

El ensordecimiento de las sibilantes: una cuestión de oído

Sibilant devoicing: A matter of hearing

Assumpció ROST BAGUDANCH

Universitat Autònoma de Barcelona

Assumpcio.Rost@uab.cat

ORCID iD: <https://orcid.org/0000-0001-8133-881X>

RESUMEN

El proceso de ensordecimiento de las sibilantes del español supone la desfonologización de los tres fonemas sonoros existentes en castellano medieval, a saber, /z/, /ʒ/, /d̪z/. Las explicaciones que se han ofrecido tradicionalmente para dar cuenta de un cambio tan drástico se centran en aspectos de reorganización fonológica o en el contacto lingüístico con el vasco, pero no suelen tener en cuenta la posibilidad de que la evolución tenga un origen puramente fonético, como defienden Alonso (1967) o Pensado Ruiz (1993). Siguiendo el camino de Widdison (1997), se plantea un acercamiento desde la fonética experimental, esencialmente desde la percepción, a partir de la gramática comparada: el catalán presenta un sistema de sibilantes similar al del castellano medieval. Los resultados apuntan a que el inicio de la neutralización de sibilantes parece hallarse en la existencia de variación fonética y en cómo se interpreta esta variación auditivamente: se da una mayor distancia perceptiva entre los pares de sibilantes fricativas (ante los que los jueces muestran más sensibilidad) que entre los pares de africadas, especialmente en el caso de las dentoalveolares, cuya distancia entre sordas y sonoras es mucho menor (y ante las que los participantes tienen una sensibilidad más baja). Así pues, una concepción del cambio no teleológica describe de forma mucho más satisfactoria y adecuada un cambio que no se contempla ya como excepcional ni extraño.

PALABRAS CLAVE

Lingüística diacrónica, fonética histórica, sibilantes, ensordecimiento, fonética perceptiva.

ABSTRACT

One of the most analysed phenomena in the evolution of Spanish is the sibilant devoicing process, which started in Medieval Spanish. It entailed the dephonologization of the three voiced segments /z/, /ʒ/, /d̪z/. Most of the traditional accounts for this process have focused on (a) the need of reorganization of the phonological system to gain symmetry, or (b) language contact with Basque. However, most of them disregard a purely phonetic origin of this change, as authors like Alonso (1967) or Pensado Ruiz (1993) maintain. Following the lead of Widdison (1997), we propose a different approach, taking as a starting point experimental phonetics and particularly a perception analysis in comparative grammar, since Catalan displays a sibilant system very similar to the Medieval Spanish one. Results suggest that sibilant devoicing is a phonetically-based sound change, in which allophonic variation and its perceptive interpretation play a crucial role. The auditory distance between the voiced and voiceless segments in the fricative pairs is clearly greater than in the affricate

pairs, especially when it comes to the dentoalveolar one. Even the sensibility to the signal of the participants is higher in the fricative sibilants than in the affricates. Therefore, a non-teleological conception of sound change seems to account for the evolution in a more satisfactory and adequate way. From this perspective, sibilant devoicing is no more seen as exceptional or rare.

KEYWORDS

Diachronic linguistics, historical phonetics, sibilants, devoicing, perceptive phonetics.

Fecha de recepción: 17/11/2023.

Fecha de aceptación: 1/1/2023.

1. Introducción

Hay varios aspectos de la evolución fónica del latín al español actual que han implicado, en distintos momentos históricos, una reestructuración sustancial del sistema fonológico. Estos aspectos, lógicamente, han sido objeto de estudio y debate desde los inicios de la fonética histórica como disciplina científica. Uno de ellos es la reorganización de las sibilantes, que sufren varios procesos fonológicos sustanciales antes de llegar a la configuración actual: ensordecimientos, desfonologizaciones, cambios en el lugar de articulación, pérdida de la sibilancia, etc. Las explicaciones sobre la cronología y las causas que pudieron llevar a tales alteraciones se han ofrecido desde múltiples perspectivas (Alonso 1967, Alarcos Llorach 1988, Lapesa 1981, Rost Bagudanch & Roseano 2021, Mackenzie 2022), pero no todo está resuelto. De hecho, las posibilidades que ofrece la investigación actual abren la puerta a nuevos enfoques en su estudio que han de contribuir a una mejor comprensión de todo el fenómeno.

Este trabajo solo pretende abordar uno de estos procesos, el del ensordecimiento. De hecho, se parte de una investigación anterior (Rost Bagudanch 2022) en que se llevaron a cabo dos experimentos desde la gramática comparada, uno de producción y otro de percepción, para dilucidar si, en sistemas análogos a los del castellano medieval, se daba un fenómeno análogo. Esta primera aproximación desde la fonética experimental permitió advertir que el cambio se podía explicar por razones puramente acústicas y perceptivas, tal como apuntaba ya Pensado Ruiz (1993). Sin embargo, los resultados obtenidos, especialmente los relativos a la percepción del sonido, no dejaban de ser preliminares, por lo que se ha decidido profundizar en su análisis empleando las herramientas que ofrece la Teoría de Detección de Señales (TDS, de aquí en adelante; *cfr.* MacMillan 2002, Abdi 2007, Marrero 2014). Asimismo, la publicación de investigaciones recientes (Mackenzie 2022, Núñez-Méndez 2022, Bradley, Morris & Jones 2023) que aparecieron con posterioridad a la que aquí sirve de base, ha de tenerse en cuenta para ofrecer una explicación y contextualización más adecuada para el proceso que lleva a la desfonologización de la mitad de las sibilantes medievales, puesto que permiten revisar, al menos en parte, la periodización del cambio tradicionalmente establecida.

De este modo, el objetivo primordial de esta investigación es examinar el papel que la percepción pudo jugar en el ensordecimiento de las sibilantes, y qué puesto ocupa este proceso en la reorganización de estas consonantes a la luz de las últimas publicaciones. Siguiendo a Pensado Ruiz (1993) y los datos iniciales obtenidos en Rost Bagudanch (2022), se cree que la pérdida de sonoridad se explica por cuestiones propiamente fonéticas en las que el componente auditivo participa de forma decisiva. Como defienden

Pierrehumbert (2001, 2002) u Ohala (2012), el reanálisis y la recategorización de la señal por parte del oyente sería el elemento clave que permitiría poner en marcha el cambio (cfr. también Beddor 2012, 2023, Goldrick & Cole 2023).

El trabajo se organiza como sigue: en los epígrafes siguientes se presentarán las nociones de base sobre las que se sustenta este artículo, que abarcan desde la descripción del proceso de ensordecimiento dentro de la reorganización del sistema de sibilantes (§ 1.1) y los aspectos fonéticos de la sonoridad en este tipo de sonidos (§ 1.2) hasta los planteamientos más teóricos sobre el cambio diacrónico y el papel que juega en él la percepción (§ 1.3). En el segundo apartado, se presenta el diseño experimental y, en el tercero, se detallan los resultados obtenidos (tanto los más descriptivos, en § 3.1, como los propios de los parámetros perceptivos que interesa estudiar aquí, en § 3.2). Se reserva el apartado 4 para la discusión y, finalmente, en el 5, se presentan unas breves conclusiones.

1.1. El proceso de ensordecimiento de las sibilantes medievales

El sistema medieval de sibilantes del español consistía, atendiendo a la bibliografía especializada, en tres pares de fonemas, dos fricativos y uno africado: las alveolares /s/ y /z/, las alveopalatales /ʃ/ y /ʒ/ y las dentoalveolares /tʃ/ y /dʒ/. Además, existía una africada palatal sorda /tʃ̥/, para la que, en principio, no había variante fonológica sonora (Lapesa 1981, Alarcos Llorach 1988, Lloyd 1993, Penny 1993, Bradley & Delforge 2006 o Núñez-Méndez 2021). Es importante notar que la oposición de sonoridad se habría desarrollado en contexto intervocálico en latín vulgar (Lapesa 1981: 124, Lloyd 1993: 238-240, 389, 423), para perderse posteriormente de nuevo en español clásico, entre los siglos xv y xvi (Alonso 1967: 312, Lapesa 1981: 283, Eddington 1987: 59, Lloyd 1993: 427-428, Cano Aguilar 2004: 833-834, Ariza 2012: 222-223, Penny 2014: 121-122), aunque algunas investigaciones apuntan a que la pérdida total de los correlatos sonoros habría de situarse hacia el xvii (Núñez-Méndez 2021: 40-49) o, en todo caso, más bien en el xvi (Mackenzie 2022), en línea con Menéndez Pidal (1985: 115).

En cualquier caso, este proceso de ensordecimiento no habría sido rápido ni uniforme y habría empezado mucho antes. Algunos autores, de hecho, sitúan su inicio en el s. xiii (en el caso del par dentoalveolar, *vid.* Sánchez-Prieto Borja 2004: 442) o en el xiv (Penny 2004: 603-604, Quilis 2005: 270-271). El análisis de los errores gráficos indicaría que las confusiones habrían comenzado en la oposición /tʃ/ - /dʒ/, a partir de la cual se habría extendido a los pares de sibilantes fricativas (Lapesa 1981: 283, Eddington 1987: 58, Quilis 2005: 168-169, Ariza 2012: 224)¹. Núñez-Méndez (2021: 19) detalla que este fenómeno se originó en el norte peninsular para difundirse de forma gradual y no homogénea hacia el sur e insiste, como es lógico en situaciones de cambio fónico, que se trataría de una situación en la que la variación y las vacilaciones habrían sido lo habitual a lo largo de estos siglos (cfr. también Lapesa 1981: 283, Sánchez-Prieto Borja 2004: 441-442, Quilis 2005: 168 o Ariza 2012: 222-224).

¹ Mackenzie (2022) cuestiona esta cronología y argumenta, a partir del estudio de un corpus documental, que el ensordecimiento es posterior a la fricativización y a la pérdida de la sibilancia de estas consonantes: /tʃ/ > /ʃ/ > /θ/ en paralelo a /dʒ/ > /ʒ/ > /ð/, debido a la similitud perceptiva entre las sibilantes dentalizadas y las no sibilantes. El ensordecimiento se produciría después de llegar a /θ/ y /ð/, siguiendo la descripción de Menéndez Pidal (1985: 113). Su propuesta supone un cambio importante y muy interesante en el enfoque de la reestructuración del sistema de sibilantes. Sin embargo, no nos vamos a detener en ello aquí, dado que este trabajo no se ocupa directamente de la fonologización de las categorías /θ/ y /x/.

Pese a que actualmente se suele asumir que cualquier cambio implica la existencia de variación, la explicación tradicional del fenómeno suele simplificarse y manifestarse a partir de fases evolutivas más o menos cerradas. Como se sabe, la oposición de sonoridad solo era operativa en posición interna intervocálica: en posición inicial y de coda silábica únicamente era posible la concurrencia del fonema sordo. La excepción la constituían las alveopalatales /ʃ/ y /ʒ/, que también podían contrastar en posición inicial. Al trazar la evolución, algunos especialistas han sugerido un estadio intermedio en que las sibilantes en posición final prevocálica también pudieran ser sonoras antes de la compleción del proceso (Penny 1993: 80; Bradley & Delforge 2006: 22-23), tal como sucede en otras lenguas románicas, como el catalán o algunos dialectos italianos, o incluso en determinadas variedades de español (español ecuatoriano de las tierras altas, por ejemplo²).

Lo interesante es que, en general, se ha considerado que el ensordecimiento es un proceso que separa el español de las demás variedades románicas (Alarcos Llorach 1988: 51-52, Lloyd 1993: 428), un proceso atípico por cuanto se debería esperar más bien la conservación de la sonoridad en lugar de su pérdida, máxime en el contexto en que se produce (Lahoz 2015: 148-149, Davidson 2016: 37). Aun existiendo evoluciones análogas en otros sistemas, como en algunos dialectos conservadores de veneciano o del francoprovenzal (Mackenzie 2022: 1, 12), esta idea de peculiaridad del castellano frente a las dinámicas generales se ha impuesto y, por lo tanto, había que darle una explicación. Se han producido diversas aproximaciones, que abarcan desde el contacto lingüístico hasta razones de corte puramente fonético³. Las posturas principales se basan en tres supuestos: o bien relacionan la pérdida de sonoridad con la influencia del vasco, lengua que carece de sibilantes sonoras (Martinet 1951-1952, Lloyd 1993: 429-437), o bien con reajustes del sistema fonológico dirigidos a optimizar su rendimiento funcional, puesto que estas oposiciones no resultaban especialmente productivas (Alarcos Llorach 1988: 51-53; Penny 1993: 81-82, Ariza 2012: 224), o bien con alteraciones encaminadas a lograr una mayor simetría, ya que */ʃ/* no tenía un correlato sonoro (Contini 1951: 179-180).

Otros autores, en cambio, proponen explicaciones distintas. Pensado Ruiz (1993) rebate los argumentos de estos trabajos clásicos tanto a partir del estudio de la gramática comparada, como a partir de argumentos fonéticos y del estudio de documentación. Esta autora defiende que no hay que acudir ni al contacto de lenguas ni a una explicación teleológica del cambio para dar cuenta del proceso porque existen razones internas que proporcionan una justificación satisfactoria de la evolución. De hecho, indica que el español no es la única lengua en la que se produce ensordecimiento de sibilantes y que estos procesos suelen iniciarse en un par concreto de ellas, habitualmente las africadas palatales o dentoalveolares (las africadas tenderían a converger antes que las fricativas y las palatales antes que las alveolares). Antes que ella, también Alonso (1967: 313) apuntaba a las características fonéticas de las sibilantes como la causa última de la neutralización. En particular, al igual que Pensado (1993: 214-216), aludía a las condiciones aerodinámicas necesarias para mantener en el tiempo y de forma simultánea la estridencia propia de la sibilancia y la vibración de las cuerdas vocales. Se trata de condiciones muy específicas que exigen un control importante de la presión intraoral y de la subglótica, algo que no siempre es fácil de conseguir y equilibrar (Solé 2003, Ohala & Solé 2010: 39-41, Žygis, Fuchs & Koenig 2012: 310-312).

² Vid. Lipski (1989) para una descripción detallada.

³ Para una revisión bastante exhaustiva de las mismas, vid. Núñez-Méndez (2021: 16-19).

1.2. La sonoridad en las sibilantes: resultados previos

Como se comentaba, se ha dicho que el ensordecimiento de las sibilantes en posición interna intervocálica resultaba una evolución atípica, aunque no todos los autores coinciden en esta valoración (Pensado Ruiz 1993, Mackenzie 2022). De hecho, antes de proseguir, es importante conocer algo más sobre la sonoridad en las sibilantes, y no solo en español. De inicio, hay que tener en cuenta que, en las lenguas del mundo, las sibilantes sonoras no son los segmentos más frecuentes: suelen predominar las sordas. La razón es que, como se comentaba, las condiciones aerodinámicas necesarias para producirlas son más complejas en las primeras que en las segundas y esto facilita que los contrastes de sonoridad neutralicen a partir de procesos fonético-fonológicos de diversa índole, que pueden abarcar desde la fricativización (en el caso de africadas), la pérdida de fricción y la conversión en aproximante, vocalización y, por supuesto, el ensordecimiento (Ohala & Solé 2010: 53-54, Žygis, Fuchs & Koenig 2012: 308-309, Hualde & Prieto 2014: 111). Todos estos fenómenos implican, *de facto*, una disminución de los segmentos sibilantes sonoros en los inventarios de los distintos sistemas.

En lo que se refiere propiamente al ensordecimiento, la bibliografía suele señalar que afecta principalmente a las sibilantes en posición de coda, aunque tampoco es raro en posición inicial (Lavoie 2001: 27, 43, Wetzels & Mascaró 2001, Blevins 2004: 103-106, 110-111, Lahoz 2015: 142, 169-171). En efecto, se sabe que no se da en la misma medida en todos los contextos en que pueden aparecer las sibilantes: Haggard (1978) ya explicaba que las fricativas eran más proclives a ensordecerse si precedían una consonante oclusiva o en posición final de palabra. Lo interesante, sin embargo, es que este mismo autor también atestigua casos de ensordecimiento entre vocales dentro del límite de la palabra. Trabajos posteriores señalan que las sibilantes pueden perder la sonoridad en cualquier posición, pese a que, ciertamente, hay contextos que favorecen más que otros este fenómeno (es más probable ante consonante sorda y en posición prosódica final), pero puede darse en cualquiera (Smith 1997, Davidson 2016). Así pues, una evolución como la que sufrió el castellano tampoco resultaría inesperada.

Durante mucho tiempo se ha hablado de la sonoridad (y, en concreto, de la de las sibilantes), como si se tratara de un rasgo absoluto: o tenemos sibilantes sordas o sonoras. La realidad de la fonética es algo distinta, por lo que resulta imprescindible entender cómo se genera y qué rasgos permiten estudiarla. Obviamente, para producir un sonido sonoro, se necesita que las cuerdas vocales vibren, por lo que un proceso de ensordecimiento implica una reducción en su actividad. Sin embargo, tanto en su producción como en su percepción, intervienen otros parámetros importantes: la intensidad, el tono glotal, las transiciones del F1 (T1) y, muy especialmente, la duración del segmento (Widdison 1995: 38-39, Smith 1997: 473, Lavoie 2001: 107, Bradley & Delforge 2006: 31). Las sibilantes sordas son más largas y presentan mayor intensidad, tienen T1 más breves y no muestran signos de vibración glotal o, al menos, no de forma continua. Algunos estudios han demostrado que tampoco las sonoras presentan vibración de las cuerdas vocales continuada a lo largo de toda la duración del sonido: en realidad, parece que la mayor parte de las sibilantes sonoras se producen parcialmente ensordecidas (Haggard 1978, Pensado 1993: 216, Widdison 1995: 38, Smith 1997: 260, Ohala & Solé 2010: 71-73, Davidson 2016). Pese a ello, los oyentes suelen identificar como sonoros los casos que solo lo son parcialmente (Widdison 1995, 1997: 260, Lavoie 2001: 107): esto demostraría que la percepción de la sonoridad en las sibilantes no descansa únicamente en la actividad de las

cuerdas vocales sino en otros indicadores. De hecho, se ha comprobado que la duración es el principal parámetro para distinguir los segmentos sonoros de los sordos (Haggard 1978: 101, Lavoie 2001: 107, Žygis, Fuchs & Koenig 2012: 323-324).

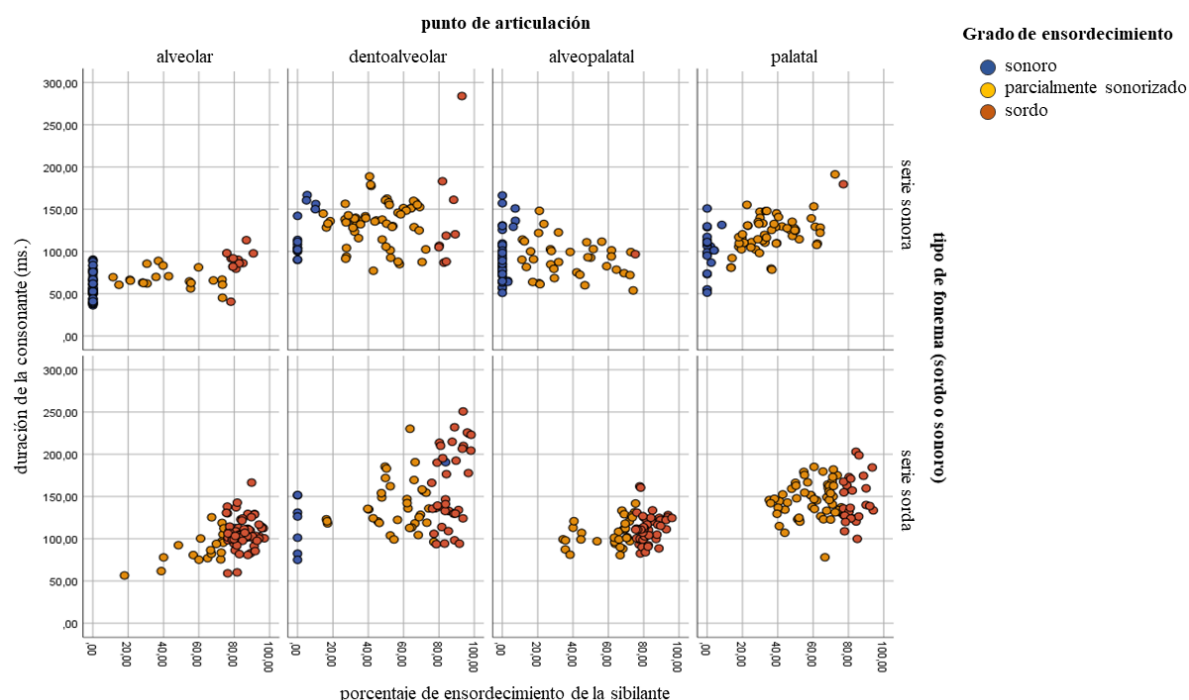
A partir de esta breve caracterización hay dos ideas principales que deben resaltarse: la primera es que la producción de las sibilantes sonoras no es uniforme y presenta variación, y la segunda, que el oyente juega un papel fundamental en la interpretación de la señal como sorda o sonora. De hecho, un estudio acústico para el catalán (Rost Bagudanch 2024), lengua que presenta un sistema de sibilantes similar al del castellano medieval, permite obtener datos relevantes sobre estas cuestiones. En este trabajo, se analizó la realización fonética de las sibilantes del catalán, es decir, de los pares /s/-/z/, /ʃ/-/ʒ/, /ts/-/d͡z/ y /tʃ/-/d͡ʒ/. Para ello, se tomaron como indicadores de sonoridad la duración del segmento y su grado de ensordecimiento⁴, el cual permitía la catalogación de los sonidos en sordos, parcialmente sonorizados y sonoros, siguiendo la clasificación propuesta por Smith (1997: 478)⁵. Pese a que los parámetros cuantitativos apuntan a diferencias claras entre los elementos sordos y sonoros de cada par (los correlatos sordos presentan mayor duración y mayor grado de ensordecimiento que los sonoros), en todas las categorías fonológicas sonoras se podían hallar alófonos parcialmente sonorizados y sordos, como puede advertirse en la figura 1. Es interesante comprobar que el par que mantiene una oposición más clara es el de las alveolares /s/-/z/, cuyas realizaciones fonéticas se encuentran más en los extremos. Sin embargo, la fricativa alveopalatal sonora /ʒ/, en hablantes de catalán central de Girona, tiende a producirse como parcialmente sonorizada, lo que podría apuntar a un ensordecimiento incipiente. Lo más interesante en lo que toca a esta investigación reside en los resultados obtenidos para los dos pares de africadas: en ambos, el grueso de las variantes alofónicas obtenido, independientemente de si corresponden a su miembro sordo o sonoro, es parcialmente sonorizado: 43,8 % para /ts/ y 70 % para /d͡z/, 67,5 % en el caso de /tʃ/ y 70 % en el de /d͡ʒ/. Se da, pues, un solapamiento entre las realizaciones de /ts/ y /d͡z/, por un lado, y de /tʃ/ y /d͡ʒ/, por el otro, que redundaría en la pérdida de sonoridad, menos evidente en el caso de las dentoalveolares, pero muy clara en las palatales.

De hecho, la sonorización que se invoca como dinámica esperable debe tomarse con una cierta precaución: en posición intervocálica se relaciona con procesos de tipo coarticulatorio sensibles a la variación temporal (Lahoz 2015: 148, Bradley, Morris & Jones 2023: 2). En efecto, Bradley, Morris & Jones (2023: 4-5), remitiendo a los estudios de Torreira & Ernestus (2012) y Hualde & Prieto (2014), explican que la sonorización de /s/ intervocálica en español, medida en términos de duración y de actividad de las cuerdas vocales, se constata más frecuentemente en posición final de palabra que en posición interna. Por lo tanto, el mantenimiento de la sordéz en esta posición no sería tan extraña e impropia.

⁴ Medido a partir de la función de Praat (Boersma & Weenink 2018) *Voice Report*, concretamente, el parámetro *Fraction of unvoiced frames* (aquí traducido libremente como grado de ensordecimiento), que evalúa la proporción de sordéz en un sonido seleccionado en términos de porcentaje respecto a su duración total.

⁵ Esta autora establece como sibilantes sonoras aquellas en que la falta de sonoridad no es superior al 10 % de la duración de la consonante, como parcialmente sonorizadas aquellas en que el ensordecimiento se encuentra entre el 11 y el 75 % del sonido, y como propiamente sordas las que no presentan sonoridad en más del 75 % del segmento.

Fig. 1. Distribución de las realizaciones sordas, parcialmente sonorizadas y sonoras para las sibilantes del catalán. La serie sonora está en la fila superior y la sorda, en la inferior (Rost Bagudanch 2024)



1.3. La percepción y el cambio fónico

Antes de proseguir, es importante señalar que aquí se parte de una concepción del cambio no teleológica, es decir, se entiende que no hay intencionalidad, por parte de los hablantes, en forjar un sistema más equilibrado, simétrico u óptimo, sino que las posibles mejoras en el sistema fonológico son consecuencia del proceso, no su causa, siguiendo las teorías generalmente aceptadas en la actualidad (*cfr.* Ohala 1993, 2012, Pierrehumbert 2001, 2002, Blevins 2004, Beddor 2012, Goldrick & Cole 2023). Esto implica entender que los cambios fónicos han de tener una causa fonética, tal como ya indicaban los neogramáticos. Asimismo, se asume que el origen fonético del cambio puede tener su raíz en cuestiones de producción del sonido, que es la perspectiva que tradicionalmente ha sido objeto de atención, pero también puede explicarse por razones perceptivas, tal como postulan autores como Ohala (1981, 2012), Pierrehumbert (2001) o Beddor (2012, 2023). Obviamente, en su difusión pueden intervenir factores extralingüísticos, aunque, atendiendo a investigaciones recientes, no debe olvidarse que un fenómeno tan común como la acomodación (o convergencia) al habla del interlocutor también puede llevar a cambios (Pardo 2006, Goldrick & Cole 2023: 4).

Resulta, pues, evidente que producción y percepción están íntimamente relacionadas y, muchas veces, es complicado poder establecer de forma clara cuál es realmente la fuente primera de las alteraciones que pueden llevar a reestructuraciones en el sistema fonológico (Beddor 2023: 10). Pese a ello, lo que parece claro a partir de los trabajos anteriormente citados es que el receptor de la señal tiene un papel

crucial en cualquier proceso de cambio. De hecho, Beddor (2023: 11) afirma que, habitualmente, y especialmente en el nivel de la comunidad, «change in perception tends to be ahead of change in production» en casos de desajustes entre producción y percepción. Estos desajustes son bastante habituales en procesos de neutralización como los que se trata en este trabajo (Labov, Karen & Miller 1991, Labov 1994: 368, Hay, Warren & Drager 2006), por lo que parece lógico pensar que la percepción de las sibilantes sonoras, aparte de su producción, pueda tener un papel clave en la explicación del ensordecimiento general de este tipo de sonidos.

Desde este punto de vista, y atendiendo a los resultados obtenidos en experimentos acústicos previos (Rost Bagudanch 2022, en prensa), parece sensato acogerse a modelos teóricos que recojan como elemento indispensable del cambio la existencia de variación y cómo esta puede influir en la percepción para dar lugar a soluciones innovadoras cuando el receptor pasa a actuar como emisor, como la teoría de ejemplares (Pierrehumbert 2001, 2002). Así, a nivel de producción, se habría dado variación en la realización fonética (acústica) de las sibilantes: los hablantes habrían ido interiorizando los distintos alófonos para cada una de las categorías, con sus diferencias en términos de sonoridad, a partir de su experiencia como oyentes. Estas nubes de ejemplares se construirían alrededor de una forma canónica, prototípica, que podría variar en función del *input* recibido: si los estímulos captados encajan con la forma prototípica, esta se mantendría como centro alrededor del que se organizan todos los ejemplares de esa categoría; sin embargo, si los estímulos percibidos pasan a tener mayoritariamente otras características, el ejemplar que hasta el momento había sido considerado canónico dejaría de serlo y ocuparía su lugar algún otro situado, hasta entonces, más en la periferia de esa nube. Su centro se desplazaría. Este sería el mecanismo que explicaría el cambio. Si, además, este cambio implicara el solapamiento con los ejemplares correspondientes a otra categoría fónica, se podría desencadenar un proceso de neutralización, en función del rendimiento funcional que pudiera tener mantener o eliminar la oposición existente (Pierrehumbert 2001, 2002, Beddor 2012, 2023, Wedel, Kaplan & Jackson 2013).

Así pues, es importante dedicar espacio para lidiar con el aspecto perceptivo de la neutralización de sonoridad en las sibilantes, tal como han hecho autores como Widdison (1995, 1997), quien ya había hecho una contribución en este terreno. El presente trabajo, pues, pretende ahondar en este aspecto, máxime teniendo en cuenta que los oyentes tienen acceso a una variación acústica importante y que esta puede condicionar el procesamiento de la señal (Pierrehumbert 2001, 2002, Beddor 2012, 2023, Kapadia, Tin & Perrachione 2023).

2. Diseño experimental

Teniendo en cuenta el objetivo central de este trabajo y tal como se ha comentado ya, se ha optado por refinar el análisis que se había iniciado en Rost Bagudanch (2022), por lo que se ha partido de los materiales de una prueba de percepción previa. Como se trata de evaluar el papel que juega la audición en el proceso de ensordecimiento de las sibilantes y comprobar hasta qué punto cada uno de los miembros de los pares de sibilantes del catalán se reconoce como sordo o sonoro, se ha llevado a cabo un test de identificación de respuesta cerrada.

Los estímulos respondían a secuencias formadas por una sibilante en posición intervocálica (del tipo [v_v]), correspondientes en su mayoría a fragmentos en interior de palabra que no coincidían con voces reconocibles.⁶ Estas secuencias fueron extraídas de las grabaciones de habla espontánea empleadas en el estudio de Rost Bagudanch (2022). Se seleccionaron los ejemplos de las distintas sibilantes en que la calidad acústica era óptima (sin ruido de fondo, con volumen adecuado y con rasgos acústicos claros en el espectrograma). Se decidió utilizar estímulos naturales, precisamente por el tipo de investigación que se lleva a cabo: el cambio fónico se produce en situación de habla natural, no de laboratorio, y por este motivo se estimó conveniente equiparar en lo posible estas condiciones, dentro de las limitaciones de un experimento de este tipo⁷. Una vez hecha la selección, se obtuvieron 8 casos útiles de [s], 10 de [z], 8 de [ʃ], 10 de [ʒ], 6 de [ts], 14 de [dz], 10 de [tʃ] y otros 10 de [dʒ], equilibrados entre sílaba tónica y átona. La excepción es el caso de la africada dentoalveolar sorda, de la que no se obtuvieron ítems en posición interna intervocálica. Tal como se advertía ya en Rost Bagudanch (2022), se trata de una sibilante con una frecuencia de aparición en interior de palabra extraordinariamente baja (un 0,0195 %, atendiendo a Rafel i Fontanals 1980: 480-481), de modo que se optó por recurrir a casos en que la africada se encontraba en posición intervocálica, pero con límite de palabra entre la vocal precedente y la consonante. Se desecharon aquellos casos en que se podía intuir alguna pausa entre estos dos segmentos que pudiera delatar la posición inicial original de la sibilante. Este es el motivo por el que el número de estímulos para [ts] es más bajo y que todos ellos se den en sílaba tónica, puesto que en contexto inacentuado la impresión de silencio entre vocal y sibilante era evidente. Como puede verse en la tabla 1, el número de estímulos fue de 76. Dado que cada uno de ellos se repetía 3 veces a lo largo de la prueba, el volumen total de ítems que la constituyen asciende a 228.

Tabla 1. Número de estímulos utilizado en la prueba de percepción

		N.º estímulos	x 3 repeticiones
Fricativas alveolares	[s]	8	24
	[z]	10	30
Fricativas alveopalatales	[ʃ]	8	24
	[ʒ]	10	30
Africadas dentoalveolares	[ts]	6	18
	[dz]	14	42
Africadas palatales	[tʃ]	10	30
	[dʒ]	10	30
Total		76	228

⁶ Cfr. Rost Bagudanch (2022: 25).

⁷ Se procuró que los estímulos presentaran variación en cuanto a los indicadores de sonoridad más robustos en Rost Bagudanch (2022), a saber, duración y grado de ensordecimiento, para abarcar al máximo el rango de variabilidad posible en la producción. Sin embargo, los resultados de ese mismo trabajo ya indicaron que ninguno de estos dos parámetros incide en las respuestas obtenidas.

El test se elaboró y distribuyó con la plataforma FOLERPA (Fernández Rei 2021), lo que permitió su distribución en línea. Los estímulos se presentaron en orden aleatorio. El silencio entre cada uno de ellos era de 1,5 segundos.

La prueba se administró a 26 jueces catalanohablantes dominantes en esta lengua⁸. Ninguno de ellos presentaba ningún problema de habla o de audición. Debido a la naturaleza del soporte elegido, los participantes efectuaron la prueba desde sus domicilios, para lo que recibieron instrucciones previas precisas sobre las condiciones en que la debían contestar: en una habitación silenciosa y utilizando obligatoriamente auriculares. Después de escuchar dos veces el estímulo, habían de responder a la pregunta «¿qué secuencia oyes?» y elegir entre las dos posibles representaciones ortográficas del ítem que se les presentaba. Para poder pasar al estímulo siguiente era indispensable proporcionar una respuesta. La prueba se dividía en 4 partes, después de cada una de las cuales se podía descansar. En total, la encuesta completa llevaba 30 minutos aproximadamente.

En el análisis de los datos se han tomado en consideración 4 variables dependientes, una de tipo cualitativo, la respuesta obtenida en la identificación de cada estímulo, y tres cuantitativas, propias del análisis según la TDS (Abdi 2007: 6, Marrero 2014: 519-529). La primera de ellas es la distancia perceptiva entre los miembros sordo y sonoro de cada par; es decir, la proximidad perceptiva relativa entre los dos miembros de cada oposición fonológica (cuanto menor sea este valor, más cercanos resultan perceptivamente, y mayor posibilidad de que puedan confundirse). En segundo lugar, se estudia el cálculo de d' , que permite testar la sensibilidad del juez a la señal (valores mayores de 1 en este parámetro indican una sensibilidad importante al estímulo). Finalmente, interesa el cálculo de C , un parámetro que permite determinar la estrategia de los participantes al contestar; es decir, si los jueces son más o menos conservadores en sus respuestas (valores cercanos a 0 apuntan a un observador ideal, objetivo y sin sesgos hacia errores o falsas alarmas; si los valores son negativos, el juez en cuestión es más liberal y tiende a las falsas alarmas, si son positivos, resulta más conservador). Los cálculos de los valores de d' y de C se han realizado a partir del software desarrollado por Serrano-Pedraza y disponible en la página web del *Laboratory of Visual Psychophysics* de la Universidad Complutense de Madrid⁹. Los de distancia perceptiva se han hecho siguiendo la fórmula aportada por Johnson (2003: 68).¹⁰ Por otra parte, como factores, se contemplaron el tipo de sibilante (cada una de las categorías fonológicas), su lugar de articulación (alveolar, alveopalatal, dentoalveolar y palatal) y el acento (sílabas tónicas o sílabas átonas)¹¹.

⁸ Para controlar que su L1 fuera, efectivamente, el catalán, se les administró el cuestionario *Bilingual Language Profile* (Birdsong, Gertken & Amengual 2012). Resultados positivos indicaban dominancia en catalán sobre el español y los valores obtenidos para cada participante encajan en este perfil. El promedio es de 57,034. Si se tiene en cuenta que valores cercanos a 0 responden a un bilingüismo perfecto y que entre -110 y 110 apuntan a un bilingüismo más o menos equilibrado, los jueces pueden ser catalogados como bilingües equilibrados, aunque se registra variación en función del participante ($sd=138.47$). En cualquier caso, esta variación oscila entre un bilingüismo equilibrado y el sesgo a favor del catalán.

⁹ <https://www.ucm.es/serranopedrazalab/software-1> (consultado 29/9/2023).

¹⁰ El cálculo de la distancia perceptiva requiere, primero, el de la similitud entre dos segmentos:

$$\text{similitud} = \frac{(\text{errores de AxB}) + (\text{errores de BxA})}{(\text{aciertos AxA}) + (\text{aciertos BxB})}$$

Una vez se obtiene este valor, la distancia perceptiva entre ellos es el negativo del logaritmo natural de la similitud: $-(\ln(\text{similitud de A y B}))$. Vid. Johnson (2003: 68) o Marrero (2014: 527-529) para información más detallada.

¹¹ En un acercamiento piloto previo, dado que las grabaciones pertenecían a locutores de dos variedades dialectales diferentes del catalán (central y mallorquín), se investigó si el dialecto podía influir en los resultados. El resultado fue negativo en todos los casos, lo cual parece lógico por cuanto el sistema de sibilantes es, en esencia, el mismo en las dos variedades. Por este motivo esta variable no se ha tenido en cuenta aquí.

El tratamiento estadístico de los datos obtenidos se ha efectuado con el programa SPSS (v. 25), empleando regresiones logísticas binomiales con efectos mixtos en el caso de la variable cualitativa (la respuesta obtenida) y modelos lineales generalizados con efectos mixtos con las variables de tipo cuantitativo (distancia perceptiva, d' y C). En el caso de la regresión logística, los efectos principales han sido el tipo de sibilante y el acento, con interacción entre ambos. Para los modelos lineales generalizados, se han establecido como efectos principales el tipo de sibilante y el lugar de articulación¹². En todos los casos, el juez y el estímulo se han considerado efectos aleatorios. El umbral de significación se ha establecido en 0,05.

3. Resultados

Se van a tratar los resultados obtenidos de forma diferenciada en función de las variables dependientes. En primer lugar, se van a ofrecer los datos descriptivos generales, que se corresponden con la evaluación por tipo de sibilante (§ 3.1). En segundo lugar, se van a presentar los propios del análisis perceptivo a partir de los valores de distancia perceptiva, d' y C (§ 3.2).

3.1. Resultados descriptivos generales por tipo de sibilante

Los resultados generales teniendo en cuenta las respuestas de los jueces ante los estímulos presentados apuntan a que, globalmente, el rasgo de sonoridad se identifica correctamente en los distintos tipos de sibilante, como se desprende de los datos de la tabla 2 y la figura 2. Así, las respuestas correctas suponen, de promedio, el 72,32 % del total, pero es importante tener en cuenta que hay confusiones entre estímulos sordos y sonoros en todas y cada una de las categorías: el volumen de errores asciende, globalmente, a un 27,67 % de los ítems, una cantidad nada desdeñable que oscila notablemente dependiendo del tipo de sibilante. Cabe señalar que las sordas tienden a presentar menor volumen de confusiones que las sonoras, salvo en el caso de las dentoalveolares.

Esta descripción general, sin embargo, debe ser matizada con una aproximación más detallada, puesto que, como se decía, se advierten diferencias significativas entre los distintos tipos de sibilante ($F(7, 5757)=14,718$, $p=0,0001$). A partir de los datos obtenidos, se observa que las fricativas presentan un comportamiento bastante homogéneo. De hecho, se reconocen mayoritariamente como se esperaba; en ellas, la tasa de aciertos supera en todos los casos el 75 % (mucho más en el caso de las alveopalatales), y la de errores no baja del 12 %, lo cual también es indicativo de que las confusiones no son extrañas. No obstante, es importante advertir que el miembro sordo de cada par se reconoce significativamente mejor que el sonoro,¹³ lo que apunta a una mayor robustez en los índices de sonoridad.

¹² No se ha incluido el acento porque en Rost Bagudanch (2022) se demostró que no afectaba a las respuestas.

¹³ Se han practicado modelos de regresión logística binomial para comprobar el comportamiento de cada par de sibilantes de forma separada (respuesta obtenida x respuesta esperada). Tanto en el caso de las sibilantes alveolares como en el de las alveopalatales, se tiende a confundir el estímulo sonoro con el sordo en mayor medida que a la inversa ($b_0=2,864$, $se=0,814$, $t=3,519$, $p<0,0001$ y $b_0=6,322$, $se=1,100$, $t=5,747$, $p<0,0001$, respectivamente).

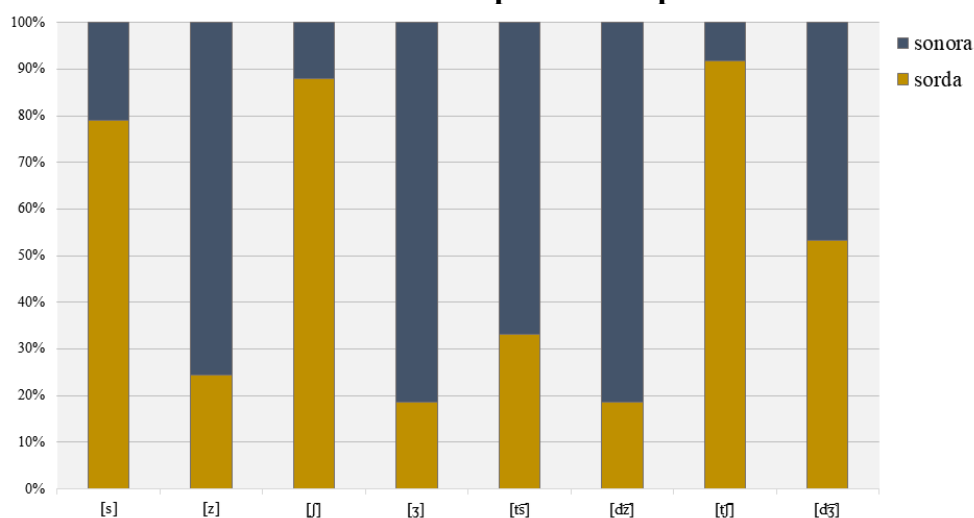
En el caso de las africadas, el panorama es distinto por cuanto se dan situaciones más extremas: en dos categorías, $[\text{ts}]$ y $[\text{dʒ}]$, el volumen de errores se incrementa (66,9 % en el primer caso y 53,3 % en el segundo), lo que es estadísticamente relevante, por cuanto $[\text{ts}]$ muestra más tendencia a interpretarse como sonora que como sorda ($\text{est}=0,284$, $\text{sd}=0,089$, $t=3,175$, $p=0,002$) y $[\text{dʒ}]$, más propensión de lo esperado a identificarse como sorda ($\text{est}=-0,430$, $\text{sd}=0,093$, $t=-4,619$, $p<0,0001$). Por su parte, en $[\text{dz}]$ y $[\text{tʃ}]$, el volumen de errores se reduce por debajo del 10 % (aciertos por encima del 80 %), lo que parece apuntar a un reconocimiento más claro incluso que en las fricativas. Estas dos últimas sibilantes, pues, parecen más robustas en lo que se refiere a la detección de la sonoridad.

Tabla 2. Resultados de la prueba de percepción, teniendo en cuenta la respuesta obtenida en función del tipo de sibilante y del acento.

Se ofrece la frecuencia absoluta y relativa para cada sibilante

		[s]	[z]	[ʃ]	[ʒ]	[ts]	[dz]	[tʃ]	[dʒ]
global	sordo	493 (79 %)	141 (22,6 %)	549 (88 %)	145 (18,6 %)	155 (33,1 %)	204 (18,7 %)	715 (91,7 %)	416 (53,3 %)
	sonoro	131 (21 %)	438 (77,4 %)	75 (12 %)	635 (81,4 %)	313 (66,9 %)	888 (81,3 %)	65 (8,3 %)	364 (46,7 %)
síl. tónica	sordo	261 (83,7 %)	67 (21,5 %)	249 (79,8 %)	109 (27,9 %)	155 (33,1 %)	56 (10,3 %)	325 (83,3 %)	26 (6,7 %)
	sonoro	51 (16,3 %)	245 (78,5 %)	36 (20,2 %)	281 (72,1 %)	313 (66,9 %)	490 (89,7 %)	65 (16,7 %)	364 (93,3 %)
síl. átona	sordo	232 (74,4 %)	74 (23,7 %)	300 (96,2 %)	36 (9,2 %)	--	148 (27,1 %)	390 (100 %)	390 (100 %)
	sonoro	80 (25,6 %)	238 (76,3 %)	12 (3,8 %)	354 (90,8 %)	--	398 (72,9 %)	--	--

Figura 3. Gráfico de barras que muestra la proporción de identificaciones de los estímulos como sordo o sonoro para cada tipo de sibilante



Al desglosar los efectos, se puede detallar algo más esta descripción global. En general, las sibilantes sordas presentan un grado de reconocimiento bastante homogéneo, que permite agruparlas perceptivamente, en términos de sonoridad: [s], [ʃ] o [tʃ] se identifican como sordas sin muchos problemas. El matiz estriba en que en la africada palatal se muestra significativamente más sólida que las demás sordas y, por este motivo, se distingue de ellas¹⁴. Algo parecido sucede con las sibilantes sonoras, que también exhiben un nivel de aciertos similar entre ellas: la tasa de identificaciones correctas no arroja diferencias relevantes entre [z], [ʒ] y [dʒ]. Como se decía, lo interesante es que sí hay diferencias entre los pares sordo y sonoro, puesto que la identificación exitosa de las sonoras es significativamente menor que en las sordas en los pares alveolar (est=-0,626, sd=0,091, t=-6,882, p<0,0001) y alveopalatal (est=-0,797, sd=0,054, t=-14,879, p<0,0001); es decir, en las fricativas.

Como se decía, los casos de [ts] y [dʒ] resultan distintos. Los estímulos de [ts] se comportan de forma parecida a los de [z], [ʒ], [dʒ]; dicho de otro modo, no presentan divergencias significativas respecto a los segmentos sonoros en términos de reconocimiento de sonoridad.¹⁵ En pocas palabras, se confunden sistemáticamente con el elemento sonoro del par dentoalveolar. En [dʒ] sucede lo contrario: sus resultados son equivalentes a los obtenidos por [s] y [ʃ]¹⁶ lo que indica que se identifica con el segmento sordo de la oposición palatal.

Aunque el acento, como efecto fijo, no parece incidir en los resultados obtenidos ni parecen darse interacciones a nivel global (F(1,5757)=0,007, p=935 y F(6,5757)=1.501, p=0,173), se advierte una tasa de aciertos superior en sílaba tónica (77 % frente a 72 % en sílaba átona). Al desglosar los resultados, se observan algunos comportamientos interesantes, principalmente en lo referente a [ts] y [dʒ]. Antes de seguir, hay que tener en cuenta que los estímulos de la dentoalveolar sorda solo pudieron conseguirse en sílaba tónica, por lo que no se ha podido llevar a cabo una correcta comparación. Sin embargo, en este contexto queda claro que [ts] se confunde con un elemento sonoro, tal como ya se ha comentado anteriormente. El caso de [dʒ] es particularmente llamativo, por cuanto no se desvía de lo esperado en contexto acentuado (los estímulos se reconocen puntualmente como sonoros de forma similar al resto de sibilantes sonoras), pero sí lo hace en contexto inacentuado. En sílaba átona se confunde siempre con [ʃ] (est=0,0001, sd=0,000, t=0,000, p=1,000), sin vacilaciones.

3.2. Parámetros perceptivos

A partir de estos datos descriptivos generales, los cálculos de distancia perceptiva, d' y C permiten afinar más el análisis. En la tabla 3 se muestran los valores obtenidos para estos parámetros en función del tipo de sibilante y, en la figura 3, se ofrecen las curvas ROC obtenidas para cada uno de los pares de sibilantes, que proporcionan información general de los resultados. En ellas, el eje horizontal representa la tasa de falsas alarmas, mientras que el vertical, la de aciertos. Las curvas muestran la tendencia de las respuestas o bien hacia los aciertos o bien hacia las falsas alarmas: cuanto más a la izquierda, mayor tasa de aciertos; cuanto más a la derecha, mayor tendencia a identificar erróneamente la señal. La zona media,

¹⁴ (est=0,163, sd=0,060, t=2,718, p=0,007) respecto a [s] y (est=0,069, sd=0,031, t=2,231, p=0,026) frente a [ʃ].

¹⁵ (est=-0,074, sd=0,121, t=-0,613, p=0,540) en comparación con [z], (est=-0,152, sd=0,109, t=-1,389, p=0,165) respecto a [ʒ] y (est=-0,131, sd=0,108, t=-1,212, p=0,225) respecto a [dʒ].

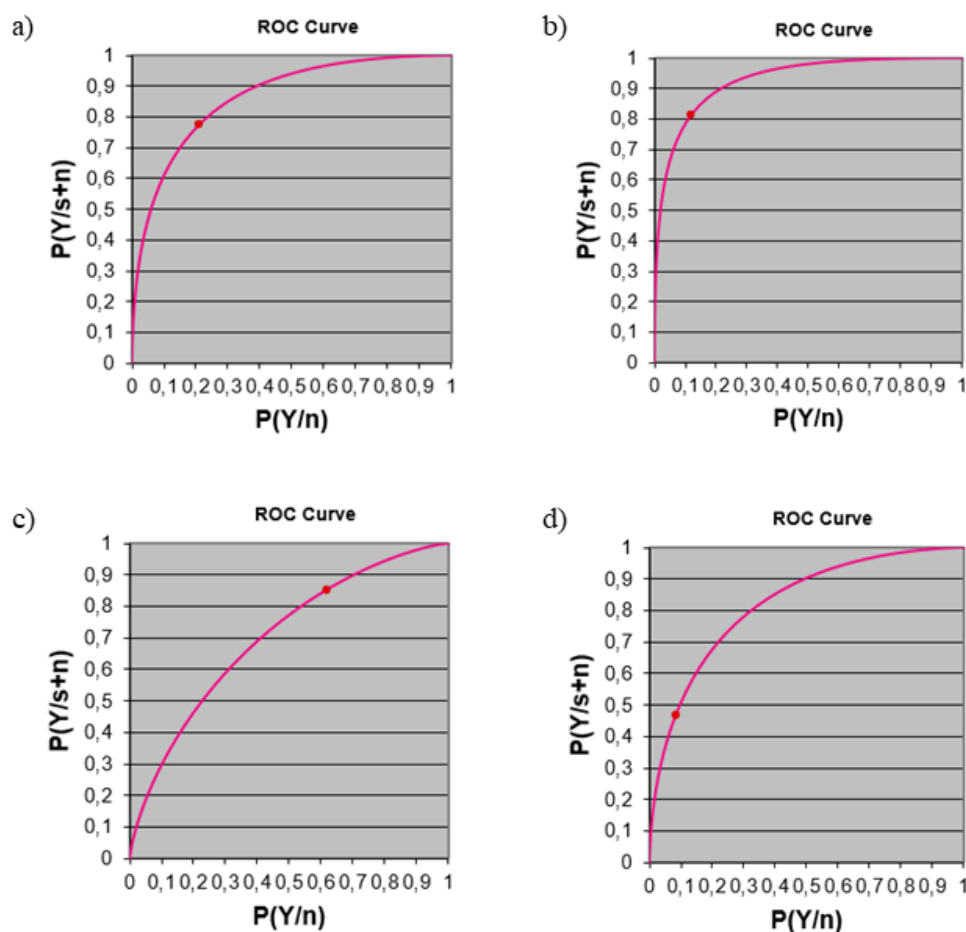
¹⁶ (est=-0,162, sd=0,060, t=-1,357, p=0,175) respecto a [s] y (est=0,001, sd=0,104, t=0,006, p=0,996) respecto a [ʃ].

que se corresponde con la diagonal que dividiría el recuadro en dos, señala el límite de aciertos por azar. También es importante fijarse en la altura de la curva: cuanto más alta, mayor sensibilidad por parte del sujeto o mayor detectabilidad de la señal (Marrero 2014: 533-535).

Tabla 4. Valores de distancia perceptiva, d' y C para cada par de sibilantes

	Sibilantes alveolares	Sibilantes alveopalatales	Sibilantes dentoalveolares	Sibilantes palatales
Distancia perceptiva	1,280	1,687	0,791	0,803
d'	1,558	2,067	0,747	1,299
C	0,027	0,140	-0,671	0,733

Figura 2. Curvas ROC que ilustran la distancia perceptiva entre los miembros de cada uno de los pares de sibilantes: (a) se corresponde con el par alveolar; (b), con el alveopalatal; (c), con el dentoalveolar y (d), con el palatal. El punto indica la sensibilidad de los sujetos o la detectabilidad de la señal



Como se puede observar, en los estímulos correspondientes a sibilantes alveolares y alveopalatales, la curva muestra un mayor sesgo hacia los aciertos, mientras que en el caso de las dentoalveolares y las palatales, lo presenta hacia las falsas alarmas y se aproxima más a la parte media del recuadro, es decir, a las respuestas por azar, especialmente en lo que se refiere a las dentoalveolares. En estos dos últimos casos, además, la curva no es tan alta, lo que apunta a una menor detectabilidad de la señal o a una menor sensibilidad por parte de los jueces. Estas primeras impresiones se corroboran a partir del estudio de los parámetros de distancia perceptiva, d' y C , que se desglosan a continuación.

3.2.1. La distancia perceptiva

En lo que se refiere al primero de estos parámetros, la distancia perceptiva entre sordas y sonoras de cada par varía significativamente en función de su punto de articulación ($F(3, 100)=22,751, p < 0,0001$): resulta mucho mayor en las alveopalatales, seguidas de las alveolares y, a mayor distancia, de las palatales y de las dentoalveolares, en una progresión alveopalatales > alveolares > palatales > dentoalveolares (cfr. tabla 3). En todo caso, se detectan diferencias relevantes entre las sibilantes alveolares ($b_0=0,696, se=0,139, t=4,993, p<0,0001$), entre las alveopalatales ($b_0=0,886, se=0,193, t=6,359, p<0,0001$) y entre las palatales ($b_0=0,818, se=0,119, t=6,872, p<0,0001$), pero no entre las dentoalveolares.

Al analizar de forma global este parámetro comparando todos los pares, se comprueba que hay diferencias importantes entre ellos, salvo en lo referente a las fricativas, que no muestran divergencias estadísticamente significativas entre ellas: los pares /s/-/z/ y /j/-/ʒ/ se comportan de forma análoga; es decir, la distancia perceptiva entre sus miembros es similar en ambos, algo que no ocurre si se comparan respecto a los miembros de los pares de sibilantes africadas.

De hecho, la distancia perceptiva entre los segmentos de la oposición alveolar es significativamente mayor que entre los de la dentoalveolar ($est=0,719, sd=0,139, t=5,162, p<0,0001$) y los de la palatal ($est=0,696, sd=0,139, t=4,993, p<0,0001$). Lo mismo se da entre los elementos del par alveopalatal respecto los del dentoalveolar ($est=0,910, sd=0,139, t=6,527, p<0,0001$) y del palatal ($est=0,886, sd=0,139, t=6,359, p<0,0001$). Finalmente, es interesante que la distancia perceptiva entre los fonos del par dentoalveolar y del palatal sea significativamente más pequeña. Estos resultados implican que, por una parte, la distinción de los segmentos sordo y sonoro de las sibilantes alveolares y alveopalatales es más nítida que en los demás pares y, por otra, que la discriminación entre dentoalveolares y entre palatales es mucho menos evidente: están más próximas en el espacio auditivo.

3.2.2. La sensibilidad a la señal: d'

En cuanto a la sensibilidad del sujeto ante el estímulo, hay que tener en cuenta que se considera importante si el valor de d' es mayor que 1. Según nuestros datos, esto ocurre en todos los tipos de sibilantes excepto en las dentoalveolares (cfr. tabla 3): 1,558 en las fricativas alveolares, 2,067 en las fricativas alveopalatales y 1,299 en las africadas palatales frente a 0,747 en las africadas dentoalveolares. De hecho, el

modelo estadístico revela que el punto de articulación de las sibilantes condiciona estos resultados ($F(3, 81)=19,416$, $p<0,0001$): d' presenta valores significativamente más altos en el caso de las sibilantes alveopalatales ($b_0=0,679$, $se=0,169$, $t=4,012$, $p<0,0001$) y palatales ($b_0=1,384$, $se=0,134$, $t=10,362$, $p<0,0001$), aunque también es estadísticamente relevante que el de las dentoalveolares sea mucho más bajo ($b_0=-0,649$, $se=0,163$, $t=-3,977$, $p<0,0001$).

Al analizar los datos de forma detallada, se advierte la existencia de divergencias importantes en el grado de sensibilidad a la señal en función de los distintos grupos de sibilantes. En primer lugar, las alveolares se distinguen de las alveopalatales y de las dentoalveolares: presentan un valor significativamente más bajo que las alveopalatales ($est=-0,564$, $sd=0,179$, $t=-3,155$, $p=0,002$) y más alto que las dentoalveolares ($est=0,763$, $sd=0,171$, $t=4,469$, $p<0,0001$). Estas últimas, a su vez, se diferencian de las alveopalatales y de las palatales, frente a quienes exhiben una d' mucho más baja ($est=-1,327$, $sd=0,176$, $t=-7,556$, $p<0,0001$ y $est=-0,649$, $sd=0,163$, $t=-3,977$, $p<0,0001$, respectivamente). Cabe notar que no hay diferencias relevantes entre alveopalatales y palatales ($est=0,114$, $sd=0,168$, $t=0,679$, $p=0,499$). Es importante tener en cuenta que el juez, como factor aleatorio, incide en los resultados obtenidos ($est=0,137$, $sd=0,068$, $Z=2,016$, $p=0,044$).

Estos datos indican que, en general, los jueces manifiestan una sensibilidad a la señal importante en prácticamente todos los tipos de sibilantes, salvo en las dentoalveolares. Esto puede facilitar la confusión entre estímulos de esta última clase. Ante las alveopalatales se muestra un grado de sensibilidad mayor, lo que podría ser una de las claves de por qué se discriminan tan bien. Ante alveolares y palatales, por su parte, se dan unos valores de d' estadísticamente similares y superiores a 1, lo que también apunta a una sensibilidad importante.

3.2.3. La estrategia de respuesta: C

Respecto a C, que mide la estrategia de los participantes al responder, deben destacarse algunos puntos. La proximidad del valor a 0 marca el observador ideal, sin sesgos en sus elecciones. Esto se da en el caso de las sibilantes alveolares ($C=0,027$), en los demás, en cambio, hay un sesgo hacia uno u otro extremo. En el caso de las sibilantes alveopalatales ($C=0,140$) y palatales ($C=0,733$), los jueces adoptan una estrategia conservadora; es decir, identifican como sordos estímulos que en realidad son sonoros. Por el contrario, con las sibilantes dentoalveolares ($C=-0,671$), su actitud es liberal, puesto que tienden a identificar como sonoros estímulos que no lo son. Los valores más alejados de 0 y que, en consecuencia, denotan un mayor sesgo en las respuestas corresponden a las sibilantes africadas.

Nuevamente, el modelo estadístico muestra que el punto de articulación del par de sibilantes incide en el valor de C ($F(3, 81)=118,493$, $p<0,0001$). Este es significativamente más cercano al observador ideal (0) en el caso de las alveolares ($b_0=-0,743$, $sd=0,080$, $t=-9,246$, $p<0,0001$), alveopalatales ($b_0=-0,758$, $sd=0,081$, $t=-9,354$, $p<0,0001$) y dentoalveolares ($b_0=-1,469$, $sd=0,078$, $t=-18,825$, $p<0,0001$); en cambio, en las palatales resulta más alejado de él ($b_0=0,773$, $sd=0,059$, $t=13,187$, $p<0,0001$).

Si consideramos las relaciones entre los tipos de sibilante, se advierte que no hay diferencias relevantes en la estrategia de respuesta entre las alveolares y las alveopalatales, pero sí la hay entre los demás grupos. Las alveolares se distinguen netamente de las dentoalveolares ($est=0,726$, $sd=0,082$, $t=8,888$, $p<0,0001$) y de las palatales ($est=-0,743$, $sd=0,080$, $t=-9,246$, $p<0,0001$); mientras que las alveopalatales lo hacen también respecto a los dos tipos africadas ($est=0,711$, $sd=0,084$, $t=8,503$, $p<0,0001$ en el caso de la comparación con dentoalveolares y $est=-0,758$, $sd=0,081$, $t=-9,354$, $p<0,0001$ en el de las palatales). Asimismo, también los valores de C resultan significativamente distintos entre dentoalveolares y palatales ($est=-1,469$, $sd=0,078$, $t=-18,825$, $p<0,0001$).

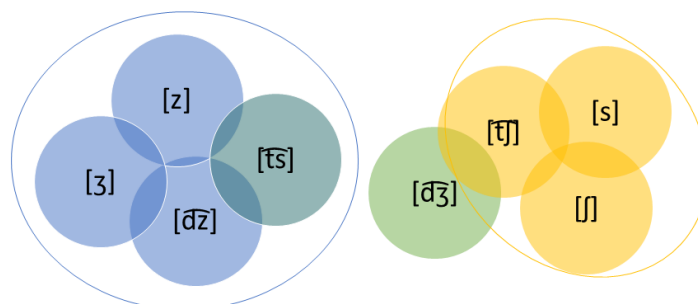
Estos resultados confirman las observaciones anteriores: en los casos en que el valor de C es más cercano a 0 (alveolares y alveopalatales, es decir, en las fricativas), el comportamiento de los jueces se puede considerar homogéneo, sin sesgos importantes, aunque su estrategia tiende a ser conservadora (identificar estímulos sonoros como sordos). Esto es así de un modo mucho más claro en el caso de estímulos correspondientes a africadas palatales. En cambio, los jueces son significativamente más liberales (reconocen como sonoros estímulos que son sordos) ante ejemplos correspondientes a africadas dentoalveolares. En cualquier caso, parece que las mayores dificultades se hallan en estos dos últimos tipos de sibilante.

4. Discusión

En este trabajo se ha intentado ahondar en el aspecto perceptivo de la desfonologización de sonoridad en las sibilantes del español. El recurso a la gramática comparada, en este caso, al estudio de las sibilantes del catalán, que presenta un sistema muy similar al del español medieval, permite reproducir las condiciones fonéticas en que se debió de dar la pérdida de los correlatos sonoros de estas consonantes (Widdison 1997: 254, Blevins 2004, Harrington 2012: 322). Aquí interesa particularmente la parte perceptiva, puesto que se ha venido advirtiendo que, en fenómenos de neutralización de una oposición fonológica, producción y percepción no siempre van a la par, sino que pueden presentar asimetrías a lo largo del proceso. Habitualmente, estas asimetrías suelen concretarse en que la pérdida de la oposición avanza más rápidamente a nivel auditivo que en el articulatorio y acústico (Labov 1994: 368, Beddor 2023: 11).

Los resultados obtenidos en el análisis de las respuestas a la prueba de percepción administrada demuestran que, mayoritariamente, los jueces son capaces de diferenciar razonablemente bien entre los segmentos sordo y sonoro de los pares de sibilantes estudiados (a saber, $/s/-/z/$, $/ʃ/-/ʒ/$, $/ts/-/dʒ/$, $/tʃ/-/dʒ/$). Pese a ello, se registran confusiones en todos los casos y, en general, de forma más pronunciada en el elemento sonoro, que tiende a ser identificado erróneamente como sordo en mayor número de casos que a la inversa. Asimismo, se ha comprobado que la capacidad de discriminación entre las consonantes que constituyen cada par no es igual en todas las sibilantes. La información que aportan estos errores, de hecho, coincide con lo pronosticado por Pensado Ruiz (1993: 210) sobre la progresión del ensordecimiento: se produce más en las africadas que en las fricativas y de forma más clara en las dentoalveolares y palatales que en las demás. Esto implica, por lo tanto, la existencia de dificultades en la percepción de la sonoridad en estos dos pares de sibilantes.

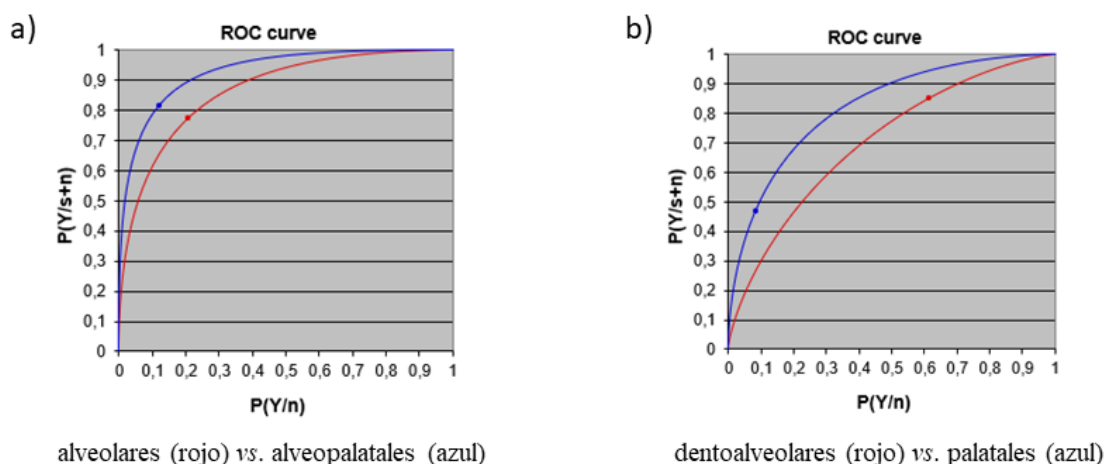
Figura 3. Gráfico que sintetiza las tendencias de respuesta de los jueces en la percepción de las sibilantes objeto de estudio. Se simbolizan en azul los estímulos correspondientes a sibilantes sonoras y, en amarillo, los de las sibilantes sordas. En verde, comportamientos intermedios



Como puede verse en la figura 4, que resume el comportamiento observado en las respuestas de los jueces, en general, las sibilantes sonoras (en azul en el gráfico) y las sordas (en amarillo) se identifican correctamente como tales, salvo en dos casos: los estímulos correspondientes a $tʃ$ tienden a percibirse como sonoros, de forma significativamente similar a z , $ʒ$ y $dʒ$; y los correspondientes a $dʒ$, como sordos, es decir, de forma parecida a s , $ʃ$ y $tʃ$. De hecho, en este caso, la confusión es completa en sílaba átona.

Los parámetros perceptivos analizados hacen posible afinar esta información. En efecto, como se sintetiza en las curvas ROC de la figura 5, en las sibilantes fricativas (5a) se da un sesgo hacia los aciertos, lo que denota que la señal se detecta de forma mucho más clara que en el caso de las africadas (5b). Los valores de d' , que indican la sensibilidad de los jueces respecto a la señal, resultan significativamente más altos para los pares $ʃ$ / $ʒ$ y s / z , aunque también es alto para el $tʃ$ / $dʒ$. El único par que se encuentra por debajo del umbral de 1 es el de las dentoalveolares, lo que señala a una menor sensibilidad de los jueces ante este tipo de estímulos.

Figura 4. Curvas ROC que permiten la comparación entre tipos de sibilante: en (a) entre sibilantes fricativas y en (b), entre sibilantes africadas.



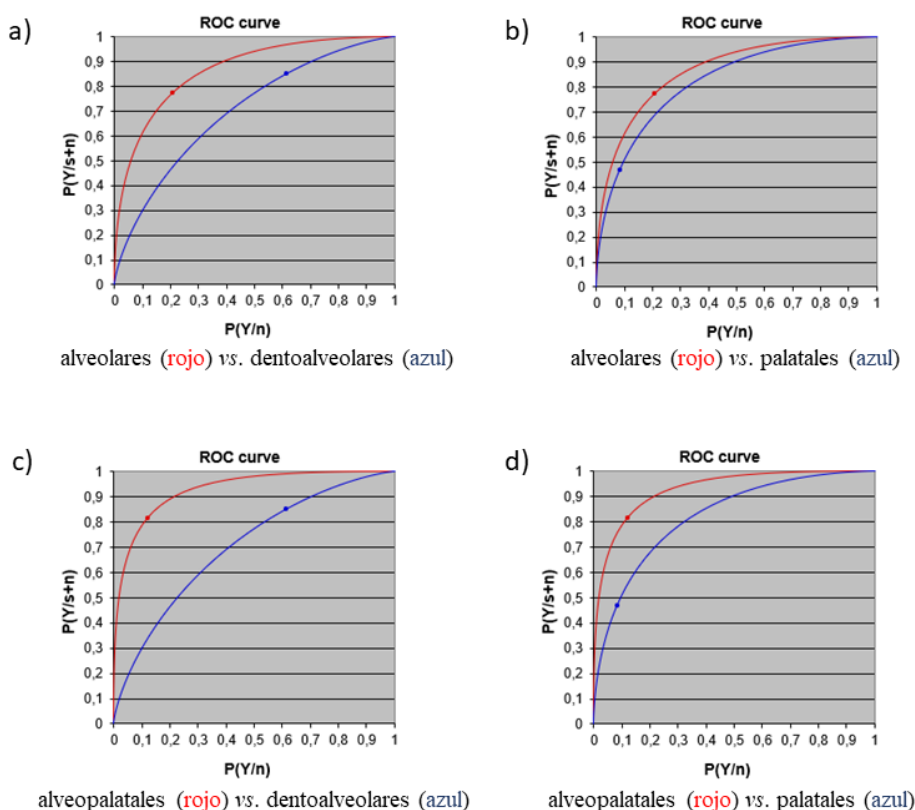
La distancia perceptiva entre los segmentos de cada oposición también sigue el mismo patrón: resulta mucho mayor entre $ʃ$ y $ʒ$ y entre s y z que entre las consonantes que constituyen los pares de

africadas. En cualquier caso, de nuevo, el que presenta menor distancia perceptiva es el $/ts/-/dz/$: estadísticamente, la distancia entre ambas categorías no es significativa, al contrario que en los demás.

Finalmente, si se tiene en cuenta la estrategia de respuesta de los jueces, también se detectan diferencias de comportamiento ante las dentoalveolares: mientras que en los demás grupos los jueces adoptan una estrategia conservadora (es decir, tienden a confundir estímulos sonoros con sordos), ante estímulos dentoalveolares la estrategia es liberal (en caso de confusión, se identifican como sonoros estímulos teóricamente sordos). En cualquier caso, sea cual sea la estrategia, los participantes exhiben sesgos más importantes hacia uno u otro extremo en el caso de las sibilantes africadas. En cambio, ante ejemplos de sibilantes fricativas (con valores de C mucho más próximos a 0) tienden a aproximarse al observador ideal, lo que indica una mayor tasa de éxito en la identificación de los estímulos.

Las diferencias entre los distintos grupos se pueden ver de forma muy clara en las curvas ROC de la figura 6: el comportamiento de las respuestas revela que las dentoalveolares se acercan al umbral del azar de forma mucho más clara que los demás tipos de sibilante. En cualquier caso, los estímulos de alveopalatales y alveolares se discriminan mucho mejor en términos de sonoridad que los de las palatales y dentoalveolares.

Figura 5. Curvas ROC que permiten la comparación perceptiva entre los distintos tipos de sibilantes



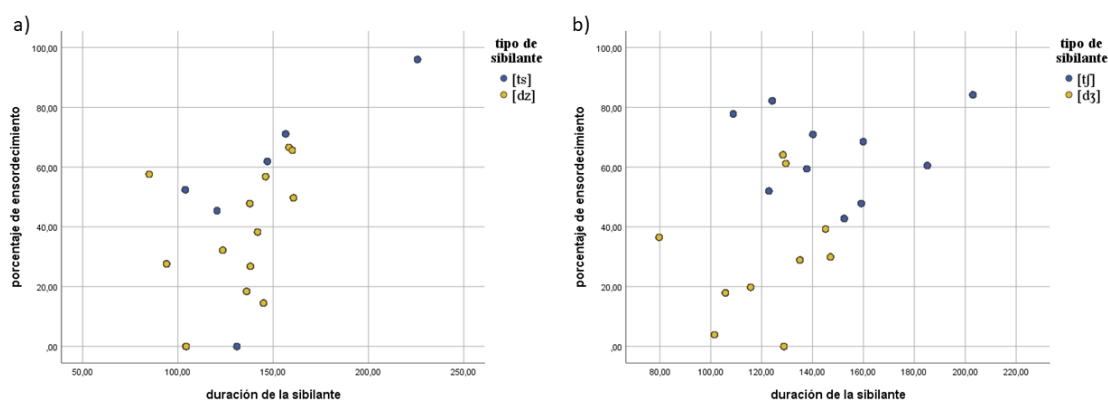
Estos datos, nuevamente, apuntan a la misma idea que los resultados descriptivos generales ya señalaban: los problemas de percepción se concentran en las categorías africadas, que presentan una menor distancia perceptiva entre los miembros de la oposición y ante las que los jueces muestran un grado menor de sensibilidad a la señal. Las dificultades, en este sentido, se dan de forma más evidente en el caso del par $/\text{ts}/$ - $/\text{d}\bar{\text{z}}/$. Esto implica que, en él, resulta más complicado interpretar correctamente los indicadores de sonoridad (la duración y el grado de ensordecimiento, en este experimento). La progresión que se puede establecer desde una mayor a menor dificultad para reconocer los estímulos (dentoalveolares < palatales < alveolares < alveopalatales) coincide, en términos generales, con Pensado Ruiz (1993: 210).

Por otra parte, es importante no olvidar una cuestión crucial. Como ya apuntaban otros estudios (Haggard 1978, Smith 1997, Davidson 2016), la sonoridad no es un rasgo binario, sino gradual. Esto puede dar lugar a variación en el *input* que llega al receptor. De hecho, aunque en las pruebas practicadas las características acústicas de los estímulos no parecían influir en la respuesta del participante (*cfr.* Rost Bagudanch 2022), Kapadia, Tin & Perrachione (2023: 215) señala que la precisión en las respuestas es mayor en casos con una variabilidad mínima entre los estímulos y en aquellos en que hay un único locutor para todos ellos. Aquí los ítems se correspondían con voces de distintos locutores, puesto que se habían extraído de conversaciones espontáneas de diferentes hablantes, y además, como se ha mencionado también en el apartado de metodología, presentaban variabilidad en sus propiedades acústicas de duración y grado de ensordecimiento. Es posible que, en determinados casos (los de las sibilantes africadas), esta variación, aunque no fuera estadísticamente significativa, supusiera un obstáculo lo suficientemente grave como para inducir el enmascaramiento del rasgo de sonoridad dado que se sabe que los hablantes son sensibles al detalle fonético (Pierrehumbert 2001, Beddor 2012, 2023, Goldrick & Cole 2023).

En este sentido, interesa revisar las características acústicas de los casos de las sibilantes africadas proporcionados a los jueces. Como se advierte en la figura 7, los correspondientes a $/\text{ts}/$ y a $/\text{d}\bar{\text{z}}/$ (7a) se agrupan en un mismo espacio acústico, y aunque los del correlato sonoro presentan un porcentaje de ensordecimiento menor que los del correlato sordo, la duración es similar en ambos. Las confusiones parecen dar la razón a Widdison (1995: 41, 1997: 260-261), quien señalaba que la duración es un indicador más robusto de la sonoridad que la propia actividad de las cuerdas vocales.

Por contra, en los ejemplos de sibilantes palatales (7b), se da una diferenciación más clara en ambos planos: los estímulos sonoros presentan valores más bajos tanto en la duración como en el porcentaje de ensordecimiento, mientras que los estímulos sordos los exhiben más altos. La explicación, por lo tanto, no podría ser exactamente la misma que en el caso de las dentoalveolares. No obstante, hay un aspecto en que los estímulos de ambos tipos de sibilante coinciden: la mayor parte de ellos se corresponden con sonidos parcialmente ensordecidos siguiendo la clasificación de Smith (1997: 478). De hecho, lo son el 67 % de los casos de $/\text{ts}/$, el 86 % de los de $/\text{d}\bar{\text{z}}/$, el 70 % de los de $/\text{t}\bar{\text{ɰ}}/$ y el 80 % de los de $/\text{d}\bar{\text{ʒ}}/$. Posiblemente, estos ítems, que no son ni propiamente sordos ni propiamente sonoros, puedan ser percibidos de forma ambigua. Su similitud en este sentido, sumado a que procedan de diversos locutores, puede llevar a la confusión (Kapadia, Tin & Perrachione. 2023: 220-221).

Figura 6. Dispersión que muestran los rasgos de duración y grado de ensordecimiento (porcentaje de ensordecimiento) en los estímulos correspondientes a sibilantes dentoalveolares (a) y palatales (b). A partir de los datos en Rost Bagudanch (2022)



Sin embargo, lo que este razonamiento no desvela es la dirección de la confusión, que resulta contraria en los dos grupos de sibilantes africadas. En este sentido, es importante no perder de vista que, dentro de la fonotáctica del catalán, este tipo de africadas no presenta una frecuencia de aparición alta. En lo referente a las palatales, parece que [tʃ] es ligeramente más común que su correlato sonoro (*cfr.* Rafel i Fontanals 1980: 487), pero no se observan diferencias entre las dos dentoalveolares (*op. cit.*: 480-481). Acústica y articulatoriamente parece que hay razones para pensar en una interpretación como sonora, puesto que se ha dado cuenta de una tendencia, en las fricativas coronales, a la sonorización contextual progresiva desde la vocal precedente ayudada por reajustes temporales en el habla (Bradley, Morris & Jones 2023: 8-9). De hecho, Hualde & Prieto (2014: 116) establecen que la sonorización de /s/ intervocálica en hablantes madrileños se verifica más en posición inicial que interna. De ser aplicable a las sibilantes en general, no debe perderse de vista que, en el fondo, los estímulos de $\sqrt{ts}$ proporcionados, pese a constituir fragmentos no reconocibles léxicamente, se corresponden con inicio de palabra.

En lo que respecta a las palatales, Widdison (1995: 39) indica que la posición de ataque silábico, y especialmente en posición tónica, puede provocar la confusión, dado que hay más tensión articulatoria y esto debería redundar en sonidos más largos e intensos, lo que es típico de las sibilantes sordas. Los resultados obtenidos coinciden con esta observación en lo referente a la posición, pero no en el acento, puesto que la confusión de sonoridad se produce de forma casi exclusiva en contexto átono, en el que estímulos de $\sqrt{dʒ}$ son asociados con $\sqrt{tʃ}$. En cualquier caso, atendiendo a Hay, Warren & Drager (2006: 480), en los procesos de neutralización perceptiva, los oyentes seleccionan las opciones fonéticas en función, entre otros factores, de su experiencia previa, de modo que es probable que los hablantes asocien la señal a $\sqrt{tʃ}$, si esta es más frecuente fonotácticamente. Esto está en la línea de lo defendido por Wedel, Kaplan & Jackson (2013: 184), quienes explican que la neutralización se produce en casos en que la carga funcional de la oposición es baja: el miembro del par que se mantiene es el más probable en términos de frecuencia de uso (aquí, el sordo, como se recordará), mientras que el otro desaparece. El debate sobre la naturaleza fonemática de estos elementos en catalán parece vivo (Wheeler 2005: 11-12), por lo que, si su capacidad distintiva y su consideración como fonemas o alófonos no está clara, no sería extraño que cupiera un proceso de convergencia hacia la categoría más robusta.

Otra cuestión que hay que tener en cuenta es que los resultados del experimento están condicionados por los participantes ($F(25, 5564)=5,015$, $p<0,0001$). Como explica Beddor (2012: 50-52, 2023: 7), los jueces no interpretan de forma uniforme la señal, sino que muestran comportamientos distintos dependiendo de su sensibilidad a los distintos rasgos acústicos de los que esa señal es portadora. Esta autora distingue entre jueces innovadores y conservadores, en función de si sus decisiones a nivel auditivo les llevan a fijar soluciones distintas a las del sistema original o no. Esto se observa de forma estadísticamente significativa en el caso de las sibilantes dentoalveolares ($F(25,1534)=2,072$, $p=0,001$), en que se detecta variabilidad en las tendencias de respuesta de los participantes. En cambio, estos actúan de forma homogénea ante los estímulos de sibilantes palatales, lo que abunda en la idea de que la identificación de los estímulos de $[d\zeta]$ como sorda en sílaba átona va más allá de un comportamiento individual y entra en el terreno de fenómeno más general. Beddor (2012: 51, 2023: 10) especifica que, para que se produzca un cambio fónico no es necesario que todos los oyentes interpreten la señal de forma análoga; basta con unos pocos jueces innovadores para que esta modificación pueda llegar a difundirse.

Todos estos datos, que pueden parecer dispersos, aportan información valiosa para entender el proceso de ensordecimiento histórico del español. No se va a entrar aquí en cuestiones de cronología: no es este el objetivo de este trabajo, que se limita a intentar dar cuenta de las razones fonéticas que pueden haber generado el fenómeno. En cualquier caso, es importante tener presente que existen paralelismos relevantes entre los datos de la bibliografía especializada y los obtenidos en Rost Bagudanch (2022) y aquí: para empezar, se advierte que, en los primeros momentos del cambio, tuvo que haber variación y que su difusión, que se habría dado de norte a sur, fue lenta, progresiva y no uniforme (Quilis 2005: 270-271, Núñez-Méndez 2021). Asimismo, a partir del estudio de las grafías parece que el fenómeno se inició con las africadas dentoalveolares y que, posteriormente, se extendió a las fricativas (Lapesa 1981: 283, Eddington 1987: 58, Quilis 2005: 168-169, Ariza 2012: 224). Esto mismo es lo que se ha encontrado en el ámbito de la producción en Rost Bagudanch (2022, en prensa) para las sibilantes del catalán, que también presentan una variación importante en su realización fonética: de hecho, se ha comprobado que esta variabilidad es mayor en los elementos sonoros de cada par de sibilantes, mientras que los sordos exhiben un comportamiento más homogéneo. El hecho de que la sonoridad sea un rasgo gradual, no binario, no hace sino aumentar las posibilidades de esta variación, que, especialmente en las africadas, se resume en el predominio de soluciones parcialmente ensordecidas sobre las propiamente sordas o sonoras. La situación en cuanto a producción, pues, podría ser análoga a la que desencadenó el cambio en castellano medieval.

De este modo, los ejemplares que constituyen el espacio acústico-auditivo de cada categoría de sibilantes son diversos. En el caso de las sibilantes fricativas se ha comprobado que estos ejemplares se concentran principalmente en realizaciones cercanas a lo esperable, a lo prototípico para este tipo de sonidos (básicamente se producen como sordas o como sonoras, según sea el caso), pero también se incluyen alófonos intermedios (parcialmente ensordecidos) en todas las categorías. Cabe destacar que únicamente se hallan ejemplares sordos en las categorías fonológicamente sonoras, lo que señala su mayor rango de variabilidad. Como se ha visto, la producción en este tipo de sibilantes suele extremar las diferencias: se trata de pares productivos, con una carga funcional importante, frecuentes en el idioma, y no parece que se den solapamientos entre ellos con facilidad. Sin embargo, la percepción muestra esa idea de asimetría que exponen Labov, Karen & Miller (1991) y que recoge Beddor (2012, 2023): pese a que no hay dudas, en

general, en la identificación de los sonidos, se detectan confusiones que pueden llegar a más del 20 % de los estímulos administrados a los jueces. Se trata de categorías más resistentes a la neutralización, puesto que la distancia perceptiva entre los miembros de las oposiciones es amplia y los jueces son altamente sensibles a la señal, pero esto no implica que, si se da un sesgo en la producción de las realizaciones a favor de soluciones con menos sonoridad o si los oyentes reinterpretan las categorías sonoras como sordas en contextos que puedan favorecer este proceso, el centro de estas nubes de ejemplares no se pueda desplazar y acabar solapándose (Pierrehumbert 2001, 2002).

Esto resulta bastante evidente en el caso de las sibilantes africadas. En ellas, se advierte que el grueso de las manifestaciones acústicas detectadas ya no se sitúa en los extremos, con lo cual la distancia entre los miembros de cada par disminuye de forma significativa. Las soluciones son más similares entre sí entre los posibles ejemplares y pueden darse solapamientos entre los alófonos de la categoría sorda y los de la sonora. De este modo se habría iniciado el proceso de desfonologización, que habría podido ir favorecido por la menor carga funcional que presentan este tipo de oposiciones (Lloyd 1993: 428, Ariza 2012: 223, Wedel, Kaplan & Jackson 2013). En este sentido, hay que recordar que, en principio, el castellano medieval no tenía oposición fonológica en términos de sonoridad entre sibilantes palatales (Núñez-Méndez 2021: 14): $[\text{d}\text{z}]$ era, simplemente un alófono de $/\text{z}/$, pero no presentaba capacidad distintiva. De hecho, a tenor de los resultados obtenidos aquí, esto parece lógico, puesto que los mismos jueces catalanohablantes parecen no distinguirla de la sorda en contexto átono en lo que al nivel perceptivo se refiere.

En lo que respecta a las dentoalveolares, que sí constituían una oposición fonológica en castellano medieval, se han podido detectar aspectos interesantes. En primer lugar, la producción de este tipo de segmentos, como se comentaba, responde principalmente a realizaciones intermedias, parcialmente ensordecidas. En el caso de $[\text{ts}]$, hay un equilibrio entre estas soluciones y las propiamente sordas, pero en el caso de $[\text{d}\text{z}]$, las parcialmente ensordecidas suponen el 70 % de los ejemplos obtenidos en el estudio acústico (Rost Bagudanch 2024). Perceptivamente, se advierte confusión, pero al contrario de lo esperado para ilustrar el proceso que nos ocupa. Ciertamente, el hecho de que los estímulos de $[\text{ts}]$ se encuentren originalmente en posición inicial (al contrario de los de $[\text{d}\text{z}]$) puede haber tenido un papel crucial en los resultados obtenidos y sesgar la prueba auditiva, que muestra una tendencia contraria a los resultados acústicos en Rost Bagudanch (2022, 2024). De todos modos, hay que recalcar que el comportamiento individual de los jueces influye en los resultados: tres de los jueces cometen significativamente más errores en la identificación de los estímulos dentoalveolares que todos los demás¹⁷: se trata de tres jueces innovadores que malinterpretan la señal de forma mucho más sesgada que el resto de participantes, y tienden a identificar mucho más como sordos los casos de $[\text{d}\text{z}]$, algo que los demás no hacen. Como indica Beddor (2023: 10), no es necesario que todos los hablantes se comporten de forma homogénea para que se desencadene el cambio: basta con que algunos de ellos reinterpreten la señal y que, en cuanto pasen de actuar como receptor a actuar como locutor, incorporen esta innovación, que habrán almacenado entre los ejemplares posibles para esa categoría fonológica, en su producción. La existencia de este tipo de oyentes demuestra que el cambio por razones perceptivas es posible y pone de manifiesto, además, que la percepción puede ir por delante de la producción en la compleción de los cambios fónicos, tal como postula Beddor (2023: 10).

¹⁷ ($b_0 = -0.892$, $se = 0.422$, $t = -2.115$, $p = 0.035$) en dos de ellos y ($b_0 = -1.135$, $se = 0.422$, $t = -2.692$, $p = 0.007$) en el tercero.

Los resultados expuestos aquí hacen posible entender que la causa del ensordecimiento de las sibilantes hay que buscarla en aspectos fonéticos de la producción y la percepción de este tipo de sonidos, tal como defendían Alonso (1967), Pensado Ruiz (1993) o Widdison (1995, 1997). En realidad, se advierte que este proceso puede estar liderado por aspectos perceptivos del habla, que serían los que tirarían de la variación alofónica existente y la llevarían, auditivamente, a ser reinterpretada en términos de sonoridad, lo que habría acabado acarreando un solapamiento entre los ejemplares de los pares sordo-sonoro de cada tipo de sibilante. El final de este solapamiento habría resultado en la desaparición del miembro menos productivo de cada par, que parece ser el sonoro atendiendo a la variabilidad que puede mostrar frente a la mayor homogeneidad de los segmentos sordos, que parecen más robustos tanto en producción como en percepción. De hecho, este tipo de procesos no son privativos del español medieval: Mackenzie (2022: 1, 12) ya ponía de manifiesto que se da en otras variedades románicas (dialectos del véneto y del franco-provenzal) y la bibliografía especializada en consonantes fricativas incluye el ensordecimiento como un fenómeno habitual entre las sibilantes sonoras, precisamente por sus características aerodinámicas y articulatorias (Solé 2003; Ohala & Solé 2010: 39-41; Žygis, Fuchs & Koenig 2012: 310-312). Los trabajos de Rost Bagudanch (2022, 2024) también evidencian que no es extraño al catalán, al menos a partir de los datos obtenidos. Así, cabría contemplar la evolución del español dentro de las dinámicas fonético-fonológicas lógicas y esperables para las lenguas del mundo, no como algo aislado. Cuestión aparte es por qué se materializa y queda fijado en la época en que lo hizo, aunque esta ya es materia para otro trabajo.

5. Conclusiones

El presente estudio quiere ser una aproximación al proceso de ensordecimiento de las consonantes sibilantes del español desde la fonética experimental, concretamente desde la perspectiva de la percepción. La intención es comprobar, más allá de los resultados obtenidos en investigaciones previas, si el aspecto auditivo podría encontrarse en el origen de los cambios que culminaron con una reestructuración importantísima en el sistema consonántico entre los siglos *xvi* y *xvii*.

Las explicaciones más extendidas sobre este proceso suelen relacionarlo con la necesidad de incorporar mejoras al sistema que supongan hacerlo más funcional, simétrico y rentable, o con el contacto de lenguas, sin contemplar causas internas, de tipo fonético. De hecho, autores como Alonso (1967), Pensado Ruiz (1993) o Widdison (1997) ya abogaban por relacionar el cambio con cuestiones articulatorias y perceptivas; es decir, con procesos de variación y cambio fónico más regulares. Un trabajo previo (Rost Bagudanch 2022), cuyos resultados apuntan en este mismo sentido, pone de manifiesto que la neutralización de sonoridad de las sibilantes se ha de atribuir a cuestiones tanto de producción como de percepción de la señal.

Así pues, se ha profundizado en el análisis de una prueba de identificación auditiva que indaga en el reconocimiento de la sonoridad en el caso de estímulos procedentes de una lengua hermana, el catalán. El recurso a la gramática comparada y a técnicas experimentales con datos sincrónicos, validado por las investigaciones recientes en el ámbito del cambio lingüístico, permite trabajar con material fónico de un sistema muy similar al del castellano medieval, punto de partida de la evolución.

Los datos obtenidos apuntan inequívocamente a causas puramente fonéticas para la explicación del ensordecimiento: se trata de la neutralización de la oposición fonológica de sonoridad en los pares de sibilantes del sistema. Como suele ser habitual en estos procesos, hay que distinguir su doble vertiente, la de la producción y la de la percepción, que no son simétricas. El análisis acústico de estas consonantes, verificado en Rost Bagudanch (2022, en prensa), muestra la existencia de variación alofónica en la realización de todas las sibilantes, aunque es en las africadas donde la producción de soluciones con un grado de sonoridad intermedio aumenta de forma importante (las fricativas, /s/-/z/ y /ʃ/-/ʒ/ presentan soluciones más extremas, como propiamente sordas o sonoras). El estudio de la percepción practicado aquí evidencia que las sibilantes fricativas presentan una distancia perceptiva mayor entre los miembros de la oposición que las sibilantes africadas, especialmente en el par dentoalveolar, cuya distancia perceptiva es muy escasa. Por otra parte, los jueces muestran mayor sensibilidad a la señal ante estímulos sibilantes fricativos, incluso ante africados palatales, pero esta sensibilidad se reduce drásticamente en el caso de /t͡s/ y /d͡ʒ/. Asimismo, recurren a estrategias de respuesta distintas: mientras que ante sibilantes fricativas su comportamiento es muy similar al del observador ideal, sin sesgos, ante las africadas se pueden rastrear sesgos, o bien a identificar como sordos estímulos sonoros (las palatales) o bien a identificar como sonoros estímulos sordos (las dentoalveolares). Todo esto pone de relieve que la percepción se desvía de la producción, puesto que el nivel de confusiones existente para las africadas excede la información de los datos acústicos, los cuales, pese a señalar una mayor similitud entre las realizaciones fonéticas, indicaban aún una cierta diferenciación entre sordas y sonoras (relativamente clara, en números globales, para las palatales, aunque ya no para las dentoalveolares).

Por otra parte, se ha podido comprobar que, en el caso de las sibilantes africadas, por donde comenzó el cambio histórico y donde los datos experimentales señalan también mayores cotas de ambigüedad fonética, el comportamiento de los jueces no es homogéneo, algo que reproduce a pequeña escala el funcionamiento de los procesos de cambio fonológico: el origen no es una alteración masiva en la interpretación y posterior producción de la señal, sino las decisiones inconscientes de unos cuantos individuos, que seleccionan, entre los ejemplares disponibles, soluciones distintas de las canónicas. A partir de ellos se puede difundir la innovación.

6. Anexo

Estímulos empleados en la prueba de percepción (cfr. Rost Bagudanch 2022: 24-25):

fonema	sílabas átonas		sílabas tónicas	
	palabra de origen	estímulo	palabra de origen	estímulo
/s/	passa 'pasa'	[a'sə]	Assajos 'ensayos'	[ə'sa]
/z/	casa 'casa'	[a'zə]	casar 'casar'	[ə'za]
/ʃ/	aixafar 'aplastar'	[əʃə]	deixar 'dejar'	[ə'ja]
/ʒ/	vaja 'vaya'	[a'ʒə]	barrejar 'mezclar'	[ə'ʒa]
/ts/	(considerav)a tsarisme 'consideraba zarismo'	[ə'tsə]	(paraul)a tsar 'palabra zar'	[ə'tsa]
/dz/	atzabeja 'azabache'	[ə'dzə]	atzar 'azar'	[ə'dza]
/tʃ/	borratxa 'borracha'	[a'tʃə]	catxat 'desmejorado'	[ə'tʃa]
/dʒ/	platja 'playa'	[a'dʒə]	assetjar 'hostigar'	[ə'dʒa]

Bibliografía

- ABDI, René (2007): «Signal Detection Theory (SDT)». En N. Salkind (ed.): *Encyclopedia of Measurement and Statistics*. Thousand Oaks: Sage, 886-889.
DOI: <https://doi.org/10.4135/9781412983907>.
- ALARCOS LLORACH, Emilio (1988): «De nuevo sobre los cambios fonéticos del s. XVI». En M. Ariza, A. Salvador & A. Viudas (eds.): *Actas del I Congreso Internacional de Historia de la Lengua Española*. Madrid: Arco / Libros, vol. 1, 47-60.
- ALONSO, Amado (1967): *De la pronunciación medieval a la moderna en español*. Madrid: Gredos, vol. I.
- ARIZA, Manuel (2012): *Fonología y fonética históricas del español*. Madrid: Arco / Libros.
- BEDDOR, Patrice S. (2012): «Perception grammars and sound change». En M.-J. Solé & D. Recasens (eds.): *The Initiation of Sound Change. Perception, Production and Social Factors*. Amsterdam / Philadelphia: John Benjamins, 37-55.
DOI: <https://doi.org/10.1075/cilt.323.06bed>.
- BEDDOR, Patrice S. (2023): «Advancements of phonetics in the 21st century: Theoretical and empirical issues in the phonetics of sound change». *Journal of Phonetics* 97, 101228, 1-17.
DOI: <https://doi.org/10.1016/j.wocn.2023.101228>.
- BIRDSONG, David, Libby M. GERTKEN & Mark AMENGUAL (2012): *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*. COERLL, University of Texas at Austin.
En línea: <https://sites.la.utexas.edu/bilingual/>.
- BLEVINS, Juliette (2004): *Evolutionary Phonology. The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.
DOI: <https://doi.org/10.1017/CBO9780511486357>.

- BOERSMA, Paul & David WEENINK (2018): *Praat: Doing Phonetics by Computer* (version 6.0.4.). <http://www.fon.hum.uva.nl/praat/>.
- BRADLEY, Travis y Ann Marie DELFORGE (2006): «Systemic contrast and the diachrony of Spanish sibilant voicing». En D. Arteaga & R. Gess (eds.): *Historical Romance Linguistics: Retrospectives and Perspectives*, Amsterdam: John Benjamins, 19–52.
DOI: <https://doi.org/10.1075/cilt.274.04bra>.
- BRADLEY, Travis, Kimberly MORRIS y Lillian JONES (2023): «Incomplete resyllabification and ambisyllabic gestural coupling in Spanish». *Loquens* 9/1-2, e094.
DOI: <https://doi.org/10.3989/loquens.2022.e094>.
- CANO AGUILAR, Rafael (2004): «Cambios en la fonología del español durante los siglos XVI y XVII». En R. Cano (coord.): *Historia de la lengua española*. Barcelona: Ariel, 825-858.
- CONTINI, Gianfranco (1951): «Sobre la desaparición de la correlación de sonoridad en castellano». *Nueva Revista de Filología Hispánica* 5, 173-182.
DOI: <https://doi.org/10.24201/nrfh.v5i2.215>.
- DAVIDSON, Lisa (2016): «Variability in the implementation of voicing in American English obstruents». *Journal of Phonetics* 54, 35-50.
DOI: <https://doi.org/10.1016/j.wocn.2015.09.003>.
- EDDINGTON, David S. (1987): «Spanish sibilant evolution». *Deseret Language and Linguistic Society Symposium* 13, 55-62.
- FERNÁNDEZ REI, Elisa (coord.) (2021): *FOLERPA: Ferramenta On-Line para ExpeRimentación Perceptiva*. Santiago de Compostela: Instituto da Lingua Galega.
En línea: <https://ilg.usc.gal/folepa/> (consulta 29/9/2023).
- GOLDRICK, Matthew y Jennifer COLE (2023): «Advancements of phonetics in the 21st century: Exemplar models of speech production». *Journal of Phonetics* 99, 101254 /1-16.
DOI: <https://doi.org/10.1016/j.wocn.2023.101254>.
- HAGGARD, Mark (1978): «The devoicing of voiced fricatives». *Journal of Phonetics* 6, 95-102.
DOI: [https://doi.org/10.1016/S0095-4470\(19\)31101-5](https://doi.org/10.1016/S0095-4470(19)31101-5).
- HARRINGTON, Jonathan (2012): «The relationship between synchronic variation and diachronic change». En A. C. Cohn, C. Fougeron & M. K. Huffman (eds): *The Oxford Handbook of Laboratory Phonology*. Oxford: Oxford University Press, 321-332.
- HAY, Jennifer, Paul WARREN & Katie DRAGER (2006): «Factors influencing speech perception in the context of a merger-in-progress». *Journal of Phonetics* 34, 458-484.
DOI: <https://doi.org/10.1016/j.wocn.2005.10.001>.
- HUALDE, José Ignacio & Pilar PRIETO (2014): «Lenition of intervocalic alveolar fricatives in Catalan and Spanish». *Phonetica* 71, 109-127.
DOI: <https://doi.org/10.1159/000368197>.
- JOHNSON, Keith (2003): *Acoustic and Auditory Phonetics*. Oxford: Blackwell.
- KAPADIA, Alexandra M., Jessica A. A. TIN y Tyler K. PERRACHIONE (2023): «Multiple sources of acoustic variation affect speech processing efficiency», *Journal of the Acoustic Society of America* 153/1, 209-223.
DOI: [10.1121/10.0016611](https://doi.org/10.1121/10.0016611).
- LABOV, William (1994): *Principles of Linguistic Change*. Vol. I: *Internal Factors*. Oxford: Blackwell.
- LABOV, William, Mark KAREN & Corey MILLER (1991): «Near-mergers and the suspension of phonemic contrast». *Language Variation and Change* 3, 33-74.
DOI: <https://doi.org/10.1017/S0954394500000442>.
- LAHOZ, José María. (2015): *Fonética y fonología de los fenómenos de refuerzo consonántico en el seno de unidades léxicas en español*. Tesis doctoral. Universidad Complutense de Madrid.
- LAPESA, Rafael (1981): *Historia de la lengua española*. Madrid: Gredos.

- LAVOIE, Lisa (2001): *Consonant Strength. Phonological Patterns and Phonetic Manifestations*. New York: Routledge.
- LIPSKI, John (1989): «/s/ voicing in Ecuadoran Spanish». *Lingua* 79, 49-71.
DOI: [https://doi.org/10.1016/0024-3841\(89\)90019-3](https://doi.org/10.1016/0024-3841(89)90019-3).
- LLOYD, Paul M. (1993): *Del latín al español*. Madrid: Gredos.
- MACKENZIE, Ian (2022): «The genesis of Spanish /θ/: A revised model». *Languages* 7/3, 191, 1-18.
DOI: <https://doi.org/10.3390/languages7030191>.
- MACMILLAN, Neil A. (2002): «Signal detection theory». En H. Pashler & J. Wisted (eds.): *Stevens' Handbook of Experimental Psychology*. Hoboken: Wiley, vol. 4, 43-90.
DOI: <https://doi.org/10.1002/0471214426.pas0402>.
- MARRERO, Victoria (2014): «Metodología de investigación en fonética perceptiva». En Y. Congosto Martín, M. L. Montero Curiel & A. Salvador Plans (eds.): *Fonética experimental, educación superior e investigación*. I. *Fonética y Fonología*. Madrid: Arco Libros, 503-541.
- MARTINET, André (1951-1952): «The unvoicing of Old Spanish sibilants». *Romance Philology* 5, 133-156.
- MENÉNDEZ PIDAL, Ramón (1985): *Manual de gramática histórica española*. Madrid: Espasa-Calpe.
- NÚÑEZ-MÉNDEZ, Eva (2021): «An overview of the sibilant merger and its development in Spanish». En E. Núñez-Méndez (ed.): *Sociolinguistic Approaches to Sibilant Variation in Spanish*. Berlin / New York: Routledge, 9-72.
- NÚÑEZ-MÉNDEZ, Eva (2022): «Variation in Spanish /s/: Overview and new perspectives». *Languages* 7/2, 77, 1-50.
DOI: <https://doi.org/10.3390/languages7020077>.
- OHALA, John (1981): «The listener as a source of sound change». En C. S. Masek, R. A. Hendrick & M. F. Miller (eds.): *Papers from the Parasession on Language and Behavior*. Chicago: Chicago Linguistic Society, 178-203.
- OHALA, John (1993): «The phonetics of sound change». En Charles Jones (ed.). *Historical Linguistics. Problems and Perceptions*. London: Longman, 237-278.
- OHALA, John (2012): «The listener as a source of sound change. An update». En M.-J. Solé & D. Recasens (eds.): *The Initiation of Sound Change. Perception, Production and Social Factors*. Amsterdam / Philadelphia: John Benjamins, 21-35.
DOI: <https://doi.org/10.1075/cilt.323.05oha>.
- OHALA, John & Maria-Josep SOLÉ (2010): «Turbulence and phonology». En S. Fuchs, M. Toda & M. Žygis (eds.): *Turbulent Sounds. An Interdisciplinary Guide*. Berlin: De Gruyter, 37-101.
DOI: 10.1515/9783110226584.37.
- PARDO, Jennifer (2006): «On phonetic convergence during conversational interaction». *Journal of the Acoustical Society of America* 119/4, 2382-2393.
DOI: 10.1121/1.2178720.
- PENNY, Ralph (1993): «Neutralization of voice in Spanish and the outcome of the Old Spanish sibilants». En D. Mackenzie & I. Michael (eds.): *Hispanic Linguistic Studies in Honour of F. W. Hodcroft*. Oxford: Dolphin, 75-88.
- PENNY, Ralph (2004): «Evolución lingüística en la Baja Edad Media: evoluciones en el plano fónico». En R. Cano (coord.). *Historia de la lengua española*. Barcelona: Ariel, 593-612.
- PENNY, Ralph (2014): *Gramática histórica del español*. Barcelona: Ariel.
- PENSADO RUIZ, Carmen (1993): «El ensordecimiento castellano: ¿un "fenómeno extraordinario"?». *Anuario de Lingüística Hispánica* IX, 195-230.
- PIERREHUMBERT, Janet B. (2001): «Exemplar dynamics: Word frequency, lenition and contrast». En J. Bybee & P. Hopper (eds.): *Frequency Effects and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins, 137-51.
DOI: <https://doi.org/10.1075/tsl.45.08pie>.

- PIERREHUMBERT, Janet B. (2002): «Word-specific phonetics». En C. Gussenhoven & N. Warner (eds.): *Laboratory Phonology VII*. Berlin / New York: Mouton de Gruyter, 101-139.
DOI: <https://doi.org/10.1515/9783110197105.1.101>.
- QUILIS, Antonio (2005): *Fonética histórica y fonología diacrónica*. Madrid: UNED.
- RAFEL I FONTANALS, Joaquim (1980): "Dades sobre la freqüència de les unitats fonològiques del català". *Estudis Universitaris Catalans* 25, 473-496.
- ROST BAGUDANCH, Assumpció (2022): «More on sibilant devoicing in Spanish diachrony: An initial phonetic approach». *Languages* 7/1, 27, 1-28.
DOI: <https://doi.org/10.3390/languages7010027>.
- ROST BAGUDANCH, Assumpció (2024): «L'ensordiment de sibilants en dos dialectes del català: Girona vs. Mallorca». En W. Elvira-García & P. Roseano (eds.): *Avances metodológicos en fonética y prosodia*, Madrid: UNED, 321-332.
- ROST BAGUDANCH, Assumpció & Paolo ROSEANO (2021): «Sobre el origen de /x/ en español: argumentos desde la fonética comparada». En L. Schøsler, J. Härmä & J. Lindschouw (eds.): *Actes du XXIXème Congrès International de Linguistique et de Philologie Romanes*. Strasbourg: Éditions de linguistique et de philologie (ELiPhi), vol. 1, 251-262.
- SANCHEZ-PIRIETO BORJA, Pedro (2004): «La normalización del castellano escrito en el siglo XIII. Los caracteres de la lengua: grafías y fonemas». En R. Cano (coord.): *Historia de la lengua española*. Barcelona: Ariel, 423-448.
- SMITH, Caroline (1997): «The devoicing of /z/ in American English: Effects of local and prosodic context». *Journal of Phonetics* 25, 471-500.
DOI: <https://doi.org/10.1006/jpho.1997.0053>.
- SOLÉ, Maria-Josep (2003): «Aerodynamic characteristics of onset and coda fricatives». En M.-J. Solé, D. Recasens & J. Romero (eds.): *15th International Conference on Phonetic Sciences*. Barcelona: ICPhS Archive, vol. 2.
En línea: 2761-64. <http://www.internationalphoneticassociation.org/icphs/icphs2003> (consulta 17/11/2023).
- WEDEL, Andy, Abby KAPLAN & Scott JACKSON (2013): «High functional load inhibits phonological contrast loss: A corpus study». *Cognition*, 128, 179-186.
DOI: <https://doi.org/10.1016/j.cognition.2013.03.002>.
- WETZELS, Leo & Joan MASCARÓ (2001): «The typology of voicing and devoicing». *Language* 77, 207-244.
DOI: 10.1353/lan.2001.0123.
- WHEELER, Max (2005): *The Phonology of Catalan*. Cambridge: Cambridge University Press.
- WIDDISON, Kirk (1995): «Physical constraints on sibilant-voice patterning in Spanish Phonology». En Jeffrey Turley (ed.), *Proceedings of the 1995 Desert Language and Linguistics Symposium*. Provo: BYU Linguistics Department, 37-42.
- WIDDISON, Kirk (1997): «Phonetic explanations for sibilant patterns in Spanish». *Lingua* 102, 253-264.
DOI: [https://doi.org/10.1016/S0024-3841\(97\)00006-5](https://doi.org/10.1016/S0024-3841(97)00006-5).
- ŽYGIS, Marzena, Susanne FUCHS & Laura L. KOENIG (2012): «Phonetic explanations for the infrequency of sibilant affricates across languages». *Journal of Laboratory Phonology* 3, 299-336.
DOI: 10.1515/lp-2012-0016.