Individual differences and sound change actuation: evidence from imitation and perception of English /str/

Mary Stevens[a], Debbie Loakes[b]

[a]Institute of Phonetics & Speech Processing, University of Munich, Germany.
[b]Centre of Excellence for the Dynamics of Language & Phonetics Laboratory, University of Melbourne, Australia.
mes@phonetik.uni-muenchen.de; dloakes@unimelb.edu.au

ABSTRACT

This study investigates the role of individual differences in the earliest stages of sound change, taking as a case study /s/-retraction in <str>. Australian English speakers completed an auditory repetition task involving isolated words with word-initial sibilants. In <str>-words the sibilant was manipulated to resemble post-alveolar [ʃ]. The same participants also completed a forced choice perception task involving sibilants. The study tests two predictions: (1) the sibilant in <str>-words should shift in the direction [s] → [ʃ] during exposure; (2) the magnitude and direction of shifts within individuals should depend on their own phonetic repertoire (production and perception). Results did not support (1) but there was partial support for (2) from production. The implications of the results for models of sound change are discussed.

Keywords: sound change, s-retraction, individual differences, imitation, perception

1. INTRODUCTION

Even individuals with very similar linguistic backgrounds can differ in terms of their production and perception of speech e.g. [4, 13]. Building on Ohala's [11] model, it has been proposed [2] that sound change might come about via chance interactions between individuals with extremely different degrees of coarticulation. More specifically, those who produce coarticulation can recognize (and filter out) its effects whereas an individual who does not coarticulate might mistake an extremely coarticulated variant as an alternate variant intended by the speaker, and adopt this variant in their own speech [2]. Thus a novel and extremely coarticulated variant might spread to other individuals, eventually culminating in permanent change to the canonical form. This account is of interest because it suggests a possible solution to the challenge of sound change actuation [15, 18]. However, the idea that listeners-turned-speakers reproduce novel variants in their own subsequent productions [2, 11] assumes a direct link between individual language users' perception and production repertoires that has been difficult to demonstrate in experimental studies e.g. [3, 6] (but see also [19]). Moreover, while there is experimental evidence that coarticulatory differences can be imitated [20], whether this might be mediated by the degree to which an individual produces coarticulation has not been directly tested. The present study investigates the role of individual differences in the earliest stages of sound change using the imitation paradigm (e.g. [1, 10]) and taking /s/-retraction in English <str> as a test case. This sound change originates in coarticulation with /r/ that causes spectral centre of gravity lowering during the sibilant [16]. /s/-retraction has progressed to completion in some varieties of English in which the sibilant in <str> resembles /ʃ/ e.g. [14]. In Australian English, synchronic variation includes coarticulatory /s/-retraction in <str>, but for most speakers the sibilant nonetheless remains acoustically closer to /s/ than /ʃ/ [16]. In the present study, a group of Australian English speakers was exposed to increased /s/-retraction in <str>-words in an auditory repetition task. Our prediction is that individuals with little /s/-retraction in their own speech will show larger shifts in the direction [str] → [ʃtr] than those who already produce /s/-retraction at baseline (for whom there is arguably little to imitate [2]). Second, the same participants took part in a perceptual categorization task (/s/ vs. /ʃ/), allowing the influence of perception vs. production repertoires on imitation behaviour to be distinguished. We test the prediction that individuals who in perception categorize an ambiguous sibilant in <str> as /ʃ/ should be more likely to shift their production target towards [ʃtr] than those who categorize the same sibilant as /s/.

2. AUDITORY REPETITION TASK

2.1 Participants

16 adult monolingual Australian English speakers (age range 29-48; 9 females) took part in an auditory

repetition task. All were long-term residents of the same small rural town (Braidwood, NSW) and none had ever lived outside of Australia. The results of a production task [16] showed that the degree of /s/-retraction in <str> varied amongst these speakers. The present study reports data from a subset of the participants who completed the production task; individual participant IDs (cf. Figs. 2 & 4) were maintained to allow comparison with [16].

2.2 Materials and recording procedure

Materials comprised twenty-nine monosyllabic English words with word-initial /s/ and /ʃ/ (target words) and fifteen monosyllabic filler words that did not contain a sibilant. The target words comprised <str>-words as well as prevocalic <s>- and <ʃ>-words (e.g. *strain, sane, Shane*). These materials were presented to participants in written and audio form. To obtain the audio materials, a female native speaker of Australian English (not one of the participants) read all words aloud from a computer screen while wearing a headset microphone. <str>-words were subsequently modified by replacing the sibilant with one that was ambiguous i.e. more /ʃ/-like. This sibilant was obtained as follows: the speaker produced the word *strain* once as [stɹæɪn] and once as [ʃtɹæɪn]. The latter is an artificial pronunciation but this method was chosen so that the post-alveolar sibilant noise would include appropriate contextual information (which would not be the case with a pre-vocalic [ʃ]). The sibilant was then spliced from both (i.e. [stɹæɪn] and [ʃtɹæɪn]) pronunciations and an 11-step [s…ʃ] continuum was obtained by interpolating between the two.[1] One interpolation from the middle of the resulting 11-step [s…ʃ] continuum (step 6) was then chosen and spliced into all target words containing <str>. The mean M1 (defined in 2.3) for this sibilant was 4943.3 Hz.

The experiment was conducted in a private home. Following [10] the recording procedure involved three consecutive parts: (1) baseline recording, (2) exposure to model talker and (3) post-exposure recording. Results from (3) are not reported here. During (1), participants read written words as they appeared on a laptop computer screen while wearing a headset microphone. During (2), participants heard over headphones the same words and were instructed to "identify the word you hear by repeating it out loud", after [1].

[1]http://www.holgermitterer.eu/HM/sample_interpolation.praat

2.3 Acoustic analysis

The acoustic analysis is based on the first spectral moment (M1), which reliably distinguishes between [s] and [ʃ] and is therefore an appropriate acoustic parameter for measuring /s/-retraction. As in [16, 17] /s/-retraction was quantified as the relative distance of sibilants in <str>-words between speaker-specific centroids for prevocalic /s/ and /ʃ/. For this study centroids were calculated on baseline data (i.e. before exposure to the model talker), and to increase the reliability of these centroids we included an additional 40 prevocalic /s/ and /ʃ/ tokens per participant (10 repetitions of *seem, sane, sheep, Shane*) from an earlier production task with the same speakers [16]. The first step was to obtain, for each sibilant token, an *M1 trajectory* and *mean M1* (averaging over the temporal middle half). Each M1 trajectory was then parameterized using DCT to obtain its mean, slope and curvature (DCT-coefficients 0, 1 and 2, respectively). We then calculated, using orthogonal projection [17], the position of each <str> token in the three-dimensional DCT-space on a line passing through the speaker-specific /s/ and /ʃ/ centroids (which were centred at 1 and -1, respectively). This gave a *distance ratio* for each <str>-sibilant token, corresponding to its relative position between the same speaker's /s/ and /ʃ/ centroids. Positive values indicate closer proximity to /s/, negative values indicate closer proximity to /ʃ/ and a value of zero indicates that the sibilant was intermediate between the two.

2.4 Results

Figure 1: By participant average difference in distance ratio for sibilants in <str> words during exposure compared to baseline.

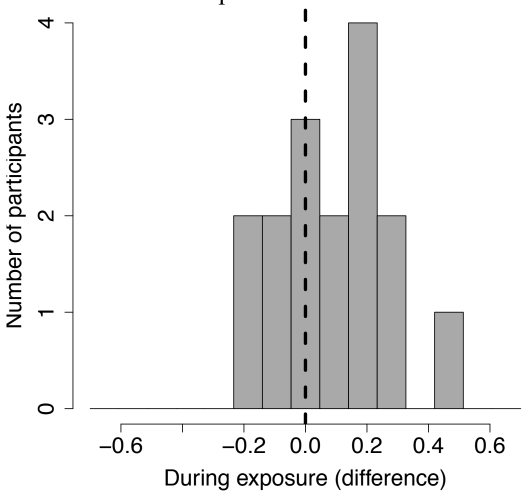


Figure 1 shows by participant differences in distance ratio for sibilants in <str>-words during exposure compared to baseline (averaged across all words). The distribution in Figure 1 is skewed to the right, which indicates a trend towards increased proximity to /s/ during exposure. A t-test confirmed that this trend was significant (one-sided t(15) = 2.2019, p = 0.022). Nonetheless, Figure 1 also shows that in some participants there was a shift towards /ʃ/ during exposure and that in others there was little change. We now investigate whether these differences during exposure might be explained by the degree of /s/-retraction in participants' own speech.

Figure 2: By participant average distance ratios for sibilants in <str>-words at baseline (x-axis) against change during exposure (y-axis).

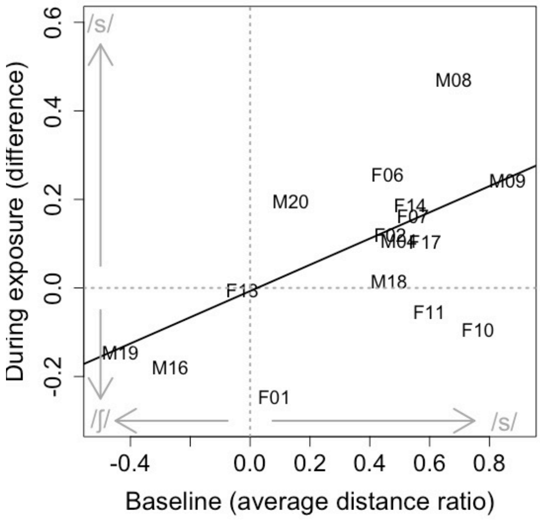


Figure 2 shows participants' average distance ratios for sibilants in <str> at baseline (x-axis) against the data from Figure 1 (during exposure). The dashed vertical line separates baseline data according to whether sibilants in <str> were on average closer to the participant's own /s/ (right) or /ʃ/ (left). The dashed horizontal line separates the data according to whether sibilants in <str> shifted towards /s/ (above) or towards /ʃ/ (below) during exposure. Most participants lie to the right of the dashed vertical line, that is, the sibilant in <str> was closer to their own /s/ (than /ʃ/) at baseline. Moreover, most of these participants lie in the top right-hand quadrant, which indicates that sibilants in <str> increased proximity to /s/ during exposure. On the other hand, the two participants for which sibilants in <str> were closer to /ʃ/ at baseline (M19, M16), showed increased proximity to /ʃ/ during exposure. A significant positive correlation between baseline and exposure data (r (14) = 0.58, p < 0.05) broadly supports the idea that baseline production targets influence behavior during exposure to increased /s/-retraction. However, results do not support the specific prediction that individuals who resist /s/-retraction at baseline should show larger shifts in the direction [str] → [ʃtr] than those who already produce /s/-retraction at baseline. On the contrary, Figure 2 suggests that individuals with little evidence of /s/-retraction in <str> at baseline show even less during exposure.

3. PERCEPTION TASK WITH <STR>

3.1 Participants

The participants were the same as in the auditory repetition task (cf. 2.1).

3.2 Materials and procedure

Materials involved /st#r/ and /ʃt#r/ as English phonotactics do not allow word-internal */ʃtr/. First, a native speaker produced *gassed rat* and *gashed rat.* The sibilant was spliced from each and a 25-step [s]…[ʃ] continuum was obtained by interpolating between the two sibilants, as in 2.2. A subset of 11 steps (5, 8 to 16, 19) was selected, concentrating on the perceptual crossover point. These sibilants were spliced back into /gæ_t ɹæt/, giving an 11-step *gassed rat–gashed rat* continuum. In order to minimize the influence on listener responses of any coarticulatory information for /s/ vs. /ʃ/ in the surrounding signal, [gæ] was spliced from *gassed rat* and [t ɹæt] from *gashed rat.* Table 1 lists the mean M1 (defined in 2.3) of the sibilant at each step.

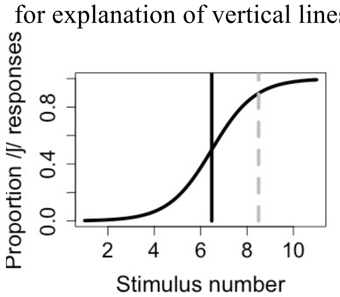
Table 1. Mean M1 (in Hz) for the sibilant in the 11 step *gassed rat–gashed rat* continuum.

1: 8022	7: 5328
2: 6897	8: 5110
3: 6538	9: 4929
4: 6192	10: 4781
5: 5871	11: 4503
6: 5582	

The continuum steps were presented in randomized order to participants who completed the experiment on a laptop while wearing headphones. Their task was to listen to each stimulus and choose whether it corresponded best to *gassed/gashed* by clicking on the corresponding button. The experiment was self-paced. Following [8], the response data were analysed by obtaining 50% *gassed/gashed* category boundaries with logistic regression by fitting a generalized linear model with Response (2 levels 'gassed'/'gashed') as dependent variable, Stimulus Number (11 levels) as a fixed factor and Participant (16 levels) as random factor. This fitted sigmoid curves to each listener, giving 16 category boundary values derived from the listener-specific intercept and slope.

3.3 Results

Figure 3. Proportion /ʃ/ responses for the *gassed rat-gashed rat* continuum (averaged on 16 listeners); see text for explanation of vertical lines.



The solid vertical line in Figure 3 shows the location of the 50% category boundary (averaged over 16 participants) on the 11-step *gassed rat – gashed rat* continuum. The grey dashed line indicates the location of the sibilant in <str>-words that listeners heard during the auditory repetition task, based on its mean M1 of 4943.3 Hz. This sibilant lies between steps 8 and 9 i.e. to the right of the pooled category boundary, which indicates that most participants would have categorized it as /ʃ/.

4. PRODUCTION AND PERCEPTION

Figure 4. By participant difference in distance ratio for sibilants in <str>-words during exposure against perceptual category boundary.

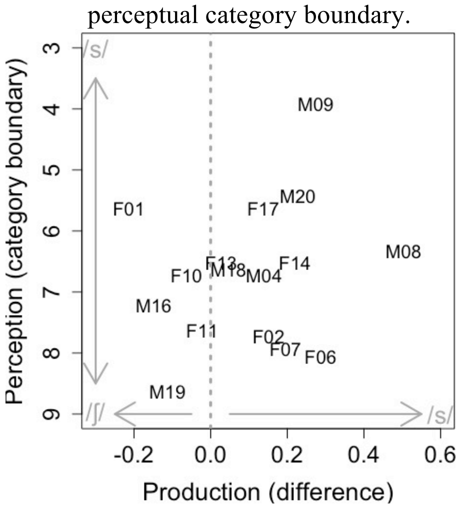


Figure 4 shows perceptual category boundaries against production data from Figure 2. There is no apparent relationship between these two variables. That is, the location of perceptual category boundaries on our *gassed rat–gashed rat* continuum does not predict production behaviour during exposure. Our main prediction in terms of the relationship between perception and production was that individuals who categorized a retracted sibilant in <str> as /ʃ/ would be more likely to shift their production target towards /ʃ/ than those who categorized the same sibilant as /s/ (perhaps filtering out coarticulatory effects). Recall from 3.3 that the sibilant in <str>-words to which participants were exposed during the auditory repetition task lay between steps 8 and 9 on the *gassed rat–gashed rat* continuum. Category boundaries for all but one participant (M19) were lower than step 8 which suggests that they would have categorized the sibilant as /ʃ/. As such, according to our prediction all participants except M19 should have shown a shift towards towards /ʃ/ during exposure (i.e. they should lie to the left of the dashed vertical line in Figure 4). This is evidently not the case: the majority of the 15 participants who categorized the ambiguous stimulus as /ʃ/ showed a shift towards /s/ during exposure.

5. GENERAL DISCUSSION

During exposure to increased /s/-retraction in <str> there was a trend towards more anterior production targets (Fig. 1). This result differs from results from a similar task in which increased coarticulatory nasalization was imitated [20]. Increased /s/-retraction in <str> was sufficient for categorization as /ʃ/ by most participants (Fig. 3), but contra predictions in [2], this did not favour imitation. The primary purpose of this study was to test whether imitation of increased coarticulation depended on individual phonetic repertoires. While there was no evidence from perception (Fig. 4), baseline production targets appeared to influence behaviour during exposure - though not as predicted. Most individuals with anterior sibilants in <str> at baseline showed increased proximity to /s/ during exposure, whereas individuals in which the sibilant in <str> was closer to /ʃ/ (M16, M19) showed a further shift towards /ʃ/. Thus increased /s/-retraction was imitated by those individuals for whom /s/-retraction already formed part of their production repertoire. This does not support predictions in [2] but is in line with evidence from some other imitation studies [7, 12] and, more generally, with the idea that *slight* pronunciation differences can drive sound change (cf. also [5]). It is possible that the perceptual salience of /s/-retraction in <str> (given categorization as a different phoneme) inhibited convergence, cf. [9], or that /s/-retraction in <str> was, for some, socio-indexical and that this influenced their behavior during exposure (e.g. [1] for vowels). Results did not support the idea that the earliest stages of sound change involve imitation of highly coarticulated pronunciations by individuals with little coarticulation in their own speech.

6. ACKNOWLEDGMENTS

Research supported by German Research Council grant N^o 387170477 (2018-21).

6. REFERENCES

[1] Babel, M. 2010. Dialect divergence and convergence in New Zealand English. *Language in Society* 39: 437-456.
[2] 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.
[3] Beddor, P. 2015. The relation between language users' perception and production repertoires. *Proc. 18th ICPhS*. Glasgow.
[4] Fuchs, S., Pape, D., Petrone, C., Perrier, P. 2015. *Individual differences in speech production and perception.* Frankfurt am Main, Peter Lang.
[5] Garrett, A., Johnson, K. 2013. Phonetic bias in sound change. In: Yu, A. (ed), Origins of sound change. Approaches to phonologization, Oxford: Oxford University Press, 51-97.
[6] Kataoka, R. 2011. *Phonetic and cognitive bases of sound change*. PhD Dissertation, UC Berkeley.
[7] Kim, M., Horton, W., Bradlow, A. 2011. Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology* 2(1): 125-156.
[8] Kleber, F., Harrington, J., Reubold, U. 2011. The relationship between the perception and production of coarticulation during a sound change in progress. *Language and Speech* 55(3): 383–405.
[9] MacLeod, B. 2014. Investigating the effects of salience and regional dialect on phonetic convergence in Spanish. In: Côté, M.-H., Mathieu, E. (eds), *Variation within and across Romance Languages: Selected papers from the 41st Linguistic Symposium on Romance Languages*. Amsterdam: John Benjamins, 351-378.
[10] Nielsen, K. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39(2): 132–142.
[11] Ohala, J. J. 1993. Sound change as nature's speech perception experiment. *Speech Communication* 13: 155-161.
[12] Olmstead, A.,Viswanathan, N., Aivar, M., Manuel, S. 2013. Comparison of native and non-native phone imitation by English and Spanish speakers. *Frontiers in Psychology* 4: 1-7.
[13] Romeo, R., Hazan, V., Pettinato, M. 2013. Developmental and gender-related trends of intra-talker variability in consonant production. *Journal of the Acoustical Society of America* 134: 3781-3792.
[14] Rutter, B. 2011. Acoustic analysis of a sound change in progress: The consonant cluster /stɹ/ in English. *JIPA* 41: 27 - 40.
[15] Stevens, M., Harrington, J. 2014. The individual and the actuation of sound change. *Loquens* 1(1). Doi: http://dx.doi.org/10.3989/loquens.2014.003
[16] Stevens, M., Harrington, J. 2016. The phonetic origins of /s/-retraction: Acoustic and perceptual evidence from Australian English. *Journal of Phonetics* 58: 118-134.
[17] Stevens, M., Harrington, J., 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), 8. Doi: http://doi.org/10.5334/gjgl.620.
[18] Weinreich, U., Labov, W., Herzog, M. 1968. Empirical foundations for a theory of language change. In: Lehmann, W., Malkiel, Y. (eds), *Directions for historical linguistics*. Austin: University of Texas Press, 95-198.
[19] Zellou, G. 2017. Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics* 61: 13-29.
[20] Zellou, G., Scarborough, R., Nielsen, K. 2016. Phonetic imitation of coarticulatory vowel nasalization. *Journal of the Acoustical Society of America* 140: 3560-3575.