

- NIETO-CASTANON, A., F. H. GUENTHER, J. S. PERKELL y H. D. CURTIN (2005): "A modeling investigation of articulatory variability and acoustic stability during American English /r/ production", *JASA*, 117, 3196-3212.
- OBEDIENTE, E. (2007): *Fonética y fonología*, Mérida, Universidad de Los Andes.
- QUILIS, A. (1981): *Fonética acústica de la lengua española*, Madrid, Gredos.
- RECASENS, D. (1991): "On the production characteristics of apico-alveolar taps and trills", *Journal of Phonetics*, 19, 267-280.
- RECASENS, D. y M. D. PALLARÉS (2001): *De la fonética a la fonología*, Barcelona, Ariel.
- RECASENS, D. y A. ESPINOSA (2007): "Phonetic typology and positional allophones for alveolar rhotics in Catalan", *Phonetica*, 63, 1-28.
- ROMÁN MONTES DE OCA, D. (en línea): *Manual de PRAAT* <http://www.domin-go-roman.net/manual_analisis_acustico.html> [consultado por última vez el 16-VI-2011].
- ROMÁN MONTES DE OCA, D. (2011): *Manual para el análisis fonético-acústico*, Santiago de Chile, Pfeiffer editorial.
- ROMERO, J. (1989): "Campos de dispersión auditivos de las vocales del castellano. Percepción de las vocales", *Estudios de Fonética Experimental*, III, 181-206.
- RYALLS, J. (1996): *A basic introduction to speech perception*, San Diego, Singular Publishing Grup Inc.
- SALA, L. y A. M. FERNÁNDEZ PLANAS (1995): "La invariación acústica en las nasales del castellano. Estudio perceptivo", *Estudios de Fonética Experimental*, VII, 161-178.
- SCHUBRIGER, M. (1970): *Introducción a la fonética*, Valladolid, Universidad de Valladolid.
- SOLÉ, M. J. (1999): "The phonetic basis of phonological structure: the role of aerodynamic factors", *Actas del I Congreso de Fonética Experimental*, Tarragona, 77-94.
- STEVENS, K. N. (1972): "Quantal nature of speech", en E. F. David y P. B. Denes (eds.), *Human communication: a unified view*, Nueva York, McGraw Hill, 51-66.
- STEVENS, K. N. (1989): "On the quantal nature of speech", *Journal of Phonetics*, 17, 1/2, 3-45.
- WITGENSTEIN, L. (2001 [1921]): *Tractatus logico-philosophicus*, París, Gallimard.

CAPÍTULO 10

EL DISIMULO DE LA CUALIDAD DE LA VOZ EN FONÉTICA JUDICIAL:

ESTUDIO PERCEPTIVO DE LA HIPONASALIDAD*

JUANA GIL
UNED-CSIC

EUGENIA SAN SEGUNDO
CSIC

1. INTRODUCCIÓN

No es fácil que el ciudadano medio capte de inmediato las repercusiones que en la vida diaria tiene la investigación científica. Por norma general, la labor de los expertos en las distintas disciplinas –salvo quizá en el caso de las ciencias de la salud– le resulta incomprensible y de dudosa aplicación al menos a corto plazo en la resolución de los acuciantes problemas de su cotidianidad. Esto es absolutamente evidente por lo que se refiere a las investigaciones y los estudios relacionados con la lingüística, y los lingüistas lo comprobaban, desgraciadamente, con mucha frecuencia. Sin embargo, la especialidad de la fonética, y por ende de la lingüística, conocida como *fonética judicial* o *fonética forense* constituye una excepción a la norma. Cualquiera de los dos adjetivos, judicial o forense, provoca en el oyente lego en la materia una serie de asociaciones mentales entre conceptos de muy variado tipo que lo lleva a pensar que ese ha de ser necesariamente un ámbito científico interesante y, desde luego, útil para la sociedad. En efecto, lo es.

* Este trabajo se enmarca en el proyecto de investigación *CIVIL* financiado por el Ministerio de Economía y Competitividad (Plan Nacional I+D, Ref. FFI10-21690). Asimismo, Eugenia San Segundo agradece al Ministerio de Educación la concesión de una beca FPU (Programa Nacional de Formación de Profesorado Universitario), con resolución del BOE de 11-07-2009, cuyo disfrute también ha facilitado la realización del presente estudio.

Existen muchas definiciones posibles de fonética forense¹, algunas más concretas o detalladas que otras, pero podrían subsumirse todas ellas en la afirmación de que la fonética judicial es la aplicación de la fonética general que tiene por objeto contribuir a la identificación de un hablante con propósitos legales a partir del análisis de su voz. Este análisis puede realizarse bien para trazar un retrato fonético del locutor (su “pasaporte vocal”), es decir, para establecer gracias a los rasgos acústicos de sus emisiones su sexo, su origen geográfico, su estatus social, etc., o bien puede llevarse a cabo con el objetivo de precisar el contenido real de los mensajes cuando estos han sido grabados y la calidad de los registros no es buena; o bien puede ser preciso para preparar adecuadamente las ruedas de reconocimiento de voces que a veces requieren los procedimientos judiciales, etc. Sobre todo, y muy especialmente, el fonetista judicial debe ocuparse, llegado el caso, de comparar una o varias muestras de habla de un locutor desconocido implicado en un delito con una o varias muestras de habla procedentes de locutor(es) conocido(s), considerados sospechosos(s), a fin de poder afirmar que ambos fragmentos de habla son compatibles con una misma fuente.

Cada una de estas complejas tareas presenta dificultades específicas y solo puede ser realizada con éxito cuando el experto responsable posee una sólida formación –lo cual lamentablemente no siempre se cumple (cf. Hollien, 1990; Eriksson y Lacerda, 2007; Butters, 2009)– y conoce los límites exactos de su función. Sin embargo, aunque estos dos supuestos se den, hay otras circunstancias que pueden condicionar de modo decisivo la labor del especialista. Pensemos, por ejemplo, en la distorsión acústica causada por la transmisión telefónica. Si tenemos en cuenta que una gran cantidad de amenazas, extorsiones, llamadas obsenas, etc., se realizan a través del teléfono, sea este fijo o móvil, es natural que la reducción que este medio inflinge en ciertos parámetros acústicos del habla suponga una dificultad añadida a la tarea del fonetista judicial, que ha de realizar sus análisis a partir de tales muestras distorsionadas (cf., por ejemplo, Guillemín y Watson, 2008).

¹ Cf., por ejemplo, Künnel (2004), Eriksson (2005), Waut (2010), entre otros muchísimos trabajos introductorios mencionables que aparecen recogidos en el apartado correspondiente de la bibliografía temática de Gil (2012).

Con todo, la gran dificultad con la que se enfrentan los expertos se deriva indudablemente del hecho de que el habla está sujeta a *variación*. Las fuentes de tal variación son de muy diverso tipo, aunque no toda ella puede considerarse propiamente *lingüística*: solo lo será la que ocurra de modo sistemático y, además, esté sujeta al control del hablante, esto es, no venga exigida por condicionamientos fisiológicos propios del ser humano (cf. Payne, 2006). Por ejemplo, si pensamos en la coordinación gestual que se precisa para emitir cualquier enunciado, cabe distinguir entre los solapamientos de movimientos en la secuencia fónica que el locutor no puede de ningún modo evitar por razones articulatorias inherentes a su propia condición física, y que alteran la realización de los distintos elementos en función del contexto, y aquellos otros detalles de su implementación que sí entrañan una elección –si bien asumida e interiorizada de modo inconsciente– por parte del emisor. La nasalización de una vocal, habitualmente oral, ante una consonante nasal, por ejemplo, será inevitable (coarticulación *automática* exigida por el entorno fónico), pero puede tener mayor o menor alcance temporal en la vocal dependiendo de la lengua o dialecto de que se trate y que el hablante haya aprendido (variación en la coarticulación de naturaleza propiamente *lingüística*).

Ambos tipos de variabilidad y los mínimos detalles fonéticos sobre los que se sustentan suponen un reto para cualquier lingüista, en particular para cualquier fonetista². Al fonetista forense, en concreto, le interesa y le afecta tanto la una como la otra, y ha de batallar con las dos, puesto que ambos tipos han de tenerse presentes al analizar cualquier muestra de habla. La variación, además, no solo se da entre personas diferentes por razones lingüísticas, sino que otras circunstancias también diversas (sexo, raza, clase social, entorno, estilo de habla, etc.) la producen en mayor o menor medida. Es un hecho comprobado que dos emisiones, incluso aunque hayan sido realizadas por la misma persona, nunca son idénticas, y puede que los cambios sean considerables con respecto a algunas dimensiones o parámetros. En una situación ideal, el fonetista judicial que, por ejemplo, ha de realizar una prueba pe-

² El reto es, sin duda, mucho mayor para los fonólogos. Sobre la cuestión, el lector interesado puede consultar, entre otros, Flemming (1997), Hawkins (2003) o Bradley (2005) y la bibliografía recogida en estos trabajos.

ricial de comparación de voces debería estar en condiciones de afirmar si las posibles diferencias observadas entre las muestras de habla comparadas son debidas con mayor probabilidad a la *variación intra-hablante*, es decir, la que indefectiblemente se da en la voz de un locutor a través de sus distintas emisiones, o responden a la *variación inter-hablante*, esto es, apuntan a la existencia de hablantes distintos.

En buena lógica cabe esperar que, si el locutor cambia, el grado de variación de los enunciados será mayor, al menos con respecto a determinadas variables fónicas, pero no siempre es así. La variación intra-hablante puede ser muy notoria y dificultar mucho la labor del experto, y ello por diversas razones: puede tratarse de una *variación no deliberada* que no dependa de la voluntad del sujeto de modificar conscientemente su voz (como sucede con la variabilidad provocada por la presencia de ruido contaminante, la elevación instintiva de la intensidad de la voz en determinadas circunstancias, el trascurso del tiempo, los cambios en el estado de salud física y psíquica, la intoxicación etílica, el consumo de estimulantes o depresivos, etc.)³, o puede ser una *variación deliberada*, deseada y buscada por el locutor. En el primer caso, el fonetista ha de conocer el tipo de alteración que las diversas condiciones, contextuales o individuales, provocan en los parámetros de estudio seleccionados para tenerlas en cuenta al extraer sus conclusiones. Por ejemplo, se sabe que en general el consumo de alcohol aumenta el valor medio de la f_0 y la duración de las pausas, por mencionar solo dos de los efectos que produce (cf. Baumeister y Schiel, 2010). Del tipo de variación con el que el experto ha de luchar en el segundo caso es del que vamos a ocuparnos en los próximos apartados de este trabajo.

2. VARIACIÓN INTRA-HABLANTE DELIBERADA: EL DISIMULO DE LA VOZ, UN RETO PARA LA FONÉTICA JUDICIAL

Masthoff (1996) ya apuntó que, de acuerdo con la información de que él disponía en el momento en que fecha su trabajo, cuando

³ Foulkes (2006) expone extensamente las principales causas, aquí solo mencionadas, de la variabilidad, y ejemplifica su trascendencia para el análisis lingüístico en general.

un acusado sabe que está siendo grabado o que su voz está siendo sujeta a examen, en un 52% de los casos trata de alterarla en la medida de sus posibilidades. Esa modificación puede ser, básicamente, de dos tipos: el hablante puede querer sonar como otra persona, es decir, adoptar otra personalidad, o el hablante puede simplemente querer oscurecer los rasgos de su propia voz para dificultar su identificación⁴. En el primer supuesto, se habla de *conversión* y es exactamente el proceso que realizan los imitadores, profesionales o aficionados, de los personajes famosos; en el segundo supuesto, estaríamos ante una *transformación*, y es lo que, de manera algo menos técnica, se conoce como *disimulo de la voz*. En el ámbito de la fonética judicial se documentan mucho más frecuentemente transformaciones que conversiones⁵ de voz, aunque estas últimas no sean imposibles.

La transformación de la voz puede conseguirse actualmente por procedimientos electrónicos y por procedimientos no electrónicos o mecánicos (de acuerdo con la muy convincente dicotomía presentada por Rodman, 1998); en ambos casos las consecuencias para el experto y/o el oyente profano que supuestamente ha de identificar la voz son muy negativas. El disimulo, en particular algún tipo de disimulo (v. *infra*), puede reducir en gran medida la capacidad de reconocimiento, y, lo que es más, a veces ni siquiera es correctamente evaluado en sí mismo si no se posee un conocimiento suficientemente amplio de la voz normal del sujeto analizado: según ejemplifica Künzel (2000), una persona puede hablar con una cualidad vocal tan nasal que lleve a un especialista no familiarizado con ese individuo a pensar que está intentando transformar su voz, cuando en realidad se trata del timbre con el que de forma natural produce todos sus enunciados⁶.

⁴ Varios son los estudios pioneros en el análisis de estas modificaciones voluntarias: Endres, Balbach y Flösser (1971), Lummis y Rosenberg (1972) y Hall (1975) son a menudo citados como tales.

⁵ El lector interesado en los procedimientos de conversión de la voz y en sus consecuencias acústico-perceptivas puede consultar sobre todo los múltiples trabajos al respecto de la investigadora Elisabeth Zetterholm, entre ellos su tesis doctoral, de 2003 (vid. cit. 2012).

⁶ Asimismo, la voz de una persona bajo un intenso choque emocional puede resultar irreconocible, sin que en principio esto se entienda como un caso de disimulo *stricto sensu*: "In effect, an extreme emotional state is a form of voice disguise", afirma Rodman (1998: 6).

Los procedimientos electrónicos, aunque cada vez resultan más accesibles para el público en general a medida que el *software* requerido para alterar la voz se comercializa y se distribuye mejor, no son todavía muy recurrentes en los casos criminales que precisan de la intervención de oyentes, expertos o profanos, como “jueces” (en el sentido técnico que se da a este término en fonética perceptiva). Los porcentajes ofrecidos a este respecto por los diversos especialistas varían: desde el 1 % aportado por Masthoff (1996) hasta el 10 % de Gfroerer (1994), pasando por comentarios no cuantitativos como los de Künzel (2000) o Clark y Foulkes (2007)⁷, quienes subrayan la rareza, desde un punto de vista relativo, de este tipo de procedimientos. Hasta cierto punto es lógico que estos mecanismos electrónicos no sean los más frecuentemente empleados en los casos criminales en los que el delincuente recurre a la transformación de su voz. Al fin y al cabo, el disimulo se documenta sobre todo en situaciones de secuestros, extorsiones y amenazas a través del teléfono, y los delincuentes, a no ser que se trate de organizaciones muy bien entrenadas y preparadas, no están a menudo en condiciones de recurrir a productos tecnológicos muy sofisticados.

Otra cuestión bien distinta son los procedimientos mecánicos de alteración de la voz, esto es, los no electrónicos, que son los más habituales. ¿Cuáles son y cómo influyen estos procedimientos en el reconocimiento de una voz?

Como ya se ha dicho al hilo de la exposición, no se ha llevado aún a cabo un gran número de estudios sobre la transformación no electrónica de la voz, pero los que existen abordan básicamente una de estas cuestiones:

2. La investigación de la capacidad de los oyentes o de los reconocedores automáticos para detectar la presencia de disimulo en una muestra de habla.

⁷ Clark y Foulkes (2007), en su trabajo –al que más adelante volveremos– dirigido precisamente a constatar los efectos del disimulo electrónico en la identificación del hablante, concluyen que este tipo de recurso influye en el éxito del reconocimiento, como era previsible, pero no lo impide necesariamente, en particular cuando la modificación introducida afecta a una única variable, en su caso la *f₀*. Este es, además, el parámetro que, junto con el tiempo (i.e. la duración de los sonidos) pero independientemente de él, suele verse alterado con mayor frecuencia por el *software* disponible al efecto.

- b. La investigación de los distintos tipos de disimulo escogidos por los hablantes y las consecuencias fonético-acústicas potenciales que esas transformaciones conllevan.
- c. La investigación de la habilidad de los oyentes para identificar las voces aun en presencia de disimulo.

Dicho de otro modo, se trata de entender si el disimulo es fácilmente detectable, qué tipo de transformaciones se emplean y qué correlatos comportan en la señal y, finalmente, hasta qué punto tales variaciones disminuyen la facultad identificadora de los oyentes. Revisaremos brevemente el progreso realizado en cada uno de estos objetivos.

- a) Por lo que se refiere a la primera línea de trabajo, una conclusión compartida por muchos autores es la de que la capacidad perceptiva humana ofrece a día de hoy un mejor rendimiento que los sistemas automáticos cuando se trata de habla disimulada. Es ya clásico el estudio de Reich (1981) en el que tanto oyentes profanos como oyentes especialistas en fonética fueron capaces de detectar el disimulo en las grabaciones de 40 hablantes que aplicaron diversos tipos de transformación a sus voces. En cambio, como bien han demostrado mucho más recientemente Künzel, González y Ortega (2004) empleando un sistema desarrollado en la Universidad Politécnica de Madrid, para ser eficaces en la tarea de detección y reconocimiento, los métodos automáticos precisan, como referencia, de amplias y variadas bases de datos de habla transformada con el mismo tipo de disimulo⁸ que se evalúe en cada caso; de no disponerse más que de poblaciones de referencia con habla natural normal, sus resultados son mucho más pobres. Conclusiones similares se extraen del estudio de Perrot *et al.* (2007) –en el cual se comprobó que los algoritmos automáticos solo diagnosticaban claramente el disimulo (71 % de acierto) si estaban entrenados con habla también transformada; en caso contrario, el porcentaje bajaba notoriamente (29 % de acierto)– y del estudio de Zhang y Tan (2008), con resultados de reconocimiento también poco satisfactorios.

⁸ En el artículo en cuestión se analiza habla modificada mediante la elevación y el descenso de la *f₀*, y también mediante el pinzamiento de la nariz, procedimiento al que volveremos más adelante.

b) A partir de la bibliografía que se ha ido publicando –por cierto, relativamente no muy abundante⁹– relacionada con la segunda línea de investigación, i.e., los distintos tipos de disimulo y su frecuencia, se puede confeccionar un esquema como el de la figura 1.

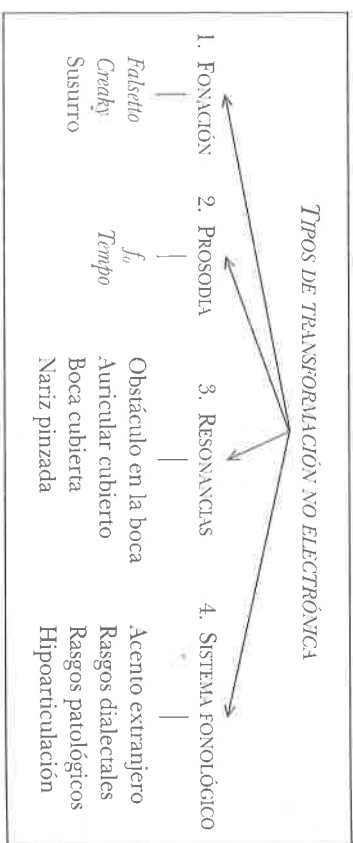


FIGURA 1. Procedimientos mecánicos más comunes empleados para transformar la voz

Según se observa en la figura, los procedimientos de disimulo más frecuentemente utilizados se reparten a *grosso modo* entre los que afectan a la fuente del sonido, esto es, los que están basados en la alteración de la fonación y de algunos rasgos prosódicos a ella ligados, grupos 1 y 2; los que modifican las resonancias orales recurriendo al empleo de cuerpos extraños o de técnicas que interfieren en la transmisión normal del habla, grupo 3, y los que se basan en la realización voluntariamente degradada de los segmentos y/o suprasegmentos de la lengua dada para que se asocien a otro sistema distinto o para que no se perciban con claridad, grupo 4. Interesa señalar que la mayor parte de los delincuentes que recurren a estas tácticas no las combinan (al menos de modo consciente y voluntario), es decir, solo tratan de usar uno de los recursos mencionados (cf. Masthoff, 1996).

Quizá el más estudiado de todos ellos haya sido hasta el momento la alteración deliberada de la f_0 , ya sea aumentándola, ya sea

⁹ Cf. Gil (2012), §16.

disminuyéndola. En ocasiones esa modificación conlleva incluso el cambio de registro de fonación, esto es, respectivamente, el paso a voz en falsete (ing. *falsette voice*) o a voz pulsada (ing. *creaky voice*). El trabajo de Künzel (2000) es indudablemente el estudio que más información ha proporcionado a este respecto, puesto que no solo analiza la influencia de tales cambios sobre la correcta identificación de la voz, sino que aporta datos interesantes sobre el papel desempeñado en esta cuestión por la variable sexo, pero hay otras aproximaciones anteriores y posteriores también relevantes (cf., por ejemplo, Endres, Rambach y Flosser, 1971 y Wagner y Köster, 1999, sobre el falsete; Hinson y Duckworth, 1993 y Moosmüller, 2001, ambas sobre el *creaky* y sus consecuencias para el reconocimiento). De todas ellas se concluye que, aun siendo todos los cambios nocivos para la identificación, la elevación de la f_0 la complica más que su bajada, y que las alteraciones en la frecuencia o los cambios de registro entrañan, concomitantemente, ciertas modificaciones en los formantes vocálicos (Moosmüller, 2001) y en el *tempo* de habla (Künzel, 2000)¹⁰. Por cierto que, con independencia de esta variación de las características temporales de la elocución, conexas con la de la f_0 , parece comprobarse que el *tempo* y el ritmo del habla son en sí mismos rasgos idiosincráticos difícilmente transformables a voluntad del locutor (cf. Dellwo, Ramyeat y Dancovicova, 2009 y Dellwo, Kolly y Leemann, 2012, por ejemplo).

Orchard y Yarmey (1995) y Yarmey *et al.* (2001) abordaron el estudio del otro modo de fonación más a menudo usado (y también poco investigado), el susurro, con resultados semejantes, en el sentido de que se manifestó igualmente como muy perjudicial para la identificación. De hecho, los sistemas automáticos se han demostrado incapaces de contrarrestar esta transformación, con 0% de acierto en el reconocimiento (cf. Zhang, 2005). Mucho más recientemente, Evans y Foulkes (2009) retoman el tema y llevan a

¹⁰ Lógicamente, las pausas abundan más y la velocidad de articulación es menor porque el hablante tiene que realizar un gran esfuerzo fonatorio y articulatorio al adoptar un registro distinto del suyo habitual. Esto nos lleva a mencionar una realidad que se ha constatado repetidas veces en todos los trabajos sobre disimulo, y es que, afortunadamente para los forenses, es muy difícil mantener la voz transformada durante periodos de tiempo prolongados, y es igualmente complicado controlar todos los aspectos que contribuyen a su caracterización: "Human voice is a complex mechanism, which can be imitated by experts. It is, however, extremely difficult to replicate all the different levels of information contained in it" (Faúndez y Monje, 2005: 11).

cabo un análisis muy detallado y exhaustivo en el cual afirman que, si bien es cierto que el habla susurrada reduce las diferencias interhablantes al eliminar la f_0 , aún permite el reconocimiento del locutor, particularmente en el caso de voces familiares y muestras de habla largas, puesto que los efectos resonadores del tracto vocal no se ven alterados y los rasgos peculiares de los hablantes (el acento, por ejemplo), tampoco. El estudio de Evans y Foulkes (2009) ofrece, pues, unos resultados más positivos que los que consiguieron Yárney *et al.* (2001), y concluye aconsejando que, ante casos de habla susurrada, se busquen nuevas "fuentes" de diferenciación interhablante, como la variación en la amplitud del susurro o las divergencias en los ajustes articulatorios y el control general del flujo aéreo espirado.

No es raro, por otra parte, que el hablante que desea evitar ser reconocido opte por intentar hablar con un acento extranjero o dialectal distinto del que le es propio (grupo 4 de la figura 1). Resulta muy interesante que este sea el procedimiento que menos disminuye la eficacia de los sistemas automáticos de reconocimiento (cf. Zhang y Tan, 2008), lo cual es comprensible si se piensa que tales sistemas pretenden ser eficientes con independencia de la lengua y del dialecto. Asimismo, es bastante general en la bibliografía la idea de que este tipo de disimulo es también detectado fácilmente por los oyentes (cf. Tate, 1979; Markham, 1999; Storey, 1996), sobre todo porque el hablante o no va a ser muy distinto ni coherente en imitar los rasgos fonéticos de la L2 durante mucho tiempo, o va a optar por no modificarlos y tratar de variar en su lugar el léxico o las estructuras gramaticales. Esto último se ha argumentado en particular para el caso de la imitación del acento extranjero, más que para el de la imitación de una coloración regional o dialectal (Künzel, 1987, *apud* Neuhauser, 2008).

A este respecto, los múltiples estudios de Sara Neuhauser y/o Adrian Simpson sobre hablantes alemanes imitando acento francés (por ejemplo, Neuhauser, 2008) no son tan optimistas y apuntan, por el contrario, en otra dirección al subrayar la dificultad que implica en ocasiones diferenciar un acento extranjero verdadero de uno imitado. Con todo, Simpson (2007a y b) y Simpson y Neuhauser (2009, 2010) abren una puerta al optimismo al buscar algún detalle fonético que perviva en el disimulo y encontrar que algunos rasgos, como por ejemplo la aparición de clics velares

en la resolución de ciertas oclusivas en alemán o la del típico golpe glotal que precede a las vocales en inicial de palabra en esa lengua, se mantienen de manera coherente durante la imitación por los germanohablantes del acento francés. Cabe suponer, a falta de estudios sobre otros idiomas, que la persistencia del detalle fonético de muy bajo nivel puede ser un indicio aprovechable para la detección del disimulo del acento en cualquier lengua, en línea con lo que explicamos en la introducción acerca de la variabilidad en la realización motivada lingüísticamente.

Otros procedimientos para transformar las características fonéticas de la voz que se han documentado hasta el momento consisten en la inserción de un cuerpo extraño en la boca (Molina de Figueiredo y Souza Brito, 2000; Horga, 2002); el recurso a la hipocarticulación (Moosmüller, 2006); la utilización de algún tipo de cobertura para la boca o el rostro (Rose y Simmons, 1996; Llamas *et al.*, 2008), y el pinzamiento de nariz (vid. *infra*, § 3). Todos ellos necesitan aún de mucho estudio, porque han sido los menos analizados en términos comparativos.

c) En cuanto a la tercera línea de trabajo, esto es, la investigación de la habilidad de los oyentes para identificar las voces aun en presencia de disimulo, se ha comprobado que esta depende de muchos factores que pueden contribuir a que la tasa de identificación correcta sea mayor o menor. El grado de familiaridad con el hablante (Hollien *et al.*, 1982; Amino y Arai, 2008), que lógicamente facilita el reconocimiento; el tamaño de la muestra de que se dispone (Ladefoged y Ladefoged, 1980), que a veces, cuando es suficientemente grande, es la auténtica clave para que el oyente pueda percibir claramente quién es el locutor; la mayor finura o sensibilidad auditiva de unos individuos con respecto a otros (Bull y Clifford, 1984, 1999; Köster *et al.*, 1998), que parece corresponderse con su mejor actuación como jueces perceptivos; el tiempo transcurrido entre la audición de la muestra original y la identificación (McGehee, 1937), cuya relación con el éxito del reconocimiento no está aún clarificada; la formación y especialización en fonética de los oyentes (Köster *et al.*, 2007); o las características peculiares y diferenciadoras de la voz escuchada (Hjortshøj Sørensen, 2009) son algunas de las variables que intervienen en el proceso y condicionan su resultado.

3. EL PINZAMIENTO DE LA NARIZ COMO RECURSO PARA EL DISIMULO DE LA VOZ: ESTUDIO PERCEPTIVO

Según se ha ido comprobando a lo largo de estas páginas, algunos procedimientos para transformar la voz alteran realmente su cualidad global, y otros no. Es decir, un cambio en el *tempo* de elocución, aunque pueda suponer una clave fiable para la identificación, como se explicó arriba, no modifica por sí solo de manera evidente la cualidad de voz del enunciado en cuestión, mientras que, por el contrario, un cambio de registro de fonación afecta de inmediato a su caracterización cualitativa. Por tanto, parecería en buena lógica que un procedimiento de disimulo como el pinzamiento de la nariz hubiera de tener serias consecuencias para el reconocimiento, puesto que modifica la conformación de parte del tracto vocal y con ello la cualidad de la voz. Sin embargo, lo cierto es que los sistemas automáticos superan bien la dificultad que este tipo de disimulo plantea (se ha registrado hasta un 65% de identificación correcta) y los informantes humanos también parecen vencer la dificultad en un alto porcentaje.

Zhang (2005) estudió la caracterización acústica del habla producida con pinzamiento de nariz, y observó que la diferencia en las frecuencias formánticas de este tipo de voz alterada con respecto a las de la voz modal solo es estadísticamente significativa por lo que se refiere al F1 y al F2 de algunas vocales. Los formantes más altos, F3 y F4, se mantienen estables y no presentan divergencias acusadas en uno y otro tipo de voz (pese a que normalmente los formantes superiores se ven afectados cuando hay cambio de registro)¹¹. Por consiguiente, lo que este procedimiento de disimulo tiene de interesante es que en él se produce un desajuste entre la percepción de la cualidad vocal claramente distinta de la modal a que da lugar, por un lado, y la ausencia de correlatos acústicos suficientes que justifiquen esa impresión auditiva. Como señalan Zhang y Tan (2008), los cambios en los parámetros acústicos no justifican en este caso la percepción generada.

Debido, pues, a la naturaleza algo paradójica de este tipo de disimulo, en el presente trabajo se decidió comprobar cuáles eran los efectos que para la identificación perceptiva de un locutor po-

¹¹ Estos resultados coinciden con los aportados por Reich y Duke (1979).

día tener el hecho de que este hable (en nuestro caso, en español) pinzándose la nariz con la mano. Para ello se concibió un experimento cuyo planteamiento y resultados se exponen en las páginas siguientes.

A. OBJETIVOS DEL ESTUDIO Y DISEÑO EXPERIMENTAL

El paradigma experimental adoptado es similar al usado en Clark y Foulkes (2007). Al igual que en dicho estudio, los oyentes que participaron en el experimento como jueces perceptivos tenían que identificar a los hablantes en un test de respuesta forzada (cf. McGuire, 2010). Puesto que, como se ha señalado más arriba, el grado de familiaridad con una voz influye en la habilidad de un oyente para identificarla, con el fin de controlar esa variable ninguno de los oyentes que tomaron parte en el experimento conocía de antemano a los sujetos cuyas voces se utilizaron como estímulos.

En segundo lugar, el estudio pretendía investigar el efecto que puede tener en la memoria un lapso temporal de 24 horas, como factor influyente en la identificación perceptiva de un hablante. Con ese objetivo en mente, se pidió a los oyentes que realizaran el test de identificación de voces en dos ocasiones: la primera, inmediatamente después de familiarizarse con ellas, y la segunda, al día siguiente (aproximadamente 24 horas después de la sesión de entrenamiento destinada a la familiarización).

La diferencia fundamental entre este estudio perceptivo y otros anteriores, como el mencionado de Clark y Foulkes (2007), estriba en que, en este caso, se realizaron dos presentaciones distintas de los estímulos en la fase de familiarización. En la primera, estos aparecen en el orden ABCD, mientras que en la segunda las voces que se escuchan aparecen en el orden BADC. Del total de 28 oyentes que participaron en el estudio, la mitad (14 oyentes, divididos en 7 hombres y 7 mujeres) realizaron la familiarización con los estímulos escuchando la primera presentación (ABCD), mientras que la otra mitad (14 oyentes, divididos en 7 hombres y 7 mujeres) se familiarizaron con ellos a través de la segunda presentación (BADC). Es sabido, a partir de los datos aportados en diversos estudios psicológicos (cf. Ebbinghaus, 1913; Deese y Kaufman, 1957;

Atkinson y Shiffrin, 1968), que no supone la misma dificultad para un sujeto recordar estímulos que se presentaron al principio o al final de una serie que acordarse de aquellos otros presentados en una posición intermedia. Es lo que se conoce como el *efecto de posición serial*, por el cual se recuerdan mejor los primeros elementos de una serie (*efecto de primacia*) y los últimos (*efecto de recencia*), mientras que quedan más expuestos al olvido los elementos que median entre las primeras posiciones y las últimas de una serie. Teniendo en cuenta, por consiguiente, estos conceptos psicológicos de *primacia* y *recencia*, el experimento se diseñó tratando de comprobar si ambos influyen en los resultados de identificación para el test. Es decir, interesaba conocer, con respecto a los oyentes que escucharon la primera presentación (orden de los estímulos ABCD), si estos identificaron mejor al primer hablante (A) y al último (D); y, si, por otro lado, los oyentes que se familiarizaron con las voces a través del segundo tipo de presentación (con orden BADC) reconocieron mejor a los hablantes B y C.

Finalmente, conviene subrayar la diferencia existente entre este estudio y otras investigaciones por lo que se refiere a la presentación de los estímulos en la fase de familiarización. Por ejemplo, Clark y Foulkes (2007) optan por presentar en la pantalla de un ordenador portátil el nombre de los hablantes a la vez que se reproduce la voz de cada uno. Los autores explican que asignaron nombres propios a cada locutor (en lugar de asignarles, por ejemplo, iniciales) con el fin de que a los oyentes les resultara más fácil recordar las voces. Sus criterios para elegir estos nombres propios fueron los siguientes:

- 1) que se tratara de nombres masculinos relativamente frecuentes en inglés, y
- 2) que los cuatro nombres elegidos fueran parecidos en cuanto a número de sílabas y patrón acentual. Los nombres propios que se utilizaron fueron *Edward, Harry, David* y *Matthew*.

Por el contrario, en el presente trabajo los oyentes debían recordar únicamente la voz mientras la escuchaban, es decir, no existía una asociación previa, establecida por los investigadores, entre la voz y un nombre propio de persona, como es el caso en

Clark y Foulkes (2007). No obstante, en la fase de familiarización, realizada mediante una presentación *PowerPoint*, los oyentes veían en la pantalla un número del 1 al 4 a la vez que escuchaban cada una de las cuatro voces, ya que posteriormente, en el test identificativo, debían escuchar una voz al azar e identificarla mediante la asignación del número correspondiente. Esto es lo habitual en los test identificativos (también llamados tipo *labelling*); es decir, debe asociarse a la voz alguna clase de etiqueta. No obstante, no parece conveniente que los investigadores vinculen previamente cada voz con un nombre de persona, puesto que el hecho de que los oyentes tengan que recordar la asociación de ambos elementos implica un grado más de dificultad en la tarea de identificación de voces, como bien apuntan Kreiman y Sidtis (2011: 176): "Name retrieval adds an extra dimension of complexity to voice recognition tasks above and beyond merely recognizing that the voice is familiar and knowing whom it belongs to".

B. HIPÓTESIS DE ESTUDIO A PARTIR DE LA REVISIÓN BIBLIOGRÁFICA LLEVADA A CABO

Las cuatro preguntas fundamentales que se quería responder a través del diseño experimental arriba descrito son las siguientes:

1. ¿Los resultados obtenidos en el test de identificación son mejores cuando los oyentes escuchan las voces de los informantes hablando sin taparse la nariz (condición normal) que cuando escuchan a estos mismos locutores hablando con la nariz pinzada?

Ya se explicó en el apartado anterior que todos los estudios realizados sobre la cuestión apuntan a que suele existir mayor dificultad para identificar una voz cuando esta presenta algún tipo de enmascaramiento. Por consiguiente, la hipótesis inicial de la que se partió fue la de que los oyentes obtendrían peores resultados en la identificación de hablantes cuando estos disimulan.

2. ¿Existen diferencias entre los resultados del test identificativo llevado a cabo justo después de la familiarización con las voces (memoria operativa) y los resultados del test identificativo llevado a cabo un día después de la familiarización?

Mencionamos anteriormente que la influencia del tiempo sobre la capacidad de identificación de una voz era una variable que había de tenerse en cuenta. La cuestión suele reformularse en términos forenses de la siguiente manera: ¿cómo influye en el reconocimiento de voces el desfase temporal entre el momento en que se comete un crimen y el momento en que tiene lugar la identificación? No está claro, al parecer, que la capacidad de identificación de una voz disminuya cuanto más tiempo haya transcurrido desde que se escuchó. Kreiman y Sidtis (2011: 257) avanzan algunas explicaciones para este hecho aparentemente desconcertante, que se resumen en la idea de que, puesto que son muchos los factores que influyen en los resultados de un test identificativo, podría ser que el factor temporal no tuviera tanto peso como nuestra intuición nos lleva a pensar (Kreiman y Sidtis, 2011: 257). En otras palabras, es posible que influya, pero no sabemos en qué medida.

3a. ¿Son algunos hablantes más difíciles de identificar que otros? Sistemáticamente, ¿se confunden más a menudo las voces de ciertos hablantes entre sí?

3b. ¿Varían los resultados identificativos en función de la posición que ocupen los hablantes (interna: posición 2 o 3; externa: posición 1 y 4) en la presentación de los estímulos durante la fase de familiarización?

Los resultados de Clark y Foulkes (2007) indican que los hablantes con la f_0 más diferente de la de los demás (es decir, el de la voz más grave y el de la voz más aguda) son aquellos a los que los oyentes confunden con menos frecuencia. En el presente caso, los locutores tienen edades muy próximas y una f_0 muy parecida, además de hablar el mismo dialecto del español. Pese a ello, podría ser que los oyentes confundieran menos a los hablantes con una f_0 media más diferente, más peculiar, al igual que en el trabajo de Clark y Foulkes (2007).

Por otro lado, también se ha pretendido comprobar, a este respecto, si los conceptos de *primacia* y *recencia* de los que antes habíamos influyen en los resultados obtenidos en el test identificativo. Por eso, se compararon los resultados obtenidos por los dos grupos de oyentes: los que se familiarizaron con los estímulos mediante la primera presentación (orden de los estímulos: ABCD) y

los que lo hicieron a través de la segunda presentación (con el siguiente orden de los estímulos: BADC).

4. ¿Obtienen mejores resultados identificativos los hombres que las mujeres?

En respuesta a esta última pregunta se documentan resultados muy diversos desde el trabajo pionero de McGehee (1937), que revelaba que los hombres son mejores que las mujeres en tareas de reconocimiento de voces no familiares. Otros autores, como Thompson (1985), concluyen, por el contrario, que no existen diferencias significativas entre ambos sexos. Teniendo en cuenta los resultados obtenidos en estudios anteriores (cf. San Segundo, 2012), se decidió partir de la hipótesis inicial de que no surgirían diferencias significativas entre la actuación de los hombres y las mujeres.

C. MÉTODO

a) Oyentes

En este estudio participó un total de 28 oyentes, de los cuales 14 eran hombres y 14 mujeres. Todos eran hablantes nativos de español como lengua materna, sin patologías del habla ni problemas de audición conocidos. En el momento en que se llevó a cabo el experimento, los oyentes residían en Madrid desde hacía más de tres meses: 9 de ellos habían nacido en Madrid, 16 procedían de distintos puntos de España y 3 de Hispanoamérica. La edad media de los oyentes era de 35,7 años, con un rango de edad comprendido entre los 25 y los 60 años. Como explicamos anteriormente, de los 28 oyentes, la mitad (7 hombres y 7 mujeres) se familiarizó con las voces a través del *PowerPoint 1*, con el orden de estímulos ABCD y la otra mitad (7 hombres y 7 mujeres) se familiarizó con las voces a través del *PowerPoint 2*, con el orden de estímulos BADC.

b) Estímulos

Los estímulos utilizados para la fase de familiarización y para los posteriores tests de identificación se extrajeron de las grabaciones de voz de cuatro locutores masculinos, residentes en Madrid, hablantes monolingües de la variedad de español centropeninsular y con edades comprendidas entre los 33 y los 36 años. Los denominaremos de ahora en adelante A, B, C y D.

Según los experimentos pilotos llevados a cabo en algunos estudios (por ejemplo, en el citado de Clark y Foulkes, 2007), el número óptimo de hablantes que se deben escuchar en un test perceptivo es cuatro, ya que los oyentes consideran que la tarea perceptiva resulta excesivamente difícil cuando el número de voces que deben recordar supera esa cifra.

HABLANTE	EDAD	F0
A	36	96 Hz
B	34	102 Hz
C	36	97 Hz
D	33	120 Hz

TABLA 1. *Edad y frecuencia fundamental media de los cuatro hablantes cuyas voces se utilizaron como estímulos perceptivos*

La tarea que tenían que realizar estos hablantes durante la grabación consistía en la lectura de un texto fonéticamente equilibrado (De Bruynickx *et al.*, 1994) en dos modos. Primero leyeron el texto en voz alta y de la forma en que lo harían normalmente, y a continuación lo leyeron pinzándose la nariz con la mano. Durante esta segunda modalidad de lectura, se pidió a los informantes que no realizaran ninguna modificación adicional con la voz (es decir, a nivel glótico, como por ejemplo, alterar la f_0), puesto que el objetivo era que la única variación que se produjera entre una lectura y otra se debiera al efecto de taparse la nariz.

Los informantes fueron grabados con un micrófono de diadema tipo condensador, omnidireccional y de respuesta plana (*Countryman E6i Omni Earset*). La grabación, además, se llevó a cabo utilizando una tarjeta de sonido externa (*Cakewalk UA-25EXCW*) y el programa de grabación *Adobe Audition* (versión CS 5.5), con la siguiente configuración: grabación en un solo canal (mono), frecuencia de muestreo de 22.100 Hz y resolución de 16 bits.

Por un lado, para la creación de los estímulos con los que debían familiarizarse los oyentes, se seleccionó el texto entero que habían leído los informantes sin pinzarse la nariz. Por otro lado, se extrajo solamente un fragmento de la grabación para constituir

los estímulos que escucharían los oyentes en los dos tests de identificación. En el primero de ellos se utilizó la parte central de la grabación con la lectura normal de cada hablante, y en el segundo test se hizo uso de la última parte de la grabación en la que los informantes leen con la nariz tapada (véase la figura 2).

La duración de los estímulos utilizados para la familiarización es de aproximadamente 32 segundos, aunque es preciso subrayar que los oyentes pudieron escuchar dos veces cada estímulo. Los empleados para los tests de identificación, tanto en el primero (voz normal), como en el segundo (nariz pinzada), duran aproximadamente 10 segundos.

TEXTO FONÉTICAMENTE EQUILIBRADO LEÍDO POR LOS INFORMANTES
El joyero Federico Vanero ha sido condenado por la Audiencia de Santander a ocho meses de arresto mayor y cincuenta mil pesetas de multa por un delito de compra de objetos robados. <i>La vista oral se celebró el miércoles pasado</i> ; durante ella, uno de los fiscales, Carlos Valáñez, <i>pidió para el joyero tres años de prisión menor y una multa de cincuenta mil pesetas. Gracias a las revelaciones de Vanero de hace dos años y medio se llegó a descubrir la existencia de una sospechosa mafia policial en España, parte de la cual se vio envuelta en el llamado "caso El Nani".</i>

FIGURA 2. *Texto fonéticamente equilibrado que leyeron los informantes*

[Se resalta en cursiva la parte de la grabación utilizada como estímulo para el test identificativo con voz normal. Aparece subrayado el fragmento usado para constituir los estímulos del test identificativo con la nariz tapada]

Se eligió la lectura de un texto, en lugar de emplear habla espontánea, con el fin de obtener datos controlados, de forma que todos los hablantes produjeran exactamente los mismos contenidos lingüísticos. Al neutralizar esta variable, siguiendo el ejemplo de otros estudios (Clark y Foulkes, 2007, por ejemplo, utilizan la lectura del cuento de *Cenicienta*), se quiso evitar que los resultados del test identificativo pudieran atribuirse al hecho de que los oyentes asociaran ciertas palabras a ciertos hablantes, en lugar de centrarse en las características de sus voces a la hora de identificarlos.

Los cuatro locutores seleccionados para formar los estímulos de este experimento perceptivo forman parte de un corpus más amplio (San Segundo, 2013). Se eligió estas voces, y no otras, porque no presentaban características especialmente peculiares en el habla y en la voz, además de, como ya se ha explicado, por pertenecer al mismo dialecto y grupo de edad, y por no diferir notablemente en la frecuencia fundamental.

Los archivos de audio originales se editaron para eliminar cualquier sonido no lingüístico, como risa o tos, así como para cortar los errores o falsos comienzos producidos por el hablante durante la lectura del texto.

c) Ejecución del experimento

Según se ha descrito anteriormente, el experimento consistía en una fase de familiarización (día 1) y en una fase de identificación, dividida a su vez en dos tipos de tests: uno en el que los oyentes escuchaban los estímulos en voz normal, y otro en el que se presentaban estímulos en los que el hablante leía con la nariz pinzada. Esta fase de identificación, que es el test perceptivo propiamente dicho, se llevaba a cabo en dos días: el que hemos llamado *día 1* (es decir, el mismo día que la familiarización) y el *día 2* (pasadas 24 horas del primer día), en el que los oyentes realizaban exactamente el mismo test que el día anterior. Por supuesto, el orden de aparición de los estímulos en cada prueba era distinto, ya que estos se reproducían de manera aleatoria.

Tanto en la fase de familiarización como en la de identificación, los oyentes escuchaban los estímulos a través de una presentación de *Microsoft PowerPoint* en un ordenador portátil (*Sony VAIO PCG-G1M / Sony VAIO VPCEA1S1E*) y mediante auriculares *AKG K-240 Studio*. El experimento se llevó a cabo en una habitación sin ruido, en la que no había presente ninguna otra persona aparte del oyente, con el fin de que este no tuviera ninguna distracción.

1. Fase de familiarización

La fase de familiarización consistió en la escucha por parte de los oyentes de las voces de los hablantes A, B, C y D, incluidas en una presentación de *PowerPoint*, tras las diapositivas que incluían las instrucciones para llevar a cabo dicha familiarización (véase la figura 3).

INSTRUCCIONES PARA LLEVAR A CABO LA FAMILIARIZACIÓN CON LAS VOCES	
DIAPOSITIVA 1:	
- A continuación vas a escuchar 4 voces que corresponden a 4 hombres diferentes.	
- Todos ellos son madrileños, con edades comprendidas entre los 33 y los 36 años.	
DIAPOSITIVA 2:	
- En concreto, lo que vas a escuchar es a estos 4 hombres leyendo el texto "El joyero Federico Vanero".	
- Lee tú mismo este texto para familiarizarte con él.	
DIAPOSITIVA 3:	
- El joyero Federico Vanero ha sido condenado por la Audiencia de Santander a ocho meses de arresto mayor y cincuenta mil pesetas de multa por un delito de compra de objetos robados. La vista oral se celebró el miércoles pasado y, durante ella, uno de los fiscales, Carlos Valcárcel, pidió para el joyero tres años de prisión menor y una multa de cincuenta mil pesetas. Gracias a las revelaciones de Vanero de hace dos años y medio se llegó a descubrir la existencia de una sospechosa mafia policial en España, parte de la cual se vio envuelta en el llamado "caso El Nani".	
DIAPOSITIVA 4:	
- A continuación te pediremos que escuches cada una de las 4 voces y te centres en recordarlas.	
- Fíjate en los rasgos característicos de cada locutor. Lo importante aquí es QUIÉN habla, no QUÉ dice. Más adelante te pediremos que identifiques las voces que vas a oír.	
DIAPOSITIVA 5:	
- Te recordamos que podrás escuchar las voces 2 veces.	
DIAPOSITIVA 6:	
- PROCEDIMIENTO	
1º) Escucharás la voz número 1, después la voz número 2, a continuación la voz número 3 y finalmente la voz número 4.	
2º) Cuando hayas terminado la primera escucha, habrá una pausa de 10 segundos y escucharás las 4 voces de nuevo, en el mismo orden.	
DIAPOSITIVA 7:	
- A continuación se escucharán las voces.	
- En cuanto avance a la siguiente diapositiva, se reproducirá la voz nº 1.	
- Para que se reproduzcan la voz nº 2 y las siguientes, simplemente avanza de diapositiva.	

FIGURA 3. Instrucciones incluidas en la presentación de estímulos para la fase de familiarización

A continuación de las instrucciones, se reproducían las voces de A, B, C y D. Cada una en una diapositiva distinta, y, al tiempo que las escuchaban, los oyentes podían leer un número en la pantalla (del 1 al 4). Después de la primera audición, tras una pausa de 10 segundos, se reproducían por segunda vez las mismas voces en el mismo orden. Recordemos que 14 hablantes se familiarizaron con el *PowerPoint 1* y otros 14 con el *PowerPoint 2*. Las instrucciones en ambos casos fueron las mismas. Lo único que cambió fue el orden de aparición de las voces de los hablantes: orden ABCD en *PowerPoint 1* y BADC en *PowerPoint 2*.

No se ofreció la posibilidad a los oyentes de tomar notas durante la fase de familiarización. Solamente uno de ellos sugirió la posibilidad de anotar las características de los hablantes, pero no se le permitió con el fin de que todos los oyentes llevaran a cabo el experimento en las mismas condiciones. Asimismo, se decidió incluir en el *PowerPoint* el texto que los jueces iban a escuchar, porque de ese modo su atención no se desviaría hacia el contenido de la grabación reproducida, sino que se centraría exclusivamente en los rasgos fónicos individuales de cada hablante. Es decir, la información proporcionada no resultaría nueva para ellos en el momento de la escucha, lo que facilitaría la tarea de retención de las voces y no del mensaje.

2. Fase de identificación

La fase de identificación del experimento tuvo lugar, como ya adelantamos más arriba, en dos días distintos. Además, los estímulos del test de identificación incluían, en ambas ocasiones, dos condiciones diferentes: lectura normal y lectura con la nariz pinzada. Hablaremos, por tanto, de dos tests de identificación, que se presentaron a los oyentes uno a continuación del otro. Tanto el día 1 como el día 2 los oyentes realizaron primero el test 1 (lectura normal) y seguidamente el test 2 (lectura con la nariz tapada).

Para llevar a cabo los tests de identificación se utilizó un "script" de *Prat* (Boersma y Weenick, 2012) que reproduce los estímulos de uno en uno a la vez que muestra una pantalla con cuatro cuadros en los que se pueden leer los números del 1 al 4. La tarea del oyente, por tanto, consiste en marcar la opción que

cree que corresponde a cada estímulo. En otras palabras, al oyente se le pide que asocie cada voz con un hablante. En este tipo de test, el informante no tiene la opción de volver a escuchar los estímulos. El tiempo de respuesta de cada participante es libre y no se ha computado a efectos de analizarlo posteriormente.

Los estímulos en cada tipo de test son cuatro: los archivos de audio correspondientes a cada uno de los hablantes A, B, C y D. El "script" de *Prat* utilizado permitió presentar tres repeticiones de cada estímulo, que se reproducen de manera aleatoria sin que un mismo estímulo aparezca dos veces seguidas, gracias a la estrategia de aleatorización de *Prat* denominada *permute balanced no doubles*.

D. RESULTADOS

1. Patrón de errores - matrices de confusión

En primer lugar, se crearon algunas matrices de confusión, que nos informan con cierto detalle de las diferencias en las respuestas de los oyentes con respecto a los cuatro hablantes. Cada columna de la matriz representa el número de predicciones (el hablante que los oyentes dijeron oír), mientras que cada fila representa los casos reales (el hablante real que escuchaban los oyentes). Las identificaciones correctas aparecen en negrita. Así, por ejemplo, la tabla 2 muestra que para los estímulos del hablante A (el primer día y en la condición normal) hubo 53 identificaciones correctas, 13 respuestas atribuidas incorrectamente al hablante B, 11 al hablante C y 7 al hablante D. La última fila y la última columna de cada tabla indican el número total de errores cometidos.

Si construimos cuatro matrices de confusión (una por cada día y condición de estudio) que muestren las identificaciones correctas y los hablantes más frecuentemente confundidos, podemos observar que los hablantes mejor identificados son siempre D (tablas 2 y 3; condición normal) o C (tablas 4 y 5; nariz pinzada) y que los hablantes que más frecuentemente confunden los oyentes son el A con el B, y viceversa, en todas las condiciones y días.

Hablañte ↓	Respuesta				Total errores
	A	B	C	D	
A	53	13	11	7	31
B	13	61	4	6	23
C	5	8	67	4	17
D	7	3	3	71	13
Total errores	25	24	18	17	

TABLE 2. Matriz de confusión, condición normal, día 1, todos los oyentes.
En negrita se marcan las identificaciones correctas

Hablañte ↓	Respuesta				Total errores
	A	B	C	D	
A	56	12	11	5	28
B	16	54	5	9	30
C	13	5	63	3	21
D	4	9	6	65	19
Total errores	33	26	22	17	

TABLE 3. Matriz de confusión, condición normal, día 2, todos los oyentes.
En negrita se marcan las identificaciones correctas

Hablañte ↓	Respuesta				Total errores
	A	B	C	D	
A	46	24	6	8	38
B	15	54	3	12	30
C	11	5	65	3	19
D	8	9	10	57	27
Total errores	34	38	19	23	

TABLE 4. Matriz de confusión, condición matriz tapada, día 1, todos los oyentes.
En negrita se marcan las identificaciones correctas

Hablañte ↓	Respuesta				Total errores
	A	B	C	D	
A	47	20	10	7	37
B	14	61	0	9	23
C	10	2	71	1	13
D	5	7	4	68	16
Total errores	29	29	14	17	

TABLE 5. Matriz de confusión, condición matriz tapada, día 2, todos los oyentes.
En negrita se marcan las identificaciones correctas

Para conocer si los resultados varían en función del grupo de oyentes, es decir, si existen diferencias entre que estos realicen la familiarización con las voces a través del *PowerPoint 1* (tablas 6-9) o que la efectúen mediante el *PowerPoint 2* (tablas 10-13), realizamos ocho matrices de confusión más, marcando una división en función del grupo de oyentes. A la vista de dichas matrices, se observa que sí, se manifiestan algunas diferencias entre los dos conjuntos.

Por un lado, el grupo 1 (que se familiarizó con las voces mediante el *PowerPoint 1*) identifica mejor al hablante C y en segundo lugar al hablante D, en todos los casos (combinación de condición experimental y día en que tuvo lugar el test) salvo en un caso (condición normal y día 1), que identifica mejor al D y en segundo lugar al C. Las confusiones entre hablantes ocurren mayoritariamente con los hablantes A y B. Los resultados, por tanto, son muy parecidos a los que obteníamos si comparábamos al conjunto entero de oyentes (tablas 2-5).

Por otro lado, si se analizan las matrices de confusión para el segundo grupo de oyentes (familiarizados con los estímulos a través del *PowerPoint 2*), se comprueba que existe alguna pequeña variación con respecto al primer grupo. En este caso, los oyentes identifican mejor al hablante D, en primer lugar, y seguidamente al hablante B. No obstante, la diferencia entre las identificaciones correctas con el hablante B y C no son muy grandes: una diferencia de tan solo un acierto más en tres de los casos (tablas 10, 11 y 12) y de tres aciertos en la tabla 13. Además, continúan produ-

ciéndose bastantes confusiones entre el hablante A y B, siguiendo el patrón que ya detectamos al contemplar las primeras matrices de confusión (en las que agrupamos a los dos grupos de oyentes) y las matrices correspondientes al primer grupo de hablantes.

Las conclusiones que se desprenderían de este primer análisis de las matrices de confusión es que el orden de aparición de un estímulo en el proceso de familiarización llevado a cabo por los oyentes no influye excesivamente en la posterior identificación de las voces que estos realizan. La única excepción la supondría el oyente B, cuya probabilidad de identificación aumenta al pasar de la posición 2 en el primer *PowerPoint* a la posición 1 en el se-

Hablañte ↓	Respuesta				Total errores
	A	B	C	D	
A	27	7	4	4	15
B	9	27	4	2	15
C	1	5	34	2	8
D	1	3	2	36	6
Total errores	11	15	10	8	

TABLA 6. *Matriz de confusión, condición normal, día 1, grupo de oyentes PowerPoint 1 (ABCD). En negrita se marcan las identificaciones correctas*

Hablañte ↓	Respuesta				Total errores
	A	B	C	D	
A	29	5	5	3	13
B	13	23	2	4	19
C	4	2	33	3	9
D	2	6	5	29	13
Total errores	19	13	12	10	

TABLA 7. *Matriz de confusión, condición normal, día 2, grupo de oyentes PowerPoint 1 (ABCD). En negrita se marcan las identificaciones correctas*

Hablañte ↓	Respuesta				Total errores
	A	B	C	D	
A	27	10	1	4	15
B	9	25	2	6	17
C	3	3	35	1	7
D	0	8	5	29	13
Total errores	12	21	8	11	

TABLA 8. *Matriz de confusión, condición matriz tapada, día 1, oyentes PowerPoint 1. En negrita se marcan las identificaciones correctas*

Hablañte ↓	Respuesta				Total errores
	A	B	C	D	
A	19	15	4	4	23
B	10	26	0	6	16
C	1	2	39	0	3
D	4	7	1	30	12
Total errores	15	24	5	10	

TABLA 9. *Matriz de confusión, condición matriz tapada, día 2, oyentes PowerPoint 1. En negrita se marcan las identificaciones correctas*

Hablañte ↓	Respuesta				Total errores
	A	B	C	D	
A	26	6	7	3	16
B	4	34	0	4	8
C	4	3	33	2	9
D	6	0	1	35	7
Total errores	14	9	8	9	

TABLA 10. *Matriz de confusión, condición normal, día 1, oyentes PowerPoint 2. En negrita se marcan las identificaciones correctas*

Hablañte ↓	Respuesta				Total errores
	A	B	C	D	
A	27	7	6	2	15
B	3	31	3	5	11
C	9	3	30	0	12
D	2	3	1	36	6
Total errores	14	13	10	7	

TABLA 11. *Matriz de confusión, condición normal, día 2, oyentes PowerPoint 2. En negrita se marcan las identificaciones correctas*

Hablañte ↓	Respuesta				Total errores
	A	B	C	D	
A	19	14	5	4	23
B	6	29	1	6	13
C	8	2	30	2	12
D	8	1	5	28	14
Total errores	22	17	11	12	

TABLA 12. *Matriz de confusión, condición nariz tapada, día 1, oyentes PowerPoint 2. En negrita se marcan las identificaciones correctas*

Hablañte ↓	Respuesta				Total errores
	A	B	C	D	
A	28	5	6	3	14
B	4	35	0	3	7
C	9	0	32	1	10
D	1	0	3	38	4
Total errores	14	5	9	7	

TABLA 13. *Matriz de confusión, condición nariz tapada, día 2, oyentes PowerPoint 2. En negrita se marcan las identificaciones correctas*

gundo *PowerPoint*. En todo caso, se hace necesario llevar a cabo un análisis estadístico para comprobar si estas diferencias, observadas en las matrices de confusión, son realmente significativas (véase el siguiente epígrafe).

2. Análisis estadístico

Tanto para averiguar si los resultados obtenidos tras realizar las matrices de confusión (véase el apartado anterior) son significativos, como para contestar a las preguntas planteadas como hipótesis de trabajo, se llevaron a cabo diversos tests estadísticos con el programa R (versión 2.15).

PREGUNTA 1: *¿Los resultados obtenidos en el test de identificación son mejores cuando los oyentes escuchan las voces de los informantes hablando sin taparse la nariz (condición normal) que cuando escuchan a estos mismos hablando con la nariz tapada?*

Para conocer la respuesta a esta pregunta, se realizó una prueba exacta de McNemar, equivalente a la prueba *t de student* pero recomendable cuando se tienen dos muestras (en nuestro caso: re-

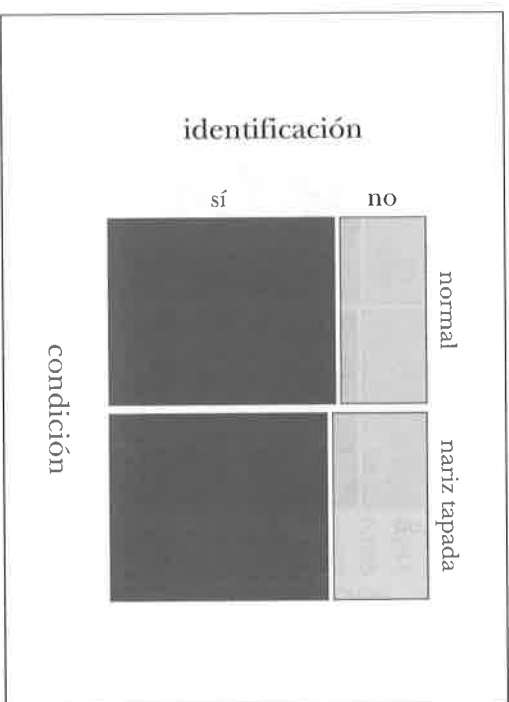


FIGURA 4. *Comparación de identificaciones correctas y erróneas en función de la condición de lectura (normal y con la nariz tapada)*

sulados de identificación para la condición "normal" y para la condición "nariz tapada") y además estas son dependientes (el grupo de oyentes que llevó a cabo los tests de identificación es el mismo en ambas condiciones). Los resultados de este test estadístico muestran que no existen diferencias significativas ($p = 0,085$) en el número de identificaciones correctas si comparamos la condición normal y la condición "nariz tapada". No obstante, el porcentaje de identificaciones correctas aumenta en la condición normal (73,06%) frente a la condición "nariz tapada" (69,64%).

PREGUNTA 2: *¿Existen diferencias entre los resultados del test identificativo llevado a cabo justo después de la familiarización con las voces (memoria operativa) y los resultados del test identificativo llevado a cabo un día después de la familiarización?*

Una prueba exacta de McNemar (con dos colas) nos permite comprobar que no existen diferencias significativas ($p = 0,64$) entre los resultados del test obtenidos por los oyentes el primer día y los resultados que obtuvieron estos un día después. Es destaca-

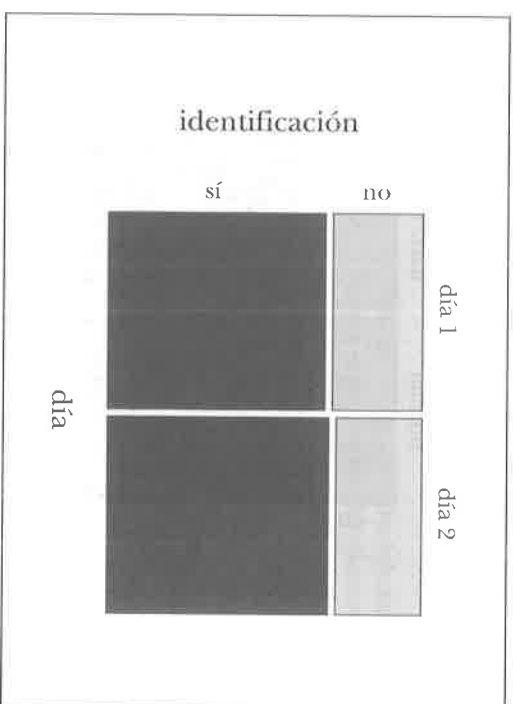


FIGURA 5. Comparación de identificaciones correctas y erróneas en función del día en que tuvo lugar la identificación: día 1 (mismo día que la familiarización) o día 2 (pasadas 24 horas desde la familiarización)

ble el hecho de que el porcentaje de identificaciones correctas aumenta el segundo día, aunque la diferencia es insignificante: 70,68% de identificaciones el primer día y 72,02% de identificaciones el segundo día.

PREGUNTA 3a: *¿Son algunos hablantes más difíciles de identificar que otros? Sistemáticamente, ¿se confunden más las voces de ciertos hablantes entre sí?*

Los resultados de una prueba *Q* de Cochran (para la comparación de proporciones en tres o más muestras relacionadas) con comparaciones *post hoc* por pares¹² indican que existen diferencias significativas o muy significativas en el número de identificaciones correctas si comparamos la mayoría de los hablantes de dos en dos (véase la tabla 14). Tan solo la comparación de las identificaciones correctas para los hablantes C y D muestra que no existen diferencias significativas. En el resto de los casos, existen diferencias significativas (*) o muy significativas (**)¹³.

HABLANTES COMPARADOS	SIGNIFICACIÓN ESTADÍSTICA
A frente a B*	$p = 0,025$
A frente a C**	$p = 1,32E-07$
A frente a D**	$p = 1,24E-06$
B frente a C*	$p = 0,011$
B frente a D*	$p = 0,025$
C frente a D	$p = 0,627$

TABLA 14. Comparación de las identificaciones correctas obtenidas por cada par de hablantes

En la figura 6 observamos que, según indica el test estadístico, las diferencias entre el número de identificaciones para el hablante C (o número 3) y para el hablante D (o número 4) son

¹² Las comparaciones *post hoc* por pares se llevaron a cabo mediante múltiples pruebas de McNemar (con corrección de Holm).

¹³ Se marcan con un asterisco los coeficientes de correlación significativos al nivel 0,05 y con dos asteriscos los significativos al nivel 0,01.

mínimas, como también pudimos comprobar en las matrices de confusión. En definitiva, ambos hablantes son los que mejor identifican los oyentes, y ambos en la misma medida. Los porcentajes de identificaciones correctas y erróneas para cada hablante se muestran en la tabla 15.

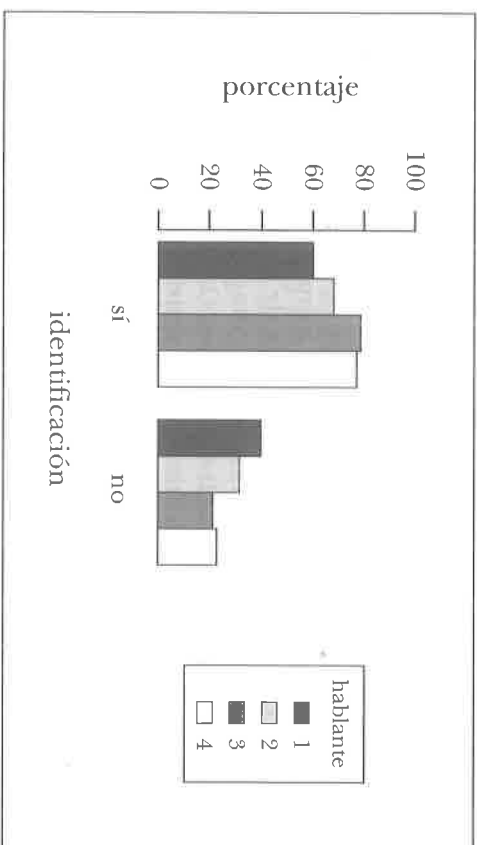


FIGURA 6. Identificaciones correctas y erróneas en función del hablante

	IDENTIFICACIONES CORRECTAS (%)	IDENTIFICACIONES ERRÓNEAS (%)
HABLANTE A	60,42	39,58
HABLANTE B	68,75	31,25
HABLANTE C	78,87	21,13
HABLANTE D	77,08	22,92

TABLA 15. Porcentaje de identificaciones correctas y erróneas en función del hablante

Finalmente, se realizó una prueba *chi cuadrado* para saber con qué otro locutor había sido confundido más frecuentemente cada hablante, y con cuánta significación estadística. Lo más relevante

de los resultados es que el hablante A fue confundido con más frecuencia con el B, y viceversa. Esta información ya la conocíamos gracias a las matrices de confusión; ahora además supimos que esto ocurre de forma muy significativa en ambos casos: $p = 4,64E-05$ para las ocasiones en que el hablante A fue identificado incorrectamente como el hablante B, y $p = 5,12E-07$ para las ocasiones en que el hablante B fue incorrectamente identificado como el hablante A. Es preciso señalar que el locutor C también se confundió sobre todo con el hablante A ($p = 7,99E-05$). Por último, el hablante D fue confundido con los locutores A, B y C a partes iguales, es decir, sin diferencias significativas ($p = 0,77$). Se confirma, por tanto, que este hablante presenta más identificaciones correctas que confusiones con otros locutores.

PREGUNTA 3b: ¿Varían los resultados identificativos en función de la posición que ocupen los hablantes (interna: posición 2 ó 3; externa: posición 1 y 4) en la presentación de los estímulos durante la fase de familiarización?

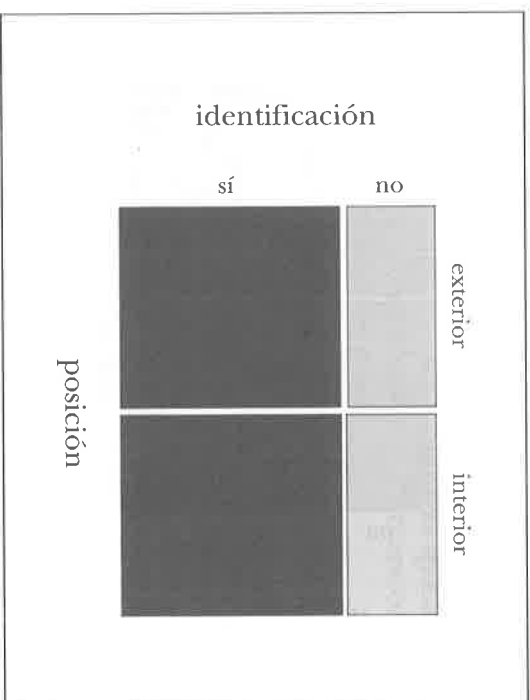


FIGURA 7. Comparación de identificaciones correctas y erróneas en función de la posición que ocupa el estímulo en la presentación utilizada para la familiarización

Tras realizar una prueba exacta de McNemar, obtenemos que las diferencias entre los dos grupos de resultados no son significativas, tanto en la prueba con dos colas (*two sided testing*; $p = 0,94$) como en la prueba con una cola (*one sided testing*; $p = 0,47$). Además, como se observa en la figura 7 (en la página anterior), al comparar las posiciones interna y externa de los estímulos no varía el número de identificaciones correctas o incorrectas obtenidas por los oyentes: cuando estos se familiarizan con las voces de los hablantes situados en una posición interior de la serie obtienen un 71,29% de aciertos, mientras que, si escuchan a los hablantes que se colocan en una posición exterior de la serie, obtienen un 71,42% de identificaciones correctas.

PREGUNTA 4: ¿Consiguen mejores resultados identificativos los hombres que las mujeres?

Los resultados de una prueba exacta de McNemar muestran que no existen diferencias significativas entre los obtenidos por

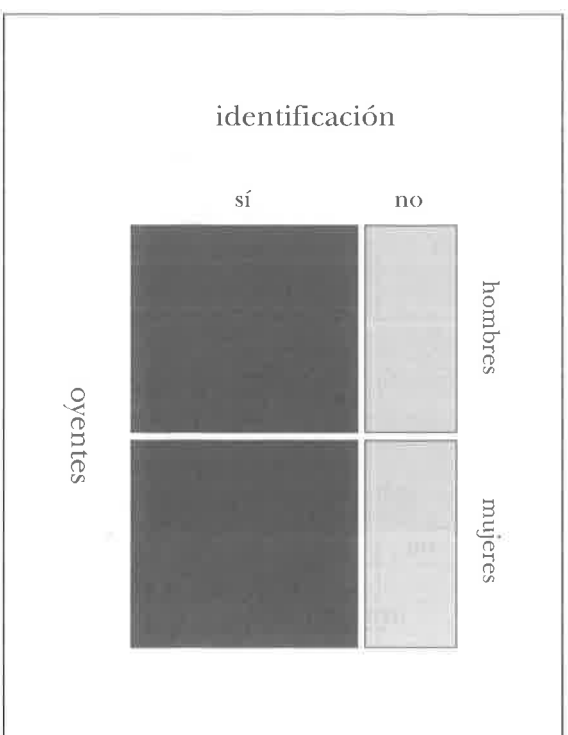


FIGURA 8. Comparación de identificaciones correctas y erróneas en función del sexo de los oyentes

los hombres y los obtenidos por las mujeres ($p = 0,95$). Los obtenidos por los hombres (71,57% de aciertos) son solo ligeramente mejores que los de las mujeres (71,13% de aciertos), como muestra el gráfico 8.

E. ANÁLISIS DE LOS RESULTADOS

• *Enmascaramiento de la voz mediante hiponasalidad*

A partir de la revisión bibliográfica llevada a cabo en los apartados iniciales del presente trabajo quedó suficientemente claro que existe mayor dificultad para identificar una voz cuando esta presenta algún tipo de enmascaramiento, así como que el pinzamiento de nariz resulta algo peculiar con respecto a otros recursos para el disimulo. Por tanto, el objetivo principal del experimento realizado fue comprobar si la hiponasalidad de la voz de un hablante, cuando este se tapa la nariz al hablar, es realmente una forma de enmascaramiento eficaz. Para ello, se compararon los resultados obtenidos por 28 oyentes en dos tests de identificación, con la única diferencia entre ambos de que en uno de los tests los hablantes leían un texto sin taparse la nariz (condición normal) mientras que en el otro los estímulos se crearon a partir de la lectura con la nariz tapada por los mismos locutores.

Los resultados del test estadístico llevado a cabo muestran que no existen diferencias significativas entre los resultados de los distintos tests, lo que corrobora la idea de que los métodos de enmascaramiento de la voz que afectan al tracto vocal, y muy en particular el pinzado de nariz, quizá no son tan eficaces como los que tienen que ver con cambios en la fonación. En cualquier caso, el enmascaramiento sin duda dificulta la identificación: en el test con la condición "nariz tapada", el número de identificaciones correctas es menor que en el test con la condición "normal". Ciertamente, el valor obtenido en el test estadístico ($p = 0,085$) no permite concluir que las divergencias sean significativas, si bien no es un valor que esté extremadamente lejos de serlo.

Una posible causa de que la diferencia entre las dos condiciones no sea más acentuada es que el test con estímulos en la condición "nariz tapada" tiene lugar siempre a continuación del test identificativo con estímulos en la condición normal. Por tanto, podría ocurrir que del primer test al segundo tenga lugar un proceso de aprendizaje. Para neutralizar esa posible influencia, en trabajos futuros se podría diseñar el experimento de forma que algunos oyentes realizaran un tipo de test (condición normal) mientras que otro grupo de oyentes, con características similares al primero, realizara el otro tipo de test (condición de hiponasalidad).

Por otro lado, no debe olvidarse que en este estudio se pidió a los informantes que, con el fin de camuflar su identidad, simplemente debían taparse la nariz al leer el texto. A pesar de las indicaciones de los investigadores en sentido contrario, algunos hablantes pretendían añadir algún cambio que afectaba a la fonación, convencidos, según explicaron, de que taparse la nariz no enmascaraba su voz suficientemente.

- *Lapso temporal entre la familiarización con los estímulos y la identificación de los hablantes*

Una interpretación simplista de la influencia del paso del tiempo sobre la memoria nos induciría a creer que se obtendrían peores resultados en los tests de identificación realizados un día después de la familiarización con las voces que en los tests llevados a cabo por los oyentes inmediatamente después de realizar la familiarización. Sin embargo, en la bibliografía específica sobre estudios forenses se encuentran resultados dispares, según ya se comentó. El hecho de que en este estudio no existan diferencias significativas entre los resultados obtenidos el primer día y el segundo podría achacarse de nuevo a un proceso de aprendizaje por parte de los oyentes, que favorecería que se obtuvieran mejores resultados identificativos el segundo día. Por otro lado, estos resultados también podrían deberse a la existencia de distintos tipos de memoria, y la primacía de un tipo u otro en los oyentes que participaron en el es-

tudio. Es decir, podría ocurrir que algunos oyentes identificaran mejor a los hablantes el primer día, debido a que tenían mejor memoria operativa, mientras que la identificación de otros locutores podría haberse visto favorecida por el lapso temporal entre la familiarización y la identificación. Estaríamos, por consiguiente, ante un fenómeno de compensación por el cual los resultados obtenidos por un tipo de hablantes contrarrestan los resultados obtenidos por otros. En un futuro, sería interesante comprobar el peso de un lapso más prolongado entre la familiarización y la identificación.

- *Efecto de la posición serial de los estímulos auditivos en la fase de familiarización*

Como se señaló en el §3.A, es ampliamente conocido en psicología que no supone la misma dificultad recordar estímulos que se presentaron al principio (efecto de primacía) o al final (efecto de recencia) de una serie que acordarse de aquellos otros estímulos presentados en una posición intermedia: estos últimos quedan más expuestos al olvido.

Pues bien, con el presente estudio se pretendió comprobar si estos conceptos realmente influyen en la identificación de hablantes. Los resultados del test estadístico demuestran que la posición en que se encuentre el estímulo durante la familiarización no afecta a los resultados identificativos. Puede ocurrir que los efectos de primacía y de recencia no tengan tanta influencia en los estímulos auditivos como en los visuales (por ejemplo, en la memorización de palabras tras leer una lista). Por otro lado, pudiera interpretarse que quizá cuatro estímulos son demasiado pocos para verse afectados por estos fenómenos. Sin embargo, tras la realización de las matrices de confusión, observamos que el hablante B es identificado en más ocasiones cuando pasa de ocupar la segunda posición (presentación *PowerPoint* número uno: orden ABCD) a ocupar la primera (presentación *PowerPoint* número dos: orden BADC). El análisis estadístico que llevamos a cabo para conocer si el efecto de la posición serial influye en las identificaciones toma en cuenta la comparación de las posicio-

nes internas (2 y 3) con las externas (1 y 4), pero el experimento no está diseñado para conocer si una posición concreta favorece la identificación. Por eso, tan solo podemos concluir que no existen diferencias significativas entre las posiciones internas y externas. En un estudio futuro, interesaría considerar cada una de las cuatro posiciones por separado, analizando cómo influye en los resultados el hecho de que una voz aparezca en una o en otra.

- *Hablantes más fáciles o más difíciles de identificar*

Los resultados, tanto de las matrices de confusión como del test estadístico, muestran que los hablantes que más confunden los oyentes entre sí son el A y el B. Esto les ocurre tanto a los oyentes que utilizaron el *PowerPoint* 1 para la familiarización con los estímulos como a los oyentes que se familiarizaron con las voces a través del segundo *PowerPoint*. Además, esto ocurre en ambas direcciones (véase la prueba *chi cuadrado* del apartado anterior). Es decir, cuando el hablante que se escucha es A, los oyentes dicen que es B, pero también cuando oyen el estímulo correspondiente a B, este es identificado con A.

Es difícil establecer la causa de la mejor identificación de los locutores C y D por los oyentes en este estudio, y además no era este uno de los objetivos de la investigación. En otros trabajos (Clark y Foulkes, 2007) se observa que los hablantes con la f_0 más diferente a la de los demás (es decir, el hablante con la voz más grave y el hablante con la voz más aguda) son aquellos a los que los oyentes confunden con menos frecuencia. En nuestro caso, D tiene claramente la f_0 más alta (120 Hz) pero C (97 Hz) tiene prácticamente la misma f_0 que A (96 Hz). Por tanto, este parámetro no ha debido ser, o no solamente, la pista acústica en la que se han fijado los oyentes para identificar a los hablantes. Por otro lado, la diferencia entre la f_0 más alta y la más baja en el estudio de Clark y Foulkes (2007) era de 35 Hz, mientras que en el presente estudio es de tan solo 14 Hz, por lo que parece bastante plausible que los oyentes tuvieran que atender a otras claves acústicas. En un futuro se podría plantear la recogida de un cuestionario con las intuiciones de los oyentes

a la hora de realizar las identificaciones. En el presente trabajo se preguntó a todos ellos qué rasgos habían considerado característicos de los hablantes para poder identificarlos. Aunque sus respuestas no se han tenido en cuenta formalmente, muchos de los oyentes destacaron haberse fijado en el ritmo de habla y en la entonación de los informantes, lo que puede ofrecer claves acerca de qué aspectos son idiosincráticos de una persona o permiten identificarla aun cuando se intente controlar el resto de variables, como la edad, el dialecto y la f_0 (recuérdese lo arriba dicho sobre el poder identificativo de las variables temporales del habla).

- *Identificaciones correctas en función del sexo del oyente*

Finalmente, como era esperable, ya que así lo señalaban la mayoría de los estudios que preceden a este en el ámbito de la fonética judicial, no existen diferencias significativas entre los resultados obtenidos por los hombres y por las mujeres que participaron como jueces perceptivos en este estudio.

4. CONCLUSIONES

Algunos especialistas ponen en tela de juicio la capacidad de la fonética judicial, en su estado actual de desarrollo, para contribuir a la identificación de hablantes (cf. Gold, 2011; Bonastre *et al.*, 2003). Uno de los argumentos en los que se basan tiene que ver directamente con la variabilidad intrínseca y extrínseca del habla, que no solo condiciona el reconocimiento de los locutores por parte de los oyentes tanto legos como expertos –por ejemplo en las ruedas constituidas al efecto– sino que también reduce el grado de comparabilidad de las muestras con las que se trabaja habitualmente en este campo. Cualquier alteración, desviación o distorsión de la voz normal de una persona complica extraordinariamente la tarea de los oyentes que han de identificarla.

Ahora bien, el hecho de que exista variabilidad en el habla no significa que la identificación sea imposible. En el caso concreto de la alteración deliberada de la voz mediante el disimulo, que es el que se ha tratado en el presente estudio, el repaso a la bibliografía demuestra fehacientemente que solo algunas modalidades de

transformación, en particular las que suponen un cambio en el registro de fonación, conducen inexorablemente al fracaso en la identificación, y aun en estos casos la capacidad perceptiva de los seres humanos ofrece un mejor rendimiento que los algoritmos de los sistemas de reconocimiento automático. En los casos restantes, cuando el disimulo afecta a las resonancias del tracto vocal más que a la fuente, o cuando se basa en la modificación de las características segmentales y/o suprasegmentales del sistema fonético-fonológico para fingir un acento idiomático ajeno, las consecuencias para la identificación no parecen ser tan graves y los porcentajes de reconocimiento acertado aumentan considerablemente.

El pinzamiento de nariz, es decir, la creación en la cavidad nasal de una caja de resonancia sin salida al exterior con la consiguiente hiponasalidad del habla, es un caso algo atípico dentro de los procedimientos mecánicos de disimulo de la voz, porque, por lo visto hasta el momento en la bibliografía especializada, resulta ser uno de los recursos menos eficaces para disfrazar al hablante, y, sin embargo, afecta claramente su cualidad de voz general. Las conclusiones que se han obtenido en el estudio aquí presentado inciden en esta misma idea y la corroboran, puesto que no se han producido divergencias significativas entre los tests de reconocimiento sobre voz normal y sobre voz hiponasal. También se han visto refrendadas otras conclusiones de estudios anteriores, como las referidas al grado semejante de agudeza perceptiva de hombres y mujeres.

Más allá de los resultados específicos del estudio perceptivo realizado, lo que resulta más interesante sin duda es que, de nuevo en este caso, se apunta la posibilidad de que algunos de los procedimientos de transformación de la voz aparentemente más exitosos puedan dejar un resquicio, unas claves robustas resistentes al disimulo que permitan el reconocimiento del locutor, en particular si las muestras de habla disponibles son lo suficientemente largas. Ciertas investigaciones sobre algunos otros recursos (por ejemplo, el empleo del susurro o la adopción de un falso acento extranjero) así lo demuestran también, según se ha explicado en las páginas anteriores. Por consiguiente, solo cabe esperar, con optimismo, que en los próximos años la fonética judicial sea capaz de descubrir más pistas reveladoras de la identidad que contribuyan a contrarrestar la alteración deliberada de la voz con propósitos criminales.

AGRADECIMIENTOS

Las autoras quieren destacar la colaboración en este artículo de Almut Braun, que realiza su tesis doctoral sobre fonética judicial en la Universidad de Marburgo, a quien agradecen encarecidamente la enorme ayuda prestada en el análisis estadístico de los datos.

REFERENCIAS BIBLIOGRÁFICAS

- AMINO, K. y T. ARAI (2008): "Effects of linguistic contents on perceptual speaker identification: comparison of familiar and unknown speaker identifications", *Acoustical Science and Technology*, 30 (2), 89-99.
- ATKINSON, R. C. y R. M. SHIFFRIN (1968): "Human memory: A proposed system and its control processes", en K. W. Spence y J. T. Spence (eds.), *The Psychology of Learning and Motivation*, vol. 2, Nueva York, Academic Press, págs. 89-195.
- BAUMEISTER, B. y F. SCHIEL (2010): "On the effect of alcoholisation on fundamental frequency", en *IATPA-2010*, Trier.
- BOERSMA, P. y D. WEENINK (2012): *Practical doing phonetics by computer*, Versión 5.3.03, URL: <<http://www.praat.org>> [21 de abril de 2012].
- BONASTRE, J.-F., F. BINBOT, L.-J. BOË, J.-P. CAMPBELL, D. A. REYNOLDS e I. MAGRIN-CHAGNOILLEAU (2003): "Person authentication by voice: A need for caution", en *EUROSPEECH 2003*, Ginebra.
- BRADLEY, T. (2005): "Systematic markedness and phonetic detail in phonology", en R. Gess y E. Rubin (eds.), *Experimental and Theoretical Approaches to Romance Linguistics*, Amsterdam, John Benjamins, págs. 41-62.
- BULL, R. y B. CLIFFORD (1984): "Earwitness voice recognition accuracy", en G. Wells y E. Loftus (eds.), *EyeWitness Testimony Psychological Perspectives*, Londres, Cambridge University Press, págs. 92-123.
- BULL, R. y B. R. CLIFFORD (1999): "Earwitness testimony", *Medicine, Science and the Law*, 39 (2), 120-127.
- BUTTERS, R. R. (2009): "The forensic linguist's professional credentials", *International Journal of Speech, Language and the Law*, 16 (2), 237-252.
- CLARK, J. y P. FOULKES (2007): "Identification of voices in electronically disguised speech", *International Journal of Speech, Language and the Law*, 14 (2), 195-211.
- CLIFFORD, B. (1983): "Memory for voice: The feasibility and quality of earwitness evidence", en S. Lloyd-Bostock y B. Clifford (eds.), *Evaluating Witness Evidence. Recent Psychological Research and New Perspectives*, Nueva York, John Wiley and Sons, págs. 189-218.

- DE BRUNICK, M., B. HARMEGNIEN, J. LISTERI y D. POCH (1994): "Language-induced voice quality variability in bilinguals", *Journal of Phonetics*, 22, 19-31.
- DEESE, J. y R. A. KAUFMAN (1957): "Serial effects in recall of unorganized and sequentially organized verbal material", *Journal of Experimental Psychology*, 54 (3), 180-187.
- DELLWO, V., S. RAMBEAD y J. DANKOVICOVA (2009): "The influence of voice disguise on temporal characteristics of speech" en *IAPPA-2009*, Cambridge.
- EBBINGHAUS, H. (1913): *Memory: A Contribution to Experimental Psychology*, Nueva York, Columbia University.
- ENDRES, W., W. BAMBACH y G. FLOSSER (1971): "Voice spectrograms as a function of age, voice disguise and voice imitation", *The Journal of the Acoustical Society of America*, 49, 1842-1848.
- ERIKSSON, A. (2005): "Tutorial on forensic speech science. Part I: Forensic phonetics", en *INTERSPEECH-2005. Eurospeech. Proceedings of the 9th European Conference on Speech Communication and Technology*, Lisboa.
- ERIKSSON, A. (2010): "The disguised voice: Imitating accents or speech styles and impersonating individuals", en C. Llamas y D. Wàrt (eds.), *Language and Identities*, Edinburgo, Edinburgo University Press, págs. 86-96.
- ERIKSSON, A. y F. LACERDA (2007): "Charlatany in forensic speech science: A problem to be taken seriously", *International Journal of Speech, Language and the Law*, 14, 169-193.
- FATÚNDEZ, M. y S. MONTES (2005): "State-of-the-art in speaker recognition", *IEEE Aerospace and Electronic Systems Magazine*, 20 (5), 7-12.
- FLEMING, E. (1997): "Phonetic detail in phonology: Evidence from assimilation and coarticulation", en K. Suzuki y D. Elzinga (eds.), *Southwest Workshop on Optimality Theory: Features in OT*, Los Angeles, Coyote Papers, págs. 1-12.
- FOULKES, P. (2006): "Phonological variation: A global perspective", en B. Aarts y B. McMahon (eds.), *The Handbook of English Linguistics*, Oxford, Blackwell, págs. 625-659.
- GIL, J. (2012): *Bibliografía temática de fonética judicial* <http://www.estudiosfonicos.cchs.csic.es/fonetica/civil/GilJ_2012_Bibliografia_Tematica_de_Fonetica_Judicial.pdf> [05 de junio de 2012].
- GROERER, S. (1994): *Häufigkeit und Art Forensischer Stimmverstellungen*, Ms. Bundeskriminalamt, Alemania.
- GOLD, E. (2011): "Forensic speaker comparison evidence: The international picture", en *IAPPA-2011*, Viena.
- GUILLEMIN, B. y C. I. WATSON (2008): "Impact of the mobile phone network on the speech signal - Some preliminary findings", *International Journal of Speech Language and the Law*, 15 (2), 193-218.

- HALL, M. (1975): *Spectrographic Analysis of Interspeaker and Intraspeaker Variability of Professional Mimicry*, Tesis Doctoral, Michigan State University.
- HAWKINS, S. (2003): "Roles and representations of systematic fine phonetic detail in speech understanding", *Journal of Phonetics*, 31, 373-405.
- HIRSON, A. y M. DUCKWORTH (1993): "Glottal fry and voice disguise: A case study in forensic phonetics", *Journal of Biomedical Engineering*, 15, 193-208.
- HJORTSHØJ SØRENSEN, M. (2009): "Voice line-ups: Some voices are easier to remember than others - and some people are better at it", en *IAPPA-2009*, Cambridge.
- HOLLEN, H. (1990): *The Acoustics of Crime. The New Science of Forensic Phonetics*, Nueva York, Plenum Publishing Corporation.
- HOLLEN, H. y W. MAJEWSKI (1977): "Speaker identification by long-term spectra under normal and distorted speech conditions", *Journal of the Acoustical Society of America*, 62 (4), 975-980.
- HOLLEN, H., W. MAJEWSKI y E. T. DOHERTY (1982): "Perceptual identification of voices under normal, stress and disguise speaking conditions", *Journal of Phonetics*, 10, 139-148.
- HORGÁ, D. (2002a): "The influence of bite-blocks on continuous speech production", en A. Braun y H. R. Masthoff (eds.), *Phonetics and its applications*, Stuttgart, Steiner, págs. 143-152.
- KÖSTER, O., M. M. HESS, N. O. SCHILLER y H. J. KÜNZEL (1998): "The correlation between auditory speech sensitivity and speaker recognition ability", *Forensic Linguistics*, 5 (2), 22-32.
- KÖSTER, O., M. JESSEN, F. KHARI y H. ECKERT (2007): "Auditory-perceptual identification of voice quality by expert and non-expert listeners", en *XVIIth International Congress of Phonetic Sciences*, Saarbrücken.
- KREIMAN, J. y D. SUDTIS (2011): *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception*, Londres, John Wiley and Sons.
- KÜNZEL, H. J. (1987): *Speecherkennung. Grundzüge forensischer Sprachverarbeitung*, Heidelberg, Kriminalistik Verlag.
- KÜNZEL, H. J. (2000): "Effects of voice disguise on speaking fundamental frequency", *Forensic Linguistics*, 7 (2), 149-179.
- KÜNZEL, H., J. GONZÁLEZ y J. ORTEGA (2004): "Effect of voice disguise on the performance of a forensic automatic speaker recognition system", en *ODYS-2004*, 153-156.
- LADFOGED, J. y P. LADFOGED (1980): "The ability of listeners to identify voices", *UCLA Working Papers in Phonetics*, 49, 43-1.
- LLAMAS, D., P. HARRISON, D. DONNELLY y D. WÀRT (2008): "Effects of different types of face coverings on speech acoustics and intelligibility", *York Papers in Linguistics*, 9, 80-104.

- LUMMIS, R. C. y A. ROSENBERG. (1972): "Test of an automatic speaker verification with intensively trained mimics", *Journal of the Acoustical Society of America*, 51, 131(A).
- MARKHAM, D. (1999): "Listeners and disguised voices: The imitation and perception of dialectal accent", *International Journal of Speech, Language and the Law*, 6 (2), 289-299.
- MASTHOFF, H. R. (1996): "A report on a voice disguise experiment", *Forensic Linguistics*, 3 (1), 160-167.
- McGENEE, F. (1937): "The reliability of the identification of the human voice", *Journal of General Psychology*, 17, 249-271.
- MCGILONE, R., H. HOLLIEN y P. HOLLIEN (1977): "Acoustic analysis of voice disguise related to voice identification", en *Proceedings of the International Conference on Crime Countermeasures*, Kentucky, University of Kentucky, págs. 31-35.
- MCCUTRE, G. (2010). *A Brief Primer on Experimental Designs for Speech Perception Research* (manuscrito). URL: <http://people.usc.edu/~gmcguit1/experiment_designs.pdf>.
- MOLINA DE FIGUEROA, R. (2000): "Algunas considerações sobre o distorcer de voz", *Estudos Linguísticos*, XXIX, 543-548.
- MOLINA DE FIGUEROA, R. y H. DE SOUZA BRITO (1996): "A report on the acoustic effects of one type of disguise", *Forensic Linguistics*, 3 (1), 168-175.
- MOOSMÜLLER, S. (2001): "The influence of creaky voice on formant frequency changes", *International Journal of Speech, Language and the Law*, 8 (1), 100-112.
- NEUHAUSER, S. (2008): "Voice disguise using a foreign accent: Phonetic and linguistic variation", *International Journal of Speech, Language and the Law*, 15 (2), 131-159.
- NEUHAUSER, S. y A. P. SIMPSON (2007): "Initiated or authentic? Listeners' judgements of foreign accents", en *Proceedings of the XVIIth International Congress of the Phonetic Sciences*, Saarbrücken, Saarbrücken Universität, 185-188.
- NEUHAUSER, S. y SIMPSON, A. P. (2008): "Realisation of German final syllable /ən/ as a cue to accent authenticity for French accents in German", *Comunicación presentada en IALPA-2008*, Lausana.
- PAVNE, E. (2006): "Phonetic motifs and the formation of sound structure", *UCL Working Papers in Linguistics*, 18, 321-343.
- PERROT, P. y G. CHOLLET (2012): "Helping the Forensic Research Institute of the French Gendarmerie to identify a suspect in the presence of voice disguise or voice forgery", en A. Neustein y H. A. Paill (eds.), *Forensic Speaker Recognition. Law Enforcement and Counter-Terrorism*, Nueva York, Springer, págs. 469-503.

- PERROT, P., G. AVERSANO y G. CHOLLET (2007): "Voice disguise and automatic detection: Review and perspectives", en Y. Stylianou, M. Faundez y A. Espósito (eds.), *Progress in Non-Linear Speech Processing. Lecture Notes in Computer Science*, Berlin, Springer Verlag, págs. 101-117.
- PERROT, P., G. PRETEUX, S. VASSEUR y G. CHOLLET (2007): "Detection and recognition of voice disguise", comunicación presentada en el Congreso de la IALPA 2007, Plymouth.
- R DEVELOPMENT CORE TEAM (2012): *R: A Language and Environment for Statistical Computing*. Versión 2.15. R Foundation for Statistical Computing, Viena. URL: <http://www.R-project.org> [21 de abril de 2012].
- REICH, A. (1975): *Certain Effect of Selected Vocal Disguises Upon Spectrographic Speaker Identification*, tesis doctoral, University of Iowa.
- REICH, A. (1981): "Detecting the presence of vocal disguise in the male voice", *Journal of the Acoustical Society of America*, 69 (5), 1458-1460.
- REICH, A. y J. DUKE (1979): "Effects of selected vocal disguises upon speaker identification by listening", *Journal of the Acoustical Society of America*, 66, 1023-1028.
- REICH, A., K. MOLL y J. CURTIS (1976): "Effects of selected vocal disguises upon spectrographic speaker identification", *Journal of the Acoustical Society of America*, 60, 919-925.
- RODMAN, R. (1998): "Speaker recognition of disguised voices", en M. Demirekler, A. Saranti, H. Altinçay y A. Paoloni (eds.), *Proceedings of the Consortium on Speech Technology Conference on Speaker Recognition by Man and Machine: Directions for Forensic Applications*, Ankara, COST 250 Publishing Arm, págs. 9-22.
- RODMAN, R. (2003): *Speaker Recognition of Disguised Voices: A Program for Research*, Raleigh, Carolina del Norte, North Carolina State University.
- ROSE, P. y A. SIMMONS (1996): "F-pattern variability in disguise and over the telephone - comparisons for forensic speaker identification", *Proceedings of the 6th Australian International Conference on Speech Science and Technology*, Adelaide.
- SAN SEGUNDO, E. (2012): "El entrenamiento musical y otros factores que pueden influir en el reconocimiento perceptivo de hablantes", en *V Congreso de Fonética Experimental*, Cáceres, 25-28 de octubre de 2011.
- SAN SEGUNDO, E. (2013): *La comparación forense de voces en gemelos monoigóticos y dicigóticos en español centropeninsular* (tesis doctoral, manuscrito).
- SIMPSON, A. P. (2007a): "Phonetic details of nonpulsed stop release in German: inter- and individual variation", comunicación presentada en *IALPA-2007*, Plymouth (UK).
- SIMPSON, A. P. (2007b): "Acoustic and auditory correlates of non-pulsed sound production in German", *Journal of the International Phonetics Association*, 37, 173-182.

- SIMPSON, A. P. y S. NEUHAUSER (2009): "Enduring nature of epiphenomenal non-pulmonic sound production under disguise - a preliminary study", comunicación presentada en *IATPA-2009*, Cambridge (UK).
- SIMPSON, A. P. y S. NEUHAUSER (2010): "The persistency of epiphenomenal sound production in foreign accent disguise", en *IATPA-2010*, Tréveris.
- STOREY, K. C. J. (1996): "Constants in auditory and acoustic voice analysis in forensic speaker identification in cases of disguised voice", en H. Kniffka y S. Blackwell (eds.), *Recent Developments in Forensic Linguistics*, Fráncfort, Lang, págs. 203-216.
- THOMPSON, C. P. (1985): "Voice identification: Speaker identifiability and a correction of the record regarding sex effects", *Human Learning*, 14, 19-27.
- WÄGNER, I. y O. KÖSTER (1999): "Perceptual recognition of familiar voices using *faletto* as a type of voice disguise", *Proceedings of the 14th International Congress of Phonetic Sciences*, 1381-1384.
- WATT, D. (2010): "The identification of the individual through speech", en C. Llamas y D. Watt (eds.), *Language and Identities*, Edinburgo, Edinburgo University Press, págs. 76-85.
- ZHANG, C. (2005): *Acoustical Study on Disguised Voices*, tesis doctoral, Nankai University.
- ZHANG, C. y T. TAN (2008): "Voice disguise and automatic speaker recognition", *Forensic Science International*, 175, 118-122.

CAPÍTULO 11

LOS EFECTOS DEL RUIDO SOBRE LA PERCEPCIÓN DEL HABLA. APLICACIONES AUDIOMÉTRICAS

VICTORIA MARRERO

UNED

MARCELO J. RODRÍGUEZ CRUZ

Centro de Rehabilitación del Lenguaje (CRL)

ALFONSO IGUALADA PÉREZ

CSIC

1. INTRODUCCIÓN

1.1. *Los ruidos y sus efectos*

1.1.1. El ruido en el sistema de la comunicación

Desde un punto de vista físico, no hay ninguna diferencia entre el ruido y cualquier otro sonido: en todos los casos se trata de movimientos en las partículas del aire que nos rodea; lo que convierte un sonido en ruido es un rasgo psicológico: "el ruido se define como un sonido indeseado" (Namba, Kuwano y Schick, 1986). Por lo tanto, en todo estudio sobre el ruido resulta inevitable considerar la percepción subjetiva de ese estímulo por parte del receptor; si nuestra atención se centra, además, en la percepción del habla en entornos ruidosos, podemos comenzar por situar el ruido dentro del sistema de la comunicación, considerándolo como la interferencia (de todo tipo, no solo sonora) que sufre la señal; el sistema utiliza los canales de información oral-auditivo y gestual-visual de forma sincronizada, generando mecanismos flexibles de compensación cuando la tarea se vuelve exigente, para asegurar el éxito de una comunicación acorde al contexto.