

Article

More on Sibilant Devoicing in Spanish Diachrony: An Initial Phonetic Approach

Assumpció Rost Bagudanch 

Department of Spanish Philology, Faculty of Letters, University of the Balearic Islands, 07122 Palma, Spain; assumpcio.rost@uib.es

Abstract: The devoicing of sibilants took place in Early Modern Spanish, a phenomenon which has been considered problematic to account for due to its occurrence context (medial intervocalic position). Traditional explanations invoked Basque influence or a structural reorganization in search for a more balanced system. However, phonetically based reasons were proposed by some scholars. This research is a preliminary attempt to support these proposals with experimental data from a comparative grammar perspective. The Catalan sibilant system, which is very similar to the Medieval Spanish one, is acoustically and perceptively studied in order to investigate the acoustic cues of voicing and to determine if devoicing is possible. Results indicate that (a) voicing relies mainly in the proportion of unvoiced frames of the segments, on its duration, and, to a lesser extent, on its intensity; (b) sibilant devoicing occurs in all voiced categories; (c) auditorily, confusion between voiced and voiceless segments can be attested for every sibilant pair, and (d) the misparsings are more common in affricate and in palatal sibilants, [dʒ] being the most prone to be labelled as unvoiced. These findings prove that the historical process in Spanish could have a phonetic basis.

Keywords: sibilant devoicing; sound change; historical phonology; Spanish diachrony; Catalan sibilants



Citation: Rost Bagudanch, Assumpció. 2022. More on Sibilant Devoicing in Spanish Diachrony: An Initial Phonetic Approach. *Languages* 7: 27. <https://doi.org/10.3390/languages7010027>

Academic Editor: Ian MacKenzie

Received: 30 November 2021

Accepted: 24 January 2022

Published: 30 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The reorganization of the sibilant system in Early Modern Spanish has received a great deal of attention, since it implied decisive changes in Spanish consonantism. As known, the “sibilant turmoil” (the term used by Kiddle 1977) involved different phonological processes which are assumed to have concluded in the 17th century. These complex processes comprise deaffrication, devoicing and changes in the point of articulation of part of the ancient sibilants, a series of changes that can be subsumed in the loss of half of the segments which constituted the original set. The present study, which is a first approach to the topic, does not aim to cover the whole evolution, but only the phenomenon of devoicing. The devoicing process of Medieval Spanish sibilants has been widely accounted for in terms of teleological sound change;¹ however, some scholars have pointed to the possibility of a phonetic-based sound change as described in Ohala (1981, 2012), Pierrehumbert (2001, 2002) and even Blevins (2004). Pensado (1993) and Widdison (1995, 1997) are also relevant examples. This paper is intended to examine their view in greater depth and to give new insights into the initiation of this sound change from an experimental perspective.

1.1. Historical Evolution

Medieval Spanish presented three pairs of sibilant consonants: dentoalveolar affricates /tʃ, dʒ/, alveolar fricatives /s, z/ and postalveolar fricatives /ʃ, ʒ/, plus a palatal voiceless affricate /tʃ/.² The voiced counterparts arose from a voicing process in Late Latin in intervocalic context (Lapesa 1981, p. 124; Lloyd 1993, pp. 238–40, 389, 423; Cano 2004, pp. 834–35). Devoicing would have taken place in the 15th to 16th centuries (Alonso 1967, p. 312; Lapesa 1981, p. 283; Eddington 1987, p. 59; Lloyd 1993, pp. 427–28; Cano 2004, pp. 833–34; Ariza 2012, pp. 222–23; Penny 2014, pp. 121–22). However, there is

some evidence that the confusion initiated much earlier, in the 13th (Sánchez Prieto 2004, p. 442, regarding the dentoalveolar pair) or 14th centuries (Penny 2004, pp. 603–4); in fact, Quilis (2005, pp. 270–71) points to an ancient phenomenon. In any case, orthographic errors in Medieval texts point to a primary confusion between the result of dentoalveolar deaffrication (/t͡s, d͡z/ > /s̺, z̺/), which would have induced devoicing in the fricative pairs (Eddington 1987, p. 58; Lapesa 1981, p. 283; Quilis 2005, pp. 168–69; Ariza 2012, p. 224).³

As we have said, voiced–voiceless pairs were only contrastive in medial intervocalic position, while only voiceless sibilants were possible in word-initial and coda positions, except for the postalveolar pair (/ʃ, ʒ/), which also contrasted in word-initial position. Some researchers suggest an intermediate stage in which word-final sibilants in prevocalic context would also have been voiced before devoicing was completed (Penny 1993, p. 80;⁴ Bradley and Delforge 2006, pp. 22–23), as occurs in other Romance varieties, such as Catalan or Italian dialects. Even some modern varieties of Spanish, such as highland Ecuadorian Spanish, display this behaviour (see Lipski 1989 for a detailed description).

The devoicing process has been considered a rare evolution within the Western Romance varieties by some scholars (Alarcos 1988, pp. 51–52; Lloyd 1993, p. 428), since in such a context voicing would be expected (Lahoz 2015; Davidson 2016, p. 37) and it is this atypical evolution that has sparked interest in the topic. The literature has put forward different explanations for it, which range from language contact to aerodynamic reasons. The most traditional accounts attribute devoicing to (a) the influence of Basque, which lacked voiced sibilants (Martinet 1951–1952; Lloyd 1993, pp. 429–37), (b) a readjustment of the system to improve its efficiency, since voiced–voiceless oppositions were not productive enough (Alarcos 1988, pp. 51–53; Penny 1993, pp. 81–82; Ariza 2012, p. 224) or (c) a readjustment of the system aiming at a more perfect symmetry, since /t͡ʃ/ did not have a voiced counterpart (Contini 1951, pp. 179–80). Pensado (1993) deals with these approaches and her conclusion is that neither language contact nor structural accounts seem to be completely satisfactory. Moreover, she argues that they are unnecessary since internal phonetic factors can justify the change. Pensado shows that sibilant devoicing is not unique to Spanish (cf. Żygis et al. 2012 for a detailed explanation of the infrequency of voiced sibilants in the world’s languages) and that, in the world’s languages, this process usually begins with a specific pair, usually palatal or dentoalveolar affricates. Affricates tend to devoice before fricatives, and palatals do so before than alveolars. However, Pensado (1993) was not the first to point to phonetic causes. Alonso (1967, p. 313) also referred to the intrinsic properties of sibilants. The important issue here is that voiced sibilants need to ensure simultaneous vocal-fold vibration and turbulence, which is not a simple task, due to the required aerodynamic conditions (broadly speaking, intraoral pressure must be less than subglottal pressure but higher than exterior pressure to maintain both mechanisms, cf. Pensado 1993, pp. 214–16; Solé 2003; Ohala and Solé 2010, pp. 39–41; Żygis et al. 2012, pp. 310–12).

It is worth noting that a parallel devoicing phenomenon is nowadays occurring in certain varieties of Spanish regarding a particular evolution of *yeísmo* (the merger process of /ɲ/ and /j/ in favour of the latter): in Rioplatense Spanish (mainly Buenos Aires area in Argentina and Montevideo in Uruguay), /ʒ/ resulting from the neutralization is being devoiced in a way similar to Early Modern Spanish. According to the literature (Fontanella de Weinberg 1987, pp. 146–50; 2000), this sound change presents the characteristics of a change from below, since it originated in middle classes and among young speakers. At present, it displays different rates of implementation. In Buenos Aires it seems to be completed among younger speakers in middle classes, and it is likely to be almost completed in younger speakers of the upper classes (Rohena-Madrado 2013, 2015). However, Montevideo has not reached the same level of extension yet (Michnowicz and Planchón 2020).

1.2. Obstruent Devoicing Process

One of the main objections by Pensado (1993) to the classical analysis is that she rejects the idea of devoicing as an atypical evolution for sibilants. Obstruent devoicing has been

widely attested, described and explained not only in coda position but also in word-initial context (Lavoie 2001, pp. 27, 43; Wetzels and Mascaró 2001; Blevins 2004, pp. 103–6, 110–11; Lahoz 2015, pp. 142, 169–71).⁵ When examining the inventories of obstruent elements in the world's languages, the general impression is that voiced segments are not common (Smith 1997, p. 473; Ohala and Solé 2010, p. 53; Żygis et al. 2012). Indeed, since the aerodynamic requirements for producing them are more complex, contrasts may be lost by means of various phonological processes whose result is a reduction in the number of voiced units (Ohala and Solé 2010, pp. 53–54; Żygis et al. 2012, pp. 308–9, 313; Hualde and Prieto 2014, p. 111). These mechanisms include changes to a voiced fricative (in the case of affricates) or a glide, defrication, vocalization, and, of course, devoicing.

Take special note that devoicing affects sibilants differently depending on their context. Haggard (1978) shows that fricatives are more likely to devoice when followed by a voiceless stop or even a voiced stop than in intervocalic position; and word-final position also seems to favour devoicing. Smith (1997), in her survey on /z/ devoicing in American English, corroborates this. Her data show that /z/ can devoice in any position, but the extent and the likelihood of the process vary depending on the neighbouring sounds and prosodic structure: a following voiceless sound and a final position in the prosodic domain increase the likelihood of devoicing. These findings are coherent with the general picture on obstruent devoicing presented above.

At this point, we must introduce some remarks on what devoicing implies from a phonetic point of view. Obviously, devoicing entails a reduction in vocal-fold activity,⁶ but it also involves other relevant parameters. Experiments in production and perception point to intensity, degree of glottal tone, F1 transitions and, specifically, duration (Widdison 1995, pp. 38–39; Smith 1997, p. 473; Lavoie 2001, p. 107; Bradley and Delforge 2006, p. 31): voiceless sibilants are usually longer and more intense, with short F1 transitions and no glottal vibration (or, at least, not constant glottal vibration). In fact, there is evidence that vocal fold vibration is not constant for the whole duration of voiced sibilants (Haggard 1978; Pensado 1993, p. 216; Widdison 1995, p. 38; Smith 1997, p. 260; Ohala and Solé 2010, pp. 71–73; Davidson 2016). Smith (1997) observes that a large proportion of /z/ utterances in her experiment were partially devoiced (she considers as partially devoiced all cases in which voicing covers a range from 25% to 90% of the sound). Similar results are found in Van de Velde and van Hout (2001), who examined the devoicing of /v, z, g/ in two varieties of Dutch.

However, in spite of partial devoicing, listeners are able to parse the stimuli with a voiced phonological category (Widdison 1995, 1997, p. 260; Lavoie 2001, p. 107), since they may ignore this cue and rely on other indicators to repair the ambiguity (Ohala 1981, 2012). After examining Polish and German sibilants, Żygis et al. (2012, pp. 323–24) show that the voicing contrast, both in word-initial and medial positions, relies mostly on durational parameters and the measurement of intraoral pressure. Voiced sibilants display a significantly shorter duration and lower intraoral pressure peaks. Their findings are in line with the literature, which coincides in focusing on duration as a highly reliable cue for detecting devoicing (Haggard 1978, p. 101; Lavoie 2001, p. 107).

This phonetic description of the phenomenon needs a final comment. Two important ideas arise from the previous explanation: (a) there is variation in the production of voiced sibilants (Widdison 1995; Smith 1997; Van de Velde and van Hout 2001; Benet et al. 2012), and (b) the listener has a key role in the interpretation of the signal, since he/she will decide which phoneme to interpret in the phonological parsing stage. These elements are the ingredients that pave the way for language change.

1.3. On Sound Change

As known, speech production entails phonetic variation, at least to a certain extent. Allophony can be conditioned by a number of factors, including speaking style, phonological context and prosodic conditions, as well as the need to guarantee communicative success by ensuring the intelligibility of the signal (Lindblom 1990; Blevins 2004; Ohala

1981, 2012). Phonetic variation, however, needs to be correctly parsed with the appropriate phonological category. Utterances more similar to the canonical form for a category do not represent a problem for their recognition and interpretation as members of that category, but exemplars which differ from this prototype and are more peripheral in its acoustic-auditory space may be ambiguous. They may even fall in the intersection of the cloud of exemplars of neighbouring phonological categories. In these cases, reanalysis and recategorization may take place. In other words, when the listener reinterprets the non-canonical input as a different category, a sound change can be initiated (Pierrehumbert 2001, 2002; Blevins 2004; Ohala 1981, 2012).

There are some extra factors that can favour the change. One of them is the frequency of usage: when peripheral exemplars become more frequent, they may gradually be identified as variants central to their category and bring about a displacement from central values in the cloud of exemplars: what was once regarded as canonical will no longer be considered so (Pierrehumbert 2001). This preference for peripheral variants can be related to the functional load of phonological contrasts: in the case of oppositions yielding a high functional load, it is important to preserve the difference between the phonological categories in order to assure communicative success. This implies the selection of phonetic variants that display greater distinction between the members of the phonological pair, speakers therefore preferring allophones with more exaggerated phonetic cues to avoid misparsing. However, in cases where the functional load is low (i.e., when the opposition is not productive), the maximization of contrast is not that important and peripheral forms can be selected without endangering communication (Wedel et al. 2013, p. 184). This could have been the initial step in voicing neutralization in Early Modern Spanish sibilants.

Another interesting aspect of sound change which has been discussed in devoicing cases is whether it corresponds to a lenition or a fortition process. Although it has usually been associated with strengthening mechanisms (Alarcos 1988, pp. 51–52; Lavoie 2001, pp. 27, 43; Blevins 2004, pp. 110–11; Lahoz 2015; Bybee and Easterday 2019, p. 270), its adscription is not always clear. Bybee and Easterday (2019, p. 271) define lenition and fortition in articulatory terms, as a reduction or increase in the magnitude and duration of articulatory movements. As mentioned before, Lavoie (2001, pp. 27, 43) generally considers devoicing phenomena as fortition processes, though she admits it may be a way to simplify production in certain contexts (Lavoie 2001, p. 107), in agreement with Smith (1997, pp. 494–96). Indeed, when occurring in weak prosodic positions (such as word-final or sentence-final positions), devoicing entails gesture magnitude decrease. This is also the case with neighbouring voiceless sounds. In the former case, the ultimate cause is the reduction in the transglottal air flow, which permits friction but prevents vocal-fold vibration; in the latter, less glottal adduction (a more open glottis) is responsible for unvoicing.

Intervocalic unstressed contexts have been associated with hypoarticulation conditions and, therefore, with weakening processes (Lindblom 1990, pp. 404–5; Lavoie 2001, p. 168; Bybee and Easterday 2019, p. 288). Nevertheless, in such contexts voicing is deemed to be the normal lenition phenomenon instead of devoicing, which has been regarded as a case of strengthening (Alarcos 1988, pp. 51–52; Jiménez and Lloret 2014). Even Haggard (1978, p. 100) or Davidson (2016, p. 40) show that fully voicing in obstruents is attested after stressed vowels, while devoicing cases are more common when the obstruent precedes a stressed vowel. However, if the explanations provided by Smith (1997) and Lavoie (2001) are right, sibilant devoicing in Spanish would satisfy the requirements of lenition, since it involves a reduction in the magnitude of gestures. This would provide the basis to challenge the traditional account of this historical change.

1.4. Scope and Hypotheses

In order to verify whether these phonetically based explanations account for historical sibilant devoicing, we assume that synchronic variation is related to or parallels diachronic sound change (Ohala 2003, p. 672; Blevins 2004 or Harrington 2012, p. 322), so that the

instrumental study of speech can be profitably used for a better understanding of linguistic changes. The main problem in this perspective is how to deal with phonological systems which have disappeared through language evolution. This seems to be the case for Spanish sibilants, at least if we focus on standard Spanish. Blevins (2004) argues that comparative grammar, dialectology or even the study of language acquisition are useful in these cases: diachronic studies should essentially be multidisciplinary.

Widdison (1995) is an example of this. In this research the author checks the perceptive aspects of the sibilant devoicing process under examination here by manipulating the duration and intensity cues of [ʃ] and [s] in two experiments, one with stimuli from Argentinian Spanish and the other from Mexican Spanish. However, the participants in these perception tasks were all Spanish speakers, whose system lacked the phonological contrast between voiced and voiceless categories.⁷ To avoid this problem, we decided to focus on comparative grammar instead of dialectology, since the literature points out the similarity between the Medieval Spanish sibilant system and the current Catalan sibilants (Penny 1993; Bradley and Delforge 2006; Jiménez and Lloret 2014).

Standard Catalan displays four pairs of sibilant consonants: two fricative pairs (alveolar/s, z/ and alveopalatal/ʃ, ʒ/) and two affricate pairs (dentoalveolar/tʃ, dʒ/⁸ and palatal/tʃ̟, dʒ̟/) which contrast in word-initial and medial onset position.⁹ They present dialectal variation at the phonetic level. Importantly for the current study, devoicing phenomena have also been detected, basically affecting affricates (in Central Valencian Catalan and in the Southern Catalonia variety) but also fricatives (Veny 1998, pp. 31, 64; Wheeler 2005, pp. 11–22; Recasens 2014, pp. 239–45, 251–53, 263–64; see also Hualde and Prieto 2014, p. 111; Hualde et al. 2015, p. 246).¹⁰ To achieve a better fit to the characteristics of Medieval Spanish and avoid dialect effects regarding devoicing, this work will draw its attention to Central Catalan from Girona and Majorcan Catalan.

In short, the core purpose of this paper is to examine in greater depth the hypothesis in Pensado (1993) and Widdison (1995, 1997), who stated that sibilant devoicing in Early Modern Spanish can be accounted for in terms of phonetic aspects of speech. As commented on before, we applied an experimental approach to a close Ibero-Romance language, Catalan, to cover the two parts of the process: production and perception. Two different experiments are thus proposed. The first one is an acoustic study that will allow us to describe the relevant features regarding voicing of each voiced–voiceless contrastive pair. In this sense, it will be necessary to determine if sibilant production is homogeneous or if there is variation (Haggard 1978). The second experiment is an identification task designed to test the parsing of different stimuli (voiced, partly devoiced or voiceless) to voiced or voiceless categories. This will allow us to check whether devoicing can be detected perceptually and, if so, to what extent this occurs in each of the contrasting sibilant pairs.

According to the literature, we expect sibilant production to display variation, so that sibilant allophones will present different degrees of voicing, which will be measurable by means of duration, intensity and vocal-fold activity. In this sense, shorter duration, lower intensity and a higher degree of vocal-fold vibration should point to voiced segments; while the contrary should signal devoicing, which is expected to be highly variable. These acoustic parameters are thus assumed to be correlated. Not all sibilants will show identical tendencies: most probably, affricates will be more prone to lenition processes leading to devoicing than fricatives, and palatals more than alveolars.

At the perception level, listeners will be able to detect the aforementioned acoustic cues, so that phonologically voiced segments will be identified as voiceless if their phonetic features differ from the prototypical characteristics of fully voiced sounds. As a result, the diachronic change could be deemed to have a phonetic basis, like most regular sound changes.

2. The Acoustic Approach

2.1. Method

The acoustic experiment investigates the characteristics regarding voicing of the sibilants /s, z/, /ʃ, ʒ/, /tʃ, dʒ/ and /tʃ̺, dʒ̺/ in Central and Majorcan Catalan. Since Catalan displays four pairs of sibilants, we decided to examine all of them, especially if we remember that Spanish also has the voiceless palatal affricate and that its voiced counterpart also existed in Medieval Spanish, even though it had no phonological status. Another reason to include the palatal affricate pair is the claim that affricates and palatals undergo devoicing earlier than other sibilants (Pensado 1993, pp. 209–10). The other central aim of the experiment is to determine whether devoicing is possible in the same position as in Early Modern Spanish, i.e., in medial intervocalic context. Thus, only this position has been taken into consideration, except for /tʃ̺/: as commented before, this consonant is very rare in this position (see note 8), so the utterances with this segment were intervocalic but in word-initial position with no speech interruption or pause between the preceding vowel and the sibilant. On the other hand, in order to avoid vowel effects on the consonant which could skew the results, the vowels surrounding the sibilants were restricted to the central ones (/a/ and /ə/).

Stress could also have been considered as a factor in the corpus configuration, since unstressed position has been claimed to favour lenition (Lavoie 2001, p. 168; Bybee and Easterday 2019, p. 288). In fact, as Davidson (2016, p. 36) comments, lexical stress seems to play a role in voicing implementation (see also Haggard 1978, p. 96). However, we finally rejected stress as a factor because it was not clearly relevant in Smith (1997, p. 490) and not relevant at all in a first statistical approach to our data.¹¹

Language change occurs in casual speech. However, using a spontaneous speech corpus would have implied major difficulties in controlling the experimental variables. For this reason and taking into account that this is a first approach to the topic, a reading task was preferred. The selected speakers had to read aloud twice 23 short sentences (8 of them distractors), which contained words including the eight Catalan sibilants. Some examples are given in the sentences below and the whole list is offered in Appendix A.

- a. La tieta es va *casar* amb un senyor de Capdepera.
[‘Auntie married a man from Capdepera’]
- b. Li agradava moltíssim anar a la *platja* el mes de setembre.
[‘He/She adored going to the beach in September’]

A total of 16 female speakers aged between 22 and 68 were recruited for this task. All of them had university education and none reported speech production problems. Ten of them were from Girona (Northern Catalonia), so they were Central Catalan speakers, while the other six were Majorcan (thus Majorcan Catalan speakers). It was considered interesting for the purpose of the research to examine whether there were differences in the production of sibilants depending on the dialect. Neither Central Catalan in Girona nor Majorcan Catalan displayed devoicing phenomena, but Majorcan Catalan tends to strengthen sibilant fricatives to affricates in medial position (Bibiloni 2016, pp. 127–28), which implies an increase in duration. Since duration is one of the relevant cues in devoicing, it will be pertinent to check whether this tendency can involve a greater degree of devoicing in Majorcan speakers.

It is important to note that all speakers identified Catalan as their L1, and declared that they were dominant in this language, with no influence of Spanish pronunciation. To confirm this, they had to answer the *Bilingual Language Profile* questionnaire (BLP, Birdsong et al. 2012), which includes aspects related to linguistic background, usage of languages, competence and linguistic attitudes of the participants. Final scores between –110 and 110 indicate that the participants are balanced bilinguals; negative results point to a preference for Spanish and positive results imply a bias in favour of Catalan. As can be observed in Table 1, the mean score was 86.309 for Girona speakers and 40.980 for Majorcan ones: all of them were balanced bilinguals and, crucially, they were all clear Catalan dominants, though

the bias in favour of Catalan is significantly stronger for Girona speakers ($t(12) = -2.347$, $p < 0.037$).

Table 1. Mean scores in the *BLP* questionnaire for Girona speakers and Majorcan speakers. Results from each section and final scores are shown.

<i>BLP</i> Section	Girona Speakers	Majorca Speakers
II. Linguistic History	17.659	12.938
III. Linguistic Use	47.088	21.232
IV. Linguistic Competence	3.629	5.107
V. Linguistic Attitudes	17.933	1.702
Total score	86.309	40.980

Recordings were made in a quiet echo-free office in the Spanish Philology Department of the University of the Balearic Islands in the case of Majorcan Catalan speakers, and in a silent echo-free room adapted for the purposes of this research in a private home in Girona in the case of Central Catalan speakers. In both cases, a Samson C01U Pro microphone (Samson Technologies, Hicksville, NY, USA) connected to an Acer laptop (Acer Inc., Taipei, Taiwan) was used. The microphone was placed approximately 20–25 cm from the speakers, on the same table they were seated at. The recording software was Praat (v. 6.0.40, [Boersma and Weenink 2018](#)), with a sampling frequency of 44,100 Hz. Spectrograms and oscillograms were analysed using the same software. Segmentation of the sibilant sounds was done manually. The sibilant intervals started at the point where friction began and finished where periodicity from the following vowel initiated in the oscillogram in the case of fricatives; for affricates, the occlusive phase was also included in the interval, since the total duration of the segment is one of the properties examined here.

We set three acoustic parameters as dependent variables: consonant duration, the intensity of the consonant, and the fraction of unvoiced frames in the sibilant. As mentioned before, duration is one of the more consistent cues of voicing, as well as its intensity and vocal-fold vibration ([Haggard 1978](#); [Pensado 1993](#); [Widdison 1995](#)). Regarding the latter, since this is an acoustic experiment, we decided to use the Voice Report function in Praat, which, among other things, measures the part of a period in which glottal pulses cannot be detected, thus offering the portion of the sound selection which is considered unvoiced. These measurements are deemed reliable for detecting the degree of voicing, according to [Eager \(2015\)](#) and [Davidson \(2016\)](#). To extract all these parameters, we adapted a pre-existing Praat script ([Elvira-García 2013](#)), following the specifications in [Eager \(2015, p. 3\)](#) concerning the pitch threshold in women’s speech (100–300 Hz) and time step (0.001 s). In order to offer a qualitative description of the possible phonetic variants, the degree of voicing is also considered a dependent variable. This variable has been set according to [Smith \(1997, p. 478\)](#), allophones being considered as fully voiced when the devoicing portion did not exceed 10% of the sound, partially devoiced when it accounted for 11 to 75% of the sibilant, and fully devoiced (or unvoiced) if it surpassed 75% of the sound.

Voicing value (voiced or voiceless category), dialectal variant of the speakers (Girona Central Catalan and Majorcan Catalan) and sibilant (every single category by point-manner of articulation combination) were established as the predictor variables. We included 2 instances in the corpus for each factor, so, as can be seen in [Table 2](#), we analysed 32 sibilants per speaker (4 dentoalveolars, 4 alveolars, 4 alveopalatals, and 4 palatals which were read twice), which yields a total of 512 tokens (32×16 speakers).

Table 2. Number of cases in the corpus for each sibilant, and number of utterances read by each speaker.

Sibilant Type		Girona Central Catalan	Majorcan Catalan
Alveolar sibilants	voiceless [s]	2	2
	voiced [z]	2	2
Alveopalatal sibilants	voiceless [ʃ]	2	2
	voiced [ʒ]	2	2
Dentoalveolar sibilants	voiceless [tʃ]	2	2
	voiced [dʒ]	2	2
Palatal sibilants	voiceless [tʃ]	2	2
	voiced [dʒ]	2	2
Total per speaker:		16 × 2 = 32	

Generalized linear mixed-effects statistical analyses were performed for each dependent variable using the IBM SPSS software package (v. 25.0, IBM Corporation, Armonk, NY, USA), with main effects (dialect and voicing) and interactions (dialect × voicing) for each pair of sibilants. Speaker and word were entered as random effects. In addition, a multinomial mixed-effects logistic regression was carried out for degree of voicing with sibilant and dialect as main effects, and interactions between sibilant and dialect. Random effects were also speaker and word. In all cases, the significance level was set at 0.05.

2.2. Results

This section shows the results obtained. First, data concerning every pair of sibilants will be given first, with respect to fricatives (Section 2.2.1) and then regarding affricates (Section 2.2.2), which will allow a description of the voicing implementations in each member of the pair. Second, comparative results across the whole set of sibilants will be commented on (Section 2.2.3). The statistical results are presented in Table 3 and will be explained in detail in the following subsections.

2.2.1. Fricative Sibilants

Results from fricative sibilants will be addressed first. Descriptive data concerning duration, intensity and fraction of unvoiced frames are presented in Table 4. Overall, voiceless fricatives have a longer duration, greater intensity and a higher proportion of unvoiced frames than their voiced counterparts. This general tendency is observed in both dialects examined. It is interesting to note that the results for voiced fricatives display more variability than those for voiceless fricatives, particularly in the fraction of unvoiced frames, but also in duration in the case of [ʒ], which is the most variable segment in this aspect. This suggests more heterogeneous realizations regarding these parameters. [z] shows more variation in the unvoiced frames portion not only globally, but also if inspected by dialect.

Inspection of the alveolar pair reveals that duration and fraction of unvoiced frames depend on the voicing value ($F(1, 2) = 89.459, p < 0.009$ and $F(1, 2) = 181.648, p < 0.003$). [s] is significantly longer (36.03 ms. on average; $b_0 = 48.143, se = 4.255, t = 11.316, p < 0.001$) and has a higher portion of unvoiced frames (54.71% more; $b_0 = 63.717, se = 4.904, t = 12.993, p < 0.0001$) than [z]. There are no significant differences between these consonants concerning intensity. However, dialect does condition intensity ($F(1, 14) = 5.294, p < 0.037$): in Majorcan Catalan, alveolar sibilants display lower intensity ($b_0 = -5.742, se = 2.218, t = -2.589, p = 0.019$) than in Girona Central Catalan. Neither word nor speaker show an effect on the results, except for intensity as a main effect ($se = 6.431, Z = 2.421, p < 0.015$). When we examine the interaction between voicing value and dialect, we obtain important information: it is statistically relevant when it comes to duration ($F(1, 108) = 20.445, p < 0.0001$) and the fraction of unvoiced frames ($F(1, 108) = 5.352, p < 0.023$). In Majorcan Catalan, [s] is significantly shorter than in Girona Central Catalan (21.24 ms less; $b_0 = -24.211, se = 5.354, t = -4.522, p < 0.0001$) and it also displays a lower proportion of unvoiced frames (7.8% less; $b_0 = -17.997, se = 7.779, t = -2.313, p < 0.023$). Furthermore, Majorcan Catalan [z]

has a significantly higher proportion of unvoiced frames (10.19% more) and less intensity (5.74 dB less) than Girona Central Catalan [z].

Table 3. Generalized mixed-effects linear model results for each sibilant pair and in cross-category comparison. Significant results are in bold.

Variables		Type of Sibilant Pair			
		Alveolar	Alveopalatal	Dentoalveolar	Palatal
Voicing	Duration	F(1, 2) = 89.459, <i>p</i> = 0.009	F(1, 2) = 6.922, <i>p</i> = 0.114	F(1, 110) = 14.973, <i>p</i> < 0.0001	F(1, 2) = 17.292, <i>p</i> = 0.051
	Intensity	F(1, 2) = 4.092, <i>p</i> = 0.179	F(1, 2) = 2.400, <i>p</i> = 0.252	F(1, 2) = 1.738, <i>p</i> = 0.312	F(1, 110) = 26.626, <i>p</i> < 0.0001
	Fraction of unvoiced frames	F(1, 2) = 181.648, <i>p</i> < 0.003	F(1, 109) = 386.661, <i>p</i> < 0.0001	F(1, 2) = 32.797, <i>p</i> < 0.023	F(1, 2) = 88.786, <i>p</i> < 0.009
Dialect	Duration	F(1, 14) = 3.249, <i>p</i> = 0.093	F(1, 14) = 1.168, <i>p</i> = 0.298	F(1, 14) = 1.536, <i>p</i> = 0.236	F(1, 14) = 1.599, <i>p</i> = 0.227
	Intensity	F(1, 14) = 5.294, <i>p</i> < 0.037	F(1, 14) = 3.681, <i>p</i> = 0.076	F(1, 14) = 1.395, <i>p</i> = 0.257	F(1, 14) = 4.172, <i>p</i> = 0.060
	Fraction of unvoiced frames	F(1, 14) = 0.026, <i>p</i> = 0.873	F(1, 14) = 8.152, <i>p</i> < 0.013	F(1, 14) = 0.098, <i>p</i> = 0.759	F(1, 14) = 13.440, <i>p</i> < 0.003
Voicing × dialect	Duration	F(1, 108) = 20.445, <i>p</i> < 0.0001	F(1, 107) = 4.316, <i>p</i> < 0.040	F(1, 110) = 0.113, <i>p</i> = 0.737	F(1, 108) = 8.555, <i>p</i> < 0.004
	Intensity	F(1, 108) = 1.849, <i>p</i> = 0.177	F(1, 107) = 0.387, <i>p</i> = 0.535	F(1, 108) = 0.443, <i>p</i> = 0.507	F(1, 110) = 37.320, <i>p</i> < 0.0001
	Fraction of unvoiced frames	F(1, 108) = 5.352, <i>p</i> < 0.023	F(1, 109) = 7.191, <i>p</i> < 0.008	F(1, 108) = 1.961, <i>p</i> = 0.164	F(1, 108) = 8.464, <i>p</i> < 0.004
Global cross-category comparison					
Phoneme	Duration	F(7, 9) = 51.970, <i>p</i> < 0.0001			
	Fraction of unvoiced frames	F(7, 9) = 82.209, <i>p</i> < 0.0001			
Dialect	Duration	F(1, 14) = 3.440, <i>p</i> = 0.085			
	Fraction of unvoiced frames	F(1, 14) = 2.072, <i>p</i> = 0.172			
Phoneme × dialect	Duration	F(7, 473) = 2.771, <i>p</i> < 0.008			
	Fraction of unvoiced frames	F(7, 473) = 6.640, <i>p</i> < 0.0001			

Table 4. Mean value and standard deviation for the dependent variables. Duration values are offered in milliseconds, intensity in decibels, and the fraction of unvoiced frames is presented as a percentage of the sound’s duration.

Variables		[s]		[z]		[ʃ]		[ç]	
		\bar{x}	sd	\bar{x}	sd	\bar{x}	sd	\bar{x}	sd
Global values	Duration	104.59	19.13	65.52	17.25	110.74	17.37	92.19	26.89
	intensity	45.07	5.57	41.67	5.82	43.62	5.89	42.36	5.57
	Fraction of unvoiced frames	81.50	11.68	24.53	32.50	73.87	15.42	21.79	24.60
Majorcan Catalan	Duration	91.31	17.31	67.38	16.31	101.62	15.78	92.36	23.93
	intensity	42.54	6.18	38.08	6.46	40.58	6.01	39.69	6.03
	Fraction of unvoiced frames	76.62	12.18	30.90	35.20	67.13	15.50	5.76	8.41
Girona Central Catalan	Duration	112.55	15.51	64.41	17.90	116.22	16.08	92.02	28.75
	intensity	46.60	4.61	43.82	4.17	45.45	5.06	43.90	4.71
	Fraction of unvoiced frames	84.43	10.47	20.71	30.58	77.91	14.05	31.01	26.15

Alveopalatals only display significant differences between voiced and voiceless category in the fraction of unvoiced frames ($F(1, 109) = 386.661, p < 0.0001$), which is statistically higher in [ʃ] (54.30% more; $b_0 = 46.091, se = 3.358, t = 13.969, p < 0.0001$) than in [ç]. Dialect also affects this parameter ($F(1, 14) = 8.152, p < 0.013$): Majorcan Catalan alveopalatals

display a lower proportion of unvoiced frames (18.18% less; $b_0 = -25.587$, $se = 9.957$, $t = -3.678$, $p < 0.002$) than Girona Central Catalan alveopalatals. No significant results were detected in duration or intensity values. The speaker affects the model regarding both intensity and fraction of unvoiced frames. There are significant interactions between voicing value and dialect concerning duration ($F(1, 107) = 4.316$, $p < 0.040$), apart from fraction of unvoiced frames ($F(1, 109) = 7.191$, $p < 0.008$). [ʃ] is significantly shorter (14.60 ms less; $b_0 = -14.643$, $se = 7.048$, $t = -2.078$, $p < 0.040$) and has a lower proportion of unvoiced frames (25.58% less; $b_0 = 14.812$, $se = 5.524$, $t = 2.682$, $p < 0.008$) in Majorcan Catalan. This seems to point to a more voiced realization compared to [ʃ] in Girona Central Catalan and, thus, less clear-cut differences between the two alveopalatal phonemes in Majorcan Catalan.

2.2.2. Affricate Sibilants

Descriptive data for affricates are shown in Table 5. As in fricatives, voiceless affricates display longer duration and a higher proportion of unvoiced frames, though when it comes to intensity, dentoalveolars and palatals do not seem to fit: the former have the same behaviour as fricatives ([tʃ] is more intense), but palatals do not, since the voiced segment shows more intensity than the voiceless one. This deviation is due to the results in Majorcan Catalan, a dialect in which [dʒ] has more intensity than [tʃ]. It is important to highlight some interesting information obtained from standard deviation values. In general, unlike fricative sibilants, voiced categories do not present more variability. In this case, [tʃ] is the most heterogeneous segment in almost every parameter, not only if we consider the whole sample, but also if we take into account the dialect factor. However, we can see that while there is great variability regarding fraction of unvoiced frames in Majorcan Catalan, there is great homogeneity in Girona Central Catalan. Possibly, the word position in this category is playing a role in its results, since it is the only case in word-initial intervocalic context, as we explained before.

Table 5. Mean value and standard deviation for the dependent variables. Duration values are offered in milliseconds, intensity in decibels, and fraction of unvoiced frames is presented as a percentage of the sound’s duration.

Variables		[tʃ]		[dʒ]		[tʃ]		[dʒ]	
		\bar{x}	sd	\bar{x}	sd	\bar{x}	sd	\bar{x}	sd
Global values	Duration	149.87	38.91	130.05	25.01	147.98	21.99	117.90	24.78
	intensity	38.31	10.27	35.98	6.01	31.09	7.94	34.92	7.71
	Fraction of unvoiced frames	66.95	25.79	38.73	23.26	64.14	14.89	30.85	21.47
Majorcan Catalan	Duration	140.64	33.87	122.94	27.45	148.57	23.73	105.96	24.34
	intensity	36.79	10.21	33.41	4.19	24.16	5.00	36.04	7.72
	Fraction of unvoiced frames	66.24	32.05	43.40	29.60	59.74	17.77	16.17	19.74
Girona Central Catalan	Duration	155.40	41.06	134.33	22.72	147.63	21.19	125.07	22.40
	intensity	39.22	10.32	37.52	6.44	35.25	6.31	34.25	7.73
	Fraction of unvoiced frames	67.38	3.42	35.92	18.33	66.78	12.35	39.65	17.40

The results for dentoalveolar affricates indicate that friction duration and the fraction of unvoiced frames are affected by the voicing quality of the consonant ($F(1, 110) = 14.973$, $p < 0.0001$ and $F(1, 2) = 32.797$, $p < 0.023$): [tʃ] is significantly longer (19.39 ms.; $b_0 = 21.078$, $se = 6.137$, $t = 3.434$, $p < 0.001$) and displays a higher proportion of unvoiced frames (27.14% more; $b_0 = 31.466$, $se = 5.034$, $t = 6.251$, $p < 0.007$) than [dʒ]. Intensity does not significantly change. Dialect does not affect the voicing cues and no interaction has been detected. However, the speaker affects both duration ($se = 158.352$, $Z = 2.043$, $p < 0.041$) and the fraction of unvoiced frames ($se = 145.542$, $Z = 2.399$, $p < 0.016$).

Neither voicing nor dialect, as main effects, condition the duration of palatal sibilants. However, the voicing value affects both intensity ($F(1, 110) = 26.626, p < 0.0001$) and fraction of unvoiced frames ($F(1, 2) = 88.786, p < 0.009$). $[d̪̺]$ displays more intensity (5.43 dB more; $b_0 = 34.250, se = 1.535, t = 22.312, p < 0.0001$) and a lower proportion of unvoiced frames (35.34% less; $b_0 = 39.658, se = 3.541, t = 11.199, p < 0.0001$) than $[tʃ]$. Dialect also has an effect on this last parameter ($F(1, 14) = 13.440, p < 0.003$): overall, Majorcan Catalan palatal sibilants show a significantly lower percentage of unvoiced frames (15.26% less; $b_0 = -23.484, se = 5.032, t = -4.667, p < 0.0001$). The speaker influences duration ($se = 73.39, Z = 2.036, p < 0.042$) and intensity ($se = 7.354, Z = 2.072, p < 0.038$). Interactions between the two factors have been detected for all the dependent variables: duration ($F(1, 108) = 8.555, p < 0.004$), intensity ($F(1, 110) = 37.320, p < 0.0001$) and fraction of unvoiced frames ($F(1, 108) = 8.464, p < 0.004$). In Majorcan Catalan, $[tʃ]$ is less intense (11.08 dB less; $b_0 = -12.875, se = 2.108, t = -6.109, p < 0.0001$) and displays a lower proportion of unvoiced frames (7.04% less; $b_0 = 16.438, se = 5.650, t = 2.909, p < 0.004$), while Majorcan $[d̪̺]$ is shorter (19.11 ms. less, $sd = 7.960, t = -2.401, p < 0.026$) and shows a lower fraction of unvoiced frames (23.48% less, $sd = 5.032, t = -4.667, p < 0.0001$) when compared to their Girona Central Catalan counterparts. These results suggest that $[d̪̺]$ is more clearly voiced in Majorcan Catalan than in Girona Central Catalan.

As can be observed from the above figures, fraction of unvoiced frames is the only parameter which is constant in distinguishing between realizations of voiced and voiceless sibilants. Nevertheless, duration is also involved in many cases, so a certain relationship between these two parameters might be inferred. To check this impression, we carried out a generalized mixed-effects linear model for each sibilant pair (duration x fraction of unvoiced frames, speaker and word as random effects). The results prove a direct relationship between the two variables in all sibilant pairs,¹² except for alveopalatal fricatives (see Figure 1). In all cases, the speaker is relevant to the statistical model. In view of these results, from now on we will focus on the fraction of unvoiced frames as a reliable cue to detect the degree of voicing and, as a subsidiary indicator, duration.

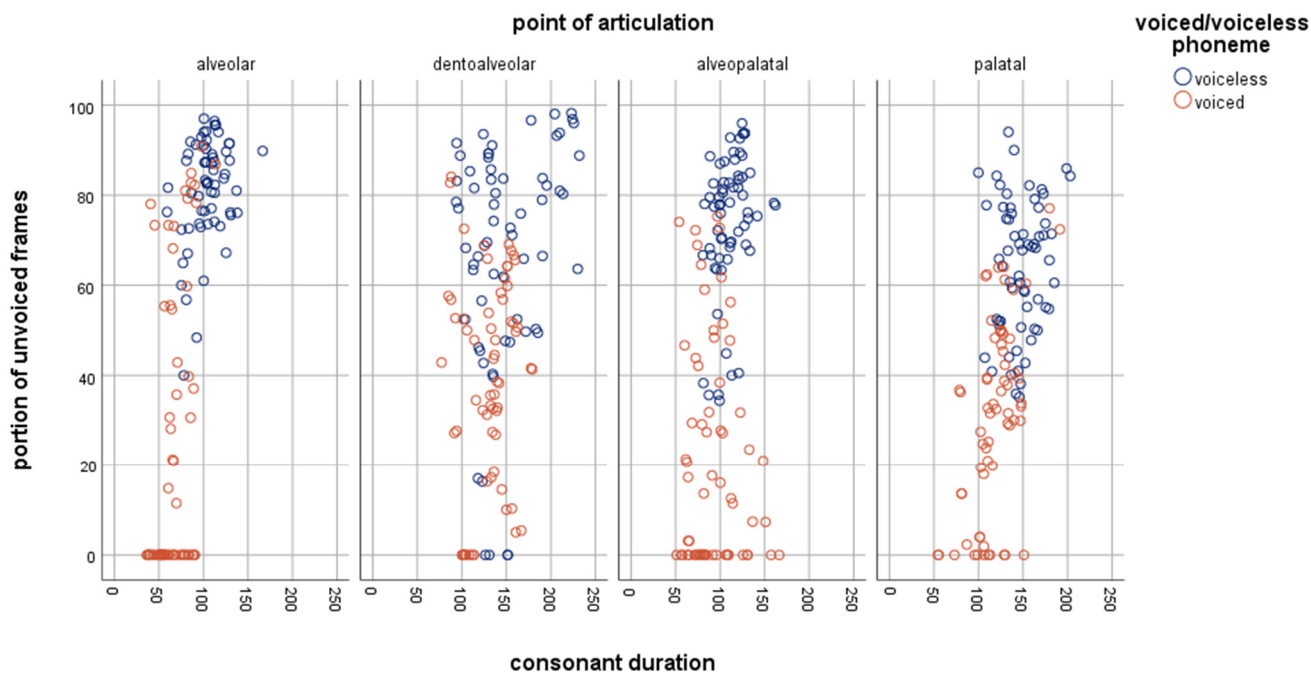


Figure 1. Scatter plots showing the relationship between proportion of unvoiced frames and duration and the voicing value in the four sibilant pairs.

2.2.3. Cross-Category Comparison

An overall comparison among all the categories has been carried out for both duration and fraction of unvoiced frames (sibilant and dialect as fixed effects, speaker and word as random variables). The model reveals a significant effect of the sibilant on duration ($F(7, 9) = 51.970, p < 0.0001$) and on fraction of unvoiced frames ($F(7, 9) = 82.209, p < 0.0001$). Dialect does not affect these parameters. In all cases, the speaker conditions the results ($se = 38.033, Z = 2.234, p < 0.025$, and $se = 36.127, Z = 2.321, p < 0.020$). Interaction of the two factors has been found for the two dependent variables ($F(7, 473) = 2.771, p < 0.008$, and $F(7, 473) = 6.640, p < 0.0001$).

As for the fixed effects, both duration and fraction of unvoiced frames distinguish all the categories¹³ except in a few cases. [s] displays similar duration to the fricative sibilants, [ʃ] is equivalent to [dʒ] and [tʃ] is also similar to [tʃ]. Regarding the fraction of unvoiced frames, [z] shows no significant difference from [ʒ] and [dʒ], nor do [s] from [ʃ] and [tʃ] from [tʃ]. In the remaining cases, voiceless counterparts have significantly longer duration and higher rates of unvoiced frames. It is worth noting that, while there is no overlap between voiced and voiceless segments if we consider fraction of unvoiced frames, there is a degree of overlap if we use duration as the criterion.

At this point, if Smith’s (1997, p. 478) classification of the degree of voicing is applied, overlapping seems to be general across categories in the intermediate range (see Table 6, Figure 2). In fact, almost all the categories present fully voiced utterances, fully devoiced ones and, more importantly for our purposes, partially devoiced realizations. Only alveolar fricatives and voiceless alveopalatals tend to be clearly realized at the extremes (fully voiced or fully devoiced), instances of the other categories being concentrated in an intermediate range from 11% to 75% of unvoiced frames. This still holds true when examined by dialect, apart from [ʃ] in Majorcan Catalan.

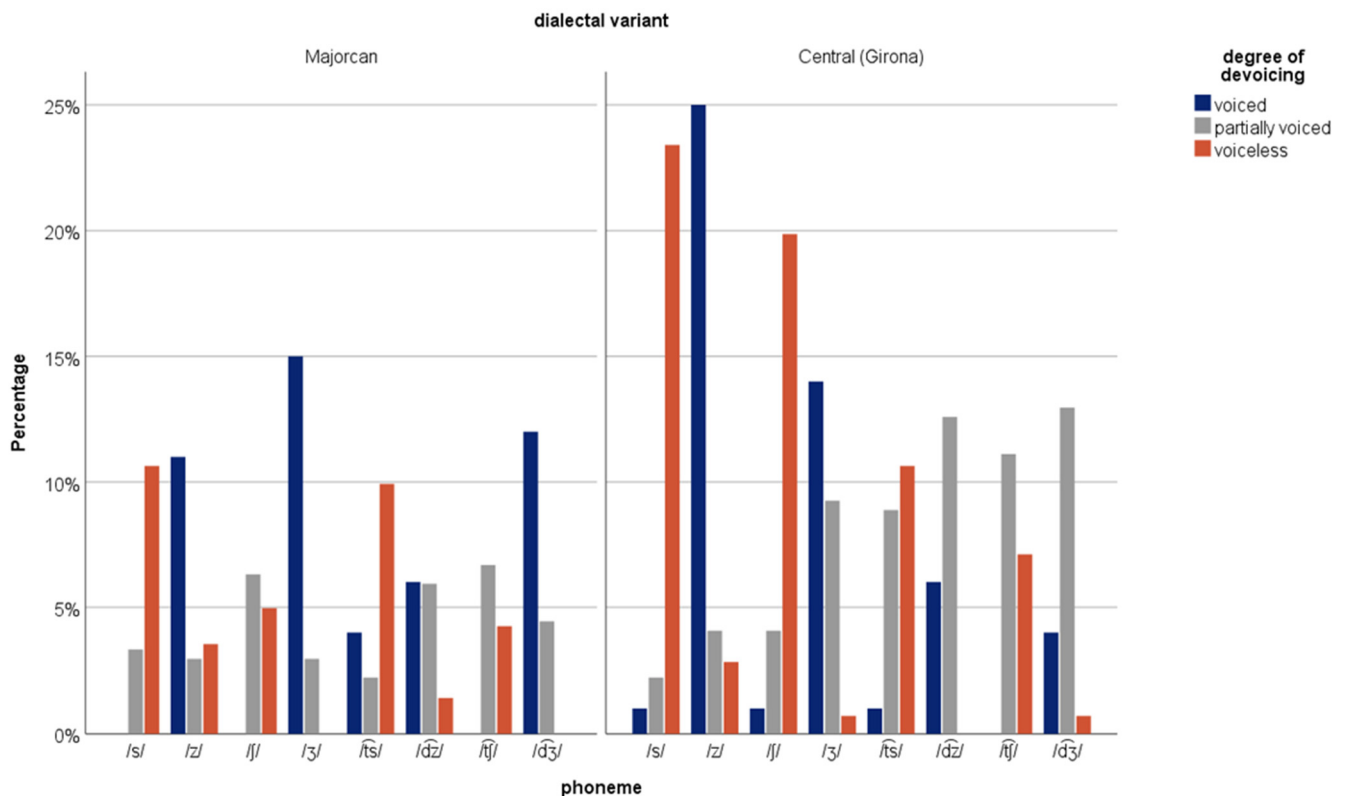


Figure 2. Bar plots showing the distribution of the three variants in Smith (1997) for each phonological category in Girona Central Catalan and in Majorcan Catalan.

Table 6. Number of instances and percentage of occurrence of the three variants determined by Smith (1997) for each sibilant category.

Degree of Voicing		[s]	[z]	[ʃ]	[ʒ]	[tʃ]	[dʒ]	[tʃ]	[dʒ]
Voiced		1 (1.6%)	36 (56.3%)	1 (1.6%)	29 (46%)	5 (5%)	12 (18.8%)	-	16 (25%)
Partially Devoiced		15 (23.4%)	19 (43.8%)	28 (43.8%)	33 (46.9%)	30 (46.9%)	50 (78.1%)	48 (75%)	47 (73.4%)
Unvoiced		48 (75%)	9 (14.1%)	35 (54.7%)	1 (1.6%)	29 (45.3%)	2 (3.1%)	16 (25%)	1 (1.6%)
Majorcan Catalan	voiced	-	11 (45.8%)	-	15 (65.2%)	4 (16.7%)	6 (25%)	-	12 (50%)
	partially devoiced	9 (37.5%)	8 (33.3%)	17 (70.8%)	8 (34.8%)	6 (25%)	16 (66.7%)	18 (75%)	12 (50%)
	unvoiced	15 (62.5%)	5 (20.8%)	7 (29.2%)	-	14 (58.3%)	2 (8.3%)	6 (25%)	-
Girona Central Catalan	voiced	1 (2.5%)	25 (62.5%)	1 (2.5%)	14 (35%)	1 (2.5%)	6 (15%)	-	4 (10%)
	partially devoiced	6 (15%)	11 (27.5%)	11 (27.5%)	25 (62.5%)	24 (60%)	34 (85%)	30 (75%)	35 (87.5%)
	unvoiced	33 (82.5%)	4 (10%)	28 (70%)	1 (2.5%)	15 (37.5%)	-	10 (25%)	1 (2.5%)

Statistical analysis confirms these observations. The model indicates that the sibilant category ($F(14, 12) = 2.957, p < 0.033$) and the interaction between category and dialect ($F(14, 479) = 2.523, p < 0.002$) have an effect on the degree of voicing. Fully devoiced variants are significantly more frequent in [s] ($b_0 = 5.321, se = 1.603, t = 3.319, p < 0.001$), [ʃ] ($b_0 = 5.078, se = 1.601, t = 3.171, p < 0.002$) and even [tʃ] ($b_0 = 4.244, se = 1.609, t = 3.558, p < 0.002$), while partially devoiced variants are significantly more common in [dʒ] ($b_0 = 2.222, se = 0.625, t = 3.319, p < 0.001$). On the other hand, partially devoiced cases are significantly less common than expected in [z] ($b_0 = -3.091, se = 0.759, t = -4.072, p < 0.001$) and, at the limit of significance, [ʒ] ($b_0 = -1.621, se = 0.742, t = -2.184, p < 0.049$). As commented, dialect as a main effect is not relevant, but if we break down the effect, Majorcan Catalan seems to display fewer partially devoiced realizations than Girona Central Catalan ($b_0 = -2.225, se = 0.732, t = -3.038, p < 0.003$). However, Majorcan Catalan [z] is more prone to partially devoiced variants than Girona Central Catalan ($b_0 = 2.762, se = 0.907, t = 3.047, p < 0.002$): the average of this kind of realizations, in fact, is significantly higher (33.3% vs. 27.5%). These figures confirm the general data found in the quantitative results.

3. The Perception Experiment

3.1. Method

It has been deemed important to complete this contribution with a preliminary survey of the perception part of the process. As seen before, most of the literature focuses not only on the production aspects of devoicing, but also on the auditory aspects, which are thought to be a key part of the sound change (Pierrehumbert 2001, 2002; Blevins 2004; Bradley and Delforge 2006; Ohala 1981, 2012; Jiménez and Lloret 2014). The results for production raise some interesting questions, especially whether the acoustic cues which have proved successful in the characterization of the degree of voicing (particularly fraction of unvoiced frames and duration) may have a role in the recognition of sibilants as voiced or voiceless (Widdison 1995, 1997; Lavoie 2001, p. 107).

With this purpose in mind, we conducted an identification test, whose stimuli were extracted from the recordings in the acoustic experiment. From the utterances analysed we selected the intervocalic sequences which displayed the best acoustic quality (with no background noise, adequate volume and clear acoustic features in the spectrogram). We preferred to use natural stimuli: apart from segmenting the [V_V] sequences from the lexical unit they belonged to in order to make word recognition difficult, there was no other manipulation. Therefore, we chose 8 instances of [s], 10 instances of [z], 8 for [ʃ], 10 for [ʒ], 6 for [tʃ], 14 for [dʒ], 10 for [tʃ] and 10 for [dʒ], which involved 34 cases of fricative sibilants and 40 for affricate sibilants, giving a total of 74 different stimuli. Each stimulus was repeated 3 times and they were presented in random order. Thus, the final number of items was 222.¹⁴ The interstimulus silence was 1.5 s. The test was prepared and distributed online with the FOLERPA platform (Fernández Rei 2021). It is important to

remark that cases of [t̪s] entailed some problems: as known, the intervocalic sequences of this sibilant were not in medial position, but in word-initial context. Although in the acoustic experiment we specified that there was no silence between the first vowel and the sibilant, auditory analysis often gave the impression of a cut between them. Consequently, most cases had to be rejected and only 6 utterances could be used.

The participants were 26 Catalan-speakers, 14 from Majorca and 12 from Girona, all of them university students. As in the acoustic experiment, we checked language dominance with the *BLP* questionnaire, whose results are shown in Table 7. In both groups, speakers were Catalan dominants, though there were some important differences by origin ($t(23) = -2.205, p < 0.038$): while Majorcan participants were considered balanced bilinguals (final score 6.785), Girona participants were not, and displayed a clear bias towards Catalan (120.987).

Table 7. Mean scores in the *BLP* questionnaire for Girona and Majorca participants. Results for each section and final scores are shown.

<i>BLP</i> Section	Girona Speakers	Majorca Speakers
II. Linguistic History	16.838	3.443
III. Linguistic Use	78.975	3.503
IV. Linguistic Competence	1.650	−6.647
V. Linguistic Attitudes	23.522	6.485
Total score	120.987	6.785

None of the subjects reported any speech or hearing problem and they were not aware of the purpose of the research. They were told to listen to the stimuli over headphones in a quiet room at home. Each stimulus was presented twice before the participants had to provide an answer. The test was a closed-ended questionnaire, so that they had to choose between two response options (voiced sibilant or voiceless sibilant) from an orthographic transcription of the stimuli. They could not proceed with the rest of the test without answering the item. The whole test was organized in 4 item blocks, after which the participants could rest for a while before continuing with the experiment. It took about 30 min to complete it.

We considered the answer (voiced or voiceless) as the only dependent variable. As factors, we took into account the sibilant type in the stimuli, the origin of the participants (Majorca or Girona), and stress (sibilant in a stressed syllable or in an unstressed syllable), since some scholars have pointed out that it could play a role in the devoicing process (Haggard 1978, p. 96; Smith 1997, pp. 489–90; Davidson 2016). We also set as predictor variables the fraction of unvoiced frames and the sibilant duration. As seen, we selected several instances of each sibilant as stimuli. These cases, apart from being of an adequate acoustic quality, were chosen according to their portion of unvoiced frames, so that we obtained a variety of stimuli for each sibilant in a descending or ascending scale at approximately 10% intervals in their fraction of unvoiced frames (the length of the scale depended on the total range in each type of sibilant and on the quality of the sequences; in fact, for this reason it was not always possible to obtain the complete scale). Duration values, on the other hand, were left as random since the previous experiment proved their direct correlation with the portion of unvoiced frames.

Mixed-effect binomial logistic regressions were also conducted using IBM SPSS (v. 25.0, IBM Corporation, Armonk, NY, USA) with main effects (sibilant, origin, stress, fraction of unvoiced frames and duration) and interactions. Participant and stimulus were entered as random effects. Significance was set at 0.05.

3.2. Results

Firstly, the results concerning the qualitative factors will be given. As can be observed in the confusion matrix in Table 8, the global tendency is to correctly label voiced and

voiceless stimuli as such. However, there are cases of misparsing in all the categories, with a confusion rate that ranges from 8.3% ([tʃ] identified as [dʒ]) to 66.9% ([ts] recognized as [dʒ]). In general, the correct answers are more than 75% of the total except in the cases of [ts] and [dʒ], in which participants' behaviour is quite different (unexpected labelling in most of the responses, as specified below). When stress is considered, different tendencies may be noticed (see Table 8 and Figure 3): alveolar fricatives, as well as dentoalveolar affricates and [dʒ], reach better rates of identification as voiced or voiceless in stressed position; however, alveopalatal fricatives and [tʃ] reach better recognition rates in unstressed syllable, though they are also correctly identified in stressed position.

Table 8. Confusion matrix with the results of the perception test regarding phonological category and interaction between phonological category and stress. Number of responses for each item and percentages are shown.

Variables		[s]	[z]	[ʃ]	[ʒ]	[ts]	[dʒ]	[tʃ]	[dʒ]
global	voiced	131 (21%)	483 (77.4%)	75 (12%)	635 (81.4%)	313 (66.9%)	888 (81.3%)	65 (8.3%)	364 (46.7%)
	voiceless	493 (79%)	141 (22.6%)	549 (88%)	145 (18.6%)	155 (33.1%)	204 (18.7%)	715 (91.7%)	416 (53.3%)
stressed	voiced	51 (16.3%)	245 (78.5%)	63 (20.2%)	281 (72.1%)	313 (66.9%)	490 (89.7%)	65 (16.7%)	364 (93.3%)
	voiceless	261 (83.7%)	67 (21.5%)	249 (79.8%)	109 (27.9%)	155 (33.1%)	56 (10.3%)	325 (83.3%)	26 (6.7%)
unstressed	voiced	80 (25.6%)	238 (76.3%)	12 (3.8%)	354 (90.8%)	-	398 (72.9%)	-	-
	voiceless	232 (74.4%)	74 (23.7%)	300 (96.2%)	36 (9.2%)	-	148 (27.1%)	390 (100%)	390 (100%)

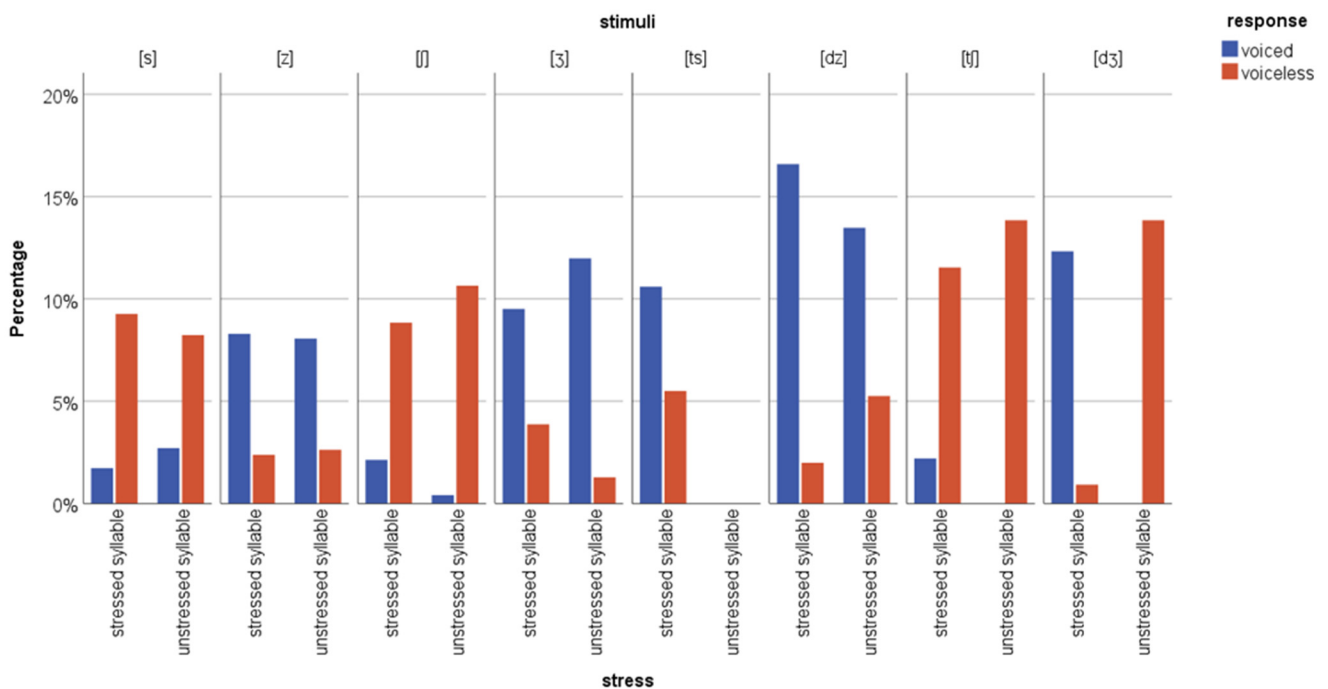


Figure 3. Bar plots representing the portion of voiced and voiceless identifications for each sibilant category depending on stress.

The statistical model (answer \times sibilant type \times dialect \times stress) was significant ($F(22, 5749) = 7.446, p < 0.0001$), though only one main effect, sibilant type, was relevant ($F(7, 5749) = 14.732, p < 0.0001$). As commented, stimuli corresponding to voiceless categories are identified significantly less often as voiced than stimuli corresponding to voiced categories, pointing to a correct identification of the items. The aforementioned exceptions are relevant: $[\text{ts}]$ displays significantly more voiced labellings than expected ($sd = 0.091, t(5749) = 3.098, p < 0.002$), and $[\text{d}\bar{z}]$ displays significantly fewer voiced identifications than expected ($sd = 0.094, t(5749) = -4.574, p < 0.0001$). In these two cases, the inclination to identify sounds corresponding to voiceless or voiced categories as such becomes somewhat blurred. For $[\text{ts}]$, most of the examples (66.9%) were classified as voiced, thus reversing the expected behaviour; for $[\text{d}\bar{z}]$ there is also an inversion, but to a lesser extent (53.3% of the stimuli were labelled as voiceless). It is important to note that both speaker and stimulus influence the results.

There is a significant interaction between phonological category and dialect ($F(7, 5749) = 2.766, p < 0.007$). Participants from Majorca tend to classify $[\bar{z}]$ instances as voiced to a significantly lesser degree than participants from Girona ($F(1, 5749) = 4.169, p < 0.041$). There is no statistically relevant interaction between phonological category and stress ($F(6, 5749) = 1.503, p < 0.173$), but some interesting tendencies deserve attention. Overall, significant results show that voiced and voiceless segments are being correctly parsed with their expected voicing feature, but if we break down the effects, we observe that cases of $[\text{d}\bar{z}]$ are always identified as voiceless (100% of the stimuli) in unstressed syllable ($sd = 0.024, t(5749) = -15.936, p < 0.0001$). This is relevant information since the misparsing rates for $[\text{d}\bar{z}]$ in stressed syllables are much lower (only 6.7%). To sum up, despite a global tendency to correctly identify the stimulus with the appropriate phonological category (i.e., with the expected voice feature), there are two affricate categories which display unexpected levels of confusion, $[\text{ts}]$ and $[\text{d}\bar{z}]$. The former tends to be interpreted as voiced while the latter is perceived as voiceless when in an unstressed context.

It was also interesting to examine whether fraction of unvoiced frames and duration, the most robust indicators of voicing in the acoustic experiment, had an effect in the domain of perception. As can be seen in Table 9 and in Figure 4, stimuli identified as voiceless are in general longer and display a higher portion of unvoiced frames than those identified as voiced in every category, the only exception being $[\text{d}\bar{z}]$. This general observation is confirmed by the statistical analysis. The mixed-effect binomial logistic regression (answer \times fraction of unvoiced frames \times duration) demonstrates that neither fraction of unvoiced frames nor duration condition the participant's response as main effects; however, two- and three-way interactions (sibilant type \times fraction of unvoiced frames, sibilant type \times duration, sibilant type \times fraction of unvoiced frames \times duration) are significant.

Fraction of unvoiced frames influences participants' answers depending on the sibilant ($F(7, 5740) = 3.747, p < 0.0001$). If the portion of unvoiced frames increases (see data in Table 8), there are significantly fewer stimuli identified as voiced among the instances of $[\text{s}]$, $[\text{z}]$, $[\text{j}]$, $[\text{ts}]$ and $[\text{d}\bar{z}]$.¹⁵ $[\bar{z}]$ also follows this tendency, but its results are not statistically relevant. $[\text{d}\bar{z}]$, however, tends to have a slightly lower portion of unvoiced frames when labelled as voiceless. Duration also determines the participants' election depending on the phonological category ($F(7, 5740) = 8.281, p < 0.0001$): when duration increases, the proportion of voiced classifications significantly decreases in all fricative sibilants and in the dentoalveolar affricates.¹⁶ Once again, palatal affricates do not show this behaviour, though the results are not statistically relevant: when parsed as voiceless, their duration is shorter than when recognized as voiced.

More interestingly, there is a correlation of fraction of unvoiced frames and duration depending on the sibilant type ($F(8, 5740) = 6.601, p < 0.0001$). In $[\text{z}]$ ($b_0 = -0.001, se = 0.0001, t = -2.562, p < 0.010$), $[\bar{z}]$ ($b_0 = -0.002, se = 0.0001, t = -5.199, p < 0.0001$) and $[\text{d}\bar{z}]$ ($b_0 = -0.002, se = 0.001, t = -2.577, p < 0.010$), longer duration and higher proportion of unvoiced frames imply less voiced parsing; on the other hand, $[\text{j}]$ shows significantly

more cases than expected labelled as voiced when fraction of unvoiced frames and duration increase ($b_0 = 0.003$, $se = 0.001$, $t = 2.476$, $p < 0.013$), though correct parsings still outweigh incorrect answers. Overall, the participants' judgements confirm the acoustic measurements regarding duration and fraction of unvoiced frames, thus showing that they are reliable parameters to establish voicing degree, which seem to be acoustically and perceptually related.

Table 9. Mean value of fraction of unvoiced frames (in percentages) and duration (in milliseconds) in the stimuli labelled as voiced or voiceless for every category. Standard deviation values are also provided.

Stimuli		Fraction of Unvoiced Frames (%)		Consonant Duration (ms.)	
		\bar{x}	sd	\bar{x}	sd
[s]	voiced	70.35	9.62	95.45	13.97
	voiceless	80.25	11.58	108.48	9.34
[z]	voiced	31.51	26.37	65.64	18.26
	voiceless	40.36	30.10	70.74	20.40
[ʃ]	voiced	50.37	17.65	95.87	11.43
	voiceless	66.26	17.60	112.82	14.87
[ʒ]	voiced	25.60	20.07	95.43	24.13
	voiceless	43.57	20.82	100.15	19.22
[tʃ]	voiced	46.19	27.79	135.20	28.25
	voiceless	71.16	24.56	172.04	45.62
[dʒ]	voiced	33.67	21.38	128.77	24.27
	voiceless	45.31	20.44	140.79	20.41
[tʃ]	voiced	51.92	13.19	146.18	16.84
	voiceless	65.75	13.20	149.61	28.30
[dʒ]	voiced	31.42	21.38	132.89	11.68
	voiceless	29.03	19.17	111.78	20.52

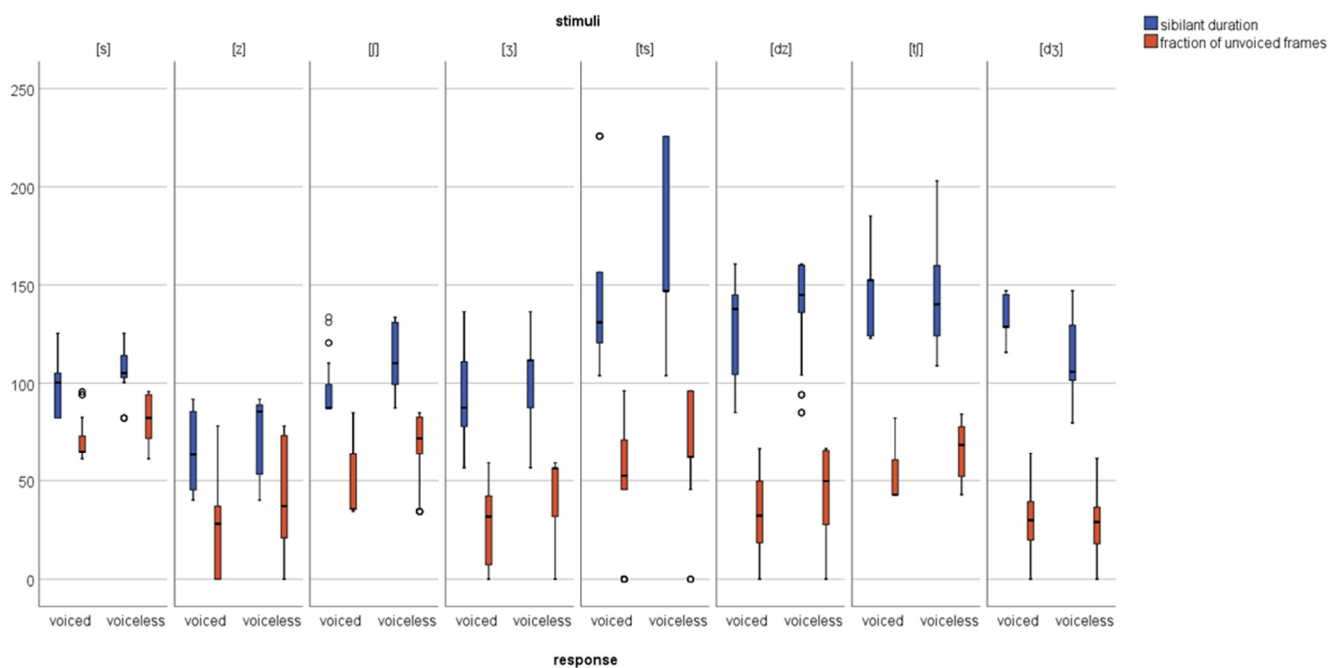


Figure 4. Box plots representing duration values (blue boxes) and fraction of unvoiced frames (orange boxes) in stimuli labelled as voiced and voiceless in each phonological category.

4. Discussion

This paper is concerned with sibilant devoicing in the history of Spanish and aims to shed some light on the topic from an experimental perspective with the aid of comparative grammar. Therefore, Catalan sibilants (/s, z, ʃ, ʒ, t͡s, d͡z, t͡ʃ, d͡ʒ/), which had been deemed very similar to those in Medieval Spanish (Penny 1993; Bradley and Delforge 2006), were acoustically and perceptively analysed in order to determine whether devoicing in intervocalic medial contexts is possible and whether this has a phonetic basis. However, it must be recalled that there are some limitations to our analysis, due to the phonemic distribution of Catalan /t͡s/, which had to be studied in intervocalic initial position. It is also important to bear in mind that the sample of speakers and participants is certainly limited. Indeed, speaker and participant as random effects do condition the statistical models both in the acoustic and in the perception experiments.¹⁷ Thus, our explanations are necessarily provisional.

The acoustic results for Catalan sibilants showed that fraction of unvoiced frames, consonant duration and also intensity (though to a much lesser extent) were relevant in determining voicing characteristics. These parameters were conditioned by the kind of sibilant and, occasionally, by dialect. Duration and fraction of unvoiced frames in particular have been shown to be robust indicators of voicing, since they distinguish voiced from voiceless segments in almost all the sibilant pairs, in line with Hualde and Prieto (2014). As a general tendency, voiced sibilants are shorter, display less intensity and have a smaller portion of unvoiced frames than their voiceless counterparts, as stated previously in the literature (Haggard 1978; Pensado 1993; Widdison 1995; Solé 2003; Ohala and Solé 2010). In this sense, our results in Catalan sibilants fit the general behaviour described for sibilants in the world's languages (Smith 1997 or Davidson 2016 for English, and Van de Velde and van Hout 2001 for Dutch are three examples) and coincide with recent research in the topic (Hualde and Prieto 2014; Hualde et al. 2015).

When applying Smith's (1997, p. 478) scale of voicing to our results, we also observe a great heterogeneity in the production of sibilants. Cases of fully voiced, partially devoiced and fully devoiced consonants are detected in every phonological category. This implies phonetic variation (see Hualde and Prieto 2014; Hualde et al. 2015): sibilants in Catalan are not uniform, which is also in line with the results reported by Smith (1997) or Davidson (2016) for English, or by Van de Velde and van Hout (2001) for Dutch. As mentioned in these works, partially devoiced utterances are the most frequent option for most sibilants, with some exceptions. Alveolar fricatives and [ʃ] tend to be produced more at the extremes ([s] and [ʃ] display more fully devoiced cases, [z] more fully voiced examples). An intermediate case would be [ʒ], in which partially devoiced utterances are less common than expected, though prevailing. At the other end of the scale, [d͡ʒ] presents a significant and clear tendency to partial devoicing (see Hualde et al. 2015, p. 261).

However, Catalan exhibits a certain variability depending on the dialect, at least according to our preliminary data concerning Girona Central Catalan and Majorcan Catalan. Majorcan speakers produce a more voiced [s] than Girona speakers, since in Majorca [s] presents shorter duration and a significantly lower portion of unvoiced frames. The duration of Majorcan [ʃ] (significantly shorter than Girona [ʃ]) also suggests a more voiced production. On the other hand, Majorcan [z] is more devoiced, since its proportion of unvoiced frames is higher. These findings could point to a certain tendency towards an intermediate acoustic space for these segments in Majorcan, while acoustic differences in Girona Central Catalan remain clearer and suggest more extreme realizations. A similar idea arises from the results in Hualde and Prieto (2014, p. 120), also with Central Catalan data. It seems as if the Majorcan options were more peripheral, nearer to a border area between phonological categories (Pierrehumbert 2001, 2002), and thus closer to a hypothetical neutralization. An explanation for this could be associated with the fact that the Majorcan speakers were not so clearly Catalan-dominant as the Girona speakers: according to the BLP scores, though they are Catalan-dominant, the bias in favour of this language was significantly less pronounced.

Affricate sibilants, however, display a different tendency. There are no remarkable divergences between Majorcan and Girona Central Catalan regarding dentoalveolars, but there are in palatals. In this case, Majorcan speakers seem to exaggerate their voicing characteristics, since [tʃ̞] is more devoiced (higher portion of unvoiced frames) and [dʒ̞] is more voiced (it is significantly shorter) than in the production of Girona speakers. This gives the impression that Girona Central Catalan utterances are not as clearly distinct as in Majorcan Catalan. If this is the case, it cannot be accounted for in terms of a lesser degree of Catalan dominance. This finding is in line with [Benet et al. \(2012, p. 402\)](#), who also noted a tendency to neutralize palatal sibilant affricates in Barcelona Catalan, a tendency that was not attributable to Spanish influence, but to internal mechanisms in the system. [Hualde et al. \(2015, p. 260\)](#) also describe the same tendency, with no connection with Spanish influence, for Central Catalan speakers. In general, considering the voicing scale proposed by [Smith \(1997, p. 478\)](#), Majorcan Catalan seems to prefer more extreme realizations than Girona Central Catalan, a global result which must be attributed to the behaviour of affricates.

One of the reasons for examining sibilant voicing in these two dialectal variants was the possibility of affrication in the production of alveopalatal sibilants in Majorcan speakers, as described in [Bibiloni \(2016, pp. 127–28\)](#). This phenomenon should entail an increase in consonant duration, which could conform to the characteristics of devoicing. However, at least from our numbers, this is not the case: Majorcan alveopalatals, especially voiced ones, are significantly shorter than affricates. We could only detect statistically relevant similarities between the duration of [ʃ] and [dʒ̞], which did not indicate an increase in the alveopalatal instances, but a decrease in the duration of affricate instances. In fact, as commented previously, the duration of Girona speakers' [ʒ] was not significantly different from that of Majorcans, and, unexpectedly, [ʃ] was significantly longer in Girona Central Catalan.

This general picture still requires some comments regarding fraction of unvoiced frames and consonant duration. Concerning the first, overlaps have been detected among almost all voiced sibilants ([z, ʒ, dʒ̞]) and among several voiceless sibilants ([s, ʃ, [ʃ̞, tʃ̞, tʃ̞]). This suggests that most voiced sibilants share a similar degree of voicing in terms of the portion of unvoiced frames, and that several voiceless sibilants also coincide in their degree of devoicing. As a consequence, it seems that there is a clear-cut difference between voiced and voiceless segments regarding fraction of unvoiced frames, since no overlap is detected among consonants belonging to the two series. This gives support to [Rohena-Madrado's \(2013, 2015\)](#) method for studying /ʒ/ devoicing, since he employed the values of /s/ percentage of voicing as a benchmark for determining the voiced prepalatal behaviour. However, data for duration indicate a more complex panorama, since there is indeed overlapping among voiced and voiceless sibilants (see [Figure 5](#)): when considered together, fricative sibilants display similar duration, as do [ʃ, dʒ̞] and even voiceless affricates.

The most interesting case may be the aforementioned [ʃ]-[dʒ̞] coincidence: since duration has been deemed one of the most important voicing cues ([Haggard 1978](#); [Pensado 1993, p. 218](#); [Widdison 1995, p. 38](#); [Lavoie 2001, p. 107](#); [Hualde and Prieto 2014](#)), this result may point to a situation in which a voiced affricate could be interpreted as equivalent to a shorter item (the fricative), which is more efficient to produce due to its voiceless character, as explained by [Ohala and Solé \(2010, pp. 53–54\)](#). If we recall that voiceless sibilants are not completely unvoiced, there may be an intermediate point at which peripheral exemplars of the two categories may overlap: the evident preference of [dʒ̞] for partially devoiced realizations supports the idea of exemplars of this category moving from the more canonical forms to a more external area in the acoustic space. It is also relevant that this convergence occurs with palatal consonants, the most prone to devoicing according to [Pensado \(1993, p. 210\)](#).

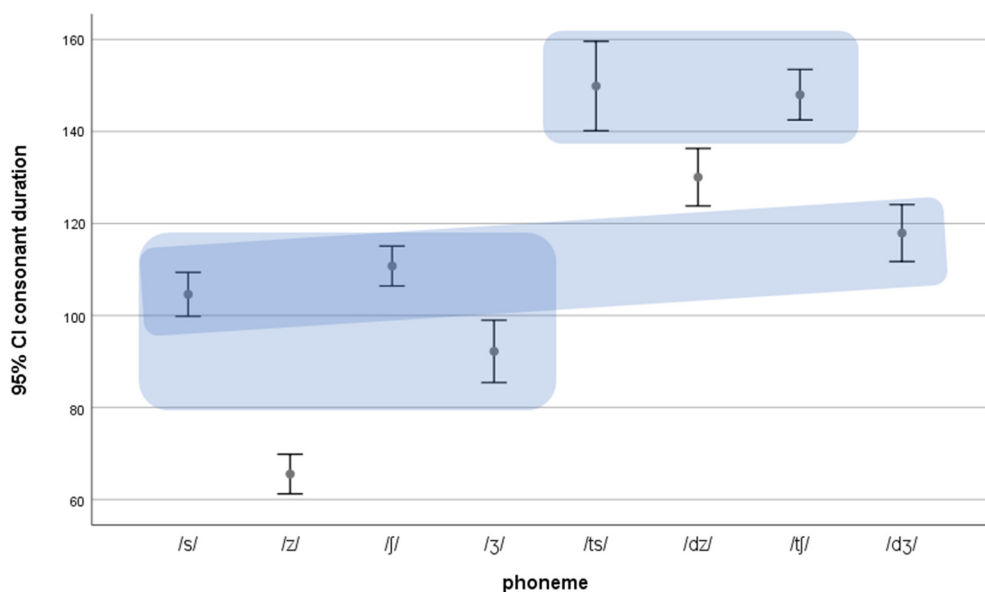


Figure 5. Error bar with duration values for every Catalan sibilant category. Blue shades indicate statistical similarities among the categories.

In the perception domain, participants usually parse the stimuli with their expected voicing value according to the phonological category of the input. On average, 72.32% out of the total stimuli were correctly labelled. However, this percentage increases to 86.23% if we ignore two categories in which misparsings prevail: [tʃ] and [dʒ]. [tʃ] is frequently identified as voiced (confusion rates amount to 66.9%) and [dʒ] is often interpreted as voiceless (53.3% of the total, 100% when in an unstressed syllable). Interestingly, stress is not affecting the perception of the stimuli, except for [dʒ]. This is an unexpected result, since Haggard (1978, p. 100), Widdison (1995, p. 39), Davidson (2016, p. 40) or, to some extent, even Smith (1997, p. 490) pointed to stress as a determining factor in devoicing. Instead, our general results coincide with Hualde et al. (2015, p. 260), who find no effects of stress on voicing (neither it has been deemed to play any role in the Spanish historical process of devoicing). As for [dʒ], its devoicing has been reported in the literature on production (Benet et al. 2012; Recasens 2014, p. 264; Hualde et al. 2015, p. 206), and, regarding its association with stress, it has been described as mostly occurring in post-tonic position (Haggard 1978, p. 100; Recasens 2014, p. 264), or when preceding stress, specially at the end of an unstressed syllable (Smith 1997, p. 490; Davidson 2016, p. 40). In any case, [dʒ] misperceptions fit the idea of devoicing in unstressed position; [z] and [dʒ] confusions are also more usual in unstressed syllable, though not statistically relevant. This fact suggests that unstressed position, as a non-salient context (Smith 1997, p. 490), may favour perceptual confusion in these voiced sibilants. As Kember et al. (2021, p. 414) indicate, prominence facilitates utterance processing.

At this point it is important to remark that confusions do exist in all sibilants, though to different extents: [tʃ] and [dʒ] display a significantly higher rate of errors than the other categories. Although the differences between them are not statistically relevant, alveopalatals, voiced dentoalveolar and voiceless palatal sibilants display lower error rates than alveolar sibilants. This is in line with Pensado's (1993, p. 210) comment on affricate sibilants being more susceptible to devoicing than fricatives (see also Hualde et al. 2015, p. 260): if confusion levels are greater, this implies a stronger inclination to reverse voicing interpretation.

If we focus on the voicing feature, we can observe a tendency to slightly increase confusions in voiced categories with respect to voiceless ones (28.3% in the former vs. 27.05% in the latter, 19.96% vs. 13.76% if we leave out [tʃ] and [dʒ]). From this we may tentatively infer a higher tendency to misperception in the case of voiced sibilants in

comparison to voiceless ones, which seem to be somewhat more stable. This may indicate that voiceless sibilants are easier to parse, which seems to be consistent with the acoustic results: as explained before, they are more homogeneous in their production ([s] and [ʃ] were basically produced as fully devoiced sibilants) than voiced sibilants, which displayed more variability. This would lead, in the case of voiceless sibilants, to the acoustic-auditory space being more crowded around similar values, thus yielding more canonical exemplars which should be easier to identify as voiceless. This is particularly evident in [s], [ʃ] and [tʃ]. As commented before, [tʃ] presents a completely different tendency, but we must remember that the context may play a role in its behaviour: [Smith \(1997\)](#) or [Davidson \(2016, p. 46\)](#) demonstrate that position in the word is an important factor in voicing. Indeed, our results are in line with [Smith \(1997, p. 487\)](#), who found that devoicing was more likely to occur in onset word-medial position than in word-initial position. We have seen that the particular limitations to the distribution of [tʃ] has some consequences in the acoustic approach, but it seems to be even more apparent at the perception level. Its results, therefore, may not be really comparable to the other cases, more so if we take into account that words including this element in Catalan are not frequently used and/or are often loanwords. A better selection of the target items in this case could have changed the results.

Voiced sibilants, on the other hand, tended to be produced as partially devoiced (only [z] was usually uttered as fully voiced, in line with [Hualde and Prieto 2014, p. 117](#)), which would mean a displacement from the expected canonical forms and a less crowded space around the prototypes. This would entail greater difficulty in the parsing process and a higher rate of confusion. This fact is evident in the case of [dʒ], where misparsings are the norm rather than the exception. Another voiced segment, [ʒ], seemed to be taking this route in Majorcan Catalan, since misperceptions are statistically significant, although not yet prevalent.

Stimuli interpreted as voiced were generally shorter and had a lower proportion of unvoiced frames (except for [dʒ]). This behaviour was overwhelmingly clear in the case of fricative and dentoalveolar affricate sibilants. Palatal affricate stimuli could be labelled as voiceless, even when duration decreased, though this result is related more to the tendency for [dʒ] to be parsed as voiceless than to the results for [tʃ].

From these observations some important ideas arise. As [Haggard \(1978\)](#), [Smith \(1997\)](#), [Van de Velde and van Hout \(2001\)](#), [Bradley and Delforge \(2006\)](#), [Rohena-Madrado \(2013, 2015\)](#) or [Davidson \(2016\)](#) argue, voicing does not seem to be a categorical feature, but a gradual one. Our results in Catalan, which generally agree with [Hualde and Prieto \(2014\)](#) and [Hualde et al. \(2015\)](#), give support to this idea. At least in sibilant consonants, there is variation in the acoustic realizations, which demonstrates it. In fact, sounds corresponding to voiced categories display different degrees of voicing and may include allophones which could be classified as fully devoiced following [Smith's \(1997, p. 478\)](#) criteria. Obviously, these examples are not the majority among these categories, but entail a certain overlap with voiceless categories. Phonetic realizations of voiceless sibilants also include partially voiced instances, although, in these cases, fully voiced exemplars are anecdotic. As mentioned above, this is consistent with duration values. Our findings lead us to the conclusion that [s, z, ʃ] are more stable and homogeneous in production and perception than [ʒ] and the affricate sibilants, as stated by [Pensado \(1993, p. 210\)](#).

The aforementioned acoustic features in affricates could pave the way for the preference of unvoiced segments due to the aerodynamic and acoustic restrictions adduced by [Solé \(2003\)](#), [Ohala and Solé \(2010, pp. 39–41\)](#), or [Žygis et al. \(2012, pp. 310–12\)](#), by ensuring turbulence at the expense of vocal-fold vibration. If there is some level of equivalence or overlap, more efficient options in aerodynamic terms will be produced. From this point of view, the idea of lenition defended by [Smith \(1997, pp. 494–96\)](#) and [Lavoie \(2001, p. 107\)](#) makes sense: broadly speaking, options in which only turbulence is required can be favoured over those in which turbulence and vocal-fold vibration are needed since it is possible to obtain a similar result, even when no influence of the context may be adduced. However, the increase in duration contradicts a lenition phenomenon. [Pensado \(1993,](#)

pp. 222–23) insisted (in line with traditional accounts of historical sibilant devoicing in Spanish) on a strengthening process, though our results seem to be partially coherent with the description of devoicing as a reduction in duration and in the magnitude of articulatory gestures, in accordance with the definition of lenition offered in [Bybee and Easterday \(2019\)](#), p. 271). This reasoning may particularly explain the results for [d̥ʒ]: ambiguity in production results in perception misparsing and, finally, leads to complete devoicing in production as the final stage in a process of change ([Pierrehumbert 2001](#); [Blevins 2004](#); [Ohala 1981, 2012](#)).

The second relevant issue concerns the perception level. Our data show an evident coincidence between acoustic properties and auditory impression in the Catalan sibilants. In fact, the perception of voicing relies on durational parameters, as [Haggard \(1978\)](#), [Widdison \(1995\)](#), [Lavoie \(2001\)](#), and [Žygis et al. \(2012\)](#) had pointed out. However, we detected that fraction of unvoiced frames was also involved in the recognition of voicing. As [Widdison \(1995, 1997\)](#) and [Ohala \(1981, 2012\)](#) remark, listeners seem to rely on more than a single cue to correctly parse the input. Certainly, we only examined two parameters, but it may be equally possible that other acoustic cues are helping to determine voicing adscription: [Bradley and Delforge \(2006\)](#), p. 31) mention F1 transitions, the duration of the preceding vowel or F₀ of the following vowel, for example. Despite this fact, misparsings are obviously possible and are even more likely when functional load is low ([Wedel et al. 2013](#), p. 184): the consideration of [tʃ] as a phoneme is not undisputed, and cases in which [d̥ʒ] may give rise to minimal pairs are not abundant ([Wheeler 2005](#), pp. 11–12). Besides, in our experiment, the stimuli were extracted from complete words, thus resulting mostly in sequences with no lexical meaning: in such cases, lexical knowledge could not help in disambiguating the signal.

All these considerations may be relevant to our understanding of the sibilant devoicing process in Early Modern Spanish. As commented, the sibilant system in Medieval Spanish was very similar to that in Catalan ([Penny 1993](#); [Bradley and Delforge 2006](#)). For this reason, we could attempt to extrapolate the information obtained for Catalan to the sound change in Spanish. In this respect, the first important topic to address is that, as stated by [Pensado \(1993\)](#) or [Widdison \(1995, 1997\)](#), sibilant devoicing is a phonetically based process. The analysis of Catalan sibilants demonstrates the overlap in the allophonic production of voiced and voiceless sibilant pairs, especially when it comes to voiced sibilant utterances. Only [z] seems to be more resistant to this general tendency. Even duration values indicate this overlap. If we admit duration as one of the most reliable cues for voicing, the resulting picture is highly complex and shows multiple areas of coincidence among sibilants: in [Figure 5 \(supra\)](#) we can observe that only two voiced categories are clearly distinct: [z], the shortest sibilant, and [d̥ʒ], which holds an intermediate place between voiceless affricates and the rest of the sibilants. The central space is occupied by [s], [ʃ], [ʒ] and [d̥ʒ]. Note that [ʒ] and [d̥ʒ] are the sibilants which showed a more evident tendency to devoice, according to our figures (in this respect, Rioplatense Spanish devoicing of /ʒ/ —[Fontanella de Weinberg 1987, 2000](#); [Rohena-Madrado 2013, 2015](#); or [Michnowicz and Planchón 2020](#)— constitutes a perfect case of study for the explanation of the diachronic evolution). This ambiguous crowded space may be the key to shedding light on the historical process in Spanish.

As is known, not all voiced sibilants were devoiced at the same time: the phenomenon began with dentoalveolar segments, which induced devoicing in the fricative series ([Eddington 1987](#), p. 58; [Lapesa 1981](#), p. 283; [Sánchez Prieto 2004](#), p. 422; [Quilis 2005](#), pp. 168–69; [Ariza 2012](#), p. 224). [Pensado \(1993\)](#), p. 210) agrees with this claim and points out that devoicing is usually gradual and starts with a specific sibilant pair before spreading to the others. She notes that the progression was from affricates to fricatives and from palatals to alveolars. Our data in Catalan support this. Ambiguities are found mostly in affricate sibilants, namely [d̥ʒ] and [tʃ]. The first case clearly demonstrates the devoicing process, particularly at the auditory level, as we commented before. The second is also very interesting, despite the limitations to its distribution. If we examine [tʃ] and [d̥ʒ] in terms of [Smith's \(1997\)](#), p. 478) degree of voicing classification, we observe an interesting result: most of the utterances of both affricates are partially devoiced (46.9% of the total in

the case of [t̪s̪] and 78.1% for [d̪z̪]). It is even more interesting to examine this by dialect: Majorcan Catalan speakers tend to produce [t̪s̪] as fully devoiced, but 66.7% of instances of [d̪z̪] are partially devoiced (only 25% fully voiced). Girona Catalan speakers show a similar pattern for [t̪s̪], but the proportion of partially devoiced instances of [d̪z̪] rises to 85% (only 6% are fully voiced). This suggests that dentoalveolars are not produced at the extremes, like alveolar fricatives, but at an ambiguous intermediate point. Their clouds of exemplars tend to overlap in the acoustic-auditory space, as described by [Pierrehumbert \(2001, 2002\)](#). Historical change may thus be accounted for as a case of phonological ambiguity ([Blevins 2004](#); [Ohala 1981, 2012](#)).

What seems to be clear is that phonetic variation regarding voicing is a fact. This is the first compulsory step for a sound change to initiate ([Lindblom 1990](#); [Ohala 1981, 2012](#); [Pierrehumbert 2001, 2002](#); [Blevins 2004](#)). After this, the presence of multiple allophones can lead to a process of reanalysis and recategorization following different paths: in the dentoalveolar pair, the acoustic ambiguity among phonetic realizations may imply the selection of the more frequent variants, which would have tended to be the more devoiced ones, so a recategorization as voiceless would follow (a case of choice, according to [Blevins 2004](#)). In the case of palatal affricates, misperception indicates that listeners reinterpret the input as voiceless, which would mean that their production would have varied to fit a voiceless category (a case of chance in [Blevins 2004](#)). Even [ʒ] seems to show an incipient tendency to devoicing. These complementary sources could have been the starting point of the whole process. In fact, though [s] and [z] seem to be the most stable segments, the volume of incorrect identifications in the perception experiment reveals that misperceptions do occur (above 20% in both categories and slightly more often in unstressed syllable), so confusion between them is also possible and may take place.

The historical process in Early Modern Spanish thus seems to have followed universal phonetic tendencies and we may account for it as a phonetically grounded sound change, rather than invoking other causes in its inception. Internal factors, as in most regular sound changes, show that the global explanation may correspond to a complex process of lenition, or, at least, not a clear strengthening case. Language contact ([Martinet 1951–1952](#); [Lloyd 1993](#), pp. 429–37) or the need for a readjustment in the phonological system ([Contini 1951](#), pp. 179–80; [Alarcos 1988](#), pp. 51–53; [Penny 1993](#), pp. 81–82; [Ariza 2012](#), p. 224) may be contributing factors (in the first case) or the result (in the second case) of the process more than the triggering factor.

5. Conclusions

This paper is a first approach to sibilant devoicing in the history of Spanish, a process that concluded in the 16th century. Since such a process has been deemed atypical, the usual explanations claimed that devoicing was the result of a reorganization of the mediaeval phonological system aiming at more symmetry and/or more efficiency. Other justifications invoked linguistic contact as the reason for the neutralization: pressure from Basque, which lacks voiced sibilants, could have induced voicing neutralization in Spanish. However, phonetic factors can explain the sound change in a more simple and efficient way, as [Pensado \(1993\)](#) or [Widdison \(1995, 1997\)](#) have pointed out.

Following recent methods in historical linguistics, we turned to comparative grammar and experimental phonetics to examine the topic in greater depth and try to determine whether this evolution could be understood as a phonetically based sound change, initiated for internal reasons. With this purpose in mind, Catalan sibilants were acoustically and perceptually analysed in terms of consonant duration, intensity and fraction of unvoiced frames, also taking into account their degree of voicing.

Our results point to a phonetically based sound change, since devoicing may be accounted for in terms of acoustic and aerodynamic factors, as well as in terms of misperception, like most regular sound changes. As has been pointed out, voiced sibilants are not common segments in the world's languages and tend to devoice due to aerodynamic constraints. This is evidenced by means of certain acoustic features, such as duration

and fraction of unvoiced frames. If the cloud of exemplars of voiced sibilant categories overlaps with the cloud of exemplars of their voiceless counterparts, the voiceless options will be preferred, since their production is more efficient. Moreover, perceptual ambiguity among exemplars in the peripheral area in the acoustic-auditory space will pave the way for reanalysis and recategorization. This process will be favoured more if the phonological opposition has a low functional load, as was the case in Medieval Spanish sibilants.

However, it is important to remark that our results are only tentative, since the sample in the acoustic experiment is clearly limited: more recordings are required in future research to guarantee solid findings. Even in the perception experiment more participants should be recruited for further analysis in order to ensure more representative results. Moreover, it will be crucial to refine the corpus, to achieve a more solid analysis of voiceless dentoalveolar affricates and more reliable comparison between the sibilants. Taking these aspects into account will be the way to obtain more substantial results to support these preliminary findings.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were waived for this study, since the data is properly anonymized and informed consent is obtained at the time of original data collection.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data available on request due to restrictions in privacy in the informed consent signed by the speakers. The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the mentioned restrictions in privacy, which prevents the authors to publicly share the obtained data with third parties.

Acknowledgments: This work would not have been possible without the generous help and interesting observations from Wendy Elvira-Rodríguez and Rubén Pérez-Ramón. I would also like to thank three anonymous reviewers for their helpful comments.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

Laboratory speech corpus employed in the acoustic experiment. The words in italics contain the segments analyzed in this paper.

1. Els *assajos* són els dissabtes a migdia. ‘Rehearsals are on Saturdays at noon.’
2. Era tan poca-vergonya que es va dedicar a *assetjar* els companys de feina. ‘He/She was such a rascal that he/she harassed his/her workmates.’
3. Explicava que calia *aixafar* el raïm per obtenir el vi. ‘He/She explained that grapes should be squashed in order to obtain wine.’
4. Va pensar que l’*atzar* el duia per camins ben imprevisibles. ‘He/She thought that fate led him along unsuspected paths.’
5. La tieta es va *casar* amb un senyor de Capdepera. ‘Auntie married a man from Capdepera.’
6. Sentir la paraula “*tsar*” i riure era tot u. ‘He/She used to laugh as soon as he/she heard the word *tsar*.’
7. Va agafar una *passa* de panxa que el va *deixar* ben aixafat. ‘He/She caught a stomach bug and he/she feels under the weather.’
8. Trobo que el pare ha *catxat* molt en el últims temps. ‘I feel that recently our/my father has aged a lot.’
9. Li agradava moltíssim anar a la *platja* el mes de setembre. ‘He/She adored going to the beach in September.’
10. Va decidir *barrejar* lleixiu i sulfurant amb resultats desastrosos. ‘He/She decided to mix bleach with hydrochloric acid, and the result was calamitous.’
11. Em va dir que vivia a *casa* d’una parenta. ‘He/She told me that he/she lived at a relative’s place.’

12. La Laura va aparèixer *borratxa* a la seva festa de comiat. ‘Laura turned up completely drunk at her farewell party’
13. Veient la contrarietat, va exclamar “*vaja*” i va marxar. ‘When he/she realised the setback, he/she exclaimed “oh!”, and left.’
14. Considerava “*tsarisme*” un préstec del rus. ‘He/She considered *tsarism* a loanword from Russian.’
15. Tenia el cabell d’un color *atzabeja* molt bonic. ‘He/she had got gorgeous jet black hair.’

Appendix B

Table A1. Stimuli used in the perception experiment.

Phoneme	Stressed Syllable		Unstressed Syllable	
	Stimulus	Fraction of Unvoiced Frames (%)	Stimulus	Fraction of Unvoiced Frames (%)
/s/	[a’sa]	94.1%	[asa]	95.6%
	[a’sa]	82.6%	[asa]	82.3%
	[a’sa]	73%	[asa]	71.9%
	[a’sa]	61%	[asa]	64.9%
/z/	[a’za]	73.3%	[aza]	78.2%
	[a’za]	37%	[aza]	30.6%
	[a’za]	28.1%	[aza]	20.9%
	[a’za]	0	[aza]	0
/ʃ/	[a’ja]	82.7%	[aʃa]	84.9%
	[a’ja]	71.9%	[aʃa]	76.1%
	[a’ja]	63.6%	[aʃa]	65.7%
	[a’ja]	35.6%	[aʃa]	34.3%
/ʒ/	[a’za]	56.2%	[aʒa]	59%
	[a’za]	42.1%	[aʒa]	47.7%
	[a’za]	31.8%	[aʒa]	31.7%
	[a’za]	13.6%	[aʒa]	7.3%
	[a’za]	0	[aʒa]	0
/t͡s/	[a’t͡sa]	96%	-	-
	[a’t͡sa]	71.1%	-	-
	[a’t͡sa]	61.9%	-	-
	[a’t͡sa]	52.4%	-	-
	[a’t͡sa]	45.4%	-	-
	[a’t͡sa]	0	-	-
/d͡z/	[a’d͡za]	66.6%	[ad͡za]	65.6%
	[a’d͡za]	57.6%	[ad͡za]	56.8%
	[a’d͡za]	49.7%	[ad͡za]	47.8%
	[a’d͡za]	38.3%	[ad͡za]	32.2%
	[a’d͡za]	27.6%	[ad͡za]	26.8%
	[a’d͡za]	14.5%	[ad͡za]	18.4%
	[a’d͡za]	0	[ad͡za]	0
/t͡ʃ/	[a’t͡ʃa]	82.2%	[at͡ʃa]	84.2%
	[a’t͡ʃa]	7.09%	[at͡ʃa]	77.8%
	[a’t͡ʃa]	60.5%	[at͡ʃa]	68.5%
	[a’t͡ʃa]	52%	[at͡ʃa]	59.4%
	[a’t͡ʃa]	42.8%	[aʃa]	47.8%
/d͡ʒ/	[a’d͡ʒa]	64.1%	[ad͡ʒa]	61.2%
	[a’d͡ʒa]	39.3%	[ad͡ʒa]	36.5%
	[a’d͡ʒa]	29.9%	[ad͡ʒa]	28.9%
	[a’d͡ʒa]	19.8%	[ad͡ʒa]	17.9%
	[a’d͡ʒa]	0	[ad͡ʒa]	3.9%

Notes

- 1 We use “teleological” in the sense described in [Blevins \(2004, p. 45\)](#) or in [Ohala \(2012, p. 24\)](#) to refer to purpose-directed sound changes, mainly oriented to an optimization of the phonological system. One reviewer points out that teleological change can also include phonetic-based sound changes, since speakers decide on the variants for ease of production or comprehension. Following [Blevins \(2004\)](#) and [Ohala \(2012\)](#), we assume that even changes related to ease of articulation or maximization of contrast (which can be associated to phonetic reasons) are not goal-directed: these are effects of the unaware actuation of the speaker and the listener in uttering and perceiving sounds (see also the results regarding speech style in [Rohena-Madrazo 2013, pp. 52–55](#)).
- 2 It should be noted that most descriptions omit /t̃/ from the explanation ([Alarcos 1988](#); [Lapesa 1981](#); [Lloyd 1993](#); [Penny 1993](#); [Bradley and Delforge 2006](#)), due to the lack of a voiced cognate: the absence of a devoicing process in this case justifies the omission. However, as one reviewer correctly points out, it is likely that [d̃] existed as an allophone of /z/ in post-pausal contexts. In fact, phonetic variation probably have existed for all the sibilant categories.
- 3 For considerations on the preservation of the voiced sibilants in some particular areas, see, for example, [Sanchis Guarner \(1949\)](#), regarding Aguaviva, a village in the border region between Aragon and Catalonia (a bilingual area influenced by Catalan), or [Salvador and Ariza \(1992\)](#) concerning Cáceres province. Recent research would be needed to verify if the maintenance of the opposition is still alive in these spots. Even in Judeo-Spanish the voiced–voiceless contrasts were preserved ([Lleal 2004, pp. 1150–51](#); [Noll 2014, p. 604](#); [Bradley 2022, p. 816](#)).
- 4 Note that [Penny \(1993\)](#) suggests that neutralization in such a context would have later extended to medial intervocalic position, where the functional load of the voiced–voiceless contrast was indeed very low. In this sense, [Wedel et al. \(2013\)](#) demonstrated statistically the relationship between functional load and neutralization processes: phonological oppositions yielding high functional load are usually maintained, while pairs which are not productive tend to neutralize.
- 5 See also [Hualde and Prieto \(2014\)](#) and [Jiménez and Lloret \(2014\)](#) for two interesting approaches from the opposite perspective (voicing in the Romance languages).
- 6 “A voiced fricative is said to be devoiced when its periodic component ceases before the friction component” ([Haggard 1978, p. 95](#)).
- 7 In spite of the presence of [j] and [z] in some varieties of Spanish, among them most of Argentinian Spanish, it is not a transparent opposition for all speakers. It should be noted that the author explains that participants had phonological training in the second experiment.
- 8 Dentoalveolar affricates are not very frequent and /t͡s/ in particular is very rare in intervocalic lexical position ([Wheeler 2005: 12](#)). According to the data in [Rafel i Fontanals \(1980, pp. 480–81\)](#), the voiceless dentoalveolar in medial position has a relative frequency of 0.0195%, while in word-initial position it increases to 0.1072%.
- 9 For further descriptions on the Catalan sibilants voicing contrasts, see [Hualde and Prieto \(2014, p. 112\)](#) regarding the alveolar pair, and [Hualde et al. \(2015, pp. 244–46\)](#), for the prepalatal fricatives and the palatal affricates.
- 10 Devoicing of sibilants has been claimed to be the result of the influence of Spanish in urban areas of Barcelona, though [Benet et al. \(2012\)](#) show that, while the devoicing of /z/ can be related to Spanish as L1, /d̃/ devoicing must be associated with internal factors, not to language contact. It seems problematic to attempt generalizations regarding the effect of Spanish on Catalan in this respect.
- 11 We carried out generalized linear mixed-effects, with speaker and word as random effects; the voicing value was set as the fixed effect and the interaction between stress and voicing value was analysed. Stress had no effect on duration ($F(2, 120) = 0.071, p < 0.932$), intensity ($F(2, 120) = 0.349, p < 0.706$) or degree of glottal fold vibration ($F(2, 120) = 0.052, p < 0.950$) in the alveolar pair, nor in the alveopalatal sibilants ($F(2, 119) = 0.129, p < 0.879$; $F(2, 119) = 0.112, p < 0.894$; $F(2, 119) = 0.031, p < 0.969$), dentoalveolars ($F(2, 120) = 0.036, p < 0.965$; $F(2, 120) = 0.147, p < 0.863$; $F(2, 120) = 0.102, p < 0.903$), or palatal sibilants ($F(2, 120) = 0.403, p < 0.670$; $F(2, 120) = 0.083, p < 0.921$; $F(2, 120) = 0.194, p < 0.824$).
- 12 ($F(1, 115) = 23.451, p < 0.0001$) in alveolar fricatives, ($F(1, 126) = 25.239, p < 0.0001$) in dentoalveolar affricates and ($F(1, 98) = 29.574, p < 0.0001$) in palatal affricates.
- 13 Duration significantly distinguishes [z] ($b_0 = -60.662, se = 6.206, t = -9.775, p < 0.0001$), [z̃] ($b_0 = -8.856, se = 6.206, t = -5.326, p < 0.0001$), [t͡s] ($b_0 = 30.331, se = 6.206, t = 4.888, p < 0.0001$), [t͡ʃ] ($b_0 = 22.553, se = 6.206, t = 3.634, p < 0.003$) and [d̃] ($b_0 = 125.077, se = 5.268, t = 23.743, p < 0.0001$). Fraction of unvoiced frames yields significant results in the cases of [s] ($b_0 = 44.777, se = 4.512, t = 9.925, p < 0.0001$), [z] ($b_0 = -18.940, se = 4.512, t = -4.198, p < 0.001$), [j] ($b_0 = 38.257, se = 4.512, t = 8.480, p < 0.0001$), [t͡s] ($b_0 = 27.731, se = 4.512, t = 6.147, p < 0.0001$) and [t͡ʃ] ($b_0 = 27.128, se = 4.512, t = 6.013, p < 0.0001$).
- 14 The complete list of stimuli is given in Appendix B.
- 15 Statistical results for [s] ($b_0 = -0.564, se = 0.251, t = -2.242, p < 0.025$), [z] ($b_0 = -0.230, se = 0.109, t = -2.108, p < 0.035$), [j] ($b_0 = -0.497, se = 0.149, t = -3.343, p < 0.001$), [t͡s] ($b_0 = -0.287, se = 0.115, t = -2.488, p < 0.013$) and [d̃] ($b_0 = -0.223, se = 0.112, t = -1.983, p < 0.047$).
- 16 [s] ($b_0 = -0.412, se = 0.153, t = -2.696, p < 0.007$), [z] ($b_0 = -0.137, se = 0.030, t = -4.646, p < 0.0001$), [j] ($b_0 = -0.455, se = 0.084, t = -5.444, p < 0.0001$), [z̃] ($b_0 = -0.099, se = 0.029, t = -3.375, p < 0.001$), [t͡s] ($b_0 = -0.179, se = 0.037, t = -4.825, p < 0.0001$) and [d̃] ($b_0 = -0.153, se = 0.029, t = -5.347, p < 0.0001$).

- ¹⁷ It is interesting to note that works on devoicing of /s/ in Rioplatense Spanish, though having a more solid sample, also report relevant individual variation (see Rohena-Madrado 2015 or Michnowicz and Planchón 2020).

References

- Alarcos, Emilio. 1988. De nuevo sobre los cambios fonéticos del siglo XVI. In *Actas del I Congreso Internacional de Historia de la Lengua Española*. Edited by Manuel Ariza, Antonio Salvador and Antonio Viudas. Madrid: Arco Libros, vol. 1, pp. 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.
- Benet, Ariadna, Susana Cortés, and Conxita Lleó. 2012. Devoicing of sibilants as a segmental cue to the influence of Spanish onto current Catalan phonology. In *Multilingual Individuals and Multilingual Societies*. Edited by Kurt Braunmüller and Christoph Gabriel. Amsterdam: John Benjamins, pp. 391–404.
- Bibiloni, Gabriel. 2016. *El català de Mallorca. La fonètica*. Palma: Lleonard Muntaner.
- Birdsong, David, L. M. Gertken, and Mark Amengual. 2012. *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*. Austin: COERLL, University of Texas at Austin. Available online: <https://sites.la.utexas.edu/bilingual/> (accessed on 8 October 2021).
- Blevins, Juliette. 2004. *Evolutionary Phonology. The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.
- Boersma, Paul, and David Weenink. 2018. *Praat: Doing Phonetics by Computer*, Version 6.0.40. Available online: <https://www.fon.hum.uva.nl/praat/> (accessed on 16 May 2018).
- Bradley, Travis G. 2022. Judeo-Spanish. In *Manual of Romance Phonetics and Phonology*. Edited by Christoph Gabriel, Randall Gess and Trudel Meisenburg. Berlin: de Gruyter, pp. 808–38.
- Bradley, Travis G., and Ann Marie Delforge. 2006. Systemic contrast and the diachrony of Spanish sibilant voicing. In *Historical Romance Linguistics: Retrospectives and Perspectives*. Edited by Deborah Arteaga and Randall Gess. Amsterdam: John Benjamins, pp. 19–52.
- Bybee, Joan, and Shelece Easterday. 2019. Consonant strengthening: A crosslinguistic survey and articulatory proposal. *Linguistic Typology* 23: 263–302. [CrossRef]
- Cano, Rafael. 2004. Cambios en la fonología del español durante los siglos XVI y XVII. In *Historia de la lengua española*. Edited by Rafael Cano. Barcelona: Ariel, pp. 825–58.
- Contini, Gianfranco. 1951. Sobre la desaparición de la correlación de sonoridad en castellano. *Nueva Revista de Filología Hispánica* 5: 173–82. [CrossRef]
- Davidson, Lisa. 2016. Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics* 54: 35–50. [CrossRef]
- Eager, Christopher D. 2015. Automated voicing analysis in Praat: Statistically equivalent to manual segmentation. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Edited by Maria Wolters, Judy Livingstone, Bernie Beattie, Rachel Smith, Mike MacMahon, Jane Stuart-Smith and Jim Scobbie. Glasgow: University of Glasgow. Available online: <http://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/proceedings.html> (accessed on 11 October 2021).
- Eddington, David S. 1987. Spanish sibilant evolution. *Deseret Language and Linguistic Society Symposium* 13: 55–62.
- Elvira-García, Wendy. 2013. *Voice Report for All Labelled Wavs in a Folder*, [Praat script].
- Fernández Rei, Elisa (coord.). 2021. *FOLERPA: Ferramenta On-Line para Experimentación Perceptiva*. Santiago de Compostela: Instituto da Lingua Galega. Available online: <https://ilg.usc.gal/folepa/> (accessed on 20 October 2021).
- Fontanella de Weinberg, María Beatriz. 1987. *El español bonaerense. Cuatro siglos de evolución lingüística (1580–1980)*. Buenos Aires: Hachette.
- Fontanella de Weinberg, María Beatriz (coord.). 2000. *El español de la Argentina y sus variedades regionales*. Buenos Aires: Edicial.
- Haggard, Mark. 1978. The devoicing of voiced fricatives. *Journal of Phonetics* 6: 95–102. [CrossRef]
- Harrington, Jonathan. 2012. The relationship between synchronic variation and diachronic change. In *The Oxford Handbook of Laboratory Phonology*. Edited by Abigail C. Cohn, Cécile Fougeron and Marie K. Huffman. Oxford: Oxford University Press, pp. 321–32.
- Hualde, José Ignacio, and Pilar Prieto. 2014. Lenition of intervocalic alveolar fricatives in Catalan and Spanish. *Phonetica* 71: 109–27. [CrossRef]
- Hualde, José Ignacio, Christopher D. Eager, and Marianna Nadeu. 2015. Catalan voiced prepalatals: Effects of nonphonetic factors on phonetic variation? *Journal of the International Phonetic Association* 45: 243–67. [CrossRef]
- Jiménez, Jesús, and María Rosa Lloret. 2014. Efectos graduales de la sonorización en las lenguas romances. *Revista de Filología Románica* 31: 55–82. [CrossRef]
- Kember, Heather, Jiyoun Choi, Jenny Yu, and Anne Cutler. 2021. The processing of linguistic prominence. *Language and Speech* 64: 413–36. [CrossRef] [PubMed]
- Kiddle, Lawrence B. 1977. Sibilant turmoil in Middle Spanish. *Hispanic Review* 45: 327–36. [CrossRef]
- Lahoz, José María. 2015. *Fonética y fonología de los fenómenos de refuerzo consonántico en el seno de unidades léxicas en español*. Doctoral dissertation, Universidad Complutense de Madrid, Madrid, Spain.
- Lapesa, Rafael. 1981. *Historia de la lengua española*. Madrid: Gredos.
- Lavoie, Lisa M. 2001. *Consonant Strength. Phonological Patterns and Phonetic Manifestations*. New York: Routledge.
- Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H Theory. In *Speech Production and Speech Modelling*. Edited by William J. Hardcastle and Alain Marchal. Dordrecht: Kluwer Academic Publishers, pp. 403–39.

- Lipski, John. 1989. /-s/voicing in Ecuadoran Spanish. *Lingua* 79: 49–71. [\[CrossRef\]](#)
- Lleal, Coloma. 2004. El judeoespañol. In *Historia de la lengua española*. Edited by Rafael Cano. Barcelona: Ariel, pp. 1139–67.
- Lloyd, Christopher. 1993. *Del latín al español*. Madrid: Gredos.
- Martinet, André. 1951–1952. The unvoicing of Old Spanish sibilants. *Romance Philology* 5: 133–56.
- Michnowicz, Jim, and Lucia Planchón. 2020. *Sheísmo* in Montevideo Spanish. Not (yet) identical to Buenos Aires. In *Variation and Evolution: Aspects of Language Contact and Contrast across the Spanish-Speaking World*. Edited by Sandro Sessarego, Juan José Colomina-Almiñana and Adrián Rodríguez Riccelli. Amsterdam: John Benjamins, pp. 163–86.
- Noll, Volker. 2014. L'espagnol en dehors de l'Europe. In *Manuel des Langues Romanes*. Edited by Andre Klump, Johannes Kramer and Aline Willems. Berlin: de Gruyter, pp. 588–607.
- Ohala, John. 1981. The listener as a source of sound change. In *Papers from the Parasession on Language and Behavior*. Edited by C. S. Masek, R. A. Hendrick and M. F. Miller. Chicago: Chicago Linguistic Society, pp. 178–203.
- Ohala, John. 2003. Phonetics and historical phonology. In *The Handbook of Historical Linguistics*. Edited by Brian D. Joseph and Richard D. Janda. Oxford: Blackwell, pp. 669–86.
- Ohala, John. 2012. The listener as a source of sound change. An update. In *The Initiation of Sound Change: Perception, Production, and Social Factors*. Edited by Maria-Josep Solé and Daniel Recasens. Amsterdam: John Benjamins, pp. 21–36.
- Ohala, John J., and Maria-Josep Solé. 2010. Turbulence and phonology. In *Turbulent Sounds. An Interdisciplinary Guide*. Edited by Susanne Fuchs, Martine Toda and Marzena Żygis. Berlin: de Gruyter, pp. 37–101.
- Penny, Ralph. 1993. Neutralization of voice in Spanish and the outcome of the Old Spanish sibilants. In *Hispanic Linguistic Studies in Honour of F. W. Hodcroft*. Edited by David Mackenzie and Ian Michael. Oxford: Dolphin, pp. 75–88.
- Penny, Ralph. 2004. Evolución lingüística en la Baja Edad Media: Evoluciones en el plano fónico. In *Historia de la lengua española*. Edited by Rafael Cano. Barcelona: Ariel, pp. 593–612.
- Penny, Ralph. 2014. *Gramática histórica del español*. Barcelona: Ariel.
- Pensado, Carmen. 1993. El ensordecimiento castellano: ¿un “fenómeno extraordinario”? *Anuario de Lingüística Hispánica* IX: 195–230.
- Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In *Frequency Effects and the Emergence of Linguistic Structure*. Edited by Joan Bybee and Paul Hopper. Amsterdam: John Benjamins, pp. 137–51.
- Pierrehumbert, Janet. 2002. Word-specific phonetics. In *Laboratory Phonology VII*. Edited by Carlos Gussenhoven and Natasha Warner. Berlin and New York: Mouton de Gruyter, pp. 101–39.
- Quilis, Antonio. 2005. *Fonética histórica y fonología diacrónica*. Madrid: UNED.
- Rafel i Fontanals, Joaquim. 1980. Dades sobre la freqüència de les unitats fonològiques en català. *Estudis Universitaris Catalans* 25: 473–496.
- Recasens, Daniel. 2014. *Fonètica i fonologia experimentals del català. Vocals i consonants*. Barcelona: Institut d'Estudis Catalans.
- Rohena-Madrado, Marcos. 2013. Variación y cambio de sonoridad de la fricativa postalveolar del español de Buenos Aires. In *Perspectivas teóricas y experimentales sobre el español de la Argentina*. Edited by Laura Colantoni and Celeste Rodríguez Louro. Madrid/Frankfurt: Iberoamericana Vervuert, pp. 37–57.
- Rohena-Madrado, Marcos. 2015. Diagnosing the completion of a sound change: Phonetic and phonological evidence for /s/ in Buenos Aires Spanish. *Language Variation and Change* 27: 287–317. [\[CrossRef\]](#)
- Salvador, Antonio, and Manuel Ariza. 1992. Sobre la conservación de sonoras en la provincia de Cáceres. *Zeitschrift für Romanische Philologie* 108: 276–92. [\[CrossRef\]](#)
- Sánchez Prieto, Pedro. 2004. La normalización del castellano escrito en el siglo XIII. Los caracteres de la lengua: Grafías y fonemas. In *Historia de la lengua española*. Edited by Rafael Cano. Barcelona: Ariel, pp. 423–48.
- Sanchis Guarnier, Manuel. 1949. Noticia del habla de Aguaviva de Aragón. *Revista de Filología Española* XXXIII: 15–65.
- Smith, Caroline L. 1997. The devoicing of /z/ in American English: Effects of local and prosodic context. *Journal of Phonetics* 25: 471–500. [\[CrossRef\]](#)
- Solé, Maria-Josep. 2003. Aerodynamic characteristics of onset and coda fricatives. Paper presented at 15th International Conference on Phonetic Sciences, Barcelona, Spain, August 3–9; vol. 2, pp. 2761–64.
- Van de Velde, Hans, and Roeland van Hout. 2001. The devoicing of fricatives in a reading task. In *Linguistics in the Netherlands 2001*. Edited by Ton van der Wouden and Hans Broekhuis. Amsterdam: John Benjamins, pp. 219–29.
- Veny, Joan. 1998. *Els parlars catalans (síntesi de dialectologia)*, 12th ed. Palma: Moll.
- Wedel, Andrew, Abby Kaplan, and Scott Jackson. 2013. High functional load inhibits phonological contrast loss: A corpus study. *Cognition* 128: 179–86. [\[CrossRef\]](#)
- Wetzels, Leo W., and Joan Mascaró. 2001. The typology of voicing and devoicing. *Language* 77: 207–44. [\[CrossRef\]](#)
- Wheeler, Max. 2005. *The Phonology of Catalan*. Cambridge: Cambridge University Press.
- Widdison, Kirk A. 1995. Physical constraints on sibilant-voice patterning in Spanish Phonology. In *Proceedings of the 1995 Desert Language and Linguistics Symposium*. Edited by Jeffrey Turley. Provo: BYU Linguistics Department, pp. 37–42.
- Widdison, Kirk A. 1997. Phonetic explanations for sibilant patterns in Spanish. *Lingua* 102: 253–64. [\[CrossRef\]](#)
- Żygis, Marzena, Susanne Fuchs, and Laura L. Koenig. 2012. Phonetic explanations for the infrequency of voiced sibilant affricates across languages. *Journal of Laboratory Phonology* 3: 299–336. [\[CrossRef\]](#)