2011) y se resume aquí. El electroglotógrafo emite una corriente de muy baja intensidad y alta frecuencia que se transmite con gran facilidad de un pliegue vocal al otro y al mismo tiempo no daña a la persona. Cuando los pliegues están separados, es decir, cuando no hay contacto entre ellos y el aire se concentra en la glotis, la corriente no puede transmitirse puesto que el aire es un mal conductor eléctrico. Cuando los pliegues se juntan dan paso a la corriente. La señal electroglotográfica refleja el aumento o disminución de la conductancia a través de aumentos y disminuciones de la amplitud de las ondas que emite. La onda llegará al máximo de su amplitud cuando la corriente pase de un electrodo al otro atravesando de un extremo al otro los pliegues vocales en contacto. Contrariamente, la onda alcanzará mínima amplitud cuando se interrumpa el paso de la corriente eléctrica de un electrodo a otro debido a la separación de los pliegues vocales (Pérez Sanz, 2011, pp. 61-68). En el *falsetto*, como se ha dicho, debido a que el contacto suele ser superficial y breve y en ocasiones los pliegues no llegan siquiera a juntarse, la señal electroglotográfica se pierde por momentos.

La técnica de EGG es muy utilizada para estudiar las patologías vocales y la voz cantada. A pesar de sus indudables ventajas y de las posibilidades de análisis que presenta este método (véase por ejemplo, Howard, 2009), en el caso del *falsetto* existen problemas para interpretar los datos obtenidos mediante esta técnica.

1.5.2. Modelos de estimación de la fuente glótica

En los últimos años se ha dado un fuerte impulso para crear modelos eficaces que reconstruyan la fuente glótica (véase por ejemplo, Monzo, 2010). Fant (1960) fue el primero en introducir el modelo de la fuente-filtro para la producción de habla. Como se ha mencionado en §1.1, la laringe actúa como fuente de sonido primaria y el sonido fuente producido en la laringe se modificará cuando pase por el filtro —el tracto vocal—, según su configuración y resonancias (véase Godino y Gómez-Vilda, 2013, para más detalles). Se han desarrollado varios modelos de la fuente glotal que intentan representar, con la menor cantidad de parámetros posible, el funcionamiento de la laringe durante la fonación. A lo largo del tiempo se han desarrollado –y se siguen

construyendo– varios tipos de modelos, unos son mecánicos y otros computacionales; algunos intentan modelar la onda glótica propiamente, otros su derivada[14].

El programa informático que se ha utilizado en esta tesis para el análisis de la fonación ha sido el *software* *BioMet®Phon*, de *BioMetroSoft®* (http://www.biometrosoft.com), el cual se explicará con mayor detalle en el Capítulo 3. Se trata de un modelo computacional de la onda glótica y de su derivada. El programa trabaja con la señal de voz como entrada, la cual se obtiene a partir de la señal microfónica y, a partir de ella aísla, mediante un filtrado inverso[15], información acústica y biomecánica[16] sobre el comportamiento de los pliegues vocales, descontando de la señal vocal en su conjunto la función de transferencia que ejerce el tracto vocal, la cual incluye también los efectos de la radiación labial que son del orden de 6 dB/octava (véase Palacios Alonso, 2018, pp. 65-67 para una explicación más detallada del proceso). *BioMet®Phon* considera tanto la onda glotal como su derivada porque la información que aporta cada una de ellas es complementaria. Este tipo de análisis resulta muy útil, lógicamente, para detectar patologías vocales (véase Gómez-Vilda, Fernández-Baillo, Rodellar-Biarge, Nieto-Lluis, Álvarez-Marquina, *et al*., 2009).

[14] Derivada del flujo o variación temporal del flujo.
[15] La técnica del filtrado inverso se utiliza para estudiar la fuente de sonido que se genera en la laringe. Idealmente, dicha técnica cancela el efecto del filtrado producido por el tracto vocal. Como se explica en Cobeta y Núñez (2013, p. 197): "El filtrado inverso es una técnica no invasiva que refleja el movimiento vibratorio de las cuerdas vocales, reconstruyendo la onda de excitación glótica mediante la creación de un filtro que revierte la influencia del tracto vocal sobre ésta."
[16] "Biomecánica: estudio de la mecánica del tejido biológico" (Núñez, 2013b, p. 611).

Capítulo 2
La percepción de la cualidad de voz y el reconocimiento de hablantes individuales

2. La percepción de la cualidad de la voz y el reconocimiento de hablantes individuales

Este capítulo está dedicado a describir, a partir de la bibliografía disponible, la forma en que se percibe la voz humana y cómo a partir de ella los oyentes son a veces capaces de diferenciar, reconocer o identificar a diferentes personas. Primeramente, se conceptualizan los términos *voz* y *cualidad de voz* (§2.1). En §2.2 se sintetizan los atributos físicos, biológicos, psicológicos y sociales que sobre el hablante puede transmitir su propia voz; en § 2.3 se aborda el tema de la cualidad de la voz y el reconocimiento de locutores, pasando revista a los tipos de procesos cognitivos subyacentes (§2.3.1), y a algunos de los factores condicionantes del reconocimiento perceptivo de hablantes individuales (§2.3.2). A partir de la segunda mitad del capítulo, el interés se centra en el estudio de dos factores que, según se ha descrito hasta ahora en la bibliografía, parecen interferir de modo muy importante en el reconocimiento de voces, cuales son el disimulo voluntario de la voz (§2.4) y el efecto del idioma (§2.5). En concreto, el recorrido bibliográfico se detendrá particularmente en el disimulo mediante la elevación drástica de la f_0 y el cambio hacia el registro de *falsetto*, en el primer caso, y, respecto del segundo factor, interesarán particularmente las instancias en las que hablantes y oyentes no comparten la misma lengua, frente a otras situaciones en las que hablantes y oyentes son hablantes nativos de la misma lengua y variedad dialectal.

2.1. La voz: su percepción

En la bibliografía especializada suele aparecer una diferenciación terminológica y conceptual entre *voz* y *cualidad de voz*. Como explican Kreiman y Sidtis (2013 [2011] [17], p. 6), ambas expresiones, además, pueden hacer referencia a conceptos más o menos amplios según los enfoques y propósitos. El término *voz* es ambiguo

[17] La obra de Jody Kreiman y Diana Sidtis (2011): *Foundations of voice studies*, Malden, MA: Wiley-Blackwell, es, muy probablemente, el mejor estado de la cuestión que existe hasta el momento sobre el estudio multidisciplinar de la voz, considerada tanto desde el punto de vista de su producción como de su percepción; precisamente por ello, con frecuencia se aludirá a ella a lo largo de este capítulo.

porque puede referir estrictamente a las características fisiológicas que determinan su producción y a sus consecuencias acústicas, o también, al efecto perceptivo que tales características provocan en el oyente. El término *cualidad* se utiliza con mayor frecuencia cuando se quiere poner el foco en la esfera perceptiva y aludir al modo en que los oyentes perciben ―evalúan, valoran, reconocen, discriminan, etc.― las voces. La perspectiva más amplia que observa qué ocurre en la producción y en la percepción es la que se adopta en esta tesis porque como sostienen Kreiman y Sidtis: "voice quality may best be thought of as an interaction between a listener and a signal" (2013 [2011], p. 9).

Por último, y al igual que ocurre con el término *voz*, la expresión *cualidad de voz* puede ser considerada en términos más o menos vastos. En ocasiones refiere a la impresión perceptiva que provocan en los oyentes las características de la voz derivadas exclusivamente del comportamiento laríngeo. Otras veces alude a los efectos perceptivos que en los oyentes despiertan las características del sonido producido en la laringe sumadas a aquellas derivadas de la configuración de las cavidades supraglóticas y aspectos prosódicos del habla. Esta última visión es lógicamente más amplia que la primera pues considera la respuesta de los oyentes frente al sonido final del habla (para más detalles consúltese Kreiman y Sidtis, 2013 [2011], pp. 5-10).

2.2. Cualidad de voz y características personales del locutor

> No sé por qué motivo, pero no pude colgarle el teléfono. En la voz de aquella mujer había algo que me llamaba la atención […]
> ―¿Es verdad que me conoces?― le pregunté.
> ―Nos hemos visto cientos de veces.
> ―¿Cuándo? ¿Dónde? […]
> ―Por lo menos dame alguna prueba. Dame una prueba de que me conoces.
> ―¿Cómo qué?
> ―Como mi edad, por ejemplo.
> ―Treinta años –respondió ella al instante―. Treinta años y dos meses. ¿Te basta con esto?
> Enmudecí. Era evidente que me conocía. Pero por mucho que rebuscara en mi memoria, no lograba recordar su voz.

> —Ahora te toca a ti adivinar cosas sobre mí –dijo en tono provocativo—. **Por la voz, imagínate cómo soy. Cuántos años tengo, dónde estoy, cuál es mi aspecto, ese tipo de cosas**. […]

[Texto extraído del Capítulo 1 de la novela *Crónica del pájaro que da cuerda al mundo* de Haruki Murakami, 2001 (edición en *eBook*). Resaltados de la tesis].

Como se presupone del diálogo entre los personajes, todos los oyentes, en mayor o menor medida, pueden deducir información de una persona prestando atención a las características de su voz porque ella transmite indicios sobre los propios hablantes. Como se verá enseguida, estos indicios o atributos se agrupan en *biológicos* y *físicos* (edad, sexo, estado de salud, apariencia física), *psicológicos* (actitud, estado emocional, personalidad) y *sociales* (educación, estatus social, origen), para una explicación más detallada de esta agrupación puede consultarse Gil Fernández y San Segundo (2014a, pp. 3-4). Todos estos aspectos han sido examinados en diversos estudios, como se comentará brevemente.

2.2.1. Información de carácter biológico, psicológico y social

Las voces de las personas vienen determinadas en cierta medida por su configuración física y anatómica. En Kreiman y Sidtis, (2013 [2011], pp. 110-155) se resumen los estudios realizados hasta la fecha de su publicación con respecto a la relación entre la voz y el aspecto físico de los hablantes (tamaño o corpulencia, aspecto o apariencia física, y, por supuesto, el sexo).

Si la cualidad de voz puede transferir indicios sobre los individuos, entonces es razonable pensar que pueda contribuir, en algunos casos, a diferenciar hablantes. Sin embargo, las impresiones que los oyentes extraen a partir de las voces no siempre son acertadas, a menudo se establecen asociaciones incorrectas, se cometen fallos inesperados, incluso entre aquellas asociaciones que a *priori* podrían resultar más evidentes. Como aparece recogido en Watt (2010, p. 79), el tono de la voz puede incluso inducir a error respecto de cuál podría ser el sexo de un individuo, puesto que los rangos de la frecuencia fundamental en los que se mueven las voces masculinas y femeninas a veces se solapan.

A esto habría que añadir que las razones por las que una voz suena de una determinada manera son variadas. Por ejemplo, una voz aguda –correlato perceptivo de valores altos de f_0– podría ser la respuesta involuntaria a un estado anímico de euforia (por ejemplo, Frick 1985; Scherer y Oshinsky 1977; entre otros). Un *falsetto* podría corresponderse con el intento de un hablante por transmitir significados estilísticos, sociales o identitarios (Podesva, 2007), o también, ser el resultado de un disimulo intencionado, como se verá más adelante con mayor profundidad (véase §2.4 y §2.4.1).

Además de variar las causas, el grado de permanencia de una determinada cualidad vocal en el habla de una persona puede ser también variable. Como explica Fernández Planas (2015, p. 59), por ejemplo, una voz disfónica podría estar provocada por una causa neurológica permanente, como el Párkinson; por causas orgánicas, más o menos transitorias, como podrían ser pólipos, edemas o nódulos; o bien por causas funcionales, como malos hábitos en la fonación, esfuerzos puntuales, etc. De modo que muchos de estos atributos personales, incluso los biológicos, que podrían tener una manifestación vocal más o menos característica y que escaparían al control del hablante, son variables y, en consecuencia, susceptibles de deducciones incorrectas o inexactas.

Las asociaciones que puedan construirse entre cualidad de voz e información psicológica del hablante resultan todavía menos sólidas, aunque es posible fijar algunas correspondencias. Se ha visto, por ejemplo, que los estados anímicos y emocionales se manifiestan de forma bastante transparente en la voz; por ejemplo, un tono de voz grave (correlato perceptivo de valores bajos de f_0) suele corresponderse con la tristeza y el aburrimiento; mientras que el miedo, el enfado y la alegría se identifican con tonos agudos relacionados con valores altos de f_0 (Fonagy, 1978; Frick, 1985; Scherer, 1974; Scherer y Oshinsky, 1977, etc.). Incluso hay estudios que han postulado que las emociones y los estados de ánimo tienen un correlato fisiológico en el hablante y, por consiguiente, en la fonación. En Kreiman y Sidtis (2013 [2011], pp. 324-328), se explica, cómo una determinada emoción o estado anímico genera por lo común varias respuestas fisiológicas, lo que provoca a su vez un resultado acústico que luego tendrá una interpretación perceptiva por parte del oyente. Por ejemplo, comentan las autoras, los estados de excitación pueden provocar un aumento de la frecuencia fundamental y de la intensidad. Los estados más relajados, de otra parte, suelen caracterizarse por valores más bajos de la f_0 y menor intensidad (véase también Fernández Trinidad, 2015).

Por último, la voz transmite información social y sociolingüística sobre el locutor porque ciertos rasgos vocales pueden ser característicos de un dialecto determinado, esto es, dan pistas sobre el origen lingüístico de las personas, o son un indicador de su nivel educativo, social o incluso económico (Campbell, 2007). Se ha señalado que algunos dialectos se caracterizan por tipos de voces específicos, por ejemplo, que la *creaky voice* en el inglés de Edimburgo es distintiva de hablantes de una posición social medio—alta al tiempo que la *harsh voice* es característica de las clases trabajadoras (Esling, 1978). También se ha observado que en Estados Unidos, por ejemplo, la *creaky voice* aparece habitualmente en el habla de mujeres americanas jóvenes (Pennock, 2005; Wolk, Abdelli-Beruh, y Slavin, 2011; Yuasa, 2010, entre otros). La información sociolingüística que puede aportar una determinada cualidad de voz es en buena medida el resultado de un aprendizaje social puesto que se adquiere y, en ocasiones, se imita la forma de hablar propia de un dialecto o estilo de habla en un intento de acomodación lingüística para sentirse más próximo a un determinado grupo social, con el propósito de adaptarse a identificarse con él (Abercrombie, 1967; Giles, 1984; Laver, 1968).

De lo dicho se deducen varias cuestiones principales que conviene tener presente y que fueron señaladas en los trabajos de Gil Fernández (2012) y Gil Fernández y San Segundo (2014a). La evaluación de la cualidad de voz permite derivar —aunque no siempre de forma correcta— una gran cantidad de datos sobre los hablantes, y de la voz se sirven los oyentes —expertos y profanos— para caracterizar, evaluar, diferenciar y reconocer locutores. Sin embargo, la cualidad de voz es mutable de acuerdo con diversos factores y circunstancias que los especialistas que abordan su estudio deben tratar de aclarar. Algunas características vocales pueden ser propias de un grupo social o lingüístico y otras podrían responder a un estado emocional puntual o a una situación comunicativa concreta. Por último, la voz se modifica por distintas causas, algunas de las cuales escapan al control de los hablantes, mientras que, en otros casos, son los hablantes los que buscan esas modificaciones de forma consciente. Como se ha explicado, la voz de una persona es diferente según su sexo, edad o su estado de salud, anímico o emocional y, en estos casos, las distintas cualidades de voz son más bien un síntoma o un resultado de estos condicionantes biológicos, físicos y psicológicos. De otra parte, es asimismo posible ejercer cierto control sobre la producción de la voz modificándola según propósitos que pueden ser muy diferentes: caracterizar un personaje, transmitir ciertos significados y emociones fingidas, etc., y también, claro

está, las personas pueden querer, en determinadas situaciones, ocultar su identidad cambiando sus características vocales distintivas.

2.3. Cualidad de voz y reconocimiento de locutores: posibilidad de reconocer voces y personas a través de sus voces

Las personas tienen voces distintas y este es un hecho fácilmente constatable por cualquiera. Como apunta Watt (2010), aunque en teoría puede concebirse la posibilidad de que dos personas pudieran tener la misma voz, la investigación en este campo viene demostrando que las probabilidades de que esto ocurra son prácticamente inexistentes. Incluso los estudios que comparan las voces de familiares han probado que estos difieren entre sí en su producción vocal a pesar de compartir características y antecedentes biológicos y sociales. Tales divergencias, opina Watt, muestran que las personas pueden mantener una singularidad vocal y que esto implica, por tanto, la posibilidad de asociar una voz con una persona, conocida o no (Watt, 2010, pp. 76-85).

La asociación entre un tipo de voz y una persona se lleva a cabo por seres humanos ―como se ha dicho, con o sin conocimientos específicos, con o sin entrenamiento auditivo―y por diversos programas informáticos especializados en el reconocimiento automático de locutores (véase el Capítulo 7 de esta tesis, sobre las implicaciones prácticas en el ámbito de la fonética forense, y parte del Capítulo 6). Todo tipo de reconocimiento (humano o automático) tiene sus limitaciones y todos pueden cometer fallos; sin embargo, parece posible determinar formas individuales de la voz y del habla por las que una persona pudiera llegar a ser reconocida; es decir, parece posible llegar a reconocer patrones vocales individualizadores. No obstante, como también apunta Watt (2010), y como se tendrá oportunidad de observar a lo largo de los apartados de este capítulo, los diversos estudios experimentales sobre la variabilidad intralocutor, sobre las condicionantes que hacen que la voz de un mismo individuo suene diferente, y también los múltiples estudios que muestran que las voces no se perciben ni se recuerdan del mismo modo exigen máxima prudencia (Watt, 2010, p. 85). En efecto, los resultados de los estudios experimentales y las anécdotas de casos reales (véase Capítulo 7) obligan a considerar con muchísimo cuidado –y algunos piensan que con escepticismo– la posibilidad de discriminar y reconocer voces y, más aún, la posibilidad de reconocer a una persona a partir de su voz o de su habla.

2.3.1. Procesos cognitivos de reconocimiento basados en patrones y basados en componentes

Los modelos de procesamiento cognitivo que intervienen en el reconocimiento de voces se agrupan tradicionalmente en dos clases: el modelo atomístico o componencial, que supone un procesamiento por rasgos, y el modelo gestáltico u holístico, que supone la identificación de un patrón global. De acuerdo con el primer enfoque, el oyente descompone las voces en elementos que se van sumando para evaluar las semejanzas y diferencias existentes entre ellas. Por tanto, este tipo de procesamiento presupone que el oyente percibe y clasifica la presencia y ausencia de una serie de rasgos vocales y va evaluando su distintividad. El enfoque holístico, en cambio, entraña un reconocimiento de patrones globales, percibidos en su conjunto (Kreiman y Sidtis, 2013 [2011], p. 158).

Como aparece explicado en los trabajos de Kreiman (1997) y Kreiman y Sidtis, pp. 178-181), en el marco del modelo componencial, se piensa que los oyentes se guían por un conjunto de rasgos preestablecidos y en ellos basan sus juicios para distinguir o identificar las voces. Los estudios que parten de esta premisa pretenden descubrir cuáles son los parámetros vocales que resultan más relevantes para el reconocimiento de la cualidad individual de la voz y, para ello, suelen manipularlos para luego observar en qué medida se ha visto afectado el reconocimiento luego de aplicar tal o cual cambio. Sin embargo, como comentan las autoras, se ha comprobado que no todos los oyentes responden de la misma forma a las mismas manipulaciones, es decir, se ha visto que la alteración de un mismo rasgo no afecta de igual manera el reconocimiento de locutores; algunas voces son fácilmente reconocibles y otras no, aun si se ha modificado el mismo parámetro vocal (consúltense, por ejemplo, los estudios de Van Lancker, Kreiman y Emmorey, 1985; Van Lancker, Kreiman y Wickens, 1985, citados por Kreiman y Sidtis, 2013 [2011], p. 182).

Esto implicaría aceptar que son múltiples los factores que intervienen en el reconocimiento de voces y que unas veces algunos parámetros son determinantes y otras veces esos mismos parámetros no lo son, como se ha señalado en varios estudios. Las claves perceptivas sobre la identidad de las personas a partir de sus voces no solo varían de una voz a otra sino que también dependen del contexto en el cual esas claves aparecen. La importancia de cualquier parámetro acústico o clave auditiva podría variar considerablemente, y de hecho lo hace, de un oyente a otro; las señales individuales de

la voz son cambiantes en función de la voz que se busca reconocer, del oyente y del contexto. Descubrimientos como estos, opinan Kreiman y Sidtis, no parecen encajar muy bien en el modelo reduccionista de reconocimiento de hablantes basado en un conjunto finito de rasgos vocales distintivos o, al menos, lo ponen en cuestionamiento, por no ser capaz de explicar el desigual comportamiento de los oyentes ante los mismos estímulos (pp. 181-183).

El enfoque holístico, por su parte, ofrece una explicación diferente del proceso de reconocimiento de voces postulando que se reconocen como patrones gestálticos. Desde esta perspectiva, no hay una clave única que por sí sola sea capaz de explicar el reconocimiento de una cualidad de voz individual (véase, por ejemplo, Kuwabara y Sagisak, 1995, trabajo citado por Kreiman y Sidtis, p. 181). Entendiéndose así el mecanismo de reconocimiento de voces, no existiría ninguna ventaja en aislar un conjunto de rasgos que pudiera subyacer a la identificación de una cualidad individual de la voz, ya que el peso que tuviera cada uno de estos parámetros no estará dado por sí mismo sino que vendrá determinado por su interacción con el patrón vocal completo en el que dichos rasgos se inscriben (Kreiman y Sidtis (2013 [2011], pp. 181-182 y Kreiman, 1997, p. 98).

Al lado de estos dos modelos básicos de procesamiento, en Kreiman y Sidtis (2013 [2011], pp. 187-191) se propone una especie de modelo mixto en el que se entrelazan ambos tipos de procesos perceptivos, el asociado a patrones globales con el asociado a rasgos individuales de la voz. De acuerdo con su propuesta, en el reconocimiento de voces familiares, así como en la discriminación de voces no familiares, intervendrían ambas estrategias cognitivas, pero el peso relativo de cada una de ellas sería diferente. En la discriminación de voces no familiares, que es el caso que interesa en este trabajo, el procesamiento perceptivo partiría de las características que emerjan efectivamente del propio estímulo, es decir, su dirección sería de “abajo a arriba”. Por ello, en la discriminación de voces desconocidas el examen de varios rasgos específicos tiene mayor peso y utilidad, al tiempo que en la identificación de voces familiares se utiliza con más éxito el reconocimiento del patrón complejo.

El grado de intervención entre un proceso u otro parece evidenciarse en la especialización cerebral hemisférica:

> Cerebral sides-of-processing for voice perception varies with familiarity, as pattern recognition mechanisms native to the right cerebral hemisphere are more successfully engaged for familiar, distinct patterns, while the feature-analytic processes inherent in discriminating unfamiliar voices are better modulated by left cerebral areas" (Kreiman y Sidtis, 2013 [2011], p.189).

2.3.2. Factores que pueden influir en el reconocimiento perceptivo de hablantes

Se asume que son varios los factores que pueden incidir en la producción y en la percepción de una voz y por tanto en la precisión de los oyentes para discriminar o reconocer locutores a partir de la escucha de sus voces. Lejos todavía de conocerse con claridad, continúan estudiándose y debatiéndose cuáles podrían ser los factores que hipotéticamente impactan o interfieren en la exactitud y fiabilidad de los reconocimientos auditivos de voces, y en qué medida lo hacen. Tal y como aparece recogido en Gil Fernández (2013a y 2013b), las investigaciones realizadas con estos propósitos suelen adoptar tres perspectivas diferentes para analizar tales factores: la del oyente, la del locutor y la que gira en torno a las características del estímulo, su transmisión y el tipo de tarea perceptiva (puede verse una completísima revisión de estos y otros factores en Kreiman y Sidtis, 2013 [2011], pp. 237-259).

Como sintetiza Gil Fernández en (2013a y b), desde la primera perspectiva, que considera las características de la persona que escucha, la mayor parte de los trabajos ha intentado esclarecer la posible influencia de factores como el sexo del oyente, su edad, su capacidad de concentración, su agudeza auditiva, su formación musical, preparación científica o profesional en la escucha y análisis de voces, el conocimiento de la lengua que habla el locutor, etc.

Los factores que se centran en el locutor se vinculan principalmente con las posibles causas de la variación individual en la producción de la voz, tales como su estado físico, emocional o anímico, y el disimulo intencionado de su habla.

Estos puntos de vista, el de la persona que habla y el de la que escucha, frecuentemente se entrecruzan (¿se reconocen mejor las voces masculinas que femeninas?, ¿es mejor el reconocimiento cuando locutor y oyente son del mismo sexo?) o incluso son interdependientes (un locutor tiene un acento extranjero o habla una lengua extranjera solo desde la perspectiva de quien lo escucha).

A pesar de los esfuerzos por clarificar los efectos que pueden tener las características de locutores y oyentes en la capacidad de estos últimos para reconocer a los primeros a través de sus voces, no se aprecian, todavía, demasiados acuerdos en los hallazgos alcanzados. Como también analiza Gil Fernández (2013a y 2013b), quizás, los resultados más consensuados son los que relacionan la precisión en el reconocimiento y las características lingüísticas y la duración de las muestras de habla. Casi todos los estudios que se han realizado en este sentido concluyen que los reconocimientos auditivos son mejores cuanto más extensas y variadas son las muestras de voces que se escuchan.

En esta tesis, circunscrita, como se ha dicho, al estudio de la discriminación perceptiva de hablantes con voces no familiares, se estudiará la posible influencia de dos factores: el disimulo voluntario de la voz mediante el registro de *falsetto* y el efecto que puede tener el que hablante y oyente compartan la misma lengua. A continuación, se revisan los principales estudios previos encontrados en la bibliografía que tratan los dos factores que se han de examinar.

2.4. Modificación voluntaria de la voz. El disimulo y sus formas

Es posible distorsionar la voz de varias formas y con diferentes propósitos, como se detallará enseguida. Los procedimientos que se usan, los rasgos vocales que se ven afectados y sus efectos en el reconocimiento de locutores resultan de especial interés en el ámbito de la investigación en fonética judicial principalmente, porque se ha observado que el disimulo empeora el reconocimiento de locutores, como se verá más adelante. Precisamente por ello este apartado se centrará en resumir los hallazgos provenientes de dicho campo de estudio aplicado.

Rodman (1998, pp. 8-9) definió el disimulo vocal como "any alteration, distortion or deviation from the normal voice, irrespective of the cause". Tal y como él mismo indica, a pesar de que esta definición resulta algo imprecisa –empezando por no esclarecer qué debería entenderse por "normal"–, permite en cambio generar una razonable clasificación de los diferentes tipos de disimulo vocal. Rodman separa el disimulo en dos dimensiones independientes pero que se entrecruzan: de una parte, disimulo *deliberado* frente a *no deliberado* y, de otra, disimulo *electrónico* frente a *no electrónico*, como se observa en la Figura 2.1, tomada de Rodman (1998, p. 9). Un

disimulo deliberado y electrónico sería, por ejemplo, el uso de distintos dispositivos electrónicos para distorsionar la voz. Esta práctica es relativamente frecuente en radio y televisión cuando se busca proteger la identidad de la persona entrevistada. El disimulo electrónico pero no deliberado englobaría las distorsiones introducidas por el canal de transmisión de la voz, especialmente aquellas provocadas por el canal telefónico. Los no deliberados y no electrónicos serían las alteraciones derivadas de algún estado físico o emocional involuntario del locutor. Los efectos de intoxicaciones por drogas y alcohol, enfermedades que afectan a la voz, determinados estados emocionales que provocan igualmente alteraciones vocales, etc., serían ejemplos de modificación no electrónica no deliberada. Finalmente, el disimulo deliberado y no electrónico comprende técnicas como la utilización de distintos registros vocales o algunos tipos de fonación (como el *falsetto*, el *creak,* la voz susurrada, etc.), la simulación de acentos dialectales, las modificaciones de la velocidad de habla, entre otros muchas (Rodman, 1998: pp. 8-10).

Tipos de disimulo	**Deliberado**	**No deliberado**
Electrónico	alteración electrónica, dispositivos	distorsiones del canal telefónico…
No electrónico	registros y tipos de fonación, acentos…	intoxicación por alcohol, drogas, enfermedades, estados emocionales

Figura 2.1. Clasificación general de tipos de disimulo tomada de Rodman (1998, p. 9).

El disimulo electrónico deliberado es bastante poco frecuente, representa apenas el 10 % de los casos de disimulo de la voz (Masthoff, 1996). Las transformaciones de la voz provocadas involuntariamente por mecanismos electrónicos (disimulo electrónico no deliberado), máxime aquellas derivadas de la transmisión telefónica fija y móvil, han sido bastante estudiadas, y sus principales efectos en la transmisión de los sonidos del habla y en la identificación de hablantes parecen conocerse suficientemente (consúltese, por ejemplo, los trabajos de Byrne y Foulkes, 2004; Künzel, 2001, y el de Jiménez Gómez (2011), para el español).

Según relata Rodman (1998, p. 9), entre los no electrónicos, el disimulo involuntario de la voz provocado por estados anímicos y emocionales o enfermedades está relativamente poco investigado y sus resultados interesan sobre todo al ámbito de la salud. Los efectos provocados por el alcohol y las drogas están siendo objeto de interés científico en el ámbito de la fonética forense y cada vez son más numerosos los

trabajos que se realizan en este campo evaluando su efecto en la prosodia (Braun y Künzel, 2003), específicamente en los parámetros temporales del habla (Barfüßer y Schiel, 2010, González Ceria, 2016) o en la frecuencia fundamental (Baumeister, y Schiel, 2015), por referir algunos de los estudios más recientes.

El disimulo no electrónico y voluntario incluye varias técnicas distintas y ha concentrado la atención de los investigadores preocupados por aclarar su perjuicio para el reconocimiento humano y automático de locutores. Como han explicado Kreiman y Sidtis (2013 [2011], p. 243), desde la perspectiva de locutor, el hablante que disimule su voz o su habla recurriendo a técnicas como estas buscará alterar de forma más o menos consciente sus propias características vocales intentando asemejar su forma de hablar a la de otra persona o intentando producir un habla no marcada, neutra, en definitiva, no distintiva. Ambos mecanismos persiguen el mismo objetivo, el de ocultar su patrón vocal característico.

En la Figura 2.2, también adaptada de Rodman (1998), se recogen algunos ejemplos de disimulo deliberado no electrónico.

Alteraciones en la fonación	**Alteraciones en el plano segmental**	**Alteraciones en la prosodia**	**Alteraciones motivadas por los cambios en el tracto vocal**
Mediante la elevación del tono	sonidos dialectales	relativas a la entonación	pinzamiento de nariz
Mediante la disminución del tono	acentos extranjeros	relativas al *tempo*	objetos en la boca
Mediante la elevación o descenso de la laringe	defectos del habla	relativas a la acentuación	objetos que cubran la boca
Mediante cambios de registro vocal		relativas al ritmo	rango de movimiento mandibular

Figura 2.2. Tipos de alteraciones en el disimulo deliberado no electrónico. Esquema tomado y adaptado de Rodman (1998, pp. 9-10).

Esta es una muestra de los distintos procedimientos de transformación intencionada de la voz que podrían encontrarse en casos reales de disimulo con fines delictivos, pero existen más, por supuesto. Como apunta Rodman (p. 9), el elenco no es

completo y ni siquiera tal vez "completable", considerando el ingenio y, podría añadirse, la buena plasticidad vocal que pueden llegar a demostrar los seres humanos.

El disimulo de la de fonación se consigue, sobre todo, alterando la configuración y funcionamiento de los pliegues vocales, lo que, en ocasiones pero no siempre, revierte en distintos registros vocales o tipos de fonación. Las alteraciones segmentales, modificaciones en la pronunciación de los sonidos del habla, estarían dirigidas o bien a modificar las características fónicas segmentales de una lengua o variedad dialectal y acercarlas a la de otra, o bien a emular un habla patológica. El disimulo de la prosodia implica la modificación de los patrones entonativos y acentuales, variar el tono medio o el rango tonal, el *tempo*, etc. Por último, la deformación del tracto vocal busca la alteración forzada de las cavidades supraglóticas, principalmente. Puede efectuarse gracias a la modificación de las cavidades de resonancia o dificultando la "normal" transmisión de la señal vocal mediante el uso de diversos objetos. Por lo común, este tipo de alteración se consigue con técnicas bastante rudimentarias como la colocación de un bolígrafo en la boca, el pinzamiento de la nariz, el uso de pañuelos, pasamontañas, o capuchas, etc. Entre todos los tipos de disimulo, probablemente sea el último de los mencionados el más simple de realizar y de sostener coherentemente por mucho tiempo (Rodman, 1998, pp. 9-10).

Esta agrupación en cuatro tipos principales de todas las variedades de disimulo deliberado y no electrónico no es la única posible y presenta cierto grado de arbitrariedad, como el propio Rodman lo admite. Por ejemplo, para imitar un dialecto o una lengua extranjera, una persona seguramente intentará modificar tanto los aspectos segmentales como los suprasegmentales; entre estos últimos, seguramente podrán alterarse los prosódicos pero incluso también los registros o los tipos de fonación, si se considerara que alguno de ellos forma parte de un hábito fonatorio propio de la lengua o dialecto que se busca emular.

En opinión de Masthoff (1996), los delincuentes no acostumbran, por lo general, a combinar estas tácticas y usan cada vez solo uno de los recursos mencionados. Sin embargo, es difícil pensar que los hablantes tengan demasiado control sobre cuáles son los aspectos que están modificando para obtener voces diferentes. De hecho, cuando se observan los resultados acústicos de los disimulos, se comprueba que un cambio determinado desencadena frecuentemente otros colaterales, por ejemplo, Künzel (2000) o Wagner y Köster (1999) han constatado que la disminución o la elevación de la f_0 pueden acarrear concomitantemente modificaciones en la velocidad de articulación o en

la velocidad del habla (un resultado similar se observa en esta tesis, véase Capítulo 4, §4.2).

2.4.1. Tipos más frecuentes de disimulo vocal

Según se expondrá a continuación, algunos trabajos conducentes a averiguar qué técnicas de disimulo humano serían las preferidas y en qué medida las características propias del hablante podrían determinar esa elección mostraron algunos resultados relevantes para la presente investigación.

Masthoff (1996) realizó un experimento en el que les solicitó a 20 personas germanoparlantes de ambos sexos que disimularan sus voces de forma de encubrir su identidad de la manera más eficaz posible pero sin alterar la inteligibilidad de la frase pronunciada. Se los ubicó en la situación de un chantajista que tenía que dar un mensaje por teléfono. En el 65% de los casos los locutores prefirieron utilizar el susurro, la elevación del tono de la voz o la disminución del tono de voz, en este orden de preferencia; es decir, todos casos de disimulo que involucran cambios en la fonación. En el resto de los casos, los participantes decidieron imitar un acento regional o un acento extranjero o recurrieron al pinzamiento de nariz. Se constató asimismo una diferencia entre sexos con respecto al disimulo consistente en la variación de tono: los hombres recurrieron únicamente a aumentar el tono, mientras que las mujeres optaron exclusivamente por disminuirlo (Masthoff, 1996, p. 166).

En otro experimento, Künzel (2000) les pidió a un grupo de 50 hombres y 50 mujeres que escogieran entre aumentar o disminuir su tono habitual o, si no se sentían capaces de realizar estas transformaciones, disimularan su voz mediante el pinzamiento de nariz. Su estudio estaba centrado en el comportamiento de la frecuencia fundamental. Respecto de la preferencia por un tipo u otro de disimulo, observó que, consecuentemente con los datos reportados en casos forenses reales, los participantes del experimento recurrieron más frecuentemente a los métodos que afectan el tono de la voz. Esta predilección probablemente se deba, expone Künzel, al hecho de que dicha técnica resulta ser tan eficiente para disimular la identidad como para preservar el mensaje lingüístico que se busca transmitir (2000, p. 173). Los resultados mostraron también que, por lo general, los locutores con un tono medio de voz superior al promedio prefirieron modificar sus voces aumentando sus valores de f_0. Los locutores con un tono de voz inferior a la media se inclinaron, en cambio, por disminuirlo. Por

ello, y en términos generales, los hombres prefirieron disimular sus voces bajando la frecuencia fundamental y las mujeres, alzándola. Sin embargo, en cifras absolutas, las mujeres optaron más frecuentemente por disimular sus voces mediante el pinzamiento de la nariz. No se observó ninguna relación, en cambio, entre la f_0 habitual de un individuo y su preferencia por el uso del mecanismo "pinzamiento de nariz" (p. 172).

Al parecer, el hallazgo de que el nivel medio de la f_0 de un locutor podría condicionar la dirección en la que alteraría dicho parámetro para disimular su tono corrobora la experiencia en casos forenses en los que también se pudo observar una relación entre la f_0 natural del locutor y el tipo de disimulo que termina utilizando durante la llamada telefónica incriminatoria (Künzel, 2000, p. 155)[18].

Cuando el disimulo elegido fue el aumento de la f_0, los hombres consiguieron elevar sus voces significativamente más que las mujeres. Al comparar las medias entre ambos sexos observó que los hombres aumentaron su f_0 promedio de 116, 6 a 223,9 Hz (11,3 semitonos, prácticamente una octava); mientras que en las mujeres se constató un incremento de 208,5 a 297,8 Hz (6,2 semitonos, correspondientes a poco más de media octava), como explica Künzel (pp. 160-161). La situación contraria se comprobó cuando el disimulo escogido había sido la disminución de la frecuencia fundamental; en ese caso, las mujeres disminuyeron su f_0 de 208,5 Hz a 169,6 Hz, es decir, un descenso de 3,56 semitonos, y los hombres consiguieron bajar su tono de voz de 116,6 a 100,9 Hz, un promedio de 2,5 semitonos (Künzel, 2000, p. 163). Este comportamiento, podría explicarse, en opinión de Künzel, por la relación entre sexo y f_0 media que le dejaría a los hablantes masculinos más espacio para moverse hacia las frecuencias más altas y, a las mujeres, más espacio para hacerlo en las frecuencias más bajas (p. 168). El aumento o la disminución del tono podían efectuarse sin salir del registro modal o, por el contrario, implicar cambios radicales, de modal a *falsetto* o a *creak* (o *creaky voice*), respectivamente. Nuevamente, se encontró una diferenciación entre sexos respecto de las estrategias fonatorias adoptadas para modificar el tono de la voz. Se observó que, en términos generales, los hombres fueron más proclives a imprimir cambios drásticos en los patrones de vibración de los pliegues vocales y decidieron mucho más a menudo que

[18] Comenta Künzel que el descubrimiento de una preferencia individual por aumentar o disminuir el tono de voz, basada en la f_0 media de cada locutor, con independencia del sexo, no contradice el resultado de Masthoff (1996) de que únicamente los hombres subían el tono y las mujeres lo bajaban. En el experimento de Masthoff los participantes eran libres de elegir cualquier disimulo a condición de encubrir al máximo su identidad y preservar la inteligibilidad del mensaje. Por tanto, interpreta Künzel, en tal situación quizás los sujetos hayan optado por intentar disfrazar su identidad sexual, los hombres buscando un tono de voz más propio de las mujeres y las mujeres adoptando un tono más propio de los hombres (p. 174).

las mujeres usar el *falsetto* cuando el disimulo consistía en elevar el tono de voz (Künzel, 2000, p. 172).

En un experimento posterior (Zhang y Tan, 2008), centrado exclusivamente en voces masculinas, también se observaron diferencias entre los locutores respecto a su capacidad para disfrazar sus voces. Así, tendieron a manifestar una mayor habilidad para distorsionar sus voces elevando la f_0 que disminuyéndola, lo cual se interpretó, al igual que en Künzel (2000), como una consecuencia de que, al tratarse en todos los casos de locutores masculinos, disponían de un margen más amplio para subir que para bajar su tono natural (Zhang y Tan, 2008, pp. 121-122).

Estas preferencias por ciertos disfraces vocales relevada de los experimentos parece constatarse también en delitos reales, principalmente en casos de secuestros y extorsión (véase por ejemplo, Zhang y Tan, 2008). Esta modalidad delictiva involucra muy frecuentemente el disimulo de la voz puesto que los delincuentes saben o sospechan que sus voces están siendo registradas en las llamadas telefónicas. En tales circunstancias, y según se recoge en varios estudios, las técnicas preferidas suelen ser el *falsetto*, la voz susurrada, la simulación de acentos extranjeros y el pinzamiento de nariz[19] (véanse por ejemplo, Figueiredo y de Souza Britto, 1996; Künzel, 1994; Künzel, 2000; Masthoff, 1996; Perrot y Chollet, 2008; Praveena y Krishna, 2015).

2.4.2. Consecuencias del disimulo en el reconocimiento del locutor, especialmente del derivado de la elevación de la frecuencia fundamental

Como es lógico, y ha sido señalado muchas veces, todos los procedimientos de disimulo, aunque en distinto grado, dificultan el reconocimiento de patrones individualizadores de los hablantes que permitan, al oído humano o a los sistemas informáticos, reconocerlos (véase, Hollien, 1990, por ejemplo). El perjuicio que los mecanismos de fonación, en particular, las alteraciones de la f_0 media habitual, provocan en el reconocimiento de locutores suscita un especial interés, máxime si se consideran la importancia de la f_0 en la percepción de la voz[20], el consecuente deterioro

[19] Sobre este tipo de disimulo, cabe mencionar el estudio para el español de Gil Fernández y San Segundo (2014b).

[20] Según se recoge en Kreiman y Sidtis (2013 [2011], pp. 160-162), la capacidad de diferenciar la voz materna de otras voces no familiares podría estar presente en los humanos desde el nacimiento o incluso previamente (véase Spence y Freeman, 1996, que ofrecen una síntesis sobre esta cuestión). Ciertos

en el reconocimiento del hablante y, como se indicaba antes, la prácticamente nula distorsión del contenido de los mensajes del habla, al menos en muchos idiomas. Aun así, los estudios sobre el perjuicio que las alteraciones o cambios en el funcionamiento de la laringe ocasionan en el reconocimiento de hablantes son todavía muy pocos, y los centrados en el aumento de la f_0, tanto en registro modal como en *falsetto*, son incluso menos frecuentes. Como ha sido señalado en varios trabajos (por ejemplo, Künzel, González-Rodríguez y Ortega-García, 2004; Perrot, Preteux, Vasseur y Chollet, 2007, y Rodman, 1998, entre otros), los estudios sobre la detección y el reconocimiento automáticos de la voz disimulada persiguen varios objetivos. De una parte, encontrar indicadores del disimulo, es decir, discriminar si la voz ha sido disimulada o no; por otra, si se constata que efectivamente ha sido disimulada, detectar qué tipo específico de mecanismo de disimulo ha sido utilizado; y, por último, y como objetivo más ambicioso, pretenden llegar a reconstruir la voz original (Rodman, 1998, p. 10). A pesar de la relevancia que tiene el disimulo en fonética judicial, siguen siendo escasos los esfuerzos de investigación dirigidos a aclarar estas cuestiones, como señalan, Perrot, Aversano y Chollet (2007, p. 109); en opinión de estos autores, principalmente a causa de la dificultad de sistematizar y analizar las alteraciones que el disimulo provoca en la cualidad de la voz y en los rasgos que ayudan al reconocimiento.

Perrot, Preteux, Vasseur y Chollet (2007), en un trabajo sobre la detección y el reconocimiento del disimulo de la voz, se ocuparon de analizar cuatro formas de disimulo: elevación y disminución de la frecuencia fundamental, pinzamiento de nariz y colocación de la mano cubriendo la boca. A partir de dos bases de datos, con voces normales y disimuladas, evaluaron el desempeño del sistema automático de reconocimiento de disimulo. Comprobaron que los sistemas basados en algoritmos automáticos fueron capaces de detectar y reconocer correctamente la voz normal (85%) cuando las muestras del test se compararon con un corpus de voz normal; sin embargo, cuando las mismas muestras fueron comparadas con un corpus de voz disimulada, el porcentaje de reconocimiento automático descendió a un 15%. De igual forma, el sistema fue capaz de reconocer correctamente el disimulo con un porcentaje de reconocimiento de 71%, cuando esas muestras se compararon con un corpus previamente entrenado en voz disimulada;

estudios que observan, por ejemplo, los movimiento fetales, los patrones de succión o miden la frecuencia cardíaca así lo indican (Gerhardt y Abrams, 2000, y Hepper, Scott y Shahidullah, 1993, entre otros). En esta capacidad parece que la frecuencia fundamental desempeña un papel decisivo. Por ello, algunos autores sugieren que la f_0 es el indicio acústico que más tempranamente se utiliza en la vida del ser humano para reconocer voces.

cuando se contrastaron con un corpus entrenado en voz normal, el porcentaje de reconocimiento fue de 29% (Perrot *et al*., 2007, p. 2). Otros estudios, como los de Perrot, Aversano y Chollet (2007) o Perrot y Chollet (2008) llegaron a conclusiones semejantes sobre el desempeño de los sistemas automáticos para reconocer el disimulo mediante las mismas cuatro técnicas de transformación de la voz.

El perjuicio provocado por el aumento de la f_0 para los sistemas de reconocimiento automáticos de locutor parece ser muy evidente. El trabajo de Künzel, González-Rodríguez y Ortega-García (2004), por ejemplo, evalúa la exactitud de los métodos automáticos en el reconocimiento de locutores con voces disimuladas mediante el aumento y la disminución de la f_0, y el pinzamiento de nariz, ya que, en el estudio anterior de Künzel (2000)[21] se había observado que estas tres técnicas habituales de disimulo generan importantes modificaciones en el comportamiento vocal del hablante y, por tanto, repercuten negativamente en el reconocimiento del locutor. De su análisis, Künzel *et al.* (2004) concluyen que el grado en el que el reconocimiento automático se deteriora depende en buena medida del hecho de que la base de datos correspondiente a la población de referencia contenga o no voces disimuladas. Cuando la población de referencia es un modelo entrenado en el mismo tipo de disimulo que afecta a la muestra de prueba, el reconocimiento automático proporcionado por el sistema disminuye un 16% en el caso del disimulo consistente en el aumento de la f_0, un 3% en el de la disminución de la f_0, y un 0% en el caso de pinzamiento de nariz. Sin embargo, si el modelo no está entrenado, el reconocimiento de locutor decrece en un 36%, 14% y 70% para las modalidades de aumento de la f_0, disminución de la f_0 y pinzamiento de nariz, respectivamente (Künzel *et al.*, 2004, p. 3).

Resulta interesante notar que los mismos autores observaron a partir de un análisis auditivo que casi todos los hablantes que habían optado por disimular su voz mediante el aumento del tono y no fueron reconocidos correctamente por el sistema automático habían cambiado al registro de *falsetto*. Los autores concluyen que se precisan trabajos de investigación más básica que profundicen en las características fonéticas definitorias de los diferentes tipos de disimulo para luego poder razonar sobre

[21]En el estudio de Künzel (2000), basado en métodos acústicos para evaluar las consecuencias del disimulo, se intentaba dilucidar, entre otros aspectos, si a pesar de la alteración de la f_0, persistía alguna información específica del hablante que pudiera ayudar en su reconocimiento. Los datos obtenidos de su análisis le permitieron concluir que, si el disimulo de la voz consistía en una disminución de la f_0, entonces era quizás posible deducir los valores medios de la frecuencia fundamental habitual del locutor, fuera este hombre o mujer. Sin embargo, si el disimulo consistía en aumentarlos, provocando o no el cambio de registro hacia el *falsetto*, tales inferencias no podrían realizarse (p. 173).

las posibilidades y las limitaciones que los sistemas automáticos de reconocimiento podrían tener en cada caso específico (Künzel *et al*., 2004, p. 4).

Zhang y Tan (2008 pp. 118-122) también estudiaron la influencia del disimulo no electrónico deliberado en el rendimiento de un sistema de reconocimiento automático de locutores. En concreto, analizaron el efecto de 10 tipos de disimulo, entre los que se encontraban la elevación y el descenso de tono, el susurro, el habla rápida y lenta, la colocación de objetos en la boca y el fingimiento de acentos extranjeros, entre otros. El estudio se centró en voces masculinas; compararon cada voz disimulada con las voces normales almacenadas en una base de datos y calcularon el porcentaje de reconocimientos correctos. Los resultados mostraron, en primer lugar, que el sistema de reconocimiento automático de locutores obtiene un buen rendimiento para el reconocimiento de voces normales y que este se ve muy menguado con las voces disimuladas. En comparación con el reconocimiento para las voces normales, las tasas de reconocimiento correctas de las voces disimuladas disminuyen significativamente en todos los casos de disimulo excepto para la modalidad de acento extranjero. En los demás casos, el grado de deterioro en el reconocimiento varió según los distintos enmascaramientos de voz. El susurro provocó el daño más importante, impidiendo el reconocimiento. El segundo disimulo más perjudicial resultó ser el incremento de la f_0, que obtuvo un 10% de reconocimientos correctos.

Resultados como estos muestran que el incremento acusado de la f_0 y el *falsetto* complican notablemente el reconocimiento automático y semiautomático de locutores, pero ¿qué ocurre con el reconocimiento llevado a cabo por humanos?, ¿qué consecuencias tiene este tipo de disimulo en la percepción auditiva humana? Como se verá a continuación, estas cuestiones son incluso más difíciles de esclarecer.

Uno de los primeros estudios acerca del perjuicio causado por diversos factores en el reconocimiento humano lo constituye el trabajo de McGehee (1937), según el cual las alteraciones en la f_0 comprometen seriamente la posibilidad de reconocer perceptivamente a los locutores. En circunstancias tales de disimulo, la tasa de identificaciones correctas se reducía en un 17%, cayendo de 80% a 63% (p. 263). En su trabajo no se aclaraba si dichos cambios implicaban o no cambios de registro vocal (*apud* Künzel, 2000, p. 151).

Como fue antes señalado, lograr reconocer auditivamente a un hablante implica un complejo proceso de toma de decisiones y en él intervienen múltiples factores como

la capacidad del oyente, las características vocales del hablante, el entorno en el que se desarrolla la escucha, las características de la tarea de reconocimiento, etc.

Wagner y Köster (1999) estudiaron la capacidad de oyentes sin ningún entrenamiento en fonética para identificar voces familiares disimuladas en *falsetto*. Participaron del experimento 8 locutores hombres, alemanes, sin acentos regionales prominentes ni hábitos distintivos al hablar, que grabaron un texto leído de aproximadamente 15 segundos de duración. El texto simulaba una llamada telefónica de un chantajista y todos los locutores lo leyeron con sus voces habituales y en *falsetto* (1999, p. 1381). Para recrear las condiciones forenses más habituales, las muestras se registraron mediante transmisión telefónica. Los oyentes conocían a cinco de los locutores (en concreto, eran compañeros de trabajo) y los otros tres les eran desconocidos. El grupo de oyentes que juzgó las voces estaba formado por empleados alemanes, sin conocimientos de fonética, que trabajaban en la misma empresa de los locutores que sirvieron como objetivos en la tarea de identificación. Los oyentes debían indicar los nombres de las personas cuyas voces habían reconocido como familiares y el grado de familiaridad que creían reconocer en las voces (1999, p. 1382). Los autores estimaron la sensibilidad de los oyentes para discriminar entre hablantes familiares y no familiares en el marco de la *Teoría de Detección de Señales* (véase Capítulo 5, §5.1). En la condición de habla no disimulada, la capacidad de discriminación entre locutores familiares y no familiares fue muy elevada (proporción de Aciertos: 0.97; proporción de Falsas Alarmas: 0.29; d' media: 2.434); sin embargo, en la condición de *falsetto* estos valores disminuyeron muy acusadamente (proporción de Aciertos: 0.04; proporción de Falsas Alarmas: 0.01 y d' media: 0.408, Wagner y Köster, 1999, p. 1383). Aunque los autores no calcularon el sesgo de respuesta, estos resultados en las tasas de Falsas Alarmas indicarían que los oyentes adoptaron una estrategia de respuesta más conservadora al evaluar las voces disimuladas probablemente a causa de la dificultad que supuso la tarea. Wagner y Köster (1999) llegan a tres conclusiones importantes. Primeramente, que cuando las voces no han sido disimuladas y la duración de las muestras permite el reconocimiento auditivo (unos 15 segundos, véase también Künzel, 1990, citado en Wagner y Köster, 1999), las personas son capaces de reconocer de forma fiable a locutores familiares, incluso cuando la calidad de la grabación se haya deteriorado a causa del filtrado telefónico (p. 1383). Otra conclusión que derivan del estudio es que en la condición de habla disimulada mediante *falsetto,* esta posibilidad de reconocimiento desaparece, es decir, en esta condición los oyentes no expertos no serían

capaces de reconocer voces familiares. Por último, que los juicios de oyentes sin conocimientos fonéticos (no entrenados) deberían ser juzgados con muchísima cautela, especialmente en casos de disimulo (1999, p. 1383), lo cual confirma, en opinión de los autores, la importancia de la participación de los fonetistas en estas tareas, importancia que había sido ya remarcada por estudios previos (véase, por ejemplo, Köster, Hess, Schiller y Künzel, 1998).

En cuanto al reconocimiento de locutores individuales, los resultados mostraron que no pudieron ser identificados dos locutores y que, entre los que fueron identificados correctamente, uno de ellos se reconoció significativamente mejor que los otros. Estas diferencias en el reconocimiento de locutores, opinan los autores, no podrían, de ningún modo, justificarse por un distinto grado de familiaridad con las voces objetivo sino que responderían más bien a razones acústicas; en concreto, estarían directamente vinculadas con los valores medios de la f_0 en la voz normal y *falsetto* de cada locutor. Las grandes diferencias de tono medio entre la voz habitual y la disimulada en *falsetto* en un mismo hablante se correlacionarán negativamente con las posibilidades de identificar exitosamente al locutor (Wagner y Köster, 1999, p. 1383). En los datos de este estudio también se observó que en el registro de *falsetto* la desviación típica de la frecuencia fundamental aumenta, respecto de la voz modal, y que el disimulo mediante el *falsetto* parece modificar otras características del habla, como son la velocidad de articulación (sílabas por segundo) y los patrones entonativos; los autores no desarrollan estas cuestiones y concluyen en que se necesitan más estudios que aborden estos aspectos (p. 1838).

Estudios análogos se hicieron para el español (Alves, Fernández Trinidad, Gil Fernández, Infante, Pérez Sanz y San Segundo, 2012; Fernández Trinidad, Infante, Lahoz-Bengoechea y Alves, 2013).

Alves *et al*. (2012) llevaron a cabo un experimento perceptivo para comprobar que en español los registros *creak* (o *creaky voice*) y *falsetto* interfieren en el reconocimiento como se había señalado en otros trabajos para otras lenguas. Plantearon tres hipótesis; 1) que los oyentes serían capaces de reconocer las voces disimuladas por encima del nivel del azar, puesto que habría algo en la fuente glotal que permaneciera no obstante el disimulo; 2) que el *falsetto* sería el tipo de fonación que más dificultaría el reconocimiento de las voces, teniendo en cuenta los resultados de estudios previos (por ejemplo, Hirson y Duckworth, 1993; Moosmüller, 2001 y Wagner y Köster, 1999) y, por último, 3) que no existirían diferencias significativas entre el desempeño de los

jueces entrenados (fonetistas) y los jueces profanos /legos en la materia para reconocer voces disimuladas en la medida en que la percepción de la voz no es analítica sino holística.

En su diseño, Alves *et al.* (2012) plantearon una tarea de discriminación con tripletas X (voz disimulada en *creak/creaky* o *falsetto*) A (voz normal) B (voz normal). Los estímulos sobre los que los jueces tenían que dar su opinión fueron producidos por 6 mujeres jóvenes, hablantes de español centro-peninsular, que durante la grabación supieron mantener de forma continuada y homogénea los distintos registros. Los jueces que participaron de la prueba fueron todos hispanohablantes, un grupo de expertos fonetistas y otro compuesto por personas sin conocimientos fonéticos.

Los resultados comprobaron la primera y la tercera hipótesis; los oyentes fueron capaces de reconocer las voces disimuladas por encima del nivel del azar a pesar del disimulo (0.60), y no hubo diferencias significativas en ninguno de los dos registros al contrastar el desempeño de los fonetistas y de los no fonetistas. Sin embargo y contrariamente a lo que se esperaba, el número de aciertos fue significativamente mayor con las voces disimuladas con *falsetto* (0.62) que con *creak/y* (0.59). Es decir, el *creak/y* fue el tipo de fonación que más dificultó el reconocimiento de locutores. El reconocimiento fue más sencillo cuando el registro de fonación utilizado para el disimulo fue el *falsetto*. Frente a este resultado, los autores avanzaron algunas posibles explicaciones que serían puestas a prueba en un segundo test perceptivo (Fernández Trinidad *et al*., 2013). Señalaron que los estudios previos de Hirson y Duckworth (1993), Moosmüller (2001) y Wagner y Köster (1999) habían utilizado voces masculinas y que quizás esto podría explicar la discordancia entre esos resultados y los obtenidos por Alves *et al.* (2012). Así pues, el *falsetto* para los hombres y el *creak/y* para las mujeres, introducen cambios más radicales con respecto a las voces normales y, además, apuntaron el hecho de que el efecto perceptivo final de las voces no concordaría mucho con lo que podrían esperar los oyentes. El *creak/y*, por darse en tonos muy graves, quedaría bastante lejos del prototipo de cómo debería sonar una voz de mujer, y, a su vez, el *falsetto* se apartaría más del prototipo de la voz de los hombres. Otros motivos por los que quizás el *creak/y* haya terminado por ser un mejor método de enmascaramiento de la voz, propusieron los autores, podría ser que es común que en ese mecanismo de fonación la frecuencia fundamental no esté disponible como clave acústica para los oyentes. Con respecto al desempeño de los dos grupos de jueces, los autores plantearon que la capacidad para reconocer las voces quizás no dependa tanto

del conocimiento fonético de los jueces, sino que más bien podría ser el entrenamiento auditivo musical o de otro tipo lo que podría, en todo caso, tener un efecto.

En Fernández Trinidad *et al.* (2013) se replica el mismo estudio perceptivo, pero con voces masculinas, 6 jóvenes hablantes de español centro-peninsular. Como jueces, participaron oyentes con y sin entrenamiento musical, todos hablantes de español como L1. Los resultados mostraron, por una parte, que el disimulo laríngeo de la voz afecta el reconocimiento, pero que, sin embargo, los oyentes nuevamente habían sido capaces de reconocer las voces disimuladas por encima del nivel del azar (0.66 para el *falsetto* y 0.67 para el *creak/y*)[22]. No se encontraron diferencias significativas entre los dos tipos de fonación –a pesar de que, desde el punto de vista acústico, las diferencias entre ellos eran marcadas y estadísticamente relevantes– ni en el desempeño de jueces entrenados y no entrenados musicalmente.

Al comparar los resultados de ambos experimentos perceptivos los autores comentan que, quizás, las voces masculinas podrían ser más fáciles de reconocer, independientemente del tipo de disimulo, apoyándose en algunos estudios previos que habían sugerido que la voz masculina es la voz biológicamente marcada[23].

Un experimento sobre el italiano (Fernández Trinidad y Rojo, 2018), que sirvió como estudio preliminar para esta tesis (véase la Introducción), examinaba en qué medida el *falsetto* comprometía la posibilidad de reconocer hablantes masculinos, hasta

[22] Cabría señalar dos aspectos importantes para calibrar mejor la interpretación de los resultados alcanzados por Alves *et al.* (2012) y Fernández Trinidad *et al.* (2013) en cuanto a los porcentajes de reconocimiento de locutor. Por un lado, hay que tener en cuenta que los porcentajes de aciertos reportados se calcularon en función del umbral del azar (0.5). Los análisis llevados a cabo por estos dos últimos estudios no pudieron aplicar un paradigma de la *Teoría de Detección de Señales* a causa de su diseño experimental, que solamente presentó ensayos en donde la señal (la voz objetivo) estaba presente. Nótese que en una tripleta XAB, cuando el oyente sabe que el objetivo está presente bien en A, bien en B, el azar permite *a priori* un 50% de aciertos, que es mucho. Además, si los jueces responden que el objetivo está en A y resulta que se encuentra en B, no solo estarían cometiendo una Falsa Alarma sino que a la vez, indisolublemente, estarían incurriendo en una Omisión y en el paradigma de la *Teoría de Detección de Señales* las Falsas Alarmas y las Omisiones deben ser independientes. Por último, y los propios autores remarcan este aspecto, ambas pruebas perceptivas tuvieron una duración bastante extensa (~30 minutos sin pausas) y esto con toda probabilidad pudo influir negativamente en el desempeño de los jueces que quizás no consiguieron mantener la concentración a lo largo de todo el test. Esta conjetura es más que plausible aunque no se comprobó si la competencia disminuía efectivamente conforme progresaba el experimento.

[23] Según exponen Kreiman y Sidtis (2013 [2011]), algunos estudios (citan por ejemplo el de Owren, Berkowitz y Bachorowski, 2007) parecen incidir en que, en general, las voces masculinas son más fáciles de reconocer que las femeninas porque quizás lo "masculino" en la voz esté de algún modo "marcado" biológicamente gracias a un cierto nivel de testosterona. Esto podría facilitar el reconocimiento de las voces masculinas puesto que la presencia de un tono más bajo indicaría casi con seguridad la presencia de una voz masculina adulta. En cambio, los valores "no marcados", en este caso, un tono agudo, resultaría más ambiguo porque podría pertenecer a la voz de una mujer, a la voz de un niño, o a la voz de adolescente (p. 133).

qué punto la discriminación de locutores se veía afectada por este cambio drástico en la fonación. Los estímulos, los mismos que sirvieron para esta tesis, consistían en pseudofrases breves en italiano en las que se había variado únicamente el registro de fonación (de modal a *falsetto*) y fueron pronunciadas por 6 locutores napolitanos jóvenes. Se llevó a cabo una prueba perceptiva de discriminación AX en la que se combinaron de forma aleatoria 60 pares de voces, de los cuales la mitad estaban formados por estímulos en voz modal (A) y en *falsetto* (B) del mismo hablante, y la otra mitad, por estímulos pertenecientes a distintos locutores (véase Capítulo 3). De esta forma, se reducía sensiblemente la duración del experimento y las respuestas obtenidas podían analizarse dentro del paradigma de la *Teoría de Detección de Señales*. En la prueba participaron un total de 57 oyentes hispanohablantes jóvenes sin conocimientos de italiano y la duración promedio del test perceptivo fue de 15 minutos. Los resultados mostraron niveles de discriminación por encima del nivel del azar aunque bajos (d'=0.82), lo cual demostró que, frente a estímulos breves producidos por hablantes con características muy homogéneas en cuanto al sexo, edad y acento, el *falsetto* no pareció comprometer demasiado el reconocimiento de la cualidad de voz individual. En sus resultados, los autores no se detuvieron en explicar la proporción de Falsas Alarmas y Aciertos ni calcularon el criterio de respuestas de los jueces. Tampoco examinaron un posible efecto del idioma puesto que no sometieron a los oyentes italianos a la misma prueba perceptiva.

2.4.3. Consecuencias o efectos concomitantes en el habla provocados por la elevación voluntaria del tono de voz

Otro aspecto poco tratado en la bibliografía sobre el disimulo de la voz mediante la elevación de la frecuencia fundamental o la utilización del *falsetto* es el relativo a los efectos acústicos derivados de tales transformaciones. Como señalaban Künzel *et al.* (2004), se requiere más investigación básica que aclare las características definitorias de cada tipo de disimulo vocal para luego poder hacer predicciones específicas sobre el perjuicio que determinada técnica podría provocar en el reconocimiento humano y automático[24] de locutores.

[24] Refiriéndose al disimulo con *falsetto* Künzel *et al.* (2004, p. 4) afirman: "From a phonetic and physiological point of view it goes without saying that such a drastic alteration of the vocal apparatus as a whole will also affect the resonance characteristics of the vocal cavities which in turn are the basis for the

Se ha comprobado que el *falsetto* supone un cambio importante en la cualidad de la voz y que implica modificaciones complejas en el comportamiento vocal del hablante pero ¿en qué consisten realmente estas modificaciones? No es mucho lo que se sabe hasta el momento, pero los hallazgos más importantes son los que se resumen a continuación.

En el análisis de Künzel (2000) se observó que el incremento del tono de la voz en casos de disimulo voluntario suele suponer un aumento de alrededor de una octava en el caso de voces masculinas, mientras que para las femeninas el incremento suele ser de algo más de media octava (Künzel, 2000, pp. 160-161 y §2.4.1). De su estudio y del de Wagner y Köster (1999) también se deduce que los disimulos consistentes en el aumento del tono natural provocan alteraciones en la variabilidad de la f_0 y en el *tempo* del habla, sobre todo si se efectúan cambios de registro vocal. Künzel constató una menor variabilidad en la frecuencia fundamental, una ralentización en la velocidad de articulación y un aumento de las pausas y su duración, respecto del habla no disimulada. En opinión de Künzel, el cambio probablemente no intencionado en la velocidad del habla es una consecuencia de la dificultad –mayor esfuerzo y concentración por parte de los locutores– que supondría un ajuste inusual de los patrones de fonación. También se observó que la intensidad aumentaba al elevarse el tono (en particular si se llegaba al *falsetto*) y que se reducía al descender este (Künzel, 2000, p. 173).

Al estudio perceptivo efectuado por Fernández Trinidad y Rojo (2018) se sumó un estudio sobre la producción de la voz, en el que se analizaba con cierto detalle el comportamiento glotal. Dicho análisis preliminar, que ha servido como punto de partida para la realización de esta tesis, buscaba indagar en los rasgos laríngeos que se veían o no alterados en el cambio de registro hacia el *falsetto* para luego discutir cuál podría ser el peso perceptivo que tales parámetros podrían tener en el reconocimiento de la cualidad individual de la voz. El estudio se llevó a cabo utilizando el programa *BioMet®Phon*. Como se explicó anteriormente en §1.5.2 y se insistirá más adelante en el Capítulo 3 (§3.3.1), este programa aísla el contenido biométrico de la señal glótica restando la función de transferencia del tracto vocal y el efecto de la radiación labial. Calcula la onda glotal a partir de la señal acústica mediante filtrado inverso y a partir de ello estima los valores de 72 parámetros relativos al comportamiento glótico (Gómez Vilda, Rodellar, Nieto *et al*., 2013). Los resultados derivados de este análisis informaron

extraction of the MFCCs used by the automatic system. Knowledge about these interrelations is crucial for the understanding of the possibilities and limitations of automatic FSR systems in these conditions".

que los parámetros de defecto de cierre glótico (grupo F) y los de temblor (grupo G) presentaban una tasa menor de variabilidad intralocutor en el cambio de registro fonatorio. Los rasgos del grupo D (parámetros biomecánicos de la glotis) se revelaron como los más definitorios del *falsetto* junto con la modificación en la frecuencia fundamental, imprescindible para que se dé el cambio de registro. En los resultados de este estudio también se constató, al igual que en los análisis efectuados por Künzel (2000) y Wagner y Köster (1999), un efecto colateral, podría decirse, del cambio de registro en la velocidad de habla, pero, a diferencia de lo señalado en Künzel (2000), no se observó un efecto homogéneo o unidireccional: mientras algunos locutores incrementaban la velocidad, otros la disminuían al pasar de voz modal a *falsetto*. En el Capítulo 4, §4.1.2 y §4.2) de la presente tesis se detallan y analizan estos resultados.

2.5. El efecto de la lengua en la percepción de la individualidad del locutor

Se ha estudiado la posible influencia del idioma en la evaluación auditiva que hacen los oyentes y, más concretamente, en el reconocimiento de locutores. El efecto de la lengua sobre la percepción de la identidad o reconocimiento del locutor a través de la escucha de su voz se ha estudiado teniendo en cuenta cuatro posibles escenarios, como resumen Kreiman y Sidtis (2013 [2011], p. 241):

a) Locutor y oyente hablan la misma lengua pero pertenecen a variedades distintas.
b) Locutor y oyente hablan la misma lengua pero uno es hablante nativo y el otro no lo es.
c) El locutor habla en una lengua extranjera para el oyente pero que este puede comprender.
d) El locutor habla en una lengua extranjera para el oyente y que este no comprende. (p. 241).

La premisa desde la que parten todos los estudios es que, si la información lingüística y sociolingüística tienen un peso muy importante en el reconocimiento de voces, el oyente familiarizado con la lengua del locutor correría con cierta ventaja,

puesto que sería capaz de distinguir más claramente entre los rasgos lingüísticos que caracterizan una variedad dialectal o incluso social de aquellos que pudieran ser idiosincráticos del locutor. Así las cosas, explican las autoras, sería lógico esperar mejores desempeños en el reconocimiento cuanto mayor sea el grado de familiaridad del oyente con la lengua del locutor. Como se verá en un momento, a pesar de la aparente divergencia de los resultados derivados de los estudios realizados, la gran mayoría apunta en esa dirección. La Tabla 2.1 recoge y adapta la síntesis de los estudios presentados en Kreiman y Sidtis (2013 [2011], p. 242).

En el repaso de la bibliografía relacionada con este aspecto se comentarán solamente los trabajos que ponen de manifiesto una mayor cercanía metodológica con el diseño experimental planteado en esta tesis para probar la influencia de la familiaridad lingüística en el reconocimiento de voces. En concreto, se prestará especial atención a los trabajos que han contemplado la condición en la que oyentes y locutores comparten la misma lengua y variedad regional y en la que los oyentes desconocen la lengua hablada por los locutores, es decir, los dos extremos de máxima y mínima familiaridad con el idioma, que son los que se examinan en esta tesis. Dichos estudios son los desarrollados por Goldstein, Knight, Bailis, y Conover (1981), Goggin, Thompson, Strube y Simental (1991); Schiller, Köster y Duckworth (1997) y Thompson (1987), y para comentar sus resultados se seguirá el excelente y detallado repaso crítico que realizan Köster y Schiller (1997, pp. 18-28) y Schiller, Köster y Duckworth (1997, pp. 1-17).

El primero de ellos, el trabajo de Goldstein *et al*. (1981), presenta un interés añadido porque establece algunas comparaciones entre los resultados obtenidos en el reconocimiento de voces y los hallados en el reconocimiento de caras, vínculo que también se intentará realizar en esta tesis (véase Capítulo 6, §6.4). Sin perjuicio de ello, se comentarán trabajos posteriores y anteriores a estos, aunque no constituyan el centro de la revisión.

Tabla 2.1. Síntesis de los estudios que examinan el efecto de la lengua en el reconocimiento de locutores presentada por **Kreiman y Sidtis, 2013 [2011], p. 242**. La tabla se reproduce con algunas modificaciones y se añaden otros datos.

Muestra de voz	Misma lengua o mismo dialecto	Dialectos regionales	Acento extranjero	Lengua extranjera familiar	Lengua extranjera no familiar	Referencias del estudio
frase larga	85% correctas	82% correctas	81 % correctas			Goldstein *et al.* (1981) Exp. 1
1 palabra	56% correctas	55% correctas	37% correctas			Goldstein*et al.* (1981) Exp. 2
frase breve			A 0.57; FA 0.18		A 0.58; FA 0.18	Goldstein *et al.* (1981) Exp.3
texto 80 palabras aprox.	A 0.65; FA 0.27		A 0.52; FA 0.27		A 0.38; FA 0.27	Thompson (1987)
párrafos de un texto (aprox.90 palabras)	d'=1.19				d'=0.58	Goggin *et al.* (1991) Exp. 1*
párrafos de un texto (aprox. 90 palabras)	d'=1.47				d'=0.73	Goggin *et al.* (1991) Exp. 2
párrafos de un texto (aprox. 90 palabras)	d'= 1.34		d'= 0.95		d'= 0.72	Goggin *et al.* (1991) Exp.3
párrafos de un texto (aprox. 90 palabras)	d'= 1.22 d'= 1.34		d'= 1.05			Goggin *et al.* (1991) Exp. 3
texto de 1 minuto				d'= 2.56	d'= 1.67	Köster y Schiller (1997)
enunciados sin sentido	A 0.72; FA 0.14 d'=1.63 c=0.249			A 0.85; FA 0.22 d'=1.80 c= -0.132	A 0.79; FA 0.22 d'=1.57 c= -0.017	Schiller, Köster y Duckworth (1997)
2 frases	A 0.88		A 0.13			Doty (1998)
frase breve	62% correctas	26% correctas				Vanags, Carroll y Perfect (2005)
no especificado (al menos 1 frase)	A 0.41; FA 0.44	A 0.34; FA 0.65				Kerstholt, Jansen, van Amelsvoort y Broeders (2006)
texto de 40-50 segundos				A 0.47; FA 0.67	A 0.47; FA 0.93	Philippon, Cherryman, Bull y Vrij (2007)

*Kreiman y Sidtis (2013 [2011], p.242) reportan el valor medio de la d' obtenidas en los experimentos 1 y 2 de Goggin *et al.* (1991); d'= 1.33 para la condición "misma lengua o dialecto" y d'=0.66 para la condición "lengua no familiar". Para el experimento 3 del mismo trabajo, reportan los valores de d' cuando los oyentes fueron los participantes monolingües de inglés.

Goldstein *et al.* (1981) ofrecen una interesante perspectiva al buscar posibles analogías entre el reconocimiento de rostros y el de voces (para una interesantísima reflexión en torno a este tipo de comparación se recomienda la lectura de Kreiman y Sidtis (2013 [2011], pp. 200-203). En el trabajo de Goldstein y colaboradores aluden a una comparación entre el reconocimiento de rostros de razas diferentes y el de voces que hablan una lengua extranjera, o la misma lengua del oyente pero con un fuerte acento extranjero. Al parecer, y según plantean estos autores, habida cuenta de los hallazgos obtenidos de las investigaciones en el reconocimiento de caras, los rostros de personas de la misma raza o grupo étnico se reconocen más fácilmente que los de otros grupos o razas (Chance, Goldstein y McBride, 1975; Malpass y Kravitz, 1969, citados por Goldstein *et al*., 1981, p. 217). Este fenómeno denominado "efecto de la raza", explican, se refleja en apreciaciones comunes del tipo "todos me parecen iguales", si se trata de reconocimiento facial, o en "todos me suenan igual", para los casos de reconocimiento de locutores que hablan una lengua desconocida (1981, p. 217). Teniendo esto presente diseñaron una serie de experimentos que enseguida se comentarán con la finalidad de probar que una influencia similar del "efecto de la raza" podría estar actuando en el reconocimiento de voces.

En el estudio de Goldstein *et al.* (1981) se plantearon tres condiciones experimentales diferentes. En el primer experimento (p. 218), un grupo de hablantes de inglés americano escuchó una frase en inglés pronunciada por hablantes taiwaneses, por negros americanos y por blancos americanos. En la prueba de reconocimiento, los oyentes –hombres y mujeres americanos negros y blancos– oían la misma frase en inglés pronunciada por 4 locutores diferentes y debían reconocer cuál de todos ellos era el locutor objetivo. Se calculó el porcentaje de aciertos en los reconocimientos y estos no resultaron ser estadísticamente diferentes entre los grupos de oyentes (taiwaneses 81%, americanos negros 82%, americanos blancos 85%). El resultado llevó a los autores a concluir que el reconocimiento de locutores con acento nativo no era mejor que el de los locutores con acento extranjero, aunque se observó una tendencia a que los hablantes chinos se reconocieran peor (1981, p. 218).

En el segundo experimento (pp. 218-219) redujeron la longitud del estímulo a una palabra asilada, lo cual no solo revirtió, como era de esperar, en una disminución significativa del reconocimiento general de los tres grupos de locutores (de 83% a 50%), sino que además se comprobó que este efecto había sido particularmente acusado en la identificación de hablantes chinos (37%), mientras que no se apreciaron divergencias

significativas entre los grupos de voces americanas negras y blancas, 56% y 55%, respectivamente. Estos datos indican, según los investigadores, que la reducción de la longitud del estímulo y la consecuente reducción en la información que conlleva afecta principalmente a las hablas con acento (1981, p. 219).

En el tercer experimento (pp. 219-220) analizaron cómo reconocían los oyentes nativos de inglés americano sin conocimientos de español a los hispanófonos nativos hablando inglés con un acento muy marcado y hablando español sin acento. Los locutores nativos de español pronunciaban una frase breve en inglés (con un acento extranjero marcado) y otra también breve en español sin acento. Luego de un intervalo de retención de 10 minutos, los oyentes americanos escucharon diez voces pronunciando la misma frase en cada una de las dos modalidades de lengua correspondiente. No se observó ningún efecto de la lengua en el reconocimiento de los locutores (0.58 y 0.57 de Aciertos para la condición español e inglés acentuado, respectivamente; y 0.18 de Falsas Alarmas en ambas condiciones).

La interpretación del conjunto de resultados, en opinión de los autores, apenas les permite aportar prueba alguna sobre el efecto de la lengua en el reconocimiento de voces. Por un lado, los resultados indican que conocer el idioma no es un requisito indispensable para el reconocimiento de los locutores a partir de sus voces y que "practically speaking, voice recognition is just as good (or as por) for foreing voices as it is for native voices" (Goldstein *et al.* 1981, p. 220). Por otro, constatan que limitar la cantidad de información que transmite la voz mediante la reducción del estímulo que se presenta a los oyentes perjudica el reconocimiento general, y que este se vuelve aún más considerable en el reconocimiento de voces con acento (1981, p. 220).

El efecto de la longitud del estímulo, comentan Goldstein *et al*., (1981, p. 217), en el reconocimiento de voces había ya sido detectado en estudios anteriores como los de Bricker y Pruzansky (1966) y Pollack, Picketi y Sumby (1954) que investigaron la capacidad para identificar voces familiares a partir de la escucha de muestras de voz de duración y contenido variables. Según demostraban estos estudios, el reconocimiento mejora conforme aumenta la duración del estímulo y el repertorio de sonidos de la muestra. Además, mostraron que los oyentes son capaces de identificar correctamente voces familiares cuando los estímulos se reproducen en reverso, aunque en estas condiciones, lógicamente, el reconocimiento empeora. Con dicha manipulación se consigue un continuo de sonido sin significado que no puede siquiera asociarse a ningún idioma en particular. De acuerdo con estos resultados es posible postular que el

reconocimiento de la voz sea en cierta medida independiente del reconocimiento o la comprensión del habla y que los oyentes muy probablemente se sirven para el reconocimiento de las voces de claves acústicas no vinculadas directamente con la lengua (Goldstein *et al*., 1981, pp. 219-220 y véase también San Segundo, Foulkes y Hughes, 2006, p. 1).

El estudio de Thompson (1987) demostró un efecto significativo del idioma en el reconocimiento de hablantes. Su descripción se realizará a partir de la que ofrecen Schiller *et al*., (1997, pp. 3-4). Thompson grabó a seis hombres bilingües inglés-español que leyeron tres versiones diferentes de dos textos breves: en inglés, en español, y en inglés con un acento español muy marcado. Los oyentes que realizaron la tarea de reconocimiento de voces (en concreto, una tarea de identificación de locutor en una rueda de reconocimiento) fueron oyentes nativos de inglés monolingües sin conocimientos de español. Los oyentes escucharon a un locutor leyendo un texto en una de las condiciones lingüísticas recién descritas. A la semana siguiente oyeron, en la misma condición lingüística, el otro texto leído por seis locutores diferentes (entre ellos se encontraba la persona que había leído el texto en la primera ocasión).

Los resultados de este experimento mostraron que los participantes monolingües de habla inglesa identificaron con mayor exactitud a los locutores cuando hablaban en inglés, que cuando lo hacían en inglés con acento español muy marcado, y, a estos últimos, mejor a su vez que cuando hablaban español (Aciertos: 0.65, 0.52 y 0.38, respectivamente). Sin embargo, las proporciones de Falsas Alarmas fueron idénticas en las tres condiciones experimentales (0.27). En el mismo trabajo, Thompson repitió el experimento simulando una rueda de reconocimiento con objetivo ausente, es decir, la voz que los oyentes debían detectar no estaba presente entre las seis que oían los participantes. Aunque la proporción de Falsas Alarmas aumentó (0.56), no se encontraron diferencias estadísticamente relevantes ni en la proporción de Falsas Alarmas ni en la de Rechazos Correctos para ninguna de las tres condiciones lingüísticas. Un tercer experimento, que utilizaba únicamente la lectura en inglés y en español del texto mostró el mismo efecto beneficioso de la lengua sobre el reconocimiento de locutores, siendo significativamente superiores las tasas de Aciertos para las voces de los angloparlantes que para la de los hispanoparlantes. Nuevamente, no se constató ninguna diferencia en la proporción de Falsas Alarmas. Frente a estos resultados, Thompson (1987) concluyó que el efecto de la lengua en la identificación de locutores se explica por la mayor capacidad que los oyentes nativos tienen para

identificar rasgos idiosincráticos de hablantes de su propia lengua (tomado de Schiller *et al*., 1997 pp. 3-4).

Un poco más tarde, Goggin, Thompson, Strube y Simental (1991) publican los resultados de un estudio que confirma que la familiaridad de la lengua de los oyentes con la de los locutores influye positivamente en la identificación de voces. Goggin *et al.* (1991) llevaron a cabo 4 experimentos para probar los efectos del conocimiento de la lengua en la identificación de voces.

En el primer experimento (Goggin *et al.*, 1991, pp. 450-452), participaron seis hombres bilingües inglés-alemán que grabaron una versión en inglés y otra en alemán de dos pasajes de dos textos breves. Los oyentes fueron un grupo de participantes nativos de inglés americano monolingües, sin conocimientos de alemán, que primero escucharon una versión del primer pasaje producido por el locutor objetivo y, después de un intervalo de retención de cinco minutos, tuvieron que identificar la voz de ese locutor en una rueda de reconocimiento. En esa rueda se incluyeron las voces de los seis locutores que producían el segundo pasaje del texto en la condición de lengua correspondiente. Los jueces realizaron el experimento en las dos condiciones lingüísticas (lengua familiar – lengua no familiar). Los resultados de este primer experimento pusieron de manifiesto un efecto significativo de la lengua porque la identificación fue mejor para las voces inglesas que para las alemanas. Los oyentes norteamericanos monolingües sin conocimiento de alemán lograron identificar a los locutores cuando estos hablaban su propio idioma (d'=1.19) y no lo consiguieron (d'=0.58) cuando esos mismos locutores hablaban alemán, es decir, una lengua desconocida para ellos (Goggin *et al.*, 1991, p. 450).

En el segundo experimento (pp. 452-453), un grupo de oyentes nativos de alemán sin ningún conocimiento de inglés escucharon las voces de los mismos hablantes bilingües inglés-alemán del experimento 1. El procedimiento fue idéntico al de la prueba anterior, con la diferencia de que el intervalo entre la escucha de la voz objetivo y la rueda de reconocimiento fue mayor. Los resultados apuntaron en la misma dirección: las voces se identificaron mejor cuando los locutores hablaban en alemán, en este caso la lengua compartida por oyentes y locutores (d'=1.47), que cuando lo hacían en inglés (d'=0.73), lengua desconocida para los oyentes (1991, p. 450).

El tercer experimento (pp. 453-455) corroboró también los resultados de los dos anteriores estudios. Con él se pretendía confirmar que, si la familiaridad lingüística era realmente un factor decisivo en el reconocimiento de locutores, entonces los oyentes

bilingües podrían identificar por igual las voces en cualquiera de los dos idiomas. Seis hombres bilingües inglés-español grabaron tres versiones diferentes de los pasajes, en inglés, en español y en inglés con fuerte acento español. Dos grupos de oyentes fueron los jueces de este tercer experimento, un grupo de estudiantes bilingües español-inglés y un grupo de monolingües anglohablantes. Los oyentes escucharon el primer pasaje en una condición de lengua particular y, después de un intervalo de retención de 30 minutos, realizaron la tarea de identificación que consistía en reconocer la voz del locutor objetivo en una rueda de seis voces distintas. Durante la rueda de reconocimiento, los jueces oyeron el segundo texto producido en la misma condición de lengua del primer texto que escucharon. Los resultados de la prueba mostraron otra vez que la familiaridad con la lengua del locutor tiene un efecto beneficioso en la identificación de voces porque los oyentes ingleses monolingües identificaron mejor las voces inglesas (d'=1.34) que las voces inglesas con acento (d'=0.95) y mejor aún que las voces de los locutores que hablaban español (d'=0.72). De otra parte, los oyentes bilingües (español-inglés) identificaron, aunque con diferencias, las voces en las tres condiciones: locutores hablando en inglés (d'=1.22), locutores hablando en inglés con fuerte acento español (d'=1.05) y locutores hablando en español (d'=1.34), (véase Goggin *et al.*, 1991, p. 450).

El resultado global de estos tres primeros experimentos abona la hipótesis de que el reconocimiento de voces es más preciso cuando el oyente está familiarizado con el idioma que se habla. Sin embargo, y como apuntan los autores, no permite desentrañar cuáles son los posibles factores lingüísticos que conducen a dicha mejora porque en la señal acústica se entrelazaban claves fonéticas, léxicas y semánticas (Goggin *et al*., 1991, p. 455).

El cuarto y último experimento del estudio de Goggin y colaboradores (1991, pp. 455-456) intenta con su diseño dilucidar cómo afecta al reconocimiento esta posible ventaja que supone la familiaridad con la lengua. Los oyentes, que dominaban muy bien el idioma inglés, tuvieron que realizar la misma tarea de identificación que en las anteriores condiciones experimentales, pero a partir de la escucha de pasajes de textos en inglés con la información fónica, semántica y sintáctica intacta, o a partir de alguna de las manipulaciones realizadas en el texto conducentes a degradar o cancelar alguna de dichas claves. Las transformaciones fueron de tres tipos: (1) se mezclaron palabras con el objetivo de que esa mezcla produjera pasajes anómalos desde un punto de vista semántico, pero en los cuales se mantuvieran algunas claves sintácticas de la lengua y se

preservara el léxico; (2) se mezclaron sílabas pero manteniendo la fonotaxis normativa de la lengua; y (3) se presentaba el texto reproducido en reverso; en este caso, se destruían las señales fónicas, semánticas y sintácticas habituales de la lengua. El mismo tipo de transformación se presentó para las comparaciones. El reconocimiento empeoró progresivamente conforme el texto se volvía más incomprensible (p. 450): d'= 1.40 (condición de texto normal, sin transformaciones), d'= 1.22 (transformación tipo 1 "palabras mezcladas"), d'=0.68 (tipo 2 "sílabas mezcladas"), y d'= 0.38 (transformación tipo 3 "habla invertida"), véase Goggin *et al*., (1991, p. 450).

El trabajo de Schiller, Köster y Duckworth (1997) reexamina los resultados de un estudio previo realizado por Schiller y Köster (1996) en el que confirmaban que la familiaridad con el idioma afecta la capacidad de reconocer voces y afirmaban que los oyentes que dominan la lengua de los locutores que pretenden reconocer no solo utilizan claves acústicas sino que también se sirven de información lingüística más general. Es así que Schiller y colaboradores parten de una nueva hipótesis: si es cierto que los oyentes decodifican información lingüística para intentar reconocer a un locutor, entonces los oyentes de lenguas diferentes deberían ser igualmente capaces de identificar a los locutores cuando la información lingüística no está disponible (pp. 5-6). Los autores recuerdan que Goggin *et al.* (1991) ya habían observado que la distorsión de los estímulos afectaba notablemente el reconocimiento de locutores y que este empeoraba conforme el habla se volvía cada más incomprensible. Goggin *et al.* (1991) valoraron este efecto en su cuarto experimento, como se ha visto antes, únicamente con oyentes que tenían un buen dominio del idioma inglés, que era la lengua que hablaban los locutores en la grabación original, antes de las manipulaciones posteriores. Por ello, remarcan Schiller *et al*. (1997, p. 6), en el estudio de Goggin y colaboradores no se llegó a probar si el efecto de la lengua en la identificación del locutor desaparece cuando los estímulos van perdiendo información lingüística hasta perder por completo su sentido. Esto es precisamente lo que pretenden demostrar Schiller *et al*. (1997) con su trabajo. Para ello, diseñaron un experimento en el que intentaron eliminar toda la información lingüística posible haciendo algunas manipulaciones en la grabación. Primeramente, seis hablantes nativos de alemán leyeron un texto de 140 palabras. Luego, todas las sílabas del texto grabado se sustituyeron por /ma/ para minimizar el efecto de la lengua. Como aún así puede conservarse información lingüística de tipo prosódico, de ese texto seleccionaron arbitrariamente algunos fragmentos de entre 4 y 8 segundos cada uno que volvieron a grabar a través de línea telefónica (que elimina

frecuencias inferiores a 300 Hz y superiores a 3.500). El procedimiento consistió en un test de identificación forzada en el cual los oyentes debían responder "sí" o "no", según reconocieran o no la voz objetivo en una rueda de reconocimiento. En el experimento perceptivo participaron de forma separada tres grupos de oyentes (oyentes nativos de alemán, oyentes nativos de inglés sin nociones de alemán y oyentes nativos de inglés con conocimiento de alemán). Los oyentes se familiarizaron primero con la voz objetivo escuchando la lectura del texto completo durante 5 minutos. Luego de un descanso, escucharon las muestras manipuladas y presentadas en un orden aleatorio y debieron decidir para cada una de ellas si pertenecía al mismo locutor objetivo con el que se habían familiarizado previamente. En cada juicio, los oyentes disponían de 5 segundos para tomar su decisión (1997, pp. 6-7). Schiller *et al*. (1997) exploraron las medidas de discriminabilidad (d') y los sesgos de respuesta (criterio c), en el marco de la *Teoría de Detección de Señales*. Un primer dato interesante aportado por este experimento, apuntan los autores, es que, como era previsible, la capacidad para reconocer las voces disminuye cuando se suprime información lingüística pero que a pesar de ello son capaces de identificar voces y locutores. No observaron diferencias estadísticamente relevantes entre los valores de d prima para los tres grupos de jueces aunque sí constataron diferencias en los criterios de respuesta, más liberal en el caso de los oyentes no nativos. De este modo concluyen que los oyentes de distintos perfiles lingüísticos se desempeñan de forma bastante similar en la tarea de reconocimiento de locutor cuando se elimina la mayor parte de la información lingüística de los estímulos. Schiller y colaboradores interpretan que estos resultados no contradicen la presunción de que la familiaridad lingüística mejora el reconocimiento de locutores precisamente porque cuando la información lingüística desaparece los nativos y no nativos se desempeñaron de forma muy parecida. Según expresan, el hecho de no haber encontrado diferencias en la discriminación de los distintos grupos de oyentes cuando prácticamente no hay información lingüística indica que el efecto de la lengua es real (1997, p. 16).

Un último estudio que se comentará es el que recientemente han publicado San Segundo, Foulkes y Hughes (2016). Aunque, a diferencia de los anteriores, este no se ocupa de realizar pruebas de identificación o discriminación de voces, resulta de mucho interés pues se centra en evaluar el efecto del idioma en la evaluación perceptiva que los oyentes realizan de la cualidad de voz de locutores muy parecidos a partir de muestras de habla breves (~3 segundos). En dicho estudio, dos grupos de oyentes (nativos de

español y nativos de inglés sin conocimientos de español) juzgan las similitudes en la voz de un grupo de hablantes con características biológicas, lingüísticas y sociolingüísticas muy semejantes (5 pares de gemelos monozigóticos masculinos nativos de español, de la variedad centro-peninsular, con edades muy próximas y con similares valores de f_0). Las calificaciones de similitud realizadas por los dos grupos de jueces se analizaron mediante la técnica *Escalamiento Multidimensional* a fin de estudiar las representaciones perceptivas generadas por las voces. La hipótesis de los autores es que las semejanzas entre ellas se podrán explicar exclusivamente por la cualidad de la voz y que serán valoradas de forma holística y muy semejante por los oyentes nativos y no nativos de español (San Segundo *et al*., 2016, p. 2).

Los resultados parecen corroborar dicha hipótesis porque tanto unos jueces como otros clasificaron las semejanzas y las distancias entre las voces de los locutores de una forma muy parecida. Según señalan, esto podría apuntar a que en oyentes de lenguas distintas están funcionando estrategias de percepción similares y que todos ellos recogen las mismas claves perceptivas cuando evalúan la semejanza de voces a partir de estímulos con escasa información lingüística. La utilización de muestras de habla breves, comentan los autores, dificulta el que los oyentes atiendan a claves dependientes del idioma y, de algún modo, obligaría a los oyentes a basar sus juicios de similitud casi exclusivamente en la voz. No obstante, aclaran los autores, los comentarios cualitativos realizados por los oyentes indicaron que ciertos aspectos como la velocidad de articulación pudieron ayudarles a percibir las semejanzas entre las voces de los distintos locutores (San Segundo *et al*., 2016, p. 4). Finalmente, concluyen que la ventaja de conocer la lengua del locutor puede lógicamente funcionar si los oyentes escuchan muestras de habla más extensas.

Este breve repaso bibliográfico es suficiente para mostrar la heterogeneidad de los diseños experimentales, la cual lógicamente dificulta la interpretación comparada de los resultados. Como se ha visto, no siempre coincide el grado de familiaridad entre las lenguas ni los grupos de oyentes que se comparan; tampoco es equivalente el tamaño de las muestras de habla sobre las cuales se realizan los experimentos perceptivos, y, por último, las medidas que dan cuenta de la habilidad de los oyentes para reconocer las voces son, igualmente, diferentes. Según puede fácilmente observarse de la Tabla 2.1 (tomada de Kreiman y Sidtis), los estímulos utilizados en los distintos estudios son a veces palabras aisladas, a veces frases aisladas de longitud variable y, en ocasiones, fragmentos de textos o textos completos, más o menos extensos. Disparidad semejante

se observa en las medidas calculadas. Mientras algunos contemplan los porcentajes de respuestas correctas, otros calculan las proporciones de Aciertos, otros consideran las proporciones de Aciertos y Falsas Alarmas y otros calculan la medida de discriminabilidad (d') que proporciona la *Teoría de la Detección de Señales*; finalmente, unos consideran además el criterio de respuesta de los jueces y otros no. Tal diversidad metodológica responde, muy probablemente, al hecho de que las distintas investigaciones están muchas veces motivadas por preocupaciones prácticas concretas, por lo común, aplicaciones en el ámbito judicial.

Aunque evidentemente las comparaciones entre los resultados de los distintos experimentos deben realizarse con cautela, es posible extraer algunas conclusiones coincidentes entre todos ellos. Como fue ya señalado por Kreiman y Sidtis (2013 [2011, pp. 241-243), la mayor parte de los resultados aportados por las distintas investigaciones reseñadas apoyan la hipótesis de que un mejor cocimiento del idioma tiene consecuencias positivas en la identificación del locutor a través de su habla y que, por tanto, la familiaridad lingüística desempeña un papel importante en el reconocimiento de las voces. El análisis de los resultados también indica que, conforme aumentan la duración y la variedad de los estímulos de la muestra, se incrementa la precisión en los reconocimientos. El hecho de que el reconocimiento de locutores mejore con el aumento de la duración de las muestras de voz podría explicarse muy probablemente porque se dispone de más indicios: un repertorio de sonidos más variado, la presencia de patrones rítmicos y entonativos, etc. En definitiva, se trataría de claves dependientes de la lengua a las que los oyentes podrían acceder, sobre todo si son hablantes nativos de esa lengua. La superioridad de los oyentes nativos frente a los no nativos en el reconocimiento de locutores sería válida, por consiguiente, principalmente en los casos en que las muestras de voces fueran relativamente extensas y variadas pero quizás no resultaría tan evidente en los casos en que la información segmental y prosódica no estuviera disponible para el oyente.

Tomados en su conjunto, la mayor parte de los estudios hasta aquí reseñados sugieren que son ciertos los aspectos que se retoman, muy resumidos, a continuación:

1) El reconocimiento de voces y de locutores mejora si se comparte o se conoce la lengua del locutor.

2) La capacidad de los oyentes para evaluar voces y reconocer locutores disminuye cuando se elimina información fonética y lingüística más general del habla.
3) Cuando se pierden claves lingüísticas aún es posible apreciar similitudes y reconocer diferencias entre voces, y los oyentes nativos y no nativos muestran desempeños muy similares en esos casos.
4) Los oyentes también perciben e interpretan características vocales que no necesitan del conocimiento del idioma en el que habla el locutor para evaluar voces y reconocer hablantes. Entre estas, podrían encontrarse las que se relacionan con la cualidad de la voz.

Capítulo 3
Metodología

3. Metodología

El presente capítulo expone la metodología seguida en esta tesis, la cual se organiza en cuatro bloques diferentes que se corresponden con las fases de la investigación. Primero, se plantean las hipótesis que esta tesis busca examinar (§3.1), luego se explica la metodología utilizada para la *generación del corpus* (§3.2), es decir, los criterios seguidos para elegir a los informantes, el material fonético utilizado y el procedimiento de obtención de las muestras sonoras. En tercer término, se exponen los criterios metodológicos aplicados al *análisis de las muestras* (§3.3), esto es, el análisis de los parámetros fisiológicos y acústicos relacionados con la actividad de los pliegues vocales, y el análisis de los parámetros acústicos derivados de la acción conjunta de la fuente y el filtro. Por último, y correspondiéndose con la fase del estudio perceptivo, se presenta la descripción de la metodología seguida para la realización de los test de discriminación de voces (§3.4). A lo largo de este capítulo, no solo vienen justificadas las decisiones sobre los procedimientos metodológicos tomadas en cada caso, sino que también se exponen las motivaciones para la elección de los programas y herramientas utilizados.

3.1.Planteamiento de las hipótesis

Esta tesis pretende profundizar en el conocimiento del *falsetto* realizando un estudio detallado de los aspectos acústicos que lo caracterizan y lo diferencian de la voz modal, así como de los posibles correlatos articulatorios a los que apuntan dichos rasgos acústicos. Parece razonable que, antes de valorar en qué medida el *falsetto* afecta la discriminación de voces y el reconocimiento del locutor y si pueden o no existir diferencias en el comportamiento de los oyentes que comparten la misma lengua con el locutor, –otro aspecto que explora este trabajo–, se establezcan primeramente qué rasgos del comportamiento laríngeo están realmente cambiando en el *falsetto* y potencialmente cuáles no se modifican. Por tanto, un primer objetivo de esta tesis es el de describir las diferencias acústicas entre los registros de voz modal y *falsetto*.

Las alteraciones que los hablantes efectúen en el mecanismo de fonación habitual, y más concretamente, los cambios de registro, complican seriamente y en ocasiones hasta impiden el reconocimiento de los hablantes (§2.4.2). Aun así, es dable pensar que, pese a ello, pudieran permanecer inalterables ciertos parámetros vocales que, por no depender del registro sino del locutor, constituyeran índices de la individualidad de la voz.

Siguiendo la hipótesis general previamente planteada en el proyecto CIVIL [25](Alves, Gil Fernández, Pérez Sanz y San Segundo, 2014) que postulaba la existencia de rasgos fonéticos idiosincrásicos que no pudieran ser susceptibles de modificación, una suerte de "huella biométrica de la voz"[26], se formula la primera hipótesis de esta investigación:

Hipótesis 1. En el paso de voz modal a falsetto, habrá rasgos laríngeos que modifiquen…

a) todos los locutores por ser intrínsecos al cambio de registro.

b) solo algunos locutores o ninguno.

Al contrastar el comportamiento de los rasgos fonatorios de los individuos en voz modal y *falsetto*, se pretende poder aislar estos dos grupos diferentes de rasgos para luego poder derivar cuáles podrían ser los rasgos definitorios del registro del *falsetto* (grupo a). Es importante aclarar que solo potencialmente los rasgos que no se modifiquen podrán vincularse con la cualidad individual de la voz de los locutores porque algunos de estos rasgos inalterados (si es que se encuentran) podrían no presentar grandes variaciones entre locutores[27].

[25] La hipótesis del proyecto CIVIL postulaba que existen parámetros vocales –principalmente los laríngeos– que resisten al disimulo voluntario de la voz en el cambio de registro modal-*falsetto*, es decir, "…existen MARCADORES INDIVIDUALES ROBUSTOS que toleran los intentos deliberados de distorsión de la cualidad de voz a través de la variación en el modo de fonación, en concreto a través de la fonación *creak* y *falsetto*, que son, como ya quedó dicho, las alteraciones que con más frecuencia se emplean en el disimulo con fines delictivos" (Alves *et al*., 2014, p. 603).

[26] "Lo que no se ha investigado hasta el momento es la posibilidad de que exista alguna 'huella biométrica' –en expresión del profesor Gómez Vilda- localizada en la onda glotal. (Alves *et al*., 2014, p. 604).

[27]En un estudio posterior se realizará también la comparación interlocutor con los parámetros del grupo b) y se buscarán contrastar los valores de cada locutor con la población de referencia para obtener dos grupos de datos: grupo de parámetros que no varían al cambiar de registro pero que no permiten identificar locutores puesto que sus valores son muy similares entre los locutores; y grupo de parámetros que podrían ser individualizadores de la voz.

Si es cierto que perviven en la voz algunas características laríngeas inmutables en el cambio de registro vocal, si se comprueba que permanecen rasgos laríngeos constantes en la voz modal y en la fingida en *falsetto*, estos podrían encontrar eco en la percepción, es decir, podrían ser advertidos por el oído humano.

Un segundo objetivo de esta tesis es tratar de dilucidar si los parámetros de la fonación que se alteran (y los que no) en el cambio de registro modal-*falsetto* son perceptivamente relevantes para los oyentes. Así, se plantea la segunda hipótesis de este estudio:

Hipótesis 2. Los rasgos laríngeos son una clave perceptiva para el reconocimiento de locutores.

Esta hipótesis prevé, por tanto, que las señales vocales que provienen de la cualidad de voz en sentido restringido, es decir, los rasgos laríngeos, serán relevantes perceptivamente en el reconocimiento de hablantes. Dicho de otro modo, que los rasgos no alterados pueden constituir una clave perceptiva para el reconocimiento de locutor en casos de cambio de registro hacia el *falsetto*.

Para comprobar esto último, los resultados de los test perceptivos deberían demostrar que los locutores se reconocen por encima del nivel del azar a pesar del disimulo en *falsetto*:

Hipótesis 3. Los oyentes pueden reconocer a un mismo locutor por encima del nivel del azar, aunque haya cambiado de registro modal-falsetto.

En §2.5 se repasaron los principales estudios dedicados a evaluar el efecto del idioma en la percepción y el reconocimiento auditivo de la individualidad del hablante y, más específicamente, por parte de oyentes que pudieran o no compartir la lengua con el locutor. Tomados en su conjunto y a pesar de la divergencia metodológica de los estudios, se observó que los oyentes tienen la capacidad de atender a características de la voz que parecen ser independientes de la lengua para reconocer rasgos individualizadores del locutor.

Los experimentos que se realizan en esta tesis intentarán dar apoyo adicional a este supuesto y pretenden probar que las propiedades no lingüísticas de la cualidad de voz, en concreto, las derivadas del comportamiento laríngeo, pueden desempeñar un

papel importante en el reconocimiento de voces y de locutores, y que, cuando esto ocurre, conocer la lengua del hablante no supone ventaja sustancial. Es decir, cuando casi solo estén disponibles para el oyente rasgos laríngeos de la cualidad de la voz, los oyentes que comparten la lengua con los locutores que hay que reconocer no tendrán una ventaja comparativa respecto de los oyentes que no la comparten o no la conocen. De este modo se formula la cuarta y última hipótesis:

Hipótesis 4. Puesto que los registros de fonación son, para ambas lenguas, aspectos extralingüísticos, los oyentes de español se van a comportar igual que los oyentes italianos en el reconocimiento de locutores italianos que disimulan su voz en falsetto.

3.2.Generación del corpus

3.2.1. Selección de informantes

Las muestras sonoras fueron recogidas de 6 locutores masculinos, hablantes nativos de italiano (italiano regional de Nápoles), con edades comprendidas entre los 25 y 35 años y con estudios universitarios concluidos. Los locutores demostraron tener una buena capacidad de disimulo de la voz mediante el registro del *falsetto*. Se trata, pues, de un grupo homogéneo de informantes.

La captación de informantes se realizó a través del envío de correos electrónicos a listas de distribución del CCHS-CSIC y del *Istituto Italiano di Cultura* de Madrid. Podía participar todo aquel que estuviera interesado siempre que cumpliera con todos los criterios de elegibilidad que se detallan a continuación:

(1) Ser mayor de edad y tener entre 20 y 40 años.
(2) Ser hablante nativo de italiano (regional de Nápoles).
(3) No tener ninguna patología vocal, ni dificultades en el habla, ni problemas de audición.
(4) No llevar ningún aparato dental, ni prótesis, ni implantes de ningún tipo en la boca.
(5) No haber consumido alcohol ni otro tipo de drogas antes de las sesiones de grabación.

Con anterioridad a las sesiones de grabación, los informantes se pusieron en contacto a través del correo electrónico. Si cumplían, en principio, con todos los requisitos, se les enviaba un documento explicativo de los propósitos del estudio y algunos detalles acerca de su participación, así como un *Power Point* con audios para que se familiarizaran con el registro del *falsetto* y con la tarea que deberían realizar. Luego, se convenía una cita en el Laboratorio de Fonética del CSIC de Madrid.

Se grabó a un total de 7 informantes. Luego, sin embargo, en la fase de escucha y análisis preliminar de los audios se observó que uno de ellos no había conseguido realizar correctamente el *falsetto* ni había sido capaz de mantener el disimulo a lo largo de todas las frases. Este informante fue, por tanto, posteriormente descartado, por lo que solo se utilizaron las grabaciones de los otros 6 informantes cuyos datos se recogen en las Tablas 3.1 y 3.2. Todos ellos cumplimentaron el formulario de datos personales y firmaron su consentimiento, y a cada uno se le pagó la cantidad de 15 euros para compensar su colaboración.

Tabla 3.1: Descripción de los sujetos que participaron como informantes.

Iniciales del nombre y del apellido	Locutor	Sexo	Edad	Nivel educativo	Fumador
A. C.	L1	Hombre	29	Titulado universitario	Sí
M. D.	L2	Hombre	34	Titulado universitario	Sí
V. P.	L3	Hombre	30	Titulado universitario	Sí
P. N.	L4	Hombre	34	Titulado universitario	Sí
M. C.	L5	Hombre	30	Titulado universitario	No
A. N.	L6	Hombre	33	Titulado universitario	No

Tabla 3.2: Descripción del perfil lingüístico de los sujetos que participaron como informantes.

Iniciales del nombre y del apellido	Locutor	Lengua nativa	Otros idiomas	Nivel (Deficiente: 1 2 3 4 5 Muy bueno)
A. C.	L1	Italiano	Español	4
			Inglés	3
M. D.	L2	Italiano	Español	4
			Inglés	4
V. P.	L3	Italiano	Español	4
			Inglés	3
P. N.	L4	Italiano	Español	4
			Inglés	3
M. C.	L5	Italiano	Español	4
			Inglés	4-5
A. N.	L6	Italiano	Español	4
			Inglés	3

3.2.2.Material fonético

El objetivo principal que guió el diseño de los estímulos fue dirigir la atención de los oyentes a la percepción de la voz derivada principalmente de la configuración o comportamiento laríngeo de los locutores.

Para no limitar la percepción del registro a una sola palabra se decidió trabajar en el dominio de la frase. Se propuso a los informantes la lectura de una única frase breve e idéntica para todos los locutores a fin de controlar, en la medida de lo posible, la aparición de rasgos temporales y entonativos del habla característicos que pudieran contribuir decididamente a la tarea posterior de discriminación de hablantes. Se buscó que la frase fuera breve pero al mismo tiempo lo suficientemente extensa como para que el cambio de registro estuviera bien asentado. En el diseño de la frase también se pretendía cancelar el significado, de forma que fuera una frase posible para esa lengua pero carente de sentido. La frase elegida fue: “Dica dadiva adagio e vada a casa”. Se trata de una pseudofrase, pues “dadiva” es una palabra sin sentido en italiano. Con ella se perseguía también una simetría con las frases del corpus CIVIL, *Cualidad Individual de la Voz e Identificación del Locutor* (Alves *et al.,* 2014; Alves, Fernández Trinidad,

Gil Fernández, Infante, Pérez-Sanz y San Segundo, 2012; San Segundo, Alves y Fernández Trinidad, 2013) en las que se da la repetición de palabras formadas por sílabas con la estructura CV, donde C es sonora y no nasal y V es [a] tónica[28].

A pesar de que existen otras soluciones factibles para intentar que un oyente se concentre en la cualidad de la voz en el sentido más restrictivo del término –cualidad resultante del comportamiento de la fonación– y se abstraiga de todo lo demás[29], en esta tesis se ha optado por tratar de hacerlo de la forma más natural posible.

Los informantes repitieron la misma pseudofrase en ambos registros, grabando un total de 30 repeticiones en voz modal y 30 en *falsetto*. Como se verá luego en el apartado dedicado al análisis del comportamiento glótico, para extraer las características fonatorias de los locutores en los diferentes registros se analizaron únicamente las vocales [a] que se encontraban en sílaba tónica y libre o abierta (sin coda), entre dos consonantes sonoras no nasales, en concreto dA (de *dadiva* y *adagio*) y vA (de *vada*).

[28]Como es habitual en los estudios sobre cualidad de voz, se eligió la vocal [a] para extraer los parámetros biomecánicos de la voz por ser la vocal que permite una mejor separación de la fuente sonora de los efectos provocados por el filtro que realiza el tracto vocal. Alonso Palacio (2018) ofrece una explicación detallada de por qué es conveniente descartar vocales cerradas, como la /i/ o la /u/, o excesivamente nasalizadas; y escoger /a/ o /e/ por su mejor predisposición para reconstruir la fuente glótica. "[…] otro aspecto importante a tener en cuenta para la reconstrucción de una buena fuente glótica es la interacción de la fuente y el filtro. Esta interacción se basa en la reverberación de la onda por su paso por los pliegues vocales hasta llegar a la zona de radiación. Las vocales cerradas como son la /i/ y la /u/, además de algunas consonantes nasales, ofrecen una mayor alteración en el patrón de fonación de la fuente glótica" (Alonso Palacio, 2018, p. 64). Sobre este punto, véase también Gobl y Ní Chasaide, 2010, p. 380). Por otra parte, se escogieron las [a] tónicas por presentar mayor duración y estabilidad.

[29] En la práctica forense se ha señalado la utilidad y el poder individualizador de la cualidad de voz y, por ejemplo en España algunos especialistas la toman en consideración en sus análisis aunque advierten la dificultad para evaluarla de forma objetiva (véase Delgado, 2001, pp. 254-262). Delgado aboga por la opción de análisis que ha convenido en denominar A.P.R.E.S. (*Aural Perception on Reverse Speech*), esto es, una escucha repetida y sistemática de las grabaciones en un orden temporal invertido (p. 256). Según expresa el mismo Delgado: "…la percepción de emisiones de habla mediante A.P.R.E.S. nos presenta la cualidad vocal desnuda, como si se tratase de una supraestructura que trasciende a la suma de sus distintos componentes. Un conjunto de rasgos sonoros que flotan en el aire carentes de otro significado que el de su propia identidad como estructura acústica" (2001, p. 258). Sin embargo, en esta tesis no se ha optado por esta técnica, pues se cree que el habla invertida, si bien destruye el significado, altera la fonotáctica de la lengua –piénsese por ejemplo que las sílabas abiertas se vuelven cerradas y viceversa– y eso podría desvirtuar la percepción de la emisión en un sentido no explorado aún. Para intentar aislar la cualidad de la voz en el sentido laríngeo también se ha recurrido a otras estrategias. Por ejemplo, en trabajos como el de Scherer, Ladd y Silverman (1984, p. 1350) se utiliza la técnica *random spliced* para volver ininteligible una frase y conservar al mismo tiempo la f_0 global y la cualidad de la voz. El procedimiento consiste en cortar la frase con contenido léxico en fragmentos de 300 milisegundos (alrededor de 3 segmentos) junto con los segmentos adyacentes superpuestos de 3 ms. Las partes solapadas se atenúan a la amplitud de cero para evitar la inclusión de transiciones. Una vez cortados los distintos fragmentos y atenuadas las transiciones solapadas se mezclan de manera aleatoria en órdenes distintos. Finalmente, en las técnicas actuales de síntesis de voz por concatenación de elementos se prefiere tomar la sílaba como unidad, puesto que permite conservar la redundancia de claves y las transiciones internas a la sílaba, ambos aspectos muy difíciles de imitar de manera adecuada en el producto sintético. De este modo, el *output* conserva la naturalidad fonotáctica de la lengua, conserva las coarticulaciones y las transiciones.

Como se expuso antes, los informantes grabaron la pseudofrase "*Dica dadiva adagio e vada a casa*"; sin embargo, tanto para el experimento perceptivo como para el análisis posterior sobre la producción de la voz se utilizó únicamente el fragmento "*Dica dadiva adagio e vada*". De este modo se excluía el final del enunciado del estímulo auditivo que posteriormente evaluarían los jueces, puesto que las inflexiones correspondientes a la modalidad oracional variaban considerablemente y de manera bastante evidente de un locutor a otro y, por tanto, podrían constituir una pista clara para los oyentes sobre la individualidad del locutor. Al mismo tiempo, se evitaba luego tener que analizar los sonidos de la última porción de la frase con *BioMet®Phon*, ya que, por lo común, resulta más difícil mantener estable un *falsetto* al final del enunciado.

3.2.3.Obtención de las muestras

Las muestras se obtuvieron en el Laboratorio de Fonética del Centro de Ciencias Humanas y Sociales del Consejo Superior de Investigaciones Científicas (CCHS-CSIC, Madrid) durante los meses de enero, febrero y marzo de 2015. Se grabó a todos los informantes en una cámara insonorizada, utilizando el mismo equipamiento y las mismas características técnicas (véase la Tabla 3.3).

Tabla 3.3: Equipamiento utilizado y ajustes de grabación.

Equipamiento		
Equipo	Modelo	Fabricante
Micrófono de condensador	Omnidireccional Earset Æ E6i	Countryman
Interfaz de audio	ÆUA-25EX	Roland
Software	Adobe Audition 1.0 para Windows	Microsoft
Ajustes de grabación		
Frecuencia de muestreo	Resolución	Formato de audio
44100 Hz	16 bits	.wav

Cada locutor grabó un total de 60 frases (30 por registro) y dichas grabaciones se realizaron en dos sesiones diferentes para cubrir la posible variabilidad intrahablante[30].

[30] Como se ha señalado repetidamente en la bibliografía, los hablantes pueden diferir en la forma en la que cada vez pronuncian un mismo enunciado; en este caso, se habla de variabilidad intralocutor. Las

Posteriormente, con la ayuda de un equipo de fonetistas, se seleccionaron 10 frases por locutor (5 en voz modal y 5 en *falsetto*), en función de la calidad de las emisiones.

3.3.Metodología para el análisis de la producción de las muestras de voz

3.3.1. Análisis del comportamiento fonatorio con BioMetroSoft®

Para analizar la fonación se utilizó el programa informático *BioMet®Phon* [versión 2.3, Diciembre 2012] (http://www.biometrosoft.com), diseñado en la Universidad Politécnica de Madrid por el grupo de ingenieros que dirige el Dr. Pedro Gomez-Vilda[31]. Este programa es capaz de aislar el contenido biométrico de la señal glótica restando, de la señal vocal en su conjunto, la función de transferencia que ejerce el tracto vocal y el efecto de la radiación labial. Calcula la onda glotal a partir de la señal acústica mediante un filtrado inverso y, a partir de ello, estima cuáles son los valores de 72 parámetros relativos al comportamiento glótico. La lista completa de los parámetros se ofrece en la Tabla 4y puede también consultarse en Gómez-Vilda y Nieto-Lluis, 2015, pp. 22-24; Palacios Alonso, 2018, pp. 68-73, entre otros trabajos).

---

diferencias que puedan existir entre diversos hablantes determinan, en cambio, la variabilidad inter-locutores.

[31] Según se explica en su página web (http://www.glottex.com/es/sobre-nosotros), BioMetroSoft® es una compañía "dedicada al desarrollo de herramientas y aplicaciones para la Biometría de la voz, y a proporcionar soluciones y servicios de consultoría basadas en la misma tecnología. Fundada a finales de 2011 como 'start-up' de la Universidad Politécnica de Madrid, tras obtener el primer premio en la VII Competición para la Creación de Nuevas Compañías de Base Tecnológica, entre otras 260 propuestas. BioMet®Soft produce soluciones en Seguridad, Peritaje Forense, y Medicina […] con una tecnología patentada por la Universidad Politécnica de Madrid. Esta tecnología se basa en la extracción y caracterización de los perfiles de fonación de los locutores, y permite el establecimiento de diferentes niveles de inferencia. BioMet®Soft coopera con hospitales de la Comunicad de Madrid, Compañías de Consultoría, Gabinetes de Abogacía y Asesoría Legal, Entidades Financieras, Empresas de Generación de Energía, etc. Las aplicaciones desarrolladas por BioMet®Soft están en uso por los Cuerpos y Fuerzas de Seguridad del Estado". Actualmente ofrece 5 productos: BioMet®Phon, BioMet®Fore, BioMet®Sing, BioMet®Scie y BioMet®Ling (http://www.glottex.com/). Entre las cinco herramientas, se eligió BioMet®Phon por tratarse del software que más detalles ofrece sobre la evaluación de la fonación. Es un programa muy utilizado por terapeutas de la voz (otorrinos, foniatras, logopedas) para evaluar posibles patologías vocales en pacientes pero también está diseñado para el análisis detallado de la fonación no patológica, de la cualidad de voz en general.

Tabla 3.4: Nombre y definición de los 72 parámetros relativos al comportamiento glótico que estima el programa *BioMet®Phon*.

Parámetro	Significado
1. *Absolute Pitch*(Hz)	Es la frecuencia de un ciclo glótico, inversa del período fundamental. Es útil, junto con otros parámetros, para distinguir la voz femenina de la masculina.
2. *Absolute Normal Jitter* (%)	Es la variabilidad de la frecuencia fundamental de un ciclo al siguiente. El cálculo se basa en el valor inverso de la diferencia entre los períodos de dos ciclos consecutivos, dividido por el promedio aritmético de ambos. Sirve, junto con otros, para detectar inestabilidad en la fonación y ayuda a caracterizar la disfonía.
3. *Absolute Normal Area Shimmer* (%)	Diferencia entre las amplitudes de dos ciclos consecutivos de la onda glótica, partida por la media aritmética de ambas amplitudes. Sirve, junto con otros, para detectar inestabilidad en la fonación y ayuda a caracterizar la disfonía.
4. *Absolute Normal Minimum Sharpness*	Grado de afilamiento o agudeza del pico del *Maximum Flow Declination Rate* (MFDR): amplitud negativa del pico dividida por su anchura. Es válido, junto con otros, para detectar inestabilidad en la fonación y ayuda a caracterizar la disfonía. Es de utilidad, junto con otros parámetros, para caracterizar la disfonía y detectar emocionalidad en la voz.
5. *Noise-HarmonicRatio, NHR* (%)	Relación entre las energías de los componentes no armónicos y armónicos del espectro de potencia de la onda glótica.
6. *Muc./AvAc. Energy, MAE*	Relación entre las energías de la onda mucosa/onda glótica. Ayuda a caracterizar la disfonía y es de utilidad, junto con otros parámetros, en la detección de posibles alteraciones neurológicas.
7. *MWC Cepstral 1*	Primer coeficiente cepstral de la onda glótica.
8. *MWC Cepstral* 2	Segundo coeficiente cepstral de la onda glótica.
9. *MWC Cepstral 3*	Tercer coeficiente cepstral de la onda glótica.
10. *MWC Cepstral 4*	Cuarto coeficiente cepstral de la onda glótica.
11. *MWC Cepstral 5*	Quinto coeficiente cepstral de la onda glótica.
12. *MWC Cepstral 6*	Sexto coeficiente cepstral de la onda glótica.
13. *MWC Cepstral 7*	Séptimo coeficiente cepstral de la onda glótica.
14. *MWC Cepstral 8*	Octavo coeficiente cepstral de la onda glótica.
15. *MWC Cepstral 9*	Noveno coeficiente cepstral de la onda glótica.
16. *MWC Cepstral 10*	Décimo coeficiente cepstral de la onda glótica.
17. *MWC Cepstral 11*	Decimoprimer coeficiente cepstral de la onda glótica.
18. *MWC Cepstral 12*	Decimosegundo coeficiente cepstral de la onda glótica.
19. *MWC Cepstral 13*	Decimotercer coeficiente cepstral de la onda glótica.
20. *MWC Cepstral 14*	Decimocuarto coeficiente cepstral de la onda glótica.

21. *MW PSD 1st Max. ABS.*	*Power Spectral Density* (PSD). Primer valor máximo de la densidad espectral de la onda glótica.
22. *MW PSD 1st Min. rel.*	Primer valor mínimo de la densidad espectral de la onda glótica.
23. *MW PSD 2nd Max. rel.*	Segundo valor máximo de la densidad espectral de potencia de la onda glótica.
24. *MW PSD 2nd Min. rel.*	Segundo valor mínimo de la densidad espectral de potencia de la onda glótica.
25. *MW PSD 3nd Max. rel.*	Tercer valor máximo de la densidad espectral de potencia de la onda glótica.
26. *MW PSD End Val. rel.*	Valor de la densidad espectral de potencia de la onda glótica a una frecuencia de muestreo del 50%.
27. *MW PSD 1st Max. Pos. ABS.*	Frecuencia a la que se sitúa el primer valor máximo de la densidad espectral de potencia de la onda glótica.
28. *MW PSD 1st Min. Pos. rel.*	Frecuencia a la que se sitúa el primer valor mínimo de la densidad espectral de potencia de la onda glótica en relación con su primer valor máximo.
29. *MW PSD 2nd Max. Pos. rel.*	Frecuencia del segundo valor máximo de la densidad espectral de potencia de la fuente glótica en relación con su primer valor máximo.
30. *MW PSD 2nd Min. Pos. rel.*	Frecuencia relativa del segundo valor mínimo de la densidad espectral de potencia de la fuente glótica en relación con su primer valor máximo.
31. *MW PSD 3rd Max. Pos. rel.*	Frecuencia relativa del tercer valor máximo de la densidad espectral de potencia de la onda glótica en relación con su primer valor máximo.
32. *MW PSD End Val. Pos. rel.*	Frecuencia de la densidad espectral de potencia de la onda glótica a una frecuencia de muestreo del 50% en relación con su primer valor máximo.
33. *MW PSD 1st Min NSF*	Grado de afilamiento o agudeza del primer valle en forma de "V" en la envolvente espectral de la densidad de potencia de la onda glótica: amplitud negativa del pico dividida por su anchura.
34. *MW PSD 2nd Min NSF*	Grado de afilamiento o agudeza del segundo valle en forma de "V" en la envolvente espectral de la densidad espectral de potencia de la onda glótica: amplitud negativa del pico dividida por su anchura.
35. *Body Mass*	Masa dinámica correspondiente al cuerpo de cada cuerda vocal implicada en cada ciclo vibratorio glótico
36. *Body Losses*	Pérdida por fricción del cuerpo de cada pliegue vocal durante cada ciclo glótico.
37. *Body Stiffness*	Tensión que presenta el cuerpo de cada pliegue vocal durante cada ciclo vibratorio glótico.
38. *Body Mass Unbalance* (%)	Diferencia entre las medidas de las masas dinámicas del cuerpo de cada pliegue vocal tomadas en dos ciclos consecutivos, dividida por su media.
39. *Body Losses Unbalance*	Diferencia entre las pérdidas por fricción del cuerpo de cada pliegue medidas en dos ciclos consecutivos,

	dividida por su media.
40. *BodyStiffnessUnbalance*(%)	Diferencia entre las tensiones del cuerpo del pliegue medidas en dos ciclos vecinos, dividida por su media.
41. *Cover Mass*	Masa dinámica equivalente de la cubierta del pliegue vocal para cada ciclo glótico.
42. *Cove rLosses*	Pérdida por fricción de la cubierta del pliegue vocal para cada ciclo glótico.
43. *Cover Stiffness*	Tensión equivalente de la cubierta del pliegue vocal para cada ciclo glótico.
44. *Cover Mass Unbalance* (%)	Diferencia entre las masas dinámicas de la cubierta del pliegue medidas en dos ciclos vecinos, dividida por su media.
45. *Cover Losses Unbalance*	Diferencia entre las pérdidas por fricción de la cubierta del pliegue medidas en dos ciclos vecinos, dividida por su media.
46. *Cover Stiffness Unbalance* (%)	Diferencia entre las tensiones de la cubierta del pliegue vocal medidas en dos ciclos vecinos, dividida por su media.
47. *Rel. Recovery 1 Time*	Relación entre el instante del primer tiempo de recuperación y la duración total del ciclo glótico.
48. *Rel. Recovery 2 Time*	Relación entre el instante del segundo tiempo de recuperación y la duración total del ciclo glótico.
49. *Rel. Open 1 Time*	Relación entre el instante del primer tiempo de abertura y la duración total del ciclo glótico.
50. *Rel. Open 2 Time*	Relación entre el instante del segundo tiempo de abertura y la duración total del ciclo glótico.
51. *Rel. Max. Amplitude Time*	Relación entre el instante en que se alcanza la máxima amplitud de la onda glótica y la duración total del ciclo glótico.
52. *Rel. Recov. 1 Ampl.*	Relación entre la amplitud en el primer tiempo de recuperación y la amplitud pico a pico de la fuente glótica.
53. *Rel. Recov. 2 Ampl.*	Relación entre la amplitud en el segundo tiempo de recuperación y la amplitud pico a pico de la fuente glótica.
54. *Rel. Open 1 Ampl.*	Relación entre la amplitud en el primer tiempo de abertura y la amplitud pico a pico de la fuente glótica.
55. *Rel. Open 2 Ampl.*	Relación entre la amplitud en el segundo tiempo de abertura y la amplitud pico a pico de la fuente glótica.
56. *Rel. Stop Flow Time*	Relación entre el instante de mínimo flujo y la duración total del ciclo glótico.
57. *Rel. Start Flow Time*	Relación entre el instante de inicio del flujo y la duración total del ciclo glótico.
58. *Rel. Closing Time*	Relación entre el instante de máximo flujo y la duración total del ciclo glótico
59. *Val. Flow GAP*	Relación entre el escape de flujo durante la fase de contacto y el flujo total durante un ciclo glótico (medido sobre el flujo).
60. *Val. Contact GAP*	Relación entre el escape de flujo durante la fase de contacto y el flujo total durante un ciclo glótico

	(medido sobre la fuente glótica).
61. *Val. Adduction GAP*	Relación entre el flujo durante la fase de aducción y el flujo total durante un ciclo glótico.
62. *Val. Permanent GAP*	Relación entre el flujo durante la fase de recuperación y el flujo total durante un ciclo glótico.
63. *1st.Order Cyclic Coefficient*	Primer coeficiente PARCOR en el modelo equivalente autorregresivo de la tensión del cuerpo del pliegue vocal eliminando su media.
64. *2nd. Order Cyclic Coefficient*	Segundo coeficiente PARCOR en el modelo equivalente autorregresivo de la tensión del cuerpo del pliegue vocal eliminando su media.
65. *3rd. Order Cyclic Coefficient*	Tercer coeficiente PARCOR en el modelo equivalente autorregresivo de la tensión del cuerpo del pliegue vocal eliminando su media.
66. *Physiological Tremor Freq.*	Primer componente de la tensión del cuerpo del pliegue vocal eliminando su media. Suele distribuirse entre 2-4 Hz y se atribuye al temblor de carácter fisiológico.
67. *Physiological Tremor Ampl.*	Amplitud del primer componente de la tensión del cuerpo del pliegue, en % respecto al valor medio de la tensión del pliegue.
68. *Neurological Tremor Freq.*	Segundo componente de la tensión del cuerpo del pliegue vocal eliminando su media. Suele distribuirse entre 5-8 Hz y se atribuye al temblor de carácter neurológico.
69. *Neurological Tremor Ampl.*	Amplitud del primer componente de la tensión del cuerpo del pliegue, en % respecto al valor medio de la tensión del pliegue.
70. *Fluttering Tremor Freq.*	Tercer componente de la tensión del cuerpo del pliegue vocal eliminando su media. Suele distribuirse entre 9-12 Hz y se atribuye al temblor de carácter neurofisiológico.
71. *Fluttering Tremor Ampl.*	Amplitud del tercer componente de la tensión del cuerpo del pliegue, en % respecto al valor medio de la tensión del pliegue.
72. *Tremor amplitude (rMSA)*	Desviación estándar de la tensión del pliegue vocal eliminando su media.

Es posible distribuir todos los parámetros en 7 grupos diferentes (véanse BioMet®PhonUser's Manual, 2014; Gómez-Vilda, Fernández-Baillo, Rodellar-Biarge, Nieto-Lluis, Álvarez-Marquina, *et al*., 2009; Gómez-Vilda, Álvarez-Marquina, Tsanas, Lázaro-Carrascosa, Rodellar-Biarge, *et al*., 2016; Gómez-Vilda y Nieto-Lluis, 2015; y Palacios Alonso, 2018, pp. 68-73, entre otros):

Grupo A. *Parámetros de perturbación*: 1-6. Se trata de parámetros relativos a la f_0 y a las distorsiones o perturbaciones asociados a ella.
Grupo B. *Parámetros cepstrales*: 7-20. Forman parte de la firma biométrica del locutor en forma compacta o global, y junto con otros son útiles para la identificación y verificación del locutor.
Grupo C. *Parámetros de perfil espectral*: 21-34. Forman parte de la firma biométrica del locutor, responden al comportamiento normativo o no normativo de este, y, junto con otros, ayudan en la identificación y la verificación del hablante, así como verificación en la determinación de la presencia de disfonía de origen orgánico.
Grupo D. *Parámetros biomecánicos de la glotis*: 35-46. Son parámetros que responden a estimaciones biométricas de la masa, de la tensión y de las pérdidas de energía de los pliegues vocales. Constituyen un conjunto robusto de descriptores del funcionamiento mecánico de la glotis, y junto con otros ayudan a determinar los tipos de fonación.
Grupo E. *Parámetros de la onda glótica de base temporal*: 47-58. Se relacionan con los coeficientes temporales que presenta la onda glotal y con los valores relativos al cierre, abertura o escape de aire que permiten los pliegues. Constituyen un descriptor completo de los instantes temporales de interés del ciclo glótico (cierre, retorno, abertura).
Grupo F. *Parámetros de defecto de cierre glótico*: 59-62. Constituyen un descriptor de los defectos en el cierre, la aducción, la abducción o la separación permanente de los pliegues vocales y, junto con otros, sirven para caracterizar las imperfecciones detectadas en el ciclo glótico. Se calculan en términos de diferencias de áreas.
Grupo G. *Parámetros de temblor*: 63-72. Proporcionan información sobre la presencia de defectos o irregularidades en la actuación del sistema neuromotor vinculado al cierre glótico, puestos de manifiesto por la aparición de temblor en la voz (controlado o incontrolado).

De cada locutor se extrajeron valores correspondientes a los 72 parámetros a partir de vocales [a] tónicas (3 [a] tónicas x 5 repeticiones x 2 registros), lo cual arrojó un total de 30 mediciones por parámetro glotal por locutor (15 en voz modal y 15 en *falsetto*). Los resultados obtenidos así como su discusión se expondrán más adelante en los Capítulos 4 y 5.

3.3.2. Análisis del tempo

A pesar de que mediante el diseño experimental se intentó controlar la variable temporal del habla de cara a la discriminación de voces, se ha señalado previamente en la bibliografía relevante que los cambios de registro pueden desencadenar cambios en la velocidad de habla o en la velocidad de articulación de los locutores (véanse muy especialmente, Künzel, 2000 y Wagner y Köster, 1999). Asimismo, se ha demostrado

en varios trabajos que el factor temporal en un sentido amplio es una clave perceptiva relevante en el reconocimiento de hablantes pues contribuye en buena medida a la caracterización individual del habla (véanse por ejemplo los trabajos de Federico, Mori y Paoloni, 2005; Künzel, 1997; Leeman, Kolly y Dellwo, 2014). Por estas razones se ha decidido no modificar la velocidad para que fuera la misma en todas las frases. Mantener la velocidad real de cada frase permitirá observar su efecto en el cambio de registro y analizar luego la velocidad característica de cada locutor para poder determinar si existen diferencias significativas entre ellos y, de ser así, estudiar si esta variable ha podido influir en su posterior discriminación.

Para el estudio de las variables temporales del habla[32] se ha recomendado utilizar la sílaba como unidad de medida (véase por ejemplo, Tauroza y Allison, 1990 y para más detalles sobre las distintas unidades de medida propuestas véase Schwab, en prensa). Puesto que en el presente estudio se utiliza la misma frase, los locutores pronunciaron todas las sílabas y no realizaron pausas, en esta tesis se ha calculado únicamente la duración promedio de cada frase; todos los resultados se presentan detalladamente en el Capítulo 4. Por comodidad expositiva se hará referencia a los aspectos temporales con las expresiones *tempo*, *velocidad de elocución* y *velocidad de habla*, aunque todos los cálculos se han realizado sobre la duración de las frases. Para profundizar en las diferencias entre velocidad de habla y de articulación véanse por ejemplo, Gil Fernández, 2007 y Schwab (en prensa).

3.4.Metodología para la aplicación de los test perceptivos

3.4.1. Diseño del test perceptivo

Para evaluar hasta qué punto los oyentes pueden ser capaces de reconocer a la misma persona hablando con su voz normal y con su voz disimulada en *falsetto* se elaboró un test perceptivo de discriminación, el cual se aplicó, como se explicará más adelante en el Capítulo 5, a jueces españoles y a jueces italianos (napolitanos).

El tipo de prueba perceptiva escogido fue un test de discriminación igual-diferente, prueba en la que se busca que los oyentes decidan si los dos estímulos que se

[32] El *tempo* o velocidad de habla considera el tiempo de articulación y el tiempo de las pausas que el locutor realiza (véase por ejemplo, Schwab, en prensa).

les presentan son iguales o distintos. Como explica Marrero (2014), esta tarea puede llevarse a cabo de varias formas. En esta tesis se decidió aplicar el paradigma AX:

> Paradigma AX (el más frecuente), en el que al sujeto se le pregunta: ¿el estímulo X y el estímulo A son iguales o diferentes? La respuesta "diferentes" se incrementa en el momento en que la diferencia entre A y X empieza a ser perceptivamente relevante […] (p. 523).

En este test se ofrecen dos estímulos (modal-*falsetto*) al oyente y este debe contestar si cree que pertenecen al mismo hablante o a hablantes diferentes. Los estímulos se combinaron de tal manera que en cada juicio los oyentes tuvieran que identificar si ambas voces pertenecían al mismo locutor o a locutores distintos. El propósito último de este estudio es detectar las características vocales de los locutores que los hacen o no reconocibles cuando disimulan su voz en *falsetto* y no interesa saber si a un locutor se lo confunde con otro/s concreto/s. El test de discriminación pretende evaluar principalmente si a un mismo locutor se lo ha reconocido o no cuando los jueces lo hayan discriminado correctamente en los pares coincidentes, es decir, en los pares modal-*falsetto* de un mismo locutor. Los pares no coincidentes se incluyeron únicamente como distractores. De este modo, se asegura que si un juez acierta por encima del nivel del azar significa realmente que ha sido capaz de reconocer al locutor. Además, el número de pares coincidentes presentados a cada juez fue igual al número de pares no coincidentes para cumplir con la suposición señalada en la bibliografía (por ejemplo, McGuire, 2010) de que, en una tarea de percepción, los jueces esperan un mismo número de pares coincidentes y no coincidentes.

Para el estudio perceptivo se utilizaron las mismas frases que para el análisis acústico, frases que, como se explicó antes, fueron seleccionadas en función de la calidad de las emisiones. Por tanto, los mismos materiales sonoros que se utilizaron antes para el estudio de la producción fueron los que luego se sometieron a la diferenciación perceptiva por parte de los jueces[33].

[33]Recuérdese que aunque los informantes grabaron la frase "Dica *dadiva* adagio e vada a casa"; se decidió utilizar solamente parte de ella, en concreto, el fragmento: "Dica *dadiva* adagio e vada" de forma de evitar la porción final del enunciado que presentaba diferencias prosódicas entre los distintos sujetos y, por tanto, podría constituir una pista importante para la discriminación de voces.

Todos los audios fueron además normalizados a -1 dB a través de un editor de audio (*Audacity-win-2-0-3*) para ajustar las grabaciones al mismo nivel de intensidad y evitar que los jueces se vieran influenciados por los distintos volúmenes.

El diseño del test consistió en 5 frases x 2 modalidades (modal-*falsetto*) x 6 locutores x 2 condiciones (parejas de locutores iguales— parejas de locutores distintos), lo cual resultó en 60 pares de voces y 120 estímulos. Las parejas de locutores iguales son pares coincidentes, donde cada miembro es el mismo locutor hablando en voz modal y en *falsetto*; las parejas de locutores distintos no coinciden, cada miembro proviene de un locutor diferente.

El orden de presentación de los pares fue aleatorizado y se mantuvo para todos los jueces. El intervalo entre el primer y segundo estímulo de cada par (*Inter-Stimulus Interval*) fue de 0,5 segundos y, entre los pares (*Inter- Trial Interval*), fue de 1 segundo, mientras que el tiempo de respuesta de cada juez fue libre. Cada oyente evaluó a todos los locutores y, como ya se ha dicho, un total de 60 pares de estímulos, de los cuales la mitad estaban formados por estímulos del mismo hablante y la otra mitad por estímulos de hablantes diferentes.

3.4.2. Plataforma utilizada para la elaboración del test perceptivo

La creación de la prueba perceptiva así como su ejecución se llevaron a cabo mediante la herramienta FOLERPA, *Ferramenta On-Line para ExpeRimentación PerceptivA* (Fernández Rei, 2014), desarrollada por el Instituto da Lingua Galega (ILG) de la Universidad de Santiago de Compostela (USC, España). Se trata de una herramienta en línea diseñada para la experimentación perceptiva en lingüística que permite tanto crear como desarrollar test de percepción. Es una herramienta de fácil uso tanto para los investigadores que crean sus test como para los jueces que los realizan; es, además, de acceso libre y gratuito. Ha sido previamente utilizada con muy buenos resultados por varios investigadores y jueces (Aguete Cajiao, Fernández Rei y Osório Peláez, 2016).

Como se ve en las Figuras 3.1 y 3.2, dicha plataforma permite dos perfiles de acceso, el de juez (Figura 3.1) y el de investigador (Figura 3.2), posibilitando, por un lado, el acceso de un mismo juez a diferentes test y, por otro, el acceso de cada investigador a los distintos proyectos que genere.

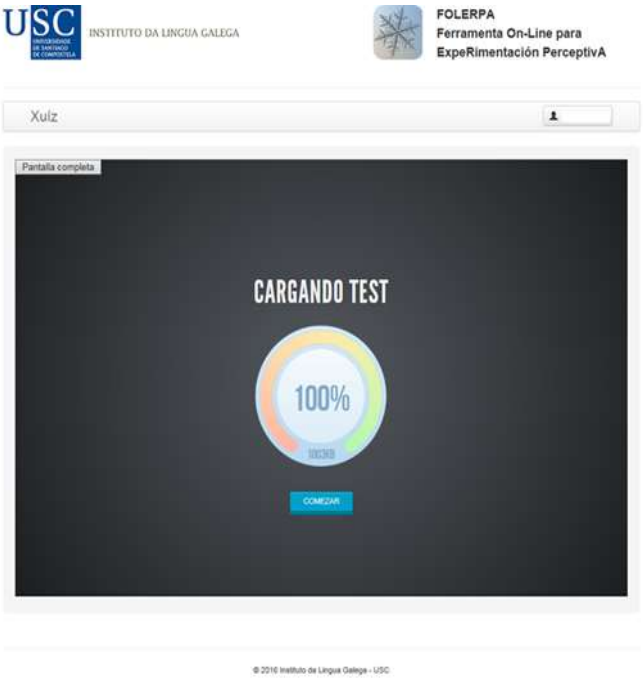


Figura 3.1. Pantalla de inicio de FOLERPA desde el perfil de juez.

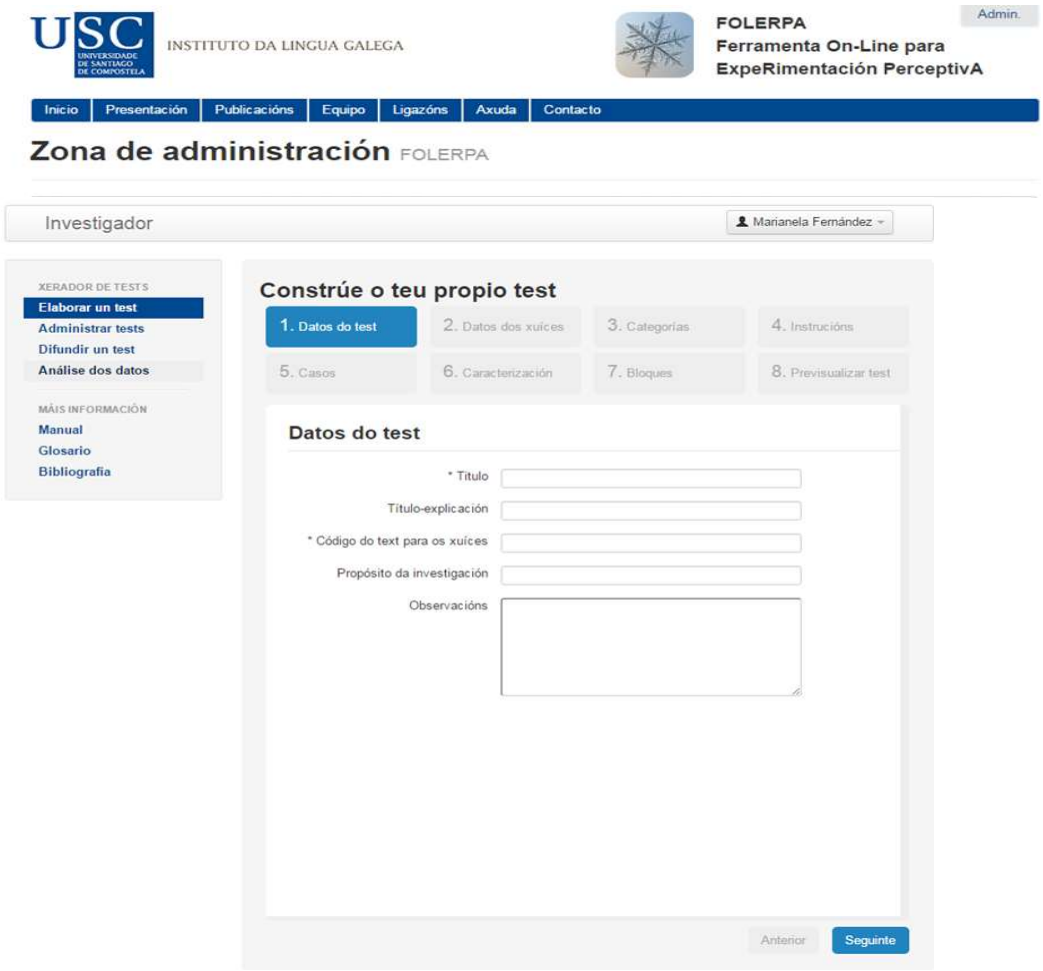


Figura 3.2. Pantalla de inicio de FOLERPA desde el perfil de investigador.

Desde el perfil de investigador se pueden ver las opciones de construcción y generación de los test. La plataforma permite confeccionar distintos tipos de test, con diseños de discriminación, identificación y mixtos. Es posible incrustar audios en formatos de alta calidad, como por ejemplo *wav*. Además de esto, la herramienta permite al investigador diseñar el cuestionario de datos acerca de los perfiles de los jueces (Aguete Cajiao *et al*., 2016; Fernández Rei y Moutinho, 2012).

En cuanto a la distribución, cada investigador puede dar acceso a cada uno de los test que haya generado a las personas que desee por medio de un código o a través de una URL. Esto ofrece a los jueces la posibilidad de realizar las pruebas en línea. Por último, cabe señalar que FOLERPA ofrece una estadística básica de los resultados obtenidos que el usuario puede descargarse inmediatamente después de que los jueces hayan realizado el test (Aguete Cajiao *et al*. 2016; Fernández Rei y Moutinho, 2012).

Como aparece recogido en el *Manual de uso* de la herramienta, una vez situados en el módulo *Elaborar un test,* el investigador tiene acceso a ocho formularios que tendrá que ir completando por pasos para construir la prueba perceptiva, (véase *supra*, la Figura 3.2). A continuación se detallan las principales características de estos ocho formularios según la información que aparece en el *Manual de uso* al tiempo que se detallan las características del test perceptivo realizado en esta tesis.

1. *Datos do test*: en este primer bloque se han de proporcionar los datos básicos de la prueba, como título y propósito, y se asigna un código, único para cada test.

2. *Datos dos xuíces*: aquí, el investigador escoge los datos que quiere obtener de los jueces que participarán en el experimento para que estos los suministren luego y pueda guardarse automáticamente dicha información. Para los test de discriminación de voces de esta tesis los jueces tuvieron que precisar su sexo, su edad, su nivel educativo, cuál era su lengua materna (español/italiano), si hablaban español/italiano como segunda lengua, si estaban acostumbrados a escuchar o evaluar voces por su trabajo o formación y si tenían problemas de audición.

3. *Categorías:* se introducen las categorías que estarán asociadas a los estímulos y se determinan los valores posibles dentro de cada una de ellas. En el presente trabajo, a los estímulos se asociaron 3 categorías: *Locutor* (con valores 1, 2, 3, 4, 5 y 6), *Registro* (modal y *falsetto*) y *Frase* (1, 2, 3, 4 y 5).

4. *Instrucións*: se redactan las instrucciones que los jueces leerán para comprender la tarea. Véase más adelante la Figura 3.6 para conocer las instrucciones que se proporcionaron a los participantes inmediatamente antes de iniciar la evaluación de las voces.

5. *Casos*: en este bloque el investigador selecciona el número de estímulos que va a aparecer en cada caso –en este test fueron pares AX, es decir, 2 estímulos por vez, lo que hace un total de 60 pares y 120 estímulos–, el número de opciones de respuesta –dos: SÍ/NO-, el orden en el que se presentarán —aleatorio y fijo para todos los jueces— y el tiempo de pausa entre las repeticiones —en el test, un mismo par se reproducía dos veces, uno inmediatamente después del otro sin mediar pausa entre ellos—; el tiempo entre el primer y el segundo estímulo de cada par fue de 0,5 segundos y entre cada uno de los 60 pares, de 1 segundo.

6. *Caracterización:* desde este formulario se pueden caracterizar todos los casos del test. Además de escribir las preguntas que se formularán a los jueces y las respuestas posibles, se suben a la plataforma los audios que se emplearán como estímulos. En los test utilizados en esta tesis, la pregunta fue siempre la misma: *¿Las dos voces pertenecen a la misma persona?*, y las opciones de respuesta también: *Sí/No*. Se fueron cargando uno a uno los audios en formato *wav*. según el orden establecido y para cada par se seleccionó manualmente la respuesta correcta.

7. *Bloques*: en este bloque el investigador determina el número de bloques en el que desea distribuir los casos del test, si quiere que esos bloques se repitan o no, el texto que presentará entre los bloques y la pausa entre ellos. La elección fue, como se dijo antes, dividir los 60 pares en dos bloques de 30 y

no preestablecer ningún tiempo de pausa entre ellos, sino dejar a los propios jueces la decisión sobre el tiempo de espera para continuar con la realización del test.

8. *Previsualizar test*: con esta opción se permite al investigador ver cómo ha quedado el test antes de que lo realicen los jueces y, además, corregir cualquier error que se pudiera detectar. Cuando se comprueba que todo es correcto, se guarda la prueba y se obtiene un esquema con toda la información relativa al test. El resumen de la información de este test de discriminación de voces se ofrece en la Figura 3.3.

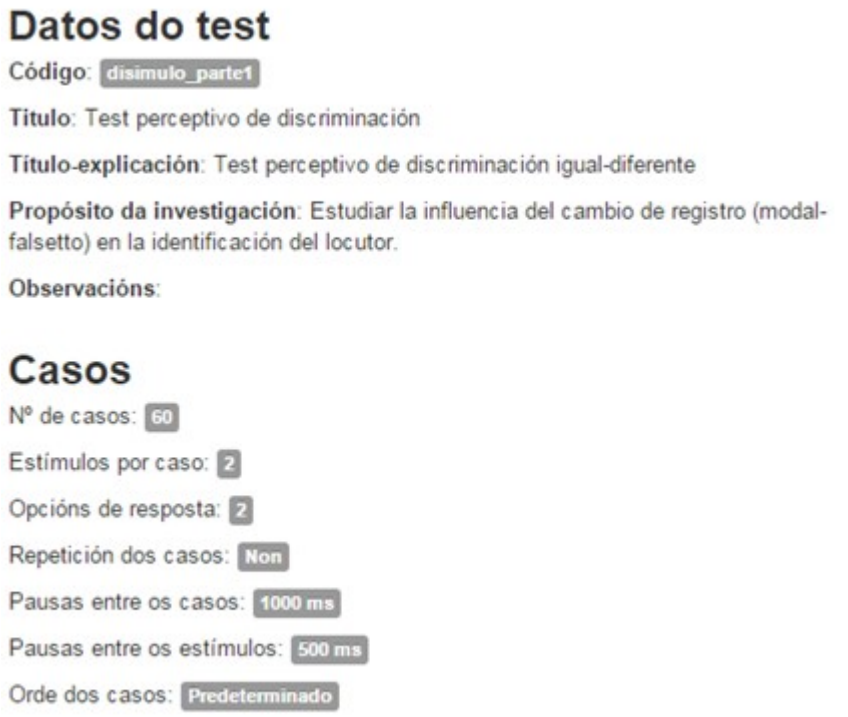


*Figura 3.3.*Esquema con toda la información del test perceptivo de discriminación de voces.

3.4.3. Jueces y ejecución del experimento (primer test)

Antes de comenzar con la tarea y por medio de un formulario se volvió a solicitar a los jueces la siguiente información: sexo, edad, nivel de educación, lengua materna e idiomas conocidos. También se les preguntó si, debido a su formación o a su experiencia laboral, tenían algún tipo de entrenamiento auditivo en la escucha y valoración de voces. Asimismo, se descartó que tuvieran problemas de audición preguntándoles expresamente sobre esto. Esta información se solicitó tanto en un formulario impreso en papel como en el propio test, de forma más reducida, como se muestra en la Figura 3.4.

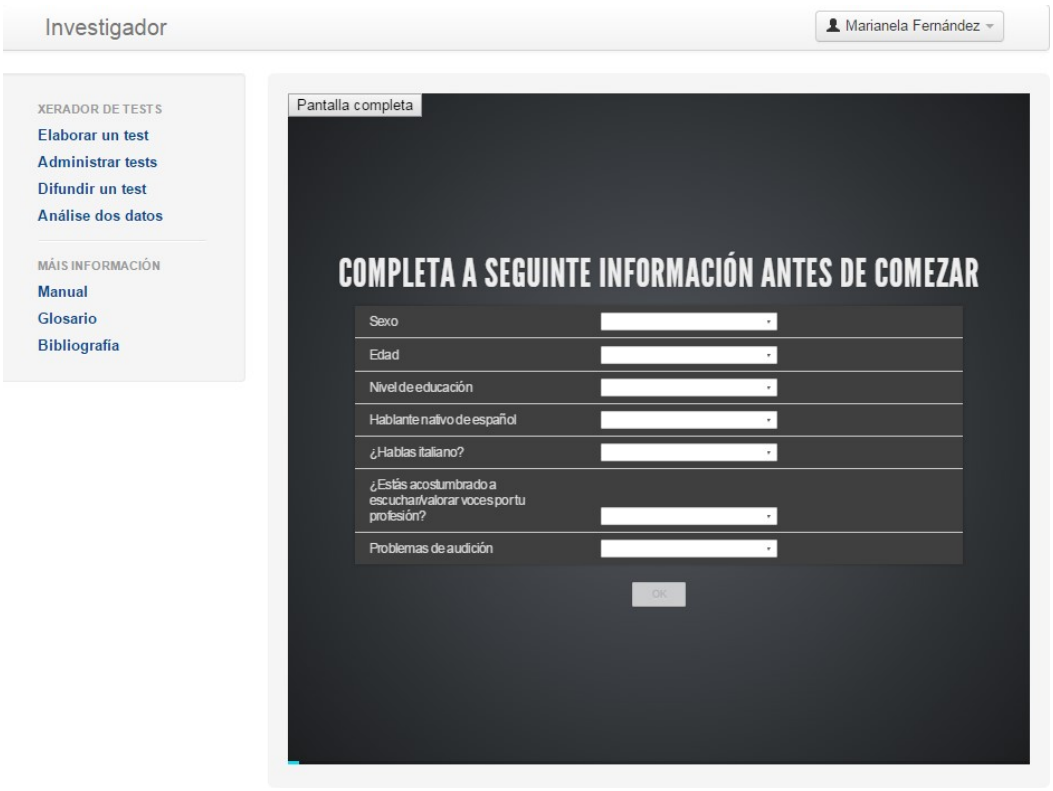


Figura 3.4. Formulario digital que los jueces debían completar antes de comenzar con el test.

Participaron en el primer experimento perceptivo 60 jueces españoles (53 mujeres y 7 hombres, con edades comprendidas entre los 18 y los 25 años (media de edad: 20,58)), universitarios, estudiantes de primer año de la Licenciatura en Logopedia, y hablantes nativos de español. Ninguno estaba acostumbrado a evaluar voces y no presentaban problemas de audición. Los jueces no conocían a ninguno de los informantes grabados.

El experimento tuvo lugar en un aula de informática de la Facultad de Psicología de la Universidad Complutense de Madrid, ubicada en el campus de Somosaguas, Madrid. Se realizó en dos mañanas diferentes, los días 20 y 21 de abril de 2015. La participación de todos los jueces fue voluntaria y se les informó, con anterioridad a la realización del experimento, de que las personas que tuvieran mayor número de aciertos recibirían una retribución económica en compensación, que fue de 50 euros por persona. Como se ve en la Figura 3.5., cada participante disponía de un ordenador y el Laboratorio de Fonética del CSIC proporcionó los auriculares necesarios.

Figura 3.5. Fotografía tomada durante el desarrollo del experimento en la Facultad de Psicología de la Universidad Complutense de Madrid.

Para tener mayor seguridad de que los jueces comprenderían de forma correcta la tarea que tendrían que realizar, y para que estuvieran familiarizados con la interfaz de FOLERPA y con el tipo de estímulos que oirían, se seleccionaron aleatoriamente cinco pares de voces como serie de práctica. Luego de este breve entrenamiento se comenzó con el test propiamente dicho. Seguidamente a las instrucciones (Figura 3.6), el programa reproduce cada par de estímulos con una repetición y a continuación ofrece al juez dos opciones de respuesta: ambos estímulos pertenecen al mismo locutor o a locutores diferentes (véase la Figura 3.7).

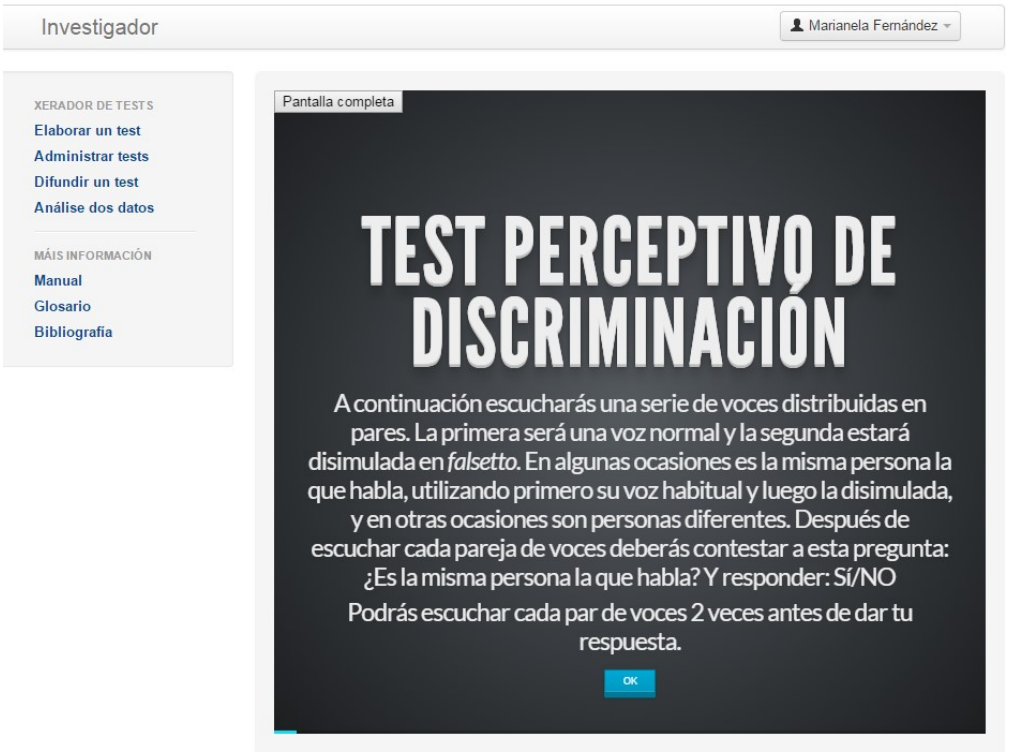


Figura 3.6. Pantalla en la que se resumen las instrucciones básicas para la realización del test perceptivo.

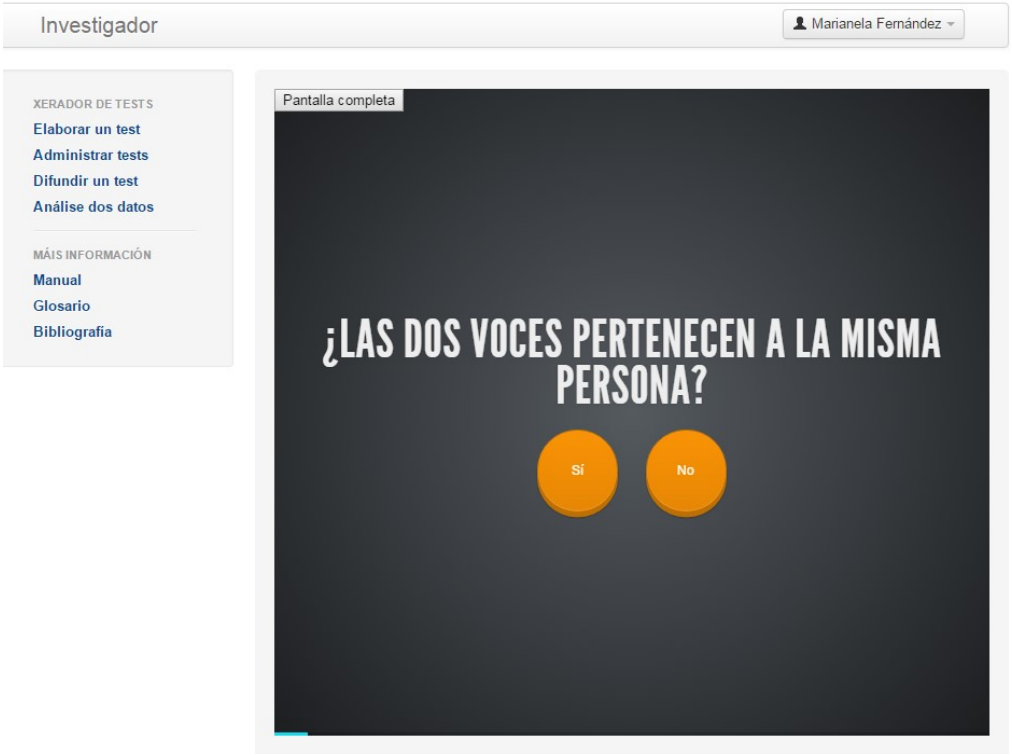


Figura 3.7. Imagen de la pregunta y de las opciones de respuesta que se ofrecían a los jueces luego de la escucha de cada par de voces (2 veces) para su valoración.

Una vez que el oyente da su respuesta, se reproduce el siguiente par de voces. Recuérdese que los 60 pares se distribuyeron en dos bloques de 30, separados por una pausa sin tiempo fijo, y que cada participante decidía el momento en que quería continuar evaluando los 30 pares restantes. Como el experimento se realizó con los participantes sentados en un aula, se pudo comprobar que las pausas de descanso fueron de 5 minutos, aproximadamente. El tiempo total del experimento fue de alrededor de 20-30 minutos por participante. Los resultados obtenidos de este experimento se analizarán y discutirán detenidamente en los Capítulos 5 y 6 de esta tesis.

3.4.4. Jueces y ejecución del experimento (segundo test)

El mismo test de discriminación se realizó con oyentes nativos de italiano (italiano regional de Nápoles). Se aplicó exactamente el mismo procedimiento y la única modificación consistió en la traducción al italiano del texto explicativo del test.

Durante las mañanas y tardes de los días 21 y 22 de abril de 2016, ochenta (80) voluntarios (58 mujeres y 22 hombres) realizaron el experimento en la *Facoltá di Filosofia e Lettere* de la *Universitá Degli Studi di Napoli Federico II* (Nápoles, Italia). Todos los participantes eran napolitanos, estudiantes del primer curso de licenciatura (*Laurea in Lingue, culture e letterature moderne europee*). Como en el caso del test 1, ninguno conocía a los locutores cuyas voces habrían de evaluar, no habían tenido experiencias previas en valoración o evaluación de voces y, no presentaban, lógicamente, problemas de audición. Del mismo modo que con el test realizado a hispanohablantes, los oyentes realizaron el experimento en un aula de la universidad, en grupos de 20, aproximadamente, y utilizaron los mismos auriculares que el grupo del test 1, como se aprecia en las Figuras 3.8 y 3.9.

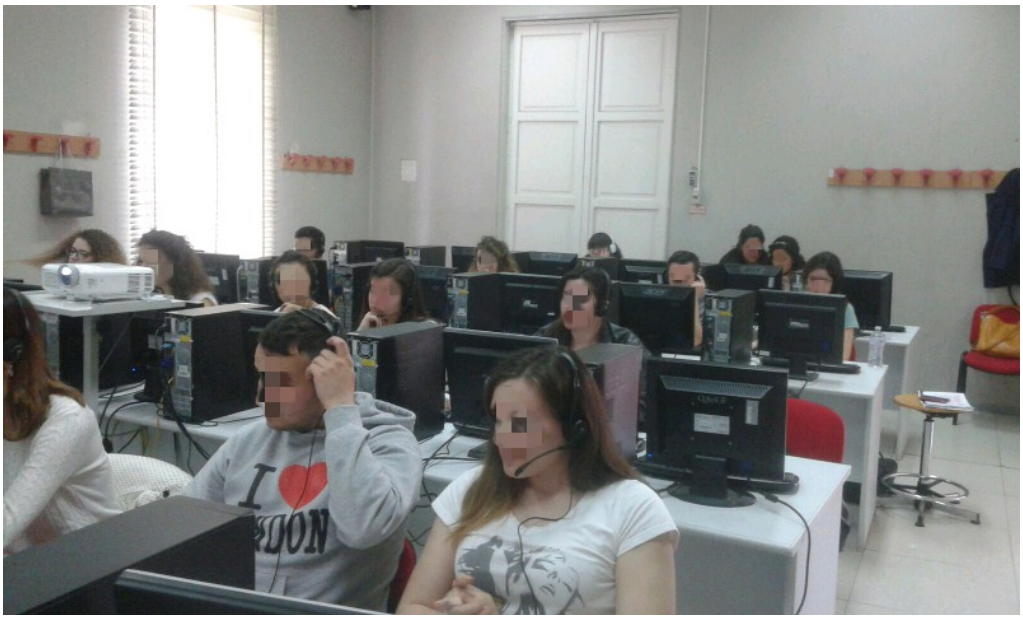


*Figura 3.8.*Desarrollo del test perceptivo 2, realizado en la *Universitá Degli Studi di Napoli Federico II*.

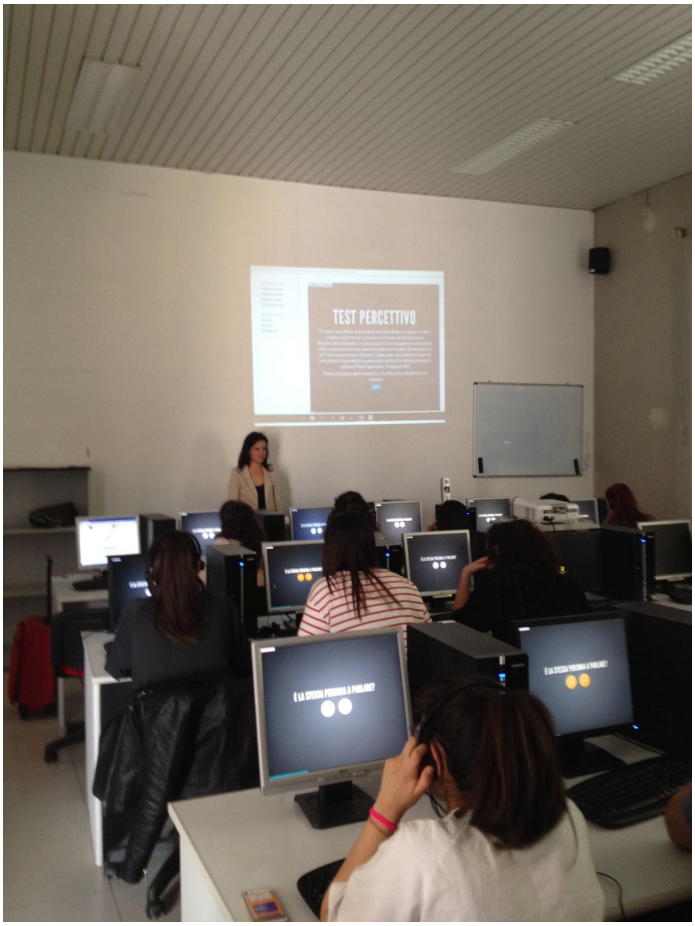


Figura 3.9. Fotografía tomada durante el desarrollo del test perceptivo 2, realizado en la *Universitá Degli Studi di Napoli Federico II*.

Capítulo 4
Resultados del análisis de la producción de la voz y del habla

4. Resultados del análisis de la producción de la voz y del habla

En este capítulo se presentan los datos extraídos del análisis de la producción de la voz y del habla y se adelanta una interpretación de los resultados obtenidos. En primer término, se reportan los valores relativos al comportamiento de la f_0 en el registro modal y en *falsetto* (§4.1.1). A continuación, se exponen los resultados derivados del estudio de la producción de voz en el sentido más restrictivo de este último término, esto es, el análisis del comportamiento laríngeo. Para realizar el análisis laríngeo se llevaron a cabo dos estudios. Por un lado, se examinó el comportamiento de los 72 rasgos glotales que permite analizar el programa *BioMet®Phon* (§4.1.2), y, a continuación, se realizó un *Análisis por Componentes Principales* (§4.1.3). Luego de los análisis de voz se presentan y discuten los resultados derivados del estudio de la velocidad de habla (§4.2). El capítulo se cierra con algunas consideraciones sobre los seis locutores, en cuanto a aspectos segmentales y prosódicos del habla (§4.3).

4.1. Análisis de la voz

4.1.1. Comportamiento de la f_0 en el cambio de registro (modal-falsetto)

Se han calculado los valores máximos, mínimos y medios de la f_0 alcanzados en los registros modal y *falsetto* para cada locutor y también el promedio general obtenido en los 6 locutores. La Tabla 4.1 contiene los datos obtenidos para el registro de voz modal y la Tabla 4.2, para el registro de *falsetto*.

Tabla 4.1. Valores de la frecuencia fundamental (máximo, media, desviación típica y mínimo) expresados en hercios para cada locutor y el promedio alcanzado por todos los locutores en voz modal.

	Registro Modal						
	L1	L2	L3	L4	L5	L6	Promedio
Máximo	150.1	119.1	120.4	117.2	162.2	153.5	137.1
Media	129.7	110.1	103.5	110.4	121.8	124.1	116.6
Desviación Típica	12.9	8.3	9.6	6.0	23.4	15.9	12.7
Mínimo	113.8	96.3	90.2	100.6	94.4	108.6	100.7

Tabla 4.2. Valores de la frecuencia fundamental (máximo, media, desviación típica y mínimo) expresados en hercios para cada locutor y el promedio alcanzado por todos los locutores en el registro de *falsetto*.

	Registro *Falsetto*						
	L1	L2	L3	L4	L5	L6	Promedio
Máximo	355.9	374.3	326.9	433.7	441.4	519.7	408.6
Media	306.4	302.8	284.1	378.3	366.3	336.4	329.0
Desviación Típica	49.4	63.7	42.2	30.1	52.4	123.2	60.2
Mínimo	145.7	172.4	153.0	319.1	281.7	218.6	216.1

Asimismo, se calcularon las distancias entre el registro modal y *falsetto* para observar el aumento del tono alcanzado por cada locutor así como el promedio global de todos ellos. En la Tabla 4.3 se reportan esas diferencias expresadas en hercios y sus equivalencias en octavas y *cents*, las cuales fueron obtenidas con un *script* para *Praat*.

Tabla 4.3. Diferencias alcanzadas en los valores de la frecuencia fundamental entre el registro de *falsetto* y la voz modal para cada locutor y el promedio obtenido de todos los locutores. Los valores se expresan en hercios, octavas y *cents*, para interpretar más fácilmente los decimales de las octavas[34].

	Falsetto - Modal						
	L1	L2	L3	L4	L5	L6	Promedio
Hercios	176.6	192.6	180.6	267.8	244.5	212.2	212.4
Octavas	1.24	1.46	1.46	1.78	1.59	1.44	1.50
Cents	1488	1751	1748	2132	1906	1726	1792

De los datos se desprende que los valores medios, considerando todos los locutores, fluctuaron entre 110 y 120 hercios en el registro de voz modal, algo esperable para voces masculinas. En el *falsetto*, la media se ubicó en torno a los 300 hercios, valores muy altos de f_0. Al observar el promedio de la desviación estándar se aprecia que en el *falsetto* aumenta también la variación, la dispersión es muy grande (60.2 Hz) si la comparamos con la observada en la voz modal (12.7 Hz). De promedio, el punto más bajo registrado en el *falsetto* fue de 216 Hz, el doble del obtenido para el modal (100.7 Hz); y el punto máximo alcanzó los 408.6 Hz en el *falsetto* contra 137.1 Hz en el registro modal.

[34]Recuérdense las siguientes equivalencias: 100 *cents*=1 semitono; 1200 *cents*= 1 octava; 12 semitonos=1 octava.

El incremento promedio entre la frecuencia fundamental de voz modal a *falsetto* es de una octava y media para todos los locutores. El Locutor 4 es el que más sube su f_0, 1 octava y 8 semitonos; y el que menos sube es el Locutor 1, consiguiendo aumentar en 1 octava y 2 semitonos su f_0 natural.

4.1.2. Análisis del comportamiento laríngeo con BioMet®Phon de *BiometroSoft®*

El análisis de los rasgos glotales de la voz con *BioMet®Phon*[35] ha sido laborioso y se ha realizado en varias fases. Para facilitar su comprensión se irá describiendo cada una de ellas y ejemplificando con los resultados extraídos del Locutor 1.

El primer paso consistió en calcular los valores medios correspondientes a los 72 parámetros a partir de una muestra de 30 vocales [a] tónicas por locutor, 15 producidas con voz modal y otras 15 producidas con voz de *falsetto*. Es decir, para cada locutor se midió 30 veces el valor de cada uno de los 72 parámetros (3 [a] tónicas x 5 frases x 2 registros). Fueron analizadas un total de 180 muestras de vocales [a] tónicas, extraídas, como se ha dicho antes, de la lectura de frases. De acuerdo con el *Manual del Usuario* de *BioMet®Phon* (www.biometrosoft.com), las muestras de voz deben tener un mínimo de 50 ms en el caso de una voz femenina típica (f_0=200 Hz) y de 100 ms en el de una voz típica masculina (f_0= 100 Hz). En el corpus, las vocales extraídas del registro de *falsetto* (f_0 media: 329 Hz) no presentaron problemas; sin embargo, para analizar las vocales de duración menor a 50 ms en el registro modal (f_0 media: 116.6 Hz) fue necesario modificar la configuración del programa para obtener una versión optimizada que permitiera analizar también vocales más breves, y con esta nueva configuración se analizaron finalmente todas las vocales de todos los locutores. En la Tabla 4.4 se ejemplifica, con algunos parámetros glotales, la forma en la que fueron extraídos los valores para el Locutor 1. A continuación, se realizó un contraste de medias (mediante la prueba *t-Student* con corrección de Bonferroni) entre la voz modal y el registro de *falsetto* para cada parámetro y cada locutor a fin de observar qué parámetros cambiaban significativamente de un registro a otro, tal y como se muestra en la Tabla 4.5.

Tabla 4.4: Ejemplo del fichero de datos (fragmento) obtenido al ejecutar *BioMet®Phon*. Nota: Se muestran los valores estimados para 14 parámetros glotales. Cada casilla corresponde a la media de los valores de 3 ejemplares de la vocal [a] para cada una de las 5 frases o repeticiones en cada registro (M: Modal y F: *Falsetto*) para el Locutor 1.

[35] Unos primeros resultados se ofrecieron en Fernández Trinidad y Rojo (2018).

Locutor	Registro	Frase	1. Absolute Pitch	2. Abs. Norm. Jitter	3. Abs. Norm. Ar. Shimmer	4. Abs. Norm. Min. Sharp.	5. Noise-Harm. Ratio (NHR)	6. Muc./AvAc. Energy (MAE)	7. MWC Cepstral 1
1	M	1	131.4179076	0.008786903	0.020931344	1.031557079	0.591286754	0.18324892	2732.674504
1	M	2	123.9153036	0.013567945	0.013351234	1.028981529	0.57759058	0.193696281	3661.235103
1	M	3	122.8024562	0.016999049	0.020675453	1.039789238	0.561828162	0.202057839	3895.901204
1	M	4	125.2189323	0.012726745	0.008300741	1.002789614	0.556561609	0.176864677	3646.621404
1	M	5	127.6300938	0.012465245	0.021662581	1.02662266	0.516157418	0.193586203	3603.551033
1	F	1	307.332851	0.015192816	0.02569918	0.991365499	0.712227033	0.161624987	1161.410133
1	F	2	306.9569739	0.034557918	0.01337262	1.005511142	0.706487969	0.062171645	1262.937981
1	F	3	290.8610461	0.070598517	0.028862215	1.003087307	0.733193103	0.446923057	1291.843401
1	F	4	317.9310123	0.054560772	0.040663973	0.977945702	0.652082783	0.421568147	1590.376441
1	F	5	291.8450665	0.026264375	0.013949793	0.994097843	0.646480332	0.05552741	1698.353271

Locutor	Registro	Frase	35. Body Mass	36. Body Losses	37. Body Stiffness	38. Body Mass Unbalance	39. Body Losses Unbalance	40. Body Stiffness Unbalance	41. Cover Mass
1	M	1	0.01962397	4.501985976	13388.28885	0.004192586	0.011674567	0.019971135	0.008062393
1	M	2	0.021686296	4.333451473	13163.75791	0.01952052	0.012916241	0.040177945	0.008986101
1	M	3	0.022845697	4.275694602	13644.94321	0.028477456	0.015639121	0.054394175	0.008167511
1	M	4	0.02149072	4.337010744	13323.93414	0.024114604	0.016766537	0.050616988	0.008367989
1	M	5	0.019911871	4.484066158	12818.27068	0.010401314	0.017966291	0.034998643	0.007933957
1	F	1	0.008174462	4.437646419	30495.27682	0.012310703	0.020826076	0.039731859	0.004282817
1	F	2	0.009534066	4.293436137	35651.67221	0.076698131	0.040255505	0.143674343	0.005550302
1	F	3	0.018873399	3.875121343	64028.22493	0.211135077	0.033966722	0.258106159	0.003132836
1	F	4	0.017620659	3.856134303	71216.67006	0.157515343	0.029518203	0.227458791	0.003377986
1	F	5	0.008650289	4.437630743	29107.20936	0.017576696	0.035937626	0.067509374	0.00536925

Tabla 4.5: Resultado del contraste de medias (prueba *t-Student* con corrección de Bonferroni) entre la voz modal y el registro de *falsetto* para los 72 parámetros en el Locutor 1.Los valores de la misma fila y subtabla que no comparten el mismo subíndice son significativamente diferentes en p<,05. Las pruebas asumen varianzas iguales. En la última columna se ha señalizado con un aspa (X) las variables que resultaron ser significativamente diferentes entre el registro modal y el *falsetto* para el Locutor 1 (52 parámetros).

Locutor		**Parámetros glotales (72)**	***Falsetto***	**Modal**	
1	1	Absolute_Pitch	306,4684905196179400$_a$	129,7938939924826500$_b$	X
	2	Abs._Norm._Jitter	,0332953070888395$_a$	,0161800694187247$_b$	X
	3	Abs._Norm._Ar._Shimmer	,0315022763703621$_a$	,0286311167275837$_a$	
	4	Abs._Norm._Min._Sharp	,9672883800055283$_a$	1,0116042017527970$_a$	
	5	Noise_Harm._Ratio_NHR	,6817603966287631$_a$	,5452782600389980$_b$	X
	6	Muc._AvAc._Energy_MAE	,1794852998369246$_a$	,1849490893388358$_a$	
	7	MWC_Cepstral_1	1412,4754358902420000$_a$	3582,3658117449836000$_b$	X
	8	MWC_Cepstral_2	796,8972235479860000$_a$	1690,0723138571325000$_b$	X
	9	MWC_Cepstral_3	354,5782365506592000$_a$	554,5801079267541000$_b$	X
	10	MWC_Cepstral_4	210,9358574349027700$_a$	371,1270973090330500$_b$	X
	11	MWC_Cepstral_5	149,0699499888258700$_a$	344,1610406419228500$_b$	X
	12	MWC_Cepstral_6	185,2135872195199500$_a$	365,3628602638067000$_b$	X
	13	MWC_Cepstral_7	72,0165257774332400$_a$	203,7862050836012000$_b$	X
	14	MWC_Cepstral_8	79,5414612721953800$_a$	179,5525938269833500$_b$	X
	15	MWC_Cepstral_9	83,4517338717330400$_a$	112,7725427279555800$_a$	
	16	MWC_Cepstral_10	71,2293351613782000$_a$	192,6910518606639500$_b$	X
	17	MWC_Cepstral_11	56,8575097291485250$_a$	122,8164716420223100$_b$	X
	18	MWC_Cepstral_12	63,1975852374604160$_a$	206,1278809049744000$_b$	X
	19	MWC_Cepstral_13	44,4450870610402300$_a$	121,0305944171278700$_b$	X
	20	MWC_Cepstral_14	61,3539965713912640$_a$	96,0581785033443700$_b$	X
	21	MW_PSD_1st_Max._ABS	18,3616265216441100$_a$	28,8941061083753220$_b$	X

	22	MW_PSD_1st_Min._rel	-15,0521288433943600$_a$	-15,5037395219282250$_a$	
	23	MW_PSD_2nd_Max._rel	-10,2435318345826970$_a$	-12,5360910478722080$_a$	
	24	MW_PSD_2nd_Min._rel	-24,6155105417084240$_a$	-23,9485118105886020$_a$	
	25	MW_PSD_3rd_Max._rel	-19,0731696814339240$_a$	-20,2278362892213720$_a$	
	26	MW_PSD_End_Val._rel	-49,1470867395111000$_a$	-55,3403995121950000$_b$	X
	27	MW_PSD_1st_Max._Pos._ABS.	2,1806030226618210$_a$	2,0166666666666666$_a$	
	28	MW_PSD_1st_Min._Pos._rel	2,0520585746659425$_a$	2,6337301587746030$_b$	X
	29	MW_PSD_2nd_Max._Pos._rel	2,5544258648884433$_a$	3,1648809524031742$_b$	X
	30	MW_PSD_2nd_Min._Pos._rel	3,5635121509761070$_a$	4,4650793650793640$_b$	X
	31	MW_PSD_3rd_Max._Pos._rel	4,1961407829153990$_a$	5,0531746031746040$_b$	X
	32	MW_PSD_End_Val._Pos._rel	18,8003291054085220$_a$	45,0666666666666700$_b$	X
	33	MW_PSD_1st_Min_NSF	12,1534473847153670$_a$	8,9805539424548420$_b$	X
	34	MW_PSD_2nd_Min_NSF	21,1155275387347000$_a$	16,9697526771338470$_b$	X
	35	Body_Mass	,0108750960734788$_a$	,0229448740065510$_b$	X
	36	Body_Losses	4,2716970784655240$_a$	4,3888187490151880$_b$	X
	37	Body_Stiffness	38742,5145021630450000$_a$	15979,0544927750420000$_b$	X
	38	Body_Mass_Unbalance	,0755519318518887$_a$	,0403730857551899$_b$	X
	39	Body_Losses_Unbalance	,0357121851170909$_a$	,0182043020293515$_b$	X
	40	Body_Stiffness_Unbalance	,1323810463157848$_a$	,0690562873521464$_b$	X
	41	Cover_Mass	,0055362131729821$_a$	,0082858464545360$_b$	X
	42	Cover_Losses	10,5953194212303340$_a$	5,9990980804353250$_b$	X
	43	Cover_Stiffness	35902,4730395951850000$_a$	8508,5550152277540000$_b$	X
	44	Cover_Mass_Unbalance	,1206199951400739$_a$	,0468502076503435$_b$	X
	45	Cover_Losses_Unbalance	,1318937047099384$_a$	,0365005371587430$_b$	X
	46	Cover_Stiffness_Unbalance	,1968122822351879$_a$	,0807698199283533$_b$	X
	47	Rel._Recov._1_Time	,1711304794292717$_a$	,0512741803969834$_b$	X

	48	Rel._Recov._2_Time	,2299419957291732$_a$	,0682720564839294$_b$	X
	49	Rel._Open_1_Time	,2176709842254143$_a$	,0553770346858424$_b$	X
	50	Rel._Open_2_Time	,3308387408618362$_a$	,0907290228702378$_b$	X
	51	Rel._Max._Ampl._Time	,4694661718279730$_a$	,1012928004508096$_b$	X
	52	Rel._Recov._1_Ampl.	,6559197588378310$_a$	,7488421813127789$_a$	
	53	Rel._Recov._2_Ampl.	,7215488480743163$_a$	,8746098415518885$_a$	
	54	Rel._Open_1_Ampl.	,6080875437984775$_a$	,7588271092314330$_a$	
	55	Rel._Open_2_Ampl.	,6822004363107516$_a$	,9206533824179172$_b$	X
	56	Rel._Stop_Flow_Time	,1772442576829232$_a$	,0442172765778271$_b$	X
	57	Rel._Start_Flow_Time	,1994012428162565$_a$	,0442172765778271$_b$	X
	58	Rel._Closing_Time	,7277406168453376$_a$	,3859081541640456$_b$	X
	59	Val._Flow_GAP	,0294970706666667$_a$	,0000000000000000$_a$	
	60	Val._Contact_GAP	,1149061050323521$_a$	,1201285757204555$_a$	
	61	Val._Adduction_GAP	,1538667042259109$_a$	,0855236606956617$_a$	
	62	Val._Permanent_GAP	,1728455604538757$_a$	,2861613698740055$_b$	X
	63	First._Order_Cyc._Coeff.	-,7749287421100824$_a$	-,3726113185702537$_b$	X
	64	Second._Order_Cyc._Coeff.	,2509627245004201$_a$	,4712086985353065$_a$	
	65	Third._Order_Cyc._Coeff.	,2831441874261534$_a$	,3532260390264593$_a$	
	66	PhysTremor_Frequency_Hz	5,8710028401944570$_a$	6,0124228242230210$_a$	
	67	PhysTremor_Est._Amplitude	2,0493833888453850$_a$	1,4581312319487196$_a$	
	68	NeurTremor_Frequency_Hz	9,8367674509252100$_a$	10,6650152586749880$_b$	X
	69	NeurTremor_Est._Amplitude	2,3063090731418283$_a$	1,3971788875101580$_b$	X
	70	FlutTremor_Frequency_Hz	13,5795636527283130$_a$	14,5255989115086220$_b$	X
	71	FlutTremor_Est._Amplitude	2,5372967075349275$_a$	1,4920727783468393$_b$	X
	72	Global_Tremor_rMSA	,0731861054139721$_a$	,0962850003513715$_a$	

Idéntico procedimiento se realizó, lógicamente, en el caso de los 5 locutores restantes, y el número total de parámetros alterados por cada uno de los 6 locutores, es decir, el número total de parámetros significativamente distintos entre voz modal y *falsetto,* se presenta resumido en la Tabla 4.6.

Tabla 4.6: Número total de parámetros glotales modificados por locutor, sobre un total de 72. Los locutores aparecen ordenados según la cantidad de parámetros que han modificado, de mayor a menor.

	Nº de parámetros modificados
Locutor 4	61
Locutor 5	58
Locutor 1	52
Locutor 3	46
Locutor 2	38
Locutor 6	35

La siguiente fase consistió en observar los parámetros glotales modificados por locutor como consecuencia del cambio de registro (ver el recuento de casos en la Tabla 4.7 y la proporción media en la Tabla 4.8 y la Figura 4.1), en función del grupo o familia de parámetros

Para ello se utilizó la organización explicada antes en §3.3.1 y que se recuerda a continuación:

- Grupo A: p.1-6: parámetros de perturbación.
- Grupo B: p.7-20: parámetros cepstrales.
- Grupo C: p. 21-34: parámetros de perfil espectral.
- Grupo D: p. 35-46: parámetros biomecánicos de la glotis.
- Grupo E: p. 47-58: parámetros de base temporal de la onda glótica.
- Grupo F: p. 59-62: parámetros de defecto de cierre glótico.
- Grupo G: p. 63-72: parámetros de temblor.

Tabla 4.7: Número de parámetros glotales que varió significativamente cada locutor al cambiar de registro (modal/*falsetto*) organizados por grupo de parámetros.

Grupo	Total de parámetros por grupo	L1	L2	L3	L4	L5	L6
A	6	3	4	3	4	6	3
B	14	13	12	10	14	14	14
C	14	9	8	11	12	11	6
D	12	12	7	7	12	10	6
E	12	9	4	8	12	11	3
F	4	1	2	2	3	2	2
G	10	5	1	5	4	4	1
Total por locutor	72	52	38	46	61	58	35

Tabla 4.8: Proporción (desviación típica y error típico) de parámetros glotales que varió significativamente cada locutor al cambiar de registro (modal/*falsetto*), organizados por grupo de parámetros.

Grupos de parámetros	Proporción	Desviación típica	Error típico
A	0.64	0.19	0.08
B	0.92	0.11	0.05
C	0.68	0.16	0.07
D	0.75	0.22	0.09
E	0.65	0.30	0.12
F	0.50	0.16	0.06
G	0.33	0.19	0.08

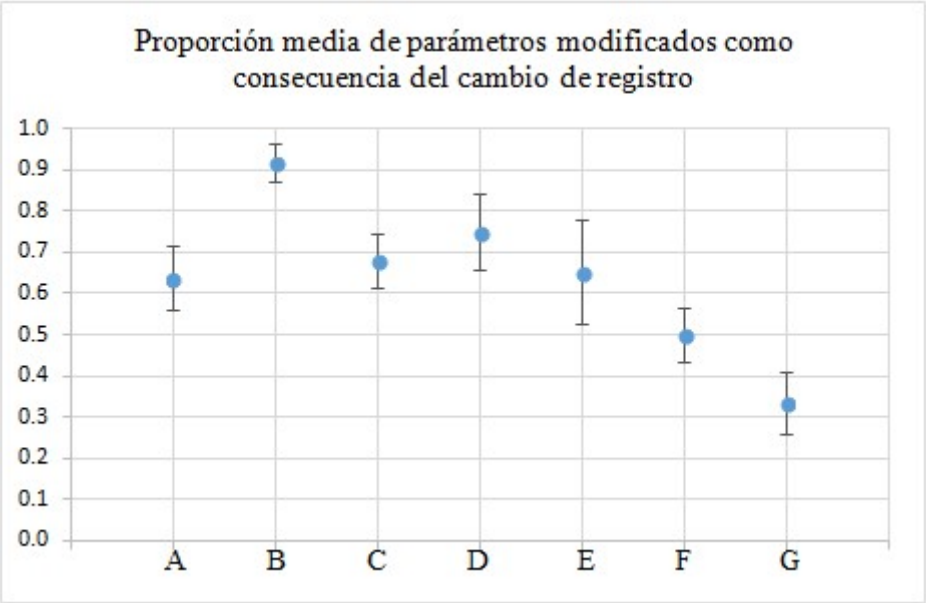


Figura 4.1. Proporción media de parámetros modificados como consecuencia del cambio de registro de fonación (modal/*falsetto*).

Algunos aspectos interesantes despuntan al comparar los valores medios calculados entre todas las proporciones de parámetros que han sido modificados por los 6 locutores al pasar de voz modal a *falsetto*.

Con una proporción media de algo más de 0.9, los parámetros del grupo B (cepstrales) han sido los que más se han modificado como consecuencia del cambio de registro. A estos les siguen inmediatamente los del grupo D (biomecánicos de la glotis) con una media de 0.75, y luego los parámetros pertenecientes a los grupos C (perfil espectral), E (base temporal de la onda glótica) y A (perturbación de la f_0), con valores medios de 0.68, 0.65 y 0.64, respectivamente.

Los grupos que presentaron menor variación fueron el F (parámetros de defecto de cierre glótico), variando la mitad de los parámetros (0.5), y el grupo G (temblor), variando solo un tercio (0.33).

En el gráfico, las barras de error representan la variación por sujetos. El grupo de parámetros que presenta mayores diferencias entre sujetos es el E y, el que menos, el B, grupo donde se ve que los sujetos son más consistentes entre ellos en variar una gran proporción de parámetros.

4.1.3. Análisis factorial: Análisis por Componentes Principales

Como se ha visto en el estudio de la producción de la voz, en concreto en el estudio enfocado al análisis de los rasgos glotales (§4.1.2), se maneja un elevado número de variables, muchas de las cuales están, además, relacionadas entre sí. Por ejemplo, si aumentan la tensión de la cubierta y del cuerpo del pliegue vocal (*Cover_Stiffness* y *Body_Stiffness*) aumentará también el tono (*Absolute_Pitch*); si la tensión soportada por la cubierta del pliegue es muy alta es esperable encontrar mayores gastos de energía en ella (*Cover_Losses*).

Por tanto, al gran número de variables se suma también un alto nivel de redundancia informativa como consecuencia de la covarianza. Esta situación no permite interpretar los datos con mucha claridad y además dificultaría, como es lógico, algunos estudios estadísticos posteriores que se pudieran querer realizar (véase el Capítulo 6), sobre todo si se pretenden diseñar modelos de regresión, en los cuales lo que se busca es predecir el comportamiento de una variable en función de otra u otras. En una situación como esta, parte de la variación que fuera capaz de explicar la variable *Absolute_Pitch*, por ejemplo, estaría también explicada por las variables *Cover_Stiffness* y *Body_Stiffness*, lo cual inflaría pero no aumentaría verdaderamente el poder predictivo del modelo.

Por estas razones, se ha optado por realizar un ***A**nálisis por **C**omponentes **P**rincipales* (o ACP) para solventar estas dificultades, reducir la dimensionalidad de los datos, e intentar determinar con más claridad qué Componentes Principales son mejores para definir los registros vocales.

El *Análisis de Componentes Principales* y otras técnicas de asociación también disponibles, como el *Análisis Lineal Discriminante,* ayudan mucho en este sentido. El *Análisis de Componentes Principales* es un método estadístico no paramétrico –es decir, un método que se aplica cuando se desconoce *a priori* cuál puede ser la distribución de los datos– capaz de extraer información relevante a partir de conjuntos complejos de datos y cuyo uso está ampliamente extendido (véase Jolliffe, 1986). Se utiliza con frecuencia para intentar conocer las causas de la variabilidad de tales conjuntos y ordenarlas en función de su contribución para explicarla, es decir, en función de la varianza explicada. Sin embargo, los beneficios que comportan la reducción de las variables y el aumento de la predictibilidad pueden ser dudosos si se consigue disminuir

la complejidad computacional pero no se logra interpretar el significado de los Componentes Principales resultantes del análisis[36] (véase por ejemplo, Godino, Gómez-Vilda y Blanco, 2006).

Como se mostrará más adelante, el rendimiento de los resultados extraídos del *Análisis por Componentes Principales* en el presente estudio ha resultado ser, sin embargo, doblemente provechoso, puesto que ha permitido reducir el número de dimensiones sin que por ello se sacrificara el significado: todos los Componentes Principales en los que ha quedado reducida la varianza de los datos son interpretables, como se verá enseguida.

Precisamente en aras de tal interpretación y para facilitar la generación de posibles nuevas hipótesis a partir de los resultados que se obtuvieran en el *Análisis por Componentes Principales*, se decidió reducir previamente el número de variables originales (72) escogiendo solo aquellos parámetros que presentaran un correlato fisiológico más claro. Es así que se seleccionaron los parámetros relativos a la perturbación de la frecuencia fundamental (Grupo A: parámetros 1, 2, 3 y 5)[37], los biomecánicos de la glotis (Grupo D: parámetros 35-46), los referidos a las variables temporales de la onda glótica (Grupo E: parámetros 47-58) y, finalmente, los concernientes al defecto de cierre glótico (Grupo F: parámetros 59-62). Solo quedaron excluidos del *Análisis por Componentes Principales* los parámetros cepstrales (Grupo B: parámetros 7-20), los de perfil espectral (Grupo C: parámetros 21-34) y los parámetros de temblor (Grupo G: parámetros 63-72). Esta exclusión se fundamenta, más detalladamente, de la siguiente manera:

- **Grupo B (parámetros 7-20):** la razón para excluir los parámetros cepstrales del *Análisis por Componentes Principales* es que no tienen una interpretación clara, la relación funcional que mantienen con los demás parámetros que sí se conocen es abstrusa. Es probable, como se ha visto, que los parámetros cepstrales sean sensibles al cambio de fonación (modal-*falsetto*) pero, al tener todavía un

[36] En el *Análisis por Componentes Principales* no se conserva el significado que tienen los parámetros o rasgos originales.

[37] Del grupo A se han ignorado los parámetros 4 y 6, es decir, *Absolute Normal Minimum Sharpness* (grado de afilamiento o agudeza del pico del *Maximum Flow Declination Rate* o *MFDR*) y *Muc./AvAc. Energy*, *MAE* (relación entre las energías de la onda mucosa/onda glótica), respectivamente, por no tener una interpretación clara hasta el momento.

significado difuso, la relación con el cambio de registro sería poco transparente, como de "caja negra" y, por tanto, no aclararía demasiado su análisis.

- **Grupo C (parámetros 21-34):** posiblemente entre ellos también podría haber alguno que manifestara sensibilidad al cambio de fonación, pero sucede lo mismo que con los cepstrales, su interpretación semántica es demasiado vaga en cuanto al cambio de registro. En voces con una frecuencia fundamental alta –como la voz femenina o más aún el *falsetto*– los armónicos están lógicamente más separados entre sí puesto que son múltiplos naturales de la f_0 (recuérdese lo explicado en §1.3.2.). Dada esta mayor separación entre armónicos, el número de puntos de datos que se puede extraer para un determinado rango de frecuencias es más limitado. Dicho de otro modo, la reconstrucción de la envolvente espectral requiere un mayor nivel de interpolación o inferencia entre los datos muestrales.

  Cuando el espectro es más compacto, porque las líneas armónicas están más juntas, como en el caso de la voz masculina modal, entonces sí se obtiene una estimación de la envolvente o tendencia espectral más definida, porque precisamente, el muestreo es más detallado. Esta es la razón por la que la información que puede ofrecer el espectro armónico en la voz femenina o en el *falsetto* (con frecuencias fundamentales elevadas en ambos casos, aunque en diverso grado) es mucho más complicada de tratar o procesar.

- **Grupo G (parámetros 63-72):** se ha prescindido de ellos para el *Análisis por Componentes Principales* sencillamente porque no son pertinentes en el análisis del *falsetto*. De hecho, el temblor es otro tipo de fonación, vinculada al ámbito del canto (*vibrato*) o al de las patologías vocales (*tremor*).

En síntesis, de los 72 parámetros originales fueron seleccionados para el *Análisis por Componentes Principales* solo aquellos más relevantes y cuya interpretación en términos fisiológicos fuera más clara, de modo que el total de parámetros analizables se redujo a 32 (grupos A[38], D, E y F). Como se ha dicho ya, los que han sido excluidos del análisis fueron descartados bien por no tener un significado claro (grupo B), bien

[38] A excepción de los parámetros 4 y 6.

por ser muy complicados de interpretar en el registro de *falsetto* (grupo C); bien por no tener nada que ver con el registro *falsetto* (grupo G).

El *Análisis por Componentes Principales* genera, a través de unas operaciones matemáticas, nuevas variables o Componentes Principales –abreviados como CPs–, cuyo número coincide siempre con el de las variables originales. Por tanto, al inicio del análisis hay tantos Componentes Principales como parámetros originales, en este caso, 32, pero, como se verá enseguida (véase, más adelante, la Tabla 4.5), no todos los CPs explican el mismo porcentaje de la varianza total de los datos, es decir, no todos alcanzan la misma relevancia explicativa. Por convención, solamente se mantienen en el análisis aquellos Componentes Principales que expliquen al menos un 5% de la varianza total de los datos; los componentes que no alcanzan ese porcentaje mínimo son, en consecuencia, descartados (véase, Baayen, 2008, p. 130).

Puesto que los Componentes Principales están calculados matemáticamente para que no exista ninguna correlación entre ellos, nunca se dará el caso de que un subconjunto de rasgos agrupados bajo el componente *x* también sea capaz de explicar el componente *y*. Esto significa que en el *Análisis por Componentes Principales* no existen solapamientos explicativos entre los diferentes CPs y, por tanto, la carga explicativa que se añada al incorporar al análisis un Componente Principal nuevo representará una ganancia neta con respecto a su poder predictivo.

Los Componentes Principales se ordenan jerárquicamente, de mayor a menor, en función de su poder explicativo, es decir, en función del porcentaje de varianza que cada uno es capaz de explicar. Luego de adoptar este criterio umbral del 5% para seleccionar los componentes más informativos, los 32 parámetros quedaron reducidos a 4 Componentes Principales y con ellos es posible dar cuenta del 81,75% de la varianza total de los datos, como se observa en la Tabla 4.9.

Tabla 4.9. Lista de los 32 Componentes Principales (CPs) en orden de importancia, de mayor a menor, según el porcentaje (%) de la varianza explicada por cada uno de ellos.

Componentes	% de varianza	% acumulado
1	**31.751**	**31.751**
2	**23.449**	**55.200**

3	**20.438**	**75.637**
4	**6.120**	**81.757**
5	4.129	85.886
6	3.471	89.357
7	1.983	91.341
8	1.610	92.950
9	1.222	94.172
10	1.112	95.284
11	.843	96.127
12	.740	96.867
13	.566	97.433
14	.460	97.894
15	.430	98.324
16	.347	98.671
17	.297	98.968
18	.264	99.232
19	.207	99.439
20	.150	99.589
21	.096	99.685
22	.077	99.761
23	.065	99.826
24	.049	99.875
25	.042	99.917
26	.031	99.947
27	.020	99.967
28	.016	99.983
29	.010	99.993
30	.006	99.998
31	.001	100.000
32	.000	100.000

Las variables originales se reparten o se apoyan en los CPs en distintas proporciones, según sea la aportación de cada variable original a la explicación del nuevo CP. Una vez que se ha decidido mantener los 4 primeros CPs por ser los más

informativos se realiza la rotación por el método *Varimax* para facilitar su interpretación, puesto que así, cada variable original intentará agruparse en el menor número de CPs posible.

En los resultados de la rotación, el primer Componente Principal es capaz de explicar el 27,30 % de la varianza de los datos, al añadirse el segundo Componente Principal se explica un 21,63 % adicional, el Componente 3 añade un 17,83 %, y finalmente, el Componente 4 es capaz de explicar un 14,98 % más, tal y como se observa en la Tabla 4.10.

Tabla 4.10. Extracción de 4 Componentes Principales y sus porcentajes de explicación de varianza luego de la rotación.

Componentes Principales	% varianza	% acumulado
1	27.307	27.307
2	21.631	48.937
3	17.838	66.775
4	14.982	81.757

Las 32 variables originales quedan pues, asociadas a 4 factores o Componentes Principales. En la Tabla 4.11 se muestra dicha agrupación y el grado de correlación entre cada una de las variables originales y el componente que las agrupa.

Tabla 4.11. Distribución de los 32 parámetros originales en los 4 Componentes Principales y sus grados de correlación. Las casillas en blanco no llegan a una correlación con valor absoluto mínimo de .300.

	Componentes Principales			
	1	2	3	4
Absolute_Pitch Media	.936			
Body_Stiffness Media	.895			
Noise_Harm._Ratio_NHR Media	.894			

Cover_Losses Media	.874		-.305	
Cover_Stiffness Media	.856			
Cover_Mass_Unbalance Media	.818			
Cover_Losses_Unbalance Media	.793			.442
Cover_Stiffness_Unbalance Media	.786			
Body_Mass Media	-.772			
Rel._Stop_Flow_Time Media	.622	.468	.463	
Val._Flow_GAP Media	-.620	.547		
Rel._Open_2_Time Media		.967		
Rel._Max._Ampl._Time Media		.960		
Rel._Closing_Time Media		.932		
Rel._Open_1_Time Media		.923		
Rel._Start_Flow_Time Media	-.437	.834		
Rel._Recov._2_Time Media	.392	.824		
Val._Permanent_GAP Media	.493	-.638		
Rel._Recov._1_Time Media	.577	.628		
Rel._Open_1_Ampl Media			-.916	
Rel._Recov._2_Ampl Media			-.883	
Rel._Recov._1_Ampl Media			-.822	
Rel._Open_2_Ampl Media		-.501	-.809	
Val._Adduction_GAP Media			.796	
Cover_Mass Media			-.668	-.354
Abs._Norm._Ar._Shimmer Media	-.489		.540	.358
Val._Contact_GAP Media		.303	.399	
Body_Stiffness_Unbalance Media				.946
Body_Mass_Unbalance Media				.921
Abs._Norm._Jitter Media			.379	.861
Body_Losses Media		.359		-.797
Body_Losses_Unbalance Media			.470	.748

Como se dijo al comienzo de este apartado, la interpretación de los Componentes Principales suele ser, en ocasiones, problemática, y en ello radica el principal desafío de este tipo de análisis. Esta dificultad de interpretación ocurre porque las nuevas variables abstractas, es decir, los CPs, son una combinatoria de las diferentes variables de origen. El significado de cada variable original se conoce, pero ahora es necesario encontrar un sentido a cada uno de los CPs resultantes del análisis.

Los parámetros originales han quedado ya agrupados en 4 Componentes Principales (CPs), como se ha explicado. En el interior de cada Componente Principal, los parámetros están ordenados por valor absoluto de correlación o magnitud (ver *supra* Tabla 4.11). El número que aparece para cada parámetro indica su correlación con el Componente Principal: cuanto más alto resulta ese valor mayor correlación mantienen

ambos entre sí, y, cuanto más bajo es ese número, menos fuerte es la correlación de ese parámetro concreto con el Componente Principal.

Además de observar el peso que cada parámetro original asume en la determinación de cada Componente Principal hay que considerar los coeficientes negativos o positivos, porque algunos parámetros estarán positivamente correlacionados mientras otros lo estarán negativamente. Así, por ejemplo, de la observación de la Tabla 4.11 se deduce que los parámetros *Absolute_pitch* y *Body_mass* están ambos correlacionados con el Componente Principal 1; sin embargo, el primero presenta una correlación positiva (.936) —es decir, el valor del CP1 aumenta cuando se incrementa el valor de la variable o parámetro original *Absolute_pitch*—, mientras que el otro presenta una correlación negativa (-.772), es decir, cuando aumenta el valor del CP1 disminuye el valor del parámetro original *Body_mass*. Este resultado es extremadamente coherente con la dinámica del sistema biomecánico, pues cuanto mayor sea la masa dinámica vibrante, menor será el tono fundamental observado. En cambio, la correlación entre el *Body Stiffness* (tensión del cuerpo del pliegue vocal) y el CP1 es de 0.895, expresando una relación dinámica alineada: a mayor tensión en el pliegue, mayor será el tono fundamental observado.

Al examinar los datos que arroja el Análisis de Componentes Principales, se aprecia que muchos de los parámetros muestran una fuerte correlación con el Componente Principal bajo el que se agrupan. Para comprender qué parámetros o rasgos contribuyen más a la configuración de cada Componente Principal, se optó por escoger aquellos cuyo valor absoluto de correlación hubiera sido superior a 0.7[39]; los parámetros que estuvieron por debajo de ese valor se descartaron por exhibir una débil correlación con su Componente Principal correspondiente. Se presentan en las Tablas 4.12, 4.13, 4.14 y 4.15 los parámetros más relevantes que componen cada uno de los 4 Componentes Principales o CPs, ordenados por valor absoluto de correlación o magnitud.

[39] No hay ningún parámetro cuya correlación sea superior a 0.7 con más de un Componente Principal (CP) a la vez.

Tabla 4.12: Rasgos originales agrupados bajo el primer Componente Principal por orden de valor absoluto de correlación o magnitud.

Componente Principal 1	
Rasgos o parámetros	**Coeficiente factorial**
Absolute_Pitch Media	.936
Body_Stiffness Media	.895
Noise_Harm._Ratio_NHR Media	.894
Cover_Losses Media	.874
Cover_Stiffness Media	.856
Cover_Mass_Unbalance Media	.818
Cover_Losses Unbalance Media	.793
Cover_Stiffness_Unbalance Media	.786
Body_Mass Media	-.772

El primer Componente Principal es el más relevante, pues, como se ha indicado antes, es capaz de explicar el 27,30 % de la varianza total de los datos. Este primer componente reúne 9 parámetros o rasgos y parece estar recogiendo información tonal. Está centrado en el *pitch* y, en general, en las diversas estrategias conducentes a aumentar o disminuir la f_0. Además del *pitch* están asociados al Componente 1 rasgos como el comportamiento de la cubierta del pliegue vocal, y de las masas y tensiones del cuerpo. La masa de las distintas partes de los pliegues vocales se pude redistribuir en distintas localizaciones y variar tanto su grado de tensión así como su comportamiento de vibración. Todos los rasgos relativos a la cubierta del pliegue presentan una elevadísima correlación positiva con el primer Componente Principal, mientras que el parámetro *Body_Mass Media* se vincula con él negativamente. Por tanto, los parámetros incluidos en el Componente Principal 1 que están vinculados con el cuerpo del pliegue vocal –tensión y masa– contribuyen a definir el CP1 de forma inversa, la tensión con una proyección de 0.895 (más tensión, más f_0), mientras que la masa contribuye negativamente con una proyección de -0.772 (a menor masa, mayor f_0), como ya se ha comentado anteriormente.

Tabla 4.13: Rasgos originales agrupados bajo el segundo Componente Principal por orden de magnitud.

Componente Principal 2	
Rasgos o parámetros	**Coeficiente factorial**
Rel._Open_2_Time Media	.967
Rel._Max._Ampl._Time Media	.960
Rel._Closing_Time Media	.932
Rel._Open_1_Time Media	.923
Rel._Start_Flow_Time Media	.834
Rel._Recov._2_Time Media	.824

El segundo Componente Principal explica el 21,63 % de la varianza. En él se agrupan 6 rasgos o parámetros originales que, como se ve, también parecen ser selectivos, puesto que dan cuenta exclusivamente de factores temporales relativos a la onda glotal. El Componente 2 recoge parámetros que definen, en el total de la duración de un ciclo glotal, el instante en el que se producen aspectos relevantes durante un ciclo de fonación, como por ejemplo, los instantes que indican los momentos de apertura y cierre glotal, entre otros (véase, para más detalles, *BioMet®Phon User's Manual*, 2014, p. 7). Los 6 parámetros originales exhiben una correlación muy alta y positiva con el CP2.

Tabla 4.14: Rasgos originales agrupados bajo el tercer Componente Principal por orden de magnitud.

Componente Principal 3	
Rasgos o parámetros	**Coeficiente factorial**
Rel._Open_1_Ampl Media	-.916
Rel._Recov._2_Ampl Media	-.883
Rel._Recov._1_Ampl Media	-.822
Rel._Open_2_Ampl Media	-.809
Val._Adduction_GAP Media	.796

El tercer Componente Principal da cuenta del 17,83 % de la varianza y muestra una tendencia también marcada, porque se observa que los 4 parámetros de mayor peso se asocian de forma muy fuerte y negativa con el Componente Principal, y todos ellos dan cuenta de las amplitudes relevantes de la onda de flujo glotal. El quinto parámetro que queda agrupado bajo el tercer factor principal con una magnitud positiva de .796 % es un defecto observado en la fase de aducción de los pliegues vocales, previa al contacto, o una aducción incompleta, que también estaría relacionado con la amplitud.

Tabla 4.15: Rasgos originales agrupados bajo el cuarto Componente Principal por orden de magnitud.

Componente Principal 4	
Rasgos o parámetros	**Coeficiente factorial**
Body_Stiffness_Unbalance Media	.946
Body_Mass_Unbalance Media	.921
Abs._Norm._Jitter Media	.861
Body_Losses Media	-.797
Body_Losses_Unbalance Media	.748

Finalmente, el cuarto y último Componente Principal explica el 14,98 % de la varianza. Aunque no es en absoluto poco lo que consigue explicar, su relevancia es bastante menor si lo comparamos con el primer Componente (CP1), como es lógico. Agrupa 5 de los rasgos originales y estos se centran en el comportamiento del cuerpo del pliegue vocal, y su influencia en la variabilidad temporal de la frecuencia fundamental conocida como *jitter*.

En resumen, la agrupación de rasgos en estos 4 componentes principales ha resultado ser realmente selectiva: cada uno de los factores principales ha seleccionado parámetros de unas características similares entre sí, constituyendo de este modo 'familias' de parámetros. El primer Componente Principal se vincula de forma muy clara con el *pitch* y con el comportamiento de la cubierta principalmente, al tiempo que el cuarto Componente Principal lo hace con el cuerpo del pliegue vocal. El segundo Componente Principal recoge la dimensión temporal de la onda glotal indicando los

instantes de apertura y recuperación en el ciclo. El tercer Componente Principal recoge la dimensión de las amplitudes correspondientes.

Para conocer si los valores de los Componentes Principales aumentan o disminuyen en modal o en *falsetto* se realizó un *test* de medias con corrección de Bonferroni y se obtuvieron también resultados muy interesantes, como se observa en las Tabla 4.16 y 4.17. Los valores de la misma fila que no comparten el mismo subíndice (a, b) son significativamente diferentes en $p< .05$.

Tabla 4.16: Resultados del contraste de medias para todos los locutores.

	Falsetto	**Modal**
	Media	Media
CP1	**.90363$_a$**	**-.90363$_b$**
CP2	.05062$_a$	-.05062$_a$
CP3	.08756$_a$	-.08756$_a$
CP4	-.05015$_a$	.05015$_a$

Tabla 4.17. Resultados del contraste de medias por locutor.

			Falsetto	**Modal**
			Media	Media
Locutor	1	CP1	.54137$_a$	-.69249$_b$
		CP2	-.17152$_a$	-2.02209$_b$
		CP3	.30977$_a$	-.06369$_a$
		CP4	.99664$_a$	-.22832$_b$
	2	CP1	.83115$_a$	-.80410$_b$
		CP2	.03742$_a$	-.22940$_a$
		CP3	-.10373$_a$	1.55220$_b$
		CP4	.64472$_a$	.28434$_a$
	3	CP1	.77149$_a$	-1.32498$_b$
		CP2	.31729$_a$	.30246$_a$
		CP3	.25842$_a$	-.69912$_b$
		CP4	-1.02627$_a$	-.56898$_b$
	4	CP1	1.47695$_a$	-1.19954$_b$
		CP2	.18501$_a$	1.46173$_b$
		CP3	-.36066$_a$	-1.03654$_b$
		CP4	-.02930$_a$	-.73767$_a$
	5	CP1	1.41513$_a$	-.66571$_b$

		CP2	.22089$_a$	-.96776$_b$
		CP3	-.81648$_a$	-1.08473$_b$
		CP4	-.45736$_a$	.47007$_b$
	6	CP1	.38567$_a$	-.73495$_b$
		CP2	-.28536$_a$	1.15135$_b$
		CP3	1.23807$_a$	.80650$_a$
		CP4	-.42930$_a$	1.08144$_b$

Según los resultados obtenidos a partir del cotejo de las medias, todas las variables que aparecen recogidas en el primer factor o Componente Principal (CP1) varían para todos los locutores y siempre en el mismo sentido (salvo en el caso del parámetro vinculado con la masa del pliegue vocal) cuando cambian de voz modal al *falsetto*. El único factor principal discriminante entre ambos registros es, pues, el primer Componente Principal (CP1), es decir, el factor asociado con el *pitch*, porque es el único diferenciador al pasar de un registro a otro en todos los locutores. En la Figura 4.2 se presentan gráficamente y para todos los locutores los resultados obtenidos en el *test* de medias en una matriz simétrica de 4 x 4.

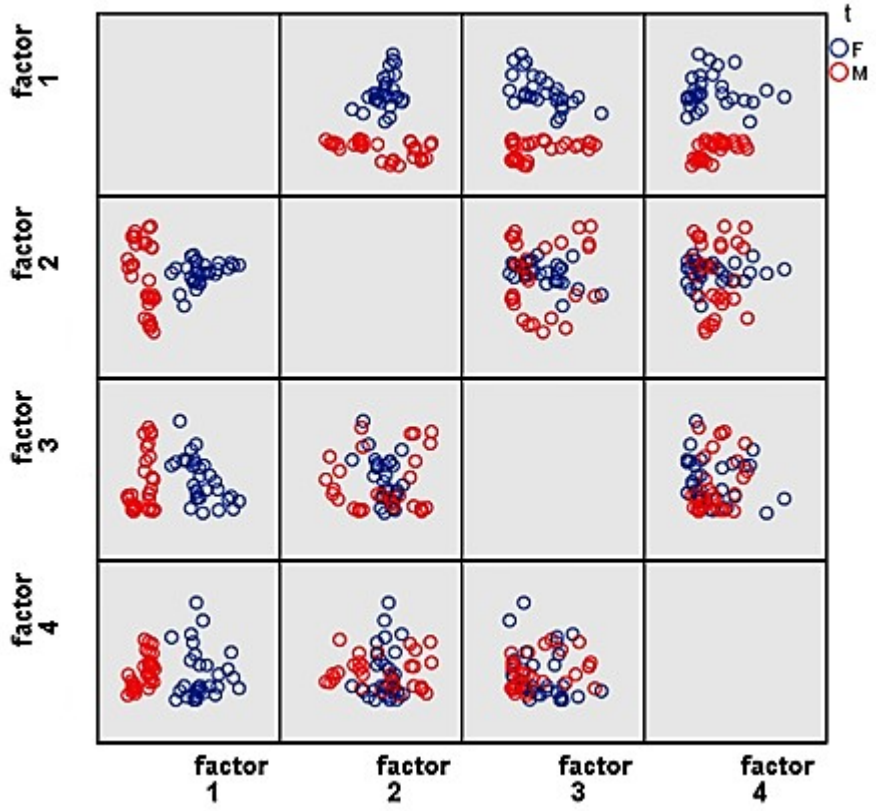


Figura 4.2. Gráfico de dispersión matricial.

Los casilleros en los que los círculos rojos (voz modal) se separan de los azules (*falsetto*) de tal forma que podría trazarse una línea recta entre ellos y dividirlos en dos grupos muestran que ese componente es un factor apto para la clasificación binaria por registros con dos decisiones: voz modal o *falsetto*. Este es el caso del primer Componente Principal (fila 2 - columna 1, fila 3 - columna 1 y fila 4 - columna 1 y también, fila 1 - columna 2, fila 1 - columna 3, fila 1- columna 4).

A continuación se explica en detalle el comportamiento del primer Componente Principal, las variables que lo determinan, cuáles son y cómo se comportan ellas mismas, puesto que el primer Componente Principal (CP1) se ha revelado como el único factor definitorio del cambio de registro entre voz modal y *falsetto* y al que, en consecuencia, se ha optado por denominar, precisamente, *registro fonatorio*.

4.1.3.1. Primer Componente Principal (CP1): registro fonatorio

Como se observa en los datos, los valores de todos los rasgos que componen el CP1 aumentan significativamente respecto de su media en el *falsetto* y disminuyen significativamente en el registro de voz modal con la única excepción del parámetro *Body Mass Media* que, como se ha referido antes, se comporta inversamente, es decir, disminuye significativamente en el *falsetto* y aumenta en el registro modal o, dicho de otro modo, a mayor tono, menor masa del cuerpo del pliegue vocal implicada en la vibración.

Parámetros del primer Componente Principal (CP1) y comportamiento de los mismos:

> *Absolute Pitch Media:* refiere al valor medio de la f_0 y es absolutamente lógico que aumente de forma significativa en el *falsetto* y disminuya también significativamente en el registro de voz modal. Se comprueba, pues, que en el mecanismo de fonación del *falsetto* sube la f_0 y, en consecuencia, el tono resultante. El *Análisis por Componentes Principales* constata este hecho detectando este parámetro como diferenciador claro de ambos registros, lo cual valida la sensibilidad del análisis.

Noise Harmonic Ratio NHR Media: da cuenta de la proporción existente entre ruido y armonicidad. El análisis muestra, una vez más, valores significativamente mayores para el *falsetto* que para el registro modal. El hecho de que exista, por tanto, un componente de turbulencia mayor en el *falsetto* podría estar indicando un peor cierre de los pliegues vocales en este registro respecto del modal.

Cover Stiffness Media: alude a la tensión soportada por la cubierta del pliegue vocal. El hecho de que su valor medio sea mayor en el *falsetto* que en el registro modal es también congruente con lo que se ha descrito en la bibliografía (véase Capítulo 1, § 1.3 y § 1.4.3)

Cover Losses Media: da cuenta de gastos de energía en la cubierta y el análisis muestra valores significativamente superiores para el registro de *falsetto*. Es esperable encontrar más pérdidas de energía en la cubierta cuando la fonación es más forzada.

Cover Mass Unbalance Media, Cover Stiffness Unbalance Media y Cover Losses Unbalance Media: son, como se ha visto antes, parámetros que hacen referencia a las irregularidades de la vibración en la cubierta. Era de suponer que tal desequilibrio fuese mayor en *falsetto* que en voz normal y esto se comprueba en los resultados. Los valores arrojados parecen indicar que puede que exista una vibración más irregular en un registro que en otro y que esta sea mayor en el *falsetto*.

Body Mass Media: este parámetro da cuenta de la cantidad de masa correspondiente al cuerpo de cada pliegue vocal implicada en la vibración. Es el único parámetro que se correlaciona negativamente con el CP1 (-.772) –a mayor tono, menor masa implicada en la fonación– y que disminuye al cambiar al registro de *falsetto*, como era también de esperar (véase nuevamente el Capítulo 1, §1.3, §1.4.3).

Body Stiffness Media: refiere a la tensión que presenta el cuerpo de cada pliegue vocal durante cada ciclo vibratorio glótico. Los datos muestran que esta aumenta al pasar al *falsetto*, como cabría esperar también (véase Capítulo 1, §1.3, §1.4.3).

En síntesis, aunque el tono sea el primer parámetro que se revela como principal diferenciador de los registros modal-*falsetto*, no puede ser, en absoluto, suficiente para determinar tal cambio. Fundamentar el cambio de registro solo con este parámetro sería incorrecto, puesto que la f_0 puede, evidentemente, ascender aun en la fonación modal; sin embargo, otros parámetros asociados a él no, como son los biomecánicos. Como se comprueba en los datos obtenidos, estos últimos son los que están realmente induciendo el comportamiento de la f_0 y son, en definitiva, los que, junto con ella, justifican el cambio de registro. Tal y como explica Titze (2000 [1994] p. 291) cuando la f_0 se incrementa por encima de un determinado punto y este punto dependerá lógicamente de cada hablante, el músculo Tiroaritenoideo no puede seguir activo en un pliegue vocal tal elongado, por lo que se relaja y se retrae y, en consecuencia, disminuye la superficie en vibración (véase también Alves *et al*., 2014, p. 599). La masa que vibra en ese momento es, pues, muy reducida, de modo que la tensión es muy elevada, por concentrarse en una menor superficie. Esta situación es precisamente la que se corrobora en los datos obtenidos del *Análisis por Componentes Principales* y del test de medias.

Los rasgos agrupados en el CP1 son, pues, muy interesantes porque implican la caracterización del modo de fonación *falsetto*: parecen ser los que verdaderamente lo determinan. Además, la pauta de comportamiento que sigue la cubierta de los pliegues vocales se revela esencial en el cambio de registro (explica la mayor parte de la varianza), lo cual significa que es el elemento diferenciador. El grado de implicación de la cubierta es muy distinto en el registro modal que en el *falsetto* (en este, prácticamente solo vibran las capas más externas de los pliegues) y esto es congruente con lo planteado previamente por Titze (2000 [1994], pp. 289-300).

El resto de las variables recogidas en los Componentes Principales CP2, CP3 y CP4 tienen un comportamiento más oscilante, unas veces cambian entre registros y otras veces no, dependiendo de cada locutor, como se aprecia en la Tabla 4.17 (véase *supra*) y en la 4.18, que resume dicha información.

Tabla 4.18: Resumen de los cambios comprobados en el comportamiento glotal agrupados en los 4 Componentes Principales y diferenciados por locutor. Las casillas sombreadas destacan los CPs que han experimentado una alteración significativa en cada locutor al cambiar de registro.

	Factores o Componentes Principales			
Locutores	CP1	CP2	CP3	CP4
L1				
L2				
L3				
L4				
L5				
L6				

4.2. Análisis del habla: *tempo*[40]

Como se ha explicado al presentar la metodología en el Capítulo 3, a través del diseño experimental se ha procurado controlar la velocidad de habla, utilizando siempre la misma pseudofrase leída. Sin embargo, y como también se ha dicho antes, los cambios de registro pueden suponer también cambios en el *tempo*, y así lo han constatado algunos estudios previos (recuérdense los trabajos de Künzel, 2000 y Wagner y Köster, 1999). A su vez, los aspectos temporales del habla en general se han revelado como muy útiles para la caracterización de hablantes individuales y, por tanto, para su reconocimiento (véanse, por ejemplo, Gold y French, 2011; Gold, 2012; Künzel, 1997; Romito, Lio y Galatà, 2005; entre muchos otros). Estas constataciones motivaron el análisis del *tempo* en los 6 locutores de este estudio. Los resultados más relevantes del *tempo* característico de los locutores fueron presentados en Fernández Trinidad y

[40] Recuérdese lo señalado previamente en el Capítulo 3 sobre la utilización de los términos *tempo*, *velocidad de elocución*, *velocidad de habla* en esta tesis. Estas expresiones se utilizarán para hacer referencia a los resultados obtenidos en esta tesis a pesar de que se haya calculado únicamente la duración promedio de las frases producidas por los locutores en modal y en *falsetto*. Esta decisión metodológica explicada en (§3.3.2) se justifican porque los locutores pronunciaron siempre la misma frase, las mismas sílabas y no realizaron pausas.

Rojo (2018) y en esta tesis se retoman pero analizándolos y explicándolos con mayor detalle y claridad.

La Tabla 4.19 muestra la duración promedio de las frases de cada locutor, hablando en modal, en *falsetto* y sin diferenciar registros (global).

Tabla 4.19: Media (y desviación típica) de la duración de las frases (en ms) por locutor, por registro vocal y en global.

	L1	**L2**	**L3**	**L4**	**L5**	**L6**
Modal	1.486 (0.062)	1.461 (0.066)	1.372 (0.014)	1.599 (0.045)	1.256 (0.036)	1.173 (0.072)
Falsetto	1.350 (0.072)	1.590 (0.073)	1.366 (0.036)	1.721 (0.061)	1.165 (0.029)	1.126 (0.031)
Global	1.418 (0.096)	1.525 (0.095)	1.369 (0.026)	1.660 (0.082)	1.211 (0.057)	1.150 (0.058)

Primeramente se realizó un análisis de la varianza (ANOVA), el cual demostró que la variable locutor ejerce un efecto principal sobre la duración global de los enunciados (fusionando el registro modal y *falsetto*), $F(5,54)=67.94$, $p<0.001$***. El análisis post-hoc con corrección de Bonferroni reveló igualmente diferencias significativas entre todos los locutores (siempre con los registros agrupados) excepto entre los locutores L1 y el L3 y entre los locutores L5 y L6. Todas las diferencias fueron significativas y alcanzaron el nivel de $p<0.001$*** menos las existentes entre los locutores L1 y el L2, que obtuvieron una significatividad menor, ($p<0.05$*). En el gráfico que se presenta en la Figura 4.3 se recogen estas diferencias.

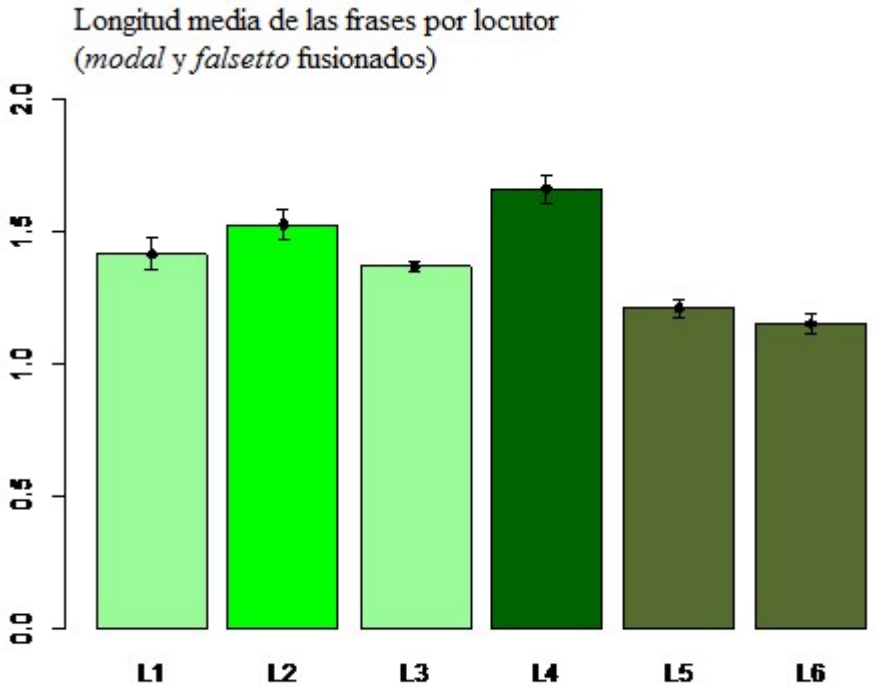


Figura 4.3. Diferencias en la duración de las frases (sin diferenciar registros) de los 6 locutores.

También se ha comprobado que la variable locutor ejerce un efecto principal sobre la duración de las frases considerando cada registro por separado, para el registro modal $F(5,24)=43.59$, $p<0.001$*** y para el *falsetto* $F(5,24)=93.61$, $p<0.001$***. Los análisis post-hoc (Bonferroni) arrojan los datos que se detallan a continuación. En el registro modal, no se aprecian diferencias significativas en la duración de las frases entre los locutores L1 y L2; L2 y L3; L5 y L6. En cambio, las parejas de locutores L1-L3; L1-L4; L3-L5 sí se diferencian con una significatividad de $p<0.05$*; las parejas L2-L4 con una significatividad mayor, valores de $p<0.001$** y, finalmente, los pares L1-L5; L1-L6; L2-L5; L2-L6; L3-L4; L3-L6; L4-L5; L4-L6 exhiben la máxima diferencia con valores de $p<0.001$***. Todos estos resultados se presentan de forma más clara en el gráfico de la Figura 4.4.

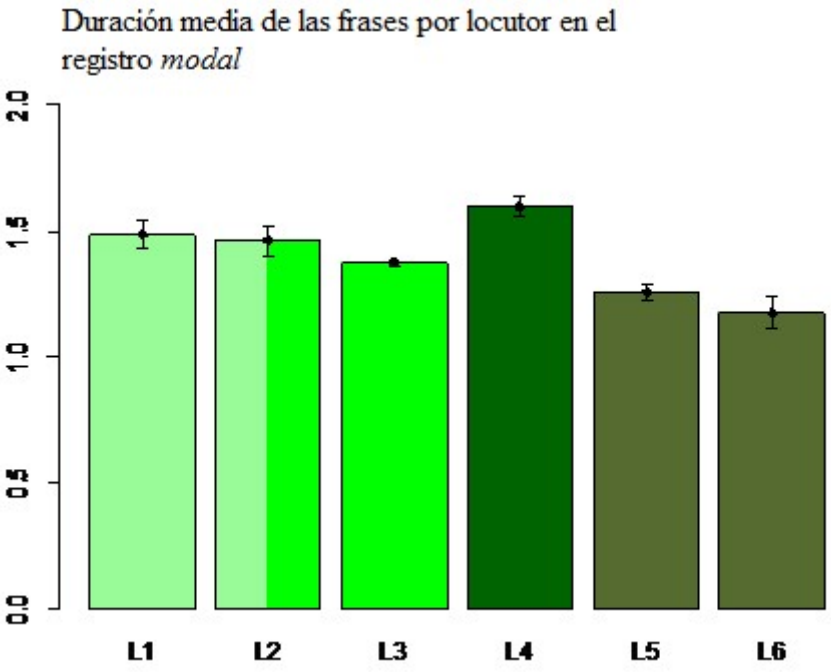


Figura 4.4. Diferencias en la duración de las frases pronunciadas en voz modal por los 6 locutores.

El análisis post-hoc efectuado en el registro de *falsetto* indica que, en cuanto a la duración, no se distinguen las frases pronunciadas por las parejas de locutores L1-L3 ni L5-L6; mientras existen diferencias significativas entre los locutores L2-L4 ($p<0.05$*) y para el resto de los casos ($p<0.001$***). Véase la Figura 4.5 para el resumen de estos resultados.

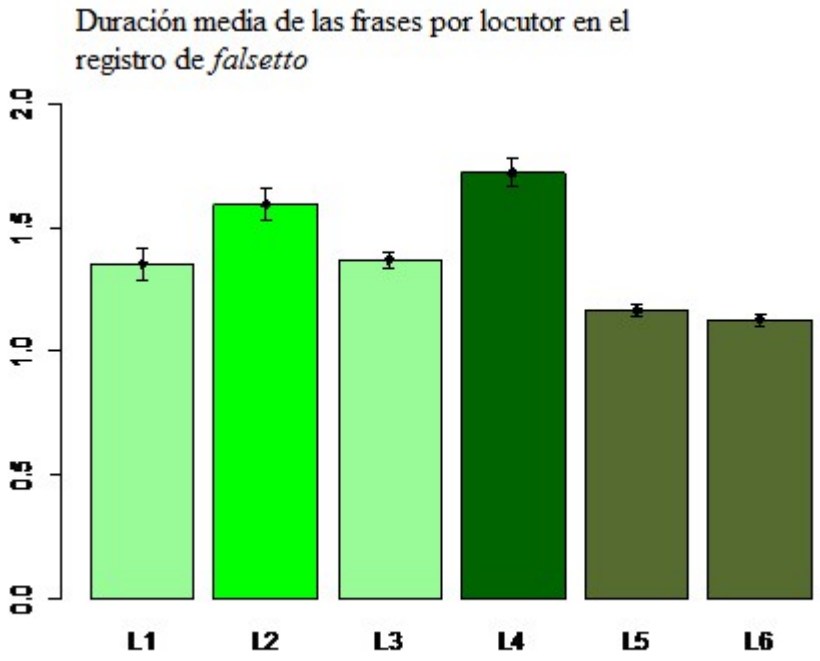


Figura 4.5. Diferencias en la duración de las frases pronunciadas en *falsetto* por los 6 locutores.

El efecto principal ejercido por el cambio de registro no fue significativo aunque sí resultó serlo la interacción entre el locutor y el registro, con valores de $F(5,48)=10.61$, $p<0.001$***. Los análisis post-hoc (Locutor*Registro) mostraron los siguientes resultados: la velocidad de las frases pronunciadas por el Locutor 1 con voz modal (M1) se diferencian de la velocidad de las frases pronunciadas por el mismo Locutor 1 cuando habla en *falsetto* (F1), es decir, M1≠F1; y lo mismo sucede con los Locutores 2, 4 y 5, es decir, M2≠F2, M4≠F4, M5≠F5 (todas estas diferencias alcanzan un nivel de significatividad de $p<0.05$*). La duración en el registro modal y en el *falsetto* para el resto de los locutores fue la misma, por tanto, M3=F3 y M6=F6.

Finalmente, todos los efectos descritos anteriormente se resumen gráficamente en la Figura 4.6.

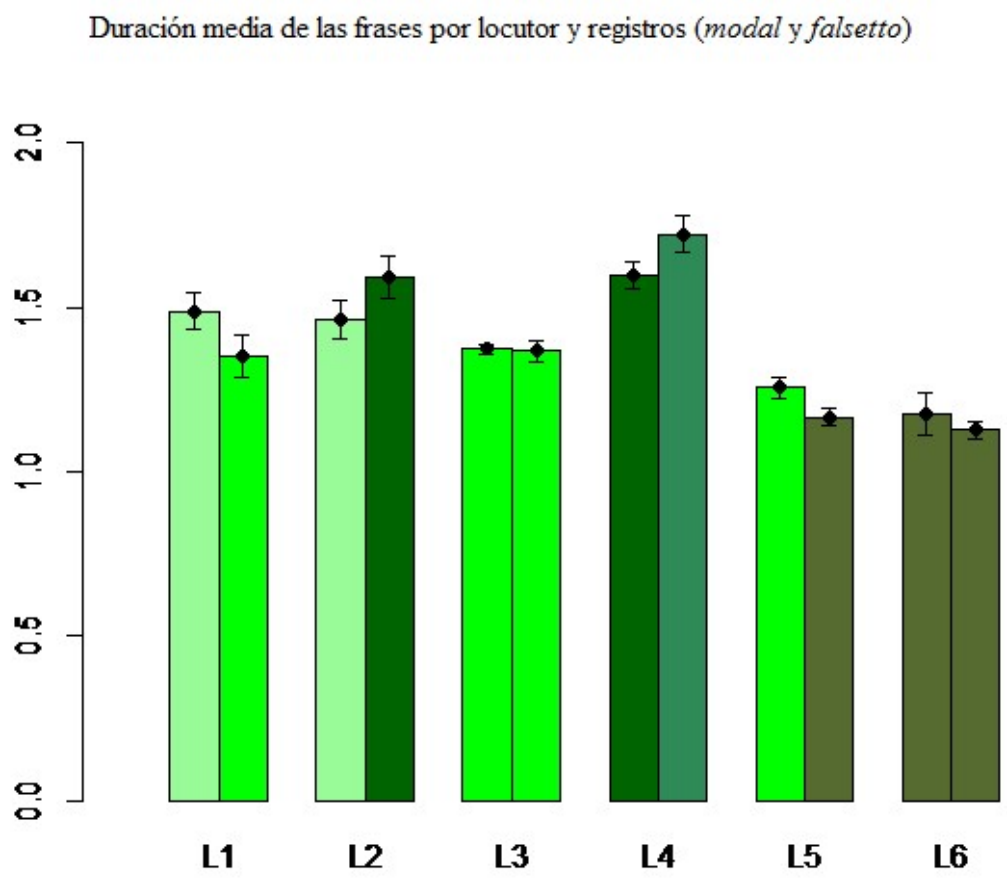


Figura 4.6. Duración media de las frases en función del locutor y registro vocal (voz modal en la columna izquierda y *falsetto* en la columna derecha). Figura tomada de Fernández Trinidad y Rojo (2018).

4.2.1. Tempo global

Al evaluar el efecto principal encontrado entre las variables locutor y duración se observó que únicamente tres locutores (en concreto, L4, L5 y L6) destacan por la duración global (registro modal+*falsetto*) de sus enunciados. El Locutor 4 tardó significativamente más que el resto en pronunciar las frases —es decir, fue el que habló más lentamente—, mientras que los Locutores 5 y 6 tardaron significativamente menos —fueron los más rápidos—. Por tanto, cabría predecir que la velocidad de habla solo podría constituir una pista perceptiva únicamente para el caso de los locutores 5 y 6

como bloque (sin poder distinguirlos entre sí) y, por otra parte, para el Locutor 4, por haber sido este el más lento de todos los locutores. Sin embargo, el factor temporal del habla difícilmente podría ser de utilidad para reconocer al resto de los locutores (L1, L2 y L3) quienes, considerando la duración global de sus frases, no hablaron ni especialmente lento ni especialmente rápido respecto de la media.

4.2.2. El tempo según los registros

A diferencia de los resultados obtenidos por Künzel (2000), en este estudio no se ha comprobado un efecto homogéneo del registro en la duración de las frases. Se verifica que algunos locutores modifican el *tempo* al cambiar de registro pero su comportamiento no es unívoco, es decir, los locutores que muestran una diferencia en la velocidad de sus frases en un registro y otro parecen estar utilizando estrategias distintas e incluso contrapuestas. Se aprecian diferencias significativas en la duración de los registros en los locutores L1, L2, L4 y L5. En los locutores L1 y L5 la duración de las frases en el registro modal es mayor a la duración de las frases en *falsetto*; por tanto, la velocidad de habla resulta ser más lenta en el registro modal para estos dos hablantes. De modo contrario, en el caso de los locutores L2 y L4, la duración en el *falsetto* es mayor a la registrada en modal; en consecuencia disminuyen de forma significativa la velocidad cuando recurren al disimulo trámite *falsetto*. Este comportamiento dispar podría estar indicando que los locutores difieren en las estrategias utilizadas, probablemente de modo involuntario o irreflexivo, para hacer frente a la complejidad que seguramente supone mantener, aun en frases breves, la configuración laríngea de una voz falsada.

4.2.3. Interacción entre el tempo global y el tempo de cada registro

Tras analizar la interacción entre el *tempo* global y el de cada registro por separado, se observó que tres locutores destacaban en relación con el *tempo* global: el Locutor 4, por hablar significativamente más lento que el resto; y los locutores L5 y L6 por hacerlo más rápidamente. Como se ha visto, el Locutor 6 no presentaba diferencias

internas entre la duración de las frases pronunciadas con voz modal y las pronunciadas con *falsetto*, algo que sin duda, debería de contribuir positivamente al reconocimiento del locutor. La presunción es que si hubiera diferencias significativas entre el registro modal y el *falsetto* de los locutores y estas diferencias se percibiesen, esto complicaría o impediría reconocer al sujeto. Sin embargo, se ha visto que los locutores L4 y L5 destacan por la duración global de sus frases —siendo el primero es más lento, y el segundo, más rápido, junto con L6— pero presentan variabilidad intralocutor, puesto que las duraciones en la voz modal y en el *falsetto* se diferencian significativamente entre sí. No obstante, se da una particularidad en el L4 y es que la duración en su registro modal es significativamente mayor que la duración que se aprecia en todos los registros modales del resto de locutores (*L4-L1; **L4-L2 y *** L4-L3, L4-L5, L4-L6) y la duración en su *falsetto* es también significativamente mayor que la del resto de los *falsetto* (*L4-L2 y ***L4-L1, L4-L3, L4-L5, L4-L6).

4.3. Observaciones en torno a posibles características segmentales y prosódicas del habla de los locutores

En el Capítulo 3 correspondiente a la metodología, en concreto, en el apartado §3.2.2, se explicaron las razones que motivaron la elección de la pseudofrase inventada, las cuales respondían fundamentalmente a dos cuestiones fundamentales: 1) la posibilidad de una futura comparación con las frases del corpus CIVIL y 2) la búsqueda por aislar razonablemente y de forma natural, es decir, no manipulada, la impresión perceptiva que pudiera provocar en un oyente la cualidad de la voz, más concretamente, aquella derivada del cambio de registro modal-*falsetto*. La finalidad que se persiguió en todo momento fue la de tratar de constreñir a los oyentes que realizaran la tarea perceptiva de discriminación de locutores a que, durante su valoración auditiva, tuvieran que centrarse sobre todo en aspectos laríngeos presumiblemente coincidentes, identificables o reconocibles en la voz modal y en la voz disimulada con *falsetto* de un mismo locutor. Como era de prever, podrían manifestarse, incluso en pseudofrases breves, algunas claves entonativas y, por ello, se suprimieron los finales de todos los enunciados. Asimismo, podría esperarse que emergieran, como ha ocurrido, diferencias en la velocidad de habla de los locutores que pudieran, por tanto, constituir también

pistas (o despistes) perceptivos para los oyentes. Las diferencias en la velocidad del habla que fueron constatadas (§4.2) prefirieron no controlarse mediante una manipulación posterior de los estímulos grabados, porque, según se ha señalado en la bibliografía, podrían ser características del pasaje de voz modal a *falsetto*, es decir, consecuencia de ese cambio de registro y, por tanto, concomitantes del *falsetto*.

Las grabaciones de las 10 frases pronunciadas por los seis locutores fueron doblemente evaluadas a fin de identificar si en ellas estaban presentes rasgos segmentales o suprasegmentales perceptivamente evidentes, al margen de los temporales y entonativos ya señalados. Una primera evaluación perceptiva de los audios se realizó por cuatro miembros del equipo de investigación del Laboratorio de Fonética del CSIC, todos ellos hablantes nativos de español. La segunda evaluación se llevó a cabo por 3 expertos fonetistas pero hablantes nativos de italiano.

Los cuatro fonetistas españoles constataron diferencias en la velocidad de habla para los locutores L4 y L6, en el mismo sentido en que fue descrito en §4.2.1, es decir, percibieron en términos globales que el Locutor 4 hablaba siempre más lentamente y el Locutor 6 lo hacía de forma más rápida. Tres de ellos, además, observaron que el mismo Locutor 4 pronunciaba las obstruyentes oclusivas con un reforzamiento y que esta característica se podía advertir en su voz modal y en su *falsetto*, como había sido señalado en Fernández Trinidad y Rojo (2018).

De otra parte, los tres fonetistas italianos, en una escucha igualmente atenta y minuciosa de los audios, observaron muy pocos rasgos peculiares que enseguida se detallarán y, muy importante, no hubo entre ellos consenso o unanimidad en sus juicios, es decir, no resaltaron los mismos rasgos fonéticos concretos al describir el perfil personal de cada hablante. Sobre el único aspecto en que coincidieron todos fue en afirmar que perceptivamente no se apreciaban rasgos de pronunciación evidentes que, *a priori*, permitieran reconocer más fácilmente a un locutor que otro.

Dos de los fonetistas italianos señalaron que unas pocas frases pronunciadas por el Locutor 3 podían llegar a percibirse ligeramente diferentes de las pronunciadas por el resto de los locutores si se atendía a algunos aspectos prosódicos. Aparentemente, su prosodia no respondería al patrón napolitano promedio. En una de las repeticiones de las frases, apuntaron, el locutor no parece haber realizado, o al menos no se puede percibir fácilmente, la duplicación fonotáctica [ev:ada] tan característica del habla napolitana o, incluso, meridional. Este aspecto concreto, sin embargo, no llamó la

atención del tercer fonetista italiano, quien solamente señaló una particularidad potencialmente distintiva para el Locutor 5. Este locutor, en opinión del tercer fonetista italiano consultado, parecía tener un ritmo "mecánico", como si pronunciara las palabras sílaba a sílaba. Este aspecto, expresa, afecta al ritmo en el sentido de que en la variedad regional del centro-sur de Italia existe reducción vocálica (como duración y como timbre) y, por tanto, el L5 se podría percibir con un ritmo diferente al presentar menor reducción vocálica.

Estas valoraciones en su conjunto avalan, por tanto, el hecho de que en las grabaciones de los seis locutores no estuvieran presente rasgos segmentales o suprasegmentales perceptivamente muy evidentes, al margen de los temporales o de los derivados del comportamiento laríngeo que ya se comentaron.

Capítulo 5
Resultados del estudio perceptivo de discriminación de voces

5. Resultados del estudio perceptivo de discriminación de voces

El presente Capítulo 5 presenta los análisis y comenta los resultados obtenidos de las pruebas perceptivas realizadas para observar cómo se desenvuelven los oyentes italianos y españoles en la discriminación de locutores italianos con voces disimuladas en *falsetto*.

Como se verá en detalle, se efectuaron análisis en varios niveles diferentes, la mayoría de ellos en el marco de la *Teoría de Detección de Señales*, razón por la cual se ha decidido describir someramente este modelo (§5.1) para facilitar la lectura de los resultados alcanzados en este trabajo. Primeramente, se estudió la exactitud de las respuestas ofrecidas por los dos grupos de oyentes, clasificándolas en Aciertos, Rechazos Correctos, Omisiones y Falsas Alarmas (§5.2.1). Posteriormente, se llevaron a cabo los análisis que dan cuenta del grado de discriminabilidad de las voces y el criterio de respuesta seguido por los jueces, según la *Teoría de Detección de Señales* (§5.2.2). También se analizaron los tiempos de reacción (TR) en función del tipo de respuesta, de la lengua de los jueces y de los estímulos presentados (§5.3). Por último se examinaron los resultados teniendo en cuenta los distintos locutores (§5.4).

5.1. *Teoría de la Detección de Señales* (TDS)

Como explican Ballesteros (1997, pp. 1-5) y Reales y Ballesteros (1995, p. 15), a diferencia de los clásicos modelos utilizados en psicofísica, la *Teoría de la Detección de Señales* (Tanner y Swets, 1954 y Green and Swets, 1966) propone considerar, además de la estimulación física, los factores personales de los propios sujetos. Esta necesidad responde a que se ha demostrado que la cantidad o intensidad de estimulación física no es el único factor determinante de la detección de una señal sino que factores personales del observador (como la experiencia, la motivación, las implicaciones que su decisión tiene para el sujeto, entre otros que se podrían mencionar) influyen en las respuestas de los sujetos. Como se expresa en Ballesteros (1997):

> Según la *Teoría de la Detección de Señales*, la captación de la información relevante del medio y la toma de decisiones de los seres humanos sobre la detección o no de una señal va a depender, por un lado, del nivel de ruido que acompañe a la señal y, por otro, de las consecuencias e implicaciones que tenga para el individuo esta decisión […] (p. 5).

Por tanto, continúa Ballesteros, la actuación de un ser humano frente a un estímulo no vendrá determinada exclusivamente por la calidad o intensidad del estímulo que se le presente sino que también jugarán un importante papel los factores cognitivos o psicológicos del propio observador, y estos factores dependientes del sujeto pueden incluso, en ciertas circunstancias, ser más determinantes que la propia intensidad del estímulo (1997, p. 6)[41].

Antes de presentar los análisis realizados, se ofrece una breve descripción de los principales presupuestos de dicha teoría en aras de una mejor comprensión de los resultados obtenidos.

5.1.1. Tipos de respuesta

En la *Teoría de la Detección de Señales* (en adelante, TDS), se le denomina "señal" (S)[42] al estímulo que el sujeto percibe y detecta, y "ruido" (R) a los demás estímulos que aparecen junto con la señal. El ruido obstaculiza la detección de la señal y, en consecuencia, complica su discriminación; por este motivo, los observadores suelen cometer errores en sus juicios (véase por ejemplo, Ballesteros, 1997).

Según los postulados de la TDS existen cuatro tipos de posibles respuestas determinadas por el tipo de ensayo presentado (señal/ruido) y por el tipo de respuesta dada por el sujeto observador (sí/no), como se esquematiza en la Tabla 5.1 (véanse por ejemplo, Macmillan y Creelman, 1991, Reales y Ballesteros, 1995).

[41] En Goldstein 2006 [1988], pp. 583-584) se ofrece también una clara explicación a este respecto.
[42] El ruido se considera como algo constante; la señal siempre estará acompañada de un cierto ruido (señal +ruido), por ello la discriminación del observador nunca es perfecta (Ballesteros, 1997, p. 16).

Tabla 5.1: Esquema de las posibles respuestas según el modelo de la *Teoría de Detección de Señales*.

		S (señal)	**R (ruido)**
Respuesta	**Sí**	Acierto (Sí/S)	Falsa Alarma (Sí/R)
	No	Omisión (No/S)	Rechazo Correcto (No/R)

Las respuestas sí/no ante la presencia de la señal (S) o la ausencia de la misma (R) dan lugar a cuatro posibles situaciones como se aprecia en la Figura 5.1. De una parte, dos clases de aciertos o decisiones correctas: Aciertos y Rechazos Correctos y, de otra, dos tipos de fallos, que podrán ser Falsas Alarmas u Omisiones. Los Aciertos (A) se dan toda vez que el observador dice detectar la señal y la señal estaba presente. Los Rechazos Correctos (RC) ocurren cuando el observador responde que no ha detectado la señal y esta, en efecto, no estaba presente. Las Omisiones se producen cuando el sujeto dice no haber detectado la señal a pesar de que sí estaba presente y, finalmente, las Falsas Alarmas son también errores que comete el sujeto al haber afirmado que la señal estaba presente cuando en realidad no lo estaba. De este esquema también se deduce que los Aciertos y las Omisiones son complementarios, dado que son respuestas que se dan en presencia de la señal; al tiempo que las Falsas Alarmas y los Rechazos Correctos también se complementan en la medida en que ambas son respuestas de los sujetos observadores ante la presencia de ruido (por ejemplo, Ballesteros, 1997).

Como se explica en Macmillan y Creelman (1991) y en Reales y Ballesteros (1995), una tasa de Aciertos del 100 % se podría alcanzar si el sujeto observador respondiera siempre que "sí"; sin embargo, este comportamiento también proporcionaría un 100% de Falsas Alarmas. De igual modo, una tasa de 0 % de Falsas Alarmas podría conseguirse si contestara siempre que "no", pero, una vez más, esta estrategia arrojaría un 0% de Aciertos. Por este motivo, en TDS se propone el cálculo de la proporción de Aciertos y Falsas Alarmas como una medida más completa y fiable para evaluar el desempeño de los sujetos.

5.1.2. Discriminabilidad o d prima

La *d prima* es el índice de sensibilidad más importante de esta teoría y se considera una buena medida para evaluar la capacidad del observador para detectar una señal determinada,

pues considera tanto Aciertos como Falsas Alarmas: la discriminabilidad aumenta conforme aumenta la tasa de Aciertos y disminuye la de Falsas Alarmas (véanse Macmillan y Creelman, 1991, pp. 9-10 y Reales y Ballesteros, 1995, p. 30). Si el valor de la *d'* es igual a 0, esto significará que el número de Aciertos y de Falsas Alarmas ha sido el mismo e indicará que la sensibilidad del observador es nula, que el rendimiento del sujeto ha sido igual al azar, y por tanto, que el sujeto es incapaz de discriminar el estímulo "señal". Si el valor de la *d'* es negativo querrá decir que la proporción de Falsas Alarmas ha sido mayor que la de Aciertos. Los valores de *d'* positivos significan que la proporción de Aciertos ha sido mayor a la de Falsas Alarmas e indican distintos niveles de discriminación, por ejemplo: *d'* =1 (69 % de A y 31 % de FA); *d'* = 1.35 (75 % de A y 25 % de FA) y *d'* = 4.65 (99 % de Aciertos y 1 % de FA, se considera el efecto techo, Macmillan y Creelman 1991, 9-10).

5.1.3. Sesgo de respuesta o índice c

Como se ha apuntado antes (§5.1), la actuación de un sujeto no depende exclusivamente de su sensibilidad para discriminar entre la señal y el ruido sino que está también influida por factores psicológicos que determinan su criterio de respuesta o de decisión (conservador/neutral/liberal). En una tarea de discriminación, los jueces pueden preferir un tipo de respuestas frente a otro. Por ello en la TDS se calcula, además del índice de discriminabilidad (*d'*), el criterio que los sujetos han tenido para dar sus respuestas. Es el *sesgo de respuesta*, identificado con el parámetro *c*[43]. Como explican Macmillan y Creelman (1991, p. 33) y Reales y Ballesteros (1995, p. 32), este parámetro es la medida básica de sesgo en TDS y representa la tendencia de los jueces a sesgar sus respuestas en una dirección u otra, hacia el "sí" o hacia el "no". Si el valor de *c* es igual a 0 se considera que no hay sesgo, entonces no se observa una preferencia del sujeto observador por dar un tipo de respuesta. Si los valores de *c* son negativos se interpretan como un sesgo hacia el "sí" y se denomina *criterio liberal*. Si en cambio, los valores de *c* son positivos, se interpretan como un sesgo hacia el "no", denominado *criterio conservador*.

[43] Aunque actualmente se considera que el índice *c* representa la mejor forma de medir el sesgo de respuesta (tendencia del observador a favorecer un tipo de respuesta frente a otra) otra medida muy utilizada fue la *razón de verosimilitud* o *ß* (véanse , Macmillan y Creelman 1991, pp. 35-36 y Reales y Ballesteros, 1995, pp. 32-33).

5.1.4. Curvas ROC

Los datos que se obtienen al realizar un experimento aplicando TDS pueden expresarse por medio del gráfico denominado *Curva Característica de Operación del Receptor (COR)*, aunque suelen utilizarse más a menudo las siglas ROC, tomadas del nombre en inglés *Receiver Operating Characteristic*. Según se explica en Macmillan y Creelman (1991, p. 42) y Reales y Ballesteros (1995, p. 34), estas curvas se consiguen dibujando la proporción de Aciertos y la de Falsas Alarmas y esta relación se grafica mediante el trazado de dos ejes: el eje vertical para los Aciertos y el horizontal para las Falsas Alarmas (ver Figura 5.1). Cada punto de la curva corresponde a las proporciones de Aciertos y de Falsas Alarmas obtenidas variando progresivamente el criterio de decisión del observador (Macmillan y Creelman 1991, p. 42; Reales y Ballesteros, 1995, p. 34).

Como se comentó antes en §5.1.2, una detección ineficiente tendrá la misma tasa de Aciertos que de Falsas Alarmas. La detección nula (*d'*= 0) se representa en el gráfico ROC mediante la diagonal llamada *línea del azar* (véase, por ejemplo, Ballesteros, 1997 y Reales y Ballesteros, 1995, pp. 34-65).

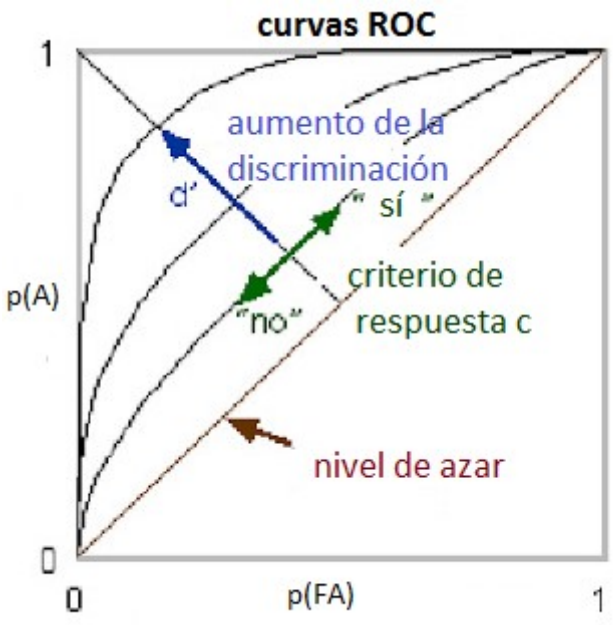


Figura 5.1. Información proporcionada por las curvas ROC en TDS. Imagen adaptada de Manzanero (http://psicologiapercepcion.blogspot.com.es/p/psicofisica-sensorial.html).

Entonces, la diagonal principal en el espacio ROC representa el nivel del azar, es decir, cuando la proporción de Aciertos y Falsas Alarmas es la misma y la discriminabilidad es nula. Según se explica por ejemplo en Reales y Ballesteros (1995, p. 34), "cada punto de la *curva ROC* corresponde a las proporciones de Aciertos y de Falsas-alarmas obtenidas variando el criterio de decisión del observador". En el marco de la TDS, la medida de discriminabilidad o *d prima* representa la distancia entre un punto específico de una curva ROC y la diagonal principal. A medida que la *d'* se incremente, esto es, a medida que la discriminación sea mejor (aumente la proporción de Aciertos y disminuya la de Falsas Alarmas), la curva se irá alejando de esa línea diagonal y se irá acercando a la esquina superior izquierda, haciéndose cada vez más prominente o convexa. La detección será mejor cuando más se aproxime la curva a esa esquina superior izquierda y la curva resulte más abombada (véase Ballesteros 1997, y Reales y Ballesteros 1995, p. 34-35).

5.2. Análisis de los resultados perceptivos en el marco de la *Teoría de Detección de Señales*

Como se ha dicho, la TDS ha sido el marco teórico seleccionado para estudiar las respuestas de los oyentes en el test de discriminación de voces realizado como prueba perceptiva. Recuérdese que el diseño escogido para la prueba perceptiva fue, como se indicó en el Capítulo 3, un test de discriminación igual-diferente AX. A todos los oyentes se les presentaron 30 parejas de voces del mismo locutor y un mismo número de parejas de distinto locutor. Tanto en las parejas iguales (AA) como en las diferentes (AB), siempre se oía en primer lugar una voz pronunciando la pseudofrase "Dica dadiva adagio e vada a" en registro modal y en segundo, la voz disimulada en *falsetto* pronunciando la misma frase. El orden de presentación de los 60 pares fue aleatorio y fijo para todos los oyentes, a fin de que todos ellos se enfrentaran a las mismas combinaciones de estímulos, esto es, a la misma dificultad.

El concepto de "señal" se relaciona con las demandas de la tarea, es decir, lo que se le ha pedido a los jueces que detecten. Por tanto, en este diseño experimental la señal se corresponde con los pares de locutores iguales (AA) y el ruido, es decir, los "cebos" que se proponen como alternativa a la señal, con los pares de locutores distintos (AB). Frente a estos ensayos de señal y ruido, 30 de cada uno, los jueces debían detectar la señal en ambas condiciones tomando una decisión dicotómica (sí/no) frente a la pregunta *¿es el mismo locutor el que habla?*

En el contexto de la TDS, las respuestas de los jueces pueden caer, como se ha explicado antes, en cuatro categorías. Si se les había presentado un par de voces del mismo locutor (condición de señal presente) y los oyentes respondieron "sí", sus respuestas se contabilizaban como Aciertos (A). Si en esta misma condición de señal presente (pares de igual locutor) los oyentes respondieron "no", se contabiliza como una Omisión (O). Si, en cambio, la señal no estaba presente y se los había enfrentado a la condición de solo ruido (pares de distinto locutor) y los oyentes respondieron "sí", se contabiliza como una Falsa Alarma (FA), y, si respondieron "no", como un Rechazo Correcto (RC). Tal y como se puede observar, las proporciones de Aciertos y de Falsas Alarmas varían de forma independiente entre sí, ya que ambas reflejan puntuaciones con respecto a diferentes clases de ensayos o condiciones, "señal" y "ruido" respectivamente. Así, los Aciertos representan las respuestas correctas en la condición de "señal", al tiempo que las Falsas Alarmas representan las respuestas incorrectas en la condición de "ruido".

Para evaluar el efecto de la variable "lengua" en las respuestas de los oyentes se llevaron a cabo dos tipos de análisis. Por un lado, se estudiaron las medidas de exactitud de las respuestas (Aciertos, Falsas Alarmas, Omisiones y Rechazos Correctos) y, por otro, las medidas de discriminabiliad (*d'*) y criterio de respuesta (*c*).

5.2.1 Medidas de exactitud de las respuestas

Se analizaron los resultados en función de la exactitud de las respuestas en el grupo de oyentes españoles e italianos a partir de los datos obtenidos en los dos test realizados. Las puntuaciones medias y las desviaciones típicas para cada tipo de respuesta pueden observarse en las Tablas 5.2 (oyentes españoles) y 5.3 (oyentes italianos).

Tabla 5.2: Valores medios y desviaciones típicas (entre paréntesis) obtenidos para las medidas de exactitud en el grupo de jueces españoles, calculados a partir de 3600 casos (n=3600).

		ESTÍMULO	
		SEÑAL (AA)	RUIDO (AB)
RESPUESTA	SÍ	Aciertos 20.4 (3.9)	Falsas Alarmas 12.1 (3.9)
	NO	Omisiones 9.6 (3.9)	Rechazos Correctos 17.9 (3.9)

Tabla 5.3: Valores medios y desviaciones típicas (entre paréntesis) obtenidos para las medidas de exactitud en el grupo de jueces italianos (n=4800).

		ESTÍMULO	
		SEÑAL (AA)	RUIDO (AB)
RESPUESTA	SÍ	Aciertos 20.8 (3.9)	Falsas Alarmas 10.5 (3.6)
	NO	Omisiones 9.2 (3.9)	Rechazos Correctos 19.5 (3.6)

Como se aprecia en las Tablas 5.2 y 5.3, los españoles cometen mayor número de Falsas Alarmas, una media de 12.1 de FA por sujeto, frente a 10.5 de los italianos. La prueba test-t indica que esta diferencia es estadísticamente significativa: $t(138)=2.57$, $p<0.05$ *. El número de Aciertos es ligeramente menor en los españoles, pero la diferencia no es significativa: $t(125.5)= -0.62$, ns.

Por si su interpretación resultara más intuitiva se ofrecen también, para cada grupo de jueces, la proporción de Aciertos sobre los 30 ensayos en los que aparecía la señal (es decir, las parejas AA), y la proporción de Falsas Alarmas sobre los 30 ensayos de ruido (parejas AB). Los resultados aparecen en la Tabla 5.4.

Tabla 5.4: Proporción de Aciertos y Falsas Alarmas (y desviaciones típicas) en los grupos de jueces españoles e italianos.

	españoles	**italianos**
proporción de Aciertos	0.68 (0.13)	0.69 (0.13)
proporción de Falsas Alarmas	0.40 (0.13)	0.34 (0.12)

5.2.2. Medidas de discriminabilidad (d') y criterio de respuesta (c)

Se han calculado para los dos grupos de jueces las medidas *d'* y *c* y los resultados se muestran en la siguiente tabla.

Tabla 5.5: Valores medios y desviaciones típicas (entre paréntesis) de *d'* y *c* para los jueces españoles e italianos.

	Españoles	**Italianos**
d'	0.77 (0.59)	0.95 (0.58)
c	−0.13 (0.24)	−0.07 (0.25)

La *d prima* promedio en los españoles es de 0.77 y, en los italianos, resultó ser ligeramente mayor, con una media de 0.95. Sin embargo, estas diferencias entre los grupos no alcanzan significatividad estadística: t(138)=1.86 p=0.07 ns.

En la Figura 5.2 se puede observar la distribución de los valores de la *d* prima para el grupo de oyentes italianos y para el de españoles.

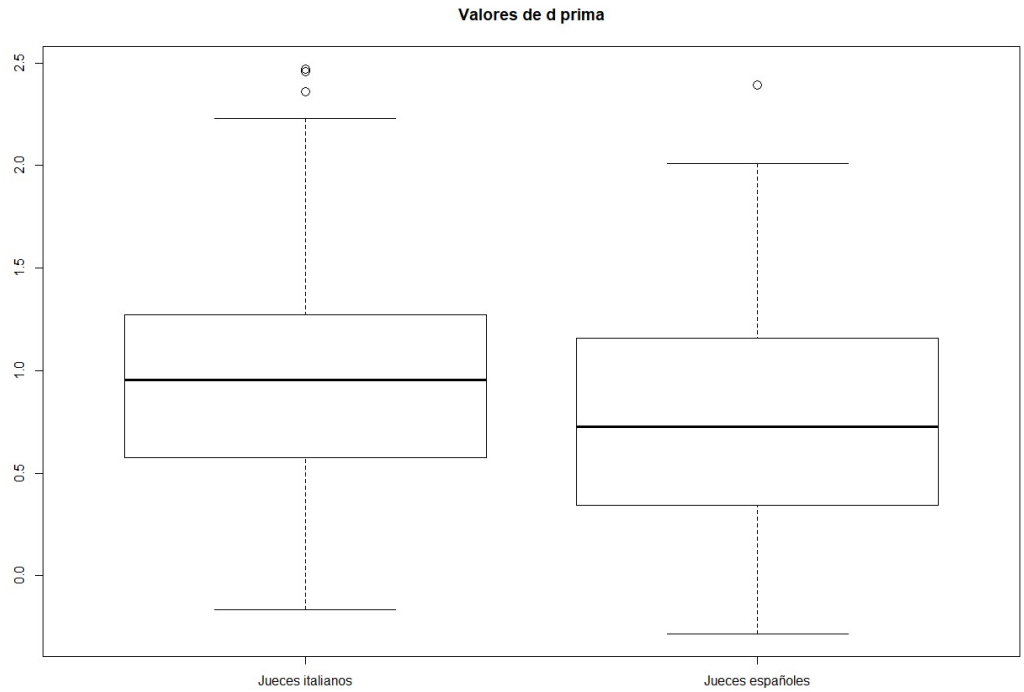


Figura 5.2. Distribución de los valores de la d prima en función de la lengua materna de los jueces (italianos y españoles).

El valor medio de *c* para los italianos es de -0.07 y el de los españoles de -0.13; sin embargo, una vez más, los resultados del test-t no arrojan diferencias significativas: t(138)=1.33, p=0.19 ns. La tendencia a una discriminabilidad más baja y a un criterio más liberal de los españoles podría explicar el peor resultado que tienen estos jueces cuando no se les presenta la señal, es decir, en los ensayos de solo ruido, lo cual se traduce en un mayor número de Falsas Alarmas. Aun así, ni las diferencias en los valores medios de *d'* ni las de los de *c* resultan ser significativas, como se ha dicho.

Por tanto, de los datos solo se puede concluir que tanto los españoles como los italianos han demostrado una capacidad similar para discriminar las voces. No existen diferencias significativas entre grupos (a excepción de las Falsas Alarmas); la discriminabilidad es débil (valores medios *d prima* cercanos a 1) aunque es significativamente superior al azar porque los valores son significativamente distintos de 0. La estrategia de respuesta también ha sido la misma para ambos grupos con un índice de *c* medio <0, lo cual indica un *criterio liberal* en las respuestas de ambos grupos, es decir, una tendencia a responder "sí". Los resultados obtenidos de la aplicación de TDS se representan en los gráficos ROC que aparecen en las Figuras 5.3 (para los españoles) y 5.4 (para los italianos).

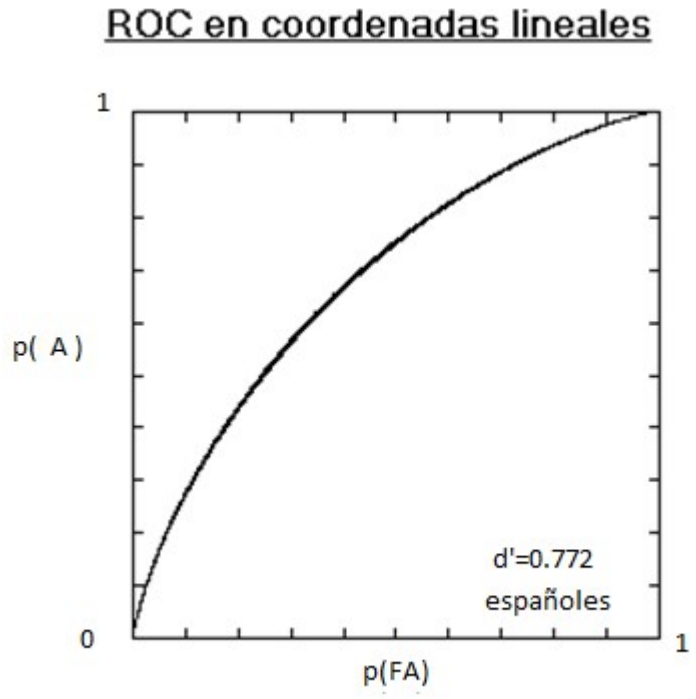


Figura 5.3. Curva ROC para los oyentes españoles.

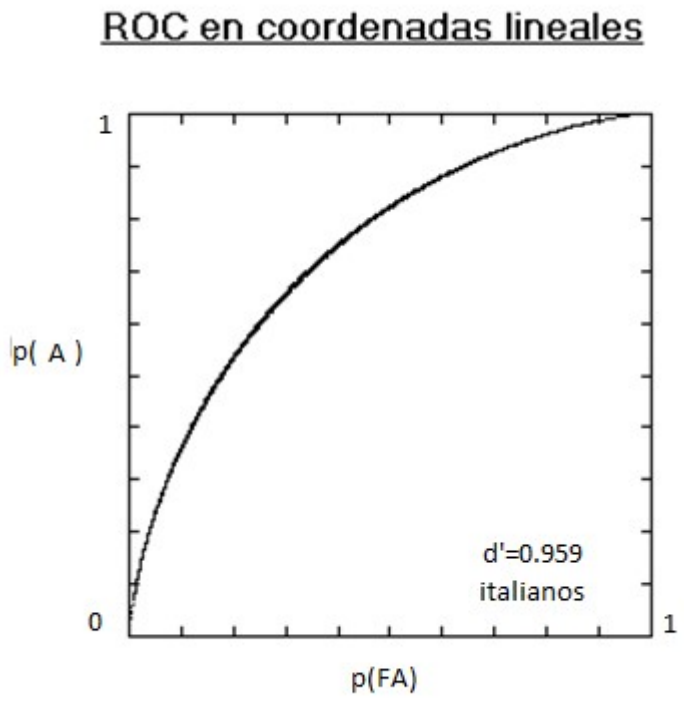


Figura 5.4. Curva ROC para los oyentes italianos.

Debe aclararse que al explorar uno a uno los datos del análisis de *d prima* se observó que 6 jueces (3 españoles y 3 italianos) habían obtenido valores negativos, y otros cinco (3 españoles y 2 italianos), valores de *d*' iguales a 0. Dichos resultados, ordenados de menor a mayor, se presentan en la Tabla 5.6.

Tabla 5.6: Jueces italianos (5) y españoles (6) cuyos valores de *d prima* fueron iguales a cero o negativos.

jueces con *d prima* negativa o cero		
juez	**lengua**	**valor de la *d prima***
crs	spa	−0.59
rql	spa	−0.28
gmplllgr	ita	−0.16
ntngrss	ita	−0.09
dnlsclbrno	ita	−0.08
strlpzdlcs	spa	−0.08
lglm	ita	0
ttvicldbngr	ita	0
nlvrzdlrcsls	spa	0
lc	spa	0
lrrqstnsrrt	spa	0

Los valores de *d prima* negativa responden a una mayor proporción de Falsas Alarmas (FA) que de Aciertos (A), lo que significa que el número de veces que estos seis jueces respondieron que la señal estaba presente (es decir, que respondieron que las voces modal y *falsetto* pertenecían al mismo locutor) en la condición de ensayo de ruido (parejas de locutores distintos) fue mayor que el número de veces que respondieron que la señal estaba presente cuando realmente lo estaba (parejas de locutores iguales). Esto sugiere que los seis jueces referidos no entendieron bien la consigna, puesto que no es cierto que los estímulos de locutores diferentes fueran más parecidos entre sí que los estímulos de locutores iguales. Los valores de *d prima* igual a 0 de los otros cinco jueces se explica, observando sus respuestas, porque han alcanzado igual número de Aciertos que de Falsas Alarmas, lo cual indica que su discriminabilidad ha resultado ser igual al azar.

De todas formas, se realizó un análisis excluyendo los datos de los seis jueces que presentaron valores negativos de *d'* y se observaron resultados prácticamente iguales. El resultado para los italianos fue de 0.96 (0.59) y para los españoles de 0.78 (0.57). Por este motivo y ante la falta de razones objetivas para eliminar los datos de los seis jueces con *d*' negativa, se decidió mantener los datos de los 140 jueces para todos los análisis.

5.3. Análisis de los resultados relativos a los tiempos de reacción (TR)

Se analizaron asimismo los tiempos de reacción (TR) que conllevaban las respuestas que dieron los oyentes en los test perceptivos y para ello se creó primero una tabla con todos los valores obtenidos, ordenados de menor a mayor y expresados en milisegundos (ms). A continuación, se muestran en la Tabla 5.7 las observaciones con valores de tiempos de reacción negativos o iguales a cero y también un valor desmesuradamente elevado, todos ellos debidos a problemas técnicos durante la ejecución del test. Por este motivo, estos seis casos fueron eliminados del análisis posterior sobre los tiempos de reacción, quedando un total de 8394 casos entre los dos grupos de jueces (4798 casos juzgados por italianos y 3596 casos juzgados por españoles).

Tabla 5.7: Valores relativos al TR (ms) que se han tenido que desestimar debidos a un error técnico en el desarrollo de la prueba perceptiva.

juez	lengua	caso	TR (ms)
vrnc	spa	3	−942
jhnndrsn	spa	10	−717
rncndls	spa	6	−30
ttvicldbngr	ita	31	0
rcs01	spa	18	0
lccppl	ita	31	123803

5.3.1. Tiempos de reacción en función del tipo de respuesta, del grupo de oyentes y del tipo de estímulo presentado

Como es habitual en muchos estudios psicolingüísticos, se transformaron los tiempos de reacción (TR) mediante logaritmos naturales (o neperianos)[44] para evitar los valores periféricos extremos, que únicamente se producen por la cola positiva, ya que no son posibles valores negativos.

Tabla 5.8: Valores medio y desviación típica (entre paréntesis) de los tiempos de reacción (logTR) en función de los cuatro tipos de respuesta para la totalidad de los jueces, sin discriminar grupos.

	(logTR)
Aciertos	6.81 (0.81)
Falsas Alarmas	7.05 (0.90)
Omisiones	7.07 (0.91)
Rechazos Correctos	6.79 (0.82)

Se encontró significatividad en el efecto principal del tipo de respuesta sobre logTR: $F(3, 8391)=58.08$, $p<0.001$***. Para comprobar las parejas de niveles significativamente distintos, se aplicó un análisis post-hoc con corrección de Bonferroni, y resultó que, salvo en el

[44] Tal conversión se realizó con las funciones que ofrece *Excel*.

caso de la pareja de FA-O y la de A-RC, todas las demás comparaciones resultaron significativas ($p<0.001$***). Por lo tanto, se observa que el logTR es significativamente distinto entre cualquier tipo de respuesta "acertada" y cualquier tipo de respuesta "errónea".

Se calcularon las medias y desviaciones típicas de los logTR por grupos de jueces: italianos 6.96 (0.90), españoles 6.80 (0.79). Se encontró que los italianos tardaron significativamente más en emitir sus respuestas, $F(1,8392)=76.58$, $p<0.001$***. El efecto principal de la lengua y de los tipos de respuesta en los TR se grafica en la Figura 5.5.

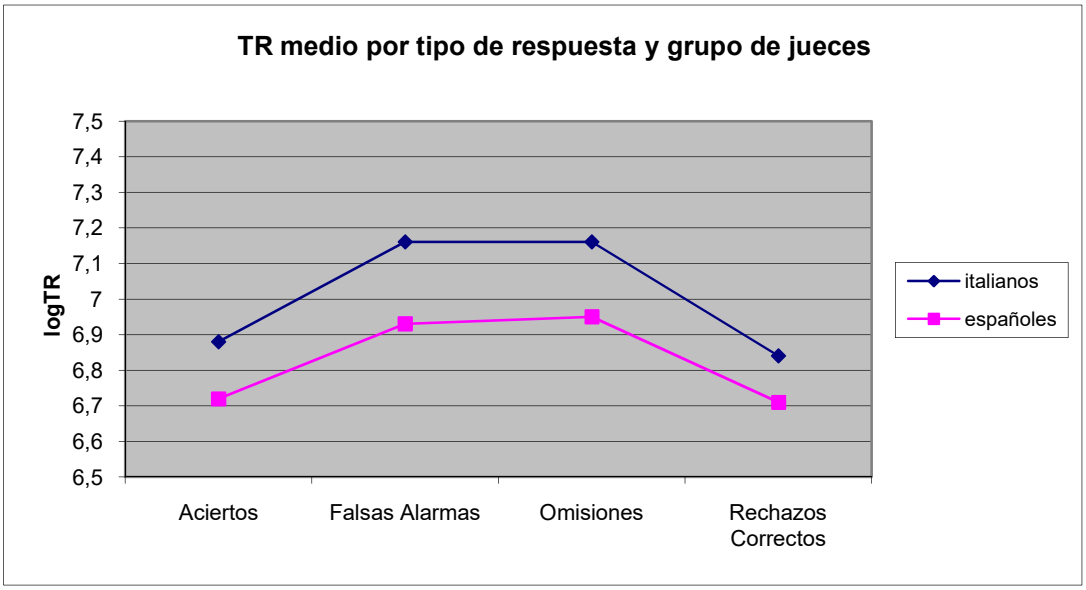


Figura 5.5. Tiempos de reacción (LogTR) medio por tipo de respuesta (A, FA, O y RC) y por grupo de jueces (italianos y españoles).

Un ANOVA de dos factores demostró que la interacción entre el tipo de respuesta y la lengua nativa de los jueces no es significativa: $F(3,8386)=1.38$,$p=0.25$ ns.

Tampoco se halló un efecto significativo del hecho de presentar como estímulos parejas de igual locutor (AA) o de locutores diferentes (AB): $F(1,8392)=0.02$,$p=0.89$ ns.

5.4. Análisis de los resultados en función de los locutores

A continuación se estudia la posibilidad de que algún locutor haya sido mejor reconocido que otro. Para ello se creó una tabla de contingencia para realizar el chi2. En una línea se calculó el número de respuestas acertadas para todos los estímulos iguales (AA), es decir, el número de Aciertos (A), por locutor. En otra línea se calculó el número de respuestas incorrectas o fallidas para los estímulos AA, es decir el número de Omisiones (O), por locutor.

Estos datos se fusionaron en una tabla para hacer el chi2. Los resultados muestran que hay un efecto principal de Locutor, $\chi^2(5)=30.68$, $p<0.001$ ***. Para conocer qué niveles concretos resultaron significativamente distintos del resto, se realizó un post-hoc Bonferroni y el resultado indica que se trata del Locutor 6, L6, $p<0.01$ **.

Luego se repitió el mismo análisis de chi2 pero separando las dos lenguas. Para los italianos, los resultados fueron: $\chi^2(5)=25.02$, $p<0.001$ ***, post-hoc Bonferroni: L6, $p<0.01$ **. Para los españoles, $\chi^2(5)=8.17$, $p=0.15$ ns. Es decir, que no hay efecto principal de Locutor en el grupo de jueces españoles.

Finalmente, en la siguiente tabla se ofrecen las medias (y desviaciones típicas) de la proporción de Aciertos para los jueces de ambos grupos fusionados, para los jueces españoles y para los jueces italianos.

Tabla 5.9: Proporción de Aciertos obtenida para cada Locutor sobre el total de respuestas AA (igual locutor). Medias (y desviaciones típicas) entre jueces (de ambos grupos fusionados), jueces españoles, y jueces italianos.

	grupos fusionados	**españoles**	**italianos**
L1	0.71 (0.24)	0.70 (0.23)	0.71 (0.25)
L2	0.68 (0.24)	0.67 (0.24)	0.68 (0.24)
L3	0.66 (0.23)	0.68 (0.23)	0.65 (0.23)
L4	0.67 (0.28)	0.67 (0.28)	0.68 (0.29)
L5	0.64 (0.24)	0.63 (0.24)	0.66 (0.24)
L6	0.76 (0.25)	0.73 (0.25)	0.79 (0.25)

Capítulo 6
Discusión final de los resultados (producción y percepción). Conclusiones y preguntas de investigación futura

6. Discusión final de los resultados (producción y percepción), conclusiones y preguntas de investigación futura

En este capítulo se discuten globalmente los resultados y se esbozan las principales conclusiones y aportaciones que se extraen de la tesis así como algunas futuras líneas de investigación que se abren a partir de ella.

El análisis de la producción vocal que se realizó sobre las muestras de voz de los 6 locutores que participaron en el experimento ha permitido profundizar en la especificación de las características del registro de *falsetto* y establecer sus principales diferencias con respecto a la voz modal. De una parte, el estudio pormenorizado de los aspectos acústicos de las voces permitió observar qué rasgos del comportamiento glotal se vieron alterados y cuáles no en el cambio de registro. Luego, se pudieron aislar los factores definitorios del registro de *falsetto* y derivar cuáles podrían vincularse quizás con la cualidad individual de la voz. Estos últimos deberán ser tratados en estudios futuros con mayor profundidad de análisis.

El presente capítulo se organiza de la siguiente forma. Primeramente se comentan los resultados derivados del estudio de la producción de la voz y del habla, explicados en detalle en el Capítulo 4, y se interpretan a la luz del conocimiento que hasta el momento se tiene sobre los rasgos que definen y caracterizan al *falsetto* (§6.1). En el siguiente apartado se razonan los resultados obtenidos de las dos pruebas perceptivas de discriminación de voces explicadas en el Capítulo 5 y, nuevamente, se interpretan en relación con lo que se sabe hasta el momento sobre el perjuicio que el *falsetto* causa en la discriminación de locutores (§6.2) y la posible influencia positiva que en la discriminación de voces tiene el hecho de que locutor y oyente compartan la lengua (§6.3). Finalmente, se presentan algunas consideraciones en torno a los tiempos de reacción (§6.4).

6.1. Sobre los rasgos que definen y caracterizan al *falsetto*

6.1.1. Análisis de la frecuencia fundamental (f_0)

Cuando los 6 locutores masculinos analizados recurrieron al *falsetto*, el aumento promedio realizado por cada uno de ellos fue de 1.5 octavas respecto de sus correspondientes valores de frecuencia fundamental en la voz modal. El locutor que consiguió llegar más alto rozó incluso las 2 octavas (1 octava y 8 semitonos) y el que llegó menos alto subió 1 octava y 2 semitonos.

La frecuencia media en el registro modal fue de 116.6 Hz y la del *falsetto*, de 329.0 Hz. Se observó asimismo que la dispersión de la f_0 se incrementaba notablemente en el *falsetto*, con 60.2 Hz de desviación típica promedio, frente a los 12.7 Hz registrados para la voz modal.

El valor más bajo de f_0 consignado durante la realización del *falsetto* fue de 216 Hz y, para la voz modal, de menos de la mitad (100.7 Hz). El punto más alto alcanzó los 408.6 Hz de promedio en el *falsetto* y 137.1 Hz en el registro modal. Las cifras relativas al valor medio de f_0 en el *falsetto* y en la voz modal, así como el aumento de tono de un registro a otro, concuerdan con los datos ofrecidos por la bibliografía anterior, por ejemplo, Colton (1969), Childers y Lee (1991), Hollien (1974), Künzel (2000), Roubeau *et al*., (2009), Švec, Schutte y Miller (1999), Wagner y Köster (1999); véanse los Capítulos 1 y 2 del presente trabajo. Los demás datos aportados en esta tesis son, hasta donde se ha podido comprobar, nuevos, pues no se habían registrado previamente en la bibliografía especializada.

6.1.2. Análisis de 72 rasgos glotales con *BioMet®Phon*

Uno de los objetivos de este trabajo era poder avanzar en el conocimiento de los rasgos laríngeos que, como consecuencia del cambio de registro modal-*falsetto*, sufrieran alteraciones significativas en los hablantes. Esto permitiría, por un lado, identificar los rasgos definitorios o determinantes del *falsetto* al tiempo que facilitaría la búsqueda futura de los rasgos laríngeos que, por mantenerse inmutables, podrían constituir una pista sobre la identidad de los locutores en casos de disimulo voluntario

de la voz trámite *falsetto*. En consonancia con la hipótesis principal del proyecto CIVIL (Alves *et al*., 2014; véase §3.1) que buscaba probar que en la voz permanecían rasgos –características laríngeas– persistentes a los intentos de disimulo mediante el cambio de registro (modal-*falsetto*, modal-*creak/creaky voice*), en esta tesis se propuso la primera hipótesis que se recuerda a continuación:

Hipótesis 1. En el paso de voz modal a falsetto, habrá rasgos laríngeos que modifiquen…

a) todos los locutores por ser intrínsecos al cambio de registro.

b) solo algunos locutores o ninguno.

En la tesis se analizaron los valores correspondientes a 72 parámetros glotales que estima la herramienta *Biomet®Phon* de *BioMetroSoft®* para conocer cuáles de ellos se veían alterados y cuáles no debido al cambio de registro. Se calculó la proporción media de parámetros laríngeos que se vieron modificados como consecuencia del cambio de registro, y se analizaron en función del grupo al que pertenecían (§4.1.2).

El grupo que se vio más afectado por el cambio de registro fue el de los cepstrales (grupo B). Prácticamente todos los parámetros de este grupo variaron significativamente entre la voz modal y el *falsetto*. Se observó asimismo que con respecto a este grupo se consignaba además la variabilidad interlocutor más baja, reflejando que, al cambiar de registro, todos los hablantes alteraban de modo coherente una proporción elevada de parámetros cepstrales.

Como también se explicó en el §4.1.2, a pesar de haber sido estos los parámetros que mayor variación experimentaron, se decidió no incluirlos en el Análisis por Componentes Principales porque aún no se conoce suficientemente su significado y, por tanto, cualquier efecto apreciable resultaría muy difícil de interpretar. Sin embargo, sería muy interesante, sobre todo para comprender mejor el rendimiento que los reconocedores automáticos de locutores tienen con las voces disimuladas, realizar en el futuro un análisis pormenorizado de este grupo. El hallazgo de que los parámetros cepstrales parecen ser muy sensibles al cambio de registro y de que todos los locutores están variando de forma constante una proporción grande de tales parámetros podría explicar el hecho de que, cuando está presente el *falsetto*, los reconocedores automáticos que se basan decisivamente en los rasgos cepstrales disminuyan su rendimiento a la hora de identificar el locutor. Estudios previos que evaluaban la exactitud de los

métodos automáticos en el reconocimiento de locutores con voces disimuladas, como los de Künzel *et al.* (2004) o Zhang y Tan (2008), a los que se hizo referencia en el §2.4.2., coincidían en que la elevación de la f_0 provocaba una disminución drástica en el reconocimiento, más incluso que otras técnicas de disimulo voluntario no electrónico. Incluso Künzel y colaboradores (2004, p. 4) comentaban en las líneas dedicadas a la investigación futura que aún quedaba pendiente la tarea más ardua, la de estudiar de manera pormenorizada las características fonéticas y fisiológicas de las distintas formas de disfraz vocal y, entre estas, muy en particular las características definitorias del *falsetto*. Estos autores mencionaban, a modo de ejemplo y para dar mayor peso empírico a esta afirmación, que durante la fase de análisis auditivo habían observado que aquellos locutores que peor habían sido reconocidos por los sistemas automáticos no solamente habían incrementado su f_0 sino que además habían cambiado de registro; y que, por tanto y muy plausiblemente, un cambio tan radical como ese afectaría también a las resonancias del tracto vocal que constituyen la base para la extracción de los MFCCs (*Mel* Cepstral *Coefficients*) utilizados por los sistemas automáticos (Künzel *et al.* 2004, p. 4).

El segundo grupo que experimentó mayores variaciones al pasar al registro *falsetto* fue el de los parámetros vinculados al funcionamiento biomecánico de los pliegues vocales (grupo D). Parece muy lógico que los parámetros pertenecientes a este grupo, parámetros que dan cuenta de la masa, la tensión y los gastos de energía de los pliegues vocales, presenten un comportamiento claramente distinto entre voz modal y *falsetto* y determinen, en buena media, el registro de fonación. De hecho, en las caracterizaciones de los registros que pueden localizarse en la bibliografía relevante al caso (por ejemplo, Gómez-Vilda y Pérez Sanz, inédito; Hirano, 1982; Hollien, 1974; Laver, 1980; y Titze, 1994, entre otros), se alude fundamentalmente a alguno de ellos, como se mencionó en el §1.4.

Tras los biomecánicos se ubicaron los relacionados con el perfil espectral (grupo C). De acuerdo con estos resultados, algunos de estos parámetros parecen ser también sensibles al cambio de registro pero al igual que ocurría con los cepstrales, son de difícil interpretación. En este caso particular, la complejidad mayor radica en que cuando se analizan voces que presentan valores altos o muy altos de f_0, como es el caso del *falsetto*, la información que puede ofrecer el espectro armónico es menos precisa o detallada, como se explicó antes en el Capítulo 4 (§4.1.3).

Los parámetros relativos a la evolución temporal del flujo (grupo E) y los de perturbación de la frecuencia fundamental (grupo A) variaron en una proporción similar de un registro a otro. Evidentemente, la modificación de la f_0 (grupo A) es imprescindible para el cambio de registro, aunque no suficiente. Como se ha explicado en los Capítulos 1 y 4 y se volverá a insistir en el próximo apartado, el tono de voz resultante estará determinado fundamentalmente por el grado de tensión que presenten los pliegues vocales, es decir, por las modificaciones que se produzcan en los parámetros del grupo D, fundamentalmente en los parámetros relativos a las estimaciones de la masa y tensión de los pliegues vocales implicadas en la vibración.

Finalmente, solo la mitad de los parámetros del grupo F (defecto de cierre glótico) se vieron modificados y, fue dentro del grupo G (parámetros de temblor) donde se registraron las menores variaciones porque solo un tercio de los parámetros pertenecientes a este grupo se vieron alterados en el cambio de registro.

Estos resultados permitían ya, por tanto, validar la primera hipótesis de esta investigación. Sin embargo, para conseguir una cabal comprensión de los rasgos del grupo a), recuérdese, aquellos intrínsecos al cambio de registro, se decidió realizar un posterior estudio mediante el Análisis de Componentes Principales (§4.1.3).

6.1.3. Análisis de los Componentes Principales

Para el Análisis de Componentes Principales se seleccionaron los parámetros que, de acuerdo con la bibliografía existente, se creyó que podían ser relevantes en el cambio de registro, y otros quedaron sin tratar, bien porque aún no se conoce suficientemente qué significan (Grupo B, parámetros cepstrales), bien porque resultan muy difíciles de interpretar en el registro de *falsetto* (Grupo C, parámetros del perfil espectral), o porque no tienen nada que ver con él (Grupo G, parámetros relativos al temblor). Estos rasgos han quedado sin explorar y deberían analizarse en estudios posteriores.

El Análisis por Componentes Principales mostró que el único factor discriminante entre ambos registros es el primer Componente Principal o CP1. Estos resultados confirman que existe una serie de parámetros que, como se preveía en la primera hipótesis, cambian de forma significativa en todos los sujetos al pasar de un registro a otro y que, esos parámetros, agrupados en el CP1, se relacionan directamente

con la f_0 y con los cambios en la masa y la tensión de los pliegues vocales, rasgos que ya habían sido descritos en la bibliografía como definitorios del registro de *falsetto* (véase por ejemplo, Titze, 1994) y que los análisis llevados a cabo en esta tesis lo ratifican de manera muy clara. Como fue explicado en el Capítulo 4, no es posible fundamentar el cambio de registro únicamente en un incremento del tono puesto que este puede incluso aumentar significativamente dentro del registro modal. Los datos derivados principalmente del Análisis por Componentes Principales han confirmado que son los parámetros biomecánicos, es decir, aquellos que responden a las estimaciones biométricas de la masa, la tensión y las pérdidas de energía de los pliegues vocales, los que verdaderamente inducen el comportamiento de la f_0 y, al hacerlo, dan lugar al cambio de registro.

En síntesis, el comportamiento de los parámetros que tras el análisis quedaron coherentemente agrupados en un único Componente Principal (CP1) es el que caracteriza y determina el cambio de registro de voz modal a *falsetto*. En este Componente Principal definitorio del registro se agruparon el tono, todos los rasgos relativos a la cubierta, y los parámetros que dan cuenta de la masa y tensión del cuerpo del pliegue vocal. Se observó que los valores de todos ellos aumentaron significativamente en el *falsetto* y disminuyeron drásticamente en la voz modal. Dentro del primer Componente Principal o CP1, solo la masa del cuerpo tuvo, como era de esperar, un comportamiento inverso, es decir, disminuyó en el *falsetto* cuando los otros aumentaron. En particular, la pauta de comportamiento que sigue la cubierta devino esencial para que se produjera tal cambio porque se confirmó que en la voz modal la implicación de la cubierta es totalmente diferente de la que presenta en el *falsetto*.

Otra conclusión interesante que se desprende del Análisis por Componentes Principales es que, aunque normalmente los Componentes Principales (CPs) son un constructo matemático muy difícil de interpretar, los resultados se muestran muy alentadores para todos los investigadores que en el futuro quieran dedicarse al estudio de la cualidad de voz. Se comprueba que es posible reducir enormemente la multidimensionalidad de los análisis teniendo en cuenta que luego los múltiples parámetros van a quedar claramente agrupados por CPs que pueden ser explicados.

6.2. Sobre el *falsetto* y su perjuicio para la discriminación de voces

Al comienzo de este trabajo se postuló, por medio de la formulación de otras dos hipótesis (hipótesis 2 y 3), que los oyentes serían capaces de reconocer a un mismo locutor cuando hablara en voz modal y *falsetto* gracias a un conjunto de rasgos o señales individuales de la cualidad de voz que pervivirían al disimulo de registro y que podrían ser perceptivamente relevantes para los oyentes. Tales hipótesis se recuerdan a continuación:

Hipótesis 2. Los rasgos laríngeos son una clave perceptiva para el reconocimiento de locutores.

Hipótesis 3. Los oyentes pueden reconocer a un mismo locutor por encima del nivel del azar, aunque haya cambiado de registro modal-falsetto.

Se esperaba que, incluso habiéndose alterado notablemente la f_0 mediante el *falsetto*, los oyentes serían capaces de reconocer las voces más allá del nivel del azar porque quedarían señales individuales de la voz a pesar del disimulo. Y que las señales vocales que provienen de la cualidad de voz en sentido restringido, es decir, los rasgos laríngeos, serían relevantes perceptivamente para el reconocimiento de hablantes.

Los estudios anteriores sobre el disimulo de la voz reseñados en el Capítulo 2 de esta tesis mostraban que los métodos humanos más frecuentes (por ejemplo, Künzel, 2000; Masthoff, 1966; Zhang y Tan, 2008) y que por lo general más afectan el reconocimiento o discriminación de voces familiares y no familiares son los que provocan cambios importantes en el habitual funcionamiento laríngeo. Entre esos cambios a nivel laríngeo, el aumento drástico de la f_0 resultó ser muy conveniente para falsear de manera rentable la propia voz. Se apuntaron fundamentalmente las siguientes razones (recuérdese, por ejemplo, Künzel, 2000, p. 173): es relativamente fácil de realizar, no interfiere en la correcta transmisión de los mensajes lingüísticos y ha demostrado una considerable influencia perjudicial para el reconocimiento de locutores, tanto para el reconocimiento perceptivo llevado a cabo por humanos (véanse en el Capítulo 2, por ejemplo, los estudios de Alves *et al.*, 2012; Fernández Trinidad *et al.*, 2013; Künzel, 2000; Masthoff, 1996; Wagner y Köster, 1999), como para el que

realizan los sistemas automáticos (por ejemplo, Künzel *et al.*, 2004; Perrot, Preteux, Vasseur y Chollet, 2007, y Zhang y Tan, 2008).

El diseño del experimento perceptivo realizado en esta tesis estuvo orientado a averiguar hasta qué punto el *falsetto* dificulta la discriminación en las parejas de igual locutor y en qué medida lo hace. Los estímulos que se presentaron en las pruebas perceptivas se escogieron de forma que los oyentes se centraran principalmente en las características de la voz derivadas sobre todo del comportamiento laríngeo y para que no hubiera más diferencias en la cualidad de voz que las provocadas por el cambio de registro. En la realización de las pruebas perceptivas los oyentes compararon dos estímulos emparejados modal-*falsetto* pronunciados por iguales y distintos locutores y debieron responder si las parejas de estímulos presentadas cada vez pertenecían al mismo locutor o a locutores diferentes (§3.4). Los datos se analizaron en el marco de la *Teoría de Detección de Señales* (§5.1) lo cual permitió evaluar adecuadamente el grado de discriminabilidad (d') y los sesgos de respuesta causados por la tendencia de los oyentes a contestar de una forma determinada (índice c).

Los datos obtenidos confirman que el disimulo laríngeo de la voz afecta[45] pero no impide el reconocimiento del locutor. Los oyentes fueron capaces de discriminar locutores hablando en voz modal y *falsetto* significativamente por encima del nivel de azar. Sin embargo, la discriminabilidad fue muy débil, con valores de d' cercanos a 1 (d' media= 0.77 para el grupo de jueces españoles y 0.95 para el de los italianos). El criterio de respuesta fue, en ambos casos, liberal (españoles, c= −0.13 e italianos −0.07), es decir, manifestaron una tendencia a responder que "sí"; en este caso, a responder que la voz modal y *falsetto* pertenecían al mismo locutor (§5.2).

Estos resultados discrepan con los anteriores que habían mostrado o bien una reducción muy drástica en el reconocimiento o bien la incapacidad de los oyentes para reconocer voces en la condición de *falsetto*. Las pruebas desarrolladas por Alves *et al.* (2012) y Fernández Trinidad *et al.* (2013) indicaban que los oyentes reconocían por encima del nivel del azar (tasa de acierto = 0.62 y 0.66 para cada estudio, respectivamente) pero estos resultados debían reconsiderarse pues, como se explicó previamente (§2.4.2), los porcentajes de aciertos fueron calculados en base al umbral del azar (0.5) y, por el mismo diseño del test perceptivo, el azar permitía *a priori* un

[45] Aunque no es posible calcular el nivel de perjuicio provocado por el *falsetto* pues en esta tesis no se evaluó la habilidad de los mismos oyentes para reconocer a los mismos locutores en la condición voz modal-voz modal.

50% de aciertos. En el experimento realizado por Fernández Trinidad y Rojo (2018) se había mostrado un resultado discriminante para los jueces, aunque también muy bajo, con una d' media de 0.82. Como en esta tesis, los locutores eran italianos pero los jueces que evaluaron las voces habían sido solo españoles.

Parece un hecho innegable que el *falsetto* interfiere negativamente en las tareas de evaluación y reconocimiento de voces, aunque los resultados difieren de los aportados por otros estudios que, como se ha visto, suelen ser inferiores al azar. El hecho de que los niveles de discriminación hallados no coincidan con los de otros estudios previos puede explicarse por la ausencia de homogeneidad o falta de sistematicidad en los procedimientos experimentales. Las discrepancias más evidentes que se han mencionado en la bibliografía (véase el Capítulo 2) se relacionan principalmente con las características del locutor (voces masculinas frente a femeninas), con las de los jueces (expertos/entrenados frente a oyentes profanos) y con el tipo de vínculo entre ellos (familiaridad con las voces o con la lengua). Los resultados de esta tesis apuntan también a otros factores o variables que convendría considerar y que podrían relacionarse directamente con la tarea y, en consecuencia, con el tipo de juicio que emiten los oyentes.

Son variadas las tareas perceptivas que pueden solicitarse a los oyentes en distintos experimentos y también variadas las tareas que, en situaciones reales, los oyentes pueden tener que realizar: evaluación, comparación, discriminación, reconocimiento, identificación, verificación de locutor (véase Gil Fernández 2014: pp. 66-67 para una discusión sobre las tareas realizadas por expertos en el ámbito de la fonética judicial y véanse Kreiman 1997, pp. 86-88 y Kreiman y Sidtis 2013[2011], pp. 156-158 para una descripción detallada de las tareas de evaluación de voces en general). En algunos de los experimentos reseñados en el Capítulo 2 (§ 2.4.2) se les pedía a los oyentes que reconocieran voces familiares, en otros, voces desconocidas. En ocasiones las comparaciones se hacían con todas las voces auditivamente disponibles para los oyentes, en otras, las voces oídas se comparaban con el recuerdo de otra voz archivada en la memoria a largo plazo[46]. Incluso entre las tareas con mayores semejanzas entre sí se apreciaron diferencias importantes porque exigieron de los oyentes juicios de distinta

[46]Como explicaban Kreiman y Sidtis (2013 [2011]), todas las tareas mencionadas son exigentes y en todas ellas los oyentes deben discernir si la variación observada entre los ejemplarios de voces se justifica mejor por la variabilidad intrínseca e inherente al propio hablante o por la variabilidad existente entre locutores distintos (p. 158).

naturaleza. Por ejemplo, las tareas perceptivas realizadas en esta tesis y en el anterior trabajo de Fernández Trinidad y Rojo (2018) fueron de discriminación de parejas AX igual-diferente y los oyentes debieron realizar juicios absolutos (¿es la misma persona la que habla? SÍ/NO). Los estudios de Alves *et al.* (2012) y de Fernández Trinidad *et al.* (2013) por su parte, utilizaron un diseño con tripletas XAB y, por tanto, los oyentes debieron realizar juicios relativos: ¿qué voz se parece más a la presentada previamente?, ¿A o B?

Por tanto, incluso si se circunscribe la tarea a la discriminación de voces no familiares la toma de decisión puede realizarse mediante dos tipos de juicios diferentes. Los oyentes pueden tener que emitir juicios absolutos -las voces son iguales o diferentes-, o pueden tener que emitir juicios relativos, decidiendo cuál de las voces que se le presentan es más similar a la voz objetivo. Se trata, en suma, de procedimientos diferentes que tienen consecuencias en los resultados. Recuérdese lo que señalaba Marrero (2014, p. 523) cuando al comparar los paradigmas AX y ABX explicaba que la segunda tarea es más exigente porque entraña mayor dificultad y una capacidad de memoria mayor. Además, se ha visto que los juicios relativos inducen a error más frecuentemente que los absolutos. Como explica Manzanero (2010, p. 146) a propósito de los procesos cognitivos involucrados en el reconocimiento de rostros, los juicios relativos facilitan mayor número de falsas identificaciones que los juicios absolutos, puesto que dentro de un grupo de personas siempre existirá una que se parezca más que otras a la persona objetivo.

De la interpretación conjunta de los datos de variación del estudio acústico y los resultados obtenidos de las pruebas de percepción pueden extraerse también conclusiones parciales que merece la pena seguir explorando.

Los resultados de las pruebas perceptivas realizadas a 140 oyentes demostraron que la alteración del primer Componente Principal o CP1 no fue suficiente para imposibilitar la discriminación de un locutor con él mismo cuando cambia de registro modal a *falsetto*. El CP1, y por ende, el tono de la voz, se reveló como una clave definitoria del registro, pero no del locutor, en los casos de comparación de voces disimuladas en *falsetto*.

Los oyentes no parecieron apoyarse demasiado en el tono para reconocer a los locutores porque los principales factores responsables de alterarlo (CP1) cambiaron significativamente en todos los locutores de modal a *falsetto*. Como todos ellos han

podido reconocerse por encima del nivel del azar, se diría que el tono no desempeña un papel tan relevante en el proceso.

Estos mismos resultados sugieren, por tanto, que, aun con un claro cambio de registro, tiene que permanecer un resto vocal que es el que permite reconocer al locutor y no al propio registro, pero, ¿qué es lo que perdura entonces de la voz si no es el tono? La evidencia de los datos aquí presentados prueba que probablemente permanecen ciertos rasgos idiosincrásicos que no pueden ser alterados. Conocer lo que ha cambiado de un registro a otro (CP1) permite restringir la búsqueda de cuáles podrían ser las invariantes presentes en la señal acústica, de modo que la capacidad de reconocer al locutor, muy plausiblemente, se decidió a partir de la atención que se les prestara. Un futuro análisis que ponga en relación estadísticamente (y no solo interpretativamente) los resultados de las pruebas perceptivas y los resultados acústicos y articulatorios seguramente contribuya de manera más decisiva a desentrañar más claramente los múltiples parámetros que definen la cualidad individual, todavía ignorados.

A pesar de que al diseñar las pseudofrases del modo en que se hizo se buscó que las diferencias entre las voces emergieran casi exclusivamente del comportamiento laríngeo derivado del cambio de registro y no de variaciones en la composición léxica o en la estructura prosódica de los enunciados, se constató que en su pronunciación seguían apareciendo otros rasgos potencialmente diferenciadores que pudieron contribuir a la discriminación de locutores. Mediante el análisis realizado en el §4.2 se observó que, aun en frases de corta duración, pueden emerger características temporales destacadas. Los locutores L4, L5 y L6 sobresalieron en la duración global de las frases (considerando ambos registros simultáneamente). El Locutor 4 habló más lentamente que el resto mientras que los locutores L5 y L6 fueron significativamente más rápidos que todos los demás. Al observar el *tempo* global (modal+*falsetto*) y el *tempo* de cada registro por separado se ha visto que el Locutor 6 no presentaba diferencias internas entre la duración de sus frases pronunciadas en voz modal y en *falsetto*; sin embargo, los locutores L4 y L5, que destacaban en duración global, presentaron variabilidad interna, puesto que las duraciones en cada registro mostraron diferencias significativas entre sí. Por tanto, es bastante probable que el *tempo* pueda haber funcionado como clave perceptiva en los juicios de los oyentes en relación con ciertos locutores.

Las circunstancias en las que el *tempo* pudo haber contribuido al rendimiento de los jueces podrían ser de dos tipos. Por un lado, puede que un locutor destaque en el *tempo* global y, a su vez, no presente diferencias significativas entre la velocidad de los

registros (Locutor 6); en definitiva, que manifieste una máxima variabilidad interlocutor y una mínima variabilidad intralocutor. Por otro lado, puede que un locutor se caracterice por un *tempo* global propio –se perciba como más lento o más rápido que los otros locutores– y que, aun presentando diferencias internas entre ambos registros, el *tempo* de cada registro destaque respecto del de todos los demás –su habla en modal se la más lenta o rápida de todas las modales, o su *falsetto* sea el más lento o rápido de todos los *falsettos*– y por ello sea posible para los oyentes, en cualquier caso, identificar perceptivamente la voz modal con su *falsetto* (Locutor 4). En consecuencia, es probable que, al menos en lo que se refiere a estos locutores, los rasgos temporales hayan podido constituir una clave perceptiva importante, ayudando a apreciar similitudes o diferencias y contribuyendo, por tanto, a una mejor discriminación.

Como se mencionó en §2.4.3., los cambios de registro pueden acarrear cambios en la velocidad de habla o de articulación de los locutores y en la entonación (Künzel, 2000 y Wagner y Köster, 1999). Por ejemplo, Künzel (2000) había observado que en casos de disimulo voluntario, el aumento del tono de la voz –sobre todo si se daban consecuentemente cambios de registro– provocaba cambios en la velocidad del habla. En su estudio observó una ralentización de la velocidad y un aumento de las pausas y de su duración. Recuérdese que estas modificaciones habían sido interpretadas por Künzel (2000, p. 173) como cambios no intencionados, sino más bien derivados de la propia dificultad que supondría para los locutores producir ajustes fonatorios poco habituales. En los datos obtenidos de esta tesis no se observa, sin embargo, un efecto homogéneo del registro en la duración de las frases. Los locutores se comportaron de forma diferente dependiendo del registro y utilizaron estrategias distintas y opuestas. Se observaron diferencias significativas en la duración de los registros en el caso de L1, L2, L4 y L5. En el Locutor 1 y en el Locutor 5 la duración de las frases en el registro modal fue mayor que la del *falsetto*; por tanto, el *tempo* resultó ser más lento en la voz modal. Inversamente, en los locutores L2 y L4 la duración fue mayor al cambiar a *falsetto*, y, por consiguiente, disminuyeron de forma significativa la velocidad en ese registro. Estos datos parecen sugerir, pues, que los locutores se sirvieron de estrategias diferentes frente a la dificultad que seguramente supuso aumentar de manera tan drástica sus valores habituales de f_0 en el registro de *falsetto*.

Además de las variaciones en la velocidad del habla recién mencionadas, una meticulosa escucha realizada por cuatro fonetistas españoles y por tres fonetistas italianos nativos había confirmado que no existían rasgos de pronunciación demasiado

evidentes que diferenciaran perceptivamente a un locutor de otro, con algunas excepciones muy poco notorias que, presando mucha atención, podrían llegar a manifestarse débilmente en algunas frases pronunciadas por los locutores L3, L4 y L5. Los fonetistas españoles advirtieron una pronunciación más marcada de las obstruyentes oclusivas en algunas frases del Locutor 4, detalle que no fue apreciado por ninguno de los fonetistas italianos. Los comentarios cualitativos realizados por dos de los tres fonetistas italianos consultados sugerían que algunos de los enunciados pronunciados por el Locutor 3 se percibían globalmente distintos del resto de locutores en lo que a la prosodia se refería. Sin embargo, este hecho no fue detectado por el tercer fonetista al que le pareció percibir, y solamente para el Locutor 5, algunas diferencias rítmicas (véase §4.3).

En suma, al igual que en el reciente estudio de San Segundo, Foulkes, y Hughes (2016), comentado en §2.5, aquí también se constata que aun en estímulos de corta duración todavía pueden sobresalir algunos rasgos suprasegmentales como el *tempo*, la entonación y el ritmo, pues los expertos, al juzgar los estímulos señalaron ciertas peculiaridades aisladas aunque poco notorias y solamente para ciertos locutores. No obstante, es muy importante recordar que los aspectos por ellos apuntados no fueron coincidentes entre sí ni tuvieron una asertividad suficientemente contundente como para pensar que dichas características pudiesen afectar de forma dramática el reconocimiento de los locutores, y, de hecho, los resultados perceptivos así lo demostraron.

Es decir, aunque como han señalado los expertos fonetistas pudiera haber alguna leve característica potencialmente definitoria en el habla de alguno de los locutores, el hecho cierto es que, a juzgar por los resultados perceptivos extraídos de las respuestas obtenidas de los 140 jueces que participaron en las pruebas perceptivas, ninguna de esas características debió de ser tan determinante para reconocer a todos ellos.

Recuérdese que en los resultados perceptivos no se registraron diferencias significativas en la tasa de reconocimiento de unos locutores frente a otros con la única excepción del Locutor 6 el que, de una forma marginal y solamente para la mitad de los jueces (oyentes italianos), fue reconocido mejor que el resto (§5.4). Recuérdese también que este locutor no presentaba rasgos entonativos ni segmentales peculiares, sin embargo, fue el que menos número de parámetros laríngeos modificó (véase §4.1.2) y además destacaba por sobre los demás en cuanto a la velocidad de habla, habiéndose revelado como el locutor más rápido de todos, junto con el Locutor 5, pero a diferencia de este último que presentaba diferencias significativas dentro de cada registro, el

Locutor 6 ofrecía un comportamiento coherente en un registro y otro, pues mantenía la misma velocidad en modal y en *falsetto* (véase § 4.2). Es importante advertir, pues, que frente a este resultado, no es posible derivar fácilmente el efecto específico de lo laríngeo del efecto provocado por la velocidad de habla en el reconocimiento. Al haber decidido metodológicamente no controlar la variable *tempo* por los motivos aducidos en §3.3.2., no es posible ahora demarcar los efectos atribuibles concretamente al comportamiento laríngeo de los debidos a los rasgos temporales de su habla o incluso, y muy probablemente, a una interacción entre ambos. Esta cuestión podrá dirimirse en futuros estudios.

6.3. Sobre el beneficio de conocer la lengua del locutor en el reconocimiento de voces

Según se observó del repaso bibliográfico realizado en el §2.5, la gran mayoría de los trabajos demostraba la tesis de que el conocimiento de la lengua beneficia la evaluación y el reconocimiento de voces, a pesar de que no siempre podían equipararse los resultados a causa de la falta de homogeneidad de los diseños experimentales adoptados para ponerla a prueba. La aparente multiplicidad de los resultados respondía fundamentalmente a diferencias en el grado de familiaridad entre las lenguas que hablantes y oyentes compartían, al tamaño de las muestras de voz y, por supuesto, a las medidas que se calculaban para dar cuenta del desempeño de los oyentes. Habida cuenta de estas divergencias evidentes, que lógicamente dificultaban la comparación de los hallazgos alcanzados por los diversos estudios, fue posible desentrañar algunas coincidencias y corroborar el beneficio de la familiaridad con la lengua del locutor para reconocerlo a través de su voz (recuérdese Kreiman y Sidtis, 2013 [2011], p. 241-243).

Esos mismos resultados mostraban que el reconocimiento mejoraba sensiblemente con el incremento de la duración y de la variedad de los estímulos. Este último punto resultaba clave, pues podría estar sugiriendo que, solo cuando verdaderamente aumenta la información dependiente de la lengua y, por tanto, está presente en la señal un inventario más amplio de sonidos o afloran patrones rítmicos y entonativos, los oyentes nativos o buenos conocedores de la lengua del locutor podrían beneficiarse de esas claves y obtener mejores resultados en el reconocimiento de voces. Así las cosas, la ventaja de los oyentes nativos no sería tan evidente si la información

dependiente de la lengua fuera de algún modo minimizada o cancelada. Algunos estudios (véase §2.5) avalaban esta presunción porque mostraban que, aunque lógicamente disminuía la habilidad para reconocer locutores en tales circunstancias, los oyentes eran capaces de registrar diferencias y similitudes entre voces que les permitieran distinguir locutores y que, en estos casos, el desempeño de nativos y no nativos no difería demasiado. Algo semejante se comprobaba en el trabajo de San Segundo *et al*., (2016) al evaluar la familiaridad de la lengua en la calificación de similitudes entre locutores a partir de sus voces. La interpretación conjunta de los resultados aludía, por tanto, a que los oyentes estarían también prestando atención a aspectos vocales no lingüísticos y que estos, en ciertas condiciones, podrían ser suficientes para reconocer la individualidad de una voz. En esta tesis se previó que tales aspectos serían sobre todo laríngeos aunque se vio que también pudieron influir los temporales. En este sentido, el desempeño de oyentes nativos y no nativos podría ser, pues, muy similar.

Uno de los objetivos que persiguió esta investigación fue, por tanto, evaluar si la cualidad de la voz podría ser suficiente para discriminar entre voces diferentes. No resulta fácil evaluar hasta qué punto el contenido segmental y lexemático puede afectar la interpretación de la cualidad de voz puesto que el oyente normalmente los percibe a la vez. Además, como se ha señalado tantas veces, conseguir separar por completo la cualidad de voz del factor léxico y fonético segmental que conllevan los mensajes es una tarea muy complicada o quizás imposible; no obstante, el diseño de los estímulos buscó acercarse lo más posible a este objetivo.

La hipótesis 4 preveía, de modo semejante a como se postulaba en el estudio de San Segundo *et al.,* (2016), que la utilización de una única frase breve y sin sentido conseguiría que los oyentes basasen principalmente sus juicios de semejanza o diferencia de voces en la cualidad de la voz y que, por tanto, la similitud de cada hablante consigo mismo cuando cambiara de registro se explicaría principalmente por las características laríngeas que, en tanto información extralingüística, serían evaluadas de forma muy semejante por los oyentes nativos y no nativos de italiano. Dicho de otro modo, cuando prácticamente solo estuvieran disponibles para el oyente rasgos de la cualidad de la voz, los oyentes que compartieran la lengua con los locutores que hubiera que reconocer no tendrían ventaja respecto de los oyentes que no la compartieran o no la conocieran. Recuérdese el modo en que quedaba expresada la cuarta y última hipótesis de esta tesis:

Hipótesis 4. Puesto que los registros de fonación son, para ambas lenguas, aspectos extralingüísticos, los oyentes de español se van a comportar igual que los oyentes italianos en el reconocimiento de locutores italianos que disimulan su voz en falsetto.

Los resultados obtenidos de las pruebas de discriminación confirman parcialmente también esta hipótesis. La Tabla 6.1 los resume según la lengua de los oyentes.

Tabla 6.1: Síntesis de los resultados acerca del efecto de la lengua sobre la proporción de Aciertos y Falsas Alarmas, medias de discriminabilidad (d') y criterio de respuesta (c). Valores medios de tiempos de reacción (TR) en ms y Logaritmos neperianos del tiempo de reacción (logTR) según grupos de oyentes italianos y españoles, ***p <.001.

L1 de los oyentes	Aciertos	Falsas Alarmas	Discriminación (d')	Criterio de respuesta (c)	TR	LogTR
Italianos	0.69	0.34***	0.95	-0.07	1737***	6.96***
Españoles	0.68	0.40***	0.77	-0.13	1377***	6.80***

Los resultados generados en este estudio han demostrado que, al limitar la cantidad de información lingüística disponible en la señal de voz, todos los oyentes, con independencia del conocimiento que tengan de la lengua del locutor, son capaces de discriminar si dos estímulos, uno producido con voz modal y otro con voz disimulada en *falsetto*, pertenecen o no al mismo locutor, aunque como se ha remarcado antes, la discriminación fue débil. Los mismos datos sugieren, por consiguiente, que no es necesario procesar el significado de las frases pronunciadas por los locutores para que se reconozca una voz o se evalúen diferencias o semejanzas entre hablantes, al igual que se concluyó en San Segundo *et al.* (2016) y se había probado antes en Schiller *et al*. (1997), entre otros (véase §2.5).

Tanto los oyentes italianos como los españoles discriminaron por encima del nivel del azar con valores de d' muy próximos. Unos y otros fueron liberales en sus criterios de respuesta, y no se constataron tampoco aquí diferencias estadísticas entre los grupos.

A pesar de ser esto así, se observó que los italianos presentaban niveles mayores de habilidad discriminatoria y tendían a aplicar un criterio menos liberal en sus juicios, aunque estas tendencias no llegaban a resultar significativas, como ya se ha mencionado en el Capítulo 5. Esto podría explicar el mejor resultado que obtienen los italianos en las Falsas Alarmas. Por tanto, según estos datos, el conocimiento de la lengua no provoca unos mejores resultados cuando la señal está presente (AA), pero sí cuando está ausente (AB); en esta condición las diferencias entre grupos afloran con significatividad: los italianos obtienen menor proporción de Falsas Alarmas y por ello se equivocan menos cuando se presentan locutores diferentes.

6.4. Sobre los tiempos de reacción (TR) y la información que de ellos se desprende

Como se acaba de explicar, puesto que lo laríngeo[47] es extralingüístico en el caso de lenguas como el español y el italiano, no se esperaban particulares singularidades en el desempeño de unos jueces y otros. Los resultados sobre la tasa de acierto y la d prima no mostraron, en efecto, diferencias significativas entre los jueces españoles y los italianos, salvo en lo que respecta a las Falsas Alarmas, menores en el caso de los italianos. Asimismo, se observó una tendencia a que este último grupo discriminara mejor. A la luz de este resultado, que valida parcialmente la hipótesis 4, se ha decidido analizar con mayor detenimiento los tiempos de reacción (§5.3) a fin de averiguar algo más sobre qué es lo que puede esconderse tras esta tendencia diferente, aunque no significativa.

6.4.1. Tiempos de reacción (TR) y conocimiento del idioma

Se calcularon los tiempos de reacción o TR por grupo de oyentes porque ello podría aportar alguna pista sobre los procesos cognitivos implicados. Como se vio en el apartado § 2.3.1, en la bibliografía pertinente se han propuesto dos estrategias básicas y distintas de procesamiento de voces: por un lado el procesamiento por separado de

[47]Con esta expresión se hace referencia al comportamiento laríngeo solo en lo que a los tipos de fonación se refiere, pues la oposición entre sordo/ sonoro es, indudablemente, también laríngea, y constituye un rasgo con valor lingüístico relevante en los sistemas fonológicos de las lenguas.

rasgos específicos y, por otro, el procesamiento global u holístico. Se ha comentado también una tercera propuesta sintética, según la cual el procesamiento de la voz sería siempre dual, en el sentido de que incluiría ambos mecanismos o fases que se sucederían en un orden diferente (de abajo a arriba o de arriba abajo) y que tendrían un peso también distinto dependiendo de si lo que se percibe es una voz familiar o desconocida. De acuerdo con lo que planteaban Kreiman y Sidtis (2013 [2011], pp. 187-188), en el procesamiento de voces desconocidas, como es el caso presente, el proceso empezaría desde abajo, es decir, el oyente comenzaría analizando las características estrictamente presentes en la señal acústica y luego buscaría interpretarlas en relación con un patrón general, dibujado mentalmente en función de los conocimientos previos sobre la voz de un locutor, entre los cuales estarían, lógicamente, los lingüísticos. Bajo determinadas circunstancias que podían variar –que el oyente conozca la lengua del locutor o que se trate de un oyente entrenado en la valoración de voces–, se podría analizar un conjunto más o menos amplio de parámetros; por ello, en el caso de discriminación de voces no conocidas los tiempos de respuesta suelen variar más que en el del reconocimiento de voces familiares (véanse de nuevo § 2.3.1 y Kreiman y Sidtis, p. 187).

A partir del análisis realizado en esta tesis, se observó un efecto del idioma en los tiempos de respuesta o de reacción, lo cual ofrece algunas pistas sobre el proceso subyacente de percepción de la cualidad de voz que parece ser, en ambos casos, tanto para los jueces italianos como para los españoles, de abajo a arriba. Los italianos emplearon tiempos superiores a los de los españoles en todos los tipos de respuesta, es decir, tardaron significativamente más en tomar sus decisiones y ofrecer sus juicios. Probablemente tanto italianos como españoles hayan tenido que deconstruir las voces para analizarlas en la primera fase de procesamiento componencial. Parece lógico que los italianos hayan tardado más si es mayor el número de parámetros a los que pueden acceder por conocer bien el modelo promedio de cómo suena un italiano (de Nápoles) y, por tanto, poder evaluar con más detalle qué rasgos de cada voz particular se alejan de ese prototipo. Sin embargo, ¿por qué, entonces, si los italianos son capaces de procesar más rasgos, no obtienen mejores resultados? Aunque hayan podido fijar su atención en más propiedades y por ello hayan podido necesitar más tiempo, tiene sentido que no hayan discriminado mejor porque en los estímulos presentados prácticamente no había rasgos lingüísticos sobresalientes que hayan podido guiar sus juicios. El reconocimiento del patrón vocal subyacente difícilmente podía basarse en el reconocimiento de rasgos

dependientes de la lengua, sino más bien en características de la cualidad de la voz, es decir, extralingüísticas. En este sentido, es poco lo que podría aportar el análisis por rasgos constitutivos y quizás por ello la familiaridad con la propia lengua no supuso un claro beneficio.

6.4.2. Tiempos de reacción (TR) y exactitud de las respuestas

Aunque no relacionado *a priori* con la influencia de la lengua, otro punto sobre el que resulta interesante reflexionar surge de los datos obtenidos de la medición de tiempos de respuesta y de su relación con la exactitud en los reconocimientos e identificaciones. Como se ha comentado en el Capítulo 5, (§5.3), los italianos tardaron significativamente más que los españoles en todos los tipos de respuesta, pero dentro de cada grupo de oyentes se observó el mismo comportamiento en cuanto al tiempo de respuesta y el hecho de acertar o fallar.

En esta tesis se constata una relación entre el tiempo de reacción y el tipo de respuesta, y se observa que los casos de acierto (entendiendo por ello Aciertos y Rechazos Correctos) presentan tiempos de respuesta menores, mientras que los fallos (Falsas Alarmas y Omisiones) se caracterizan por tiempos de respuesta mayores. Todos los oyentes, independientemente de su lengua e independientemente de si estaban evaluando pares de igual locutor (AA) o de locutores distintos (AB), tardaron significativamente más cuando fallaron que cuando acertaron (véase §5.3).

Una mirada a los estudios sobre reconocimiento facial podría ayudar a entender mejor este comportamiento. Algunos trabajos dedicados a observar el comportamiento de los testigos en las ruedas de reconocimiento facial habían llegado antes a resultados semejantes observando una relación inversamente proporcional entre la probabilidad de acertar y el tiempo de reacción de las respuestas. En concreto, estos trabajos comprobaron que, para el reconocimiento de caras, los tiempos de respuesta más breves suelen indicar aciertos y los tiempos más prolongados señalan errores (para conocer más detalles véanse por ejemplo, Brewer, Caon, Todd, y Weber, 2006; Dunning y Stern, 1994; Smith, Lindsay y Pryke, 2000; Sporer, 1992, 1993, 1994; Weber, Brewer, Wells, Semmler y Keast, 2004, entre varios que aparecen en la bibliografía).

De hecho, la medición de los tiempos de respuesta se llegó a recomendar como una posible técnica para evaluar la exactitud en las identificaciones de caras (véanse por

ejemplo los trabajos de Manzanero Farias-Pajak, Igual y Quintana, 2011; Weber *et al*., 2004, entre otros, que comentan la regla de los 10-12 segundos propuesta por Dunning y Perretta, 2002).

No obstante, en la práctica, se cuestiona la utilidad de los tiempos de reacción como indicador de la exactitud por varias razones (véanse nuevamente Brewer *et al*., 2006; Manzanero, 2010; y Manzanero *et al*., 2011, entre otros).

Sporer (1993 y 1994) llegó a señalar un aspecto interesante que matiza la correlación entre exactitud y tiempos de respuesta al observar los resultados obtenidos del análisis de una rueda de reconocimiento secuencial. Según se recoge en Weber *et al*. (2004, pp. 139-140), para Sporer, la correlación entre la velocidad de la respuesta y su precisión surge únicamente en el caso de los sujetos electores (*choosers*), es decir, cuando se fuerza a que los testigos señalen a uno de los sujetos de la rueda (Manzanero *et. al*., 2011, p. 111). En estos casos, los Aciertos son significativamente más rápidos que las Falsas Alarmas. Sin embargo, en el caso de los jueces no electores (*no-choosers*), no se da tal relación, pues cuando rechazan correctamente a un sujeto sus decisiones no son más rápidas que cuando lo rechazan incorrectamente (2004, pp. 139-140). Según los datos de Sporer, son más rápidos los Aciertos (Sí/S, acierto) que las Falsas Alarmas (Sí/R, error), pero al considerar los casos en los que los testigos responden "no", la velocidad de respuesta es la misma con independencia de su exactitud, es decir, no existen diferencias en el tiempo de respuesta de los Rechazos Correctos (No/R, acierto) y de las Omisiones (No/S, error), (véanse Kneller, Memon y Stevenage, 2001 y Weber *et al*., 2004). Por tanto, sus resultados atenúan en parte la premisa anterior de que los mayores tiempos de reacción indican siempre cualquier tipo de respuesta fallida o errónea (Falsas Alarmas y Omisiones) mientras que los menores tiempos de reacción son índices de cualquier tipo de respuesta acertada o correcta (Aciertos y Rechazos Correctos), (véase Kneller *et al*., 2001, p. 661)[48].

[48] Tal y como comentan Weber *et al*., (2004), en opinión de Sporer, este hecho responde al proceso de decisión de los jueces. Como explican los autores, en los Aciertos (Sí/S) la decisión se toma más rápidamente porque la imagen de la persona objetivo encuentra en la comparación muchos rasgos en común con la imagen almacenada en la memoria del testigo. De modo contrario, en una rueda con objetivo ausente, el tiempo de respuesta para la Falsa Alarma (Sí/R) es mayor, pues el emparejamiento se da más lentamente, ya que son pocos los rasgos compartidos entre el cebo de la rueda y la imagen almacenada en la memoria del testigo, todo lo cual justificaría que los Aciertos resulten ser más rápidos que las Falsas Alarmas. Sin embargo, los rechazos (respuestas "no") requieren siempre que todos los miembros de la rueda sean considerados antes de ser definitivamente descartados, por ello, todos los rechazos, independientemente de que sean correctos (No/R) o no lo sean (No/S) serían igualmente lentos (Weber *et al*., 2004, pp. 139-140). Por ello, en los casos de rechazo no se daría la relación entre el tiempo

Los resultados obtenidos de esta tesis coinciden con los más generales señalados por los estudios que afirman que las respuestas correctas (Aciertos y Rechazos Correctos) se corresponden con tiempos de reacción inferiores en comparación con las respuestas incorrectas (Falsas Alarmas y Omisiones), pues, como se recordará, aunque los italianos se demoraron más en todos los tipos de respuesta respecto de los españoles, dentro de cada grupo se observó el mismo comportamiento, a saber: siempre se tarda más cuando se falla que cuando se acierta.

No parece posible, pues, considerar que exista una única causa responsable del aumento de los tiempos de reacción y, muy probablemente, como apuntan Manzanero *et al.* (2011, p. 112), la exacta relación entre tiempo de reacción y precisión de la respuesta sigue siendo por ahora poco conocida. Sin embargo, resulta muy interesante comprobar, observando los resultados de esta tesis y los anteriormente señalados, cómo en modalidades de percepción diferentes (visual/auditiva) se ponen de manifiesto procesos cognitivos aparentemente similares.

Investigaciones como la realizada por Manzanero *et al.* (2011) sostienen, por ejemplo, que varios factores, como la calidad de la imagen mental que el testigo recuerda del delincuente, el criterio de decisión (relativo frente a absoluto), o la forma de presentación de los sujetos (simultáneo frente a secuencial) podrían estar influyendo en la relación entre el tiempo de reacción y la precisión de la respuesta, condicionándola en buena medida. De todos esos factores y otros que podrían también influir se hará una breve y final referencia a los dos últimos.

---

y la precisión de las respuestas y los Rechazos Correctos y las Omisiones presentarían los mismos tiempos de reacción (véase también Kneller *et al*., 2001, p. 661).

6.4.3. Tiempos de reacción (TR), modo de presentación de los estímulos (simultáneo frente a secuencial) y tipo de juicio (relativo frente a absoluto)

El tiempo utilizado en la toma de decisión de los jueces también se ha puesto en relación con el modo en el que se ofrecen o presentan los estímulos. Una vez más, los principales datos provienen de trabajos dedicados al reconocimiento visual, más concretamente, en ruedas de reconocimiento facial. Nuevamente, se hará referencia a ellos puesto que su estudio enriquecerá seguramente la discusión de los resultados obtenidos para el reconocimiento de voces.

En una rueda de reconocimiento, y también en experimentos perceptivos, pueden ser dos las formas de presentar los estímulos. En una rueda simultánea, todos los sujetos se presentan a la vez. El testigo debe decidir, por tanto, si alguna de esas personas que ahora se le presentan conjuntamente coincide con la imagen que recuerda del delincuente. Naturalmente, el testigo puede responder que "sí", e indicar con cuál de ellas coincide su recuerdo, o responder que "no" coincide con ninguno. En una rueda secuencial, por el contrario, cada miembro de la rueda se presenta separadamente. El testigo compara el recuerdo del delincuente con un sujeto cada vez y toma su decisión. En ambas modalidades, el sospechoso o persona objetivo puede estar o no presente, es decir, las ruedas de reconocimiento secuencial y simultánea pueden constituirse con el objetivo presente o con el objetivo ausente, dependiendo de si se incluye o no al sospechoso en la rueda (Kneller *et al.,* 2001, p. 660).

Como se aclara precisamente en el mismo artículo, la rueda más utilizada es la de presentación simultánea a pesar de que la investigación que compara ambos procedimientos señala que precisamente es la que facilita un mayor número de reconocimientos erróneos (Falsas Alarmas) a causa de que promueve que el testigo utilice una estrategia relativa durante su evaluación y posterior decisión, como se detallará enseguida.

Según se expone en el trabajo de Kneller y colaboradores, al comparar los resultados de aciertos y equivocaciones en experimentos realizados con presentaciones secuenciales y simultáneas (citan por ejemplo los trabajos de Lindsay, Lea y Fulford, 1991; Lindsay y Wells, 1985), se observó que en ambas se produce el mismo número de Aciertos pero que, en la secuencial, se reducen significativamente las Falsas Alarmas (2001, p. 660). Asimismo, se observó que en la presentación secuencial, aun no

habiendo más Aciertos, cuando los hay son más rápidos que los que se consiguen en la presentación simultánea. Esta constatación podría justificarse, según estas mismas investigaciones, por el tipo de decisión (absoluta o relativa) tomada por los sujetos. La presentación secuencial favorecería que las decisiones se tomaran con mayor rapidez gracias a una valoración absoluta mientras que la simultánea estaría obligando a los jueces a que decidieran su respuesta a partir de valoraciones relativas, que son más lentas (2001, pp. 660-661). A su vez, y como se ha explicado oportunamente en (§6.2), las evaluaciones relativas conducen más fácilmente al error, en concreto, a mayor número de Falsas Alarmas, y los mismos juicios relativos necesitarían de mayores tiempos de evaluación, es decir, presentan tiempos de reacción más largos.

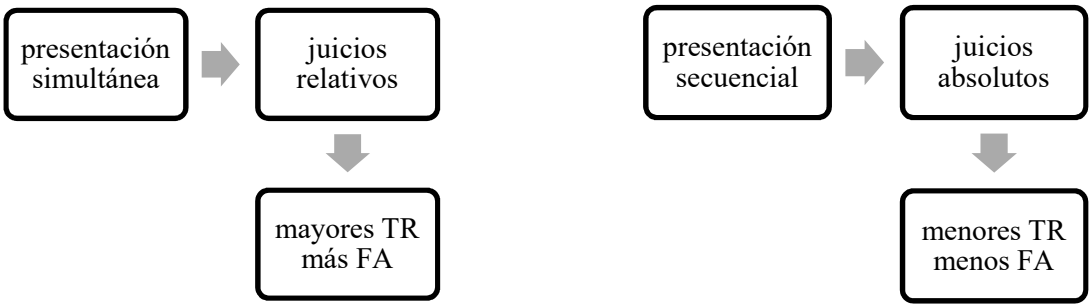


*Figura 6.1.*Relación esquemática entre presentación de estímulos (simultánea/secuencial), tipo de juicio (relativo/absoluto) y tiempos de reacción (TR).

No obstante, el estudio de Kneller *et al*. (2001, p. 670), ya mencionado, sugiere que la estrategia de decisión basada en juicios absolutos no se da exclusivamente en presentaciones secuenciales, puesto que mostró que los jueces recurren también a juicios absolutos frente a una presentación simultánea de los estímulos. A pesar de ello, los autores sí comprueban el hecho de que, si los jueces basan sus decisiones en valoraciones absolutas tienen más posibilidades de tomar una decisión correcta.

Es posible considerar la tarea perceptiva de esta tesis como de presentación secuencial, así como sería plausible asemejar, muy *grosso modo*, la distorsión provocada por el cambio de registro a la distorsión que supone el recuerdo de una

imagen (visual o vocal) almacenada en la memoria a largo plazo. Sería interesante diseñar en un futuro un experimento perceptivo con estas mismas voces pero presentándoselas todas al oyente de forma conjunta, simulando una rueda de reconocimiento simultánea, para poder trazar comparaciones entre los hallazgos relevados en ambos experimentos realizados con voces, y observar en qué medida coinciden con los derivados de los estudios previos llevados a cabo sobre reconocimiento de caras. Sin duda, coincidencias como las que se han apuntado anteriormente alientan a proseguir en la búsqueda de otras analogías y animan a intensificar el diálogo entre los que estudian el reconocimiento de voces y los que examinan el reconocimiento facial.

Capítulo 7
Síntesis de las conclusiones e implicaciones prácticas

7. Síntesis de las conclusiones e implicaciones prácticas

Con los resultados del presente trabajo se pretendía mejorar en algo el conocimiento básico del registro de *falsetto* y llegar también a una mejor comprensión de la percepción humana de la cualidad de la voz. Sin embargo, son indudables también sus implicaciones prácticas, puesto que muchos de sus hallazgos permiten realizar una transferencia de ese conocimiento al ámbito de la fonética forense o judicial. Como se verá enseguida a lo largo de este capítulo, parece bastante plausible, a partir de los resultados obtenidos, contribuir a algunas reflexiones en torno a cómo abordar en casos periciales un disimulo mediante *falsetto,* y sobre cómo puede influir en el proceso de reconocimiento de voces el hecho de que el locutor y el oyente hablen la misma lengua. Para poder discutir de forma más clara y sencilla tales implicaciones y con la esperanza de que su consideración pueda ser de utilidad en la práctica real —casos de cotejos de voces con fines judiciales, realización de ruedas de reconocimiento—, se ofrece primero una síntesis de las principales conclusiones [49] extraídas de esta tesis (§7.1) y posteriormente (§7.2) se examina su provecho en el campo aplicado, considerando la influencia de la lengua del locutor (§7.2.1), y el perjuicio causado por el disimulo mediante *falsetto* (§7.2.2), aspectos ya discutidos antes en la bibliografía sobre fonética judicial.

7.1. Síntesis de las principales conclusiones extraídas de la tesis

Un primer objetivo de esta tesis fue el de caracterizar acústicamente las diferencias provocadas por el cambio de registro modal-*falsetto* y buscar sus posibles correlatos fisiológicos; en este sentido, los principales resultados, muy resumidos, se recuerdan a continuación:

1) Cuando pasan de voz modal a *falsetto,* los locutores analizados aumentan su f_0 de promedio 1,5 octavas. La desviación típica media es mucho mayor en el *falsetto* que en la voz modal.

[49] En este capítulo se retoman de forma muy sintética las conclusiones que se expusieron en el capítulo anterior para poder pensarlas en clave aplicada al ámbito de la fonética judicial.

2) El *falsetto* provoca concomitantemente otras modificaciones, como el cambio de la velocidad de habla, aunque la tendencia observada no es clara: al comparar un registro fonatorio y otro, se comprueba que en la mitad de los locutores la velocidad se acelera, mientras que en la otra mitad disminuye.

3) El análisis con la herramienta *BioMet®Soft* ha resultado muy provechoso para el estudio de los registros modal y *falsetto,* puesto que ha permitido observar en detalle y profundidad el comportamiento glotal de los locutores a partir de muestras de voz grabadas en laboratorio. Los valores correspondientes a los parámetros glotales fueron fácil y cómodamente extraíbles a partir de la señal.

4) El estudio de los 72 rasgos glotales mostró que los parámetros cepstrales (grupo B) fueron los que más resultaron afectados como consecuencia del cambio de registro, sufriendo alteraciones significativas en todos los hablantes. A este grupo le han seguido los rasgos que dan cuenta del funcionamiento biomecánico de los pliegues vocales (grupo D). También se vieron alterados los rasgos vinculados con el perfil espectral (grupo C), mostrando que son también sensibles al cambio de registro. Por último, se verificaron cambios significativos en los parámetros de base temporal de la onda glótica (grupo E) y en los vinculados con la perturbación de la f_0 (grupo A). Los grupos que presentaron una variación significativamente menor fueron los de defecto de cierre glótico (grupo F) y los relacionados con el temblor (grupo G).

5) Del *Análisis por Componentes Principales* también se han extraído conclusiones importantes y alentadoras, pues se ha demostrado que es posible abordar el estudio de la cualidad de voz, en concreto de los parámetros que dan cuenta del comportamiento laríngeo, disminuyendo el número de factores a unos pocos Componentes Principales, independientes e interpretables.

   i. Los parámetros que definieron el primer Componente Principal (CP1) están centrados en la f_0 y en el comportamiento biomecánico de los pliegues tendente a aumentarla o disminuirla (tensión, masa,

irregularidades en la vibración y gastos de energía de la cubierta del pliegue, y tensión y masa del cuerpo). Todos estos rasgos que recogen información tonal aumentan su valor en el *falsetto,* a excepción de la masa del cuerpo puesta en vibración, que disminuye.

ii. Los parámetros recogidos por el segundo Componente Principal (CP2) son los que definen, en el total de la duración de un ciclo glotal, el instante en el que se producen aspectos relevantes durante un ciclo de fonación, como son, los instantes que indican los momentos de apertura y cierre glotal, entre otros.

iii. Los parámetros del tercer Componente Principal (CP3) se asocian principalmente con las amplitudes de la onda.

iv. Los rasgos del cuarto y último Componente Principal (CP4) se centran en el comportamiento del cuerpo del pliegue vocal, y su influencia en la variabilidad temporal de la frecuencia fundamental, conocida como *jitter.*

v. De los resultados obtenidos a partir del contraste de medias entre registros, se concluye que el único componente diferenciador entre registros para todos los locutores analizados es el primer Componente Principal (CP1). Todos los rasgos que aparecen recogidos en el CP1 varían para todos los locutores aumentando su valor (salvo en el caso del parámetro vinculado con la masa del pliegue vocal que disminuye) cuando pasan de voz modal a *falsetto.*

vi. El resto de parámetros recogidos en los CP2, CP3 y CP4 presentaron un comportamiento fluctuante, unas veces cambian entre registros y otras veces no, dependiendo de cada locutor.

6) Resultaron también interesantes los resultados del análisis perceptivo realizado por oyentes profanos para discriminar locutores a partir de voces normales y disimuladas en *falsetto*. Estos mostraron que los oyentes son capaces, bajo ciertas circunstancias, de reconocer si dos voces (modal/*falsetto*) fueron pronunciadas por el mismo locutor. Debe tenerse presente, por supuesto, que estos resultados se dieron en un contexto en el que los oyentes tenían una buena motivación para realizar la tarea, que sus decisiones no tenían una implicación trascendente, se dispuso de óptimas condiciones de grabación

y de escucha, y las muestras de voces eran muy similares excepto por lo que concierne al cambio de registro de fonación.

7) Los oyentes españoles e italianos reconocieron las voces por encima del nivel del azar aunque con un nivel de discriminación muy débil en los dos casos. Ambos grupos de oyentes aplicaron un criterio liberal (tendencia a responder que las voces pertenecían al mismo locutor) aunque entre los españoles esta tendencia fue más acusada. La única diferencia estadísticamente significativa al evaluar el efecto de la lengua en la discriminación de voces fue que los españoles cometieron más Falsas Alarmas, es decir, respondieron más veces que las dos voces pertenecían al mismo locutor cuando procedían, por el contrario, de locutores diferentes.

8) Tanto en el caso de los oyentes-jueces españoles (que desconocen la lengua de los locutores) como en el de los italianos (que la comparten), parece predominar el procesamiento por rasgos sobre el holístico.

9) Los italianos tardaron significativamente más que los españoles en todos los tipos de respuesta aunque dentro de cada grupo de oyentes se apreció el mismo comportamiento en cuanto al tiempo empleado en contestar y el hecho de acertar o errar.

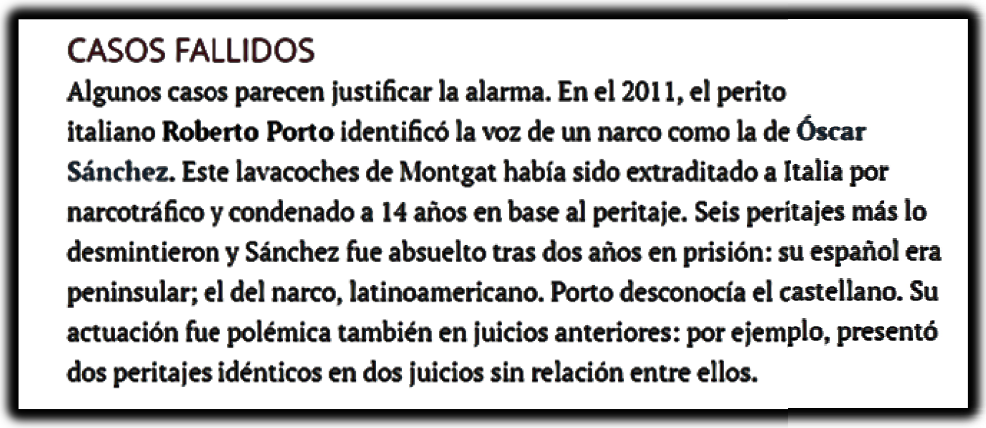
10) Por último, esta investigación corrobora también un interesante hallazgo sobre los tiempos de reacción (TR) y el tipo de respuesta descubierto antes en el ámbito del reconocimiento facial. Los oyentes, independientemente del idioma y de si evaluaban pares de igual locutor o de locutores diferentes, tardaron significativamente más cuando fallaron (Falsas Alarmas y Omisiones) que cuando respondieron correctamente (Aciertos y Rechazos Correctos). Coincidencias como esta animan a seguir confrontando los estudios realizados sobre modalidades perceptivas diferentes.

7.2. Hacia unas implicaciones prácticas de este estudio

Los principales resultados derivados de la presente investigación sobre las dos cuestiones examinadas, la influencia del idioma y el perjuicio del *falsetto* en el reconocimiento de voces, resultan relevantes en el contexto de la fonética forense, como se precisará a continuación.

7.2.1. La influencia de la lengua del locutor

En un artículo periodístico *¿Qué hay en una voz?* realizado por Catanzaro Viciano, Hummel y Tola puede leerse:



CASOS FALLIDOS

Algunos casos parecen justificar la alarma. En el 2011, el perito italiano Roberto Porto identificó la voz de un narco como la de Óscar Sánchez. Este lavacoches de Montgat había sido extraditado a Italia por narcotráfico y condenado a 14 años en base al peritaje. Seis peritajes más lo desmintieron y Sánchez fue absuelto tras dos años en prisión: su español era peninsular; el del narco, latinoamericano. Porto desconocía el castellano. Su actuación fue polémica también en juicios anteriores: por ejemplo, presentó dos peritajes idénticos en dos juicios sin relación entre ellos.

Figura 7.1. Extracto de la nota periodística *¿Qué hay en una voz?*

En la nota periodística, Catanzaro et *al*. (2015) relatan que, en 2011, un perito italiano identificó la voz de un reconocido narcotraficante como la de Óscar Sánchez, un lavacoches de la ciudad catalana de Montgat. Como consecuencia de esa pericia, Sánchez fue extraditado a Italia y condenado a 14 años de prisión. Seis peritajes posteriores, fundamentados en la observación de que el español de Sánchez era peninsular mientras que el del narcotraficante era uruguayo, refutaron la primera valoración. Al parecer, y pese a lo sorprendente que resultó el hecho al finalizar el juicio, el perito italiano desconocía los diferentes acentos regionales y, aún más,

ignoraba el español en cualquiera de sus variedades[50]. Como apuntó el Profesor Cicres en una entrevista[51], el perito italiano "ignoró los códigos éticos de las asociaciones forenses lingüísticas internacionales".

En efecto, y aunque este acontecimiento represente quizás un caso límite, rayano en lo inaudito, en el sexto punto de su código de práctica profesional, aprobado en Helsinki en 2004, es decir siete años antes de la detención de Oscar Sánchez, la *Asociación Internacional de Fonética Forense* (*IAFPA*, por sus siglas en inglés) ya se señalaba:

> Members should exercise particular caution if carrying out forensic analysis of any kind on recordings containing speech in languages of which they are not native speakers[52].

Como comentan Schiller, Köster, Duckworth (1997, p. 1), situaciones como estas son cada vez más corrientes; los peritos expertos que trabajan en el ámbito de la fonética forense participan cada vez más frecuentemente en casos reales en los que se comparan voces de hablantes extranjeros con el fin último de aportar pruebas conducentes a su identificación. Una situación similar se presenta cuando un testigo o una víctima de un delito deben reconocer la voz del delincuente que habla un idioma desconocido en una rueda de reconocimiento.

Las conclusiones de los estudios previos así como los resultados derivados de esta tesis confirman que la familiaridad con el lenguaje efectivamente contribuye a lograr una mayor precisión en el reconocimiento de patrones vocales individuales por parte de los oyentes. Incluso cuando las muestras de voces que se comparan son breves y no están disponibles muchas claves dependientes del idioma, aunque los oyentes nativos y no nativos pudieran ser capaces de reconocer a los locutores a partir de sus voces y, en términos generales, no difieran demasiado en su desempeño, los oyentes no nativos cometen mayor número de Falsas Alarmas. Evidentemente, en situaciones en las

[50]Tomado de *¿Qué hay en una voz?,* nota periodística firmada por Michele Catanzaro, Astrid Viciano, Philipp Hummel y Elizabetta Tola y publicada en "El Periódico", edición en línea, el 11 de noviembre de 2015, disponible en http://www.elperiodico.com/es/noticias/dominical/uso-las-pruebas-voz-los tribunales-4664419 [Consultado el 12/04/16].
Sobre la detención y posterior liberación de Oscar Sánchez, véase también http://www.lavanguardia.com/vida/20120321/54275798315/libertad-montgat-oscar-sanchez.html

[51] http://www.lavanguardia.com/economia/tu-espacio-profesional/20140606/54408749360/linguistia-forense-prueba-oscar-sanchez-carcel.html

[52] https://www.iafpa.net/code.htm [Consultado el 12/04/16]. La referencia a esta advertencia de la IAFPA fue también señalada por Köster y Schiller (1997, p. 26) y por Schiller, Köster, Duckworth (1997, p. 1).

que los oyentes estuvieran expuestos a muestras de habla más extensas y más variadas, el beneficio de compartir o conocer suficientemente la lengua del locutor objetivo sería aún mayor.

Es importante en este punto tener presente la diferencia entre las nociones de percepción y evaluación de la voz y las de reconocimiento e identificación del locutor a través de la voz. Piénsese, por ejemplo, en las comparaciones de voces con fines judiciales. En Gil Fernández (2012 y 2014b, pp. 66-73) se recuerda el procedimiento que puede aplicarse en una situación de comparación de voces con fines judiciales. Primeramente, a partir de una escucha minuciosa de las muestras recibidas, los expertos lingüistas/fonetistas deducen distinto tipo de información sobre la persona que escuchan para trazar un perfil personal (sexo, edad aproximada, estado de salud general, etc.), lingüístico (dialecto, rasgos léxicos, morfosintácticos, discursivos) y social (nivel educativo, estrato social, etc.). En segundo lugar, además de la descripción inferencial de esta imagen del hablante confeccionada a partir de las particularidades encontradas en las muestras de voz, es decir, además del "pasaporte vocal", en la denominación dada por Gil Fernández (2014b, p. 66) a esta tarea, los expertos pueden también realizar otro análisis. A partir de repetidas escuchas de las grabaciones buscan concentrar sus esfuerzos perceptivos únicamente en la voz abstrayéndose lo más posible de los rasgos lingüísticos ya analizados y registrados en el pasaporte vocal. Para este segundo análisis, suele emplearse algún protocolo perceptivo, como por ejemplo, el *Vocal Profile Analysis*, VPA, (Laver, Wirz, Mackenzie, Hiller, 1991; Mackenzie Beck, 2007, etc.), *Grade, Roughness, Breathiness, Asthenia, Strain Scale,* más conocido como *GRBAS*, por sus siglas en inglés,[53] (por ejemplo, Hirano, 1981), o cualquier otro. Para conocer más detalles sobre los protocolos perceptivos disponibles para evaluar la cualidad de la voz pueden consultarse los trabajos de Gil Fernández y San Segundo (2014a) y (2015).

[53] Se trata en ambos casos de protocolos perceptivos que analizan la cualidad de voz de forma componencial basándose en la fisiología de la producción. En el GRBAS se analiza únicamente la fonación mientras que el VPA considera el tracto vocal en su conjunto. En España, el GRBAS es quizás más frecuentemente utilizado en el ámbito médico mientras que el *Vocal Profile Analysis* ha resultado más interesante para el análisis y la comparación de voces con fines judiciales. Este último protocolo fue diseñado inicialmente para trabajar con hablantes de inglés. Camargo y Madureira (2008) realizaron una adaptación a la lengua portuguesa y, más tarde, Infante y Fernández Trinidad (2014) realizaron una primera propuesta de adecuación al español. Trabajos recientes como los de San Segundo, Foulkes, French, Harrison, y Hughes (2016), San Segundo y Mompeán (2017), entre otros, acomodan el esquema propuesto por el VPA para que responda más exactamente a las necesidades del trabajo realizado en la comparación forense de voces.

Por supuesto, la ventaja de compartir la lengua es muy alta, principalmente para la realización del "pasaporte vocal" o para intentar reconocer la identidad de un locutor. En casos como estos, se vuelve imprescindible la participación de personas conocedoras del idioma o incluso de hablantes nativos de la misma lengua y variedad regional del locutor, al igual que para armar ruedas de reconocimiento de voces de forma equilibrada y sin sesgos. Ahora bien, para evaluar cuestiones relativas a la cualidad de la voz, la actuación de expertos que no tengan un gran manejo o conocimiento de la lengua bajo análisis podría ser provechosa igualmente en la medida en que se podrían quizás percibir con mayor facilidad los rasgos laríngeos, ladeando consustancialmente aquellos que sean propiamente lingüísticos.

La importancia de analizar la cualidad de la voz en el ámbito de la fonética judicial fue señalada ya por varios expertos, en Gil Fernández y San Segundo (2014a, pp. 14-20) se comentan, por ejemplo, los trabajos de French (1994), Jessen (2008), Künzel (1994b), Nolan (1983, 2005 y 2007), Rose (2002). Como también ha expresado Gil Fernández (2012, 2013a y b, 2014a), la cualidad de la voz debería ser considerada más seriamente en las comparaciones o cotejos forenses pero quizás por ser el resultado de la combinación de varios parámetros, compleja de analizar y de interpretar, se la excluye a menudo de tales análisis[54]. Sin embargo, su poder discriminante parece ser importante (recuérdese nuevamente González-Rodríguez *et. al*., 2014) y esta tesis demuestra que es perfectamente posible reducir considerablemente la dimensionalidad de los rasgos laríngeos que la caracterizan y la determinan.

Aunque los análisis perceptivos desarrollados en la presente investigación mostraron que la magnitud global del efecto de la lengua sobre la discriminación no resultó significativa, se observó, sin embargo, que la proporción de Falsas Alarmas aumentaba significativamente en el grupo de oyentes españoles y que estos mostraban asimismo una tendencia mayor a señalar el objetivo, es decir, a responder afirmativamente que las voces pertenecían al mismo locutor. Esto podría ser, una vez más, de suma relevancia práctica. Como se viene señalando en múltiples trabajos, ciertamente son las Falsas Alarmas los casos más preocupantes en la medida en que, en situaciones reales, implican el reconocimiento erróneo de una persona inocente y, al mismo tiempo, la libertad del delincuente (véase Wells, Small, Penrod, Malpass, Fulero y Brimacombe, 1998).

[54] Para conocer más razones por las que en ocasiones se excluye el análisis de la cualidad de la voz en la práctica judicial consúltense las obras de Nolan (2005 y 2007).

Es claro que un mejor conocimiento de los factores que pueden influir en la exactitud de los reconocimientos contribuiría a minimizar los errores que puedan cometer los testigos auditivos. Esta investigación comprueba que, entre tales factores o variables, es muy importante considerar la lengua del oyente/testigo, principalmente en los casos de objetivo ausente, así como la complejidad de la tarea perceptiva. En futuras aproximaciones al tema, se buscará corroborar estos resultados realizando el análisis inverso: locutores nativos de español centro-peninsular serán juzgados por oyentes del mismo dialecto y por oyentes italianos sin conocimientos de español en tareas perceptivas de discriminación igual-diferente. Los locutores se seleccionarán entre los registrados en el corpus *Cualidad Individual de la Voz e Identificación del Locutor (CIVIL)*, realizado en el Laboratorio de Fonética del CSIC[55], puesto que cuenta con grabaciones de laboratorio en voz modal y *falsetto*, integradas por la lectura de frases equivalentes a las analizadas en esta tesis, todo lo cual favorecerá enormemente la comparación.

7.2.2. El perjuicio causado por el falsetto

Nuevamente, en la misma nota periodística de Catanzaro se relata[56]:

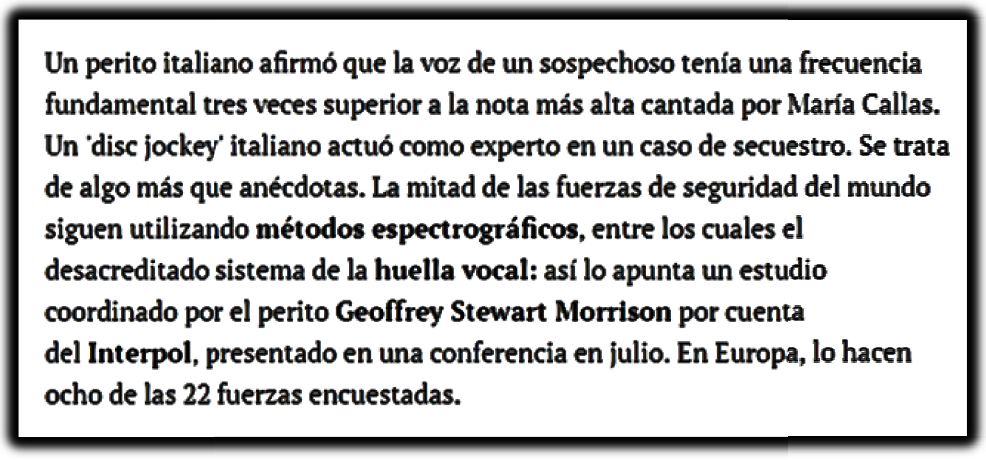

Un perito italiano afirmó que la voz de un sospechoso tenía una frecuencia fundamental tres veces superior a la nota más alta cantada por María Callas. Un 'disc jockey' italiano actuó como experto en un caso de secuestro. Se trata de algo más que anécdotas. La mitad de las fuerzas de seguridad del mundo siguen utilizando métodos espectrográficos, entre los cuales el desacreditado sistema de la huella vocal: así lo apunta un estudio coordinado por el perito Geoffrey Stewart Morrison por cuenta del Interpol, presentado en una conferencia en julio. En Europa, lo hacen ocho de las 22 fuerzas encuestadas.

Figura 7.2. Extracto de la nota periodística *¿Qué hay en una voz?*

[55] Recuérdese que puede encontrarse una descripción de este corpus en San Segundo, Alves y Fernández Trinidad (2013).
[56] Tomado de *¿Qué hay en una voz?,* nota periodística realizada por Michele Catanzaro, Astrid Viciano, Philipp Hummel y Elizabetta Tola y publicada en "El Periódico", edición online, el 11 de noviembre de 2015, disponible en http://www.elperiodico.com/es/noticias/dominical/uso-las-pruebas-voz-los tribunales-4664419 [Consultado el 12/04/16].

El cotejo de voces con finalidades judiciales puede realizarse por expertos que aplican en sus análisis el método "fonético clásico o tradicional" y el automático o bien exclusivamente por medio de sistemas automáticos de reconocimiento (véase Gil Fernández 2014b; Künzel, 1994; Praveena y Krishna, 2015; Watt, 2010; para mayores detalles). Los métodos automáticos son, lógicamente, más rápidos y objetivos pero también fallan. Se ha visto que, en los casos en los que las grabaciones tienen muchas interferencias, en aquellos en los que las muestras que se comparan no se han grabado a través de los mismos canales de transmisión y en aquellos otros en los que una de las muestras es una voz disimulada (máxime si el disimulo es laríngeo), los métodos automáticos no son mejores que los métodos humanos (véase por ejemplo, Gil Fernández, 2014a; Gil Fernández, 2014b, pp. 70-73, y González-Rodríguez *et al*., 2014)[57].

Los resultados de los experimentos llevados a cabo en esta tesis, por una parte, llaman la atención sobre la capacidad perceptiva de las personas, en escuchas cuidadosas y muy controladas, para reconocer matices vocales (a veces sutiles) que les permiten reconocer diferencias y similitudes entre las voces. De otra parte, los estudios acústicos realizados revalidan la utilidad de los expertos fonetistas para los análisis que

[57] Resulta particularmente interesante resumir aquí el reciente trabajo de González-Rodríguez *et al*. (2014, pp. 33-40, también presentado en Gil Fernández 2014a). Un grupo de fonetistas junto con uno de ingenieros realizaron un experimento para intentar esclarecer hasta qué punto el oído humano entrenado es capaz de detectar información relevante sobre las voces que permita descartar la hipótesis de que dos emisiones de habla fueron pronunciadas por la misma persona. Más aún, ante estímulos vocales de distintos locutores que los sistemas de reconocimiento automáticos no fueron capaces de reconocer como tales sino que, al contrario, identificaron equivocadamente como producidos por un mismo locutor. Se trabajó a partir de 66 conversaciones telefónicas espontáneas en inglés, las cuales habían sido previamente utilizadas en las sesiones organizadas por el *National Institute of Standards and Technology* (NIST, 2010) precisamente para poner a prueba la eficacia de los sistemas de reconocimiento automáticos. En concreto, los fonetistas trabajaron con pares de voces erróneamente identificadas como la misma persona por los sistemas automáticos. Se trató de un análisis eminentemente perceptivo confirmado por un estudio acústico posterior a la escucha. Primeramente, los dos fonetistas expertos oyeron a la vez las 66 grabaciones, todas ellas conversaciones telefónicas (fijo o móvil) con una duración aproximada de 5 minutos, de las dos voces identificadas erróneamente como pertenecientes a la misma persona. En una segunda fase, los expertos fonetistas seleccionaron 18 pares de voces correspondientes a 9 mujeres y 9 hombres. Luego, a partir de una escucha pormenorizada conjunta, reflejaron las características vocales de cada voz siguiendo el protocolo perceptivo propuesto por Kreiman y Sidtis (2011). En una cuarta fase, realizaron por separado una escucha adicional y cotejaron las impresiones perceptivas de cada uno. Por último, realizaron un análisis acústico de corroboración con *Praat* (pp.35-36). Los autores explican cuáles fueron los principales parámetros que permitieron el reconocimiento de los pares de voces como pertenecientes a locutores diferentes. Entre ellos los autores del trabajo mencionan aspectos segmentales, como la pronunciación particularmente constante de algún sonido específico, las características temporales del habla, las prosódicas y asociadas con la modulación de la f_0 y, sobre todo, la presencia sostenida de algún modo de fonación distinto a la fonación modal (*creak/creaky voice*, *falsetto*, *breathy voice*, entre otras). Este último aspecto de la cualidad de voz, comentan los autores, se reveló como fundamental para la discriminación de voces y, en consecuencia, de locutores (pp.36- 39).

involucran comparaciones de voces que tienen como finalidad esclarecer aspectos relativos al reconocimiento de locutores. Finalmente, y al igual que en el trabajo de González-Rodríguez y colaboradores recién comentado, reclaman la utilización de información no cepstral para la discriminación de voces y el reconocimiento de locutores (véase González-Rodríguez *et al*., 2014, p. 39).

Una consecuencia interesante derivada de los resultados de esta tesis es que quizás no deberían rechazarse *a priori* pericias en las que ha habido disimulo de la voz mediante *falsetto*, un claro ejemplo en el que las muestras de habla dubitada e indubitada son muy diferentes entre sí. Aunque por ahora los reconocedores automáticos no consiguen obtener resultados muy buenos con habla disimulada en *falsetto*, el método perceptivo-acústico, sin embargo, podría ofrecer un rendimiento más aceptable. Cuando se cotejan muestras de voz, una de las primeras tareas que hay que realizar es determinar qué aspectos van a ser estudiados y comparados. No existe acuerdo absoluto sobre cuáles podrían ser los parámetros más importantes para establecer la individualidad de una voz y por tanto se prefiere hablar de una importancia relativa de todos ellos en función de las distintas circunstancias (véase para una revisión bibliográfica completa a este respecto Battaner, Gil, Marrero, Llisterri, Carbó, Machuca, …Ríos, 2003). Como ha sido señalado por ejemplo en Gil Fernández (2012, 2014a y b) y se explicó en el Capítulo 2 (§2.3.1), no existe una lista preestablecida de tales parámetros ordenados por relevancia o poder discriminador, sino que la selección de los rasgos o parámetros que van a tratarse normalmente varía porque depende de las características de cada caso particular. Gil Fernández en varios de sus trabajos (2013a y 2013b, 2014b) comenta que para seleccionar los parámetros que se habrán de analizar, se valora que cumplan con ciertos requisitos, como por ejemplo, que presenten un buen número de observaciones en las muestras de habla, que sean fáciles de calcular y de interpretar; que presenten mínima variación intralocutor pero máxima variación entre locutores diferentes, que no sean parámetros interdependientes y que no se puedan disimular (Gil Fernández, 2014b, pp. 74-75)[58]. No se ha podido asilar ―probablemente porque no existe―un parámetro que cumpla con todas estas exigencias, por ello en cada caso los expertos sopesan *pros y contras* que supondría tratar uno en concreto y su capacidad para aportar información útil sobre la individualidad del locutor, es decir, su poder discriminante (Gil Fernández, 2014b, p. 75). Como se ha señalado en varios

[58] Esta lista de requisitos está a su vez tomada y adaptada de Nolan (1983) y Rose (2002), como indica la autora.

estudios (por ejemplo, Gil Fernández 2012, 2014b; Marrero, Gil, Battaner, 2003) la frecuencia fundamental parece tener un alto poder discriminante. En todo peritaje se analizan la f_0 media junto con los valores de dispersión y distribución a ella asociados (su rango, la f_0 máxima y la mínima, la desviación típica, etc.) aunque los resultados nunca son definitorios.

En el caso de disimulo por medio de *falsetto* evidentemente los valores relativos a la f_0 serán, como se ha visto, muy diferentes y difícilmente comparables entre las muestras de voz cotejadas. El análisis acústico detallado realizado en esta tesis confirma que, lógicamente, los valores de f_0 —así como los biomecánicos de los pliegues que contribuyen de manera más importante a aumentar o disminuir el tono, es decir, que lo determinan de forma decisiva— varían en todos los locutores debido al cambio de registro y, el mismo análisis explica en qué medida se ven modificados. Sin embargo, parece que es posible aislar otros parámetros que no varían de forma sistemática o que se mantienen constantes en los locutores a pesar del cambio de registro y que, por ello, podrían estar asociados con su individualidad. Tales parámetros parecen ser rasgos de mínimo detalle, probablemente más difíciles de percibir pero también de reproducir, disimular o imitar. Según se ha demostrado a partir del análisis con la herramienta *BioMet®Soft*, los rasgos resultan relativamente fáciles de extraer y de medir, y su posterior Análisis por Componentes Principales no solo permite reducir enormemente la gran dimensionalidad de la voz, sino que además pone de manifiesto la existencia de componentes ortogonales, es decir, independientes entre sí.

A partir de los resultados de este estudio es posible aislar, más claramente, los parámetros que son sensibles al cambio de registro y no considerarlos en el cotejo o, más bien, intentar descontar su efecto en el análisis, de modo análogo a como se busca hacer cuando hay otro tipos de distorsiones provocadas por el filtrado telefónico o por el habla alcoholizada, por ejemplo. Este trabajo confirma que el registro de *falsetto* altera necesariamente los parámetros del Componente Principal 1 y que, por tanto, la información laríngea identificadora dependiente del locutor, si la hay, tiene, necesariamente que encontrarse en los restantes componentes y rasgos. Investigaciones y trabajos forenses futuros deben seguir explorando conjuntamente esta posibilidad. Según se ha desprendido claramente del Análisis por Componentes Principales realizado en el presente estudio, en el *falsetto*, es fundamental el papel que desempeñan los rasgos del Componente 1 (f_0 y comportamiento de la cubierta), mientras que es pobre la influencia de los rasgos que constituyen el Componente 4 (desempeño del

cuerpo del pliegue); como se ha explicado, la intervención del cuerpo en el *falsetto* es muchísimo menor.

Un último argumento para pensar que se podría estar algo más cerca de atenuar las dificultades derivadas del disimulo mediante la elevación drástica de la f_0, y un aliciente para continuar investigando los efectos de este mecanismo de fonación en la percepción de la cualidad de voz y en el reconocimiento de locutores, es la existencia del corpus CIVIL, antes mencionado. Se trata de un corpus grande que sirve como población de referencia para ambos registros de fonación. Aunque el corpus ha sido grabado en español, las diferencias entre lenguas para describir el comportamiento laríngeo no parecen tener demasiada importancia, al menos para casos de lenguas tipológicamente muy similares.

A pesar de algún posible adelanto que pueda entrañar este trabajo, no debe olvidarse que aún queda mucho por descubrir sobre los efectos del *falsetto* y de la lengua en el reconocimiento de hablantes. Esta tesis pretende, así, ser una pequeña contribución para avanzar en el recorrido de aquellos que estén interesados en estudiarlos.

8. Referencias bibliográficas

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: University of Edinburgh.

Aguete Cajiao, A., Fernández Rei, E., y Osório Peláez, C. (2016). FOLERPA: A tool for building and conducting perceptual experiments. *Dialectologia: revista electrònica*, *VI*, 245-275.

Allen, E. L., y Hollien, H. (1973). Vocal fold thickness in the pulse (vocal fry) register. *Folia Phoniatrica*, *25*, 241–250.

Alves, H., Fernández Trinidad, M., Gil Fernández, J., Infante, P., Lahoz-Bengoechea, J. M., Pérez Sanz, C., y San Segundo, E. (2012). Disguised voices: A perceptual experiment. Presentado en 3rd European Conference of the International Association of Forensic Linguistics, Porto, 15-18 Octubre, 2012.

Alves, H., Gil Fernández, J., Pérez Sanz, C., y San Segundo, E. (2014). La cualidad individual de la voz y la identificación del locutor: el proyecto CIVIL. En Y. Congosto, M. L. Montero Curiel, y A. Salvador Plans (Eds.), *Fonética experimental, educación superior e investigación* (Vol. 1, pp. 591-612). Madrid: Arco/Libros.

Baayen, R. (2008). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge: Cambridge University Press.

Ballesteros, S. (1997). *Evaluación de la actuación humana con la Teoría de Detección de Señales* [vídeo+libro]. Madrid: Universidad Nacional de Educación a Distancia.

Barfüßer, S., y Schiel, F. (2010). Disfluencies in alcoholized speech [resumen]. Presentado en IAFPA 2010.

Battaner, E., Gil, J., Marrero, V., Llisterri, J., Carbó, C., Machuca, M. J., . . . Ríos, A. (2003). VILE: Estudio acústico de la variación inter e intralocutor en español. In *SEAF 2003. Actas del II Congreso de la Sociedad Española de Acústica Forense*. (pp. 59-70). Barcelona: Sociedad Española de Acústica Forense. http://liceu.uab.cat/~joaquim/phonetics/VILE/ VILE_SEAF03.pdf

Baumeister, B., y Schiel, F. (2015). Fundamental frequency and human perception of

alcoholic intoxication in speech. *Proceedings of the 18th International Congress of Phonetic Sciences* (ICPhS 18-Glasgow, 10–14 August 2015).

BioMetroSoft®. (2014). BioMet®Phon: Tool for de Evaluation of Voice Quality and Biometri. User's Manual [versión 2.3, December 2012].

Braun, A. y Künzel, H. (2003). The effect of alcohol on speech prosody. En M. J. Solé, D. Recasens y J. Romero (Eds.). *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 15-Barcelona)*. Barcelona: Causal Productions.

Brewer, N., Caon, A., Todd, C., y Weber, N. (2006). Eyewitness identification accuracy and response latency. *Law and Human Behavior*, *30*(1), 31-50.

Bricker, P., y Pruzansky, S. (1966). Effects of stimulus content and duration on talker identification. *The Journal of the Acoustical Society of America*, *40*(6), 1441-1449.

Byrne, C., y Foulkes, P. (2004). Themobile phone effect'on vowel formants. *International Journal of Speech Language and the Law*, *11*, 83-102.

Camargo, Z., y Madureira, S. (2008). Voice quality analysis from a phonetic perspective: Voice profile analysis scheme (VPA) profile for brazilian portuguese. En P. Barbosa, S. Madureira, y C. Reis (Eds.). *Proceedings of Speech Prosody* (pp. 57-60).

Campbell, N. (2007). Changes in voice quality due to social conditions. *Proceedings of the 16th ICPhS, Saarbrücken*, 2093-2096.

Catanzaro, M., Viciano, A., Hummel, P., y Toa, E. (2015). ¿Qué hay en una voz? *El Periódico*.

Catford, J. C. (1964). Phonation types: the classification of some laryngeal components of speech production. En David Abercrombie, D. B. Fry, P. MacCarthy, N. Scott, y J. Trim (Eds.), *In Honor of Daniel Jones* (pp. 26–37). London: Longmans Green.

Cervera, J., y Núñez, F. (2013). Estructura histológica de la cuerda vocal. En *Patologías de la voz* (pp. 47-54). Barcelona: Marge Medica Books.

Chance, J., Goldstein, A., y McBride, L. (1975). Differential experience and recognition memory for faces. *The Journal of Social Psychology*, *97*(2), 243 253.

Childers, D., Hicks, D., Moore, G., y Alsaka, Y. (1986). A model for vocal fold vibratory motion, contact area, and the electroglotogram. *The Journal of the Acoustical Society of America*, *80*, 1309-1320.

Childers, D., Hicks, D., Moore, G., Ezkenazi, L., y Lalwani, A. (1990). Electroglottograpy and vocal fold physiology. *Journal of Speech, Language, and Hearing Research*, (33), 245-254.

Childers, D., y Krishnamurthy, A. (1985). A critical review of electroglottography. *Journal, Critical Reviews in Biomedical Engineering*, *12*(2), 131-164.

Childers, D., y Lee, C. K. (1991). Vocal quality factors: Analysis, synthesis, and perception. *the Journal of the Acoustical Society of America*, *90*(5), 2394-2410.

Cobeta, I., y Núñez, F. (2013). Laboratorio de voz. Análisis de la señal acústica. En I. Cobeta, F. Núñez, y S. Fernández (Eds.), *Patologías de la voz* (pp. 188-198). Barcelona: Marge Medica Books.

Coll, R. (2013). Valoración logopédica del paciente disfónico. En *Patologías de la voz* (pp. 135-145). Barcelona: Marge Medica Books.

Colton, R.H. (1969). Some acoustic parameters related to the perception of modal-*falsetto* voice quality. *Folia Phoniatrica*, *25*, 270-280.

Colton, R. H. (1973). Vocal intensity in the modal and falsetto registers. *Folia Phoniatrica*, *25*, 62-70.

Colton, R.H., y Hollien, H. (1973). Perceptual differentiation of the modal and falsetto registers. *Folia Phoniatrica et Logopaedica*, *25*(4), 270–280.

Davis, S. (1979). Acoustic characteristics of normal and pathological voices. En N. Lass (Ed.), *Speech and language: Advances in basic research and practice* (pp. 271-335). New York, NY: Academic.

Delgado, C. (2001). *La identificación de locutores en el ámbito forense* (Tesis de Doctorado), Universidad Complutense de Madrid, Madrid.

Doty, N. (1998). The influence of nationality on the accuracy of face and voice recognition. *The American Journal of Psychology*, *111*(2), 191-214.

Dunning, D., y Perretta, S. (2002). Automaticity and eyewitness accuracy: A 10- to 12-second rule for distinguishing accurate from inaccurate positive identifications. *Journal of Applied Psychology*, *87*, 951–962.

Dunning, D., y Stern, L. (1994). Distinguishing accurate from inaccurate eyewitness identifications via inquiries about decision processes. *Journal of Personality and Social Psychology*, *67*(1994), 818-835.

Esling, J. (1978). The identification of features of voice quality in social groups. *Journal of the International Phonetic Association*, *8*(1-2), 18-23.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Federico, A., Mori, L., y Paolini, A. (2005). La velocità di articolazione come parametro identificativo: potenzialità e limiti per la caratterizacione del parlante. En P. Cosi (Ed.), *AISV 2004. Misura dei parametri. Aspetti tecnologici ed implicazioni nei modelli linguistici. Atti del 1o Convegno Nazionale AISV - Associazione Italiana di Scienze della Voce. Università di Padova, 2-4 Diciembre 2004* (pp. 869-876). Brescia: EDK Editore.

Fernández Planas, A. M. (2015). ¿Qué tiene que ver la fonética con la práctica clínica y logopédica? Algunos ejemplos de colaboración interdisciplinar. *Normas*, *5*(1), 51-65.

Fernández Rei, E. (2014). *FOLERPA: Ferramenta On-Line para ExpeRimentación PerceptivA*. Recuperado a partir de http://ilg.usc.es/FOLERPA/

Fernández Rei, E, y Moutinho, L. (2012). Contributo para o estudo percetivo de variantes dialectais prosódicas na fronteira galaico-portuguesa. Presentado en I Jornadas de Ciências da Linguagem, Aveiro (Portugal), 20 de junio de 2012.

Fernández Trinidad, M. (2015). La percepción de la cualidad de voz y los estereotipos vocales. *Revista Española de Lingüística*, *45*(1), 45-72.

Fernández Trinidad, M.; Gil Fernández, J.; Infante Ríos, P., y Lahoz-Bengoechea, J. M. (2014). "El *falsetto* como procedimiento de disimulo de la voz: rasgos alterables y rasgos permanentes", comunicación presentada en el *VI Congreso Internacional de Fonética Experimental*, Valencia 5-7 de noviembre de 2014.

Fernández Trinidad, M., Infante, P., Lahoz-Bengoechea, J. M., y Alves, H. (2013). *Falsetto* as a disguise method in male voices. Presentado en 31st International Conference AESLA, Universidad de La Laguna, Tenerife.

Fernández Trinidad, M., y Rojo Abuín, J. (2018). Perceptual cues for individual voice quality. En Juana Gil Fernández y M. Gibson (Eds.), *Romance Phonetics and Phonology*. Oxford: Oxford University Press.

Figueiredo, R., y de Souza Britto, H. (1996). A report on the acoustic effects of one type of disguise. *Forensic Linguistics*, *3*, 168-175.

Folerpa: Ferramenta On-Line para ExpeRimentación PerceptivA (s.f). *Manual de uso.*

Fonagy, I. (1978). A new method of investigating the perception of prosodic features. *Language and Speech*, *21*, 34-49.

French, P. (1994). An overview of forensic phonetics with particular reference to speaker identification, *Forensic Linguistics: The International Journal of Speech, Language and the Law* 1 (2), pp. 197-206.

Frick, R. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin*, *97*(3), 412-429.

Gerhardt, K., y Abrams, R. (2000). Fetal exposures to sound and vibroacoustic stimulation. *Journal of Perinatology*, *20*, 1-21.

Gil Fernández, J. (2007). *Fonética para profesores de español: de la teoría a la práctica*. Madrid: Arco/Libros.

Gil Fernández, J. (2012). *La cualidad de voz y la identificación del hablante*. Ponencia plenaria inaugural, II Jornadas de Lingüística Forense, Universidad Autónoma de Madrid.

Gil Fernández, J. (2013a). *El papel de la fonética en la identificación del locutor*. Ponencia plenaria inaugural, I Congreso Colombiano de Jóvenes Lingüistas Instituto Caro y Cuervo / Universidad Nacional de Colombia, Bogotá.

Gil Fernández, J. (2013b). *Voz e identidad*. Ponencia de clausura, 31 Congreso de la Asociación Española de Lingüística Aplicada Universidad de La Laguna, Tenerife.

Gil Fernández, J. (2014a). *El hombre y la máquina en fonética judicial*. Ponencia plenaria, VI Congreso Internacional de Fonética Experimental Universidad de Valencia, Valencia.

Gil Fernández, J. (2014b). Más allá del «efecto CSI»: Avances y metas en Fonética Judicial. En Y. Congosto, M. Montero, y A. Salvador Plans (Eds.), *Fonética experimental, educación superior e investigación* (Vol. 1, pp. 63-111). Madrid: Arco/Libros.

Gil Fernández, J., Fernández Trinidad, M., Infante, P. y Lahoz-Bengoechea, J. M. (2017). "Obtaining speech samples for research and expertise in forensic phonetics". En: Orletti, F. y Mariottini, L. (Eds.) *Theories, Practices, Instruments of Forensic Linguistics* (pp. 27-50). Cambridge: Cambridge Scholars Publishing.

Gil Fernández, J. y San Segundo, E. (2014a). La cualidad de voz en fonética judicial. En E. Garayzábal, M. Jiménez y M. Reigosa (Coords.), Lingüística Forense. *La Lingüística en el ámbito legal y policial* (pp. 154 -199). Madrid: Euphonía Ediciones.

Gil Fernández, J. y San Segundo, E. (2014b). El disimulo de la cualidad de voz en fonética judicial: estudio perceptivo de la hiponasalidad. En A. Penas (Ed.), *Panorama de la fonética española actual* (pp. 321-366). Madrid: Arco / Libros. Consultado en: https://www.researchgate.net/publication/266891191_La_cualidad_de_voz_en_fonetica_judicial (26 pp.)

Gil Fernández, J., y San Segundo, E. (2015). Nuevas aportaciones al estudio de la percepción del habla. En J. Gil Fernández y E. San Segundo (Eds.), *La percepción del habla. Revista Española de Lingüística, 41(1)* (pp. 7-21). Madrid: Sociedad Española de Lingüística.

Giles, H. (1984). *The dynamics of speech accommodation*. Berlin: Mouton.

Gobl, Ch., y Ní Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication*, *40*(1), 189–212.

Gobl, C., y Ní Chasaide, A. (2010). Voice source variation and its communicative functions. En W Hardcastle, J. Laver, y F. Gibbon (Eds.), *The handbook of phonetic sciences* (2.ª ed., pp. 378–423). Oxford: Wiley-Blackwell.

Godino, J., y Gómez-Vida, P. (2013). Notas sobre acústica vocal. En I. Cobeta, F. Núñez, y S. Fernández (Eds.), *Patologías de la voz* (pp. 76-109). Barcelona: Marge Medica Books.

Godino, J., Gómez-Vida, P., y Blanco, M. (2006). Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters. *IEEE Transactions on Biomedical Engineering*, *53*(10), 1943–1953.

Goggin, J., Thompson, C., Strube, G., y Simental, L. (1991). The role of language familiarity in voice identification. *Memory y cognition*, *19*(5), 448-458.

Gold, E. (2012). Articulation rate as a discriminant in forensic speaker comparisons. En *UNSW Forensic Speech Science Conference, Proceedings*. Sydney.

Gold, E., y French, P. (2011). International practices in forensic speaker comparison.

International Journal of Speech, Language and the Law, *18*(2), 293-307.

Goldstein, A., Knight, P., Bailis, K., y Conover, J. (1981). Recognition memory for accented and unaccented voices. *Bulletin of the Psychonomic Society*, *17*(5), 217-220.

Goldstein, B. (2006 [1988]). *Sensación y percepción* (6.ª ed.). Madrid: Thomson.

Gómez-Vilda, P., Álvarez-Marquina, A., Tsanas, A., Lázaro-Carrascosa, C., Rodellar-Biarge, V., Nieto-Lluis, V., y Martínez-Olalla, R. (2016). Phonation Biomechanics in Quantifying Parkinson's Disease Symptom Severity. En *Recent Advances in Nonlinear Speech Processing, Smart Innovation, Systems and Technologies* (pp. 93-102). Springer International Publishing.

Gómez-Vilda, P., Fernández-Baillo, R., Rodellar-Biarge, V., Nieto-Lluis, V., Álvarez-Marquina, A., Mazaira-Fernández, L., M., … Godino-Llorente, J. (2009). Glottal source biometrical signature for voice pathology detection. *Speech Communication*, *51*(9), 759-781.

Gómez-Vilda, P., y Nieto-Lluis, V. (2015). *Informe técnico para documentar la funcionalidad de la aplicación informática BioMet®Phon* (Documento interno). Grupo de informática aplicada al procesamiento de señal e imagen (GIAPSI).

Gómez-Vilda, P., y Pérez Sanz, C. (s. f.). *Factores biomecánicos de la voz pulsada, crepitante, modal y falsetto (inédito)*.

Gómez-Vilda, P., Rodellar-Biarge, M., Nieto-Lluis, V., Martínez-Ollalla, R., Álvarez-Marquina, A., Scola Yurrita, B., … Fernández-Fernández, M. (2013). BioMet®Phon: A System to Monitor Phonation Quality in the Clinics. *Proceedings eTELEMED 2013: The Fifth Int. Conf. on e-Health, Telemecidine and Social Medicine* (pp. 253-258). Nice, France.

González Ceria, J. (2016). Un estudio acústico sobre los aspectos temporales del discurso bajo la influencia del alcohol en hablantes del español. *Loquens*, *3*(1), 10-3989.

González-Rodríguez, J., Gil, J., Pérez, R., y Franco-Pedroso, J. (2014). What are we missing with i-vectors? A perceptual analysis of i-vector based falsely accepted trials. *Proceedings of Odyssey* (pp. 33-40).

Gordon, M., y Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. *Journal of Phonetics*, *29*(4), 383–406.

Green, D., y Swets, J. (1966). *Signal Detection Theory and Psychophysics*. New York, NY: Wiley.

Gunter, H. (2003). A mechanical model of vocal-fold collision with high spatial and temporal resolution. *The Journal of the Acoustical Society of America*, *113*(2).

Hanson, H. M. (1997). Glottal characteristics of female speakers: Acoustic correlates. *The Journal of the Acoustical Society of America*, *101*(1), 466–481.

Henrich, N. (2006). Mirroring the voice from Garcia to the present day: some insights into singing voice registers. *Logopedics Phoniatrics Vocology*, *31*(1), 3–14.

Henrich, N., d'Alessandro, C., Castellengo, M., y Doval, B. (2005). Glottal open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency. *The Journal of the Acoustical Society of America*, *117*(3), 1417-1430.

Henton, C. G., y Bladon, A. (1987). Creak as a sociophonetic marker. En L. M. Hyman y C. N. Li (Eds.), *Language, speech and mind* (pp. 29-58). London: Routledge.

Hepper, P., Scott, D., y Shahidullah, S. (1993). Newborn and fetal response to maternal voice. *Journal of Reproductive and Infant Psychology*, *11*(3), 147-153.

Hirano, M. (1981). *Clinical examination of voice*. New York, NY: Springer Verlag.

Hirano, M. (1982). The role of the layer structure of the vocal fold in register control. En P. Hurme (Ed.), *Vox humana* (pp. 50-62). Jyvaskyla: University of Jyvaskyla.

Hirano, M., Yoshida, Y., y Tateishi, O. (1985). Strobofiberscopic video recording of vocal fold vibration. *Annals of Otology, Rhinology y Laryngology*, *94*(6), 588–590.

Hirson, A., y Duckworth, M. (1993). Glottal fry and voice disguise: A case study in forensic phonetics. *Journal of Biomedical Engineering*, *15*(3), 193-208.

Hollien, H. (1974). On vocal registers. *Journal of Phonetics*, *2*, 125-143.

Hollien, H. (1990). *The Acoustics of Crime: The New Science of Forensic Phonetic*. New York, NY: Plenum Press.

Hollien, H, y Colton, R. (1969). Four laminagraphic studies of vocal fold thickness. *Folia Phoniatrica et Logopaedica*, *21*(3), 179–198.

Hollien, H., Girard, G., y Coleman, R. (1977). Vocal fold vibratory patterns of pulse register phonation. *Folia Phoniatrica et Logopaedica*, *29*(3), 200-205.

Hollien, H., y Michel, J. F. (1968). Vocal fry as a phonational register. *Journal of Speech,*

Language, and Hearing Research, *11*(3), 600–604.

Holmes, J. (1973). The influence of glottal waveform on the naturalness of speech from a parallel formant synthesizer. *IEEE transactions on Audio and Electroacoustics*, *21*(3), 298–305.

Howard, D. (2009). Electroglottography/Electrolaryngography. En M. Fried y A. Ferlito (Eds.), *The larynx*. San Diego, USA: Plural Press.

Infante, P. (2015). ¿Son distintos el creak y la voz creaky? *Revista Española de Lingüística*, *45*(1), 105-128.

Infante, P., y Fernández Trinidad, M. (2014). ¿Cómo trabajar con el *Vocal Profile Analysis* en español? Presentado en XI Congreso Internacional de Lingüística General, Universidad de Navarra, Pamplona 21-23 de mayo de 2014.

Ishi, C. T., Sakakibara, K.-I., Ishiguro, H., y Hagita, N. (2008). A method for automatic detection of vocal fry. *IEEE Transactions on Audio, Speech, and Language Processing*, *16*(1), 47–56.

Jessen, M. (2008). Forensic Phonetics, *Language and Linguistic Compass*, 2 (4), pp. 671-711.

Jiang, J. (2008). Physiology of voice production: How does the voice work? En M. Benninger y T. Murry (Eds.), *The singer's voice* (pp. 15-24). San Diego, USA: Plural Publishing.

Jiménez Gómez, J. (2011). Estructura formántica y campo de dispersión de las vocales del español en telefonía móvil. *Estudios Fónicos/Cuadernos de Trabajo*, *1*, 39-58.

Jolliffe, I. (1986). Principal component analysis and factor analysis. En *Principal component analysis* (pp. 115-128). New York, NY: Springer.

Keating, P., y Garellek, M. (2015). Acoustic analysis of creaky voice. *Sessão especial sobre voz crepitante no Encontro Anual da Linguistic Society of America em Portland (OR)*.

Keating, P., Garellek, M., y Kreiman, J. (2015). Acoustic properties of different kinds of creaky voice. *Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow*, August, 2015.

Kerstholt, J., Jansen, N., van Amelsvoort, A., y Broeders, A. P. (2006). Earwitnesses: Effects of accent, retention and telephone. *Applied Cognitive Psychology*, *20*(2),

187-197.

Kitzing, P. (1982). Photo-and electroglottographical recording of the laryngeal vibratory pattern during different registers. *Folia Phoniatrica et Logopaedica*, *34*(5), 234–241.

Kneller, W., Memon, A., y Stevenage, S. (2001). Simultaneous and sequential lineups: Decision processes of accurate and inaccurate eyewitnesses. *Applied Cognitive Psychology*, *15*(6), 659-671.

Köster, O., Hess, M., Schiller, N., y Künzel, H. (1998). The correlation between auditory speech sensitivity and speaker recognition ability. *Forensic Lnguistics: The International Journal of Speech, Language and the Law*, *5*, 22-32.

Köster, O., y Schiller, N. (1997). Different influences of the native language of a listener on speaker recognition. *Forensic Linguistics*, *4*, 18-28.

Kreiman, J. (1997). Listening to voices. Theory and practice in voice perception research. En K. Johnson y J. Mullenix (Eds.), *Talker Variability in Speech Processing* (pp. 85-108). San Diego, CA: Academic Press.

Kreiman, J., Gerratt, B. R., y Antonanzas-Barroso, N. (2007). Measures of the glottal source spectrum. *Journal of speech, language, and hearing research*, *50*(3), 595–610.

Kreiman, J., Gerratt, B. R., Garellek, M., Samlan, R., y Zhang (2014). Toward a unified theory of voice production and perception. *Loqu*ens, 1(1) e009

Kreiman, J, y Sidtis, D. (2013 [2011]). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. Malden, MA: Wiley-Blackwell.

Künzel, H. (1989). How well does average fundamental frequency correlate with speaker height and weight? *Phonetica*, *46*((1-3)), 117-125.

Künzel, H. (1990). *Phonetische Untersuchungen zur Sprecher-Erkennung durch linguistisch naive Personen*. Stuttgart: Franz Steiner Verlag.

Künzel, H. (1994). Current Approaches to Forensic Speaker Recognition. *Proceedings of ESCA Workshop on Automatic Speaker Recognition* (pp. 135-141). Martigny (Switzerland).

Künzel, H. (1994b). *Sprecherkennung: Grundzüge forensischer Sprachverarbeitung*. Kriminalistik Verlag. Heidelberg.

Künzel, H. (1997). Some general phonetic and forensic aspects of speaking tempo. *Forensic Linguistics*, *4*(1), 48-83.

Künzel, H. (2000). Effects of voice disguise on speaking fundamental frequency. *Forensic Linguistics*, *7*, 149-179.

Künzel, H. (2001). Beware of the 'telephone effect': The influence of telephone transmission on the measurement of formant frequencies. *Forensic Linguistics*, *8*(1), 80-99.

Künzel, H., González-Rodríguez, J., y Ortega-García, J. (2004). Effect of voice disguise on the performance of a forensic automatic speaker recognition system. En *ODYSSEY 04. The Speaker and Language Recognition Workshop.* (pp. 1-4). Toledo.

Kuwabara, H., y Sagisak, Y. (1995). Acoustic characteristics of speaker individuality. *Speech Communication*, *16*(2), 165-173.

Laver, J. (1968). Voice quality and indexical information. *International Journal of Language y Communication Disorders*, *3*(1), 43-54.

Laver, J. (1976). Language and nonverbal communication. En E. . Carterette y M. . Friedman (Eds.) (Vol. 7, pp. 345-361). New York, NY: Academic Press.

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge: Cambridge University Press.

Laver, J. (1994). *Principles of phonetics*. Cambridge: Cambridge University Press.

Laver, J., Wirz, S., Mackenzie, J., y Hiller, S. (1991). A perceptual protocol for the analysis of vocal profiles. En J. Laver (Ed.), *The Gift of Speech* (pp. 265-280). Edinburgh: Edinburgh University Press.

Leemann, A., Kolly, M.-J., y Dellwo, V. (2014). Speaker-individuality in suprasegmental temporal features: Implications for forensic voice comparison. *Forensic Science International*, *238*, 59-67.

Lindsay, R. C., Lea, J. A., y Fulford, J. A. (1991). Sequential lineup presentation: Technique matters. *Journal of Applied Psychology*, *76*(5), 741.

Lindsay, R. C., y Wells, G. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology*, *70*(3), 556-564.

Lirio, P. (2016). Análisis acústico de la voz creaky deliberada en mujeres españolas.

Estudios de Fonética Experimental, *25*, 193-232.

Mackenzie-Beck, J. (2007). *Vocal Profile Analysis Scheme: A User's Manual*. Edinburgh: Queen Margareth University College-QMUC, Speech Science Research Centre.

Macmillan, N. A., y Creelman, C. D. (1991). *Detection Theory: A User's Guide*. Cambridge: Cambridge University Press.

Malpass, R., y Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology*, *13*(4), 330-334.

Manzanero, A. L. (2010). *Memoria de testigos: obtención y valoración de la prueba testifical*. Madrid: Editorial Pirámide.

Manzanero, A. L. (s. f.). Blog: http://psicologiapercepcion.blogspot.com.es/p/psicofisica-sensorial.html

Manzanero, A. L., Farias-Pajak, K., Igual, C., y Quintana, J. (2011). Exactitud en la identificación de caras y tiempo de respuesta. *Anuario de Psicología Jurídica*, *21*, 107-113.

Marrero, V. (2014). Metodología de investigación en Fonética Perceptiva. En Y Congosto, M. Montero, y A. Salvador Plans (Eds.), *Fonética experimental, educación superior e investigación* (Vol. 1, pp. 503-541). Madrid: Arco/Libros.

Marrero, V., Gil, J., y Battaner, E. (2003). Inter-Speaker Variation in Spanish. An Experimental and Acoustic Preliminary Approach. *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, 2003* (pp. 703-706).

Masthoff, H. (1996). A report on a voice disguise experiment. *Forensic Linguistics*, *3*(1), 160-167.

McGehee, F. (1937). The reliability of the identification of the human voice. *Journal of General Psychology*, *17*, 249-271.

McGuire, G. (2010). A brief primer on experimental designs for speech perception research. Laboratory Report, 77, pp. 1-18.

Monsen, R. B., y Engebretson, A. M. (1977). Study of variations in the male and female glottal wave. *The Journal of the Acoustical Society of America*, *62*(4), 981–993.

Monzo, C. (2010). *Modelado de la cualidad de la voz para la síntesis del habla expresiva* (Tesis de Doctorado), Escola Tecnica Superior d'Enginyeria Electronica i Informatica La Salle-Ramon Llull.

Moore, P., y von Leden, H. (1958). Dynamic variations of the vibratory pattern in the normal larynx. *Folia Phoniatrica et Logopaedica*, *10*(4), 205–238.

Moosmüller, S. (2001). The influence of creaky voice on formant frequency changes. *Forensic Linguistics*, *8*(1), 100-112.

Murakami, H. (2001). *Crónica del pájaro que da cuerda al mundo (edición en eBook)*.

Ní Chasaide, A., y Gobl, C. (1997). Voice source variation. En *The handbook of phonetic sciences* (pp. 427–461).

Nolan, F. (1983). *The Phonetic Bases of Speaker Recognition*. Cambridge: Cambridge University Press.

Nolan, F. (2005). Forensic speaker identification and the phonetic description of voice quality. En W.J. Hardcastle y J. Mackenzie-Beck (Eds.) *A Figure of Speech. A Festschrift for John Laver*. Lawrence Erlbaum Associates. New Jersey, pp. 385–411.

Nolan, F. (2007). Voice quality and forensic speaker identification, *GOVOR* 24 (2), pp. 111-128.

Núñez, F. (2013a). Fisiología de la fonación. En I. Cobeta, F. Núñez, y S. Fernández (Eds.), *Patología de la voz* (pp. 55-75). Barcelona: Marge Medica Books.

Núñez, F. (2013b). Glosario. En I. Cobeta, F. Núñez, y S. Fernández (Eds.), *Patologías de la voz* (pp. 611-615). Barcelona: Marge Medica Books.

Ogden, R. (2001). Turn transition, creak and glottal stop in Finnish talk-in-interaction. *Journal of the International Phonetic Association*, *31*(01), 139–152.

Ogden, R. (2004). Non-modal voice quality and turn-taking in Finnish. En E. Couper-Kuhlen y C. Ford (Eds.), *Sound Patterns in Interaction: Cross-Linguistic Studies from Conversation* (pp. 29–62). Amsterdam: John Benjamins.

Owren, M., Berkowitz, M., y Bachorowski, J. (2007). Listeners judge talker sex more efficiently from male than from female vowels. *Attention, Perception, y Psychophysics*, *69*(6), 930-941.

Palacios Alonso, D. (2018). *Contribución al estudio de selección de parámetros para identificación de estrés en la voz* (Tesis de Doctorado), Universidad Politécnica de Madrid, Madrid.

Pennock-Speck, B. (2005). The changing voice of women. En *Actas XXVIII Congreso*

Internacional AEDEAN, J. J. Calvo García de Leonardo, J. Tronch Pérez, M. del Saz Rubio, C. Manuel Cuenca, B. Pennock Speck, and M. J. Coperías Aguilar (Eds.), pp. 407–414.

Pérez Sanz, C. (2011). Ajustes laríngeos y estilos de fonación en radio y TV (Tesis de Doctorado), Universidad Complutense de Madrid - Instituto Universitario de Investigación Ortega y Gasset, Madrid.

Pérez Sanz, C. (s. f.). *Precisiones sobre los tipos de fonación: problemas conceptuales y terminológicos* (inédito).

Perrot, P., Aversano, G., y Chollet, G. (2007). Voice disguise and automatic detection: review and perspectives. En Y. Stylianou, M. Faundez-Zanuy, y A. Esposito (Eds.), *Progress in nonlinear speech processing* (pp. 101-117). Berlin: Springer.

Perrot, P., y Chollet, G. (2008). The question of disguised voice. *The Journal of the Acoustical Society of America*, *123*(5), 3878.

Perrot, P., Preteux, C., Vasseur, S., y Chollet, G. (2007). Detection and recognition of voice disguis. *Proceedings, IAFPA 2007* (pp. 1-3). The College of St Mark y St John, Plymouth, UK.

Philippon, A., Cherryman, J., Bull, R., y Vrij, A. (2007). Earwitness identification performance: The effect of language, target, deliberate strategies and indirect measures, *Applied Cognitive Psychology, 21,* 539-559.

Podesva, R. J. (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics*, *11*(4), 478–504.

Pollack, I., Pickett, J., y Sumby, W. (1954). On the identification of speakers by voice. *Journal of the Acoustical Society of America*, *26*, 403–406.

Poyatos, F. (1993). *Paralanguage: A linguistic and interdisciplinary approach to interactive speech and sounds* (Vol. 92). Amsterdam: John Benjamins Publishing.

Poyatos, F. (1994). *La comunicación no verbal*. Madrid: Istmo.

Poyatos, F. (2002). *Nonverbal Communication Across Disciplines: Paralanguage, kinesics, silence, personal and environmental interaction*. Amsterdam: John Benjamins.

Praveena, J., y Krishna, Y. (2015). Identifying speaker from disguised speech using aural perception and Mel-frequency cepstral coefficient. *Journal of Indian Speech Language y Hearing Association*, *29*(2), 28-34.

Reales, M., y Ballesteros, S. (1995). *TDS. Un programa de ordenador para la teoría de la detección de señales*. Madrid: Editorial Universitas.

Rodman, R. (1998). Speaker recognition of disguised voices: A program for research. *Proceedings of the Consortium on Speech Technology in Conjunction with the Conference on Speaker Recognition by Man and Machine: Directions for Forensic Applications* (pp. 1-22). Ankara, Turkey: COST250 Publishing Arm.

Romito, L., Lio, R., y Galatà, V. (2005). Fluency articulation and speech rate as new parameters in the speaker recognition. En *Atti del convegno III Congreso de Fonética Experimental* (pp. 26-24). Santiago de Compostela.

Rose, P. (2002). *Forensic Speaker Identification*. London: Taylor y Francis.

Rothenberg, M. (1973). A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. *The Journal of the Acoustical Society of America*, *53*(6), 1632–1645.

Roubeau, B. (1993). *Mécanismes vibratoires laryngés et contrôleneuro-musculaire de la fréquence fondamentale* (Thèse de doctorat). Université Paris-Orsay. Recuperado a partir de http://www.theses.fr/1993PA112273

Roubeau, B., Henrich, N., y Castellengo, M. (2009). Laryngeal vibratory mechanisms: the notion of vocal register revisited. *Journal of voice*, *23*(4), 425-438.

San Segundo, E., Alves, H., y Fernández Trinidad, M. (2013). CIVIL Corpus: Voice quality for speaker forensic comparison. *Selected Papers from the 5th International Conference on Corpus Linguistics (CILC2013), Procedia - Social and Behavioral Sciences*, *95*, 587-593.

San Segundo, E., Foulkes, P., French, P., Harrison, P., y Hughes, V. (2016). Voice quality analysis in forensic voice comparison: developing the vocal profile analysis scheme. En *The 5th Conference of the International Association for Forensic Phonetics and Acoustics, York, July 24–27*.

San Segundo, E., Foulkes, P., y Hughes, V. (2016). Holistic perception of voice quality matters more than L1 when judging speaker similarity in short stimuli. *Proceedings of the 16th Australasian Conference on Speech Science and Technology (ASSTA)* (pp. 1-4). University of Western Sydney, Australia.

San Segundo, E., y Gil Fernández, J. (2014). Voces disimuladas mediante pinzamiento de

nariz: ¿Dificultan la tarea de identificación de un hablante en el ámbito de la fonética judicial? Presentado en *XLII Simposio de la Sociedad Española de Lingüística*, Centro de Ciencias Humanas y Sociales, CSIC, Madrid.

San Segundo, E., y Mompeán, J. (2017). A simplified Vocal Profile Analysis Protocol for the assessment of voice quality and speaker similarity, *31*(5), 644.e11-644.e27.

Sañudo, J. R., Maranillo, E., y León, X. (2013). Anatomía del sistema fonatorio. En I. Cobeta, F. Núñez, y S. Fernández (Eds.), *Patologías de la voz* (pp. 29-46). Barcelona: Marge Medica Books.

Scherer, K. (1974). Acoustic concomitants of emotional dimensions: Judging affect from synthesized tone sequences. En S. Weitz (Ed.), *Non-verbal communication* (pp. 105-111). New York, NY: Oxford University Press.

Scherer, K., Ladd, D., y Silverman, K. (1984). Vocal cues to speaker affect: Testing two models. *The Journal of the Acoustical Society of America*, *76*(5), 1346-1356.

Scherer, K., y Oshinsky, J. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, *1*, 331-346.

Schiller, N., y Köster, O. (1996). Evaluation of a foreign speaker in forensic phonetics: A report. *Forensic Linguistics: The international Journal of Speech, Language and the Law*, *3*, 176-185.

Schiller, N., Köster, O., y Duckworth, M. (1997). The effect of removing linguistic information upon identifying speakers of a foreign language. *Forensic Linguistics. The International Journal of Speech, Language and the Law*, *4*, 1-17.

Schwab, S. (en prensa). La velocidad de elocución: descripción fonética y variación. En Juana Gil Fernández y J. Llisterri Boix (Eds.), *Fonética y Fonología de la Lengua Española*.

Smith, S., Lindsay, R. C., y Pryke, S. (2000). Postdictors of eyewitness errors: Can false identifications be diagnosed? *Journal of Applied Psychology*, *85*, 542–550.

Spence, M., y Freeman, M. (1996). Newborn infants prefer the maternal low-pass filtered voice, but not the maternal whispered voice. *Infant Behavior and Development*, *19*(2), 199-212.

Sporer, S. (1992). Post-dicting eyewitness accuracy: Confidence, decision-times and person descriptions of choosers and non-choosers. *European Journal of Social Psychology*,

22(2), 157-180.

Sporer, S. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology*, *78*, 22-33.

Sporer, S. (1994). Decision times and eyewitness identification accuracy in simultaneous and sequential lineups. En D. Ross, J. Read, y M. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments*. Cambridge: Cambridge University Press.

Stevens, K. N. (1998). *Acoustic Phonetics*. Cambridge, MA: MIT Press.

Sulter, A., y Albers, F. (1996). The effects of frequency and intensity level on glottal closure in normal subjects. *Clinical Otolaryngology y Allied Sciences*, *21*(4), 324–327.

Sundberg, J., Titze, I., y Scherer, R. (1993). Phonatory control in male singing: A study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source. *Journal of Voice*, *7*(1), 15–29.

Švec, J., Schutte, H., y Miller, D. (1999). On pitch jumps between chest and falsetto registers in voice: Data from living and excised human larynges. *The Journal of the Acoustical Society of America*, *106*(3), 1523-1531.

Tanner, W., y Swets, J. (1954). A decision-making theory of visual detection. *Psychological review*, *61*(6).

Tauroza, S., y Allison, D. (1990). Speech rates in British English. *Applied Linguistics*, *11*, 90-105.

Thompson, C. (1987). A language effect in voice identification. *Applied Cognitive Psychology*, *1*(2), 121-131.

Titze, I. R. (1990). Interpretation of the electroglottographic signal. *Journal of voice*, *4*, 1-9.

Titze, I. R. (2000 [1994]). *Principles of voice production* (2ª edición). Iowa, City: National Center for Voice and Speech.

Titze, I. R. (2006). Theoretical analysis of maximum flow declination rate versus maximum area declination rate in phonation. *Journal of Speech, Language, and Hearing Research*, *49*(2), 439-447.

Van den Berg, J. (1968). Mechanism of the larynx and the laryngeal vibrations. En B. Malmberg (Ed.), *Manual of phonetics* (pp. 278–308). Amsterdam: North-Holland.

Van Lancker, D., Kreiman, J., y Emmorey, K. (1985). Familiar voice recognition: patterns and parameters. Part 1: Recognition of backward voices. *Journal of Phonetics*, *13*, 19-38.

Van Lancker, D., Kreiman, J., y Wickens, T. (1985). Familiar voice recognition: parameters and patterns. Part II. Recognition of rate-altered voices. *Journal of Phonetics*, *13*, 39-52.

Vanags, T., Carroll, M., y Perfect, T. (2005). Verbal overshadowing: A sound theory in voice recognition? *Applied Cognitive Psychology*, *19*(9), 1127-1144.

Wagner, I., y Köster, O. (1999). Perceptual recognition of familiar voices using falsetto as a type of voice disguise. *Proceedings of the XIVth International Congress of Phonetic Sciences* (pp. 1381-1385). San Francisco.

Watt, D. (2010). The identification of the individual through speech. En C. Llamas y D. Watt (Eds.), *Language and identities* (pp. 76-85). Edinburgh: Edinburgh University Press.

Weber, N., Brewer, N., Wells, G., Semmler, C., y Keast, A. (2004). Eyewitness identification accuracy and response latency: the unruly 10-12 second rule. *Journal of Experimental Psychology: Applied*, *10*(3), 139-147.

Wells, G., Small, M., Penrod, S., Malpass, R., Fulero, S., y Brimacombe, E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior*, *22*(6), 1-39.

Wendahl, R. W., Moore, P., y Hollien, H. (1963). Comments on vocal fry. *Folia Phoniatrica*, *15*, 251-255.

Whitehead, R. L., Metz, D. E., y Whitehead, B. H. (1984). Vibratory patterns of the vocal folds during pulse register phonation. *The Journal of the Acoustical Society of America*, *75*(4), 1293–1297.

Wolk, L., Abdelli-Beruh, N. B., y Slavin, D. (2011). Habitual use of vocal fry in young adult female speakers. *Journal of Voice*, *26*, 111-116.

Yuasa, I. P. (2010). Creaky Voice: A new feminine voice quality for young urban oriented upwardly mobile American women? *American Speech*, *85*, 315-337.

Zhang, C., y Tan, T. (2008). Voice disguise and automatic speaker recognition. *Forensic Science International*, *175*(2), 118-122.

Anexo

- Los ficheros de datos que se utilizaron para el análisis de la producción y la percepción de la voz estarán disponibles para los miembros del Tribunal a través del siguiente enlace de Google Drive:

  https://drive.google.com/drive/folders/1_dj_6g9LQQpsEzqf7mFwGdxell7IffrH?usp=sharing