Predicting sound change: acoustic evidence from a longitudinal study of /s/-retraction in English /str/

Mary Stevens

Institute for Phonetics and Speech Processing (IPS), LMU Munich
mes@phonetik.uni-muenchen.de

ABSTRACT

This study is concerned with the predictability of regular sound change at the population level and within individual grammars. Taking /s/-retraction in Australian English /str/ as a case study, it presents an acoustic analysis of longitudinal data from production experiments with eighteen speaker participants from a single speech community. Coarticulatory /s/-retraction was evident for all speaker participants in /str/ at the baseline. After eight intervening years, incremental shifts in the same direction (towards /ʃ/) were evident for most speakers. Comparison with the outcome of simulated interaction in an agent-based model suggests that regular sound change is predictable at the population level but not at the level of the individual speaker.

Keywords: sound change, longitudinal studies, individual variation, agent-based models, /s/-retraction

1. BACKGROUND

Population-level sound change can be examined by comparing speech produced by older and younger speakers of the same variety, who are understood to represent earlier and later stages, respectively, in the progression of a sound change (e.g., [2, 9, 12]). Such apparent time studies rest on the assumption that pronunciations are stable throughout adulthood – and can therefore be compared between generations. However, there is evidence from case studies that adult speakers' pronunciations can change incrementally in the direction of sound changes going on in the language spoken around them (e.g., [10]), as well as due to biological ageing (e.g., [15]) – thus potentially complicating comparisons between different age groups. There is also evidence that the tendency to shift in the direction of population-level sound changes varies between individuals. Longitudinal studies of sound changes in progress involving multiple speaker participants are rare, but [17] analysed Montreal French speakers who were recorded once in 1971 and again thirteen years later, during which time Montreal French was undergoing a sound change from [r] -> [R]. Based on the authors' auditory categorization, most speakers' rhotic pronunciation was found to have remained stable – only nine (28%) speakers were found to have significantly increased their use of innovative [R] in the later recordings, even though a sound change was in progress at the population level. The relative rarity of change within individual participants' lifespans may be due to the nature of the [r] -> [R] sound change in Montreal French, which has its source in dialect contact (rather than coarticulatory variation, as is instead the case for regular sound change).

The present study concentrates on the effects of regular sound change on speakers' production repertoires in adulthood. The sound change chosen for investigation was /s/-retraction in English /str/, whereby e.g., the sibilant in *street* comes to resemble that in *sheet*. /s/-retraction in /str/ has been shown to have its origins in coarticulation ([18]) and has been the subject of several phonetic investigations in various varieties of English (e.g., amongst others, [2, 3, 16, 18, 20]). This sound change has not yet been the subject of a longitudinal study with real speakers, but in [19] interaction was simulated amongst a group of speakers (agents) in an agent-based model (ABM). Before simulated interaction, the speakers' /str/ pronunciations showed coarticulatory /s/-retraction in /str/; afterwards, there was acoustic evidence of a population-level shift in the direction of /ʃ/. This result suggests that regular sound change is the result of the gradual accumulation of synchronic variation, in combination with density of communication [5]. However, the question remains as to whether the outcomes of simulated interaction match longitudinal changes within real speakers' production targets, i.e., whether regular sound change proceeds via incremental changes to individual grammars, even in adulthood (which would conflict with an assumption of the apparent time method, as noted above). The present study addresses this question with an acoustic analysis of sibilants in /str/ that were produced by the same speakers in two separate recording sessions, eight years apart. In section 4.2, the predictability of regular sound change is addressed by comparing our results concerning the effects of real-world interaction on speakers' /str/ pronunciation with those reported in [19], in which interaction was simulated between the same speakers (agents) in an agent-based model.

2. METHODS

2.1 Participants and recording setup

Participants were all adult monolingual speakers of Australian English and long-term residents of Braidwood, a semi-rural town in New South Wales. They were originally recruited for production experiments conducted in 2014. The present study is based on data for eighteen speakers who took part in a follow-up recording session in early 2022. Participants were paid for their participation in both recording sessions. The participants (thirteen female, five male) were aged between 29 and 49 years at the time of the first recordings. They were mostly known to each other and some reported interacting with each other daily. In terms of contact with non-local varieties, only one participant reported having spent more than one of the intervening eight years (in total) outside of the state of New South Wales.

2.2 Recording procedure and stimuli

All recording sessions took place in the same room in a private house. Participants wore a headset microphone and read words aloud as they appeared on a laptop computer screen. The words in Table 1 were presented to participants in isolation (no carrier phrase) in a randomized order, with ten repetitions in the first recording session and nine repetitions in the second. The segmental and prosodic contexts for the target sibilant were matched as closely as possible across the three (/s/ and /ʃ/, and /str/) word types. Due to a small number of speech errors and recording errors in both recording sessions, 7369 sibilant tokens from the first recording session and 6640 sibilant tokens from the second recording session (of a possible 7380 and 6642, respectively) were available for analysis.

/s/	*assault, assembly, fascinating, messy, Minnesota, motorcycle, policy, possible, sane, seem, soak*
/ʃ/	*dishevelled, information, machine, overshadowed, passionate, perishable, polishing, sheep, Shane, show, tissue*
/str/	*administrate, astringent, astronaut, astronomy, catastrophe, catastrophic, chemistry, claustrophobic, destroy, district, gastro, oestrogen, pastrami, pedestrian, orchestra, stream, restrict, strong, strut*

Table 1: Target words containing prevocalic /s/ or /ʃ/, or a /str/ cluster.

2.3 Acoustic analysis

The signal data from each recording session were labelled semi-automatically using WebMAUS (e.g., [11]) and converted into two EMU databases ([21]). Segment boundaries were corrected manually. The first spectral moment (M1), which indexes place of articulation, was calculated as a trajectory between the temporal onset and offset of each sibilant token. M1 trajectories were high pass filtered to remove all spectral information below 500 Hz and were z-score transformed (across all data frames for all sibilants, separately by speaker and recording session) using the speaker-normalization method in [13]. Rather than tracking the absolute location of sibilants in acoustic space using e.g., the M1 trajectories, /s/-retraction was measured relatively, consistent with other recent studies ([2, 19]). More specifically, it was measured in terms of the *orthogonal projection* (OP), which tracks the relative location of any sibilant token between the same speaker's /s/-words (centred at +1) and /ʃ/-words (centred at –1). The OP can be interpreted similarly to M1 or spectral centre of gravity, i.e., higher/lower OP indicates a more /s/- or /ʃ/-like pronunciation, respectively. To obtain the OP for each sibilant token, the speaker normalized M1 trajectories were parameterised into three cepstral coefficients using Discrete Cosine Transformation (DCT) (e.g. [7]), corresponding to their height, slope, and curvature. Each sibilant token's position in this three-dimensional DCT space on a line passing through speaker-specific centroids for /s/- and /ʃ/-words, namely the OP, was then calculated. Note that the OP was calculated within each recording session separately, that is, the position of each sibilant token was calculated relative to centroids for /s/- and /ʃ/-words produced during the same recording session (and by the same speaker).

3. RESULTS

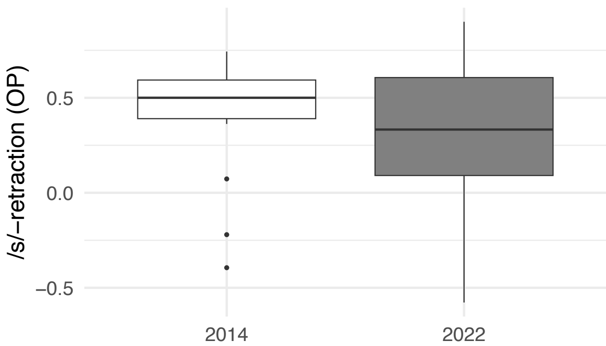

Figure 1: OP for sibilants in /str/, by recording session. Each data point depicts one speaker's mean OP in all /str/-words and repetitions. Lowering of OP in 2022 indicates increased /s/-retraction.

Figure 1 shows the mean OP for sibilants in /str/ in the two recording sessions. Lowering of OP in 2022 (dark grey) indicates increased acoustic proximity to /ʃ/. A mixed effects model was run to test the effect of the intervening years on sibilants in /str/ with OP as the dependent variable, Recording session (two levels, 2014 or 2022) and Gender (two levels: male or female) as fixed factors, and Word (41 levels) and Speaker (18 levels) as random factors. The results showed a significant effect for Recording session (Chi Square = 4.8, $p < 0.05$) which confirmed that OP was significantly lower (= increased acoustically measurable /s/-retraction) in 2022. Neither the fixed factor Gender nor its interaction with Recording session was found to have a significant effect on OP. It is evident in Figure 1 that there was an increase in the interquartile range from 2014 to 2022, which suggests divergence between speaker participants over time. To examine individual speakers' behaviour more closely, the data are shown separately by speaker in Figure 2.

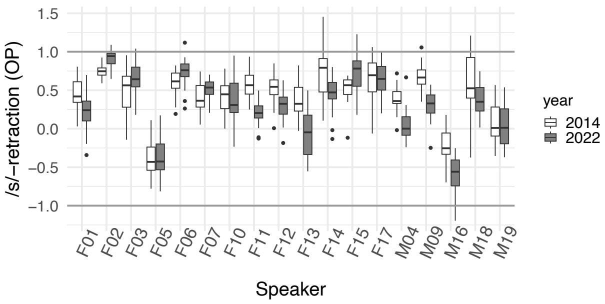

Figure 2: The same data from Figure 1, by speaker. The two thick horizontal lines indicate the OP for sibilants in /s/-words and /ʃ/-words, centred at 1 and -1, respectively.

First, all participants' OP for sibilants in /str/ was lower than that for pre-vocalic /s/, which was centred at 1 (cf. section 2.3) and is indicated by the upper horizontal line in Figure 2. Second, most – but not all – speakers' median OP showed a further downwards shift towards /ʃ/ (indicated by the lower horizontal line) in 2022. This shift towards /ʃ/ is consistent with the group-level pattern seen in Figure 1. In other words, all speakers showed acoustically measurable /s/-retraction in /str/ in 2014 and most speakers showed an increase in the degree of /s/-retraction over time. However, upward shifts in OP (i.e., increased proximity to /s/) can be seen for five speakers (F02, F03, F06, F07, F15) and two further speakers' (F05, M19) OP in /str/ remained almost identical over time. Thus, it appears that some individual speakers did not participate in the sound change evidently underway at the population level (Fig. 1), i.e., in the language spoken around them.

Note that the five male speakers (M04-M19, on the right in Fig. 2) all showed a downwards shift from 2014 to 2022, whereas female speakers (F01-F17) differed in terms of whether they showed an increase or a decrease in the degree of /s/-retraction over time. This dataset was not balanced for speaker gender, with thirteen females but only five males, and it is unclear whether resistance to shifts in the direction of /s/-retraction might be specific to female speakers or simply an artefact of our unbalanced dataset with too few male participants. In terms of the literature on /s/-retraction in English /str/, patterns differ between varieties and it is not clear whether males or females typically lead the change (see e.g., [20]). The tendency for some individuals to resist shifting in the direction of the population-level sound change is discussed further in section 4.

Some phonetic models of sound change propose that participation in a sound change involves a categorical change to an individual's phonetic repertoire, replacing a conservative variant with an innovative variant (e.g., [3, 14]). Patterns in Fig. 2 show that distributions for /str/ in the two recording sessions mostly overlapped and there was no evidence of a speaker making a categorical change from a conservative pronunciation (in this case very /s/-like) in 2014 to an innovative pronunciation (in this case very /ʃ/-like) in 2022. Thus /s/-retraction in this variety appears to involve incremental changes to speakers' pronunciation targets, in line with the idea that sound change involves a gradual shift from one gestural coordination pattern to another (e.g., [6]) and with usage-based models (e.g., [4]) in which change is incremental as language users' representations are updated over time.

It has also been suggested [3] that the tendency to shift in the direction of /s/-retraction might be related to the degree to which a speaker coarticulates in their own speech production. However, in the present dataset, speakers with relatively similar degrees of /s/-retraction in 2014, such as F01 and F07, did not behave similarly over time: a downwards shift was evident for speaker F01, for example, while speaker F07 shifted upwards; see also the speaker pair F07 and F13, or the speaker pair F05 and M16. Thus, there is no evidence in Figure 2 that the tendency for an individual to participate in the sound change involving /s/-retraction depended on the degree of coarticulation in their own speech production.

4. DISCUSSION

4.1 Longitudinal change in real speakers' /str/

This study explored whether a coarticulatory tendency towards /s/-retraction that was common to speaker participants in 2014 increased over a period of eight intervening years. Acoustic results showed evidence of a population-level shift in the direction of /s/-retraction which was also evident within most individual speakers. The relatively high proportion of speakers whose /str/ pronunciation shifted with the (shifting) population-level norm differs from [17] in which only a minority of speakers were found to have significantly increased their use of innovative [R] over an intervening 13-year period. Indeed Sankoff [17, pp. 576)] concluded that the sound change [r] -> [R] in Montreal French is *not* driven by “speakers slowly increasing their use of [R] across their lifetimes”, but rather through new speakers (with more innovative pronunciations) being added to the population. These differences in terms of the proportion of individuals whose production repertoires were found to have changed might best be attributed to the different sound changes under investigation: regular sound change with its origins in coarticulatory variation (here) vs. sound change due to dialect contact [17].

In the present study, the five speakers whose /str/ pronunciation shifted in the opposite direction to the population-level shift were all female, but a higher number of male participants would be necessary to determine whether participation in this sound change is conditioned by speaker gender (and see e.g. [20]).

In order to better understand the factors that shape population-level sound change, future studies might also explore why real speakers differed in terms of their longitudinal production behaviour. For example, resistance to shift in the direction of population-level sound change may be governed by awareness of the sound change (e.g., [1]), the ability to detect /s/-retraction in perception or to control one’s own speech production.

4.2 Effects of real vs. simulated interaction

Here we compare our longitudinal data for real speakers with the outcomes of simulated interaction in an agent-based model (ABM) [19]. The input data to the ABM were the parameterized M1 trajectories (cf. section 2.3) from the 2014 recording session, with additional data from one male speaker who was not available for the follow-up recordings in 2022. Each speaker was represented in the ABM as an agent, with their own sibilant production data stored in memory. The ABM involved 60,000 speaker-listener interactions and was repeated 100 times; here we consider the outcome of a randomly chosen simulation run. Figure 3 shows the mean OP for sibilants in /str/ in 2014 (on the left), after eight years of real-world interaction (2022, in the middle) and after simulated interaction (ABM, on the right).

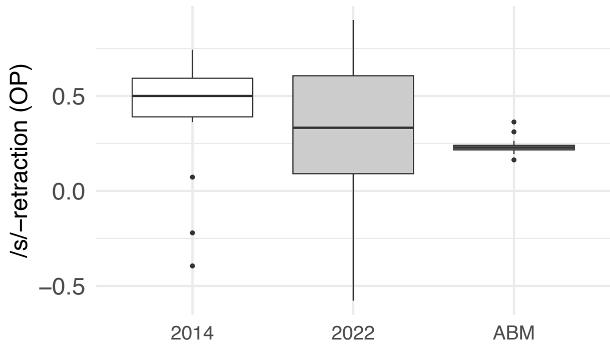

Figure 3: The same data as in Figure 1, with the addition of the right-hand box showing the outcome of simulated interaction in an agent-based model (ABM). Lowering of OP indicates increased /s/-retraction. 19 speaker-agents interacted in the ABM, data for 18 of these are shown at the baseline (2014) and after eight years (2022).

It is evident from Figure 3 that both simulated (ABM) and real-world interaction (2022) resulted in a downwards shift for sibilants in /str/, compared to the baseline (2014). Thus, the ABM accurately predicted the direction of the population-level shift that took place from 2014 to 2022. This pattern is consistent with results reported in [8], which also compared the effects of simulated versus real-world interaction. In that study, population-level vowel shifts resulting from interaction amongst an isolated group of speakers were – to a certain extent – accurately predicted by the ABM. However, most of the changes observed in [8] were found to be due to the influence of a small number of speakers (with different accents). Here, on the other hand, we have seen that tendencies that are common amongst speakers (like coarticulatory /s/-retraction) can accumulate via interaction – real and simulated, causing population-level shifts. Compared to the baseline (i.e., 2014), the interquartile range increased in the real world (2022) but decreased in the ABM. This indicates that there was divergence between real speakers but convergence between speaker-agents in the ABM (each data point in Fig. 3 shows the mean OP for one speaker/agent’s sibilants in /str/). In other words, the ABM accurately predicted the direction of the population-level change that occurred between 2014 and 2022, but not the changes that took place within individual grammars. The evident convergence between individuals in the ABM is primarily due to the structure of the model: agents interacted only with each other and exchanged only the lexical items in Table 1, whereas real interactions would have involved a larger number of interlocutors and lexical items. As noted in section 4.1, a fruitful approach for future research might be to explore the reasons why some real speakers’ /str/ pronunciations did not shift with the (shifting) population-level norm.

5. ACKNOWLEDGMENTS

This research was supported by German Research Council Grant No. 387170477.

6. REFERENCES

1 Babel, A. 2016. *Awareness and control in sociolinguistic research.* Cambridge, Cambridge University Press.

2 Bailey, G., Nicholls, S., Turton, D., Baranowski, M. 2022. Affrication as the cause of /s/-retraction: Evidence from Manchester English. *Glossa: a journal of general linguistics* 7, (1): 1–30.

3 Baker, A., Archangeli, D., and Mielke, J. 2011. Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change* 23: 347–374.

4 Beckner, C., Blythe, R., Bybee, J., Christiansen, M.H., Croft, W., Ellis, N.C., Holland, J., Ke, J., Larsen-Freeman, D., and Schoenemann, T. 2009. Language is a complex adaptive system: position paper. *Language learning* 59, (s1): 1–26.

5 Bloomfield, L. 1933. *Language.* George Allen & Unwin.

6 Browman, C. and Goldstein, L. 1992. Gestural structures and phonological patterns. In Mattingly, I.G., and Studdert-Kennedy, M. (Eds.) *Modularity and the Motor Theory of Speech Perception*. Erlbaum.

7 Harrington, J. 2010. *The phonetic analysis of speech corpora*. Wiley-Blackwell.

8 Harrington, J., Gubian, M., Stevens, M., and Schiel, F. 2019. Phonetic change in an Antarctic winter. *Journal of the Acoustical Society of America* 146: 3327–3332.

9 Harrington, J., Kleber, F., and Reubold, U. 2008. Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *Journal of the Acoustical Society of America* 123, (5): 2825–2835.

10 Harrington, J., Palethorpe, S., and Watson, C.I. 2000. Does the Queen speak the Queen's English? *Nature* 408: 927.

11 Kisler, T., Reichel, U., Schiel, F. 2017. Multilingual processing of speech via web services. *Computer, Speech and Language* 45, 326–347, doi:10.1016/j.csl.2017.01.005.

12 Kuang, J., and Cui, A. 2018. Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. *Journal of Phonetics* 71. 194-214.

13 Lobanov, B.M. 1971. Classification of Russian Vowels Spoken by Different Speakers. *Journal of the Acoustical Society of America* 49: 606.

14 Ohala, J. 1993. The phonetics of sound change. In Jones, C. (Ed.) *Historical Linguistics: Problems and Perspectives.* Longman: 237-278.

15 Reubold, U., Harrington, J. 2017. The influence of age on estimating sound change acoustically from longitudinal data. In Wagner, S.E. and Buchstaller, I. (Eds.) *Panel Studies of Language Variation and Change*. Routledge. 129-152.

16 Rutter, B. 2011. Acoustic analysis of a sound change in progress: The consonant cluster /stɹ/ in English, *Journal of the International Phonetic Association* 41. 27 – 40.

17 Sankoff, G., and Blondeau, H. 2007 'Language change across the lifespan: /r/ in Montreal French'. Language 83, 3. 560-588.

18 Stevens, M., and Harrington, J. 2016. The phonetic origins of /s/-retraction: Acoustic and perceptual evidence from Australian English. *Journal of Phonetics* 58. 118-134.

19 Stevens, M., Harrington, J., and Schiel, F. 2019. Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English. *Glossa: a journal of general linguistics* 4 (1). 1–30.

20 Stuart-Smith, J., Sonderegger, M., Macdonald, R., Mielke, J., McAuliffe, M., and Thomas, E. 2019 Large-scale acoustic analysis of dialectal and social factors in English /s/-retraction'. *Proc. ICPhS 2019*, Melbourne.

21 Winkelmann, R., Harrington, J., and Jänsch, K. 2017. EMU-SDMS: advanced speech database management and analysis in R. Computer Speech & Language 45. 392-410.