

Pre-consonantal /s/-retraction

Mary Stevens, Véronique Bukmaier, Jonathan Harrington

Institute of Phonetics and Speech Processing, LMU Munich
mes;bukmaier;jmh@phonetik.uni-muenchen.de

ABSTRACT

In pre-consonantal position historical /s/ has changed to post-alveolar /ʃ/ in several different languages. This paper explores whether such cases of /s/ > /ʃ/ might share a common phonetic origin. We test whether /s/ is more retracted before /t/ in production and in perception. Production data show that word-medial /s/ has a lower first spectral moment (henceforth, M1) before /t/ than before vowels. Perception results show a bias towards retraction in pre-consonantal position.

Keywords: sound change, /s/-retraction, speech perception, speech production, typology

1. INTRODUCTION

1.1 Pre-consonantal /s/ > /ʃ/: historical, descriptive and phonetic evidence

In the historical development of Standard German, /s/ came to be produced as post-alveolar /ʃ/ in syllable-initial position before another consonant, e.g. *slagen* > *schlagen* ‘to beat’ [ʃla:gən]. In Swabian German historical /s/ is also realised as /ʃ/ in coda clusters, e.g. *fast* ‘almost’ [faʃt] (vs. Standard German [fast], [5]). Sibilant /s/ is similarly retracted to /ʃ/ before consonants in Standard European Portuguese and in some Brazilian dialects ([12, pp. 236]), e.g. *agosto* ‘August’ [a.ˈgoʃ.tu]. There is additional descriptive evidence of /s/-retraction before certain consonants (typically one or more of the voiceless stops /p, t, k/) in other languages spoken in Europe including Rhaeto-Romance, varieties of Italian and Slovenian, and in some Iranian languages ([12, pp. 232-7]). In many varieties of English /s/ is retracted in /str/, e.g. *strong*; this particular case of /s/-retraction is typically assumed to originate in anticipation of the rhotic during production of the sibilant (e.g. [1, 18]). The phonetic origin of /s/-retraction before singleton consonants, which is the focus of this paper, is less obvious. Phonetically motivated sound changes arise because of “physiological and psychological factors common to all human speakers at any time” ([15, pp. 238]), also known as phonetic biases (cf. [6]). If pre-consonantal /s/-retraction is a phonetically motivated

sound change, we should find a bias towards retraction in synchronic data for contexts and varieties in which this sound change has not yet occurred (cf. e.g. [2, 8] on the link between synchronic variability and diachronic sound change). That is, in English, for example, we should find that:

- (1) Speakers tend to produce a slightly more retracted /s/ in *steam* than in *seam*, and/or
- (2) Listeners should tend to hear the same [s] token as more /ʃ/-like in *steam* than in *seam*.

In terms of the experimental literature, there are some published studies relevant to (1), whereas the perception of /s/ in pre-vocalic and pre-consonantal contexts, i.e. (2), has not been tested. Most phonetic studies on fricative + stop sequences (e.g. [14, 17]) have concentrated on the contextual effects of the fricative on the stop, rather than the other way around. To the best of our knowledge only two experimental studies, both based on American English [1, 10] have compared /s/ in pre-consonantal and pre-vocalic contexts. In [1] the spectral properties of /s/ were measured in the word-initial clusters /sp, st, sk, spr, str, skr/ and before vowels (e.g. *seep*). The spectral centre of gravity (CoG) during the middle half of /s/ was reported to be lower – i.e. more /ʃ/-like – before consonants than before vowels. Acoustic data reported in [10], for first spectral moment (M1) trajectories during /s/, show a similar pattern i.e. lower values before /p, t, k/ than before vowels. These acoustic data, though not explicitly discussed in [1, 10], favour the idea that /s/-retraction in pre-consonantal contexts might be a phonetically motivated sound change. However, it is not clear whether acoustic differences in M1/CoG reflect articulatory variation by speakers, i.e. (1), and precisely which articulator would be involved. Complex articulatory-acoustic relations mean that lower spectral energy for /s/ could be due to speakers lowering the jaw, retracting the tongue tip or rounding the lips (or any combination of these) during /s/ (cf. [10, pp. 950]). Notably [10] reports that lip aperture and tongue dorsum gestures differ between pre-consonantal (in *special, school*) and pre-vocalic /s/ (in /sVd/ words), but they found no difference between tongue tip gestures.

Turning to (2), acoustic data reported in [1, 10] suggest that listeners could potentially confuse /s/ with /ʃ/, i.e. that pre-consonantal /s/-retraction might be partially listener-driven. However, listeners may

not be sensitive to acoustic variability during the sibilant noise, or they may be able to compensate for this contextual variation in perception (cf. e.g. [3, 9]). In light of these manifold possible explanations, this paper compares /s/ in pre-consonantal and pre-vocalic contexts. We report on two production and two perception experiments. We limit comparisons to /st/ vs. /sV/ (where V = vowel) because typological evidence suggests that /s/-retraction is more likely to occur before voiceless stops than other consonants and to avoid confounds due to consonant place of articulation effects.

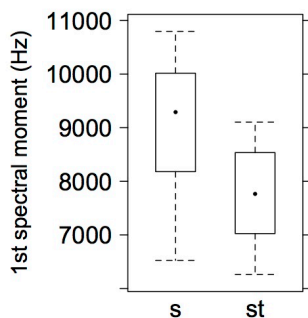
2. PRODUCTION

2.1 Production I: German

16 native speakers of German were recorded producing the words *Liste* ‘list’ /listə/, *Kiste* ‘box’ /kistə/, and *wisse* ‘He/she may know’ /visə/. These isolated word data were drawn from a larger corpus (cf. [5]). Here we analysed 368 /s/-tokens in total ((2 /st/-words × 10 repetitions + 1 /sV/-word × 3 repetitions) × 16 speakers). The data were segmented and labelled automatically using the Munich Automatic Segmentation System [19]; manual corrections were made in PRAAT [4]. Using EMU [7], we obtained the power spectrum for each target fricative by applying a 256-point discrete Fourier transform with a 40 Hz frequency resolution, 5 ms Blackman window and a frame shift of 5 ms. We then obtained the first spectral moment (M1) over 500-15000 Hz, and calculated the mean M1 within the temporal middle half (25-75%) of each fricative.

Fig. 1 shows that M1 for /s/ was lower before /t/ than in intervocalic position. A mixed model with M1 as independent variable, context as fixed factor and speaker as random factor confirmed this pattern ($\chi^2[1]=16.8, p < 0.001$).

Figure 1. M1 (in Hz) for word-medial /s/ before schwa (s) and /t/ (st). One data point for each speaker’s mean for each context.

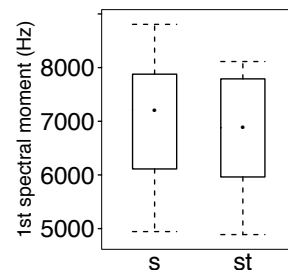


This result indicates that in Standard German, word-medial /s/ is slightly retracted before /t/ compared to pre-vocalic position. We tested whether these differences influence listener perception of /s/ in §3.2 but first we tested whether we can replicate acoustic results for word-initial /s/ in [1, 10].

2.2 Production II: English

Australian English speakers read the words *seem*, *sane* and *steam* in the carrier phrase “Any...” (3 words × 10 repetitions × 21 speakers). Procedures for data post-processing and analysis were the same as described in §2.1. Results plotted in Fig. 2 show one data point for each speaker’s mean. This plot shows that M1 during /s/ is lower preceding a voiceless stop /t/ than before vowels, i.e. there is measurable acoustic retraction in /st/ clusters. A mixed model with M1 as the independent variable, context as a fixed factor (2 levels) and speaker as a random factor showed that this difference in M1 is significant ($\chi^2[1]=22.4, p < 0.05$). This result is consistent with §2.1 and replicates existing studies on word-initial /s/ in (American) English [1, 10] which also found acoustic evidence of retraction before voiceless stops.

Figure 2. M1 (in Hz) for word-initial /s/ in pre-vocalic position (s) and preceding /t/.



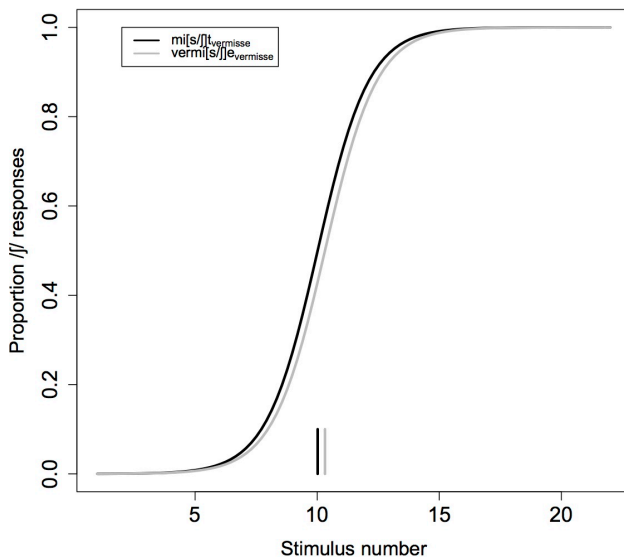
3. PERCEPTION

3.1 Perception I

Here we compare the perception of /s/ in word-medial codas preceding /t/ vs. /ə/. We tested whether listeners show a bias towards retraction in pre-consonantal position that may drive a sound change. To do so we created two synthetic continua between /s/ and /ʃ/: one between *Mist* (/mɪst/ ‘dung’) and *mischt* (/mɪʃt/ ‘to mix 3sg.’) and another between *vermisse* (/vəˈmɪsə/ ‘to miss 1sg.’) and *vermische* (/vəˈmɪʃə/ ‘to mix 1sg.’). We began with productions of *vermisse* and *vermische* spoken by a native speaker of Standard German and trained phonetician. Using STRAIGHT Tandem [11], these sound files were first morphed to create a 22-step continuum between /vəˈmɪsə/ and /vəˈmɪʃə/. The

[mis]...[miʃ]_{vermisse} sequence (i.e. the continuum) was cut out of each stimulus and sixteen were selected for the experiment (for further methodological details see also [5]). These 16 [mis]...[miʃ]_{vermisse} sequences were prepended to *_t* to create a 16-step *Mist-mischt*_{vermisse} continuum and spliced between *ver__e* to create a 16-step *vermisse-vermische*_{vermisse} continuum (subscript _{vermisse} denotes that [mis]...[miʃ] was synthesized from *vermisse/vermische* productions). The resulting stimuli were presented to 16 Standard German listeners in a randomized order with ten repetitions in a two-alternative forced-choice perception experiment. Listener responses were analysed with a logistic regression model with stimulus response (/s/ vs. /ʃ/) as the dependent variable, stimulus (16 levels: stimulus 1, 2, ...16) as the independent variable and listener as a random factor.

Figure 3. Psychometric curves for the proportion of /ʃ/ responses as a function of stimulus number (1 = [s]-endpoint and 22 = [ʃ]-endpoint) for the *mi[s/ʃ]t* (black) and *vermi[s/ʃ]e* (grey) continua.



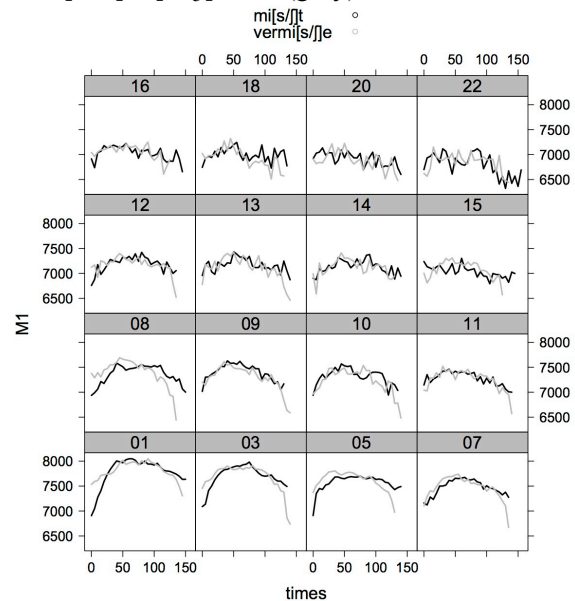
The output of this analysis was used to derive psychometric curves and decision boundaries separately for both continua. Fig. 3 shows a slight shift to the left for the *mi[s/ʃ]t*_{vermisse}-continuum, which indicates more /ʃ/-responses, i.e. a bias towards retraction in pre-consonantal vs. pre-vocalic position. However, a mixed model with decision boundary as the dependent variable, listener as a random factor, and continuum (*mi[s/ʃ]t*_{vermisse} vs. *vermi[s/ʃ]e*_{vermisse}) as fixed factor showed that this shift was not significant. At this stage we attribute this lack of a significant difference to our methods. More specifically, sibilant stimuli were derived from natural productions (*vermisse/vermische*) and then spliced into pre-consonantal and pre-vocalic contexts. This means that the sibilant noise may

have contained information for an upcoming vowel (cf. §2.1). If listeners had attended closely to the sibilant noise this would explain the lack of a significant difference between their responses to the two continua.

3.2 Perception II

We ran a second perception experiment to test whether listeners are sensitive to acoustic differences during the sibilant noise (recall from §2.1 that M1 is lower before /t/) and how this might interact with the effect on perception of the surrounding context just described at §3.1. To do so, we recorded the same speaker producing *Mist* and *mischt* and morphed these according to the same procedures described for *vermisse-vermische* at §3.1. This gave us a second 16-step [mis]...[miʃ]_{Mist} continuum, this time derived from natural pre-consonantal *Mist/mischt* productions (denoted with subscript ‘_{Mist}’). M1 trajectories for both 16-step /s...ʃ/ continua (i.e. derived from pre-consonantal vs. pre-vocalic contexts) are shown in Fig. 4.

Figure 4. M1 for sibilants in [mis]...[miʃ]_{Mist} (black) and [mis]...[miʃ]_{vermisse} (grey) continua.



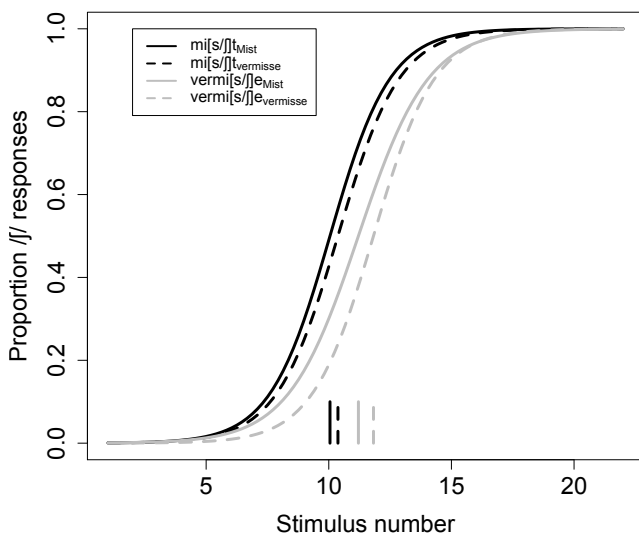
The [mis]...[miʃ] stimuli were spliced into *ver__e* and *_t* contexts resulting in four continua:

1. *mi[s/ʃ]t*_{Mist}
2. *mi[s/ʃ]t*_{vermisse}
3. *vermi[s/ʃ]e*_{vermisse}
4. *vermi[s/ʃ]e*_{Mist}

Continuum 2 and continuum 3 were compared in the first experiment; these were included again here because the listener group was comprised of different individuals. Psychometric response curves were fitted to the response data for 20 Standard German listeners using the same procedure

described at §3.1. Fig. 5 shows more /ʃ/-responses for *mi[s/ʃ]t* stimuli (in black), irrespective of whether the sibilant was originally pre-vocalic (dashed) or pre-consonantal (unbroken). The results also showed more /ʃ/-responses for sibilants originally produced in pre-consonantal (unbroken) than pre-vocalic (dashed) contexts. A mixed model with decision boundary as dependent variable, original context ('*vermisse*' vs. '*Mist*'), word (*mi[s/ʃ]t* vs. *vermi[s/ʃ]e*) as fixed factors and listener as a random factor showed a significant influence of original context ($\chi^2[2]=6.9$, $p < 0.05$) and of word ($\chi^2[2]=42.7$, $p < 0.001$) but no interaction between the two.

Figure 5. Psychometric curves for the proportion of /ʃ/ responses as a function of stimulus number for the four continua.



These results show that contextual differences during the sibilant noise for /s/ (as seen in the trajectories in Fig. 4) influence listener decisions: there were more /ʃ/-responses for sibilants originally produced in *Mist-mischt* ('*Mist*') than those extracted from *vermisse-vermische* ('*vermisse*'). Moreover, word also influenced listener decisions: inserting [mis]...[miʃ] into ___t contexts increased /ʃ/-responses. These results show that listeners were sensitive to differences during the fricative noise because '*Mist*' stimuli gave more /ʃ/-responses. However, listeners did not compensate for this acoustic information in perception, because in that case they would have attributed lower M1 to the presence of the /t/, giving more /s/-responses for *mi[s/ʃ]t_{Mist}* than *vermi[s/ʃ]e_{Mist}*. Instead Fig. 5 shows the opposite pattern.

5. DISCUSSION

This paper explored whether there is any evidence of a phonetic bias towards /s/-retraction in pre-

consonantal position that might explain its attested occurrence in several different unrelated languages. We found acoustic evidence for a bias towards pre-consonantal /s/-retraction in production: M1 measured during the middle half of /s/ was lower before /t/ than before vowels. This was the case for word-medial and word-initial /s/, measured in German and English, respectively. Notably, our result for word-initial /s/ in English replicates [1, 10]. We are currently conducting an articulatory (EMA) experiment with German and English speakers to investigate how acoustic differences in M1 relate to articulatory variation (tongue tip, lip or jaw movements) during the production of /s/ in different contexts.

Perception results showed that listeners hear the same sibilant token as more retracted when it occurs before /t/ than before schwa (§3.1); this effect was replicated for a different group of listeners in the second experiment. A similar pattern was reported in [17] in which an /s...ʃ/ continuum was prepended to [ta] and [ka]; more /ʃ/-responses were also reported for the /t/ context. We have no explanation at present as to why a following /t/ should favour more /ʃ/-responses from listeners. Our second perception experiment showed additionally that listeners were sensitive to contextual information during the sibilant noise itself. That is: /s/ was acoustically more retracted before /t/ (cf. Fig. 1) and listeners gave more /ʃ/-responses for sibilants extracted from such pre-/t/ contexts. Thus the presence of a following /t/ and spectra with a lower centre of gravity during the sibilant itself could together drive a listener-driven sound change /st/ > /ʃt/. Here we only tested the perception of word-medial /st/ vs. /sə/; further experiments are needed to determine whether there is a similar bias towards /ʃ/ in the perception of word-initial /st/.

Further research is also needed to clarify the effect of syllable-position on the production and perception of /s/. The idea that /s/ varies according to syllable position conflicts with the notion that it is extremely resistant to coarticulation [16]. Nonetheless, in Portuguese /s/ has changed diachronically to /ʃ/ in all syllable codas [12, 13] e.g. *tirar fotos* [tɨrɐˈfotʃ] 'to take photos' (i.e. not only pre-consonantly as in *agosto*), and the typological overview in [12] suggests that coda position before voiceless stops is a favourable condition for /s/-retraction. On the other hand, historical /s/-retraction in Standard German is confined to onset position e.g. *Stimme* 'voice' [ˈʃtɪ.mə] vs. *fast, lustig* 'almost, funny' [ˈfast], [ˈlus.tɪk], and production data (in §2.2 as well as [1, 10]) suggest that retraction in word-initial position might be phonetic in origin, as well.

7. REFERENCES

- [1] Baker, A., Archangeli, D., & Mielke, J. (2011). Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change*, 23, 347-374.
- [2] Beckman, M., De Jong, K., Jun, S.-A., & Lee, S.-H. (1992). The interaction of coarticulation and prosody in sound change. *Language and Speech*, 35, 45-58.
- [3] Beddor, P. S., & Krakow, R. A. (1999). Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *J. Acoust. Soc. Am.*, 106(5), 2868-2887.
- [4] Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer (Version 5.3.27).
- [5] Bukmaier, V., Harrington, J., & Kleber, F. (2014). An analysis of post-vocalic /s-/ neutralization in Augsburg German: evidence for a gradient sound change. *Frontiers in Psychology*, 5, 1-12.
- [6] Garrett, A., & Johnson, K. (2013). Phonetic bias in sound change. In A. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (pp. 51-97): Oxford University Press.
- [7] Harrington, J. (2010). *The phonetic analysis of speech corpora*. Oxford: Wiley-Blackwell.
- [8] Harrington, J. (2012). The relationship between synchronic variation and diachronic change. In A. Cohn, C. Fougerson & M. Huffman (Eds.), *Handbook of Laboratory Phonology* (pp. 321 - 332). Oxford: Oxford University Press.
- [9] Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *J. Acoust. Soc. Am.*, 123(5), 2825-2835.
- [10] Iskarous, K., Shadle, C. H., & Proctor, M. I. (2011). Articulatory-acoustic kinematics: The production of American English /s/. *J. Acoust. Soc. Am.*, 129(2), 944-954.
- [11] Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., and Banno, H. (2008). Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation, *Proc. ICASSP*, Las Vegas, USA, 3933-3936.
- [12] Kümmel, M. J. (2007). *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen*. Wiesbaden: Reichert.
- [13] Mateus, M. H., & Andrade, E. (2000). *The Phonology of Portuguese*. Cambridge: Cambridge University Press.
- [14] Nguyen, N., Wrench, A., Gibbon, F., & Hardcastle, W. (1998). Articulatory, acoustic and perceptual aspects of fricative/stop coarticulation. *Proc. ICSLP*, Sydney, Australia (no page numbers).
- [15] Ohala, J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical Linguistics: Problems and Perspectives* (pp. 237-278). London: Longman.
- [16] Recasens, D. (2004). The effect of syllable position on consonant reduction (evidence from Catalan consonant clusters). *Journal of Phonetics*, 32, 435-453.
- [17] Repp, B. H., & Mann, V. A. (1981). Perceptual assessment of fricative-stop coarticulation. *J. Acoust. Soc. Am.*, 69(4), 1154-1163.
- [18] Rutter, B. (2011). Acoustic analysis of a sound change in progress: The consonant cluster /stɪ/ in English. *J. International Phonetic Association*, 41, 27-40.
- [19] Schiel, F. (1999). Automatic Phonetic Transcription of Non-Prompted Speech. *Proc. 14th ICPhS*, San Francisco, 607-610.