# Open Library of Humanities

# Individual variation and the coarticulatory path to sound change: agent-based modeling of /str/ in English and Italian

**Mary Stevens,** Institute of Phonetics and Speech Processing, Ludwig-Maximilians-University, Schellingstraße 3, Munich, DE, mes@phonetik.uni-muenchen.de

**Jonathan Harrington,** Ludwig-Maximilians-Universität München, Germany, jmh@phonetik.uni-muenchen.de

This study is concerned with sound change in single populations. Sources differ as to whether such sound changes are attributed to the accumulation of coarticulatory variation, i.e. shared tendencies, or whether idiosyncratic variation is necessary to shift population norms. Using agent-based modeling, this study compares effects of interaction amongst members of single populations with (English-speaking) and without (Italian-speaking) coarticulatory variation involving /s/-retraction in /str/. After interaction, directional shifts /s/ -> /ʃ/ are found in English but not Italian /str/. Manipulating starting conditions to remove outlier individuals from the group of interacting agents has little effect on population-level outcomes. These results provide empirical support for the idea that sound change originates in phonetic variation that is common, rather than exceptional, in a population.

# 1. Introduction
## 1.1 Language change in single populations

The main reason why languages change and diversify is that synchronic variation can accumulate amongst speakers who interact with one another (Levinson and Gray 2012; Ringe and Eska 2013; Trudgill 2011). Referring to this process as "neutral change", Gavin et al. (2013: 513) outline how "just as in biological evolution, neutral change in heritable language units can accumulate in speech communities [and] may eventually disrupt mutual comprehension with other communities." This inherent tendency to change is often, particularly within sociolinguistic frameworks (e.g. Kerswill et al. 2008) referred to as *drift* – following Sapir's (1921: 150) observation that "a language moves down time in a current of its own making. It has a drift". Thus a contrast can be drawn between linguistic changes with their origins in language contact (e.g. Sankoff 2001) versus those that occur in single populations and can be attributed to drift, which involves "the tendency of a language to keep changing in the same direction over many generations" (Trask 2000: 98) and can affect both lexical form and grammatical structure (Ellison & Miceli 2012). Within historical linguistics, attested cases of linguistic change in single populations are often attributed to the idea that variation between individuals can develop into variation between languages (e.g., Sievers 1876 in Hinskins 2021:21; François 2014 amongst many others). Evolutionary approaches to language change also point out that "diversity, ultimately, depends on the concept of the idiolect" (Samuels 1972: 89). Individual differences are also crucial to the spread of sound change in sociolinguistic frameworks (Labov 1963; Milroy and Milroy 1985; Trudgill 1986), which rest on the idea that individuals copy pronunciations that they associate with social categories. Nonetheless, the idea that idiosyncratic behaviour plays a role in shaping languages remains largely untested and in the case of sound changes, in particular, an alternative explanation involves the accumulation of coarticulatory variation i.e. variation within – rather than between – individual speakers.

## 1.2 Sound change as the accumulation of synchronic phonetic biases

Many sound changes are understood to have their roots in synchronic variation caused by phonetic forces that are an inherent part of the transmission of language between human speakers and listeners (e.g., Paul 1888; Harrington 2012; Garrett & Johnson 2013). These phonetic forces apply to the perception of speech as well as to the way that articulatory gestures are planned and executed in time, producing variation that is non-random and directional. For example, vowels are predictably nasalized before nasal consonants as in *tan*, because speakers tend to lower the velum in anticipation of the upcoming nasal while still producing the vowel. The degree to which velar lowering precedes the oral closure gesture shows language specific patterns (e.g. Solé 2007) but the tendency is common across languages. The development into sound change comes about when this synchronic tendency towards anticipatory vowel nasalisation becomes an

obligatory, contrastive feature in a language (e.g. Latin /manus/ > French *main* /mɛ̃/ 'hand'). A substantial body of experimental research has documented fine-grained (synchronic) phonetic biases in production and perception and highlighted parallels with attested sound changes in the same direction such as the development of contrastive nasalization (Beddor 2009; Carignan et al. 2021), tonogenesis (Coetzee et al. 2018; Kirby 2014); /u/-fronting (Harrington et al. 2008), /l/-vocalization (Recasens 2012, Lin et al. 2014), /s/-retraction (Stevens & Harrington 2016) and velar palatalisation (Guion 1998).

Phonetic models of sound change differ as to whether the conversion of a phonetic bias into a permanent, obligatory feature of a language happens primarily for perceptual or articulatory reasons. In Ohala's (1993) influential model, a listener mistakes a coarticulated pronunciation for a categorically different sound and copies this (misperceived) sound in future interactions, while other models (e.g. Lindblom et al. 1995; Bybee 2012) assign more weight to the role of the speaker in this process. Phonetic models also differ in terms of whether the change within an individual language user's grammar is abrupt – as in Ohala's model in which the listener mistakes one category for another – or gradual – as in usage-based models in which change is gradual as phonological categories are updated through use (and see also e.g. Soskuthy 2015: 44 on the differences between these two frameworks). Nonetheless, there is general agreement in the literature that synchronic coarticulatory variation can, via everyday speaker-listener interactions, cause sounds to change permanently. Strictly speaking, this idea remains an assumption because there is very little empirical evidence from direct observation of the beginnings of a sound change in a single community of speakers (but see e.g. Kuang & Cui 2018 on Register change in Southern Yi). This is primarily because historical sound change is rare: phonetic biases can accumulate to cause permanent sound change in the language spoken by one population at one particular point in time, while remaining dormant in most other cases (Weinreich et al. 1968). It is also because the synchronic phonetic variation from which sound changes are thought to emerge can be very subtle and difficult to detect. Finally, sound changes also tend to happen over generations (e.g. Bloomfield 1933: 365; Weinreich et al. 1968), which means that they do not lend themselves to classic laboratory experiments (but see e.g. Hay et al. 2015 for a corpus-based phonetic study of sound change over multiple generations of New Zealand English speakers).

## 1.3 Individual variation as a catalyst for sound change

Because of the rarity of sound change, it has been suggested that a model relying on the accumulation of phonetic biases via interaction must be insufficient (e.g., Baker et al. 2011) and that some additional factor is needed to catalyze sound change in a particular speech community at a particular time. Echoing ideas from historical and evolutionary linguistics outlined in section 1.1, research on sound change has recently converged on the idea that differences between individual members of a population might play a crucial role in its earliest stages (e.g. Hall-Lew

et al. 2021; Stevens & Harrington 2014; Yu & Zellou 2019 for discussions). Garrett & Johnson (2013) reason that since "bias factors are in principle present throughout a language community, in the speech of one or more individuals there must be a deviation from the norm for some reason". Garrett and Johnson (2013) suggest that these kinds of deviation concern differences in social awareness. They propose that certain individuals might be especially inclined to attach social meaning to variation arising initially for purely phonetic reasons and to modify their own production patterns accordingly (e.g. to signal group membership). Other sources suggest that the conversion of a phonetic bias into a sound change might depend on individual differences in the magnitude of the phonetic bias (Baker et al. 2011) or in perceptual processing style (e.g. Yu 2010).

The idea that interactions between individuals might play a role in the very earliest stages of sound change (by amplifying pre-existing phonetic biases) finds support in the literature on imitation, which shows that speakers imitate fine-grained phonetic details of the speech that they hear (e.g. Nielsen 2011; Pardo et al. 2012; Zellou et al. 2016). However, imitation is not an automatic consequence of interactions between individuals (see e.g. Harrington et al. 2016; Nguyen & Delvaux 2015 for discussions). Indeed, evidence from the literature on perceptual learning shows that listeners are very good at compensating for idiosyncrasies in the speech signal (Norris et al. 2003) and do not imitate pronunciations that they deem unusual or idiosyncratic in their own subsequent productions (Kraljic et al. 2008). The extent to which variation between interacting members of a speech community contributes to the accumulation of phonetic biases, i.e. to the very earliest stages of sound change, remains to be shown.

## 1.4 Agent-based modeling of sound change

Computational agent-based models (ABMs) provide a structured way of probing how everyday interactions between individual speakers and listeners (represented by agents) might cause incremental change to shared pronunciation norms. ABMs allow researchers to control parameters such as who interacts with whom, what is said, and over what time-span, and to repeat simulation runs to test the effect of different parameter settings. Several studies have employed ABMs to test a range of hypotheses about sound change (e.g. Blevins & Wedel 2009; Stanford & Kenny 2013; Garrett & Johnson 2013; Kirby & Sonderegger 2013; Sóskuthy 2015, Harrington and Schiel 2017; Harrington et al. 2019; Stevens et al. 2019; Todd et al. 2019). The information exchanged by agents in these ABMs is typically some sort of symbolic, artificially generated representation of speech, e.g. vowels might be represented in a one-dimensional numeric space. In the interactive-phonetic (IP) ABM developed by Harrington and colleagues (Harrington & Schiel 2017; Harrington et al. 2018; 2019a), agents exchange parameterized speech signals that are calculated on speech produced by real speakers and thus include the kind of fine-grained and dynamic phonetic differences that are crucial to sound change.

The concern in Harrington and Schiel (2017) was with the effects of contact between different groups of speakers on vowel categories – in their case older and younger Standard Southern British English speakers, for whom /u/ is retracted [u] or fronted [ʉ], respectively. Interaction involved the exchange of parameterized formant trajectory data and word labels between a randomly chosen agent-speaker and agent-listener pair. A new token was absorbed into the agent-listeners' memory if it was probabilistically closest to their own distribution for that vowel. After interaction, older agents' /u/ showed a shift towards the front of the vowel space, whereas the younger agents' /u/ (which was already fronted before interaction) remained unchanged. Thus Harrington and Schiel (2017) found evidence of a large shift for one group in the direction of the historical sound change known as /u/-fronting. The asymmetrical nature of the shift, whereby one group shifted much more than the other, was attributed to the orientation of the vowels with respect to each other in acoustic space before interaction. More specifically, the younger speakers' vowels showed a compact distribution, to which the older speakers' wider distribution was skewed at baseline, and to which the older speakers' /u/ shifted during interaction.

The same ABM architecture was extended and applied to investigate sound change in a single population in Stevens et al. (2019). The challenge in this case was to model the conversion of a pre-existing phonetic bias into a permanent sound change without using contact with another group of speakers as a catalyst for change. Participants were all monolingual long-term residents of the same town. This homogeneity ensured that any differences between speakers could be attributed to idiosyncratic pronunciations rather than to regional/dialectal variation, and that the interactions to be simulated took place between members of a single population. The sound change under investigation was /s/-retraction, in which e.g. the initial sound in *street* comes to sound like that in *sheet,* which is documented to have taken place in the historical development of German (compare e.g. English *swan* and German *Schwan* /ʃvaːn/) and in a number of other languages (e.g. Kümmel 2007; Kokkelmans 2020). Synchronic phonetic variation in Australian English includes a bias towards /s/-retraction in /str/, the effects of which can be seen in **Figure 1**.

**Figure 1** shows the baseline (i.e. before agent interaction) distribution in acoustic space of pre-vocalic /s/ (in black) and of pre-vocalic /ʃ/ (in grey), and /s/ in /str/ clusters (in red). The x-axis shows the orthogonal projection which calculates the position of a sibilant on a line passing through speaker-specific centroids for pre-vocalic /s/ (centred at 1) and /ʃ/ (centred at –1) and thus indicates the degree of /s/-retraction (see section 2.1 for more detail). The important point here is that the /str/ distribution is left-shifted with respect to that of pre-vocalic /s/, i.e., samples from /str/ contexts show increased proximity to /ʃ/. This tendency towards /s/-retraction in /str/ was found to be common to all speaker participants and can be readily explained in terms of assimilation to the upcoming alveolar/retroflex rhotic approximant [ɹ]/[ɻ] in English (e.g. Shapiro 1995; Rutter 2011).
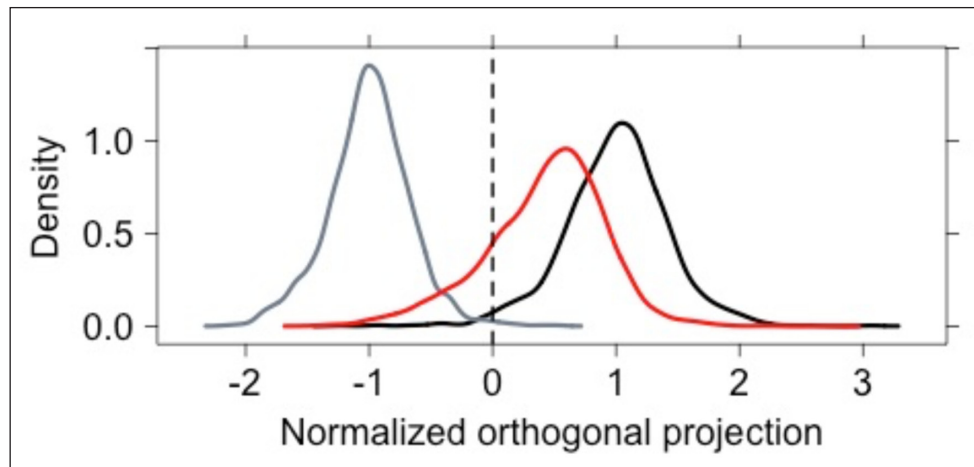
**Figure 1:** Baseline distribution in acoustic space of Australian English sibilants, with samples from pre-vocalic /ʃ/ in grey, pre-vocalic /s/ in black and /str/ in red. The x-axis shows the normalized orthogonal projection (OP), which is defined in section 2.1 but can be understood as degree of /s/-retraction: lower OP values indicate increased proximity to /ʃ/. Data aggregated across 19 speaker participants, taken from Stevens et al. (2019).

After agent interaction was simulated using a modified version of the ABM developed by Harrington and Schiel (2017), the /str/-distribution was found to shift further leftwards i.e. in the direction of the sound change /s/ -> /ʃ/. The sound change did not run to completion, which would have involved completely overlapping /str/- and /ʃ/-distributions after agent interaction. Nonetheless, the crucial breakthrough was that the change to the population-level norm, i.e. the further shift in the direction of sound change /s/ -> /ʃ/ for sibilants in /str/, emerged intrinsically from the pre-existing phonetic bias and the simulated interaction between members of a single population. No additional parameter was used to catalyze the shift in the direction of sound change: there was no contact with a different group of speakers (as in e.g. Harrington & Schiel 2017), and agent listeners did not evaluate incoming signals for social meaning (as in e.g. Garrett & Johnson 2013). Furthermore, no parameter was added to the model to drive the sound change in a particular (phonetic) direction (as in e.g. Kirby & Sonderegger 2013; Todd et al. 2019). As such, the results of this study on Australian English /str/ appear to provide empirical evidence for stochastically-based, mechanistic models in which the first stages of sound change are a consequence of the incremental accumulation of phonetic biases in combination with communication density (e.g. Bloomfield 1933), i.e. who speaks to whom and how often.

Nevertheless, consistent with a model in which sound change emerges due to the coarticulatory idiosyncrasies of one individual (e.g. Baker et al. 2011; Yu 2021), the change in the population might have happened because of the influence of a speaker with an especially /ʃ/-like sibilant in /str/, rather than the presence of a shared phonetic bias at the baseline. The sibilant /s/ is known to be particularly idiosyncratic (e.g. Gordon 2002) and, while the tendency towards

/s/-retraction in /str/ was common to all participants that interacted in the agent-based model, it varied considerably across participants in terms of degree. More specifically, in a separate study involving the same speaker participants, Stevens & Harrington (2016) found that sibilants in /str/ were for many speakers closer in proximity to their own pre-vocalic /s/ whereas one speaker showed almost complete acoustic overlap with /ʃ/.

## 1.5 The present study

As just described, the exact role of individual variation in the earliest stages of sound change remains unclear, including whether a single individual can cause change to population norms as speakers interact. The present study seeks to clarify this issue by applying the IP-model ABM architecture (Harrington et al. 2018) to two different populations and manipulating the starting conditions in order to disentangle the effects of phonetic biases in the direction of a potential sound change from idiosyncratic variation between speakers.

The first experiment examines the impact of idiosyncratic variation on population norms as speakers interact, without any phonetic bias in the baseline data. The prediction to be tested here is that there should be no such population-level shift in the direction /s/ -> /ʃ/ after agent interaction. The first challenge was to obtain appropriate baseline data for such an experiment. One possibility would have been to manipulate the Australian English speech signals to remove the coarticulatory effects on sibilants in /str/ e.g. by replacing the sibilant in /str/ with samples from pre-vocalic /s/. However, this would have meant foregoing one of the main strengths of the original study, which was that the input to the ABM constitutes real speech signals produced by real speakers. Vogt & de Boer (2010: 6) noted that studies applying computational models to language change should "initialize the model's parameters as close as possible to reality based on empirical data. Otherwise, any correlation may be an artifact or bias of the parameter settings, rather than a demonstration of the correctness of the model." With this in mind, production data were collected from Italian which, like English, shows a contrast between an anterior sibilant /s/ e.g. *senza* 'without' and a more posterior sibilant /ʃ/ e.g. *scienza* 'science', and allows /str/ clusters e.g. *stretto* 'narrow'. Crucially however, Italian is not known to show a synchronic tendency towards /s/-retraction in /str/. This is because the rhotic /r/ is typically trilled or tapped [r]/[ɾ] in Italian (Bertinetto & Loporcaro 2005) and taps and trills do not exert the long-distance coarticulatory effects that are associated with rhotic approximants in English (e.g. West 1999 for articulatory and acoustic evidence of long-distance coarticulatory effects associated with English /r/). Exceptionally, /s/-retraction in /str/ (i.e. /str/ -> /ʃtr/) is reported for some dialects spoken in Sicily and Calabria (Rohlfs 1966: 225), but only in dialects in which the rhotic is a retroflex approximant – not a tap/trill – and for which there is therefore a plausible phonetic explanation involving assimilation to the upcoming (retroflex) rhotic (Rohlfs 1966: 259), as in varieties of English. In addition to the tap/trilled rhotic pronunciation, the lack of a tendency

towards /s/-retraction in Italian /str/ may also be partly due to the place of articulation of /s/, which is lamino-dental and/or the intervening stop, which is dental. This is because there is articulatory evidence that the tongue tip slides *forward* during /s/ before dental /d/, at least in Catalan (Recasens et al. 1995), i.e. the opposite direction to /s/-retraction. Nonetheless, many regional varieties of Italian appear to show a seemingly sporadic tendency towards /s/-retraction in pre-vocalic position and /sC/ clusters (Rohlfs 1966: 224, 257; see also Maiden & Parry 1997) and in certain words produced with emphasis e.g. /s/ > [ʃ] in *stupido!* 'stupid!' (Silvia Calamai, personal communication). There is otherwise little descriptive or empirical evidence available for /s/-retraction in Italian and for this reason, it was first necessary to verify the assumption that Italian speakers do not show a tendency towards /s/-retraction in /str/, before testing the prediction for Italian /str/ that there should be no evidence of a shift towards /ʃ/ after simulated interaction. As well as constituting a more realistic test case than if the phonetic bias were simply removed from the Australian English data, as noted above, the Italian data are important for the second experiment which addresses the question of whether population-level change depends on idiosyncratic variation.

The second experiment is concerned with the extent to which the effects of interaction vary between individual agents. The question here is whether individual agents simply converge to the mean during interaction or whether – consistent with the idea that sound change is driven by individuals whose pronunciation deviates from the population-level norm, as outlined in section 1.3 – certain agents may influence the direction of any observed population-level shifts more strongly than others. We address these questions by examining the effects of simulated interaction on individual members of the Italian-speaking and Australian English-speaking populations (the same Australian English data that were first analyzed in Stevens et al. 2019). Experiment 2 is also concerned with the question of whether outlier (innovative) speakers are necessary for the magnification of phonetic biases – where they exist – into population-level change. To address this question, the starting conditions were manipulated to remove innovative speakers from the Australian English population. Comparing the outcomes of simulated interaction amongst the entire population and amongst this reduced population will show whether population-level shifts in the direction of sound change depend on especially innovative speakers.

In addition to analyzing a different language, Italian, the present study extends previous research in two further ways: (a) by comparing the outcomes of individual (repeated) simulation runs and (b) by examining effects of interaction over the entire time course of a run (or repeated runs) in which agent interaction occurs. The IP model, like any other ABM, is best used to predict what happens on average rather than in any one instance and the aggregation of data over multiple simulation runs is necessary to avoid reporting an outlier result. Nonetheless, given the two different populations (English and Italian), we are now in a position to compare the range of possible outcomes depending on whether a phonetic bias is present (in which the directionality

of acoustic shifts in /str/ should be consistent over repeated simulation runs) or not (in which it is expected to vary randomly). Turning to (b), the effects of agent interaction in the IP model have to date been investigated by comparing snapshots at two points in the simulation run: before and after agent interaction (i.e. in Harrington & Schiel 2017; Harrington et al. 2019 and Stevens et al. 2019). However, to distinguish the effects of sound change from random variation it is helpful to consider what happens *during* agent interaction i.e., over the course of the entire run. This would be true for any language change attributed to drift, which, as noted earlier, is understood to involve incremental shifts in a particular direction. It is especially important in the case of sound change because, since phonetic biases are directional, then their accumulation into sound change is expected to involve incremental shifts in the same direction (i.e., in the same direction as the pre-existing phonetic bias) over the course of the interaction. Such directional shifts can be distinguished from acoustic fluctuations that are expected as individual agents interact with each other and absorb new samples into memory.

## 2. Experiment 1: Verifying the coarticulatory path to sound change
### 2.1 Input data: Speakers, materials, acoustic signals

Speaker participants and materials for Italian were chosen in order to match as closely as possible the Australian English dataset analyzed in Stevens et al. (2019). Twenty adult first-language Italian speakers (13 female, 7 male) were recruited in Arezzo, Tuscany, which lies about 80 kilometres south-east of Florence (as noted in section 1.4, it was important to recruit participants from the same town in order to be able to attribute any phonetic differences between participants to idiosyncratic rather than regional/dialectal variation). Standard Italian is learned as the native language in Arezzo, as in other parts of Central Italy. That is, Tuscan speakers are not known to code-switch between a local and a more standard variety (Calamai 2017), which would otherwise complicate data collection and analysis. However, contemporary spoken standard Italian is strongly influenced by regional features and, as Crocco (2017: 93) points out, "Italians, unless specifically trained, speak their language with a more or less marked regional accent". Standard Italian as spoken in Arezzo is likely to show pronunciation features typical of Tuscany such as the lenition of affricate /tʃ/ to [ʃ] and the spirantization of intervocalic stops (e.g. Giannelli 2000; Bertinetto & Loporcaro 2005). Aged between 26 and 50 years, all participants were born, raised, and living in Arezzo at the time of recording and reported having none or only very minimal knowledge of a language other than Italian (e.g. a couple of years of French or English at school). Thus, while there are differences between the two participant groups (the Arezzo participants were not, to our knowledge, known to each other) in terms of their age range, sex, and linguistic experience, the participant group was a close match to the English-speaking participants in Stevens et al. (2019). Speakers were paid for their participation. Their task was to read isolated words aloud as they appeared one after the other on a computer

screen. The set of Italian words, listed in **Table 1**, was matched as closely as possible to the English dataset in terms of the target sibilant's proximity to lexical stress, syllable position and the identity of the surrounding vowels.

| <s>-words | *così, possibile, assenza, principessa, Massimo, messa, sana, sede, sole, passeggiata, enfasi* |
|---|---|
| <str>-words | *mostrò, dimostrato, minestrone, maestrale, estremo, astratto, l'inchiostro, l'apostrofo, finestra, maestro, destra, nostro, strega, strofa, strada, claustrofobia, l'astronave, costruire, istruttore* |
| <ʃ>-words | *uscì, l'ambasciata, lasciare, crescere, liscio, scienza, scena, sciolto, asciugamano, discendenza* |

**Table 1:** Italian words containing sibilants selected for the present study. Primary stress falls on the penultimate syllable in all words except *così, mostrò, uscì* (final syllable) and *l'apostrofo, enfasi, crescere, possible* and *Massimo* (antepenultimate syllable).

The main difference between the English and Italian sets of materials was due to the presence of a consonant length contrast in Italian. Ideally, the Italian materials would have contained only singleton sibilants /s, ʃ/ to match the English words, but the post-alveolar sibilant is always phonetically long in Italian (as a so-called 'inherent geminate' e.g. Bertinetto & Loporcaro 2005) and the inclusion of geminates in the Italian materials was therefore inevitable. As such, the target sibilant is phonologically long in all the <ʃ>-words in Table 1 and phonologically short in all <str>-words, while there are differences amongst the <s>-words in terms of the phonological length of the target sibilant. Note that because of these differences in phonological length, the labels *<s>-words, <str>-words* and *<ʃ>-words*, as listed in Table 1, will be used hereafter (rather than e.g. /ʃ/) unless referring specifically to phonemic length distinctions. Consonant length is evident from the orthography, e.g. *assenza* vs. *sana* in Table 1, with a geminate and a singleton, respectively. In intervocalic position, singleton /s/ is predictably realized as a voiced fricative in Italian e.g. *enfasi* /ɛnfazi/. These differences in phonological length and voicing were expected to affect the acoustic properties of sibilants and were removed from the signals before analysis and before input into the ABM with time normalization and high pass filtering, respectively, as described below.

Speech signals were recorded using SpeechRecorder (Draxler & Jänsch 2004) on a laptop computer with participants wearing a headset microphone. The words in Table 1 were presented to participants in a randomized order with ten repetitions, giving a possible 8000 tokens. Due to a recording error, there were 7999 tokens available for acoustic analysis.

The acoustic analysis of the recorded signals followed procedures described in Stevens et al. (2019) and was as follows. The signal data were labeled semi-automatically using WebMAUS (e.g. Kisler et al. 2016) and converted into an EMU database (Winkelmann et al. 2017) for further

processing. Segment boundaries were corrected manually. The signals were time-normalized, so that the duration of all sibilant tokens started at zero and ended at one (cf. e.g. **Figure 5**, x-axis). This meant that the agents did not receive any information about differences in phonetic duration or phonological length of sibilant tokens (especially relevant here because, as shown in Table 1, the Italian dataset contained both singleton and geminate sibilants). The spectral centre of gravity or first spectral moment (M1), which indexes place of articulation, was calculated as a trajectory between the temporal onset and offset of each sibilant token, using the R package emuR (Winkelmann et al. 2017). These M1 trajectories were calculated between 500–15000 Hz, i.e., with high pass filtering to remove all spectral information below 500 Hz, including fundamental frequency. Thus differences between sibilant tokens in terms of voicing (such as between predictably voiced /z/ in *enfasi* vs voiceless /s/ in *strada*) were removed from the signals while spectral information in the higher frequencies that is crucial to distinguishing place of articulation for sibilants was maintained. M1 is an absolute measure and M1 trajectories (between the temporal onset and offset of a sibilant) are easy to interpret, with lower M1 indicating a more posterior pronunciation and higher M1 indicating a more anterior pronunciation. The M1 trajectories were z-score transformed (across all data frames for all sibilants for each speaker) using the speaker-normalization method in Lobanov (1971), to control for inherent differences between sibilants produced by male (lower M1) and female (higher M1) speakers (e.g. Weirich & Simpson 2015). The normalized M1 trajectories for all sibilants were parameterised into three cepstral coefficients using Discrete Cosine Transformation (DCT) (e.g. Harrington 2010), which reduces each of them to a point in a three-dimensional space. The dimensions of the DCT-coefficients DCT-0, DCT-1 and DCT-2 correspond to the height, slope and curvature of each M1 trajectory. The primary purpose of parameterizing the M1 trajectories in this way is that they provide the main input to the agent-based model. The orthogonal projection (*OP*) of a sibilant token's position in this three-dimensional DCT space, on a line passing through speaker-specific centroids for <s>- and <ʃ>-words, was also calculated (see Appendix in Stevens et al. 2019 for mathematical details). *OP* indicates the relative location of any sibilant token between a speaker's <s>-words (centred at +1) and <ʃ>-words (centred at –1) and is thus an acoustic measure of /s/-retraction.

## 2.2 The agent-based model

The present study applied the same ABM architecture as described in Harrington and Schiel (2017) and extended in Stevens et al. (2019), with exactly the same parameter settings as in the latter study. The basic structure was as follows: one agent represented each speaker participant, with its own production data (sibilant tokens) stored in memory. Each sibilant token consisted of the word label (e.g. *sole*) and the three DCT coefficients for the height, slope and curvature of the M1 trajectory for the sibilant, e.g. the sibilant /s/ in *sole*. The agents interacted with each other by exchanging sibilant tokens. For each interaction, two agents were chosen at random,

one who is the agent-talker and one who is the agent-listener. The agent-talker chose a word at random (e.g. *sole*) and generated a sibilant token by sampling a Gaussian model based on all the sibilant tokens for that word currently stored in their own memory. The agent-listener received the sibilant token and decided whether to store it in memory. The procedure for this was based on passing a test of maximum posterior probability in which the agent listener established the sub-phonemic class (see below) with which the perceived sibilant in *sole* (for this example) was affiliated. The incoming signal was then added to that class, if the probability of membership to it was greater than to any other of the agent-listener's sub-phonemic classes.

The present study is concerned with the effects of agent interaction on the location of sibilants in acoustic space, i.e. their relative proximity to /s/ or /ʃ/. Nonetheless, it is important to note that the sound change involving /s/-retraction of course involves phonological reorganization: the sibilant in *street*, for example, is no longer categorized as the sibilant in *seat* but rather that in *sheet*. An integral part of the ABM architecture applied in this study is that, consistent with usage-based models of sound change, the association between words and phonological classes can vary between individuals and over time (i.e. over simulated interactions between agents). This is because the IP model allows the possibility of phonological re-categorization within individual agents before interaction and after the absorption of new acoustic signals (i.e. during interaction). This requires phonological classes to be defined as *sub-phonemic*. Following the terminology of Stevens et al. (2019), all agents initially shared the same phonology, which consists of two phonological classes: **S** (all < ʃ >-words) and **s + str** (all < s >- and < str >-words). At initialization, and prior to any interaction, merge and split algorithms were repeatedly applied to the sibilant tokens of each agent separately to create a number of sub-phonemic classes. Hypothetically, a sub-phonemic class $str_1$ may comprise *finestra, strega* and *destra* (some < str >-words only), $str_2$ may comprise *dimostrato, maestrale* and *mostrò* (other < str >-words only), while $s + str_1$ may comprise *così, possibile* and *costruire* (a mixture of < s >- and < str >-words), and **S** may comprise all < ʃ >-words. Thus these sub-phonemic classes can comprise any (and at least two) of the words stored in an agent's memory but within any agent all tokens for a particular word, e.g. *sole*, must belong to the same sub-phonemic class. Note that the subscript numbers in this hypothetical example indicate sub-phonemic classes of the same type, in the sense that $str_1$ and $str_2$ both contain only < str >-words but differ in terms of the exact word items. Thus there are essentially seven possible types of sub-phonemic class: **s, str, S, s + str, s + S, str + S** and **s + str + S** (which were termed *equivalence classes* in Stevens et al. 2019). A merge between any pair of sub-phonemic classes (within any agent) was made if the (posterior) probability of class membership to each of these classes separately was less than to the union of these classes. Conversely, a split occurred if a given sub-phonemic class, *P*, could be divided into two classes, $P_1$ and $P_2$, such that the probability of class membership to $P_1$ and to $P_2$ was greater than to the original class *P* (see Stevens et al. 2019: 11–12 for detail on these algorithms). During

interaction, sub-phonemic classes were used for deciding whether or not an agent memorized a (new) perceived sibilant token, and the test of whether any pair of sub-phonemic classes should be split or merged (within any agent) was applied to all agents after every 100 pairwise interactions. Further details are given in Stevens et al (2019). Abstracting from these details, the critical point here is that agents and by implication their corresponding real speakers are assumed to differ slightly in how the association between words and sibilant tokens is categorized phonologically, and moreover that this phonological categorization can be updated dynamically as a consequence of interaction. Such (typically slight) differences in phonological categorization even between agents of the same dialect and language are therefore considered in the IP model to be one of the factors that can contribute to change. As noted above, for practical reasons the present study is confined to the acoustic effects of agent interaction on sibilants (but see Stevens et al. 2019 for phonological effects of agent interaction on sibilants) and instead expands the scope of the analysis in new ways to address differences between multiple (repeated) simulation runs and over time (i.e. as agent interaction unfolds).

The ABM was run with the parameter settings from Stevens et al. (2019), most notably:

- Memory loss: the absorption of a new token into an agent-listener's memory was accompanied by removal of the oldest existing token in the same word class (and the same agent). The oldest existing token was defined as the one with the lowest time-stamp value (Harrington & Schiel 2017: 433–434). Time-stamp values were assigned randomly to each sibilant token (per agent and per word class) before running the ABM, and during interaction each new sibilant token to be absorbed into an agent-listener's memory was assigned a time-stamp value of the maximum (of any of the agent's stored tokens for the same word class) plus one.

- There was the potential for splitting and merging of phonological (sub-phonemic) classes, as described above.

- Each run consisted of 60,000 pairwise agent interactions and there were 100 separate runs. Results were aggregated over the 100 (repeated) runs to avoid reporting an outlier result, but the outcomes of individual runs are also addressed in section 2.4.

## 2.3 Quantification of the effects of agent interaction on sibilants

The effect of agent interaction on sibilants in <str>-words was tested in terms of their location in acoustic space, which was in the case of Italian predicted to remain unchanged after agent interaction. To do so, <str>-words were compared before and after agent interaction in terms of two acoustic parameters: the first spectral moment (M1) and the *orthogonal projection* (*OP*), which is a relative rather than an absolute measure and was defined earlier in section 2.1. M1 and

OP can be interpreted similarly and are, for the present purposes, essentially interchangeable: lowering of M1 and lowering of *OP* can both be interpreted as increased proximity to /ʃ/.

## 2.4 Results

First the baseline data (i.e. before agent interaction) were examined. **Figure 2** illustrates a sibilant in Italian in pre-vocalic position (below) and in /str/ (top). Both sibilants show a concentration of energy in the highest regions of the spectrogram (above 7000 Hz), which indicates a very anterior pronunciation, and there is little visible differentiation between the sibilant in /str/ vs. prevocalic context. Quantitative evidence of the lack of contextual effects on sibilants in Italian /str/ can be seen in **Figure 3**.
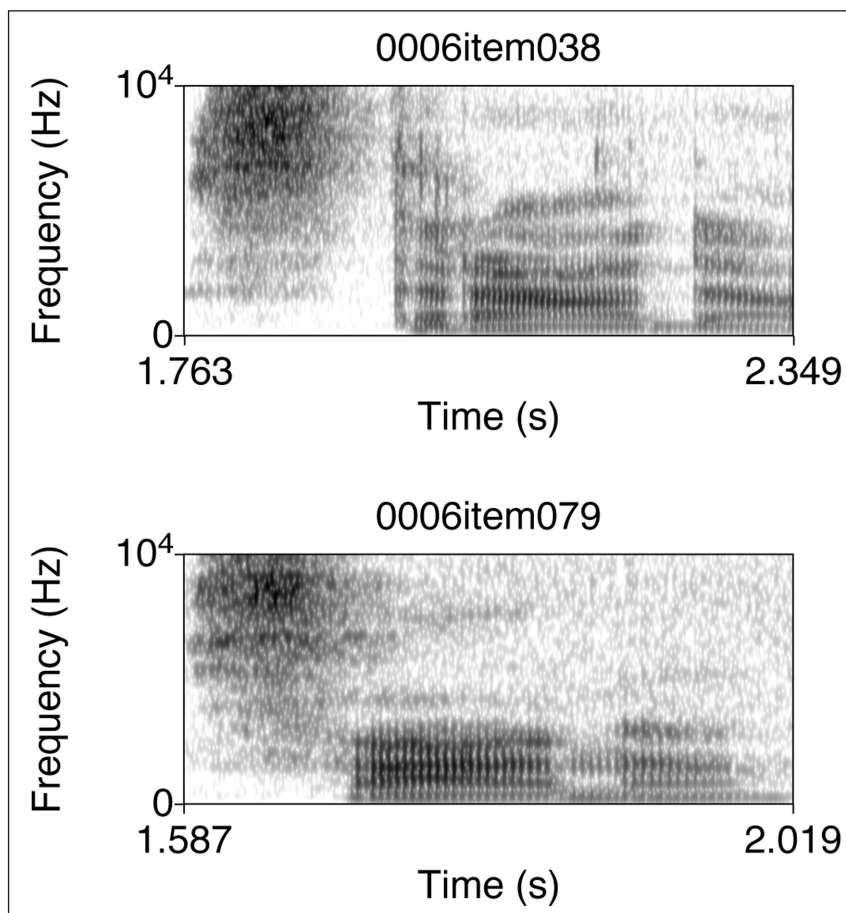


**Figure 2:** Spectrograms of word-initial /s/ in the cluster /str/ and in pre-vocalic position, produced in *strada* 'street' (above) and *sana* 'sane' (below) by a female participant speaker.
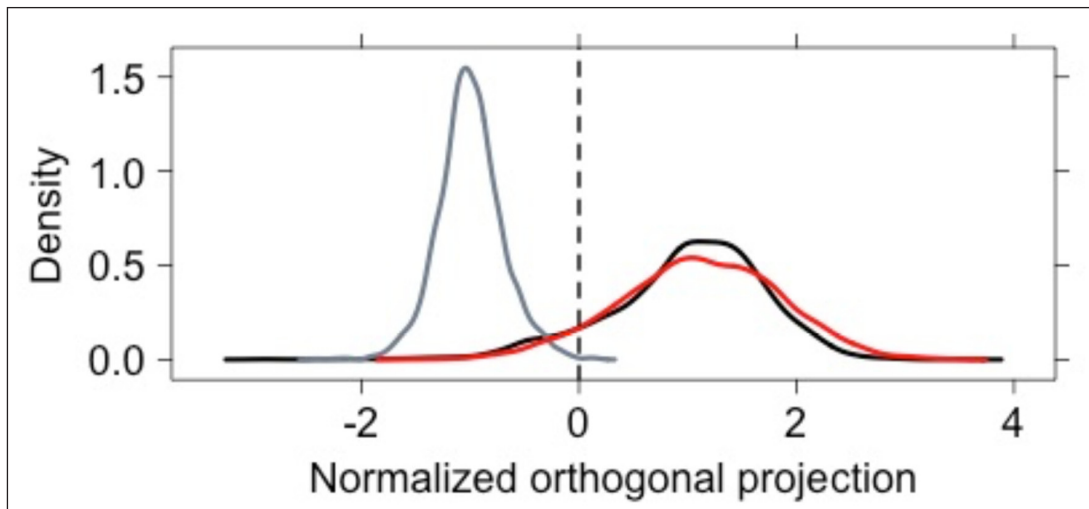
**Figure 3:** Distribution of Italian sibilants in acoustic space at baseline (before agent interaction), with samples from <ʃ>-words in grey, <s>-words in black and <str>-words in red, aggregated over 20 speaker participants.

**Figure 3** shows the baseline distributions in acoustic space of samples from Italian <s>-words in black, <str>-words in red and <ʃ>-words in grey, aggregated across the 20 speaker participants. Here the x-axis again shows the *OP* as in **Figure 1**, which measures the relative location of sibilants between speaker-specific <s>- and <ʃ>- centroids in a three-dimensional acoustic space (and see section 2.1). Note the close overlap between the black and red distributions; there is no evidence of a leftwards shift in <str> (compared to <s>) that would indicate increased proximity to <ʃ> i.e. a pre-existing tendency towards /s/-retraction in <str>-word contexts.

Close overlap between the distributions of samples from <s>- and <str>-words at baseline was also evident within most individual speakers, as shown in **Figure 4**. Differences between <s>- and <str>-distributions were visible in some speakers as a rightwards shift in <str>-words (e.g. S08, S14, S18) and for other speakers as a leftwards shift in <str>-words (e.g. S12, S15, S22). Thus, there does not appear to be any evidence of a population-level tendency towards /s/-retraction in <str>-words in these baseline data, and patterns within individual speakers suggest random variation in both directions as to whether sibilants in <str>-words are (slightly) more anterior or posterior than those in <s>-words. For this reason, these baseline data are suitable to test the prediction that (without a pre-existing shared bias towards /s/-retraction in <str>), there should be no shift in the direction /s/ -> /ʃ/ in <str> as the speaker agents interact with each other.

It is important to note the individual variation in the sibilant data in **Figure 4**: some speakers showed a clear distinction between the distribution of samples from <ʃ>-words versus those from <s>- and <str>-words (e.g. S19, S20), whereas other speakers (e.g. S06, S07, S08) showed much overlap between the distribution of phonemically posterior vs. anterior sibilants. In addition, distributions of <s>- and <str>-words were bimodal or multimodal for some speaker participants (e.g. S05, S06) but unimodal for others, evidence of individual differences in the degree of within-category variability. These differences are not unexpected for sibilants which are known to show idiosyncratic variation (as noted in section 1.4) and can be attributed to differences in vocal tract geometry (e.g. Yoshinaga et al. 2019). The important point is that these Italian baseline data are broadly comparable with the Australian English dataset in terms of idiosyncratic sibilant productions. The substantial difference between the two datasets concerns the presence (Australian English) or absence (Italian) of a pre-existing phonetic bias towards /s/-retraction in /str/.

Turning from the baseline data to the effects of agent interaction, **Figure 5** shows M1 trajectories for samples from Italian <str>-words before (solid) and after agent interaction (dashed), aggregated over all 20 participants. It is evident that M1 is higher after agent interaction towards the temporal offset (dashed trajectory), which is indicative of a more anterior pronunciation. However, M1 was otherwise slightly lower in the dashed trajectory and mean M1 hardly differed between the two (0.395 at baseline and 0.402 after agent interaction). Thus the trajectories in **Figure 5** differ primarily in terms of shape (slope), not height; there is no evidence of M1-lowering after agent interaction that would indicate increased proximity to <ʃ>-words.
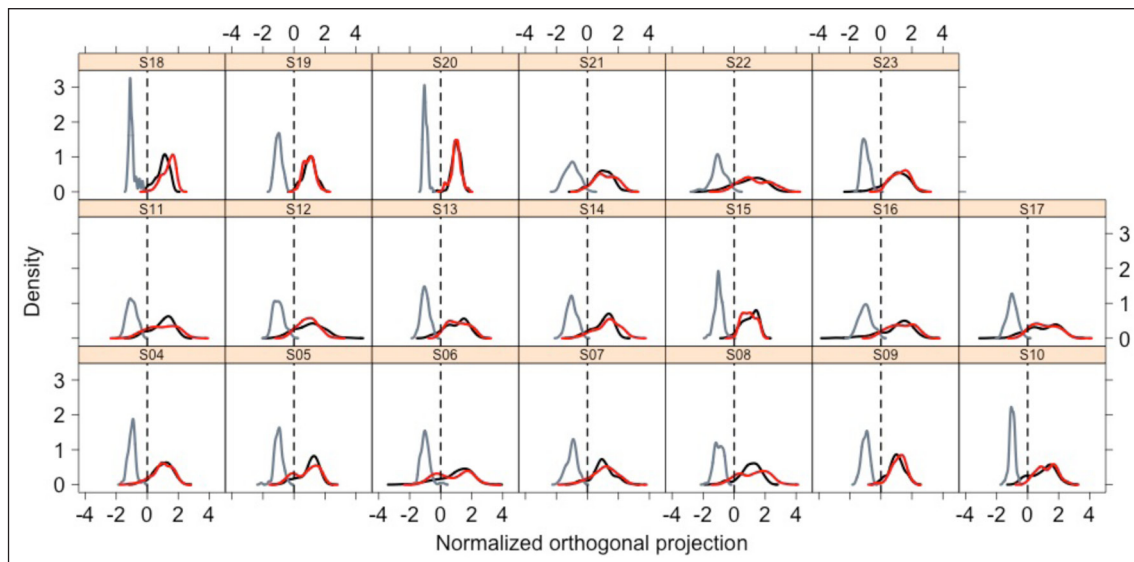


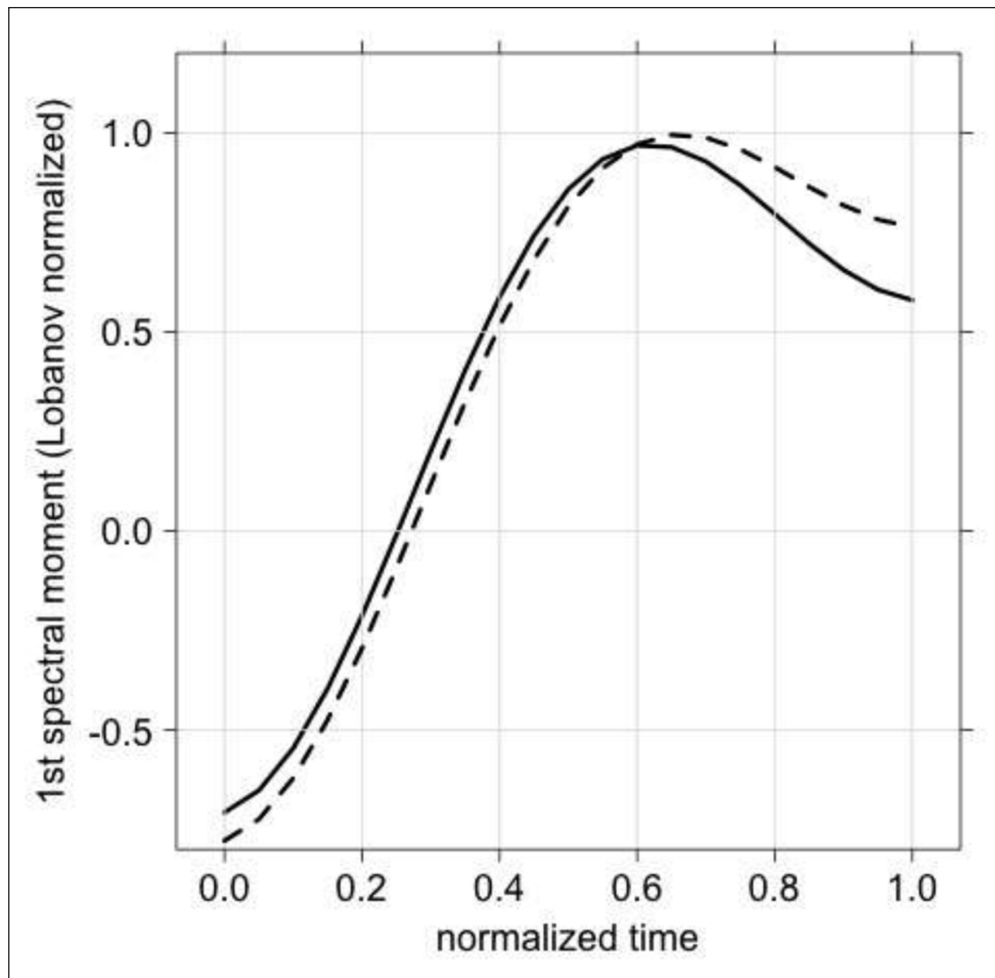**Figure 4:** The data from Figure 3, by individual speaker.

**Figure 5:** DCT-smoothed M1 trajectories aggregated over sibilants in <str>-words in the baseline (solid) and after simulated interaction between agents (dashed).

The effect of agent interaction on the distribution of <str>-words in acoustic space is shown in the combined boxplot and violin density plot in **Figure 6**. Here the measurement parameter was *OP* (Section 2.1), which incorporates information about the height, slope and curvature of M1 trajectories (i.e. the trajectories that were aggregated in **Figure 5**, above). Results can be interpreted similarly to those concerning M1 in **Figure 5**: lower (*OP*) values indicate increased proximity to <ʃ>-words. The shape of the distribution changed to multimodal after agent interaction but there was otherwise little visible difference between the two distributions. In particular, neither the entire distribution nor the median shifted visibly along the y-axis, i.e. agent interaction had no apparent effect on location in acoustic space of <str>-words relative to <s> or to <ʃ>-words.
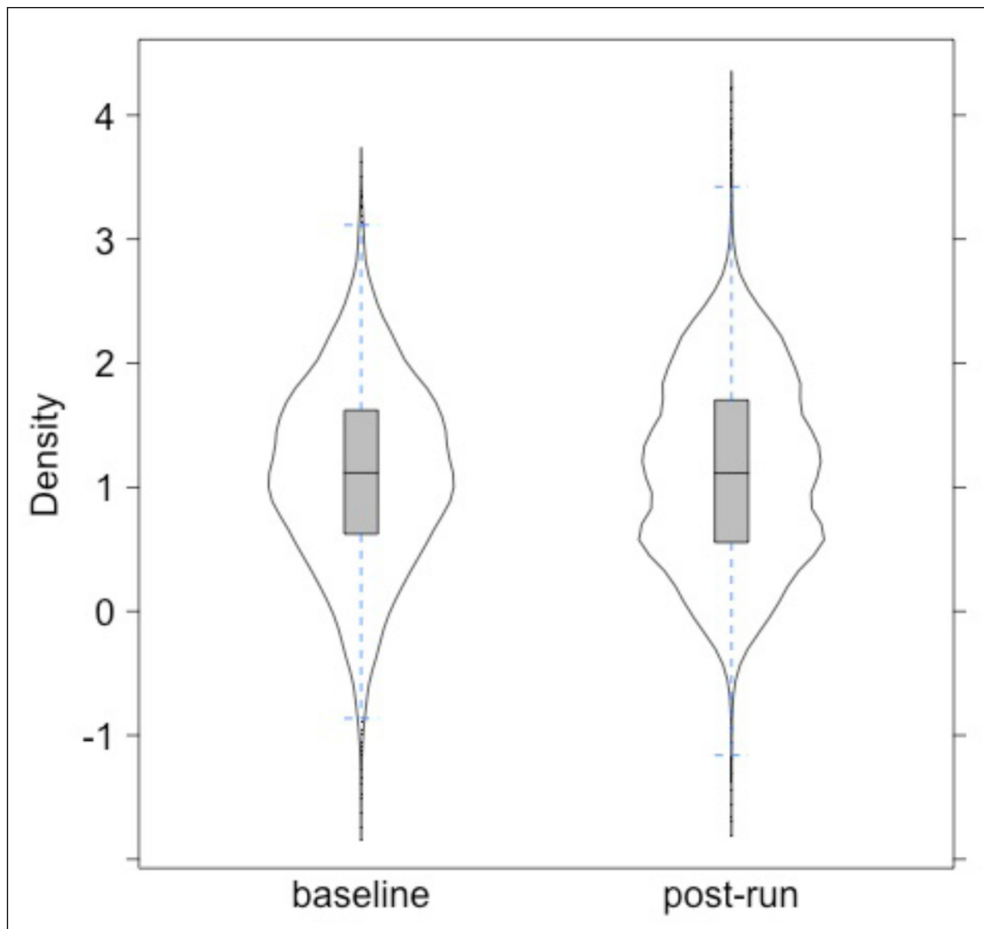
**Figure 6:** Normalized orthogonal projections *op* of sibilants in Italian <str>-words in the baseline (left) and post-run i.e. after agent interaction (right). Negative values are closer to <ʃ> and positive values are closer to <s>.

A linear mixed effects analysis was carried out to test the effect of agent interaction on sibilants in Italian <str>-words, using the same model that was applied to the Australian English dataset in Stevens et al. (2019). Here again, the dependent variable was the normalized orthogonal projection *OP*, there was one fixed factor Condition (two levels, baseline vs. post-run) and three random factors (Word, Agent, Interaction Run), and p-values were obtained by likelihood ratio tests of the full model with factor Condition against the model without factor Condition (a confound between Run and Condition was addressed by including a term that declared individual runs, but not the baseline, as a random factor; see Stevens et al. 2019 for detail on the statistical model). The results showed a significant effect for Condition (Chi Square = 13.8, p < 0.001). This was unexpected given the similarity of the two distributions in

**Figure 6**, and it must be noted that the magnitude of the effect of Condition on <str>-words was very small (*d*Cohen 0.027). The linear mixed effects analysis found a significant effect of Condition possibly as a result of differences in trajectory shape (as opposed to trajectory height i.e. proximity to <ʃ>-words) that were visible in **Figure 5**. The model would have been sensitive to differences in trajectory shape because, as noted above, the dependent variable (*OP*) contained information about the height (DCT-0), slope (DCT-1) and curvature (DCT-2) of M1 trajectories. The significant effect of Condition may also have been due to differences in distribution shape (cf. **Figure 6**) and sample size at baseline vs. post-run. The sample size differed due to the aggregation of the post-run data over 100 separate simulation runs (*n* baseline = 3799 vs. *n* post-run 379929).

Although the stochastic nature of the model means that results for any single run must be interpreted with caution, as noted earlier it is also the case that the extent of variation between individual (repeated) simulation runs can help to differentiate directional sound change from random variation. **Figure 7** shows the location in acoustic space of sibilants in Italian <str>-words in each individual simulation run. Here the acoustic parameter is the first DCT coefficient, DCT-0. This parameter was chosen, rather than M1 or *OP,* for purely practical reasons. The acoustic information exchanged by agents comprised three DCT coefficients and the first of these, DCT-0, is proportional to the height of an M1-trajectory and is therefore most relevant in terms of the location of <str>-words in acoustic space. Thus, it was most efficient to extract DCT-0 directly from the output of the ABM for this part of the acoustic analysis, which concerns multiple time points (over 60,000 agent interactions) and multiple simulation runs. DCT-0 is nonetheless closely related to the two other measurement parameters and can be interpreted in the same manner: lower DCT-0 (or *OP* or M1) indicates a more posterior pronunciation and higher DCT-0 indicates a more anterior pronunciation.

**Figure 7** shows that the effects of agent interaction on sibilants in Italian <str>-words varied over repeated simulation runs. More specifically, DCT-0 increased during some runs and decreased during others, indicating population-level shifts towards a more anterior or posterior pronunciation, respectively, while many tracks showed little visible change in DCT-0 from the baseline. Moreover, it is also evident from **Figure 7** that the direction of any population-level shifts also changed course during some individual runs. All this suggests that the effect of agent interaction on DCT-0 of sibilants in <str>-words was essentially random, involving slight acoustic fluctuations around the mean. (The acoustic differences can be considered slight because, even in the simulation run with the biggest decrease in DCT-0, in which it dropped to 0.51, sibilants in <str>-words would not have encroached on <ʃ>-words for which mean DCT-0 was –1.29 at baseline).
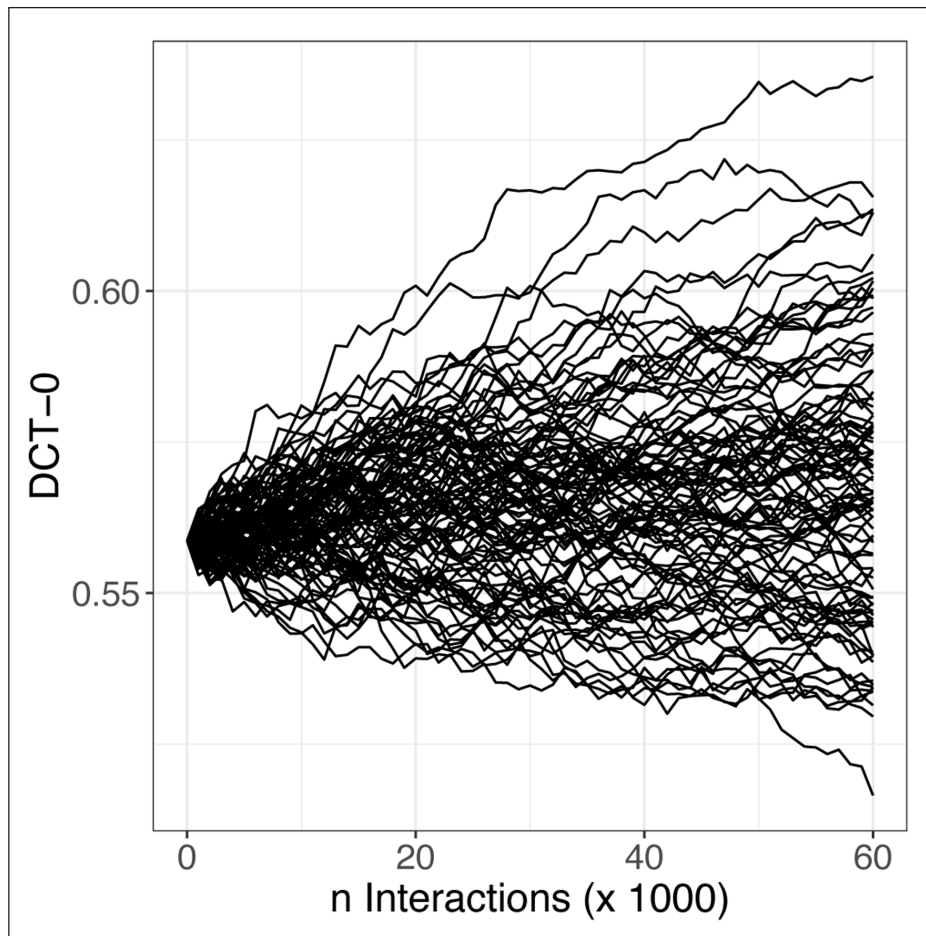
**Figure 7:** Mean DCT-0 in sibilants in Italian <str>-words during agent interaction. Each track shows mean DCT-0 in one simulation run, aggregated across 20 agents and (multiple repetitions of) the 19 <str>-words listed in Table 1. DCT-0 is proportional to the mean of the (DCT-smoothed, time-normalized) M1 trajectory between the temporal onset and offset of the sibilant. Lower DCT-0 values indicate a more posterior pronunciation and higher values indicate a more anterior pronunciation.

**Figure 8** shows comparable data from 100 simulation runs with baseline data from Australian English, in which there was a pre-existing tendency towards /s/-retraction in <str>-words, as outlined in Section 1.4. It is immediately apparent in **Figure 10** that the effect of agent interaction on sibilants in Australian English <str>-words involved incremental lowering of DCT-0 and that this was the case in all simulation runs. Lowering of DCT-0 indicates a more posterior pronunciation i.e. a further shift in the direction of the phonetic bias towards /s/-retraction that was already present at the baseline in the Australian English data.

Thus sibilants in Australian English <str>-words predictably undergo change at the population-level, whereas the effects of agent interaction on Italian <str>-words are small and

random. The reason for the different outcomes can be attributed to the difference between the two sets of starting conditions, which concerned the presence (Australian English) or absence (Italian) of a coarticulatory tendency towards /s/-retraction. In section 3 we address the extent to which the population-level change that was evident in Australian English <str>-words may have been driven by certain` individuals.
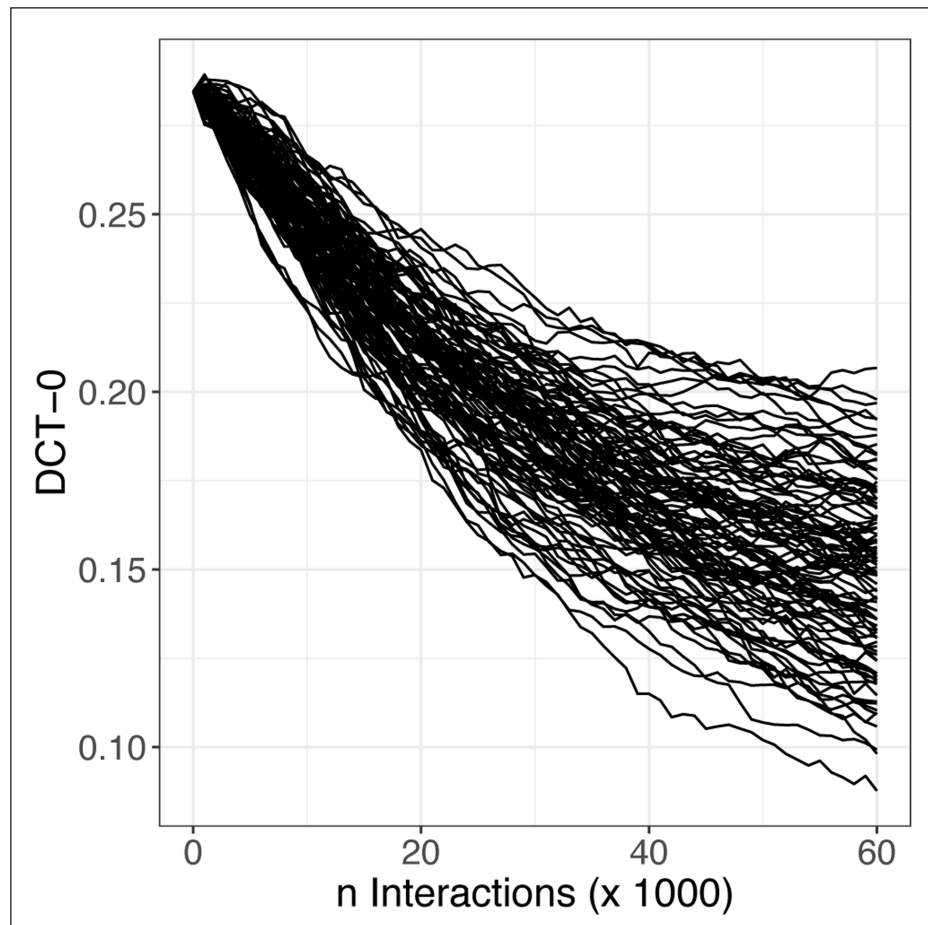


**Figure 8:** Mean DCT-0 in sibilants in Australian English <str>-words during 100 simulation runs. Each track shows data from one simulation run, aggregated over 19 agents and (multiple repetitions of) 19 <str>-words. The data were taken from the study reported in Stevens et al. (2019), in which only aggregated data (over all 100 simulation runs) were reported.

## 3. Experiment 2: Individual variation and population-level change

This experiment is concerned with the effects of interaction on individual agents. It also tests whether population-level shifts in the direction of sound change are dependent on innovative individuals.

## 3.1 Input data: Speakers, materials, acoustic signals

This section draws on two sets of input data, namely the Italian data from section 2 and the Australian English dataset first analyzed in Stevens et al. (2019).

## 3.2 The agent-based model

The model was run under three different starting conditions but with the same parameter settings in all cases, as described in Stevens et al. (2019) and in section 2.2. The starting conditions were as follows: (a) the entire population of Italian speaker agents; (b) the entire population of Australian English speaker agents; and (c) a reduced population of Australian English speaker agents from which especially innovative agents were removed.

## 3.3 Results

**Figure 9** shows mean DCT-0 in sibilants in <str>-words during simulated interaction among Italian-speaking agents (left) and among Australian English-speaking agents (right). Each track shows the data for one speaker agent averaged over 100 simulation runs.
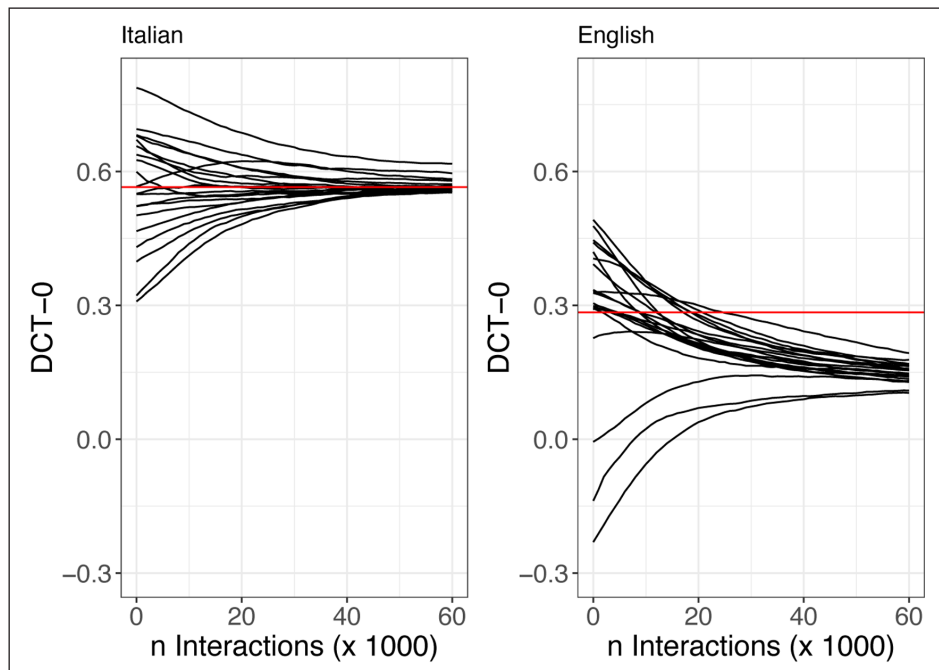


**Figure 9:** Mean DCT-0 in sibilants in <str>-words during 100 simulation runs with Italian (left) and (Australian) English (right). Each track shows data from one speaker agent, aggregated over all 100 simulation runs, <str>-words and repetitions. The red line shows the average across all speakers at baseline (*n* Interactions = 0). The Australian English data were taken from the study reported in Stevens et al. (2019).

Although somewhat greater in Australian English, **Figure 9** shows that there was variation between the individual members of both populations in terms of the acoustic properties of sibilants in <str>-words at baseline (i.e., $n$ interactions = 0). Over the course of the interaction, the Italian agents (left-hand panel) more or less converged towards the average i.e. the red line that falls in the centre of the band of Italian speakers at baseline. Thus no individual Italian speaker appeared to influence the population-level outcome of interaction more than any other. In Australian English, on the other hand, most speakers were above the red line at baseline, yet all were below this line after 60,000 interactions. Thus the population-level norm for Australian English <str>-words shifted downwards towards lower DCT-0 values and indeed towards a minority of speakers (the three "innovative" speakers for whom DCT-0 was exceptionally low at baseline, using "innovative" in the sense of Beddor et al. 2018 to refer to an individual at the extreme end of a continuum of coarticulatory variation in a population). Because the Australian English baseline data showed both innovative speakers (the three with especially low DCT-0 in <str>) and coarticulatory variation (all speakers shared a tendency towards lower DCT-0 in <str>-words), it is unclear which of these factors caused the population-level downwards shift for <str>-words that was observed during interaction. To disentangle the effects of these two factors, we re-ran the Australian English simulations under exactly the same conditions but without baseline data from the three "innovative" speakers (i.e. those with especially low DCT-0 in <str>-words at baseline). If the downwards shift in the original population was due to the influence of the three speakers with especially low DCT-0 at baseline, then there should be no such shift when these speakers are removed from the population. On the other hand, if the downwards shift in the original population was primarily due to the accumulation of the shared tendency towards lower DCT-0 in <str>-words, then a downwards shift should still be observed during interaction between the remaining speakers.

**Figure 10** shows that DCT-0 was lower for all sixteen speakers after interaction than at baseline. Thus the same downwards shift was observed in Australian English whether the population included the especially "innovative" speakers – for whom DCT-0 was already low at baseline (**Figure 9**) – or not (**Figure 10**). This result supports the idea that population-level shifts of the kind observed for <str>-words in the original study with Australian English originate in shared tendencies which can accumulate during interaction and need not depend on the influence of especially innovative speakers. The magnitude of the population-level shift was smaller for the reduced population (averaged across all agents, DCT-0 dropped to 0.189 after interaction in **Figure 10** but down to 0.149 after interaction in **Figure 9**, right-hand panel). This suggests that especially innovative agents may accelerate directional shifts but that such shifts are still likely without their influence on the group of interacting agents.
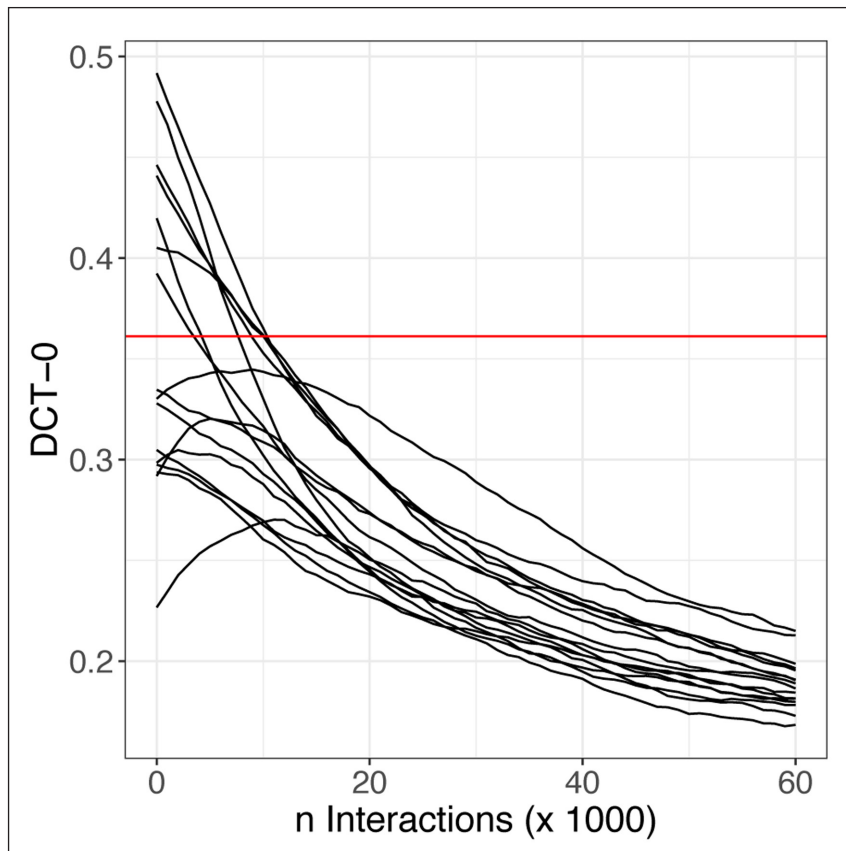
**Figure 10:** Mean DCT-0 in sibilants in <str>-words during 100 simulation runs with 16 (rather than 19) Australian English speakers (see text). Each track shows data from one speaker agent, aggregated over all 100 simulation runs, <str>-words and repetitions. The red line shows the average across all speakers at baseline (*n* Interactions = 0).

## 4. Discussion

This study applied computational modeling to explore sound change in single populations (i.e. in the absence of language contact) and the role of coarticulatory variation, individual variation, and interaction between speakers in this process. Results for Italian supported the two-fold prediction that there would be no evidence of a phonetic bias in the direction /s/ -> /ʃ/ at baseline and population-level stability despite simulated agent interaction. Close overlap was found between the distribution in acoustic space of samples from Italian <str>- and <s>-words at baseline (**Figure 3**), which showed that there was no pre-existing tendency towards /s/-retraction in <str>-words. This lack of /s/-retraction in Italian <str>-words was attributed to the tapped/trilled pronunciation of the rhotic, which does not exert the sorts of anticipatory effects associated with the English alveolar/retroflex approximant (West 1999). Although agent interaction produced/caused differences in the shape of M1 trajectories, there were no changes to their mean

and therefore none to the sibilant's place of articulation (this being closely associated with M1 mean frequency). Agent interaction was expected to affect the acoustic properties of sibilants to some degree given the stochastic nature of the model and the idiosyncratic production patterns of the speaker-agents who interacted with each other. However, the place of articulation of sibilants in <str>-words was found to show only small random acoustic fluctuations within and between simulation runs. Such stability for Italian <str>-words despite agent interaction is not surprising: the idea that sound change originates in coarticulatory variation is certainly not new (eg Ohala 1989; Kiparsky 2003; Lindblom et al. 1995; Beddor 2009; Garrett & Johnson 2013), and modeling of exemplar dynamics (Pierrehumbert 2001) has shown evidence that only systematic production biases (and not random variation) can lead to directional shifts. However the coarticulatory path to sound change has never been tested and verified with a study of this sort and the contributions of the present study were twofold. First, stability was the outcome of simulated interaction, rather than a theoretical prediction. Second, stability was maintained despite agents exchanging (parameterized) speech signals produced by actual speakers. Thus, whereas simulations of sound change typically use constructed representations, the stimuli in the present study included the kinds of phonetic variation that are present in any population and could have conceivably threatened the population-level stability of speech sounds over time.

In contrast to the results for sibilants in Italian <str>-words, which essentially fluctuated around the baseline mean during agent interaction (**Figure 9**), a directional shift /s/ -> /ʃ/ was found for <str>-words in every (repeated) simulation run involving English-speaking agents (**Figure 10**). Because all other factors were kept constant between the two datasets, the language-specific outcomes were attributed to the presence (Australian English) or absence (Italian) of a phonetic bias in <str>-words towards /ʃ/ in the baseline data. Results were also compatible with observed language-specific patterns: /s/-retraction in <str>-words is common in many varieties of English (e.g. Baker et al. 2011; Kraljic et al. 2008: 56), but not in varieties of Italian. The mechanism by which coarticulatory variation accumulates in the IP model is consistent with usage-based models (e.g. Beckner et al. 2009) in which change is incremental as listeners' representations are updated over time, rather than involving an abrupt change due to listener error (e.g. Ohala 1993).

The second part of the present study turned to the role of idiosyncratic variation in the earliest stages of sound change. What sorts of variation cause population norms to change? Must a phonetic bias be found in the speech of all members of the population in order for the bias to accumulate as these speakers interact? Can one idiosyncratic speaker cause population-level norms to shift? As outlined in the introduction, sources differ as to whether sound changes should be attributed to the accumulation of a shared phonetic bias (e.g. Bloomfield 1933; Paul 1888) or to a minority of individuals who deviate from the population norm (e.g. Baker et al. 2011; Francois 2014; Garrett & Johnson 2013; Samuels 1972; Yu & Zellou 2019; Yu 2021). The

primary aim of the present study was to disentangle these two different sources of variation, both of which were present in the baseline data for the earlier study on Australian English <str>-words. The Italian baseline data showed no coarticulatory variation in <str>-words but variation between individual speaker participants (in the sense that they were slightly more anterior than in <s>-words for some speakers and slightly more posterior for others). After interaction between the Italian speaker-agents, outcomes mirrored the random nature of the baseline variation: sibilants in <str>-words were more anterior after agent interaction in some simulation runs and more posterior in others. Thus the effect of agent interaction varied between simulation runs depending on localized interactions but no individual Italian agent-speaker consistently influenced the population-level norm. The lack of influence of any one individual in the Italian data was most evident in **Figure 9** (left panel), in which, over the course of the interaction, individual speakers' <str>-word trajectories were seen to converge to the population mean at baseline. The same was not true for Australian English in which the population-level mean for <str>-words showed a downwards shift, approximating a small number of speakers for whom values were already relatively low at baseline. Nonetheless, the downwards shift was observed whether the population of interacting Australian English speakers included especially innovative speakers (**Figure 9**) or not (**Figure 10**). In other words, the population-level change that was observed for Australian English <str>-words could not be attributed to the influence of especially innovative speakers. This result conflicts with sources on language change, that, as outlined in section 1.1, have tended to highlight the role of idiosyncratic behaviour in shaping languages. It also conflicts with the idea that sound change, in particular, is dependent on individual variation, which was outlined in section 1.3 and which was developed most fully in Baker et al. (2011). Instead, the results for Australian English sibilants in /str/ support the idea that sound change involves the incremental accumulation of directional phonetic biases that are common, rather than exceptional, to members of a particular population (e.g. Paul 1888:52; Bybee & Easterday 2019:268). Because all members of the Australian English population showed at least some degree of /s/-retraction in /str/, the question of exactly *how* common a phonetic bias must be amongst interacting members of a population in order for it to accumulate must be left for future investigations (which could simulate interaction amongst populations in which phonetic biases are present in some but not all members).

In the IP model, the listener chooses whether to memorize an incoming token based on whether it fits their own distributions of memorized samples rather than their attitudes to the speaker or to the incoming pronunciation (e.g. Drager 2010). Of course, it may well be that listener choice and social factors do play some role in sound change, perhaps at a later stage when listeners begin to copy (potentially idiosyncratic) pronunciations that they associate with social meaning (Labov 1963; Milroy and Milroy 1985; Trudgill 1986). There is evidence that social network structure, in particular, plays an important role in the propagation of novel

variants (Dodsworth 2019; Blythe & Croft 2021) and the IP model could be fruitfully extended to include information about social ties and thus the frequency with which individual members of a population interact. Nonetheless, the present study verifies that the earliest stages of a sound change can be modeled as the accumulation of coarticulatory variation via interaction, without relying on additional factors such as an association with a social category or individual speakers who deviate from population-level norms.

## Acknowledgements

## Competing Interests

The authors have no competing interests to declare.

## References

Baker, Adam & Archangeli, Diana & Mielke, Jeff. 2011. Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change* 23. 347–74. DOI: https://doi.org/10.1017/S0954394511000135

Beckner, Clay & Blythe, Richard & Bybee, Joan & Christiansen, Morten H. & Croft, William Ellis, Nick C. & Holland, John & Ke, Jinyun & Larsen-Freeman, Diane & Schoenemann, Tom. 2009. Language is a complex adaptive system: position paper. *Language learning* 59. 1–26. DOI: https://doi.org/10.1111/j.1467-9922.2009.00533.x

Beddor, Patrice Speetor. 2009. A coarticulatory path to sound change. *Language* 85. 407–28. DOI: https://doi.org/10.1353/lan.0.0165

Beddor, Patrice Speetor & Coetzee, Andries W. & Styler, Will & McGowan, Kevin B. & Boland, Julie E. 2018. The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. *Language* 94. 931–68. DOI: https://doi.org/10.1353/lan.2018.0051

Bertinetto, Pier Marco & Loporcaro, Michele. 2005. The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of the International Phonetic Association* 35. 131–51. DOI: https://doi.org/10.1017/S0025100305002148

Blevins, Juliette & Wedel, Andrew. 2009. Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica* 26. 143–83. DOI: https://doi.org/10.1075/dia.26.2.01ble

Bloomfield, Leonard. 1933. *Language* London: George Allen & Unwin.

Blythe, Richard A. & Croft, William. 2021. How individuals change language. *PLoS ONE* 16(6): e0252582. DOI: https://doi.org/10.1371/journal.pone.0252582

Bybee, Joan. 2012. Patterns of lexical diffusion and articulatory motivation for sound change. In Solé, Maria-Josep & Recasens, Daniel (eds.), *The initiation of sound change. Perception, production and social factors*. Amsterdam/Philadelphia: John Benjamins. 211–34. DOI: https://doi.org/10.1075/cilt.323.16byb

Bybee, Joan & Easterday, Shelece. 2019. Consonant strengthening: A crosslinguistic survey and articulatory proposal. *Linguistic Typology* 23. 263–302. DOI: https://doi.org/10.1515/lingty-2019-0015

Calamai, Silvia. 2017. Tuscan between standard and vernacular: a sociophonetic perspective. In Cerruti, Massimo & Crocco, Claudia & Marzo, Stefania (eds.), *Towards a New Standard: Theoretical and Empirical Studies on the Restandardization of Italian*. Berlin: De Gruyter Mouton. 213–41. DOI: https://doi.org/10.1515/9781614518839-008

Carignan, Chris & Coretta, Stefano & Frahm, Jens & Harrington, Jonathan & Hoole, Phil & Joseph, Arun & Kunay, Esther & Voit, Dirk. 2021. Planting the seed for sound change: evidence from real-time MRI of velum kinematics in German. *Language* 97. 333–64. DOI: https://doi.org/10.1353/lan.2021.0020

Coetzee, Andries W. & Beddor, Patrice Speeter & Shedden, Kerby & Styler, Will & Wissing, Daan. 2018. Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics* 66. 185–216. DOI: https://doi.org/10.1016/j.wocn.2017.09.009

Crocco, Claudia. 2017. Everyone has an accent. Standard Italian and regional pronunciation. In Cerruti, Massimo & Crocco, Claudia & Marzo, Stefania (eds.) *Towards a New Standard: Theoretical and Empirical Studies on the Restandardization of Italian*. Berlin: De Gruyter Mouton. 89–117. DOI: https://doi.org/10.1515/9781614518839-004

Dodsworth, Robin. 2019. Bipartite network structures and individual differences in sound change. *Glossa: a journal of general linguistics* 4(1). DOI: https://doi.org/10.5334/gjgl.647

Drager, Katie. 2010. Sociophonetic variation in speech perception. *Language and Linguistics Compass* 4. 473–80. DOI: https://doi.org/10.1111/j.1749-818X.2010.00210.x

Draxler, Christoph & Jänsch, Klaus. 2004. SpeechRecorder – A universal platform independent multichannel audio recording software. *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, Lisbon, Portugal.

Ellison, T. Mark & Miceli, Luisa. 2012. Distinguishing Contact-Induced Change from Language Drift in Genetically Related Languages. *Proceedings of the EACL 2012 Workshop on Computational Models of Language Acquisition and Loss*. Avignon, France.

Francois, Alexandre. 2014. Trees, waves and linkages: models of language diversification. In Bowern, Claire & Evans, Bronwen (eds.), *The Routledge Handbook of Historical Linguistics*. London: Routledge. 161–89.

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

Gavin, Michael C. & Botero, Carlos A. & Bowern, Claire & Colwell, Robert K. & Dunn, Michael & Dunn, Robert R. & Gray, Russell D. & Kirby, Kathryn R. & McCarter, Joe & Powell, Adam & Rangel, Thiago F. & Stepp, John R. & Trautwein, Michelle & Verdolin, Jennifer L. & Yanega, Gregor. 2013. Toward a Mechanistic Understanding of Linguistic Diversity. *BioScience* 63. 524–35. DOI: https://doi.org/10.1525/bio.2013.63.7.6

Giannelli, Luciano. 2000. *Toscana*. Pisa: Pacini.

Gordon, Matthew & Barthmaier, Paul & Sands, Kathy. 2002. A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association* 32. 141–74. DOI: https://doi.org/10.1017/S0025100302001020

Guion, Susan G. 1998. The Role of Perception in the Sound Change of Velar Palatalization. *Phonetica* 55. 18–52. DOI: https://doi.org/10.1159/000028423

Hall-Lew, Lauren & Honeybone, Patrick & Kirby, James. 2021. Individuals, communities, and sound change: an introduction. *Glossa*: *a journal of general linguistics* 6(1). DOI: https://doi.org/10.5334/gjgl.1630

Harrington, Jonathan. 2010. Acoustic phonetics. In Laver, John & Hardcastle, William (eds.). *The Handbook of the Phonetic Sciences*. Oxford: Wiley-Blackwell. 81–129. DOI: https://doi.org/10.1002/9781444317251.ch3

Harrington, Jonathan. 2012. The relationship between synchronic variation and diachronic change. In Cohn, Abigail & Fougeron, Cecile & Huffman, Marie (eds.), *Handbook of Laboratory Phonology*, 321–32. Oxford: Oxford University Press.

Harrington, Jonathan & Gubian, Michele & Stevens, Mary & Schiel, Florian. 2019. Phonetic change in an Antarctic winter. *Journal of the Acoustical Society of America* 146. 3327–32. DOI: https://doi.org/10.1121/1.5130709

Harrington, Jonathan & Kleber, Felicitas & Reubold, Ulrich. 2008. Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *Journal of the Acoustical Society of America* 123. 2825–35. DOI: https://doi.org/10.1121/1.2897042

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

Harrington, Jonathan & Kleber, Felicitas & Reubold, Ulrich & Schiel, Florian & Stevens, Mary. 2019a. The phonetic basis of the origin and spread of sound change. In Katz, William F. & Assmann, Peter F. (eds.), *The Routledge Handbook of Phonetics*. Oxford: Routledge. 401–26. DOI: https://doi.org/10.4324/9780429056253-15

Harrington, Jonathan & Kleber, Felicitas & Stevens, Mary. 2016. The relationship between the (mis)-parsing of coarticulation in perception and sound change: evidence from dissimilation and language acquisition. In Esposito, Anna & Faundez-Zany, Marcos (eds.), *Recent advances in nonlinear speech processing: smart innovation systems and technologies.* Berlin: Springer Verlag. 15–34. DOI: https://doi.org/10.1007/978-3-319-28109-4_3

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

Hay, Jennifer B. & Pierrehumbert, Janet B. & Walker, Abby J. & LaShell, Patrick. 2015. Tracking word frequency effects through 130 years of sound change. *Cognition* 139. 83–91. DOI: https://doi.org/10.1016/j.cognition.2015.02.012

Hinskins, Frans. 2021. The Expanding Universe of the Study of Sound Change. In Janda, Richard D. & Joseph, Brian D. & Vance, Barbara S. (eds.), *The Handbook of Historical Linguistics*. Oxford, Wiley-Blackwell. 7–46. DOI: https://doi.org/10.1002/9781118732168.ch2

Kerswill, Paul & Torgersen, Eivind & Fox, Susan. 2008. Reversing 'drift': Innovation and diffusion in the London diphthong system. *Language Variation and Change* 20. 451–91. DOI: https://doi.org/10.1017/S0954394508000148

Kiparsky, Paul. 2003. The phonological basis of sound change. In Joseph, Brian D. & Janda, Richard D. (eds.), *Handbook of Historical Linguistics*. Oxford: Blackwell. 313–42. DOI: https://doi.org/10.1002/9780470756393.ch6

Kirby, James P. 2014. Incipient tonogenesis in Phnom Penh Khmer: Acoustic and perceptual studies. *Journal of Phonetics* 43. 69–85. DOI: https://doi.org/10.1016/j.wocn.2014.02.001

Kirby, James & Sonderegger, Morgan. 2013. A model of population dynamics applied to phonetic change. *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. 776–81.

Kisler, Thomas & Reichel, Uwe & Schiel, Florian & Draxler, Christoph & Jackl, Bernhard & Pörner, Nina. 2016. BAS Speech Science Web Services – an Update of Current Developments. *Proceedings of the 10th International Conference on Language Resources and Evaluation* (LREC 2016). Portorož, Slovenia.

Kokkelmans, Joachim. 2020. Middle High German and modern Flemish s-retraction in /rs/-clusters. In Vogelaer, Gunther & Koster, Dietha & Leuschner, Torsten (eds.), *German and Dutch in Contrast*. Berlin/Boston: De Gruyter. 213–38. DOI: https://doi.org/10.1515/9783110668476-008

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

Kuang, Jianjing & Cui, Aletheia. 2018. Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. *Journal of Phonetics* 71. 194–214. DOI: https://doi.org/10.1016/j.wocn.2018.09.002

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

Labov, William. 1963. The social motivation of a sound change. *Word* 19. 273–309. DOI: https://doi.org/10.1080/00437956.1963.11659799

Levinson, Stephen C. & Gray, Russell D. 2012. Tools from evolutionary biology shed new light on the diversification of languages. *Trends in Cognitive Sciences* 16. 167–73. DOI: https://doi.org/10.1016/j.tics.2012.01.007

Lin, Susan & Beddor, Patrice Speeter & Coetzee, Andries W. 2014. Gestural reduction, lexical frequency, and sound change: a study of post-vocalic /l/. *Laboratory Phonology* 5. 9–36. DOI: https://doi.org/10.1515/lp-2014-0002

Lindblom, Björn & Guion, Susan & Hura, Susan & Moon, Seung-Jae & Willerman, Raquel. 1995. Is sound change adaptive? *Rivista di Linguistica* 7. 5–37.

Lobanov, Boris M. 1971. Classification of Russian Vowels Spoken by Different Speakers. *Journal of the Acoustical Society of America* 49(2B). 606–608. DOI: https://doi.org/10.1121/1.1912396

Maiden, Martin & Parry, Mair. (eds.) 1997. *The Dialects of Italy*. New York: Routledge.

Milroy, James & Milroy, Lesley. 1985. Linguistic change, social network and speaker innovation. *Journal of Linguistics* 21. 339–84. DOI: https://doi.org/10.1017/S0022226700010306

Nguyen, Noël & Delvaux, Véronique. 2015. Role of imitation in the emergence of phonological systems *Journal of Phonetics* 53. 46–54. DOI: https://doi.org/10.1016/j.wocn.2015.08.004

Nielsen, Kuniko. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39. 132–42. DOI: https://doi.org/10.1016/j.wocn.2010.12.007

Norris, Dennis & McQueen, James M. & Cutler, A. 2003. Perceptual learning in speech. *Cognitive Psychology* 47. 204–38. DOI: https://doi.org/10.1016/S0010-0285(03)00006-9

Ohala, John. 1993. The phonetics of sound change. In Jones, Charles (ed.), *Historical Linguistics: Problems and Perspectives*. London: Longman. 237–78.

Ohala, John J. 1989. Sound change is drawn from a pool of synchronic variation. In Breivik, Leiv E. & Jahr, Ernst H. (eds.), *Language Change: Contributions to the study of its causes*. Berlin: Mouton de Gruyter. 173–98. DOI: https://doi.org/10.1515/9783110853063.173

Pardo, Jennifer S. & Gibbons, Rachel & Suppes, Alexandra & Krauss, Robert M. 2012. *Phonetic convergence in college roommates. Journal of Phonetics* 40. 190–97. DOI: https://doi.org/10.1016/j.wocn.2011.10.001

Paul, Hermann. 1888. *Principles of the history of language* London: Swan Sonnenschein, Lowrey & Co.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In Bybee, Joan & Hopper, Paul (eds.) *Frequency effects and the emergence of lexical structure*, 137–57. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/tsl.45.08pie

Recasens, Daniel. 2012. A cross-language acoustic study of initial and final allophones of /l/. *Speech Communication* 54. 368–83. DOI: https://doi.org/10.1016/j.specom.2011.10.001

Recasens, Daniel & Fontdevila, Jordi & Pallares, Maria Dolors. 1995. A production and perceptual account of palatalization. In Connell, Bruce & Arvaniti, Amalia (eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*. Cambridge: Cambridge University Press. 265–81. DOI: https://doi.org/10.1017/CBO9780511554315.019

Ringe, Don & Eska, Joseph F. 2013. *Historical linguistics: Toward a twenty-first century reintegration*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511980183

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

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. DOI: https://doi.org/10.1017/S0025100310000307

Samuels, Michael L. 1972. *Linguistic evolution* Oxford: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781139086707

Sankoff, Gillian. 2001. Linguistic Outcomes of Language Contact. In Trudgill, Peter & Chambers, John & Schilling-Estes, Natalie (eds.), *Handbook of Sociolinguistics*. Oxford: Basil Blackwell. 638–68. DOI: https://doi.org/10.1002/9780470756591.ch25

Sapir, Edward. 1921. *Language: An introduction to the study of speech* New York: Harcourt Brace & World.

Shapiro, Michael. 1995. A Case of Distant Assimilation: /str/ → /ʃtr/. *American Speech* 70. 101–07. DOI: https://doi.org/10.2307/455876

Solé, Maria-Josep. 2007. Controlled and mechanical properties in speech: a review of the literature. In Solé, Maria-Josep & Beddor, Patrice S. & Ohala, Manjari (eds.), *Experimental Approaches to Phonology*. Oxford: Oxford University Press. 302–21.

Sóskuthy, Márton. 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 N. & Kenny, Laurence A. 2013. Revisiting transmission and diffusion: An agent-based model of vowel chain shifts across large communities. *Language Variation and Change* 25(2). 119–53. DOI: https://doi.org/10.1017/S0954394513000069

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

Stevens, Mary & Harrington, Jonathan. 2016. The phonetic origins of /s/-retraction: Acoustic and perceptual evidence from Australian English. *Journal of Phonetics* 58. 118–34. DOI: https://doi.org/10.1016/j.wocn.2016.08.003

Stevens, Mary & Harrington, Jonathan & Schiel, Florian. 2019. Associating the origin and spread of sound change using agent-based modelling applied to /s/-retraction in English. *Glossa: a journal of general linguistics* 4(1). 1–30. DOI: https://doi.org/10.5334/gjgl.620

Todd, Simon & Pierrehumbert, Janet B. & Hay, Jennifer. 2019. Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model. *Cognition* 185. 1–20. DOI: https://doi.org/10.1016/j.cognition.2019.01.004

Trask, Robert L. 2000. *The dictionary of historical and comparative linguistics.* Edinburgh: Edinburgh University Press.

Trudgill, Peter. 1986. *Dialects in contact.* New York: Blackwell.

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

Vogt, Paul & de Boer, Bart. 2010. Editorial: Language Evolution: Computer models for Empirical Data. *Adaptive Behavior* 18. 5–11. DOI: https://doi.org/10.1177/1059712309350968

Weinreich, Uriel & Labov, William & Herzog, Marvin. 1968. *Empirical foundations for a theory of language change.* In Lehmann, Winfred & Malkiel, Yakov (eds.), *Directions for Historical Linguistics*. Austin: University of Texas Press. 95–198.

Weirich, Melanie & Simpson, Adrian. 2015. Gender-specific differences in sibilant contrast realizations in English and German. *Proceedings of the International Congress of the Phonetic Sciences*, Glasgow.

West, Paula. 1999. Long distance coarticulatory effects of British English /l/ and /r/: an EMA, EPG and acoustic study. *Proceedings of the International Congress of the Phonetic Sciences*. San Francisco. 1901–04.

Winkelmann, Raphael & Harrington, Jonathan & Jänsch, Klaus. 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

Yoshinaga, Tsukasa & Nozaki, Kazunori & Wada, Shigeo. 2019. Aeroacoustic analysis on individual characteristics in sibilant fricative production. *Journal of the Acoustical Society of America* 146. 1239–1251. DOI: https://doi.org/10.1121/1.5122793

Yu, Alan C. L. 2010. Perceptual compensation is correlated with individuals' "autistic" traits: Implications for models of sound change. *PLoS ONE* 5. e11950. DOI: https://doi.org/10.1371/journal.pone.0011950

Yu, Alan C. L. 2021. Toward an individual-difference perspective on phonologization. *Glossa* 6(1). 14. DOI: https://doi.org/10.5334/gjgl.661

Yu, Alan C. L. & Zellou, Georgia. 2019. Individual Differences in Language Processing: Phonology. *Annual Review of Linguistics* 5. 131–50. DOI: https://doi.org/10.1146/annurev-linguistics-011516-033815

Zellou, Georgia & Scarborough, Rebecca & Nielsen, Kuniko. 2016. Phonetic imitation of coarticulatory vowel nasalization. *Journal of the Acoustical Society of America* 140. 3560–75. DOI: https://doi.org/10.1121/1.4966232