RESEARCH

Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English

Mary Stevens, Jonathan Harrington and Florian Schiel
Institute of Phonetics and Speech Processing, Ludwig-Maximilians-University, Schellingstraße 3, Munich, DE
Corresponding author: Mary Stevens (mes@phonetik.uni-muenchen.de)

The study explored whether an asymmetric phonetic overlap between speech sounds could be turned into sound change through propagation around a community of speakers. The focus was on the change of /s/ to /ʃ/ which is known to be more likely than a change in the other direction both synchronically and diachronically. An agent-based model was used to test the prediction that communication between agents would advance /s/-retraction in /str/ clusters (e.g. *string*). There was one agent per speaker and the probabilistic mapping between words, phonological classes, and speech signals could be updated during communication depending on whether an agent listener absorbed an incoming speech signal from an agent talker into memory. Following interaction, sibilants in /str/ clusters were less likely to share a phonological class with prevocalic /s/ and were acoustically closer to /ʃ/. The findings lend support to the idea that sound change is the outcome of a fortuitous combination of the relative size and orientation of phonetic distributions, their association to phonological classes, and how these types of information vary between speakers that happen to interact with each other.

1 Introduction

Phonetic variation is very often directional such that the likelihood of a shift in a phonetic distribution *Y* towards another distribution *X* is usually not matched by an equally probable change in *X* towards *Y* (Garrett & Johnson 2013). High back vowels are, for example, typically more likely to front than front vowels are to retract, possibly because the tongue positions of the former are much further from the centre of the vowel space (Harrington et al. 2011). Domain-final voiced stops in e.g. *tab* are synchronically shifted towards their voiceless counterparts with a greater probability than the other way round (Jongman et al. 1992; Pierrehumbert 2001). Before high front vowels, velar stops are far more likely to be misperceived as coronals than in the other direction (Winitz et al. 1972). These and many other types of synchronic asymmetries can be associated with historical sound changes. Thus the diachronic fronting of back vowels is attested in numerous languages and forms part of one of the general principles of chain shift in Labov (1994). The neutralisation of the voicing contrast in domain-final stops towards the voiceless counterpart (e.g. Port & O'Dell 1985; Warner et al. 2004) and velar palatalisation (Guion 1998; Chang et al. 2001) are also sound changes found in many languages that have their origin in the synchronic asymmetries mentioned above.

A far less well understood issue is how stability and change are related: that is, how stable conditions in which these types of phonetic asymmetry are maintained synchronically

are related to unstable conditions in which there is a directional sound change. The main aim in the present study is to develop a model of the relationship between stability and change in terms of phonetic variation that is modified when members of a community of speakers talk to each other. The starting point of the model is that speakers are phonetically idiosyncratic even if they are of the same community and spoken accent; and that occasionally diachronic change can emerge if the idiosyncrasies of multiple speakers happen to fall along the path of a shared phonetic asymmetry (Harrington & Schiel 2017). At the core of the model is the idea that sound change can emerge out of communication density (Bloomfield 1933; Labov 2001; Trudgill 2008) – that is, which speakers happen to talk with each other most frequently – combined with directional biases in phonetic variation as sketched above. Thus the model includes individual differences in the initial state, because these exist in real communities of speakers, but it does not dictate whether these individual differences necessarily lead to historical sound change.

Agent-based modelling is a computational technique that can be used to address the question of how phonetic asymmetries and communication density in combination drive sound change. Agent-based modelling is more generally relevant to explaining how community change and interaction between individuals that make up the community are related (Castellano et al. 2009). In research on sound change, agent-based models are often constructed such that there is a stochastic relationship between the lexicon, phonology and speech signals that can be updated in speech perception. Such models have been used to address different aspects of sound change. The concern in Ettlinger (2007) was to show how vowel chain shifting emerges as a consequence of stored and updated exemplars. Kirby's (2014) computational model was used to argue for functional considerations in ongoing sound changes affecting plosive contrasts in the Phnom Penh variety of Khmer. Blevins & Wedel's (2009) computational simulation showed how sound change typically fails to create homophones in minimal pairs that cannot be reliably disambiguated by pragmatic information. Stanford & Kenny (2013) used their agent-based model to test various aspects of Labov's (2007) theory of sound change by transmission (to the next generation of speakers), incrementation (brought about when children increment sound change from one generation to the next), and diffusion (brought about principally by contact between adults). Harrington & Schiel (2017) used an agent-based model to test whether the outcome of the interaction between two groups of speakers of Standard Southern British with different pronunciations of /u/ was influenced by how the two pronunciations were oriented with respect to each other in an acoustic space. They showed that, because older speakers' retracted [u] was oriented towards that of younger speakers' fronted [ʉ] (but not the other way round), the influence was correspondingly asymmetric with a large shift following interaction in the older speakers' retracted variants towards the front of the vowel space.

The present study uses the ABM architecture developed by Harrington and Schiel (2017) and extends their research in two main ways. First, the sound change to be examined involves variants of separate phonemes rather than, as in Harrington & Schiel (2017), two variants of the same phoneme. Second, the sound change to be examined takes place in the language spoken by one single group of speakers with a relatively homogeneous accent. The importance of extending this line of research to a homogeneous group is because sound change evidently does take place not only in situations of dialect or language contact but also in isolated communities – but possibly over longer time-scales (Trudgill 2011: 103).
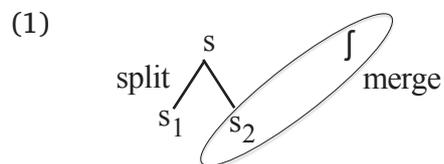
The sound change to be analysed in the present study is /s/-retraction in clusters containing a rhotic (e.g. *string*) in English (see also Mielke et al. in press on /s/-retraction). The retraction of sibilants is of particular interest in this study because of the evidence

Stevens et al: Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English

Art. 8, page 3 of 30

that it is asymmetric. For example, across word-boundaries, /s/→/ʃ/ assimilation is more likely than in the other direction (e.g. Recasens & Mira 2013). There is also evidence of spectral central of gravity lowering of /s/ in clusters even when there is no rhotic in words like *steam* (Baker et al. 2011; Stevens et al. 2015; Stevens & Harrington 2016). Compatibly, there are many /s/→/ʃ/ sound changes in consonant clusters (also without rhotics) that have taken place in various languages (Rohlfs 1966; Kümmel 2007: 232), but little evidence of changes in the other direction. Specific to /s/-retraction in /str/ clusters, there is evidence that this sound change has taken place in many varieties of American English (Labov 2001; Kraljic et al. 2008; Rutter 2011), in which the sibilant in /str/ has changed from /s/ to /ʃ/. A tendency towards /s/-retraction in /str/ clusters has also been documented for speakers of Australian English (Stevens & Harrington 2016), some British English varieties (Cruttenden 2014) and New Zealand English (Warren 1996; Lawrence 2000). The analysis of Australian English in Stevens & Harrington (2016) showed spectral centre of gravity lowering in /str/ for all speaker participants. Such lowering of the spectral centre of gravity was also found to be perceptible: listeners often perceived /ʃ/ (*sheet*) rather than /s/ (*seat*) when the sibilant was spliced from /str/ clusters (*stream*) and pre-prepended to /i:t/. These results were interpreted as evidence of a synchronic phonetic asymmetry in Australian English (in contrast with other varieties of English in which there is evidence that sound change involving /s/-retraction in /str/ clusters has already progressed to completion).

The present study models the development of /s/-retraction in /str/ in Australian English. To do so, it uses the ABM architecture developed by Harrington and Schiel (2017), which is an implementation of the interactive-phonetic (IP) model of sound change (Harrington et al. 2018). The IP model is defined in Harrington et al. (2018) but for the purposes of the present study it is important to outline the minimal set of assumptions about human speech processing and its relationship to the development of sound change from phonetic variation on which the IP model is based. In the IP model, words and phonological classes are statistical models calculated over parameterised stored speech signals. This part of the IP model is related to aspects of exemplar theory in which each word is associated with a phonetic distribution that depends on previously encountered and remembered experiences of that word (Pierrehumbert 2001; 2016) and in which a phonological unit is a statistical distribution over the union of the relevant sub-parts of remembered exemplars of words (Pierrehumbert 2003). In common with a modular feedforward architecture (Bermúdez-Ortero 2012; 2013), the phonological units in both the IP model and in exemplar theory are abstract coding elements that are smaller than the number of words in the lexicon. However, in contrast to the modular feedforward system, phonological units in the IP-model and in exemplar theory emerge from dense regions of intersection of remembered word trajectories in the parametric space – for example, in a region of dense intersection of high F2 and low F1 trajectories for /i/ (Pierrehumbert 2003).

To apply the IP model to the development of /s/-retraction in /str/ in Australian English, it was necessary to make one major structural modification. In Harrington & Schiel (2017), a listener's phonological category could only be updated by perceived samples that were probabilistically closer to that category than to any other. For example, an agent-listener was prevented from updating /u/ in a perceived *food* if the perceived vowel was probabilistically closer to the listener's /i/. This type of filtering is necessary to prevent a collapse of phonemes in a phonetic space; it also models the idea that sound changes are, if not prevented then at least inhibited by phonemic contrasts. However, this type of top-down phonological filtering on the absorption of incoming signals into memory would also block a phonological sound change such as /s/ → /ʃ/ or indeed any kind of merger: this

is because /s/ could never be updated by samples that were probabilistically closer to /ʃ/ than to /s/. The proposed way around this impasse, presented in this study, is to allow phonological categories to split and to merge. A phonological change such as /s/ → /ʃ/ can then be recast as (1), in which /s/ splits into different sub-phonemic categories of which one merges with /ʃ/:

(1)

Thus, in the revised version of the IP model phonological units (which are referred to in the remainder of the study as *phonological categories* or *phonological classes*) are sub-phonemic.

In this respect the revised version of the IP model resembles some versions of exemplar theory (e.g. Pierrehumbert 2003) in which the phonological units are also sub-phonemic and in which there are connections between sub-phonemic units and lexical items. This idea is also consistent with the view that non-contrastive phonetic variants can form part of the lexicon (Kiparsky 2015; 2016) as well as with studies of perceptual learning (Reinisch et al. 2014; Reinisch & Mitterer 2016), second language acquisition (Polka 1991), and new dialect acquisition (German, Carlson, & Pierrehumbert 2013) in which listeners have been shown to adapt to (or learn from) units that are less abstract than the classical phoneme. More generally, the existence of a sub-phonemic level of processing in the IP model is compatible with the idea that there are units of speech processing in language that are marginal i.e. that are neither phonemic nor allophonic (Ladd 2006; Scobbie & Stuart-Smith 2008; Hall & Hall 2016) and that such finely differentiated phonetic variants form an integral part of lexical access (as shown e.g. by studies of incomplete neutralisation as in Ernestus 2006; Ernestus & Baayen 2006). In the revised version of the IP model, a sound change such as the historical fortition of post-vocalic obstruents in German resulting in neutralisations of word-pairs like *Rad/Rat* (Jespersen 1913; Fourakis & Iverson 1984; Wiese 1996) does not operate all at once, but rather via numerous sub-phonemic classes that are intermediate between a full voiced /d/ and a voiceless /t/. Thus splits and mergers in the IP model occur between sub-phonemic units that are phonetically very close and possibly even (as in the case of incomplete neutralisation: Kleber et al. 2010; Roettger et al. 2014) scarcely distinguishable perceptually. For example, the /u/ vowels in the English words *fool, fooling,* and *hula,* which differ minimally and yet consistently for some speakers depending on the presence or absence of a morpheme boundary (Strycharczuk & Scobbie 2016) could be examples of the different types of sub-phonemic units that are conceived of in the IP-model as candidates for splitting and merging.

In order to illustrate how splitting and merging operate in the revised version of the IP model, we ran a very simple simulation of speaker-listener interactions based on artificial data. Figure 1a shows one listener's distribution of stored samples of two separate phonological categories, *X* (grey) and *Y* (black) on an acoustic parameter *P*. Note that the listener's categories are asymmetrical: the probability density distribution based on the listener's stored samples of *Y* (black) is more skewed towards *X* (grey) than the other way around (i.e. the left-hand slope of *Y* is flatter than the right-hand slope of *X*). The listener's phonological categories *X* and *Y* could be modified through simulated interaction with other language users based on two rules:

Stevens et al: Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English

Art. 8, page 5 of 30

R1. A perceived sample is only absorbed into a phonological category if it is probabilistically closer to that category than to any other (as in Harrington & Schiel 2017).

R2. Each time a perceived sample is absorbed into a phonological category, a test is made of whether:

(a) the category's distribution is more likely to be bimodal than unimodal (based on a test for bimodality in Hartigan 1985). If so, the category is *split* (based on *k*-means clustering: Hartigan & Wong 1979) into two new categories.

(b) any pair of existing categories is more likely to form a unimodal than two separate distributions. If so, the pair of categories is *merged*, which means that samples in these two categories are pooled together.

At the first stage of the simulated interaction, a speaker produced 100 samples of category X, s_X, and 100 samples of category Y, s_Y. Both of these sets of samples were randomly drawn from a normal distribution with a mean at -0.5 (at the arrow shown in Figure 1a); thus s_X and s_Y are indistinguishable and have a mean somewhere between the listener's X and Y distributions. This simulates what happens when a listener perceives samples of X and Y from a speaker who has neutralised the difference between these two categories. The outcome therefore depends on the classification of these samples by the listener into X and Y. The posteriori probability is on average higher for Y than it is for X, due to the skewed shape of the distributions, as outlined above. (The assumption in the simulation is that the listener nevertheless correctly identifies the categories X and Y from the speaker's
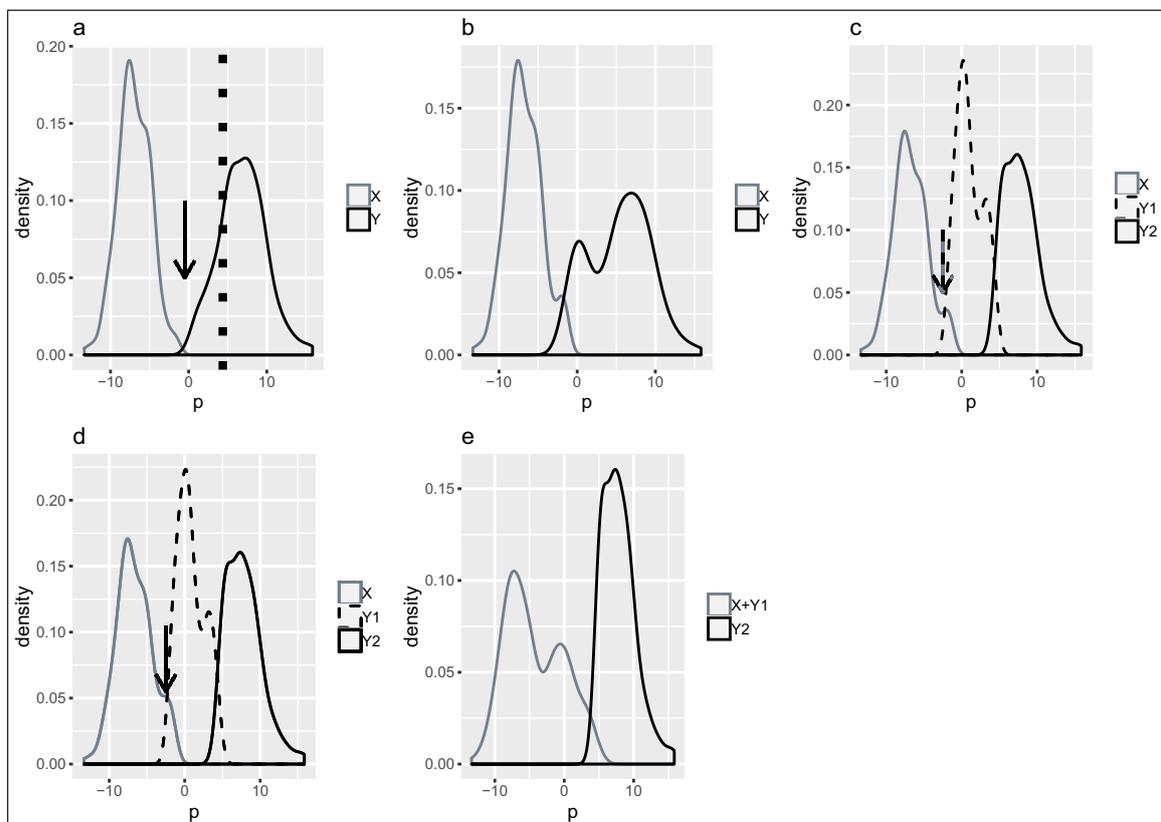

Figure 1: Simulated probability density distributions of a listener's phonological categories X (grey) and Y (black) on an acoustic parameter P, before (a) and during (b–e) interaction with other language users. Points to the left of the dotted line in (a) that originally formed part of Y are in the new merged category *X+Y1* in (e).

productions due to semantic/pragmatic context). After this first stage of the simulation, the listener's distribution of X was scarcely updated whereas most s_Y were absorbed into the listener's Y distribution, causing it to become bimodal (Figure 1b). The reason for this difference is because of a combination of the asymmetry between X and Y and rule R1: most of s_X were probabilistically closer to Y and so were not absorbed into X; by contrast, most of s_Y were absorbed into Y because they were probabilistically closer to Y than to X. After the uptake of 60 s_Y, the listener's Y-distribution became sufficiently bimodal so that rule R2(a) applied causing Y to split into two new phonological categories Y_1 and Y_2 (Figure 1c). Following this split, the simulation was continued, this time with a speaker who produced samples s_X and s_{Y1} that were once again randomly drawn from a normal distribution centred at the vertical arrow (p = −3.5) in Figure 1c, approximately halfway between the distribution of X and the new category Y_1. Because the listener's X was now more skewed towards Y_1 than in the other direction (Figure 1c), R1 applied to allow most of s_X to be incorporated into X but blocked almost all of s_{Y1} from being absorbed into Y_1 (Figure 1d). In other words, this preferential absorption into X happened because s_{Y1} samples were probabilistically closer to X than to Y_1. The simulation was continued and after the uptake of just 16 s_X, X and Y_1 were merged into a single category (Figure 1e) because their combined distributions were found to be unimodal by rule R2(b).

The simulation shows that if two of a listener's phonological categories X and Y are in close proximity and Y extends into the X-space more than the other way round, then X acts as an attractor that can pull Y apart (Figure 1c) and may then incorporate some of the sub-phonemic categories into which Y has been split (see also Pierrehumbert 2001). Thus an asymmetrical relationship between phonological categories combined with interaction is hypothesised to be the cause of both phonetic change and phonological re-categorisation. It is important to point out that an ambiguous input in this model is sufficient to cause a categorical change to a listener's phonology: that is, change is not dependent on input from speakers whose X and Y have already undergone re-categorization. Notice that all the original samples of Y to the left of the vertical dotted line in Figure 1a are at exactly the same value on the parameter P in Figure 1e but have been assigned to a different phonological category. This aspect of the model is consistent with the idea that listener uncertainty about phonological categorisation can precede any changes to speech production in the early stages of sound change (see in particular Beddor 2009; also Ohala 1993).

The present study is based on the idea that Australian English /ʃ/ and /s/ are analogous to X and Y in Figure 1, respectively, whereby samples of /s/ are more likely to extend in the direction of /ʃ/ than the other way around (because of the tendency towards /s/-retraction documented in Stevens & Harrington 2016). The overall prediction to be tested is that interaction between speakers should be more likely to cause /s/ to be phonologically reclassified and to shift acoustically in the direction of /ʃ/ than vice versa. Section 2 presents an overview of the data and section 3 describes the agent-based model that was implemented to test the main prediction of this study (the code to run the agent-based model is available at ftp://ftp.bas.uni-muenchen.de/pub/BAS/ABM/). Section 4 describes the methods that were used to analyse phonological reclassification and acoustic change. Section 5 presents the results of agent-based modelling from the perspectives of phonological reclassification (5.2), acoustic change (5.3) and bidirectional effects between the two (5.4).

2 Data

2.1 Speakers, materials, acoustic speech signals

The study was based on speech produced by 19 of the 20 Australian English speakers (6 male, 13 female; age range 29–49 years) that had been analysed for /s/-retraction in Stevens & Harrington (2016) and that were available for the present study. The speakers reported no speech or hearing problems and were all from the same rural heritage town of

Stevens et al: Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English

Art. 8, page 7 of 30

Braidwood (population of 1000) located about 300 km southwest of Sydney. The speakers all knew each other, with some talking to each other daily.

In Stevens & Harrington (2016), the speakers had produced a range of English words containing sibilants in pre-vocalic and various pre-consonantal positions, including five of the monosyllables shown in Table 1. These five monosyllables were selected for the present study because they contained sibilants in pre-vocalic position (*seem, sane, sheep, Shane*) and in /str/ clusters (*stream*). We then expanded the data set, increasing the number of words containing sibilants in these particular contexts. This was necessary because we wanted the initial state of the agent-based model to represent a more realistic snapshot of the synchronic variation that exists for sibilants in /str/ clusters in the language spoken by this particular community and also for reasons to do with splitting and merging of phonological classes (see 3.1). To this end, the same speakers took part in an additional recording session in which they produced an additional four monosyllables (*soak, strut, show, strong*) and all the polysyllabic words in Table 1. The polysyllabic words varied between 2 and 4 syllables and in their rhythmic pattern depending on whether the target sibilant was preceded or followed by a lexically stressed or unstressed syllable.

The procedure for these additional recordings was the same as in Stevens & Harrington (2016) and was as follows. The data were recorded in the same private home using a headset microphone and a MacBook Pro installed with the SpeechRecorder software (Draxler & Jänsch 2004). In both recording sessions, words were presented to participants in randomized order on a laptop computer screen with ten repetitions for each word. Whereas participants read words aloud in the carrier phrase "Any_____" for Stevens & Harrington (2016), for these additional recordings participants read words aloud in isolation (the carrier phrase was deemed unnecessary because target sibilants were primarily in postvocalic position). The experiment was self-timed: after participants produced a word, they (or the experimenter) clicked the mouse button so that the next word appeared. The materials available for analysis in the present study comprised 19 (speakers) × 41 (words) × 10 (repetitions) = 7790 sibilants. Some speakers had only produced 9 repetitions and one speaker only 8 repetitions of some words so that the actual number of sibilants analysed in the present study was 7779.

Table 1: The target sibilants according to word type (first three columns) and lexical stress pattern. The final column shows the number of syllables. (s) and (u) denote whether the vowel following the target syllable was in a lexically stressed (in bold) or unstressed syllable.

\<s\> words	\<str\> words	\<ʃ\> words	n
weak_strong (s)			
Minnes**o**ta	astr**o**nomy, catastr**o**phic	oversh**a**dowed	4
ass**e**mbly	pastr**a**mi, astr**i**ngent	dish**e**veled	3
ass**au**lt	destr**oy**, restr**i**ct	mach**i**ne	2
strong_weak (u)			
f**a**scinating	cat**a**strophe, ped**e**strian	inform**a**tion	4
p**o**ssible	**a**stronaut, **oe**strogen	p**a**ssionate	3
m**e**ssy	g**a**stro, d**i**strict	t**i**ssue	2
weak_weak (u)			
mot**o**rcycle	adm**i**nistrate, claustroph**o**bic	p**e**rishable	4
p**o**licy	ch**e**mistry, **o**rchestra	p**o**lishing	3
monosyllable (s)			
s**a**me, s**ee**n, s**oa**k	str**ea**m, str**o**ng, str**u**t	Sh**a**ne, sh**ee**p, sh**ow**	1

2.2 Acoustic parameters

The recordings were down-sampled to 32000 Hz and labelled semi-automatically with the Munich Automatic Segmentation System for Australian English (Kisler et al. 2017). The acoustic boundaries of the 7779 sibilants at the acoustic onset and offset of frication were manually checked and hand-corrected where necessary. A power spectrum was calculated for each sibilant using a DFT with a 40 Hz frequency resolution, a 5 ms Blackman window, and a 5 ms frame shift. The first spectral moment (M_1) was then calculated on each of the power spectra between 500 Hz and 15000 Hz using the emuR package (Winkelmann et al. 2017), resulting in one M_1 trajectory for each sibilant. M_1 provides an effective acoustic parameter for distinguishing between [s] and [ʃ] (Forrest et al. 1988), which comes about because of their different front cavity lengths (Heinz & Stevens 1961; Stevens 1998; Jongman et al. 2000; Shadle 2012). The M_1 trajectories were z-score transformed using the speaker-normalization method in Lobanov (1971).

A discrete cosine transformation (DCT) was then applied to each speaker-normalized M_1 trajectory. The reason for representing M_1 trajectories with DCT coefficients is because of the evidence that sibilants in different contexts vary not just at the target but also dynamically (Haley et al. 2010; Iskarous, Shadle, & Proctor 2011; Koenig et al. 2013). The first three DCT coefficients are proportional respectively to the mean, linear slope, and curvature of the trajectory and thereby encode dynamic information between the acoustic onset and offset of each (speaker-normalised) M_1 trajectory (Watson & Harrington 1999). In addition, DCT coefficients can very effectively distinguish between alveolar and post-alveolar fricatives. For visual inspection of the acoustic results (section 5.3), reconstructed M_1 trajectories were then obtained by applying an inverse DCT (Watson & Harrington 1999).

3 Architecture of the agent-based model

The agent-based model that was implemented in this study was based on Harrington & Schiel (2017). As with any study using agent-based modelling (see Section 1 for other studies), the outcome in this study will depend on certain necessary decisions in terms of the structure of the model and any parameter settings. It is necessary to keep this in mind, as different choices by modellers can lead to different outcomes. Here and elsewhere, we attempt to be as explicit as possible about our assumptions and the decisions that were necessary in adapting Harrington & Schiel's (2017) agent-based model to /s/-retraction, and readers should note that both the input data and the code that we used to run our agent-based model are available at ftp://ftp.bas.uni-muenchen.de/pub/BAS/ABM/.

In Harrington & Schiel's (2017) agent-based model agents represent real speakers (in the present study, one for each of the 19 speakers described in 2.1). Each agent was equipped with four types of information: (i) the 41 word classes shown in Table 1 (ii) up to 410 signals consisting of triplets of DCT-coefficients (one triplet per sibilant) (iii) a phonology consisting of phonological classes (initially two, but potentially variable in number) linking the word classes and signals and (iv) a time-stamp that was randomly assigned to each repetition of each word.

Thus each agent's memory was initialised with different DCT-triplets (because each agent's DCT-triplets were calculated on signals spoken by a different speaker) and the same word classes (those shown in Table 1). Each word class was further defined as a Gaussian statistical generalisation over the signals (DCT-triplets) with which it was associated. Thus, since there was a triplet of DCT-coefficients per repetition and typically 10 repetitions per word class, then each word class was represented as a 3 dimensional ellipsoid whose mean and covariance matrix were derived from the cloud of (up to) 10 points with which they were associated. Each phonological class was also defined as a

Stevens et al: Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English

Art. 8, page 9 of 30

Gaussian statistical generalisation over the union of the words with which the phonological class was associated. For example, if there was a phonological class that included the sibilants in the words *stream, strong,* and *strut* only, then the mean and covariance matrix for that phonological class were derived from (typically) $3 \times 10 = 30$ points in a three-dimensional DCT space.

3.1 Splitting and merging of phonological classes

As outlined in Section 1, in the revised version of the IP model agents could split and merge their phonological classes. A split or merge was always binary i.e. a single phonological class could be split into two separate phonological classes; or two phonological classes could be merged into a single class. The principles by which the splitting algorithm operated are illustrated in Figure 2 in a two-dimensional $C_0 \times C_2$ space for 27 data points. The 27 data points in Figure 2 show three repetitions each of nine words (*assault, destroy, district, gastro, messy, restrict, soak, strong, strut*) that initially belong to the phonological class *P* (solid ellipse). The first stage of splitting was to apply the technique of *k*-means clustering (Hartigan & Wong 1979) which breaks up the data points into two maximally distinct classes, P_1 and P_2. The second stage was to readjust class membership to disallow any word from being a member of two separate phonological classes. This was done because, as outlined in the earlier simulation and (1), phonological classes repel each other in the
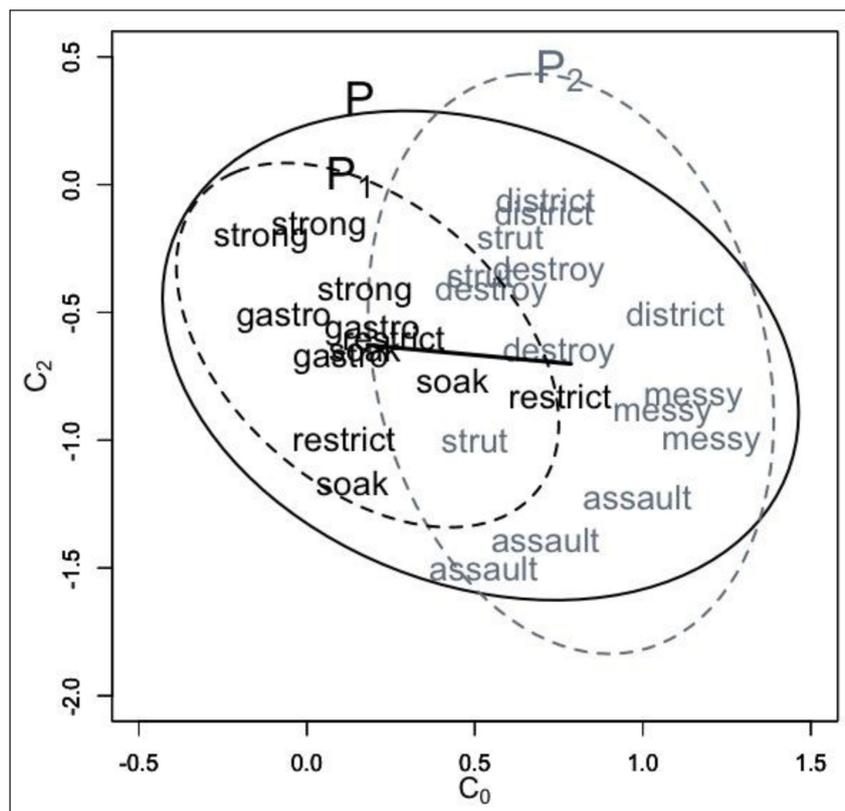

Figure 2: The distribution together with 95% confidence ellipses of 27 values from 3 repetitions × 9 words for one agent in the parameter space $C_0 \times C_2$ showing their membership to an original class *P* (solid black) and their new class membership to P_1 (dashed black) and to P_2 (dashed grey) after the application of *k*-means clustering to split *P* into two classes. The *strut* token at the intersection of the P_1 and P_2 classes (at coordinates 0.5, –1) was initially assigned to P_1 by *k*-means clustering but was subsequently reassigned to P_2 containing the other two *strut* tokens. The solid line shows the Euclidean distance between the class centroids of P_1 and P_2.

ABM. They repel each other because an agent-listener does not incorporate a perceived signal into a phonological class if it is probabilistically closer to another phonological class. Therefore, if a word were allowed to be a member of two phonological classes, then the memorized repetitions of the same word would repel each other. Moreover, lexical items would no longer be distinguishable using a small number of coding elements (Pierrehumbert 2016) with the undesired consequence that phonology would not perform its main function of facilitating the mapping between the acoustic speech signal and lexicon. Class membership was readjusted as follows: if the data points for a given word were distributed between the two classes following application of the k-means clustering algorithm, then a reassignment was made based on whichever class contained the majority of the word's data points. For example, the *strut* token at the intersection of the P_1 and P_2 ellipses in Figure 2 was initially classified as P_1 by the k-means algorithm but was then re-assigned together with a majority (2/3) of the other *strut* tokens to P_2. In the third stage it was necessary to decide whether the 27 data points were better modelled by one or two Gaussian distributions (i.e. whether the phonological class P should be split or not). To do so, a pair of logarithmic posterior probabilities was calculated (applying the standard multivariate normal distribution density function, e.g. https://en.wikipedia. org/wiki/Multivariate_normal_distribution) for each of the 27 data points. The first of these log. posterior probabilities is the probability that the data point is a member of the original class P; the second is the probability that it is a member of whichever new class (P_1 or P_2) to which it had been assigned after k-means clustering and readjustment as described above. The fourth stage was to aggregate these log. probability pairs by word: in the example in Figure 2, this results in 9 aggregated log. probability pairs, one pair for each of the 9 words. The fifth stage was to apply a paired-sample t-test at an alpha level of $p = 0.05$ to test whether the aggregated log. probabilities were significantly greater for the split classes (P_1 or P_2) than for the original class P. If so, then the class was split; if not, no splitting occurred.[1] The splitting algorithm was applied separately to each agent's phonological classes and, if a split occurred, it always resulted in a pair of classes. The output of applying the splitting algorithm to an agent that has n phonological classes was therefore maximally $2n$ new classes. Finally, because a phonological class was defined as a generalisation across words in the lexicon, no split was allowed if the output of k-means clustering resulted in a phonological class with only one word.

The merge algorithm worked on the principle that if there was no significant difference between the by-word aggregated probabilities of class membership to (i) two classes (e.g. to P_1 and P_2 in Figure 2) and (ii) the union of the two classes (to P in Figure 2), then the two classes were merged. The merge algorithm was applied iteratively to all possible paired combinations of a given agent's phonological classes. Thus, if an agent has 4 phonological classes P, Q, R, S, then there are $4!/2!2! = 6$ possible combinations of phonological class pairs to be considered for merging. Here the ordering of the pairs matters because, if a pair of classes (P, Q) is merged, then there is no possibility of subsequently merging the pair (P, R). A principle therefore had to be established for ordering the class pairs. This was based on the inter-Euclidean distance between the centroids of the two classes. In Figure 2, the inter-Euclidean distance between P_1 and P_2 is the length

---

[1] It is not known how biological systems decide whether a data distribution is uni- or bi-modal. For this reason, we have used a statistical test for this purpose. The alpha level influences the likelihood of a split. The paired-sample t-test that we have used aggregates across word repetitions (so that the comparison is between a pair of values per word class). We will follow a Reviewer's suggestion and explore the alternative approach in future versions of the ABM of using a likelihood ratio test based on a mixed model with Word as a random factor. Pilot tests suggest so far that this alternative approach makes splitting and merging somewhat less likely than with the aggregated paired-sample t-test we have used here.

Stevens et al: Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English

Art. 8, page 11 of 30

of the straight line shown in the two dimensional $C_0 \times C_2$ DCT space. Just such distances were calculated between all class pairs to be considered for merging, except that this was done in a three dimensional $C_0 \times C_1 \times C_2$ DCT space. The merge algorithm was then applied to class pairs ordered from lowest to highest Euclidean distance between the class centroids (because class pairs with a low value on this parameter typically show more overlap and are hence more likely to merge). If, for this example, the ordering from lowest to highest pairs for the 4 phonological classes P, Q, R, S based on this Euclidean distance is $(Q, R), (P, Q), (Q, S), (R, S), (P, R), (P, S)$, then the merge algorithm applies first to the pair (Q, R) and, if no merge is made, subsequently to (P, Q) and so on through all 6 class pairs. The iterative aspect of the algorithm is that these merging principles are re-applied if a merge was made. Thus suppose for this example that (Q, R) does not merge but then (P, Q) merges resulting in three phonological classes $P + Q, R, S$. Following this $P + Q$ merge, the algorithm starts over again working through the pair-wise combinations of the three classes $(P + Q, R), (P + Q, S), (R, S)$, once again in the order from smallest to highest Euclidean distance between the class centroids. Although this very rarely happened for any data in the present study, the merging algorithm could therefore merge n separate phonological classes into one class (merge e.g. P, Q, R, S into $P + Q + R + S$) by this iterative procedure.

3.2 Agent-internal phonological reclassification

As outlined in 3.1, agents could split and merge their phonological classes. The likelihood of phonological re-classification taking place depends broadly on two separate forces. The first is agent-internal: a split or merge is likely to apply if an agent has, respectively, a bimodal phonological class or two phonological classes that are very close to each other in an acoustic space. The second is agent-external: interaction between agents might cause an agent's phonological classes to shift closer together or to drift further apart in an acoustic space and then to merge or to split.

In order to separate these agent-internal from agent-external forces that act on phonological re-classification, the split and merge algorithms were repeatedly applied to the phonological classes of each agent separately *prior* to any interaction between agents until there was no further change. To establish these *baseline* phonological classes, all agents started with the same two phonological classes: the sibilants in all $<s>$ and $<str>$ words (see Table 1) were in the same phonological class that will be denoted by /s + str/; and the sibilants in $<ʃ>$ words were in a separate phonological class /ʃ/. (We use henceforth $<\ >$ when referring to the sibilants that are members of the three groups in Table 1; // will be used to denote phonological classes in the agent based model). There could never be any acoustic changes during this agent-internal phonological reclassification: the only possible changes were in how a sibilant was phonologically classified.

Two types of phonological classes could be formed as a result of this agent-internal phonological reclassification. Firstly, new phonological classes could be formed within any of the $<s>$ words, within any $<str>$ words, or within any $<ʃ>$ words. For example, the sibilants of the $<s>$ words in *motorcycle* and *policy* might be in one class /s/$_1$ and other (and not necessarily all) $<s>$ word sibilants might be in another class /s/$_2$. The notation /s/$_1$, /s/$_2$ ... /s/$_n$ will henceforth be used to denote separate phonological classes that contain different (non-intersecting) groupings of $<s>$ words only. (Analogously, /str/$_1$, /str/$_2$ ... /str/$_n$ denote sibilants of $<str>$ words only that are in different phonological class subsets; and /ʃ/$_1$, /ʃ/$_2$ ... /ʃ/$_n$ are phonological class subsets of $<ʃ>$ words only). Secondly, phonological classes could be formed across the three ($<s>$, $<str>$, $<ʃ>$) word types. For example, the phonological classes P_1 and P_2 in Figure 2 are /s + str/$_1$ and /s + str/$_2$ respectively. This is because P_1 includes sibilants from at least one $<s>$ word

(*soak*) and from at least one <str> word (*gastro, strong, restrict*). P_2 is /s+str/$_2$ because it includes different words from those in P_1 and also contains at least one <s> word (*assault, messy*) and at least one <str> word (*destroy, district, strut*). The other logically possible types of phonological classes across word types are: /s+ʃ/$_n$ (a phonological class that includes sibilants from at least one <s> word and at least one <ʃ> word), /ʃ+str/$_n$ (a phonological class that includes sibilants from at least one <ʃ> word and at least one <str> word) and /s+str+ʃ/$_n$ (a phonological class that includes sibilants from at least one <s> word, and at least one <str> word, and at least one <ʃ> word). A hypothetical example of agent-internal phonological reclassification is shown in Figure 3.

The phonological classes that result from agent-internal phonological reclassification, exemplified for one hypothetical agent in Figure 3, differ between agents. This is because each agent was initiated with different DCT-triplets (derived from one speaker's production data); thus agents differ in terms of the degree of overlap between phonological classes and their orientation in the DCT-space, and also in terms of the likelihood of application of the split and merge algorithms described in 3.1.

3.3 Interaction between agents

Each interaction was always between a pair of agents randomly chosen from the available 19 agents. One of the agents was the agent talker the other the agent listener. The agent talker randomly selected a word class from the lexicon of 41 words and sampled a three-dimensional Gaussian trained on their memorized samples for that word to generate a signal consisting of a DCT-triplet. The selected word class and derived signal were transmitted to the agent listener. No phonological class label was transmitted because,
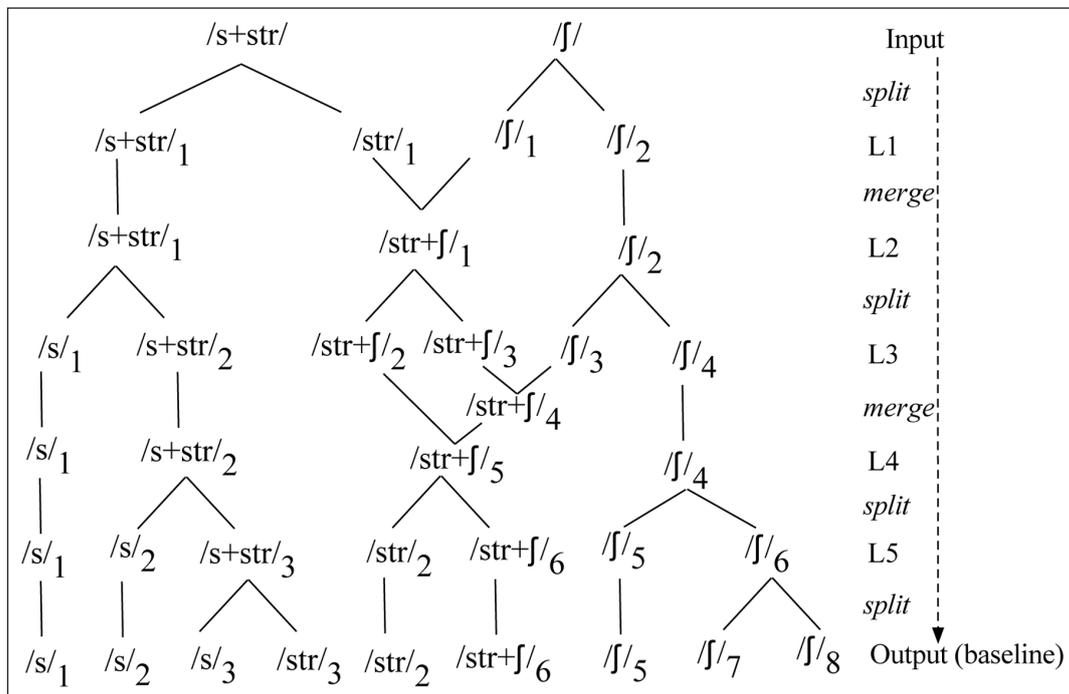

Figure 3: An example of agent-internal phonological reclassification due to the repeated application of the split and merge algorithms. The input consists of the two phonological classes /s+str/ (that includes sibilants in all <s> and <str> words) and /ʃ/ (that includes sibilants in all <ʃ> words). The output phonological classes or *baseline*, derived in this hypothetical example via 5 intermediary levels (L1–L5), are the phonological classes that are input to any interaction between agents. The derivation of the phonological classes at L4 shows an example of iterative merging in which the merged class /str+ʃ/$_4$ is once again merged with /str+ʃ/$_2$ to derive /str+ʃ/$_5$.

Stevens et al: Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English

Art. 8, page 13 of 30

for reasons to do with the agent-specific application of the split and merge algorithms (3.1 and 3.2 above), phonological systems could (and generally did) differ between the agent talker and the agent listener. Because the word class is transmitted with the signal to the listener, there is never any possibility of word misperception in this model. More specifically, if the agent talker transmits *stream* then this is the word that is unambiguously perceived by the agent listener. Of course, lexical confusion does occur in speech perception, but since such confusion is not considered to have any major role in explaining the nature and direction of sound change, lexical confusion (e.g. hearing *seem* instead of *stream*) was not included in this agent based model.

The signal was incorporated into the same word class of the agent listener if it was probabilistically closest to the agent listener's phonological class with which the word class transmitted by the talker was associated. For example, if (i) the agent talker had transmitted the word class *stream* and (ii) the agent listener's *stream* was associated with the phonological class $/s + str/_1$, then the incoming signal was only added to the signals associated with the listener's *stream* if the (posterior) probability of the signal belonging to $/s + str/_1$ was greater than to any other of the agent listener's phonological classes. Thus phonology in this agent based model acts as a brake to restrict perceived signals from being added to memory if their probability of phonological class membership is too low (based on the above criterion).

If a signal was absorbed into the agent listener's memory as described above, then it received a time stamp (the most recent time stamp of all the tokens in the agent's memory). The signal with the oldest time stamp was then removed from the same word class. This ensured that the number of memorized signals per word class remained the same following an interaction.

After such an interaction, an agent talker and agent listener were again randomly selected for the next interaction from the pool of 19 available agents.

The algorithms for splitting and merging phonological classes were applied to all agents' phonological classes after 100 such pairwise interactions had taken place: the value of 100 was chosen based on a pilot study showing that this was roughly the lowest number at which a split of an agent's phonological class was not immediately undone by a merge.

4 Experimental method

4.1 Interaction runs

The experimental results were based on 100 *interaction runs*. A single interaction run consisted of 60,000 agent interactions following the procedure in 3.3. This number of interactions was chosen because pilot experiments had shown that this was typically the point after which no further observable changes to the phonological classes or acoustic signals took place.

The degree of change between the *baseline* and *post-run* was quantified by comparing the phonological classes (Section 4.2) and acoustic signals (Section 4.3) stored in all agents' memories. The baseline refers to the phonological classes and acoustic signals prior to any interaction between agents (but after the agent-internal phonological reclassification in 3.2); the post-run denotes the phonological classes and acoustic signals following 60,000 agent interactions. Given that a post-run is necessarily stochastic (since agent pairs and word classes are randomly selected), 100 separate post-runs were obtained. That is, we ran the agent-based model 100 times (each with 60,000 interactions) in order to incorporate a high degree of the possible (stochastic) variation between post-runs into the analysis.

4.2 Testing for phonological reclassification

The overall hypothesis to be tested in this study was that interaction would cause sibilants in <str> words to shift along the direction of the sound change /s/ → /ʃ/. In terms of phonology, the associated prediction is that sibilants in <str> words should split from <s> word sibilants and increasingly merge with <ʃ> word sibilants. Testing for the generality of such phonological re-classification required abstracting away from the agent-specific phonological classes (e.g. $/str/_1$, $/str/_4$, $/str/_7$) that were seen in Figure 3. For this purpose, the hypotheses were tested by pooling the agents' phonological classes into one of seven so-called equivalence classes shown in Table 2.

An equivalence class is defined in terms of whether or not it includes any <s> words, any <str> words, or any <ʃ> words. Suppose, to take the earlier example, an agent has two phonological classes: $/s/_1$ that includes sibilants in *motorcycle* and *policy* and $/s/_2$ that includes sibilants in some other <s> words. Then both $/s/_1$ and $/s/_2$ are in the same equivalence class **s** because they both include at least one sibilant in <s> words (and none from any <str> or <ʃ> word). Analogously the phonological classes $/s+str/_1$ and $/s+str/_2$ referred to in 3.2 and Figure 3 are both in the equivalence class **s+str** because they both include sibilants from at least one <s> word and from at least one <str> word (and none from any <ʃ> word). With equivalence classes of this kind, the predictions can be recast as follows: if interaction causes phonological reclassification in the direction of the sound change /s/ → /ʃ/, then there should be a decrease in **s+str** and an increase in either **str** or **ʃ+str** (or both) equivalence classes in the post-run compared with the baseline.

4.3 Testing for acoustic change

Acoustic change for sibilants in <str> words was measured by comparing baseline and post-run data in two ways. First, aggregated, smoothed and time-normalized M_1 trajectories were calculated from the DCT-triplets in agents' memories for the baseline and (multiple) post-run conditions following the procedure described in Section 2.2; lowered M_1 trajectories in the post-run data can be interpreted as evidence of a shift in the direction of the sound change /s/ → /ʃ/. Second, the position of each memorized <str> sibilant in the three-dimensional DCT-space was calculated as *op* which is the normalized orthogonal projection onto the line passing through the speaker-specific centroids for sibilants in <s> and <ʃ> words in the same condition *baseline/post-run* (see Appendix for the mathematical details). For any sibilant, positive *op* values denote proximity to typical <s> sibilants (centred at 1), while negative values denote proximity to typical <ʃ> sibilants (centred at -1). A value of *op = 0*

Table 2: The relationship between equivalence classes and word types. 1 and 0 denote, respectively, whether the equivalence class includes or excludes sibilants from the corresponding word types (e.g. [1, 1, 0] for **s+str** means that this equivalence class includes sibilants from a least one <s> word, from at least one <str> word and that it has no sibilants from <ʃ> words).

<s> words	<str> words	<ʃ> words	Equivalence classes
1	0	0	**s**
1	0	1	**s+ʃ**
1	1	1	**s+str+ʃ**
1	1	0	**s+str**
0	1	0	**str**
0	1	1	**ʃ+str**
0	0	1	**ʃ**

Stevens et al: Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English

Art. 8, page 15 of 30

represents sibilants positioned on an orthogonal plane that is equidistant between the <s> and <ʃ> centroids. Thus increasingly negative *op* values can be interpreted as evidence of a shift in the direction of the sound change /s/ → /ʃ/.

5 Results

5.1 Baseline distributions of /s/ and /ʃ/

This section tests whether the data from Australian English speakers are suitable to test our main prediction that the uptake of new signals during interaction will promote the conversion of a synchronic phonetic asymmetry into sound change.

Figure 4 shows the distribution of the two phonological classes /s/ and /ʃ/ in an acoustic space at the baseline i.e. before agent interaction. Note that the phonological class /s/ includes samples from all <str> and <s> words while /ʃ/ includes those from <ʃ> words. The shape of the distributions is asymmetrical: /ʃ/ is compact and symmetrical, while /s/ is wider and is flatter on the left than the right flank. The dashed line at zero indicates the intermediate point between /s/ and /ʃ/ on this measure: evidently, there are more /s/ than /ʃ/ tokens on this zero line. Moreover, the peak of the /ʃ/ distribution is centred at −1 whereas the peak of the /s/ distribution is not at 1, but rather it is left-shifted towards zero. Thus there is evidence that /s/ is skewed towards /ʃ/, analogous to the way that *Y* is skewed towards *X* in Figure 1 panel (b).

Figure 5 shows the data from Figure 4 separated by speaker. Many speakers show a similar pattern to the pooled data, i.e. an asymmetry between the distributions of /s/ and /ʃ/. This suggests that the asymmetry in Figure 4 is not an artefact of unique speaker data (as for example might be the case were one speaker to produce /s/ with a post-alveolar articulation in all contexts). Rather, the asymmetry in the pooled data reflects a tendency that is shared among many speakers. Thus /s/-retraction in Australian English appears to constitute the kind of directional phonetic variation that can be converted into sound change (as sketched for other historical sound changes in Section 1).

While there is evidence in Figure 5 that /s/-retraction is a shared tendency, the asymmetry between the distributions of /s/ and /ʃ/ is nonetheless especially apparent in some individuals, such as F05 and M19 in which the /s/ distributions are bimodal, and F01 and F07 whose /s/-distributions are centred closer to zero (i.e. skewed towards /ʃ/). These individual differences visible in Figure 5 highlight the importance of allowing the possibility of agent-internal phonological reclassification before the agents interacted with each other (cf. Section 3.2). It is not clear whether the individual differences in these data
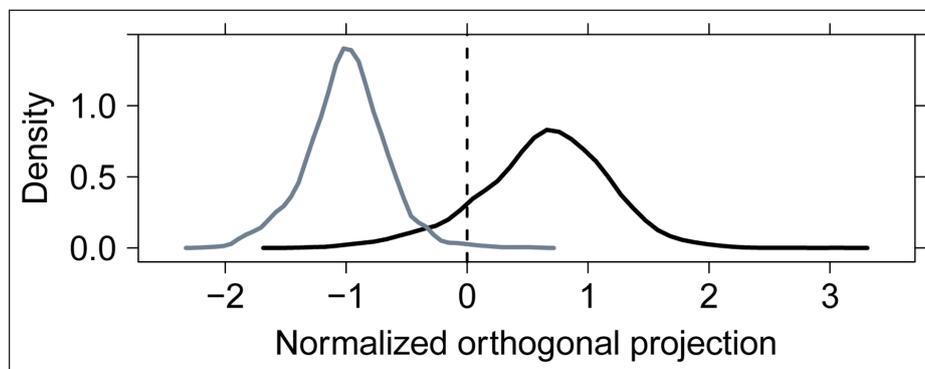

Figure 4: Distributions for normalized orthogonal projections of /s/ (black) and /ʃ/ (grey) at the baseline (i.e. before agent interaction). Note that in this figure /s/ includes sibilants in all <s> and <str> words and /ʃ/ includes sibilants in all <ʃ> words. Tokens on the dashed vertical line at zero are equidistant between /s/ and /ʃ/ centroids.
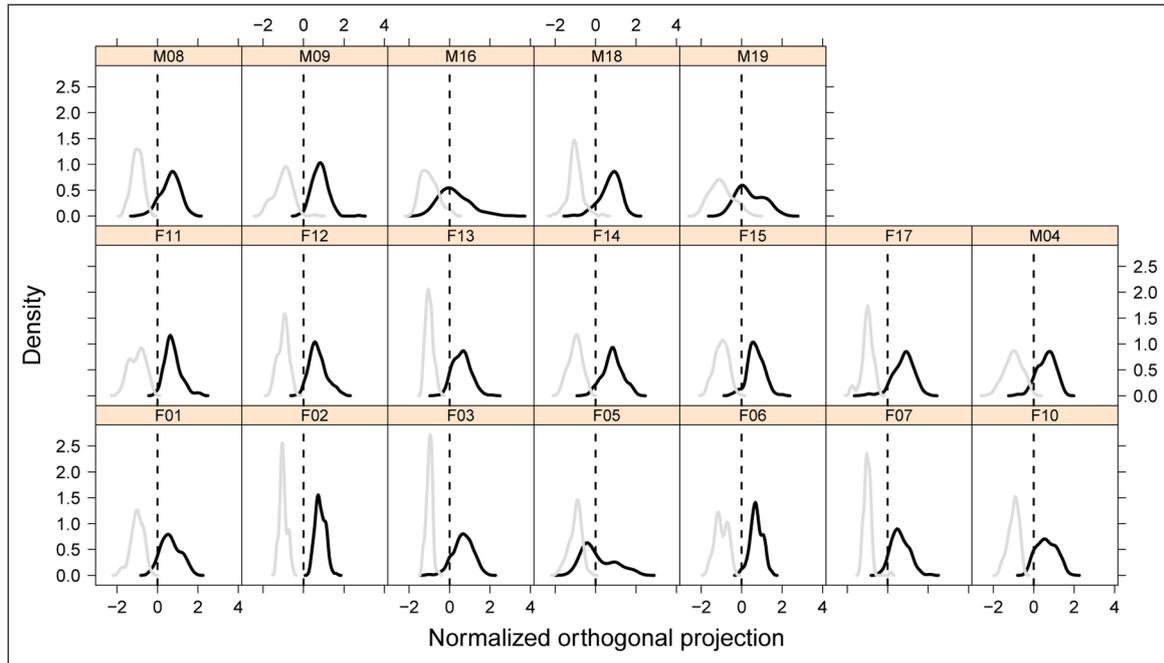
Figure 5: The same data as in Figure 4, shown separately by speaker.

are typical for stable synchronic variation or whether they indicate that the sound change from /s/ to /ʃ/ may already be underway in this community and may have progressed further in some individuals than others. However, the main question in terms of the present study is whether those individuals in which the asymmetry is especially apparent influence other speakers (more than vice versa) when they interact over an extended period of time.

5.2 Interaction and phonological reclassification

Figure 6 shows the proportion of each equivalence class at the baseline and after agent interaction and contains data for <s>, <str> and <ʃ> words. If interaction causes phonological reclassification in the direction of the sound change /s/ → /ʃ/, then there should be a decrease in **s + str** and an increase in either **str** or **ʃ + str** equivalence classes in the post-run compared with the baseline. Consistent with these predictions, a comparison of the equivalence classes between the baseline and post-run in Figure 6 shows a decrease in **s + str** and an increase in **str** due to interaction. There is also a very small increase in **ʃ + str**. If interaction had completed the /s/ → /ʃ/ sound change, then **ʃ + str** should be dominant in the post-run. The pattern of results in Figure 6 therefore points to an *intermediary* change in the direction of the /s/→/ʃ/ sound change for sibilants in <str>-words but without this sound change being completed. This intermediary stage is one in which a large proportion sibilants in <str> words breaks away from a phonological class that is shared with sibilants of <s> words. In terms of the simulation in Figure 1, the sound change caused by interaction has reached panel (d) in which Y has separated into Y_1 and Y_2 (i.e in which sibilants in <s> and <str> words have split and no longer share the same phonological class) but not yet panel (e) in which $X + Y_1$ is formed from a merger of X and Y_1 (in which **ʃ + str** is derived from the merger of sibilants in <ʃ> and in <str> words).

Two separate logistic mixed effect models were carried out using R (R Core Team 2014) and lme4 (Bates, Maechler & Bolker 2015) to test whether between the post-run and baseline there was an increase in **str** or an increase in **ʃ + str** (i.e. in those equivalence classes that were in the direction of the sound change). The binary dependent variable coded
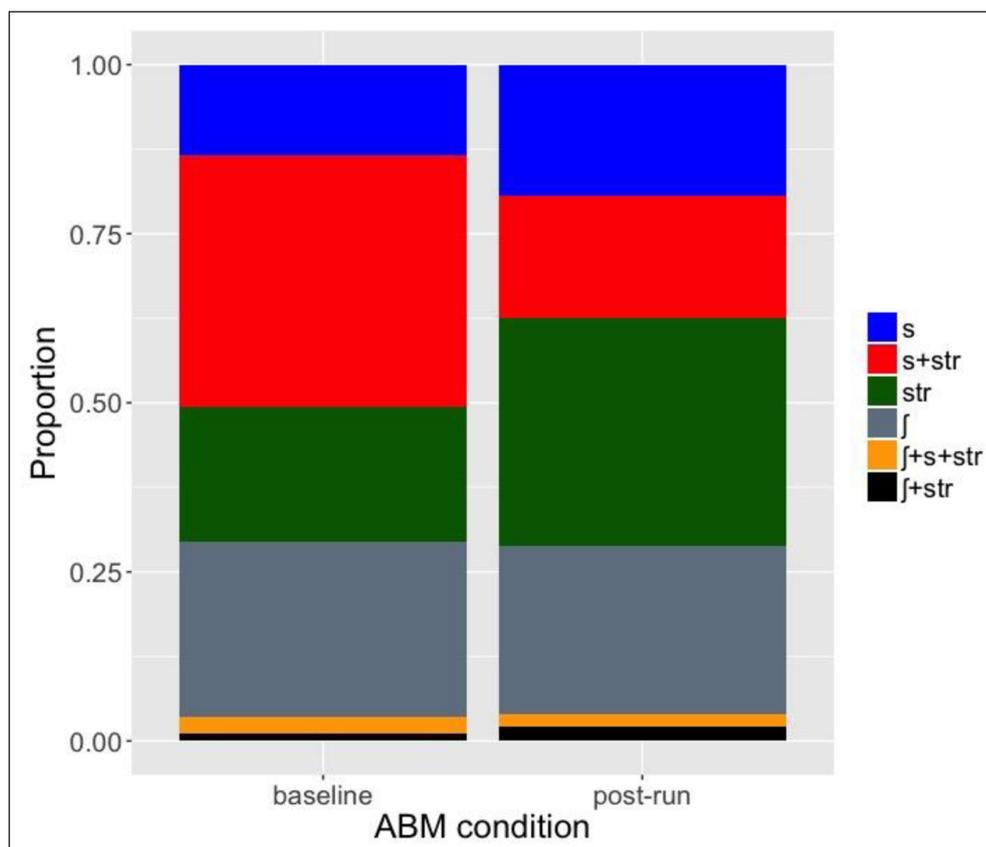
Stevens et al: Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English

Art. 8, page 17 of 30

Figure 6: The proportional occurrence of equivalence classes in the baseline and post-run across all words (i.e. <s>, <str> and <ʃ>), repetitions, speakers, and post-runs. The equivalence class **s+ʃ** that occurred in the post-run only with a proportion of less than 0.001 is not shown. The proportions are E/n where E is the number of occurrences of any one of the 6 equivalence classes (Table 2) across all agents and words and n is the total number of agent-word combinations.

the relevant equivalence class label (**str** or **ʃ+str**) as 1 and all other equivalence classes as 0. In both models, there was one fixed factor (Condition with two levels: baseline vs. post-run) and three random factors: Word (19 levels corresponding to the <str> words in the middle column of Table I), Agent (19 levels), and Interaction Run (a random factor with values of 0, 1, 2, … 100 such that 0 corresponded to the baseline and such that the other values corresponded respectively to the 100 separate post-runs). Consistently with Figure 6, the results showed proportional increases in **str** (Chi Square = 108.4, p < 0.001) and in **ʃ+str** (Chi Square = 9.1, p < 0.05) between the baseline and the post-run.[2]

The by-word bar-charts in Figure 7 show that <str>-words differed in the extent to which they had **str** in the baseline. There was a pattern to these differences. With the exception of *strong,* those words with more than 50% **str** in the baseline (the last 9 words in Figure 7) all had a lexically unstressed vowel following /str/: thus, *astronaut, gastro, claustrophobic, administrate, catastrophe, chemistry, orchestra, oestrogen.* By contrast, and with the

---

[2] The model used was **glmer(Sclass ~ Condition * Stress + (cond-1|ABM) + (1|Word) + (Condition|Agent, family = binomial)** in which **Sclass** was the equivalence class (two levels: σ, ^ σ), **Condition** was before and after the ABM run (two levels: baseline, post-run), **Stress** was lexical stress (two levels: unstressed, stressed), **ABM** was a numerical value 0, 1, 2…100 in which 0 was the baseline and 1, 2, …100 the separate 100 post-runs, and **cond** was a numeric 'dummy' variable with 0 (corresponding to the baseline) and 1 (corresponding to all 100 post-runs). As one reviewer rightly pointed out, the interaction between **ABM** and **Condition** is confounded in a problematic way. The way we dealt with this was with the term **(cond-1|ABM)** that can be used to declare only the multiple post-runs but not the baseline of the ABM as a random factor.
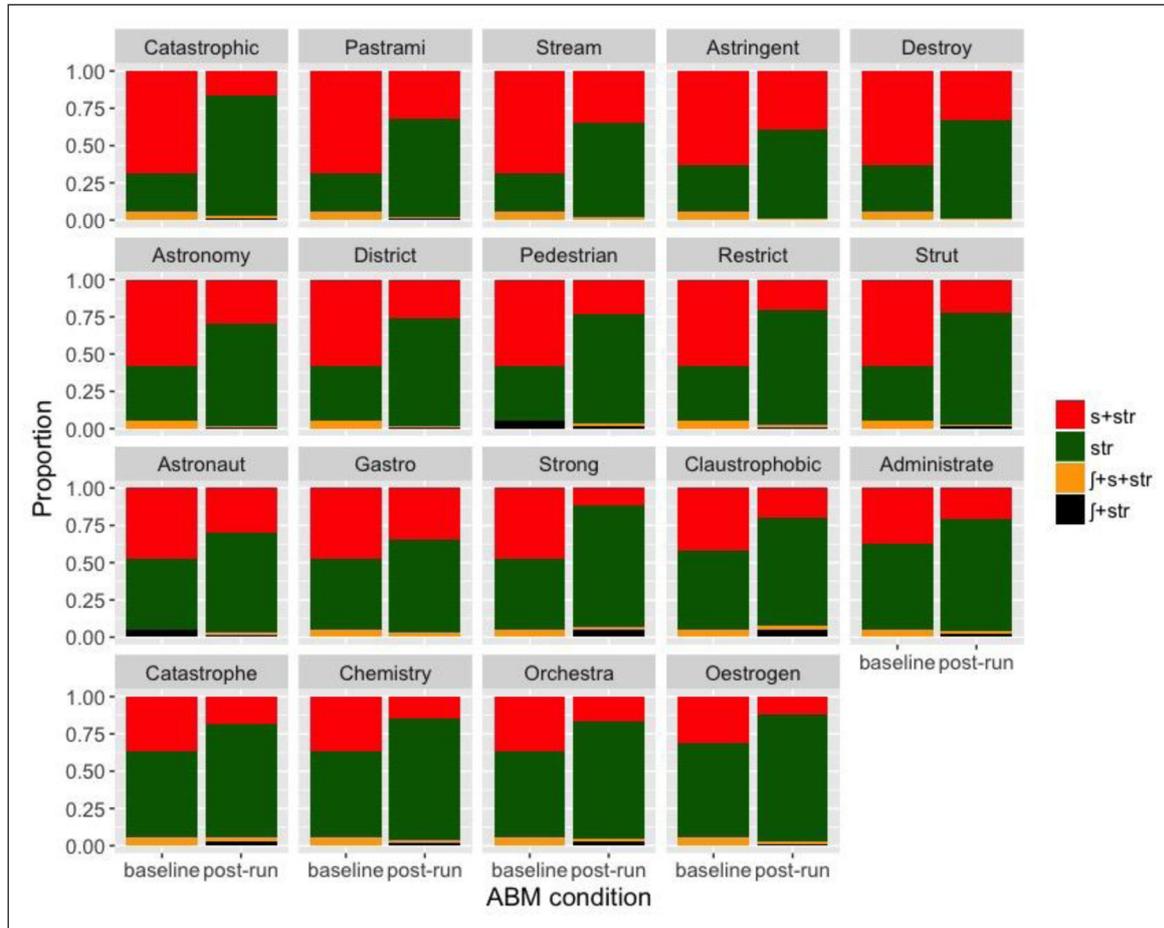
Figure 7: The proportional occurrence of equivalence classes in the baseline and post-run across all repetitions, speakers, and post-runs in <str> words shown separately by word class that are arranged in decreasing order of the proportional occurrence of baseline **s+str**.

exception of *district* and *pedestrian,* words with a lexically stressed vowel following /str/ (*catastrophic*, *pastrami*, *stream*, *astringent*, *destroy*, *astronomy*, *restrict*, *strut*) had less than 50% **str** in the baseline. Following interaction, there was an increase in **str** for all words but the effect was greater in those words with a lexically-stressed vowel following /str/, so that the by-word differences were much smaller after interaction than in the baseline.

5.3 Interaction and acoustic change

The aim in this section is to test whether agent interaction brought about acoustic change for <str> words in the direction of the /s/ → /ʃ/ sound change.

The aggregated M_1-trajectories for <str> words as a function of time in Figure 8 show the absolute difference between the baseline and post-run. The trajectory is lower for the post-run relative to the baseline. Such lowering of M_1 is consistent with the hypothesis that interaction should cause sibilants in <str> words to shift acoustically towards <ʃ> sibilants.

Figure 9 shows combined violin-density and boxplots of the *op* (see Section 4.3) for the conditions baseline (left) and the 100 post-runs (right). A lowered *op* distribution in the post-run compared to the baseline condition is consistent with the hypothesis that sibilants in <str> are closer to <ʃ> following interaction. Notice that both the baseline and the post-run distributions in Figure 9 were well within the positive region. This means that, although <str> sibilants shifted towards <ʃ> due to interaction, they were acoustically still closer to <s> in both conditions.
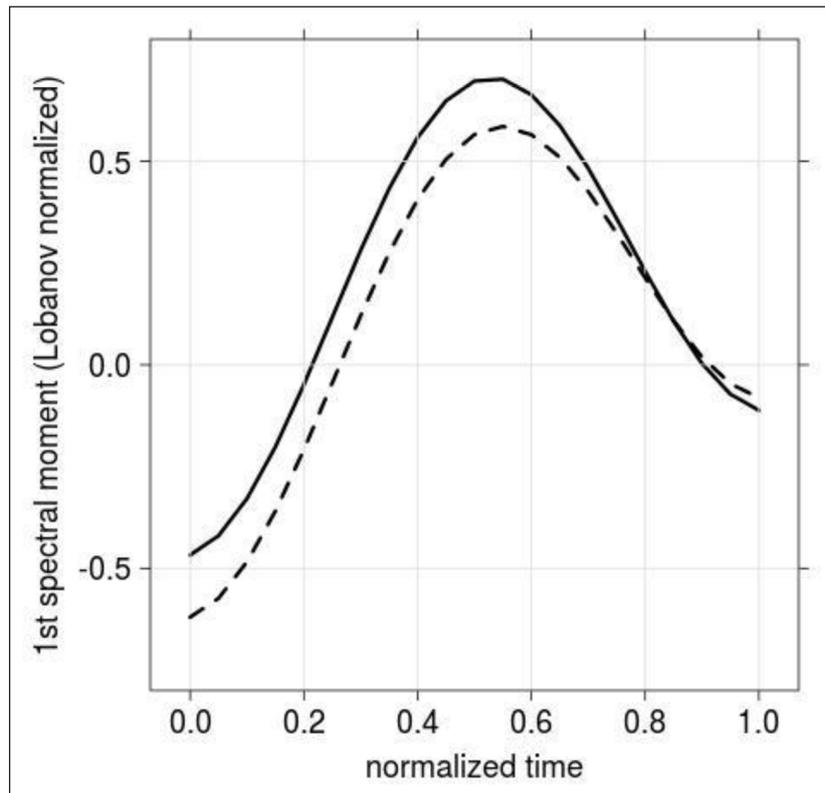
Stevens et al: Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English

Art. 8, page 19 of 30

Figure 8: DCT-smoothed M$_1$ trajectories aggregated over sibilants in <str> words in the baseline (solid) and post-run (dashed).
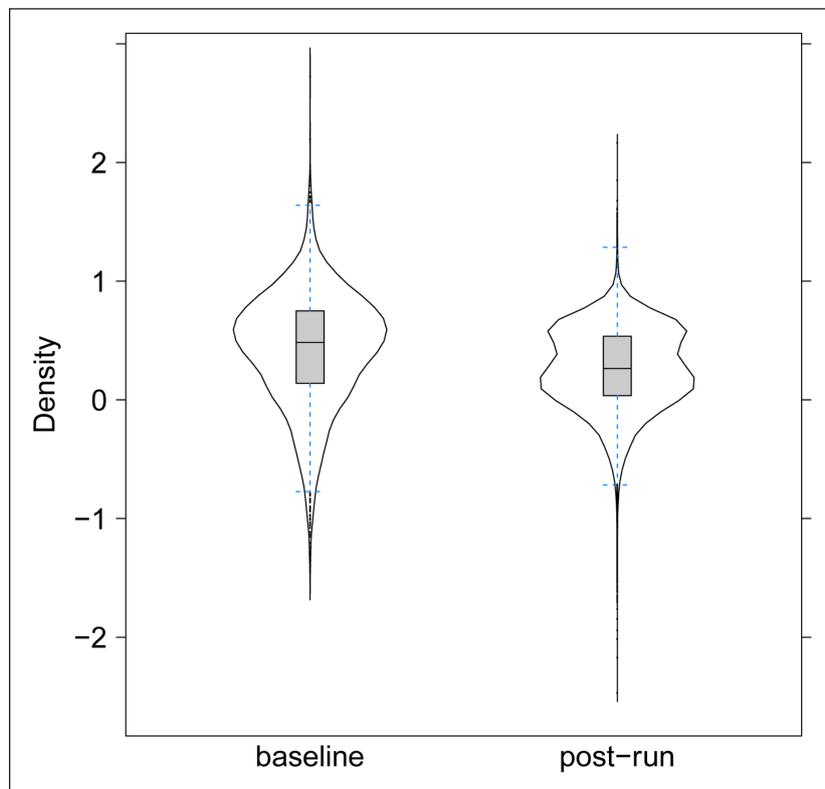

Figure 9: Normalized orthogonal projections *op* of sibilants in <str> words in the baseline (left) and in the post-run (right). Values of zero are equidistant between sibilants in <s> and <ʃ> words; negative values are closer to <ʃ> and positive values are closer to <s>.

A linear mixed effects analysis was carried out to test whether <str> was acoustically closer to <ʃ> following interaction. The dependent variable was the normalized orthogonal projection *op* (cf. Section 4.3). There was one fixed factor Condition (two levels, baseline vs. post-run) and three random factors (Word, Agent, Interaction Run), as was the case for the phonological analysis in 5.2. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. P-values were obtained by likelihood ratio tests of the full model with factor Condition against the model without factor Condition. The results showed a significant effect for Condition (Chi Square = 382.5, $p < 0001$): that is, <str> sibilants were significantly closer to <ʃ> sibilants in the post-run compared with the baseline.[3]

5.4 Bi-directional effects of phonological re-classification and acoustic change

Agent interaction caused a shift in the direction of the /s/ → /ʃ/ sound change for sibilants in <str>, both when examined from a phonological perspective (section 5.2) and from an acoustic perspective (section 5.3). An important aspect of the present model is that it is bi-directional: gradual acoustic change due to the uptake of new signals facilitates phonological re-classification, while phonological re-classification enables the uptake of new signals. Thus it is reasonable to assume that the effects of phonological re-classification and acoustic change might reinforce each other, promoting continued shifts in the direction of the /s/ → /ʃ/ sound change. If so, then the magnitude of the acoustic shift from baseline to post-run should be dependent on phonological re-classification (and vice versa).

Figure 10 shows the normalized orthogonal projection *op* for sibilants in <str> words at the baseline and post-run according to the equivalence class (**s + str** or **str**) that they were associated with. If the density functions were overlapping, then there would be no acoustic change between the baseline (solid lines) and post-run (dashed lines) and, more importantly, there would be no acoustic differences between sibilants that had (**str**) or had not (**s + str**) split from <s> words. Instead, Figure 10 shows a leftwards shift (i.e. increasingly negative normalized orthogonal projection *op* values) in the post-run (dashed) relative to the baseline (solid), which is consistent with the acoustic shift /s/ → /ʃ/ for <str> words described in 5.3. The difference between baseline and post-run distributions for sibilants associated with **s + str** (black) appears smaller than for those associated with **str** (grey). This difference in the effect of agent interaction on **s + str** versus **str** supports the idea that phonological reclassification (in the form of a split from <s> words) promotes acoustic change in the direction of /s/ → /ʃ/ for <str> words. Thus acoustic change and phonological re-categorization appear to reinforce each other during interaction.

The results of a linear mixed effects analysis showed that this interaction between Condition (baseline/post-run) and equivalence class (**s + str, str**) was significant (Chi Square = 107.2, $p < 0.001$).[4] Posthoc tests with Bonferroni correction revealed that all four possible combinations between Condition and equivalence class (cf. densities in Figure 10) were significantly different with $p < 0.05$.

---

[3] The model used was **lmer(op ~ Condition + (1|Agent) + (1|Word) + (cond-1|ABM))** in which **op** was the normalized orthogonal projection, **Condition** was before and after the ABM run (two levels: baseline, post-run), and **ABM** was a numerical value 0, 1, 2 …100 in which 0 was the baseline and 1, 2, …100 the separate 100 post-runs. See Footnote 2 for a definition of **cond-1**.

[4] The model used was **lmer(op ~ Condition * Eclass + (1|Agent) + (1|Word) + (cond-1|ABM))** in which **Condition** was before and after the ABM run (two levels: baseline, post-run), **Eclass** was the binary factor equivalence class (**s + str, str**), and **ABM** was a numerical value 0, 1, 2…100 in which 0 was the baseline and 1, 2, …100 the separate 100 post-runs. See Footnote 2 for a definition of **cond-1**.

Stevens et al: Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English
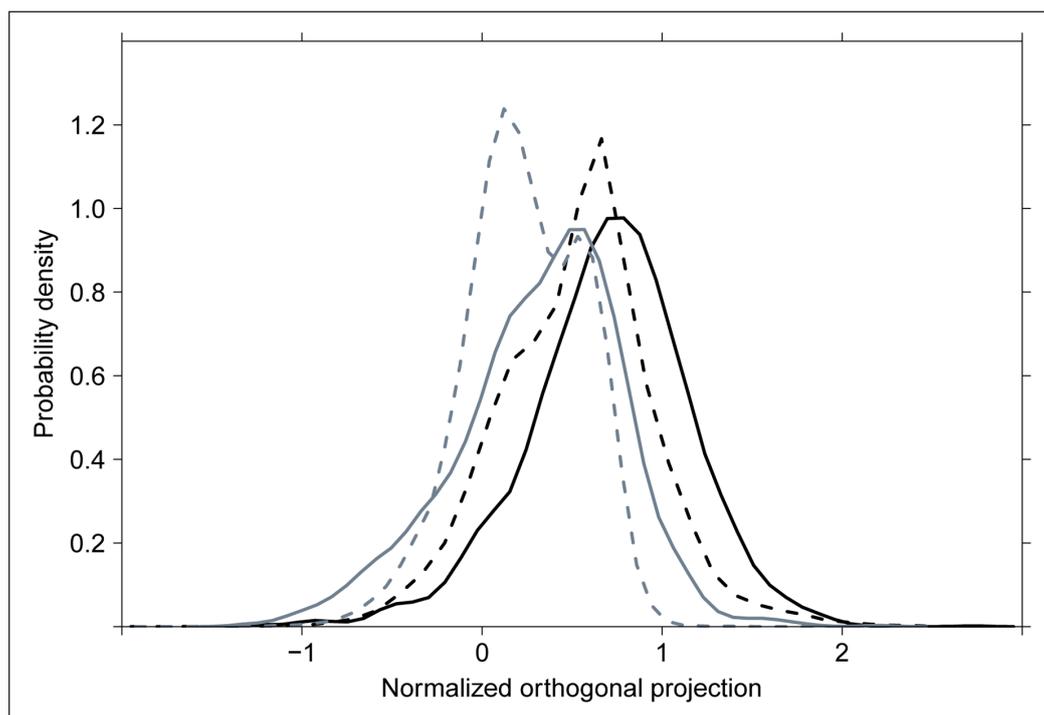
Art. 8, page 21 of 30

Figure 10: Density functions of *op* for sibilants in <str> words according to equivalence class and condition. Equivalence class **s+str** is shown in black and **str** in grey; at the baseline (solid) and post-run (dashed). Leftwards shifts indicate acoustic changes in the direction of <ʃ> sibilants.

6 General discussion and conclusions

Agent interaction in the present model caused a phonological reclassification in the direction of /s/ → /ʃ/ in <str> sibilants. This is because there was a proportional increase of both **str** and **ʃ+str** equivalence classes (i.e. in those classes that were not shared with any <s> words) between the baseline and the post-run. If interaction had completed the sound change, then all <str> sibilants should have been in shared phonological classes with <ʃ> sibilants after interaction: that is, the bar on the right in Figure 6 should have consisted of just two classes, **s** and **ʃ+str**. This was evidently not the case. There was instead evidence from the phonological analysis for an intermediate sound change in the direction of /s/ → /ʃ/ such that <str> sibilants were predominantly in their own class after interaction.

Consistent with the results of phonological reclassification, the results from the acoustic analysis of sibilants in <str> words showed acoustic changes associated with a shift from an alveolar to a more retracted pronunciation, but not with a complete shift to a post-alveolar sibilant [ʃ]. This was seen in Figure 9, which showed that although interaction caused a significant approximation of <str> towards <ʃ> sibilants, they nevertheless remained acoustically closer to <s> after interaction. This is further evidence that the sound change induced by interaction progressed to an intermediate stage in which <s> and <str> sibilants formed their own separate classes but without <str> merging with <ʃ>.

The present study also shows that the changes to <str> words due to interaction are both gradient and quantal. The uptake of new signals caused the distribution of sibilants in <str> words to shift gradually in the direction /s/ → /ʃ/. This sort of gradient pattern was seen in Figure 10, for example, in which **s+str** was closer to /ʃ/ in the post-run than at the baseline. Meanwhile, the incorporation of new signals was occasionally sufficient to trigger a categorical split from <s> words (the change from **s+str** to **str** from baseline

to post-run, seen in Figure 6). Moreover, we saw evidence that these gradient and quantal patterns appear to reinforce each other. A promising avenue for future research would be to determine exactly how these two aspects of sound change reinforce each other during interaction: the present analysis compared only baseline and post-run, but it would of course be possible to track the progression of acoustic and phonological changes over the time course of the simulations.

As far as /s/-retraction in Australian English is concerned, results for the baseline (section 5.1) were consistent with those of Stevens & Harrington (2016) for the same speakers; both studies showed by different means that /s/ in /str/ clusters is acoustically positioned between /s/ and /ʃ/ but closer to /s/. The present study attempted to test whether this pre-existing synchronic tendency towards /s/-retraction might be converted into a permanent sound change /s/ → /ʃ/, if these speakers were to interact with each other over an extended time period. As summarised above, results of the simulations showed that interaction caused an intermediate shift but that the sound change essentially ran out of momentum before completion. This result does not speak against the idea that the sound change /s/ → /ʃ/ in <str> words could occur in Australian English, nor, more broadly, does it suggest that interaction between members of single, isolated communities is insufficient to drive a sound change to completion. Dividing the number of interactions (60,000) by the number of agents (nineteen) and the number of words (41) gives 77.02. Thus an agent would have perceived approximately 77 new samples per word during any interaction run. Moreover, the absorption of these new samples into a listener agent's memory was dependent on phonological classification (Rule 1 of the simulation). Thus it appears entirely plausible that in the present study the absorption of new samples during interaction was sufficient to cause a measurable change to agents' phonetic distributions and the associated phonological classes but not for the completion of the sound change /s/ → /ʃ/ in <str> words. The simulation of further progression of this sound change could be done by increasing the number of interactions (although this would not have helped in the present study because, as noted earlier, 60,000 interactions was the point at which there was typically no further change to agents' phonetic distributions or phonological classes). An increase in the number of interactions in each simulation run before the agents predictably converge to each other could only be brought about by increasing the size of each agent's lexicon and/or the number of agents who interact with each other.[5]

We now consider the extent to which individual differences play a role in the initiation and propagation of sound change (e.g. Baker et al. 2011; Yu et al. 2013; Stevens & Harrington 2014; Zellou 2017). Yu in press, for example, argues that sound change is best understood as a process by which "individuals who exhibit unique perceptual and production strategies relative to the local community of practice may serve as innovators who can sustain the introduction of stable new phonetic variants." In the present study, there was indeed evidence that individual speaker-agents differed in terms of production patterns at the initial state of the simulation, whereby the tendency towards /s/-retraction in <str>-words was evidently stronger in some speakers than others (cf. Figure 5). After the simulated interactions, we observed that there was, on average (i.e. across speakers and ABM runs), a shift in the direction of the /s/ → /ʃ/ sound change for sibilants in

---

[5] Neither of these steps is possible based on our materials (cf. Section 2.1). A Reviewer suggested increasing the number of agents by duplicating speakers or creating new ones (by sampling from a distribution based on the present materials), but this would mean that the model would no longer simulate interactions between real speakers. A more fruitful approach for future research would be to include the possibility that each of the 19 agents interacts with other speakers e.g. their own family members, rather than interacting in a closed group. This would make the model more realistic and might help to maintain more of the individual differences between our 19 agents (of the sort visible in Figure 5) as they interact over longer time periods.

Stevens et al: Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English

Art. 8, page 23 of 30

<str> words. Thus the study shows that innovative speakers (i.e. individuals who are further along the path of the sound change) influenced other speakers more than vice versa. However, this attraction towards innovative speakers does not come about because agents attach social significance (Garrett & Johnson 2013) to the phonetic innovation, but because of the interaction of population dynamics (who speaks to whom and how often) and the orientation of the phonological classes in an acoustic space. While individual differences are crucial to the change we observe, we do not narrow our focus to a single agent or interactions between particular agents. This is partly because it would be misleading to compare an individual agent's distributions between baseline and post-run (all agents' distributions at post-run, aggregated over 100 runs, look very similar to the pooled data), and also because it would be practically impossible to pinpoint the particular sequences of interactions that may have caused an individual agent to have different outcomes in one ABM run versus another. More broadly, processes with random elements like our model are unlikely to be explained by looking at the microscopic behaviour of a single agent, but rather – as in thermodynamics where the behaviour of a gas cannot be predicted by the behaviour of a single molecule – by considering the total system of interacting agents.

The results of the simulations in the present study and in Harrington & Schiel (2017) are relevant to the very many types of sound change that develop out of directional phonetic variation (Garrett & Johnson 2013). Our simulations have shown how interaction causes a retuning of the relationship between word classes, phonological classes and speech signals which can, in turn, cause a category shift in the direction of phonetic variation. The first stage along the path of sound change has been modelled as the development of a bimodal distribution in a phonological class leading to a split. This is the stage at which a phonological class is likely to break up into sub-phonemic classes along context-dependent lines: thus alveolar and post-alveolar variants of /s/ (Baker et al. 2011; Stevens & Harrington 2016), fronted and retracted variants of /u/ (Hall-Lew 2011; Harrington & Schiel 2017), and fronted and retracted velars of which the former palatalise diachronically (Guion 1998). In the case of /s/-retraction, the sound change is then completed by a merge between a retracted /s/ and /ʃ/. Given that each sub-phonemic unit maps to a unique set of lexical items, then sound change in the IP-model (Harrington et al. 2018) is both phonetically and lexically gradual (Bybee 2001; see also Schryver et al. 2008 for a discussion and compatible evidence). The possibility is currently being explored of extending split and merge to cases of phonologisation such as the development of contrastive vowel nasalisation from a sequence of an oral vowel and a post-vocalic nasal consonant (Beddor 2015). The first stage of the sound change would be a split in which there might be separate sub-phonemic classes for vowels before oral (/V/: *said*) and nasal (/Ṽ/: *send*) consonants. The second stage of the sound change might then be a progressive merge of separate sub-phonemic classes in the sequence /Ṽn/. This second stage would then correspond to findings in perception in which listeners perceive nasalisation in the coda but cannot unambiguously attribute it to the vowel or following consonant (Beddor 2012). Listeners have been shown to vary both in parsing this type of coarticulatory nasalisation (Beddor 2009) and in normalising for coarticulation (Kataoka 2011; Yu et al. 2013). The basis for this type of listener variation in the present model is that, since split and merge are agent-specific, agents (i.e. listeners) even of the same variety do not have the same sets of phonological classes. Whether splits and mergers of such finely differentiated sub-phonemic units operate within the same individual as implied by the IP-model is unclear in the absence of many longitudinal studies (but see Harrington 2006 for some evidence which shows how within the same adult over 30 years the final vowel in *happy* is more likely to be associated with /i/ in FLEECE than in /ɪ/ in KIT).

A long-standing issue in developing models of sound change is how to represent the relationship between stability and change (Sóskuthy 2015), and in particular how to devise a model in which there is synchronic variation but in which change is not inevitable (Baker et al. 2011). Indeed, as Ohala (2012) emphasises, only a very small fraction of synchronic change becomes diachronic change. In the IP-model that was applied to /s/-retraction in the present study, stability is provided by phonological classes of low variability: the tighter the clustering of signals that belong to a phonological class, the less likely it is to be influenced by and to split or merge with other classes. Another mechanism of stability in the present model is that a signal is absorbed into a phonological class only if it is not probabilistically closer to another phonological class (Rule 1 of the simulation). On the other hand, the absorption of new signals into a phonological class creates the possibility of change.

Research in the last 20 years has demonstrated that phonetic variation can be word-specific (Goldinger 1996; Pierrehumbert 2001; 2016; Hay & Bresnan 2006) and in some models of human speech processing based on exemplar theory (e.g. Hay et al. 2013) as well as in the IP-model, individual words are associated with their own phonetic distributions. In both Pierrehumbert (2002) and in the IP-model, there is a bias towards individual words in speech production. This does not mean that speech production is based on an entire word template. Instead, the IP-model gives expression to this bias by sampling from word-specific distributions of sub-phonemic classes (e.g. from repetitions over /s/ in *seem* in producing *seem*). By contrast, speech perception must be more flexible and abstract in its sampling. As discussed in Hay et al (2013), whereas production is less likely to draw from the extreme ranges of phonetic variation, perception requires a greater degree of generalisation (i.e. often sampling from the edges of the distribution) for the listener to be able process the vast range of speaking styles and sociophonetic variation (see also Beddor 2015, for a similar view). In the IP-model, there is commensurately a greater degree of abstraction in speech perception than in speech production: in speech perception, an incoming speech signal is evaluated for its acceptability across all the words in the lexicon that share the same sub-phonemic class.

As in Harrington & Schiel (2017), the IP model was implemented in the present study with just sufficient complexity in order to test a specific hypothesis about sound change. It does not model factors such as the strength of network ties between agents (Milroy & Milroy 1985; Wasserman & Faust 1994; Mühlenbernd & Quinley 2013; Dodsworth in press). Its handling of memory retention in which the oldest signal from a distribution is discarded is clearly too simple given the way that words remembered at different stages of the lifespan influence phonetic output (Hay & Foulkes 2016). We are also currently testing various versions of the split-and-merge algorithm on very many different types of sound changes in progress including cases in which there is an unmerging of phonemes (Harrington et al. 2012), a merging of phonemes as in the sound change in progress by which the New Zealand English *ear/ear* contrast has collapsed into one category for younger individuals (Maclagan & Gordon 1996; Watson et al. 1998; Hay et al. 2006), and for a type of metathesis in which pre-aspiration has shifted to post-aspiration in Andalusian Spanish (Ruch & Harrington 2014). The algorithm will be varied by replacing *k*-means clustering as used here with Gaussian mixture modelling (e.g. Kirby 2014) and the aggregated t-test over words with some form of likelihood ratio test (Górriz et al. 2010). The type of split-and-merge algorithm that is ultimately selected will be based on whichever one is most compatible with all these different types of sound change and their relationship to synchronic variation. Finally, while DCT coefficients between an acoustic onset and offset are appropriate for modelling sound changes such as /u/-fronting and /s/-retraction, a different type of representation for modelling the mapping between phonological classes and the acoustic signal will be needed for sound changes that more obviously depend on sequences of phonemes such as phonologisation, metathesis, and dissimilation.

Stevens et al: Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English

Art. 8, page 25 of 30

Abbreviations

ABM = agent-based model, DCT = discrete cosine transform, M_1 = first spectral moment

Additional File

The additional file for this article can be found as follows:

- **Appendix.** Normalized Orthogonal Projection. DOI: https://doi.org/10.5334/gjgl.620.s1

Acknowledgements

Competing Interests

The authors have no competing interests to declare.

References

Baker, Adam, Diana Archangeli & Jeff Mielke. 2011. Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change* 23. 347–374. DOI: https://doi.org/10.1017/S0954394511000135

Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67. 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Beddor, Patrice. 2009. A coarticulatory path to sound change. *Language* 85. 785–821. DOI: https://doi.org/10.1353/lan.0.0165

Beddor, Patrice. 2012. Perception grammars and sound change. In Maria-Josep Solé & Daniel Recasens (eds.), *The Initiation of sound change: Perception, production, and social factors*, 37–55. Amsterdam: John Benjamin. DOI: https://doi.org/10.1075/cilt.323.06bed

Beddor, Patrice. 2015. The relation between language users' perception and production repertoires. In The Scottish Consortium for ICPhS 2015 (eds.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow: University of Glasgow.

Bermúdez-Ortero, Ricardo. 2012. Cycles and continua: On unidirectionality and gradualness in language change. In Terttu Nevalainen & Elizabeth Traugott (eds.), *The Oxford handbook of the history of English*, 691–720. Oxford: Oxford University Press.

Bermúdez-Ortero, Ricardo. 2013. Amphichronic explanation and the life cycle of phonological processes. In Patrick Honeybone & Joseph Salmons (eds.), *The Oxford handbook of historical phonology*, 374–399. Oxford: Oxford University Press.

Blevins, Juliette & Andrew Wedel. 2009. Inhibited sound change: an evolutionary approach to lexical competition. *Diachronica* 26. 143–183. DOI: https://doi.org/10.1075/dia.26.2.01ble

Bloomfield, Leonard. 1933. *Language*. New York: Holt.

Bybee, Joan. 2001. *Phonology and language use*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511612886

Castellano, Claudio, Santo Fortunato & Vittorio Loreto. 2009. Statistical physics of social dynamics. *Reviews of Modern Physics* 81. 591–646. DOI: https://doi.org/10.1103/RevModPhys.81.591

Chang, Steve, Plauché, Madelaine & John Ohala. 2001. Markedness and consonant confusion asymmetries. In Elizabeth Hume & Keith Johnson (eds.), *The role of speech perception in phonology*, 79–101. San Diego, CA: Academic Press.

Cruttenden, Alan. 2014. *Gimson's pronunciation of English*, 8th edition. Oxford/New York: Routledge. DOI: https://doi.org/10.4324/9780203784969

de Schryver, Johan, Anneke Neijt, Pol Ghesquière & Miriam Ernestus. 2008. Analogy, frequency, and sound change. The case of Dutch devoicing. *Journal of Germanic Linguistics* 20. 159–195.

Dodsworth, Robin. In press. Bipartite network structures and individual differences in sound change. *Glossa: A Journal of General Linguistics. Special Collection: Individuals, communities, and sound change.*

Draxler, Christoph & Klaus Jänsch. 2004. SpeechRecorder – A universal platform independent multichannel audio recording software. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, 559–562.

Ernestus, Miriam. 2006. Statistically gradient generalizations for contrastive phonological features. *The Linguistic Review* 23. 217–233. DOI: https://doi.org/10.1515/TLR.2006.008

Ernestus, Miriam & R. Harald Baayen. 2006. The functionality of incomplete neutralization in Dutch: The case of past-tense formation. In Louis Goldstein, Douglas Whalen & Catherine T. Best (eds.), *Laboratory Phonology* 8. 27–49. Berlin: Mouton de Gruyter.

Ettlinger, Marc. 2007. An exemplar-based model of chain shifts. In Jürgen Trouvain & William J. Barry (eds.), *Proceedings of the 16th International Congress of the Phonetic Sciences*, 685–688. Saarbrücken: Universität des Saarlandes.

Forrest, Karen, Gary Weismer, Paul Milenkovic & Ronald Dougall. 1988. Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America* 84. 115–124. DOI: https://doi.org/10.1121/1.396977

Fourakis, Marios & Gregory K. Iverson. 1984. On the 'incomplete neutralization' of German final obstruents. *Phonetica* 41. 140–14. DOI: https://doi.org/10.1159/000261720

Garrett, Andrew & Keith Johnson. 2013. Phonetic bias in sound change. In Alan Yu (ed.), *Origins of sound change*, 51–97. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199573745.003.0003

German, James, Katy Carlson & Janet Pierrehumbert. 2013. Reassignment of consonant allophones in rapid dialect acquisition. *Journal of Phonetics* 41. 228–248. DOI: https://doi.org/10.1016/j.wocn.2013.03.001

Goldinger, Stephen. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22(5). 1166–1183. DOI: https://doi.org/10.1037/0278-7393.22.5.1166

Górriz, Juan Manuel, Javier Ramírez, Elmar Lang, Carlos García Puntonet & Ignacio Turias. 2010. Improved likelihood ratio test based voice activity detector applied to speech recognition. *Speech Communication* 52. 664–677. DOI: https://doi.org/10.1016/j.specom.2010.03.003

Guion, Susan. 1998. The role of perception in the sound change of velar palatalization. *Phonetica* 55. 18–52. DOI: https://doi.org/10.1159/000028423

Haley, Katarina, Elizabeth Seelinger, Kerry Mandulak & David Zajac. 2010. Evaluating the spectral distinction between sibilant fricatives through a speaker-centered approach. *Journal of Phonetics* 38. 548–554. DOI: https://doi.org/10.1016/j.wocn.2010.07.006

Hall, Daniel Currie & Kathleen Currie Hall. 2016. Marginal contrasts and the contrastivist hypothesis. *Glossa: A Journal of General Linguistics* 1(1). 50. 1–23.

Hall-Lew, Lauren. 2011. The completion of a sound change in California English. In Wai Sum Lee & Eric Zee (eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*, 807–810. Hong Kong: City University of Hong Kong.

Harrington, Jonathan, Felicitas Kleber & Ulrich Reubold. 2012. The production and perception of coarticulation in two types of sound change in progress. In Susanne

Stevens et al: Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English

Art. 8, page 27 of 30

Fuchs, Melanie Weirich, Pascal Perrier & Daniel Pape (eds.), *Speech production and speech perception: Planning and dynamics*, 33–55. Bern: Peter Lang.

Harrington, Jonathan, Felicitas Kleber, Ulrich Reubold, Florian Schiel & Mary Stevens. 2018. Linking cognitive and social aspects of sound change using agent-based modeling. *Topics in Cognitive Science* 10(4). 1–21. DOI: https://doi.org/10.1111/tops.12329

Harrington, Jonathan & Florian Schiel. 2017. /u/-fronting and agent-based modelling: The relationship between the origin and spread of sound change. *Language* 93(2). 414–445. DOI: https://doi.org/10.1353/lan.2017.0019

Harrington, Jonathan, Philip Hoole, Felicitas Kleber & Ulrich Reubold. 2011. The physiological, acoustic, and perceptual basis of high back vowel fronting: Evidence from German tense and lax vowels. *Journal of Phonetics* 39. 121–131. DOI: https://doi.org/10.1016/j.wocn.2010.12.006

Hartigan, John & Max Wong. 1979. A k-means clustering algorithm. *Applied Statistics* 28. 100–108. DOI: https://doi.org/10.2307/2346830

Hay, Jennifer & Joan Bresnan. 2006. Spoken syntax: The phonetics of giving a hand in New Zealand English. *The Linguistic Review* 23. 321–49. DOI: https://doi.org/10.1515/TLR.2006.013

Hay, Jennifer, Katie Drager & Brynmor Thomas. 2013. Using nonsense words to investigate vowel merger. *English Language and Linguistics* 17. 241–269. DOI: https://doi.org/10.1017/S1360674313000026

Hay, Jennifer & Paul Foulkes. 2016. The evolution of medial /t/ over real and remembered time. *Language* 92(2). 298–330. DOI: https://doi.org/10.1353/lan.2016.0036

Hay, Jennifer, Paul Warren & Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34. 458–484. DOI: https://doi.org/10.1016/j.wocn.2005.10.001

Heinz, John & Kenneth Stevens. 1961. On the properties of voiceless fricative consonants. *Journal of the Acoustical Society of America* 33. 589–596. DOI: https://doi.org/10.1121/1.1908734

Iskarous, Khalil, Christine Shadle & Michael Proctor. 2011. Articulatory-acoustic kinematics: The production of American English /s/. *Journal of the Acoustical Society of America* 129. 944–954. DOI: https://doi.org/10.1121/1.3514537

Jespersen, Otto. 1913. *Lehrbuch der Phonetik,* 2nd edition. Leipzig: G. B. Teubner.

Jongman, Allard, Joan Sereno, Marianne Raaijmakers & Aditi Lahiri. 1992. The phonological representation of [voice] in speech perception. *Language and Speech* 35. 137–152. DOI: https://doi.org/10.1177/002383099203500212

Jongman, Allard, Ratree Wayland & Serena Wong. 2000. Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America* 108. 1252–1263. DOI: https://doi.org/10.1121/1.1288413

Kataoka, Reiko. 2011. *Phonetic and cognitive bases of sound change*. Berkeley, CA: The University of California, Berkeley PhD dissertation.

Kiparsky, Paul. 2015. Phonologization. In Patrick Honeybone & Joseph Salmons (eds.), *The Oxford handbook of historical phonology*, 563–582. Oxford: Oxford University Press.

Kiparsky, Paul. 2016. Labov, sound change, and phonological theory. *Journal of Sociolinguistics* 20. 464–488. DOI: https://doi.org/10.1111/josl.12196

Kirby, James. 2014. Incipient tonogenesis in Phnom Penh Khmer: Computational studies. *Laboratory Phonology* 5. 195–230. DOI: https://doi.org/10.1515/lp-2014-0008

Kisler, Thomas, Uwe Reichel & Florian Schiel. 2017. Multilingual processing of speech via web services. *Computer Speech & Language* 45. 326–347. DOI: https://doi.org/10.1016/j.csl.2017.01.005

Kleber, Felicitas, Tina John & Jonathan Harrington. 2010. The implications for speech perception of incomplete neutralization of final devoicing in German. *Journal of Phonetics* 38. 185–196. DOI: https://doi.org/10.1016/j.wocn.2009.10.001

Koenig, Laura, Christine Shadle, Jonathan Preston & Christine Mooshammer. 2013. Toward improved spectral measures of /s/: results from adolescents. *Journal of Speech, Language, and Hearing Research* 56. 1175–1189. DOI: https://doi.org/10.1044/1092-4388(2012/12-0038)

Kraljic, Tanya, Susan Brennan & Arthur Samuel. 2008. Accommodating variation: Dialects, idiolects, and speech processing. *Cognition* 107. 54–81. DOI: https://doi.org/10.1016/j.cognition.2007.07.013

Kümmel, Martin. 2007. *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen*. Wiesbaden: Reichert.

Labov, William. 1994. *Principles of linguistic change. Vol. 1: Internal factors*. Oxford: Blackwell.

Labov, William. 2001. *Principles of Linguistic Change. Vol. 2: Social factors*. Oxford: Blackwell.

Labov, William. 2007. Transmission and diffusion. *Language* 83. 344–387. DOI: https://doi.org/10.1353/lan.2007.0082

Ladd, Robert D. 2006. 'Distinctive phones' in surface representation. In Louis Goldstein, Douglas Whalen & Catherine T. Best (eds.), *Laboratory Phonology* 8. 3–26. Berlin: Mouton de Gruyter.

Lawrence, Wayne. 2000. /str/->/ʃtr/: Assimilation at a distance? *American Speech* 75(1). 82–87. DOI: https://doi.org/10.1215/00031283-75-1-82

Lobanov, Boris. 1971. Classification of Russian vowels by different speakers. *Journal of the Acoustical Society of America* 49. 606–608. DOI: https://doi.org/10.1121/1.1912396

Maclagan, Margaret & Elizabeth Gordon. 1996. Out of the AIR and into the EAR: Another view of the New Zealand diphthong merger. *Language Variation and Change* 8. 125–147. DOI: https://doi.org/10.1017/S0954394500001095

Mielke, Jeff, Bridget Smith & Michael J. Fox. In press. A corpus and articulatory study of covert articulatory variation and its phonological consequences in Raleigh, NC English. *Glossa: A Journal of General Linguistics. Special Collection: Individuals, communities, and sound change.*

Milroy, James & Lesley Milroy. 1985. Linguistic change, social network, and speaker innovation. *Journal of Linguistics* 21. 339–384. DOI: https://doi.org/10.1017/S0022226700010306

Mühlenbernd, Roland & Jason Quinley. 2013. Signalling and simulations in sociolinguistics. *University of Pennsylvania Working Papers in Linguistics* 19. 129–138.

Ohala, John. 1993. The phonetics of sound change. In Charles Jones (ed.), *Historical linguistics: Problems and perspectives*, 237–278. London: Longman.

Ohala, John. 2012. The listener as a source of sound change: An update. In Maria-Josep Solé & Daniel Recasens (eds.), *The initiation of sound change: perception, production, and social factors*, 21–36. Amsterdam: John Benjamins.

Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In Joan Bybee & Paul Hopper (eds.), *Frequency effects and the emergence of lexical structure*, 137–157. Amsterdam: John Benjamins.

Pierrehumbert, Janet. 2002. Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory Phonology* 7. 101–139. Berlin: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110197105.101

Pierrehumbert, Janet. 2003. Probabilistic phonology: Discrimination and robustness. In Rens Bod, Jennifer Hay & Stefanie Jannedy (eds.), *Probabalistic linguistics*, 177–228. Cambridge, MA: MIT Press.

Stevens et al: Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English

Art. 8, page 29 of 30

Pierrehumbert, Janet. 2016. Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics* 2. 33–52. DOI: https://doi.org/10.1146/annurev-linguistics-030514-125050

Polka, Linda. 1991. Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America* 89. 2961–2977. DOI: https://doi.org/10.1121/1.400734

Port, Robert & Michael O'Dell. 1985. Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13. 455–471.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for statistical computing, Vienna. http://www.R-project.org/.

Reinisch, Eva, David Wozny, Holger Mitterer & Laurie Holt. 2014. Phonetic category recalibration: What are the categories? *Journal of Phonetics* 45. 91–105. DOI: https://doi.org/10.1016/j.wocn.2014.04.002

Reinisch, Eva & Holger Mitterer. 2016. Exposure modality, input variability and the categories of perceptual recalibration. *Journal of Phonetics* 55. 96–108. DOI: https://doi.org/10.1016/j.wocn.2015.12.004

Roettger, Timo, Bodo Winter, Sven Grawunder, James Kirby & Martine Grice. 2014. Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics* 43. 11–25. DOI: https://doi.org/10.1016/j.wocn.2014.01.002

Rohlfs, Gerhard. 1966. *Grammatica storica della lingua italiana e dei suoi dialetti: Fonetica*. Pisa: Einaudi.

Ruch, Hanna & Jonathan Harrington. 2014. Synchronic and diachronic factors in the change from pre-aspiration to post-aspiration in Andalusian Spanish. *Journal of Phonetics* 45. 12–25. DOI: https://doi.org/10.1016/j.wocn.2014.02.009

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

Scobbie, James & Jane Stuart-Smith. 2008. Quasi-phonemic contrast and the indeterminacy of the segmental inventory. In Peter Avery, B. Elan Dresher & Keren Rice (eds.), *Contrast in phonology*, 87–114. Berlin: de Gruyter.

Shadle, Christine. 2012. The acoustic and aerodynamics of fricatives. In Abigail Cohn, Cecile Fougeron & Marie Huffman (eds.), *The Oxford handbook of laboratory phonology*, 511–526. Oxford: Oxford University Press.

Sóskuthy, Marton. 2015. Understanding change through stability: A computational study of sound change actuation. *Lingua* 163. 40–60. DOI: https://doi.org/10.1016/j.lingua.2015.05.010

Stanford, James & Laurence Kenny. 2013. Revisiting transmission and diffusion: An agent-based model of vowel chain shifts across large communities. *Language Variation and Change* 25. 119–153. DOI: https://doi.org/10.1017/S0954394513000069

Stevens, Kenneth. 1998. *Acoustic phonetics*. Cambridge, MA: MIT Press.

Stevens, Mary & Jonathan Harrington. 2014. The individual and the actuation of sound change. *Loquens* 1(1). DOI: https://doi.org/10.3989/loquens.2014.003

Stevens, Mary & Jonathan Harrington. 2016. The phonetic origins of /s/-retraction: acoustic and perceptual evidence from Australian English. *Journal of Phonetics* 58. 118–134. DOI: https://doi.org/10.1016/j.wocn.2016.08.003

Stevens, Mary, Veronique Bukmaier & Jonathan Harrington. 2015. Pre-consonantal /s/ retraction. In The Scottish Consortium for ICPhS 2015 (eds.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow: University of Glasgow.

Strycharczuk, Patrycja & James Scobbie. 2016. Gradual or abrupt? The phonetic path to morphologisation. *Journal of Phonetics* 59. 76–91. DOI: https://doi.org/10.1016/j.wocn.2016.09.003

Trudgill, Peter. 2008. On the role of children, and the mechanical view: A rejoinder. *Language in Society* 37. 277–280. DOI: https://doi.org/10.1017/S0047404508080354

Trudgill, Peter. 2011. *Sociolinguistic typology*. Oxford: Oxford University Press.

Warner, Natasha, Allard Jongman, Joan Sereno & Rachèl Kemps. 2004. Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics* 32. 251–276. DOI: https://doi.org/10.1016/S0095-4470(03)00032-9

Warren, Paul. 1996. /s/-retraction, /t/-deletion and regional variation in New Zealand English /str/ and /stj/ clusters. In Paul Warren & Catherine Watson (eds.), *Proceedings 11th Australian International Conference on Speech Science & Technology*, 466–471. Canberra: Australasian Speech Science and Technology Association Inc.

Wasserman, Stanley & Katherine Faust. 1994. *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511815478

Watson, Catherine & Jonathan Harrington. 1999. Acoustic evidence for dynamic formant trajectories in Australian English Vowels. *Journal of the Acoustical Society of America* 106. 458–468. DOI: https://doi.org/10.1121/1.427069

Watson, Catherine, Jonathan Harrington & Zoe Evans. 1998. An acoustic comparison between New Zealand and Australian English vowels. *Australian Journal of Linguistics* 18. 185–207. DOI: https://doi.org/10.1080/07268609808599567

Wiese, Richard. 1996. *The phonology of German*. Oxford: Clarendon Press.

Winitz, Harris, M. E. Scheib & James Reeds. 1972. Identification of stops and vowels for the burst portion of /p, t, k/ isolated from conversational speech. *Journal of the Acoustical Society of America* 51. 1309–1317. DOI: https://doi.org/10.1121/1.1912976

Winkelmann, Raphael, Jonathan Harrington & Klaus Jänsch. 2017. EMU-SDMS: Advanced speech database management and analysis in R. *Computer Speech & Language* 45. 392–410. DOI: https://doi.org/10.1016/j.csl.2017.01.002

Yu, Alan. In press. Reconceptualising phonologization. *Glossa: A Journal of General Linguistics. Special Collection: Individuals, communities, and sound change.*

Yu, Alan, Carissa Abrego-Collier & Morgan Sonderegger. 2013. Phonetic imitation from an individual-difference perspective: Subjective attitude, personality, and 'autistic' traits. *Plos One* 8(9). e74746. DOI: https://doi.org/10.1371/journal.pone.0074746

Zellou, Georgia. 2017. Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics* 61. 13–29. DOI: https://doi.org/10.1016/j.wocn.2016.12.002