



NUEVAS APORTACIONES AL ESTUDIO DE LA PERCEPCIÓN DEL HABLA

JUANA GIL - EUGENIA SAN SEGUNDO

Consejo Superior de Investigaciones Científicas - University of York

En un muy citado artículo sobre percepción del habla, Noël Nguyen 2005 hace una afirmación que resume de modo extraordinariamente sucinto un hecho incuestionable sobre el que se han escrito ininidad de tratados: «la forme sonore du langage est d'abord appréhendée à travers l'oreille» (p. 425). Es decir, no puede entenderse el estudio de las propiedades fonéticas y de los sistemas fonológicos de las lenguas, ni su relación con los procesos cognitivos implicados en el fenómeno del habla, sin tener en cuenta ineludiblemente el papel del oyente y el propio mecanismo de la percepción. Esta realidad, ya por sí sola, justifica toda la atención que pueda brindarse al fenómeno perceptivo, pero sucede, además, que en los últimos años se han realizado importantes avances en este terreno, y se han abierto nuevas perspectivas de análisis —algunas de ellas representadas en este volumen, como se verá— que hacen aún más recomendable, si cabe, tomarlo en consideración.

1. INTELIGIBILIDAD Y MULTIMODALIDAD

Entre otros enfoques interesantes, conviene destacar aquí, en primer lugar, dos grandes líneas de investigación que se han venido desarrollando en tiempos recientes: la que se refiere a la inteligibilidad del habla y a los diversos factores que la condicionan, y —vinculada, en realidad, con la anterior— la que concierne a las distintas fuentes de información, integradas ya en los estadios iniciales del procesamiento¹, en las que se apoyan los oyentes para percibir e interpretar los mensajes.

1. En palabras de Schroeder y otros 2003: «Recent findings in both monkeys and humans indicate that multisensory convergence occurs in low-level cortical structures generally believed to be unisensory in function» (p. 5). Esto es, las señales visuales, propioceptivas y auditivas convergen en los momentos iniciales del procesamiento auditivo cortical.





En la primera línea se inscriben, por ejemplo, los múltiples estudios relativos a cómo el grado de similitud de los estímulos percibidos y su integración en estructuras o redes mentales del oyente (cf. Luce y Pisoni 1998), el tamaño o a la longitud de estos mismos estímulos (cf. Pickett y Pollack 1963), el hecho de que hayan sido producidos por uno o por varios locutores (cf. Mullenix, Pisoni y Martin 1989), su frecuencia de uso (cf. Connine y otros 2008; Pitt, Dilley y Tat 2011), el sexo del locutor (cf. Bradlow, Torretta y Pisoni 1996), y otros diversos factores influyen en la inteligibilidad de los mensajes y en su percepción correcta. Dos variables más que se suman a las anteriores como elementos relevantes para poder alcanzar la correcta percepción serían el estilo de habla y la presencia o ausencia de ruido de fondo. A ambos se refiere, precisamente, el estudio de Marrero recogido en este volumen, cuyo objetivo ha sido investigar la relación entre el habla hiperarticulada generada por la locución en un entorno ruidoso (el denominado «efecto Lombard») y la identificación de la señal verbal en ruido por niños con déficit auditivo.

El efecto del ruido sobre la señal sonora se conoce como «enmascaramiento», porque el ruido «enmascara» o interfiere parcialmente en el mensaje transmitido². En este sentido, se sabe desde hace mucho tiempo que tal enmascaramiento resulta nocivo para todo tipo de hablantes, pero que aún lo es más para aquellos que están afectados de algún problema de audición (véanse, por ejemplo, Plomp 1978; Festen y Plomp 1983), y por ello los investigadores buscan permanentemente recursos válidos para contrarrestarlo (cf., entre otros, Brown y Bacon 2010; Wang y otros 2009). Uno de los posibles sistemas para hacerlo es, justamente, la producción de habla muy articulada: es decir, aquella que se produce persiguiendo y cuidando la inteligibilidad de la señal con el fin de lograr una decodificación óptima por parte del oyente (Lindblom 1996). En aquellas situaciones en las que el hablante está sometido a ruido de fondo, este efecto se consigue de manera natural elevando el tono fundamental y la intensidad, así como disminuyendo la velocidad de elocución. Pues bien, en el marco de las investigaciones llevadas a cabo conjuntamente por el Programa Infantil Phonak (PIP) y por la Universidad Nacional de Educación a Distancia, se crearon unas pruebas de audiometría verbal en ruido, entre cuyos objetivos se encontraba la búsqueda de un patrón de resultados distintos según el nivel de pérdida auditiva de niños entre 6 y 7 años: hipoacusia media-severa frente a profunda. Tras aplicar de forma progresiva distintas intensidades de ruido a una lista de frases, se elaboró un conjunto de estímulos perceptivos con el fin de presentárselos a un grupo de sujetos normoyentes y de sujetos hipoacúsicos. Además,

2. Más exactamente, Marrero, en el § 1.1. de su trabajo, define el enmascaramiento como el fenómeno que se produce cuando un estímulo interfiere y dificulta la percepción de otro.





se emplearon dos tipos de locución: por un lado, una locutora leía las frases sin ruido de fondo (condición «silencio»); por otro lado, la misma locutora debía leer las frases mientras recibía mediante auriculares un ruido de conversaciones multihablante. De este último modo se conseguía un estilo inconscientemente hiperarticulado (condición «Lombard»).

Los resultados de este experimento muestran que los efectos del ruido resultan especialmente perjudiciales para los niños con mayores pérdidas auditivas (casi en su totalidad usuarios de un implante coclear), que serían los grandes beneficiados de la locución hiperarticulada: pasan de identificar un 27% de las palabras clave en las listas en silencio a un 41% en las listas Lombard. Ello representa un incremento del 14%, mayor que el que se produce en los demás grupos, incluido el de los normoyentes. Por lo tanto, se concluye, entre otras cosas, que el habla hiperarticulada, ligada al efecto Lombard, consigue aumentar la inteligibilidad del habla en niños hipoacúsicos, de modo que puede resultar una técnica especialmente beneficiosa para el tratamiento de estos niños con hipoacusia severa e implantación coclear binaural.

Marrero termina su trabajo apuntando el interés que tendría en el futuro comprobar, en el marco de las investigaciones sobre percepción, el papel de la lectura labial como estrategia de compensación cuando existe pérdida auditiva, es decir, considerar el papel de apoyo que pueden desempeñar en la percepción otro tipo de claves, como son las visuales. Winn 2011 ya hizo este estudio para el inglés, y concluyó que precisamente las claves referidas al redondeamiento labial resultaban ser muy útiles para los sujetos con implantes cocleares e incluso para los normoyentes cuando se hallan en contextos que deterioran su audición. Por consiguiente, parece ser que los oyentes se sirven, cuando lo necesitan, de la covarianza (multimodal) que de modo natural acompaña a la señal de habla. Esto nos retrotrae a la segunda línea de investigación que mencionábamos más arriba, esto es, la referida a la integración de fuentes de información variadas o multimodales para la culminación exitosa del proceso perceptivo.

En efecto, un desarrollo interesantísimo y muy floreciente de los estudios sobre percepción del habla es el que concierne al análisis del papel desempeñado en ella por claves no necesariamente sonoras, sino también visuales, y a la interrelación que entre sí mantienen tales modalidades. Resulta ciertamente evidente que nuestro procesamiento y comprensión de las emisiones articuladas se basan no solo en el sonido que percibimos, sino también en otras claves visuales concomitantes, como pueden ser, por ejemplo, los gestos de la cara. Las regiones cerebrales que tradicionalmente se creía eran sensibles tan solo a los sonidos del habla parecen ahora responder a un cierto tipo de aducto visual que de inmediato se integra con el auditivo en diferentes contextos.





Como explica Hawkins 2004, ya a mediados del siglo pasado Sumby y Pollack 1954 demostraron que la inteligibilidad de ciertas frases aumentaba notablemente si se suministraba en paralelo a los oyentes apoyo visual, especialmente cuando existía ruido ambiental que hacía más necesario y pertinente este segundo tipo complementario de información. Desde entonces el fenómeno de la percepción multimodal del habla ha ido alcanzando cada vez mayor relevancia y se ha visto abordado desde diversas ópticas (cf., por ejemplo, los trabajos de Thompson y Massaro 1986; Munhall y otros 2004; Bahrack y Lickliter 2009; entre otros muchísimos mencionables). En España, destacan los estudios realizados en esta dirección por el grupo de investigación liderado por la Dra. Pilar Prieto, que ha abordado en numerosas publicaciones³ el análisis de la multimodalidad. En la línea marcada por este grupo se inscribe el artículo de Santiago González incluido en el presente volumen. En él se pone de manifiesto la relevancia de la percepción multimodal para comprender el fenómeno lingüístico de la ironía verbal. En efecto, aquí la percepción audiovisual –y no solo la auditiva– desempeña un papel fundamental a la hora de explicar las interacciones entre las características prosódicas y los aspectos gestuales que acompañan los enunciados irónicos. Se trata de un estudio de caso realizado sobre un corpus de 21 enunciados irónicos extraídos de varios monólogos humorísticos interpretados por un conocido presentador de televisión español. Un test perceptivo previo a los análisis cuantitativo y cualitativo de esta investigación permitió seleccionar tan solo los enunciados irónicos más prototípicos; es decir, tan solo aquellos valorados con un grado de ironía superior a 4,5 puntos (en una escala del 1 al 5) por cuatro informantes diferentes. Pese a su carácter preliminar y fundamentalmente exploratorio –debido a que tan solo se estudia un único sujeto– esta original investigación desde la perspectiva de la «prosodia audiovisual» permite comprobar cómo los enunciados irónicos y no irónicos contrastan en el número y en el modo en el que aparecen las marcas prosódicas y otras marcas gestuales, como los movimientos de la cabeza o de las cejas y otros elementos no verbales como, por ejemplo, la sonrisa o la risa.

2. LA PERCEPCIÓN, LA ADQUISICIÓN Y LA DIDÁCTICA DE LA FONÉTICA DE LENGUAS EXTRANJERAS

A diferencia del de González, el trabajo de Planas que se incluye en este volumen, plantea un test de percepción que no se sirve de ningún tipo de apoyo visual, como pueda ser el uso de imágenes o la presentación

3. Véase, por ejemplo, el trabajo reciente de Igualada, Bosch y Prieto 2015 sobre la coordinación y sincronización de los gestos y el habla en niños de doce meses de edad.





de los enunciados por escrito, sino que es exclusivamente auditivo. La primera finalidad de la autora, al realizarlo, fue conocer el grado en el que los estudiantes chinos de español como lengua extranjera son capaces de identificar enunciados interrogativos de duda y de enfado sin pistas adicionales visuales o de otra naturaleza. Asimismo, como segundo objetivo de la investigación, se pretendió encontrar qué tipo, o tipos, de inflexión tonal final de enunciado asocian los alumnos con más frecuencia a las intenciones comunicativas objeto de estudio: enfado-irritación-indignación, por un lado, y duda-extrañeza-incredulidad, por otro. En el estudio, se aprecia el intento por conseguir la mayor naturalidad y espontaneidad posibles en la locución de las frases que constituyen los estímulos del test perceptivo. Por ejemplo, las oraciones, extraídas de una obra de teatro, se grabaron dentro del contexto en el que las producían los personajes de la obra, y no de forma aislada, y se prestó especial atención a que los locutores imprimieran la intención comunicativa apropiada a cada enunciado interrogativo.

Una aportación del trabajo es la constatación de que existen seis tipos posibles de inflexión final para las dos modalidades de enunciados interrogativos expresivos que son objeto de estudio. Además, la autora recoge el porcentaje de aparición de cada una de esas seis inflexiones finales en los dos tipos de enunciados, pues no se puede delimitar un contorno único para cada intención comunicativa, sino que existe cierto solapamiento entre las distintas opciones. En cuanto a la percepción de estas intenciones comunicativas (el enfado y la duda) por parte de los alumnos chinos, alrededor del 50% de los enunciados interrogativos presentados en el test serían identificados correctamente, de acuerdo con las intenciones comunicativas correspondientes. Puesto que el porcentaje de acierto es algo mayor en las frases de enfado que en las de duda (un 63% frente a un 48%), se apunta al tono de voz como posible pista acústica que podría ayudar a los alumnos a identificar estos enunciados. La consideración de esta variable —o de otras como la velocidad de elocución— se mantendría, por tanto, como una posible perspectiva futura de trabajo.

El diseño de estudios perceptivos como el de Planas resulta especialmente útil como primer paso para la creación de material didáctico para alumnos cuya lengua materna no es entonativa, puesto que son los que suelen tener dificultades para discriminar y pronunciar los distintos contornos melódicos, en particular en su fase final. Por tanto, el objetivo último que guía a la autora es mejorar la enseñanza de la entonación expresiva del español a estudiantes orientales, en concreto a chinos. Esta finalidad directa o indirectamente didáctica es bastante común en todos los trabajos que, en los últimos años, han seguido incidiendo en la investigación del fenómeno perceptivo relacionándolo con el de la adquisición de lenguas extranjeras, todos ellos tendentes a dilucidar hasta qué punto



la percepción del habla se ve condicionada por las expectativas que la estructura de su propia lengua materna genera en el oyente⁴.

Relacionado asimismo con el ámbito de la enseñanza de una lengua extranjera, el artículo de Santamaría Busto incluido en este volumen revela la importancia de la percepción en este terreno, más concretamente en lo que respecta al papel de la evaluación perceptiva de la pronunciación de una L2. En su investigación, este autor señala cinco constructos o dimensiones que condicionan dicha evaluación, a saber: la inteligibilidad, la comprensibilidad, el acento extranjero, la fluidez y la irritabilidad. A partir de una rigurosa revisión bibliográfica llevada a cabo por el autor se desprende la dificultad de distinguir con nitidez algunas de tales dimensiones; no obstante, la investigación de Santamaría Busto consigue arrojar cierta luz sobre estas variables por medio de un análisis estadístico fundamentalmente basado en descubrir sus posibles correlaciones. Por ejemplo, hasta qué punto los resultados de la evaluación de la «inteligibilidad» del aprendiente de español como L2 se correlacionan con los resultados de una evaluación que mida su «comprensibilidad». En cuanto al diseño del experimento perceptivo, cabe señalar que la inteligibilidad se midió por medio de transcripciones, llevadas a cabo por evaluadores con el español como lengua materna mientras que el resto de aspectos relacionados con la evaluación perceptiva (comprensibilidad, acento extranjero, fluidez e irritabilidad) se midieron a través de escalas Likert de 9 niveles. Este método conlleva una graduación numérica con extremos que se oponen, por ejemplo *sin acento extranjero* y *acento extranjero extremadamente marcado*. En cualquier caso, la evaluación perceptiva se llevó a cabo a partir de la escucha de 24 frases producidas por ocho anglohablantes. Considerando que el tipo de interlocutor en español que los hablantes de L2 pueden encontrar más frecuentemente en un contexto de inmersión es el de no experto (o sea, sin experiencia en la docencia ni formación universitaria en lingüística), Busto Santamaría incluye en este experimento tanto evaluadores expertos como no expertos. Los resultados de este trabajo reflejan, entre otros aspectos, que los evaluadores tienden a sobrevalorar su capacidad para entender las muestras, que la inteligibilidad y la comprensibilidad guardan entre sí una estrecha relación, y que el acento extranjero es el constructo que más marca la diferencia en las valoraciones perceptivas entre jueces expertos y no expertos.

4. Sobre la influencia de la L1 en la percepción y la producción de la L2 existe una bibliografía inabarcable. Pueden encontrarse algunas síntesis útiles de los trabajos más influyentes sobre la cuestión en Best y Tyler 2007, Flege 1995, Gil 2007, Munro y Bohn 2007 o Strange 1995, aunque sin duda cabría citar muchas más obras.



Se sabe, a este respecto, que la percepción de un acento extranjero se deriva muy frecuentemente de la sensibilidad del oyente hacia el detalle fonético de bajo nivel⁵, y que solo la capacidad para atender a esas cuestiones, de detalle y graduales, que pueden ser relevantes en la lengua meta e irrelevantes en la materna o al revés, pero que en todo caso las distinguen, habilitará a los oyentes/hablantes a adquirir adecuadamente el componente fónico del nuevo idioma y evitar o, al menos, disminuir el acento no nativo⁶.

No obstante, esta noción de detalle fonético desborda ampliamente el ámbito de los estudios sobre percepción de lenguas extranjeras, y se generaliza a muchas otras investigaciones que tratan de averiguar en qué grado los oyentes, en general, son sensibles a índices fonéticos perceptivamente menos prominentes que los tradicionalmente asociados con las categorías y las oposiciones fonológicas, y, asimismo, en qué medida se sirven de tales claves para acceder al léxico y para entender el mensaje que perciben.

3. LA PERCEPCIÓN DEL DETALLE FONÉTICO

El detalle fonético, que —en palabras de Nguyen 2012 citando a Hawkins 2010— es la «subphonemic phonetic variability that contributes to phonological or other contrasts that distinguish meanings, but not necessarily lexical items» (p. 13), y que en los modelos teóricos más clásicos⁷ no constituía objeto de estudio independiente, ha adquirido en los últimos tiempos considerable importancia⁸, en gran parte derivada de

5. Por ejemplo, en inglés el sonido [ʃ] se articula con más redondeamiento labial que en francés (Payne 2006), y, en español, la [t] es una consonante dentoalveolar mientras que en inglés es plenamente alveolar. La incapacidad del oyente para percibir y, posteriormente, producir esas variaciones paramétricas graduales, no esenciales y no categóricas, provoca que el habla resultante no sea similar a la nativa.

6. Sobre los factores que contribuyen, en general, al mayor o menor grado de acento extranjero, el lector interesado puede consultar la revisión que realizan Piske, MacKay y Flege 2001.

7. Como es sabido, en Chomsky y Halle 1968 y en muchas de las obras que siguieron a esta, las propiedades fonéticas «físicas» (entre ellas, las concernientes a la coarticulación, la sincronización de los articuladores, etc.) se consideraban extragramaticales. La fonética quedaba, así, fuera de la gramática generativista de los primeros tiempos, y esta no fue, por tanto, tan extremadamente radical en su ruptura con el estructuralismo —que separaba nítidamente la fonología de la fonética— como se ha querido interpretar después: las matrices de rasgos generativistas se aproximaban bastante a la idea de fonema, y las reglas complementarias de variación alofónica de bajo nivel no se llegaron nunca a formular de manera explícita, ya que se asumía que los principios universales de realización fonética proveían automáticamente muchos de estos detalles al final de todo el proceso de derivación de la secuencia fónica.

8. Esto no solo ha ocurrido en el ámbito de los estudios fónicos, sino también en el de la psicolingüística. Por ejemplo, McMurray y otros 2003 escriben: «Research in speech perception has been dominated by a search for invariant properties of the signal that correlate with lexical



las más recientes concepciones fonológicas en las que las representaciones léxicas son de naturaleza fonética, de un alto grado de complejidad y no sujetas a ningún proceso de eliminación de redundancias⁹.

Ha de entenderse, además, que el análisis de tal detalle fonético no se circunscribe ahora solo a aquellas propiedades o peculiaridades que se asociarían con el léxico (por ejemplo, el punto exacto de articulación de los sonidos, la forma y duración de las transiciones acústicas entre ellos, la diversa extensión de la coarticulación, o la duración variable del *Voice Onset Time*), sino que abarca todos aquellos otros rasgos de mayor alcance que contribuyen igualmente a la comprensión y la correcta interpretación de los enunciados contextualizados en las interacciones comunicativas, como, por ejemplo, las características relacionadas con la cualidad de voz o con ciertas microfluctuaciones tonales. En este sentido, los artículos firmados por Beatriz Blecua y Assumpció Rost, por una parte, y por Infante, por otra, representan, respectivamente, los dos modos de explorar el detalle fonético y de precisar sus funciones.

En el trabajo de Blecua y Rost, las autoras investigan los efectos perceptivos de la variación acústica que presenta la fricativa labiodental sorda [f] en español. Tras comprobar en un primer análisis acústico que existen cuatro realizaciones alofónicas de dicho fonema en habla espontánea, las autoras diseñan dos pruebas perceptivas para averiguar hasta qué punto los oyentes perciben esa variación de detalle; es decir, se pretende investigar las repercusiones perceptivas de un fenómeno acústico descrito previamente por las mismas autoras. Los cuatro alófonos posibles de la fricativa labiodental sorda se organizarían a lo largo de un continuo desde la hipo- a la hiperarticulación.

La primera prueba perceptiva, de identificación fonética –realizada por oyentes con conocimientos básicos de fonética– pone de manifiesto que las diferencias entre las distintas manifestaciones acústicas de este fonema son perceptibles. El segundo test perceptivo se lleva a cabo con los mismos estímulos que la primera prueba pero, en este caso, los oyentes –que

and sublexical categories. We argue that this search for invariance has led researchers to ignore the perceptual consequences of systematic variation within such categories and that sensitivity to this variation may provide an important source of information for integrating information over time in speech perception. Data from a study manipulating VOT continua (...) indicate that lexical access shows graded sensitivity to within-category variation in VOT and that this sensitivity has a duration sufficient to be useful for information integration» (p. 77).

9. Así, por ejemplo, en los modelos de ejemplares, cada palabra se asocia en el léxico con una lista de ejemplares que reflejan sus distintas realizaciones dependientes del contexto, y que incluyen propiedades codificadas referidas a las características no solo estrictamente fonéticas, sino también físicas, psicológicas y sociales de los hablantes, las cuales en modelos anteriores quedaban completamente excluidas de las representaciones mentales abstractas de carácter puramente fonológico (véase, a este respecto, el último apartado del presente trabajo).



carecen por completo de conocimientos de fonética— deben identificar los distintos alófonos eligiendo entre dos posibilidades gráficas: <f> y . Se trata, por tanto, de una prueba de identificación fonológica. Los resultados de ambos experimentos vienen a demostrar que la variación acústica de detalle tiene una repercusión perceptiva clara. Por un lado, no se da una correspondencia unívoca entre las distintas realizaciones acústicas y el fonema /f/. Por otro, los resultados indicarían que algunas de las realizaciones sonoras pueden interpretarse como alófonos tanto del fonema fricativo labiodental sordo /f/ como del bilabial sonoro /b/. De esta forma, se ve confirmada la hipótesis principal del estudio, esto es, que la variación fonética de un fonema puede llevar a confusión en la categorización de sus distintas realizaciones, llegando, en el caso concreto de /f/, a identificarse las variantes más relajadas con un fonema sonoro.

Estos resultados se discuten en relación con el marco teórico de la fonología evolutiva de Blevins 2004, y permiten arrojar cierta luz sobre el fenómeno diacrónico que llevó del fonema labiodental sordo existente en latín a la solución aproximante bilabial del castellano actual en palabras como *raphanum* > rábano.

Por otro lado, en lo que se refiere al trabajo de Infante, la autora explora la posible diferenciación, desde el punto de vista perceptivo, entre la cualidad de voz derivada del modo de fonación conocido como *creak* y la voz *creaky*¹⁰, esta última como resultado de un ajuste laríngeo que se produciría conjuntamente con la voz modal. Es decir, lo que se pretende en este artículo, no es ya averiguar si los oyentes perciben más o menos sutiles variantes segmentales, sino comprobar si el grupo de jueces que interviene en el experimento diseñado, todos ellos expertos en fonética, es capaz de discriminar perceptivamente dos tipos de fonaciones. Para ello, debían clasificar cada uno de los treinta estímulos escuchados en *creak*, *creaky* o *no creak* (y). Tales estímulos consistían en enunciados de unos 40 segundos de duración, extraídos de conversaciones semiespontáneas y producidos por hablantes masculinos de español centropeninsular. Los resultados muestran diferencias entre los distintos tipos de jueces que realizaron el experimento perceptivo. Así, el subgrupo de fonetistas entrenados específicamente en cualidad de voz *creak* (y) mostró el menor grado de desacuerdo en sus juicios, seguido del subgrupo de jueces entrenados en el manejo del protocolo VPA en

10. Como se explica en el propio artículo de Infante, los términos ingleses *creak* y *creaky voice* no son los únicos con los que se hace referencia en la bibliografía anglosajona a este tipo de fonación, sino que a menudo se emplean otros como *laryngealization*, *glottal fry*, *glottal stop*, *stiff voice*, *pulse register* o *pulse phonation*. De igual modo, en español las traducciones que se han propuesto para dar cuenta del fenómeno son variadísimas: voz rota, voz crepitante, voz pulsada, frito vocal, etc. En muchas ocasiones, por ello, es preferible mantener el término inglés original y más inequívoco —*creak/creaky*— para no confundir al lector.



general (*Vocal Profile Analysis*)¹¹ y del subgrupo de jueces-fonetistas cuya lengua materna no es el español.

Indudablemente este estudio preliminar aborda una cuestión poco estudiada pero, como bien indica la autora, sería necesario ampliar la muestra de jueces perceptivos para obtener resultados más fiables. Es ciertamente importante comprobar hasta qué punto los oyentes son sensibles a los cambios en la cualidad de voz, sean estos más o menos sutiles, porque tales modificaciones pueden aportar información muy valiosa. En primer lugar, ya se ha comprobado (Ogden 2001, 2002) que un modo de fonación no modal, más precisamente el *creak*, adquiere valor lingüístico en finés, lengua en la que actúa como clave para la cesión del turno de palabra en las conversaciones, con independencia de que se empleen otros recursos sintácticos o entonativos. Algo similar se ha señalado también con respecto al inglés («a low falling intonation is also often terminated by creak or creaky voice. In English, termination of this sort is sometimes used for a regulative function ... with the speaker using a creaky termination as a signal of yielding the floor to the other speaker, at the end of the speaker turn», Laver 1994, p. 196), pero aún no existe ningún estudio sobre esta cuestión, que sepamos, con respecto al español. Comprobar si en lenguas enmarcadas en otras familias lingüísticas diferentes, como sería el caso de las románicas, ocurre algo semejante tendría repercusiones interesantes en el ámbito de los universales fónicos y de los estudios tipológicos.

En segundo lugar, la percepción de los cambios, incluso de detalle, en la cualidad de voz tiene muchas implicaciones que van más allá del campo de la fonética, para entrar de lleno en el de la pragmática y la psicolingüística, en cuanto que la cualidad vocal contribuye a crear determinados estereotipos socioculturales que no son sino las expectativas que las comunidades de individuos se forman y comparten acerca de la personalidad de los hablantes y sobre los atributos vocales básicos que la determinan (agresividad, atractivo, bondad, etc.). Tales estereotipos no moldean únicamente la percepción de las características idiosincráticas del locutor sino que, por extensión asociativa, condicionan la percepción y la interpretación de los mensajes, de modo que, en el mejor de los casos, no debería existir contradicción entre los rasgos vocales con los que se transmiten y su contenido.

11. El protocolo descriptivo de la cualidad de voz conocido con las siglas VPA (correspondientes a la expresión inglesa *Vocal Profile Analysis*) fue ideado por John Laver 1980 y es ampliamente comentado en Ball, Esling y Dickson 2000 o Mackenzie Beck 2005, entre otras muchas obras mencionables. Básicamente permite analizar la voz del hablante disociándola en componentes, entendidos como ajustes musculares de largo alcance que conllevan consecuencias perceptivas y que, juntos, conforman el resultado final, esto es, su cualidad a medio-largo plazo.



Este es precisamente el objeto de estudio del trabajo de Fernández Trinidad que se integra en el presente volumen. Se aborda en él, en efecto, la cuestión de los estereotipos vocales ligados a la cualidad de voz; más en concreto, se estudian dos tipos de fonación: la llamada voz *pressed* y la voz *breathy*¹², opuestas en cuanto al esfuerzo laríngeo que requiere su producción: la primera presenta un elevado grado de tensión y supone un importante esfuerzo laríngeo, mientras que el segundo tipo de fonación, considerada hipofuncional, se produce con escaso esfuerzo laríngeo. Las posibles combinaciones de ambos tipos de fonación con un tono agudo o grave se presentan en un test perceptivo a un conjunto de jueces con el fin de evaluar las posibles asociaciones de las que son susceptibles; en otras palabras, qué tipos de voces se vinculan de manera preferente con determinadas características o atributos. Este experimento perceptivo se enmarca dentro del ámbito de la publicidad y para ello se seleccionaron cuatro tipos de productos, de tal forma que la tarea de los jueces que participaron en el estudio consistía en elegir qué tipo de voz les resultaba más apropiada para anunciar cada producto. Los resultados de este experimento ponen de manifiesto que un mismo tono o un mismo tipo de fonación pueden suscitar asociaciones diversas. No obstante, el análisis multidimensional llevado a cabo revela que los oyentes seguirían unas pautas de combinación bien establecidas, no aleatorias, para establecer asociaciones perceptivas a partir de una determinada cualidad de voz.

4. REPRESENTACIONES Y PROCESAMIENTO

Aunque en el presente volumen no hay cabida para todas ellas, la investigación actual en percepción del habla abarca otras muchas líneas. Una de las más importantes y de las que más estudios han generado es la relativa a la representación de los sonidos, de los patrones fónicos y de las palabras en la mente —en la memoria— de los hablantes y de los oyentes, y al modo en que unos y otros acceden a tales representaciones. A decir verdad, este es un tema central o nuclear en los estudios de percepción porque, dependiendo de la postura que se adopte con respecto a él, pueden abordarse los otros aspectos relacionados con el proceso, incluidos los que aquí se han tratado.

Básica y muy simplificada, la cuestión tiene que ver con el papel que se le conceda a la variabilidad observada en el habla: ciertos modelos distinguen con claridad entre, por un lado, formas abstractas, invariantes y almacenadas en el léxico mental de los individuos, y, por otro, formas

12. De nuevo en este caso, las traducciones al español de estos términos no son siempre coincidentes (véase la nota 4 del trabajo de Fernández Trinidad), por lo que parece aconsejable mantenerlos aquí en su forma inglesa original.



fonéticas superficiales no presentes en el léxico y con un enorme grado de variación debida a razones contextuales o individuales (cf. Ingram 2007, Reiss 2007); otros modelos, por el contrario, defienden que las representaciones mentales incorporan ya toda la variabilidad tanto contextual como individualizadora de los hablantes, es decir, incluyen no solo rasgos vinculados al entorno en el que aparecen los sonidos en la secuencia, sino también propiedades asociadas a las características anatómicas y socioculturales de los locutores que afectan a la producción de su habla, lo que en inglés se denomina *indexical properties* (cf. Goldinger 1996, 1998; Johnson y Mullenix 1997).

En efecto, durante mucho tiempo se había pensado —en gran medida por la influencia de los enfoques fonológicos que sistemáticamente dejaban fuera de las representaciones subyacentes las variaciones debidas a la realización superficial en la secuencia o a las peculiaridades del locutor— que, en el proceso de percepción, los oyentes abstraían toda la fluctuación que las características físicas de los sonidos pueden presentar y toda la información referida al hablante que los produce para concentrarse en las categorías fonológicas contrastivas, abstrayendo lo esencial de lo contingente. Desde hace unas décadas, sin embargo, estas concepciones abstraccionistas han ido dejando paso a otros planteamientos en los que los aspectos relativos a la variabilidad y al detalle fonético, del que se ha hablado antes en estas mismas páginas, han pasado a ocupar un papel fundamental en el proceso de percepción y procesamiento del habla, en la medida en que se considera que en la memoria de los hablantes la representación de las palabras puede retener información fonética muy detallada (relativa, por ejemplo, a la voz de la persona que las pronunció) así como información de otros tipos (visual, táctil, etc., cf. Johnson y otros 1999).

Pues bien, si se retoman por un momento presupuestos teóricos como el de que la percepción del habla es multimodal, en el sentido que se ha explicado anteriormente, se comprueba que tal planteamiento es fácilmente asumible si se parte de un enfoque no abstraccionista según el cual en la memoria de los hablantes se almacenan diversas huellas de los distintos elementos léxicos percibidos por el oyente con todas sus características, es decir, ejemplares en sí mismos multimodales: cuando el oyente se ve expuesto a un estímulo, la experiencia de una determinada modalidad activa la de otras y el significado de lo que ese oyente está escuchando finalmente se deriva de la combinación de todas estas asociaciones correspondientes a, y vinculadas con, un único estímulo. Por explicarlo con un ejemplo, cuando oímos la palabra *rosa*, no solo se activaría en nuestro cerebro la forma sonora almacenada más similar, en todos sus detalles, al vocablo escuchado, sino también otros muchos recuerdos gestuales, visuales, táctiles u olfativos ligados al concepto de *rosa*.



Entender el proceso de la percepción de este modo, holístico, permite comprender mucho mejor algunos de los trabajos incluidos en este volumen por las razones expuestas al presentarlos en esta breve introducción. Con independencia de ello y de modo general, el debate sobre la naturaleza de las representaciones mentales y de los procesos de *mapping* o proyección entre ellas y los estímulos percibidos no está en modo alguno cerrado y es una línea de investigación sumamente relevante; por consiguiente, es de esperar que, en el marco de los estudios sobre el español, este enfoque se afiance con nuevas e interesantes aportaciones. Aún queda mucho trabajo por hacer y restan muchas interrogantes que responder.

Madrid, 15 de noviembre de 2015.



5. REFERENCIAS BIBLIOGRÁFICAS

- BAHRICK, L. E. y LICKLITER, R. (2009): «Perceptual development: Intermodal perception», en Goldstein, B. (ed.), *Encyclopedia of perception*, Vol. II, Newbury Park, CA, Sage Publishers, pp. 753-756.
- BALL, M. J., ESLING, J. y DICKSON, C. (2000): «Transcription of voice», en Kent, R. D. y Ball, M. J. (eds.), *Voice quality measurement*, San Diego, CA., Singular Publishing Group, pp. 45-58.
- BEST, C. T., McROBERTS, G. y GOODELL, E. (2001): «Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system», *Journal of the Acoustical Society of America* 109, pp. 775-794.
- BEST, C. T. y TYLER, M. D. (2007): «Nonnative and second-language speech perception: Commonalities and complementarities», en Munro, M. J. y Bohn, O.-S. (eds.), *Language experience in second language speech learning*, Amsterdam, John Benjamins, pp. 13-34.
- BLEVINS, J. (2004): *Evolutionary phonology*, Cambridge, Cambridge University Press.
- BRADLOW, A. R., TORRETTA, G. M. y PISONI, D. (1996): «Intelligibility of normal speech I: global and fine-grained acoustic-phonetic talker characteristics», *Speech communication* 20, pp. 255-272.
- BROWN, C. A. y BACON, S. P. (2010): «Fundamental frequency and speech intelligibility in background noise», *Hearing research* 266, pp. 52-59. Consultado en <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2885573/> " \l "R15.
- CHOMSKY, N. y HALLE, M. (1968): *The sound pattern of English*, Nueva York, Harper & Row.
- CONNINE, C.M., RAMBON, L. y PATTERSON, D. (2008): «Processing variant forms in spoken word recognition: The role of variant frequency», *Perception and psychophysics* 70, pp. 403-411.
- FESTEN J. M. y PLOMP, R. (1983): «Relations between auditory functions in impaired hearing», *Journal of the Acoustical Society of America* 73, pp. 652-62.
- FLEGE, J. E. (1995): «Second-language speech learning: Theory, findings, and problems», en Strange, W. (ed.), *Speech perception and linguistic experience*, Timonium, MD, York Press, pp. 233-277.
- GIL, J. (2007): *Fonética para profesores de español: de la teoría a la práctica*, Madrid, Arco/Libros.
- GOLDINGER, S. D. (1996): «Words and voices: episodic traces in spoken word identification and recognition memory», *Journal of experimental psychology: learning, memory and cognition* 22, pp. 1166-1183.
- , (1998): «Echo of echoes? An episodic theory of lexical access», *Psychological review* 105, pp. 251-279.

- HAWKINS, S. (2004): «Puzzles and patterns in 50 years of research on speech perception», en Slifka, J., Manuel, S. y Matthies, M. (eds.), *From sound to sense*, Boston, MA., Research Laboratory of Electronics at MIT, CD Rom.
- , (2010): «Phonetic variation as communicative system: Perception of the particular and the abstract», en Fougeron, C., D'Imperio, M., Kühnert, B. y Vallée, N. (eds.), *Papers in laboratory phonology X*, Berlín, Mouton de Gruyter.
- IGUALADA, A., BOSCH, L. y PRIETO, P. 2015: «Language development at 18 months is related to communicative strategies at 12 months», *Infant behavior and development* 39, pp. 42-52.
- INGRAM, J. C. L. (2007): *Neurolinguistics: An introduction to spoken language processing and its disorders*, Cambridge, Cambridge University Press.
- JOHNSON, K. y MULLENIX, J. W. (eds.) (1997): *Talker variability in speech processing*, Nueva York, Academic Press.
- JOHNSON, K., STRAND, E. A. y D'IMPERIO, M. (1999): «Auditory-visual integration of talker gender in vowel perception», *Journal of phonetics* 27, pp. 359-384.
- LAVER, J. (1980): *The phonetic description of voice quality*, Cambridge, Cambridge University Press.
- , (1994): *Principles of phonetics*, Cambridge, Cambridge University Press.
- LINDBLOM, B. (1996): «Role of articulation in speech perception: clues from production», *Journal of the Acoustical Society of America* 99, pp. 1683-1692.
- LUCE, P. y PISONI, D. (1998): «Recognizing spoken words: the neighborhood activation model», *Ear and hearing* 19, pp. 1-36.
- MACKENZIE Beck, J. (2005): «Perceptual analysis of voice quality: The place of vocal profile analysis», en Hardcastle, W. J. y Mackenzie Beck, J. (eds.), *A figure of speech. A festschrift for John Laver*, Londres, Lawrence Erlbaum Associates.
- MACMURRAY, R., TANENHAUS, M. K., ASLIN, R. N. y SPIVEY, M. J. (2003): «Probabilistic constraint satisfaction at the lexical/phonetic interface: Evidence for gradient effects of within-category VOT on lexical access», *Journal of psycholinguistic research* 32, pp. 77-97.
- MULLENIX, J., PISONI, D. y MARTIN, C. (1989): «Some effects of talker variability on spoken word recognition», *Journal of the Acoustical Society of America* 85, pp. 365-378.
- MUNHALL, K. G., JONES, J. A., CALLAN, D. E., KURATATE, T. y VATIKIOTIS-BATESON, E. (2004): «Visual prosody and speech intelligibility. Head movements improves auditory speech perception», *Psychological science* 15, pp. 133-137.
- MUNRO, M. J. y BOHN, O.-S. (2007): «The study of second language speech learning: A brief overview», en Bohn, O.-S. y Munro, M. J. (eds.), *Language experience in second language speech learning*, Ámsterdam, John Benjamins, pp. 3-12.
- NGUYEN, N. (2005): «La perception de la parole», en Nguyen, N., Wauquier-Gravelines, S. y Durand, J. (eds.), *Phonologie et phonétique: Forme et substance*, París,

- Lavoisier, pp. 425-447. Consultado en HAL archives-ouvertes.fr. Id: hal-00142953 <https://hal.archives-ouvertes.fr/hal-00142953>.
- , (2012): «Representations of speech sound patterns in the speaker's brain: Insights from perception studies», en Cohn, A. C., Fougeron, C. y Huffman, M. K. (eds.), *The Oxford handbook of laboratory phonology*, Oxford, Oxford University Press, pp. 359-369. Consultado en <http://www.researchgate.net/publication/242586713>.
- OGDEN, R. (2001): «Turn-holding, turn-yielding and laryngeal activity in Finnish talk-in-interaction», *Journal of the International Phonetics Association* 31, 139-152.
- , (2002): «Creaky voice and turn-taking in Finnish», *Proceedings of the British Association of Academic Phoneticians Colloquium*, 25-27 de marzo, University of Newcastle-upon-Tyne. Consultado en <http://www.baap.ac.uk/colloquium/2002.html>
- PAYNE, E. (2006): «Phonetic motifs and the formation of sound structure», *Working papers in linguistics. University College London* 18, pp. 322-343.
- PICKETT, J. M. y POLLACK, I. (1963): «Intelligibility of excerpts from fluent speech: Effects of rate of utterance and duration of excerpt», *Language and speech* 6, pp. 151-164.
- PISKE, T., MACKAY, I. R. A. y FLEGE, J. E. (2001): «Factors affecting degree of foreign accent in an L2: a review», *Journal of phonetics* 29, pp. 191-215.
- PITT, M. A., DILLEY, L. y TAT, M. (2011): «Exploring the role of exposure frequency in recognizing pronunciation variants», *Journal of Phonetics* 39, pp. 304-311.
- PLOMP R. (1978): «Auditory handicap of hearing impairment and the limited benefit of hearing aids», *Journal of the Acoustical Society of America* 63, pp. 533-49.
- REISS, C. (2007): «Modularity in the 'sound' domain: Implications for the purview of the Universal Grammar», en Ramchand, G. y Reiss, Ch. (eds.), *The Oxford handbook of linguistic interfaces*, Oxford, Oxford University Press, pp. 53-79.
- SCHROEDER, C. E., SMILEY, J., FU, K. G., O'CONNELL, M. N., MCGINNIS, T. y HACKETT, T. A. 2003: «Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing», *International journal of psychophysiology* 50, pp. 5-17.
- STRANGE, W. (1995): «Cross-language studies of speech perception: A historical review», en Strange, W. (ed.), *Speech perception and linguistic experience*, Timonium, MD, York Press, pp. 3-45.
- SUMBY, W. H. y POLLACK, I. (1954): «Visual contribution to speech intelligibility in noise», *Journal of the Acoustical Society of America* 26, pp. 212-215.
- THOMPSON, L.A. y MASSARO, D. (1986): «Evaluation and integration of speech and pointing gestures during referential understanding», *Journal of experimental child psychology* 42, pp. 144-168.



- WANG, D., KJEMS, U., PEDERSEN, M. S., BOLDT, J. B. y LUNNER, Th. (2009): «Speech intelligibility in background noise with ideal binary time-frequency masking», *Journal of the Acoustical Society of America* 125, pp. 2336-2347. Consultado en <http://web.cse.ohio-state.edu/~dwang/papers/WKPBL.jasa09.pdf>
- WINN, M. B. (2011): *The use of acoustic cues in phonetic perception: Effects of spectral degradation, limited bandwidth and background noise*, Tesis doctoral, Universidad de Maryland. Consultado en <http://gradworks.umi.com/34/95/3495595.html>.