/u/-fronting and agent-based modeling: The relationship between the origin and spread of sound change

Jonathan Harrington, Florian Schiel

Language, Volume 93, Number 2, June 2017, pp. 414-445 (Article)

Published by Linguistic Society of America
DOI: *https://doi.org/10.1353/lan.2017.0019*

/u/-FRONTING AND AGENT-BASED MODELING: THE RELATIONSHIP BETWEEN THE ORIGIN AND SPREAD OF SOUND CHANGE

JONATHAN HARRINGTON
Ludwig Maximilians University of Munich

FLORIAN SCHIEL
Ludwig Maximilians University of Munich

This study is concerned with whether an asymmetric phonetic overlap between speaker groups contributes to the directional spread of sound change. An acoustic analysis of speakers of Southern British English showed that younger speakers' fronted /u/ was probabilistically closer to that of older speakers' retracted /u/ distributions than the other way around. Agent-based modeling based on the same data showed an asymmetric shift of older toward younger speakers' fronted /u/. The general conclusion is that sound change is likely to be propagated when a phonetic bias within an individual is further magnified by a difference between speaker groups that is in the same direction.*

Keywords: sound change, agent-based modeling, vowels, speech dynamics, imitation

1. INTRODUCTION. The study of how historical sound change develops out of synchronic variation has been pursued following, broadly, two separate and to a large extent nonoverlapping objectives over the last forty to fifty years. The first is predominantly concerned with bias factors internal to the language and how they can cause continuous phonetic variation to become unstable, leading to a categorical sound change. Much of the research in this area, inspired by Ohala's (1981, 2012) research over several decades, has used laboratory studies to understand the conditions that give rise to sound changes across languages of the world, such as umlaut (Beddor et al. 2002), tonogenesis (Hombert et al. 1979, Kirby 2014), and nasalization (Beddor 2009, Solé 2014). An important characteristic of this research is that it is analyzed and tested in terms of a cognitive model of speech communication: some of the issues considered in this regard include (i) how coarticulation is controlled in speech production and then parsed or indeed misparsed in speech perception (Beddor 2012, Ohala 1993), and (ii) the emergence of sound change when listeners who perceive hypoarticulated speech—that is, a more casual speaking style in which consonants are often lenited and vowels are centralized or even deleted—exceptionally direct their attention to the signal rather than the content/semantic properties (Lindblom et al. 1995).

The focus of the second is on factors external to the language and in particular on how sound change is related to interactions between different speaker groups: for example, those due to gender (Eckert 1989, Labov 1990), adult/child (Beckman et al. 2014, Kerswill 2003, Labov 2007), social (Milroy & Milroy 1993), or dialect (Jacewicz & Fox 2012, Wolfram & Schilling-Estes 2003) differences. In the wake of the ground-breaking research by Weinreich and colleagues (1968), the evidence in such studies has often been founded on apparent-time studies (Bailey et al. 1991), in which sound change is inferred by comparing younger and older speakers from the same community, and to a lesser extent on longitudinal studies of change within the same individual(s) over several decades (Harrington et al. 2000, Quené 2003, Sankoff & Blondeau 2007). This type of research has been applied far more than the first type to understanding how sound change, once it

* We thank Megan Crowhurst, Kathryn Campbell-Kibler, James Kirby, and an anonymous referee for their helpful and insightful comments. Some of the functions for the agent-based modeling carried out in this article were written by Raphael Winkelmann. This research was supported by European Research Council grant number 295573 'Sound change and the acquisition of speech' (2012–2017).

takes hold, is propagated through a community (Labov 2007). It has been largely underpinned by social rather than cognitive models of human interaction, although in recent years there have been many more attempts to understand the cognitive bases of the spread of sound change due to social factors (e.g. Clopper 2014, Hay et al. 2015).

One of the main functions of the present study is to begin to move toward developing a unified model of sound change that is informed both by the conditions that give rise to sound change and by its propagation through the community (Harrington, Kleber, Reubold, & Stevens 2016). In order to do so, we make use of evolutionary computer models to test how the type of sound change that arises out of phonetic variation develops as speakers modeled as interacting agents imitate each other. We impose from the outset two further requirements on such a model. The first, which is designed to ground the model in phonetic data, is that the starting conditions for the agents' interactions are actual speech signals, rather than synthesized or generated, collected in our case from speakers spanning two generations in an apparent-time investigation of diachronic /u/-fronting. The second is that the input data are dynamic in the sense that the agents' interactions involve exchanging and imitating speech signals that change in time. This second requirement is important principally because so much of internal (or regular) sound change emerges from dynamic processes such as coarticulation, that is, from the fine-grained variation in how speech sounds overlap with and influence each other, in time. We now consider some further issues concerned with /u/-fronting, imitation, and computer models of sound change that form the background to this study.

1.1. /u/-fronting. High back vowel fronting in both the tense (goose, /u/) and lax (foot, /ʊ/) lexical sets in the standard accent of England, Southern British English, has been extensively studied over the last fifty years, both via comparisons of two or more age groups within the apparent-time paradigm (Harrington et al. 2008, Hawkins & Midgley 2005, Henton 1983, McDougall & Nolan 2007) and also longitudinally within the same speaker (Harrington et al. 2000). The phonetic basis of diachronic /u/-fronting is likely to be the synchronic fronting of /u/ in the context of tongue-tip or palatal consonants, a context effect to which listeners have been shown to be sensitive (Harrington et al. 2008, Lindblom & Studdert-Kennedy 1967, Ohala & Feder 1994). /u/-fronting is also reported to be a sound change in progress in several varieties of English (e.g. American: Fridland 2008, Australian: Cox & Palethorpe 2001, and New Zealand English varieties: Gordon et al. 2004). Beyond English, there is also evidence that back vowels are far more likely to front than front vowels are to retract diachronically. This asymmetry is incorporated by Labov (1994) as one of the principles accounting for the directionality of vowel chain shifts. The phonetic basis of this asymmetry was investigated by Harrington and colleagues (2011): their analyses of tongue position and movement for combinations of three places of articulation with almost all German vowels showed that mid and high back vowels were physiologically more peripheral, that is, further removed from the center of the vowel space, than were front vowels.

1.2. Imitation. Spontaneous imitation has been shown to occur—sometimes in the absence of any social pressure to do so—in many recent studies (Delvaux & Soquet 2007, Nielsen 2011, Pardo et al. 2010, Pardo et al. 2012). In Trudgill's (2004, 2008) dialect mixture model, imitation in speech, which he associates with a more general tendency toward behavioral coordination (see e.g. Sebanz et al. 2006, Shockley et al. 2009), causes sound change of the kind that has been documented in varieties such as New Zealand English in the nineteenth century, when the original settlers were isolated from other communities. Trudgill (2008) reasons that the outcome is deterministic and

mechanistic and that there was no social force behind the development of this variety, such as a need to aspire to a common New Zealand identity. Labov (2001), based in part on Bloomfield's (1933) principle of density, suggests that the diffusion of linguistic change is a predictable consequence of speakers interacting with each other. All of these deterministic models are consistent with the findings in Harrington et al. 2000, which shows that, over thirty years, Queen Elizabeth II's accent shifted from an aristocratic toward a more middle-class variety of Southern British English, but without attaining the middle-class vowel positions. This would be precisely the outcome expected in a deterministic model driven by spontaneous imitation, on the probable assumption that the Queen increasingly came into contact with more middle-class speakers during the two or three decades after the 1950s, when the dividing lines of the rigid class structure were becoming more blurred (Cannadine 1998).

Incremental sound change due to mutual imitation is also predicted by exemplar models of speech (Bybee & Beckner 2010, Pierrehumbert 2003a, 2006), in which phonological categories are modeled as statistical abstractions across remembered exemplars. In these models, imitation in production is the consequence of perceived exemplars of speech being folded into word distributions, from which phonological (and social) categories emerge: the phonological category of the hearer-turned-speaker would then shift slightly toward that of the interlocutor. Repeated exposure to a new variant should therefore result in a shift in the direction of that variant, possibly in an entirely deterministic manner, following the ideas sketched above. Of course, this is not to deny that social factors are also involved in sound change: social factors may in any case have some role to play in any model of sound change based on the mutual imitation of speech, given the recent findings showing that the extent to which one person imitates another may also be dependent on the attractiveness and typicality of the interlocutor's voice (Babel 2012, Babel et al. 2014). Nevertheless, our present study is based on building in the first instance a deterministic model in which sound change evolves as a result of population interaction, independently of social factors (see also Pierrehumbert et al. 2014).

1.3. Agent-based modeling. Although Labov (2001) suggests that much of linguistic diffusion is mechanistic and may reduce to a simple calculation, predicting the actual output when, for example, different dialect groups come into contact with each other is far from straightforward. One of the most developed models in this respect is that of Trudgill (2004, 2008), who applies deterministic principles to predicting the characteristics of the new dialect of English formed after settlement of New Zealand in the nineteenth century. One of the suggestions in Trudgill 2008 is that the phonetic outcome of this kind of dialect mixture is dependent on whichever variant is numerically superior. For example, while New Zealand was certainly settled by speakers of varieties from northern England, who produced a lax high back /ʊ/ in words like *strut* (to rhyme with *foot*), this pronunciation did not take hold in the development of New Zealand English because such speakers (with a /ʌ, ʊ/ merger) were in the minority (Trudgill et al. 2000). When two dialect groups with different variants are in roughly equal proportion, then one possibility is that both variants survive (Trudgill 2010). Another possibility is that neither variant wins, leading to a form of phonetic averaging. An example of such averaging is the development of a pronunciation in Irish English for 'each' that Trudgill (2008:245) transcribes orthographically as *euch*, which 'was intermediate between the two major competing forms in the mixture that developed in Ireland, *uch* and *ech*'. As another example, Trudgill (1999) notes that an earlier pronunciation in East Anglia of *boat* as /bu:t/ has given way to a pronunciation /bout/, because this is along the trajec-

tory between /bu:t/ and the /bæut/ found in many London varieties with which speakers of the East Anglian variety came into contact. Thus the important point here is that the variants of the dialects accommodate to each other, so that the outcome can be an 'intermediate phonetic form' (Trudgill 1999:6) characteristic of neither dialect that went into the mixture. At the same time, Trudgill (1999:7) emphasizes that the resulting form does 'not necessarily have to be intermediate in any simple or straightforward way', although he does not give many details about what a non-straightforward phonetic outcome might be. In the present article, our aim is to make use of agent-based modeling to predict the outcome of contact, based on the statistical properties of the distributions that are the input to the mixture. Our approach is influenced by developments in the last decade within usage-based or exemplar models of perception in which phonological categories, defined as statistical abstractions over perceived exemplars, are updated through interpersonal contact (Pierrehumbert 2003a, 2006).

As discussed by Gilbert (2007:1): 'In comparison with variable-based approaches using structural equations, or system-based approaches using differential equations, agent-based simulation offers the possibility of modeling individual heterogeneity, representing explicitly agents' decision rules … It allows modelers to represent in a natural way … the emergence of structures at the macro or societal level from individual action'. Some agent-based model studies, such as those of Fagyal and colleagues (2010), are concerned with using such models to explain how the social interconnectedness of individuals can drive linguistic change. The Pierrehumbert et al. 2014 model is also relevant to this issue, but in contrast to Fagyal et al. 2010 shows that prestige or the degree to which an individual is connected is less important for propagating linguistic change, which is instead driven by clusters of connected individuals who introduce innovative changes, analogous to the linguistic changes that have been shown to originate and to be propagated in closely knit communities (Milroy & Milroy 1993).

1.4. Computer models of sound change. Computational models provide one of the ways of exploring how sound or language change can accrue incrementally from interactions between agents that represent the speakers of a population. In some models (e.g. Baxter et al. 2009, Fagyal et al. 2010, Pierrehumbert et al. 2014), the input to the system consists of discrete categories (e.g. categorically different pronunciation variants in modeling sound change), while in others (e.g. Blevins & Wedel 2009, Garrett & Johnson 2013, Kirby 2014, Pierrehumbert 2001) the input is based on continuous parameters (e.g. duration, formant values). These latter models have various aims that include: demonstrating the mechanisms by which incremental category changes can occur within the exemplar paradigm (Pierrehumbert 2001; see also Wedel 2007); testing Labov's (2007) model of sound change, which is predicated on a distinction between how sounds are transmitted from caregivers to children and diffused between adults (Stanford & Kenny 2013); demonstrating how acoustic cues that become socially significant can effect category change (Garrett & Johnson 2013); understanding the occurrence of so-called anti-homophony, by which words whose meaning is not otherwise resolved by context tend not to merge diachronically (Blevins & Wedel 2009); explaining phonologization as an emergent consequence of a combination of precision loss and enhancement (Kirby 2013, 2014; see also Kirby & Sonderegger 2013); and quantifying the relationship between the conditions that give rise to sound change and its actuation (Sóskuthy 2015).

The computational model in this study is designed to test a hypothesis about the outcome when two groups of speakers, *A* and *B*, with different phonetic exponents of a

phoneme, exchange and imitate each other's isolated word productions. The prediction to be tested is that if the direction of variation is asymmetric, such that *B*'s exponents are oriented toward those of *A* to a greater extent than those of *A* are toward *B*, then *B*'s phonetic exponents should shift through interaction toward those of *A* (see also Fig. 1 below). For example, if *A* and *B* are speaker groups who have nasalized and oral vowels, respectively, in nasal contexts such as *man*, *sand*, and *send*, then the prediction is that group *B*'s oral vowels will become increasingly nasalized through contact with group *A*. This is because *B*'s phonetic exponents, while predominantly oral, may occasionally nevertheless be slightly nasalized in nasal contexts (at e.g. faster speech rates), but there is no phonetic reason why *A* should have any bias for producing oral vowels in the same nasal contexts. Accordingly, there is an asymmetric bias, which, by hypothesis, should result in a convergence toward nasalized vowels, when *A* and *B* interact with each other.

The theory is neutral (and equally applicable) to the many different kinds of social-indexical variation between the groups. For example, *A* and *B* might be two different dialect groups. In this case, the prediction is that if the dialects come into contact with each other, and if there is the type of asymmetric variation outlined above, then there will be a greater convergence of *B* toward *A* than the other way around. Such a model suggests a possible extension to Trudgill's (1999, 2004, 2008) dialect mixture model, as outlined earlier. Thus, whereas the outcome for Trudgill when dialects come into contact with each other may be some form of phonetic averaging between the two, the idea in the model being proposed here is that the outcome of convergence is also strongly influenced by the direction of the spread, that is, of how the exponents from the two dialects are oriented with respect to each other.

The same prediction applies to phonetically driven sound changes (sometimes referred as 'change from below'), which are often analyzed using older and younger speakers in apparent-time analyses common in sociolinguistics (Bailey et al. 1991, Weinreich et al. 1968). In such studies, the assumption is that older speakers have participated minimally in the sound change in progress. Nevertheless, if the sound change is phonetically motivated, then the phonetic exponents of older speakers should show some bias in the direction of the sound change that eventually takes place, that is, toward those of younger speakers. Consequently, if older and younger speakers interact with each other, then there should be convergence and a greater shift of phonetic exponents toward those of younger speakers. Perhaps this is one of the reasons why some longitudinal analyses show that older adult speakers shift their pronunciation in the direction of sound changes that are taking place, or have recently taken place, in the community (Harrington et al. 2000).

These predictions are tested in the present study using data from a previous apparent-time study concerned with /u/-fronting in Standard Southern British (Harrington et al. 2008). The sound change is likely to be phonetically motivated based both on the much greater distance of the tongue dorsum from the center of the vowel space for high back than for high front vowels, and because high back vowels in English and indeed in many languages of the world so often occur in coronal consonantal contexts, which are likely to induce tongue fronting (see Harrington et al. 2011 for further details). Older speakers, even if they have participated minimally in the sound change in progress, should have variants of /u/ that are skewed toward those of younger speakers, precisely because /u/ tends synchronically to be fronted in consonantal fronting contexts, including after palatals (*few* /fju/) and in the context of alveolars (*soon* /sun/), resulting in F2 raising of /u/ (Harrington et al. 2008, Hawkins & Midgley 2005). Younger and older speakers are, then, like the *A* and *B* groups outlined earlier: if they were therefore to in-

teract with each other, the prediction is that the older speakers' phonetic exponents of /u/ should shift toward those of the younger speakers, that is, that there should be a convergence in the direction of the sound change that has taken place.

The aim of the first part of this study is to test whether a group-dependent asymmetry exists such that the variation in older speakers' /u/-variants is in the direction of those of younger speakers. The aim of the second part of the study is to determine whether this asymmetry influences the degree to which the two groups accommodate to each other by applying agent-based modeling to (dynamic encodings of) the actual F2 values of these younger and older speakers, each separately represented by an agent.

2. GROUP OVERLAP IN /u/-FRONTING.

2.1. METHOD.

SPEAKERS AND MATERIALS. The speakers and materials were the same as those in Harrington et al. 2008. The speakers included a younger group consisting of fourteen subjects (three male, eleven female), between eighteen and twenty years old (mean age 18.9 years), and an older group of thirteen subjects (seven male, six female), between fifty-two and seventy-four years old (mean age 69.2 years). All speakers were impressionistically judged to be of a Standard Southern British (SSB) variety. These twenty-seven speakers produced isolated, randomized words, presented individually on a computer monitor, that covered most of the SSB vowel space. A total of 540 words were produced per speaker in this way from a randomized list of ten repetitions of fifty-four words. Any words that were mispronounced were excluded from the analysis. The analysis was based on the eleven words shown in Table 1, which contains combinations of four initial consonants, /f, s, k, h/, with following /i, ju, u/ (with the exception of /sju/, since there are no /sju, su/ contrasts in English).

	/f/	/s/	/k/	/h/
/i/	feed	seep	keyed	heed
/ju/	feud	—	queued	hewed
/u/	food	soup	cooed	who'd

TABLE 1. The word types used in the present study and produced by older and younger speakers in Harrington et al. 2008.

A motivating factor in choosing these subsets of words was to include minimal pair triplets across /i, ju, u/ (see Table 1). We included /i/ for two reasons: first, since no sound change has been reported in the last fifty years in SSB for this vowel, /i/ can provide a baseline for assessing the magnitude of the difference between younger and older groups on /ju, u/, which is likely to have come about due to sound change; and second, /i/ was included in order to classify probabilistically the entire high vowel space in SSB in the subsequent agent-based modeling study. The total number of produced tokens that formed the basis of this study was 11 (word types) × 10 (repetitions) × 27 (speakers) = 2,970 tokens. However, thirty-three tokens were excluded due to mispronunciations, leaving 2,937 tokens that were analyzed in this study. As described in Harrington et al. 2008, the words were acoustically segmented between the acoustic onset and offset of voicing for /i, ju, u/, and there was no further subsegmentation of the /ju/ trajectory.

ACOUSTIC PARAMETERS. The subsequent parameterization of these data was exactly as described in Harrington et al. 2008 and is summarized briefly here. The first four formant frequencies were calculated using a frame shift of 5 ms and a 30 ms Blackman window. Only F2 was analyzed in the present study: any obvious formant-tracking errors (such as when F2 tracked a lower or higher formant) were manually corrected. F2

in Hertz was converted to a Bark scale using the formula in Traunmüller 1990. Each trajectory was linearly time-normalized, that is, converted to proportional time values between 0 (voicing onset) and 1 (voicing offset).

A discrete cosine transformation (DCT) was applied to each (time-normalized) F2 trajectory using 1. For an F2 trajectory $F(n)$ of N points, $0 \leq n \leq N-1$, the m^{th} DCT-coefficient C_m ($m = 0, 1, 2$) was calculated from 1.

$$(1)\quad C_m = \frac{2k_m}{N} \sum_{n=0}^{N-1} F(n) cos\left(\frac{(2n+1)m\pi}{2N}\right)$$

$$k_m = \frac{1}{\sqrt{2}},\ m = 0;\ k_m = 1,\ m \neq 0$$

The three coefficients C_m are proportional to the F2 trajectory's mean, linear slope, and curvature, respectively, and thereby encode dynamic information in the formant trajectory (Watson & Harrington 1999). The corresponding smoothed F2 trajectories were derived by applying an inverse DCT to the output from 1.

There are several reasons for representing the formant trajectories using the coefficients of the DCT rather than sampling F2 values at, for example, the vowel target or at several time points (see also Harrington et al. 2008, Watson & Harrington 1999). First, as outlined earlier, the DCT is appropriate for modeling many kinds of sound change that are often associated with inherently dynamic synchronic processes, such as coarticulation and undershoot. Second, the DCT avoids the difficulty of having to identify a target, especially for trajectories, such as /ju/, that may have no steady-state component (see e.g. van Son & Pols 1990 for a detailed discussion). Third, in the subsequent agent-based modeling carried out in the second part of this study, the shape of the trajectory can change as a consequence of interactions between the agents. Shifts in the trajectory shape are appropriate for modeling many kinds of diachronic, dynamic vowel changes, such as the change from hiatus to diphthongs in Romance languages (Chitoran & Hualde 2007), the development of onglides in high vowels (as in Australian English; e.g. Cox & Palethorpe 2001), and the emergence of diphthongs from monophthongs (as in the Great Vowel Shift; Jespersen 1909). None of these could be modeled in an agent-based model in which a trajectory was represented by one or more static slices extracted from a vowel.

Classification. Classifications were in the three-parameter space using the DCT-coefficient vectors C derived from 1, and they were designed to determine whether the probabilistic distance of younger speakers to an older speaker's distribution was less than the probabilistic distance of older speakers to a younger speaker's distribution. For this purpose, the distribution for a speaker i on class V (V = /i, ju, u/) was defined by the mean and covariance matrix $\mu_{i.V}$ and $\Sigma_{i.V}$, respectively (thus three parameterized distributions per speaker, one per class). The squared Mahalanobis distance ($mdist_{V.j.i}$; see appendix) of any given DCT-triplet $C_{V.j}$ sampled from the same class V and from speaker j to speaker i's distribution was given by 2.

$$(2)\quad mdist_{V.j.i} = (C_{V.j} - \mu_{i.V})^T \Sigma_{i.V}^{-1} (C_{V.j} - \mu_{i.V})$$

The T superscript denotes transposition, and the -1 superscript denotes matrix inversion of the covariance matrix $\Sigma_{i.V}$. The Mahalanobis distance $(mdist_{V.j.i})^{1/2}$ is always positively valued and is equal to the (dimensionless) number of ellipsoid standard deviations from the distribution's centroid (the lower the value, the closer a DCT-triplet to the centroid). In the calculations in 2, j varied over all of the speakers from the other age group in relation to speaker i. Thus, if speaker i belonged to the YOUNGER group, 2 was calculated for each DCT-triplet that occurred in class V produced by each OLDER speaker. Finally, the resulting Mahalanobis distances were aggregated by word. The end

result was eleven aggregated distances for speaker *i*, $\bar{x}_{W.i}$, one per word (*W* = *cooed*, *feed*, *feud*, *food* ...): thus, if speaker *i* was young, then $\bar{x}_{queued.i}$ was the Mahalanobis distance aggregated across older speakers' /ju/ in *queued* to speaker *i*'s /ju/ distribution.

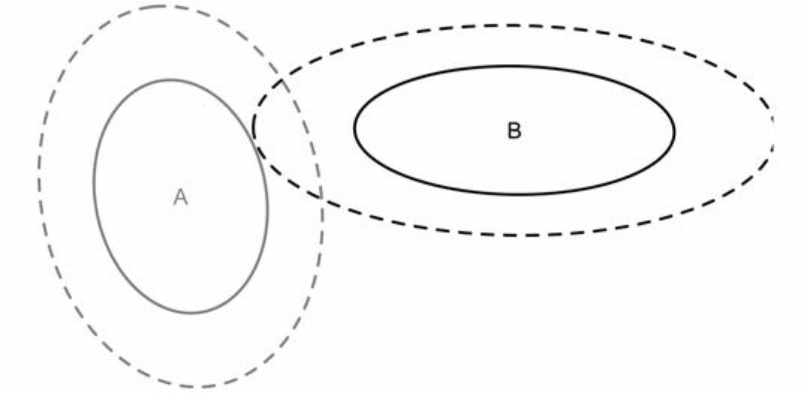

FIGURE 1. The hypothetical distribution of the phonetic variants of a phonological class produced by two speakers, *A* and *B*, from different varieties. The solid and dotted ellipses are at two and four standard deviations, respectively.

The hypothesis to be tested was that $\bar{x}_{W.i}$ was less in the /ju, u/ sets of words for older than for younger speakers. A two-dimensional schematic outline of this hypothesis is shown in Figure 1 for a hypothetical distribution of tokens produced by two speakers from two different varieties, such that speaker *B*'s distribution is oriented toward that of speaker *A*. In this case, samples from *A*'s distribution have a smaller Mahalanobis distance and therefore a greater posterior probability of belonging to *B* than do samples from *B* of belonging to *A*: this is evident from Fig. 1, where *B*'s distribution encroaches on *A*'s at four ellipse standard deviations (the dashed ellipses) but not the other way around. In terms of the present hypothesis to be tested, *A* and *B* are a younger and older speaker, respectively, and the distributions encompass samples from either /ju/ or /u/.

The further basis of this hypothesis is that, as outlined in §1, coarticulation and/or undershoot are more likely to front a retracted /u/ (characteristic of an older speaker) than to retract a fronted /u/ (characteristic of a younger speaker), so that the older speaker's ellipse should be skewed toward that of the younger speaker, as in Fig. 1.

2.2. RESULTS. The DCT-smoothed F2 trajectories in Figure 2, aggregated by speaker in the three separate vowel contexts, show that older speakers had lower F2 trajectories in /ju, u/, which is consistent with the evidence in Harrington et al. 2008 that /u/ was phonetically more retracted for older than for younger speakers.

The issue to be determined here is whether there is any evidence of more variation in the older speakers' /ju, u/ trajectories in the direction of those of younger speakers than the other way around. As an initial step in this direction, the trajectories were pooled by age group, and then the variance was calculated across the data at each of twenty-one equally spaced time points between voicing onset and offset, for the /ju, u/ classes separately. The results of this calculation, given in Figure 3, where the variance is plotted at ±1.96 standard deviations from the group means, show a much wider variation for the older than the younger age group. There is also some evidence from this display that the data from the older speakers are more likely to spill over into the younger speakers' distributions: for example, the distance between the upper variation band for older speakers and the mean trajectory for younger speakers is smaller than is the distance between the lower band of variation for younger speakers and the mean trajectory of older speakers.

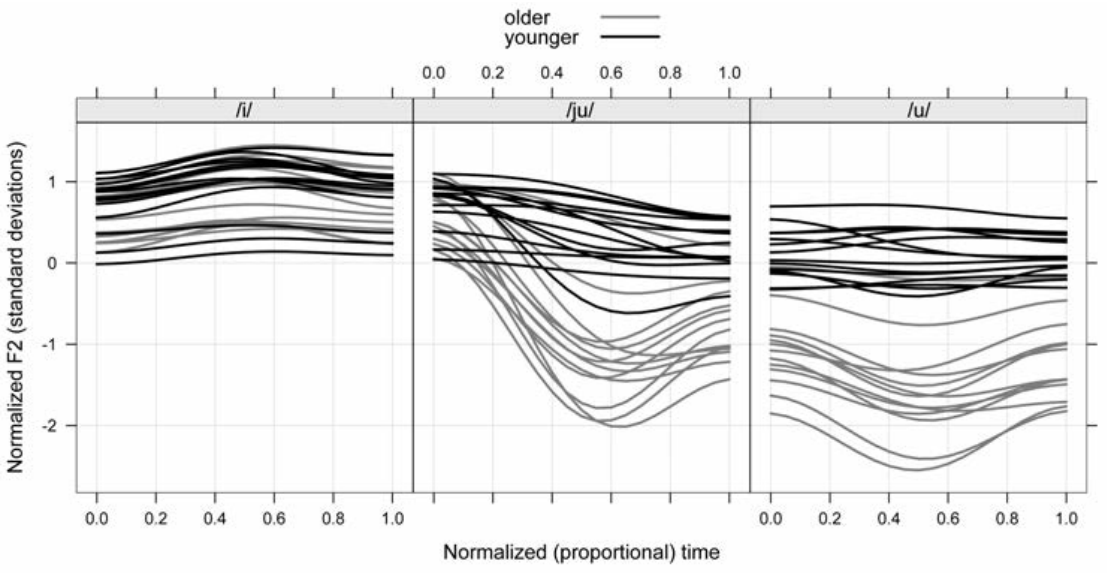

FIGURE 2. Time-normalized trajectories between voicing onset and offset in /i, ju, u/. The trajectories are aggregated by speaker (thus one trajectory per speaker) and shown separately for the older (gray) and younger (black) age groups.

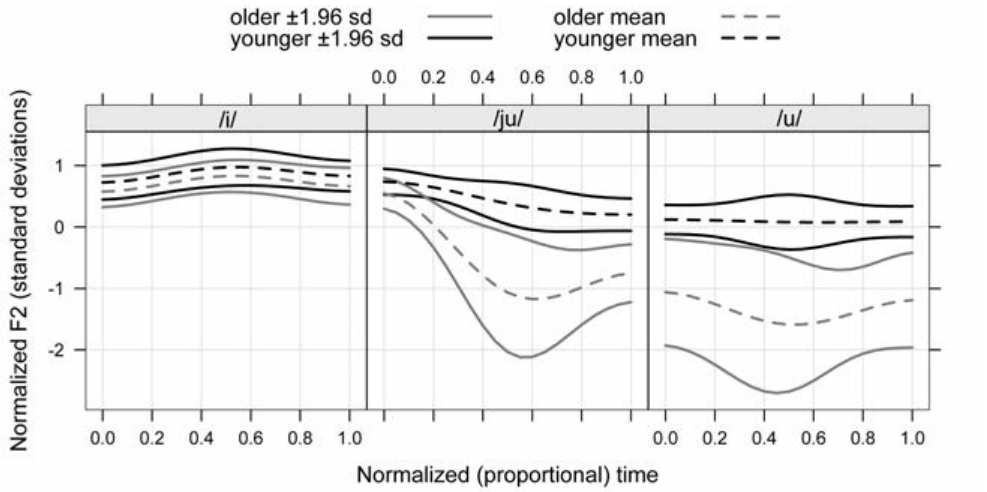

FIGURE 3. The solid lines are the F2 trajectories aggregated across the two age groups (gray/black for older/younger speakers respectively) at ±1.96 standard deviations from the mean, shown as a dashed line, as a function of time for the three vowel classes separately.

Some further evidence that the data from the older speakers were oriented to a greater extent toward those of younger speakers is shown for /ju, u/ in the DCT space in Figure 4 (from which the smoothed trajectories in Fig. 3 were derived). For /u/ (Fig. 4, left panel) and for /ju/ (Fig. 4, right panel), there is evidently a tighter clustering of points for the younger speakers and some evidence especially along the C_0 and C_1 axes that older speakers' points stray more into the younger speakers' space than the other way around.

There is, however, an obvious confound in Figs. 3 and 4 in that they combine within the age groups the variation that occurs both within and between the groups: that is, the wider band of variation in Fig. 3 for older speakers may simply come about because, as Fig. 2 shows, there are two older speakers whose /ju, u/ trajectories fall within the trajectory space of the younger group. In order to resolve this confound, mean Mahalanobis distances in the DCT space were calculated from older speakers to each

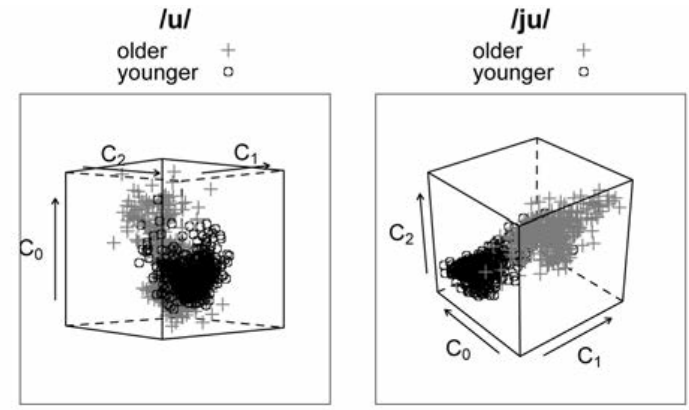

FIGURE 4. Distribution of data points from older (gray plus signs) and younger (open black circles) speakers for /u/ (left) and /ju/ (right) in the DCT $C_0 \times C_1 \times C_2$ space; data points are proportional, respectively, to the F2 mean, slope, and curvature as a function of time.

younger speaker's vowel category, and from younger speakers to each older speaker's vowel category, for each word separately, in the manner described in §2.1 'Classification'. The distances were also logarithmically transformed to reduce the skew in the raw distances (in which there was a large discrepancy between the mean and median of the data).

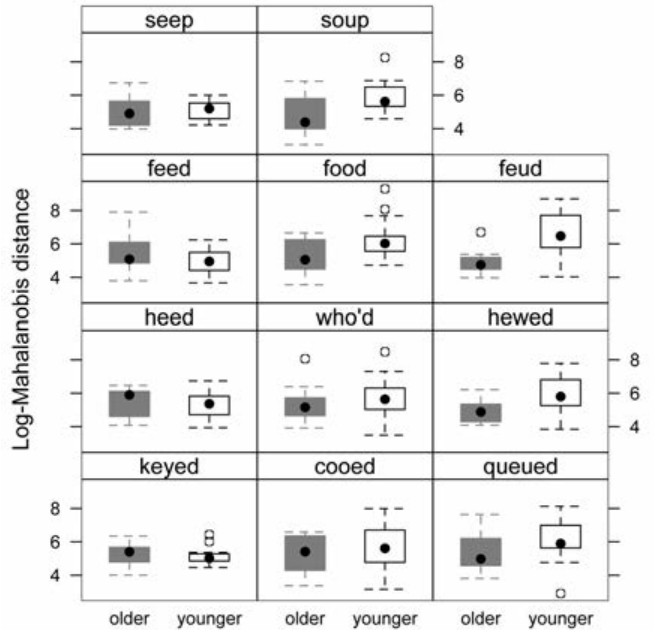

FIGURE 5. The distribution in each boxplot shows the log-Mahalanobis distances aggregated across younger speakers to the centroid of older speakers ('older') and aggregated across older speakers to the centroid of younger speakers ('younger'). Each boxplot consists of one data point per speaker.

As Figure 5 shows, there was a difference between the age groups on this measure in the predicted direction: that is, for /ju/ (column 3) and to a lesser extent for /u/ (column 2), the Mahalanobis distances from older speakers' tokens to the younger speakers' categories ('younger') were larger than the corresponding distances from younger speakers' tokens to the older speakers' categories ('older'). It seems unlikely that this trend is an artifact of biological age differences between the groups, given that, as Fig. 5 also shows, there were no consistent differences between the age groups on /i/.

A mixed model with the log-Mahalanobis distance as the dependent variable, with fixed factors vowel (two levels: /ju, u/) and age (two levels: younger, older), and with random factors word (the seven word types with /ju, u/ nuclei in Fig. 5) and speaker (for which distances to the other age group had been calculated) showed a significant difference for age ($\chi^2(1) = 6.0$, $p < 0.05$) but no interaction between these factors.[1] The result of this statistical test supports the evidence in Fig. 5 of a greater Mahalanobis distance to younger than to older speakers' categories.

2.3. Discussion. The motivation for the analysis has been to test whether there is an asymmetry between the groups in the direction of /ju, u/ variation. The analysis has shown this to be so: the data suggest that there is a greater probability of younger speakers' /ju, u/ falling within the distribution of older speakers' vowel classes than the other way around. This finding is not due to biological age, given that no such asymmetry could be detected in /i/ (for which there is also no reported sound change in SSB). The source of this asymmetry is likely to be that coarticulation and vowel undershoot push older speakers' predominantly retracted /u/-variants toward the phonetically more fronted /ju, u/ space of younger speakers. Figure 5 provides further evidence for this view: the difference between younger and older speakers is especially marked in those contexts (/ju/ and *soup*) in which the anterior consonant exerts a strong coarticulatory fronting effect on a retracted /u/. We have argued that the results in Fig. 5 are due to differences in the direction of variance, that is, that older speakers' spaces are oriented toward those of younger speakers to a greater extent than the other way around, and there is some evidence for this position from Fig. 4. But the greater scatter size in the three-dimensional DCT space for older than younger speakers may be another factor that contributes to the greater probability of a younger person's /ju, u/ being absorbed probabilistically into the distribution of older speakers than vice versa.

The further issue to be explored in the next section is the possible outcome when speakers from these two groups imitate each other, where the imitation is simulated through agent-based modeling. The prediction is that the outcome follows the direction in which sound change has taken place, that is, that older speakers' /ju, u/ will shift toward the younger speakers' fronted /ju, u/ variants.

3. Agent-based modeling. An agent-based speech communication model was implemented with a minimum number of assumptions in order to test a specific hypothesis about the direction of sound change when two speaker groups with different variants of the same phonological class interact with each other. Some of the assumptions are related to ideas in exemplar theory, in particular that there is a statistical association be-

[1] The model was run in R using the lmer() function from the lme4 package with the following command.

(i) model = lmer(mdist ~ Age * Vowel + (Vowel | Speaker) + (1 | Word))

Where mdist is the log-Mahalanobis distance, Age (age group) and Vowel the two fixed factors, and Speaker and Word the two random factors. As shown in (i), a slope and intercept were included for the Vowel term on the Speaker but not for Age, which is a nonrepeated (between-subjects) factor with respect to Speaker (in the sense that each speaker is either old or young). Only an intercept could be calculated for Word since the model would not converge when a slope was calculated with respect to Age. Calculating a slope with respect to Vowel is not appropriate since Vowel is a nonrepeated factor with respect to Word (in the sense that each word has either an /i/ or a /ju/ or a /u/ vowel).

The significance of the factors was tested by dropping terms and using a chi-squared test of significance to compare the new model with the original model above. There was no significant difference when the Age : Vowel interaction term was dropped. Age was shown to be significant when Age was dropped from the model without interaction.

tween classes (words and phonological units) and multiple signals that are stored in memory (Pierrehumbert 2003a, 2006). Consistent both with the results from various studies on imitation in speech (Babel et al. 2014) and with some ideas developed for exemplar theory (Pierrehumbert 2001, 2002), imitation in production in the agent-based model of the present study is a consequence of a perceived speech signal being incorporated into the distribution (stored in the agent-listener's memory) over which a phonological category is defined.

The mapping between a word class and remembered signals is indirectly mediated via phonology in our computational model. This indirect association reflects an idea in some episodic models of speech that phonology is a further abstraction across word classes and, importantly, does not emerge directly from remembered signals: thus, as Pierrehumbert (2003b:180) comments, '[p]honology represents generalizations over the word-forms in the lexicon, which are in turn generalizations over speech. Hence, phonology does not abstract over speech directly, but rather indirectly via the abstraction of word-forms'. Compatible with this idea, there are in our model two types of statistical generalizations: first, each word class (e.g. *few*) is a statistical model across its corresponding remembered signals; second, each phonological class is a statistical model across the remembered signals of all word classes with which it is associated (e.g. one model across the /ju/ signals of *feud*, *hewed*, *queued*, etc.).

In our computational model, phonology puts a brake on absorbing signals into memory and amalgamating them with a word class. More specifically, if a perceived signal is probabilistically too remote from the corresponding phonological class, then it is not remembered. This part of the model—in which a signal is not absorbed into memory if it is on the edge of the corresponding probability distribution (as a result of which a signal closer to /i/ would not be incorporated into the /ju/ distribution)—is, as Garrett and Johnson (2013) comment, consistent with Labov's (1994:586–87) idea that misunderstood tokens have a negligible influence on speech production; the same idea is compatible with models conceived within the episodic framework, in which a signal is not absorbed into a phonological class if it is more likely to belong to another phonological class (Hay et al. 2015, Silverman 2006). The more general point is that phonology has an indirect influence on speech production in our computational model. This is because phonology regulates the uptake of new signals into the statistical model, which not only defines the mapping between a word class and its remembered signals but also is used for sampling a new signal in speech production.

We incorporate the idea that word classes are transmitted without error from a speaker to a listener based on both the model of sound change in Lindblom et al. 1995 and episodic models of speech perception as outlined above, in which a word class has a statistical association with multiple pronunciation variants stored in memory. In these models, sound change comes about because a speaker samples from a word class that has been modified by one or more new perceived pronunciation variants (Pierrehumbert 2002). But such a modification is by definition possible only if the word class has been correctly identified (since otherwise the word class's model would not be updated).

The computational model was run with three phonological classes /i, ju, u/ (as in *feed*, *feud*, *food*). There were separate classes for /ju, u/ in our simulations because, first, given that there are no clear boundaries between /j/ and /u/ on which to base an acoustic division, /ju/ was not subsegmented; and second, because we require minimal pairs (like *feud* and *food*) to be phonologically distinctive (whereas they would not be, if there were one class for /ju, u/). We do, however, explore the implications for sound change of modeling /ju, u/ as a single phonological class in the general discussion in §5.

3.1. METHOD.

ASSOCIATION BETWEEN SPEAKERS AND AGENTS. The two speakers from the older group whose /ju, u/ were shown in Fig. 2 to pattern with the younger group were removed from further consideration. In order for the results of the agent-based modeling to not be biased by a greater proportion of speakers with fronted /ju, u/-variants, three speakers selected at random from the younger group were also removed, thereby leaving twenty-two speakers: eleven older speakers with a retracted /ju, u/, and eleven younger speakers with a fronted /ju, u/.

In the agent-based modeling carried out here, the information from each of these twenty-two speakers that had been analyzed in the previous section was incorporated into an agent's memory, with one agent per speaker. Each agent had initialized in memory the speaker's word classes (consisting of the eleven word types shown in Fig. 5), their associated vowel classes (/i, ju, u/), and triplets of DCT-coefficients, with one triplet per production. Thus, each agent had stored in memory eleven word classes (four of these with the /i/, three with the /ju/, and four with the /u/ vowel classes) × up to ten (repetitions) = 110 items, and such that each of these 110 items was associated with three DCT-coefficients calculated between voicing onset and offset in the manner explained in §2.1 'Acoustic parameters'.

Agent-based modeling was run (a) with only the eleven speakers from the younger group, (b) with only the eleven speakers from the older group, and (c) with all twenty-two speakers together. The modeling consisted of up to 50,000 iterations (after which there was negligible change; see Fig. 6 below): a single iteration was defined as the transmission of a single item from an agent-talker to an agent-listener. An *item* was defined as a set of five elements consisting of a single word-class label, its corresponding vowel-class label, and the associated DCT-triplet.

The prediction was that there would be little change in the /ju, u/ variants when agent-based modeling was carried out within the relatively homogeneous groups (a) or (b), but that in (c) there would be a shift of /ju, u/ from the older group toward the /ju, u/ variants characteristic of the younger group. No change was expected in /i/.

COMMUNICATION BETWEEN AGENTS. For the agent-based modeling, a single iteration consisted of the random selection of a pair of agents from the population of $n = 22$ agents (when speakers from both age groups were incorporated into the model), irrespective of their age grouping. One of the members of the pair was defined as the agent-talker and the other as the agent-listener. The agent-talker produced a single item (consisting of a DCT-triplet keyed to a vowel and word class). The agent-listener typically absorbed this item into memory, but only after discarding from memory an item of the same word class. The further details are as follows.

Production. For production, one of the word classes was selected at random from the agent-talker's memory. A Gaussian model was then constructed from all items of the same word class, with the DCT-triplets as parameters. Thus, for example, for a randomly selected word class *queued*, the mean and covariance matrix were calculated in a three-dimensional DCT space from all (typically $n = 10$) items with word class *queued* in that agent's memory. This three-dimensional Gaussian distribution was then used for the random generation of a single DCT-triplet: it was this randomly generated DCT-triplet that was transmitted to the agent-listener, together with the corresponding vowel and word classes (i.e. /ju/ and *queued* for this example).

Perception. The item that was transmitted, $item_{perc}$, was incorporated into the agent-listener's memory if there was at least a one-in-three posterior probability of class

membership of $item_{perc}$ with the corresponding agent-listener's vowel distribution. More specifically, three Gaussian models were constructed from all existing DCT-triplets in the agent-listener's memory, one for each of the vowel classes, in order to calculate $p(item_{perc}|\text{i})$, $p(item_{perc}|\text{ju})$, and $p(item_{perc}|\text{u})$. For example, if the word class of $item_{perc}$ was *queued*, then $item_{perc}$ was accepted into memory only if $p(item_{perc}|\text{ju}) > 0.333$, that is, greater than chance.[2] This was done in order to prevent perceived items from being absorbed into a vowel class if acoustically $item_{perc}$ was probabilistically closer to one of the agent-listener's other vowel classes (to prevent e.g. a perceived item from being absorbed into the agent-listener's /ju/ class if the DCT-triplet was acoustically closer to the agent-listener's /i/ or /u/ classes).

Agent perception also involved removing an existing item, $item_{rem}$, from memory before absorbing any $item_{perc}$ into memory that had passed the above threshold. The removal was based on two criteria. First, $item_{rem}$ and $item_{perc}$ had to have the same word class. Second, $p(item_{rem}|\text{V})$ had to be minimal—that is, $item_{rem}$ had to be the most marginal in the agent-listener's vowel space. For example, if the word class of $item_{perc}$ was *queued* (thus with vowel class /ju/), then according to the first criterion, one of the *queued*-items in the agent-listener's memory had to be discarded. According to the second criterion, the *queued*-item that was actually discarded was the one for which $p(item_{rem}|\text{ju})$ was smallest. $Item_{perc}$ was only added to memory once $item_{rem}$ had been discarded according to these two criteria. For this reason, both the total number of items and the number of items per vowel or word class remained constant in an agent-listener's memory after any iteration.

The motivation for discarding an item that was probabilistically marginal in combination with the uptake of a new item was the method that was used here to ensure a certain degree of stability in the distribution of an agent's vowel class over the DCT space, while at the same time guaranteeing a marginal change due to the uptake of a newly perceived item (see also Kirby 2014 and Sóskuthy 2015 for a discussion on this theme of stability vs. change in agent-based modeling of sound change). A high degree of stability following each iteration was guaranteed with this method, precisely because an item was only ever discarded if it was at the probabilistic margin of an agent-listener's vowel class. But there was nevertheless a relatively small shift in the distribution, through both the uptake of a new item and the loss of another each time agent perception occurred.

QUANTIFYING CHANGE. Change over the modeling was assessed by comparing the items in an agent's memory following iteration with those of the baseline, the latter consisting of the original speech data analyzed in §2 and initially stored in the agents' memories. The extent of change was quantified in two ways: by visual inspection of the mean F2 trajectories after applying an inverse DCT to triplets of DCT-coefficients, and by calculating using the Euclidean distances of all items from the baseline in a DCT space after n iterations, as in 3.

$$(3)\quad d_{w.j.n} = \left(\sum_{m=0}^{m=2}(C_{m.w.j.n} - C_{m.w.j.0})^2\right)^{1/2}$$

In 3, $d_{w.j.n}$ is the Euclidean distance and $C_{m.w.j.n}$ is the m^{th} DCT-coefficient (derived from 2) for the vowel in word w and in agent j's memory after n iterations. The vector $C_{m.w.j.0}$ $m = 0 \ldots 2$, is the centroid at baseline, that is, the means of DCT-coefficients in the same word and for the same agent before the modeling. The degree of change was assessed

[2] Classification was carried out based on the assumption of equal prior probabilities (of one in three) in the three vowel classes.

by comparing $d_{w.j.0}$ with $d_{w.j.n}$: that is, by comparing how far away the items of a given word were from the baseline centroid before (0) and after n iterations.

In terms of the predictions formulated in §3.1, $d_{w.j.n} - d_{w.j.0}$ was expected to be (not significantly different from) zero when either only older or only younger agents interacted with each other. When all agents interacted with each other, then $d_{w.j.n} - d_{w.j.0}$ was expected to be greater in /ju, u/ for older than for younger agents: that is, the words with /ju, u/ vowels were predicted to be further away from their original centroids after a certain number of iterations in older than in younger agents.

3.2. Results.

Within-group simulations. We present first the results from the within-group simulations. Recall that in these simulations, agent-based modeling was carried out separately based on speakers from the older and younger groups. No change was expected in either of these simulations, on the assumption that the within-group distributions of /ju, u/ were quite homogeneous. Figure 6 shows the Euclidean distance calculated with 3 in a DCT space from the centroid of the original utterances, with an increasing number of iterations aggregated by age group and separately by word. As Fig. 6 shows, for most words stability—that is, no further change—was obtained beyond roughly 10,000 iterations. The one exception was /u/ in *soup*, which continued to show change in older but not younger agents at up to 30,000 iterations.

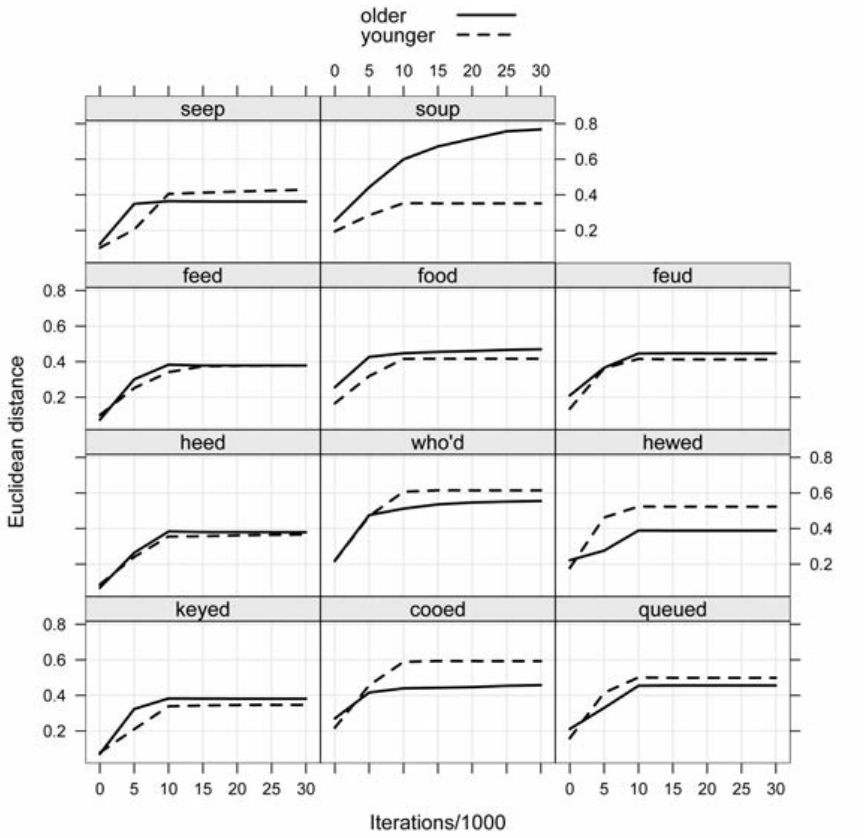

Figure 6. The Euclidean distances to the word- and speaker-specific centroids in the original data, calculated using 3 at intervals of 5,000 iterations, shown separately by word and by age group.

The data in Figure 7 show aggregated F2 trajectories in the original data and after 30,000 iterations. The upper and lower solid lines are of the original data at ±1.96 standard deviations from the mean (of the original data). The dashed lines show the aggregated F2 trajectory after 30,000 iterations.

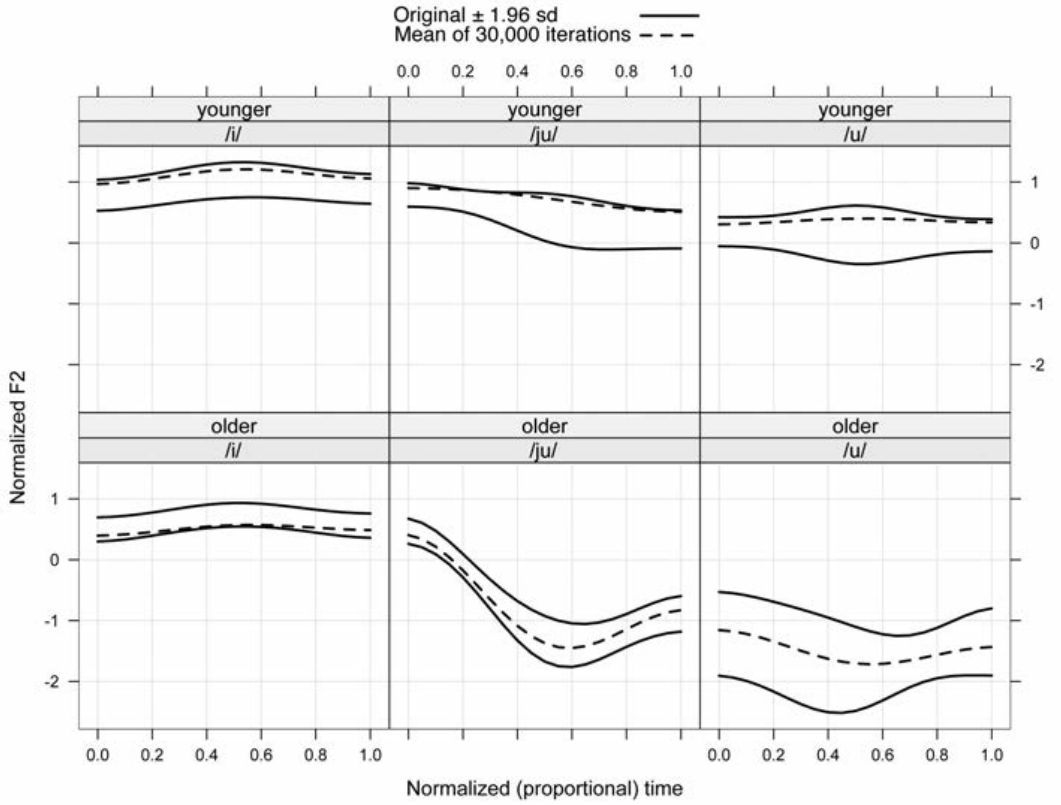

FIGURE 7. The solid lines are F2 trajectories at ±1.96 standard deviations from the mean in the original data, aggregated across speakers separately for the two age groups. The dotted line is the aggregate after 30,000 iterations.

There was evidently a tendency after 30,000 iterations for F2 to rise in all three vowel classes of younger agents, while F2 fell slightly in older agents in /i/. However, Fig. 7 shows evidence that the trajectories after 30,000 iterations were within the ±1.96 standard deviation band of the original data, which is consistent with the predicted interpretation of minimal change, when agent-based modeling is run only within the younger or the older group.

In order to quantify whether there was any significant change, the Euclidean distance was measured to the centroid in the original data using the metric in 3: this was done both for the original data and for the data after 30,000 iterations. For younger speakers, an ANOVA with log-Euclidean distance as the dependent variable and with within-subjects factors vowel (three levels: /i, ju, u/) and condition (two levels: original data, 30,000 iterations) showed a significant effect for condition ($F(1,10) = 15.2$, $p < 0.01$) and for vowel ($F(1.9,19.2) = 8.2$, $p < 0.01$), but no significant interaction between these factors. The analysis with the same variables for older agents also showed a significant effect for condition ($F(1,10) = 28.0$, $p < 0.001$) and for vowel ($F(1.9,19.0) = 41.6$, $p < 0.01$), as well as a significant ($F(1.4,13.7) = 6.4$, $p < 0.05$) interaction between these factors. Post-hoc Bonferroni-corrected *t*-tests for the older agents showed, consistent with the bottom panel of Fig. 7, a significantly greater change between 30,000 iterations and the original data in /i/ ($t(10) = 6.2$, $p < 0.01$) than in /ju/ ($t(10) = 3.9$, $p < 0.05$) or /u/ ($t(10) = 4.2$, $p < 0.05$).

Thus, the overall conclusion is that, contrary to the hypothesis, there was indeed a significant change between the original data and after 30,000 iterations; but as Fig. 7 suggests, this change, while significant, was nevertheless modest, at least as far as the aggregated F2 trajectories are concerned.

SIMULATIONS WITH ALL AGENTS. Figure 8 shows F2 trajectories aggregated by condition and by vowel class for up to 50,000 iterations, in 10,000-iteration steps, when all twenty-two agents interacted with each other.

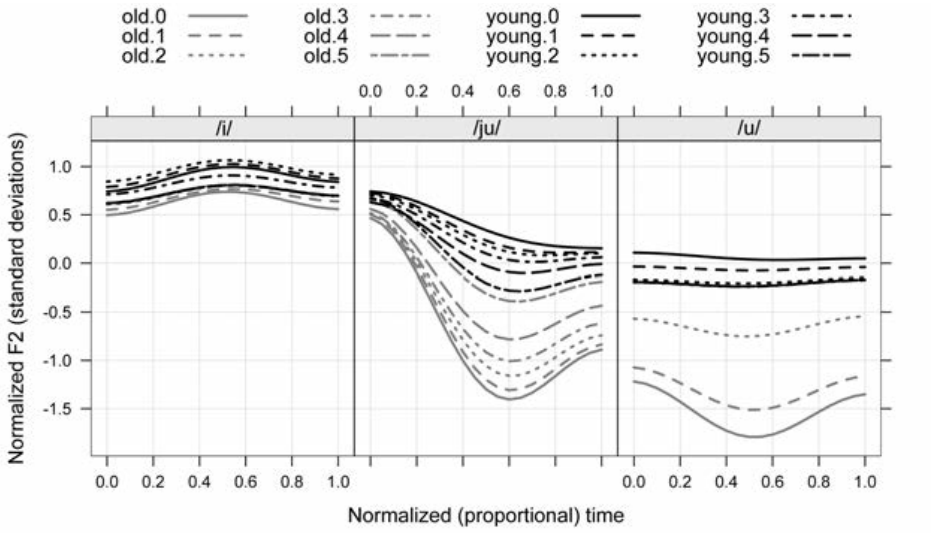

FIGURE 8. F2 trajectories aggregated in the original data (.0) and then at intervals of 10,000 (.1) iterations, up to 50,000 (.5) total, in the older ('old', gray lines) and younger ('young', black lines) agents for the three vowel classes.

Recall that the hypothesis to be tested was that the older agents' /ju, u/ should show a greater shift toward younger agents' /ju, u/ than the other way around. No such asymmetry was expected for /i/. There is very clearly support for these hypotheses from the data in Fig. 8: for /ju/, there was a relatively constant shift of the aggregated trajectory of older agents toward that of younger agents; for /u/ there was a large shift for older agents between the original data and 10,000 iterations, and then a convergence at higher iterations toward a lowered F2 trajectory of younger agents. For both /ju/ and /u/, there is strong evidence from Fig. 8 that the trajectories of older speakers at 50,000 iterations were more similar to those of younger agents' original trajectories than were younger agents' trajectories at 50,000 iterations similar to older agents' original trajectories.

In order to compare the relative size of the shift of the two age groups, the mean distance after 50,000 iterations to the centroid of the original data was calculated separately for each speaker and each word in the DCT space using 3 in §3.1. The hypothesis to be tested was that the distance was greater for the older than for the younger agents. The results of this calculation, summarized in Figure 9, show greater distances on this measure in /ju, u/ for the older agents. Consistent with Figs. 8 and 9, an ANOVA—with the log-Euclidean distance as the dependent variable and the within-subjects factor vowel (two levels: /ju, u/) and between-subjects factor age group (two levels: younger, older)—showed a significant effect for age group ($F(1,20) = 32.5$, $p < 0.001$), a nonsignificant effect for vowel, and a significant interaction between these factors ($F(1,20) = 33.4$, $p < 0.001$). Post-hoc Bonferroni-corrected t-tests showed significant differences in Euclidean distances between older and younger groups on both /ju/ ($t(19.3) = 3.4$, $p < 0.05$) and /u/ ($t(12.6) = 6.7$, $p < 0.001$).

Overall, the results thus show a greater approximation of the older toward the younger agents in /ju, u/.

3.3. DISCUSSION. When only younger agents or only older agents interacted with each other, there was a minor change to the positions of /i, ju, u/. Although these vowels had

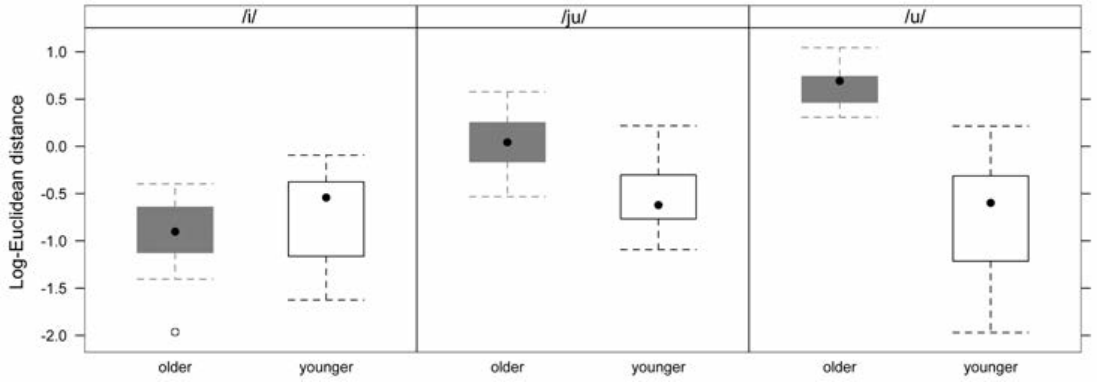

FIGURE 9. The log-Euclidean distance calculated using 3 in a DCT-space for the three vowel classes between the positions after 50,000 iterations and the word- and speaker-specific centroids in the original data. There is one value per agent in each of the boxplots.

shifted significantly relative to their original productions, the changes were small and within the range of variation of the original data. This initial study thus shows that change is not inevitable, or rather that there can be stability for populations in which the phonetic variation between speakers is small. By contrast, when the interactions were based on all twenty-two speakers across both age groups, there were asymmetric shifts such that the older speakers' /ju, u/ were fronted toward the positions of the younger speakers; by contrast, the degree of younger speakers' /ju, u/ retraction was comparatively small. There was no approximation toward either age group in /i/, which confirms that the asymmetric shift across the age groups in /ju, u/ is not an inevitable consequence of interaction as modeled here.

The asymmetry in /ju, u/ is likely to come about because, as the results of the first part of this study show, there was a corresponding asymmetry between the groups in the direction of variation in these vowels: older speakers' phonetically backed /ju, u/ distributions in the DCT acoustic space were oriented toward the fronted /ju, u/ distributions of younger speakers to a greater extent than the other way around. This asymmetry in orientation comes about because of the phonetic propensity for a retracted /u/ to front due to coarticulation and vowel undershoot: both of these effects have a strong tendency to raise the second formant frequency. This tendency for /u/ to front is the ORIGIN OF SOUND CHANGE, that is, the phonetic conditions for sound change to take place. The results of these simulations suggest that the PROPAGATION or SPREAD OF SOUND CHANGE amplifies just such conditions that can (but need not) give rise to sound change. The amplification through propagation comes about if the variation due to the origin of sound change in one group is in the same direction in which a phonetic variant of the other group is positioned.

Such a model could begin to explain why New Zealand English (NZE) neutralizes contrasts to /ə/ in prosodically weak final syllables, as exemplified by pairs such as *boxers*/*boxes*, *dancers*/*dances*, *Rosa's*/*roses*, which in SSB are differentiated by an /ə, ɪ/ opposition, respectively (i.e. *boxers*/*boxes* are both /bɒksəz/ in NZE but /bɒksəz/ ~ /bɒksɪz/, respectively, in SSB). In Trudgill's (2004) model, this is a puzzle because the majority of original settlers to New Zealand in the nineteenth century came from a dialect background that had an /ə, ɪ/ contrast, just as in present-day SSB, with only some 30% of speakers coming from dialects in which this contrast was neutralized to /ə/ as in present-day NZE; similarly, present-day SSB has /ɪ/ but present-day NZE /ə/ in the second syllable of words like *certainly*, *character*, *recipe*. The puzzle in Trudgill's deterministic model is that present-day NZE should have maintained an /ə, ɪ/ contrast, if the

outcome of the dialect mixture is determined by whichever form is produced by the majority. The agent-based simulations carried out here could provide an explanation for this anomaly. In varieties like SSB that make this /ə, ɪ/ distinction, /ɪ/ in weak syllables is more likely to centralize toward /ə/ due to vowel undershoot at faster rates of speech, or at least the shift from /ɪ/ to /ə/ due to fast or spontaneous speech is more likely than a change in the other direction from /ə/ to /ɪ/. As a result, the direction of variation of SSB /ɪ/ in weak syllables is toward /ə/. Consequently, according to the simulations presented here, if SSB speakers—or those from varieties that make the /ə, ɪ/ contrast—come into contact with speakers who only have /ə/ in weak syllables, then /ɪ/ should be pulled toward /ə/ analogously to how older agents' retracted /ju, u/ in the simulations of this study were pulled toward younger agents' fronted variants. That is, the conditions for sound change to occur (for /ɪ/ to shift toward /ə/) are amplified by interaction and imitation, precisely because the direction of synchronic variation in weak /ɪ/ is toward /ə/, or at least far more so than /ə/ is likely to be oriented synchronically toward /ɪ/.

There are, nevertheless, some problems with the implementation of the agent-based model that need to be addressed. The most pressing issue is that after a certain number of iterations, all of the agents converged on the same value for a given vowel class: that is, the variation that existed in the original data was progressively reduced to zero. More specifically, whereas the trajectory at starting conditions in Fig. 7 is an aggregate across different trajectories, with a high degree of variation between them, over an increasing number of iterations this variation was reduced, so that by 50,000 iterations, there was virtually no variation left in the data (i.e. all agents converged on the same trajectory). It could be argued in defense of this model that just this is the expected outcome in the theoretical case in which speakers accommodate to each other eventually: over a perhaps unrealistically large number of iterations, the speakers must all converge on the same phonetic values. The reason why this does not happen in reality is because this type of convergence is held in check by so many other types of variation that we have not modeled here: for example, the fact that not all speakers interact with each other to the same degree, that they meet other speakers from different social, dialect, and language backgrounds, that speakers have different vocabularies and vocabulary sizes, that the effect of imitation on children and adults may well be very different (e.g. Nielsen 2014), and that biological age differences influence acoustic output, to mention but a few. The most important finding from the modeling so far is that the agents, while admittedly converging on the same value per vowel class after a large number of iterations, did not meet in the middle: that is, there was a bias such that the convergence in /ju, u/ was significantly closer to the younger than older speakers' original data. Nevertheless, it is possible that the greater change from original positions in older agent-talkers is an artifact of the greater variation in /ju, u/ for the older than for the younger group: if the variation is greater, then the degree of change between the original positions and final convergence toward a single value may be caused by the greater spread (in addition to the orientation) of variation in the data for older than younger agents.

For this reason, the agent-based model was rerun but in such a way as to guarantee variation following interaction rather than convergence to a single value, by changing the type of memory loss following imitation. Recall in the earlier model that an item was discarded from an agent-listener's class prior to the uptake of another item from a different agent through interaction. Moreover, the discarded item was the one at the probabilistic edge of the agent-listener's distribution. Since there is no, or very little, cognitive evidence that memory changes in this way, a different model was tested by which the OLDEST items in memory were discarded prior to imitation, as in Kirby 2013.

A similar time-decay approach based on Pierrehumbert 2001 has been proposed by Wedel (2006), with the difference that in Wedel's model no exemplars are actually removed from memory, but their base-activation decreases over time. There is admittedly little cognitive evidence to support this time-decay model either (see Hay & Foulkes 2016 for a recent analysis and discussion of this issue), but in the absence of any definitive account of how the uptake of new information and discarding of old information from memory are related in speech communication, the procedure adopted here is instead to implement different types of memory loss in order to determine whether the same or similar results are obtained, as far as the hypothesis of asymmetric convergence in /ju, u/ across the age groups is concerned.

Finally, in the agent-based model run so far, the outcome has been documented for a single run of up to 50,000 iterations and in which eleven of fourteen younger speakers were randomly selected to match the number of speakers in the older group. It is possible therefore that the outcome is specific to this particular choice of eleven younger speakers and indeed to these 50,000 iterations. In order to ensure that the result is generalizable, we also reran the agent-based model 100 times (at 50,000 iterations each) with a different random selection of eleven of fourteen younger speakers on each run.

4. AGENT-BASED MODELING WITH VARIATION.

4.1. METHOD. The revised perception model was applied to simulations based on the same twenty-two agents (from both age groups, as described in §3.1 'Association between speakers and agents'). The methodology was exactly the same as in the original model in §3.1 'Communication between agents' except in two respects.

First, in the original model, an agent-listener only incorporated a perceived item, $item_{perc}$, into memory if the posterior probability of vowel-class membership was greater than chance, that is, greater than one in three. The uptake of $item_{perc}$ depended in this original model on how it was classified with respect to all three vowel classes in the lexicon. In the revised model presented here, $item_{perc}$ was only incorporated into memory if the Mahalanobis distance to the corresponding vowel class in the agent-listener's memory was under a certain threshold. For example, if the word class of $item_{perc}$ was *queued*, then $item_{perc}$ was incorporated into memory if the Mahalanobis distance of the DCT-triplet of $item_{perc}$ to the agent-listener's /ju/ class centroid was less than a threshold. This procedure (in contrast to the original model) therefore takes no account of the probabilistic distance to the other vowel classes (in this example to /i, u/) in the agent-listener's memory. The threshold in the revised model presented here was set corresponding to a cumulative probability of just over 0.99. Thus if the probability of vowel-class membership (based on the Mahalanobis distance calculation) of $item_{perc}$ was less than 0.01, then $item_{perc}$ was not incorporated into memory.

Second, and just as in the original model, an item of the same vowel and word class as $item_{perc}$ was removed from memory if $item_{perc}$ was incorporated into memory in the manner described above. In the original model, $item_{rem}$ (the item to be removed) was at the probabilistic margin of the agent-listener's vowel class. In the revised model, $item_{rem}$ was the oldest item in memory of the same vowel and word class. Word age was defined as follows. In initial conditions (i.e. prior to any interaction), a time stamp was randomly assigned to each of the (typically ten) items separately per agent and per word class. For example, prior to running the model, each of the agent's ten *queued* items was associated with a randomly assigned numerical time stamp. Any $item_{perc}$ incorporated into memory as a result of interaction was assigned a time stamp whose value was the maximum time stamp of any existing item in memory of the same word class plus one. The item removed

from memory after incorporating $item_{perc}$ was the one with the lowest time-stamp value (i.e. the oldest *queued* item in this example). Both of these modifications guaranteed that a certain degree of variation was maintained (i.e. that there was no convergence to a single point) irrespective of the number of iterations: this is because the main driving force for convergence in the model in §3—that is, the removal of items at the edge of the vowel distribution—was attenuated with this revised methodology.

In order to test for robustness of this revised model, we reran this revised agent-based model 100 times, where a single run was defined as a model of 50,000 iterations containing a random selection of eleven of the fourteen younger speakers and the eleven older speakers, and we tested whether across these 100 models there was a greater convergence of older agents' /ju, u/ to younger agents' distributions than the other way around.

4.2 Results. The aggregated means and variances for the two age groups and three vowel classes in Figure 10 show the same general trend that was observed in the original simulations: there was a greater shift in /ju, u/ for the older toward the younger group than the other way around.[3]

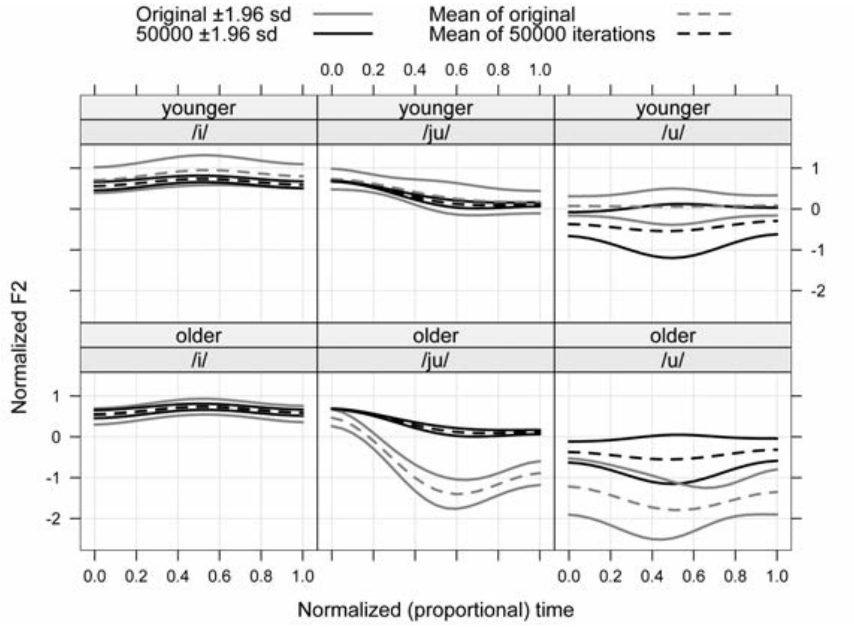

Figure 10. F2 trajectories aggregated by vowel class and speaker group (younger above, older below), shown separately for the original data (gray) and after 50,000 iterations (black) for the first run in the 100 runs (each consisting of 50,000 iterations) that took place. Dashed lines are the means and solid lines are at ±1.96 standard deviations from the mean.

The relative size of the shift of the two age groups was determined as before by comparing the mean distance after 50,000 iterations to the centroid of the original data for each speaker and each word separately in the DCT space, using 3 (§3.1 'Quantifying change'). Consistent with Fig. 10, an ANOVA with the log-Euclidean distance as dependent variable and with within-subjects factor vowel (two levels: /ju, u/) and between-subjects factor age group (two levels: younger, older) showed a significant effect for age group ($F(1,20) = 51.1$, $p < 0.001$), a significant effect for vowel ($F(1,20) = 25.8$,

[3] For animations of the change in the positions of the vowels in the DCT space for an increasing number of iterations, please see ftp://ftp.bas.uni-muenchen.de/pub/BAS/ABM/Animations/StartHere.html.

$p < 0.001$), and a significant interaction between these factors ($F(1,20) = 25.1$, $p < 0.001$). Post-hoc Bonferroni-corrected *t*-tests showed significant differences in Euclidean distances between older and younger groups for both /ju/ ($t(15.3) = 8.6$, $p < 0.001$) and /u/ ($t(19.5) = 4.4$, $p < 0.05$). Thus, once again there is evidence of a greater approximation of the older toward the younger agents in /ju, u/.

A similar set of results was obtained in aggregates across the 100 runs of the model. More specifically, after 100 runs there were 100 (runs) × 10 (repetitions) = 1,000 items per agent × word class × vowel class combination after 50,000 iterations. The dependent variable was the distance of these items aggregated by agent, word class, and vowel class to the centroid at starting conditions using 3 (§3.1 'Quantifying change'). An ANOVA with the same factors as above showed a significant effect for age group ($F(1,24) = 163.1$, $p < 0.001$) and vowel class ($F(1,24) = 90.9$, $p < 0.001$), and a significant interaction between these factors ($F(1,24) = 38.1$, $p < 0.001$). Post-hoc Bonferroni-corrected *t*-tests showed significant differences in Euclidean distances between older and younger groups for both /ju/ ($t(22.8) = 12.6$, $p < 0.001$) and /u/ ($t(22.5) = 7.7$, $p < 0.001$). Thus these results attest to the robustness of the findings of the asymmetric shift of the older toward the younger group of agents in /ju, u/.

5. General discussion. This has been one of first studies in the computational modeling of sound change in which the starting conditions for the agents in the model were equivalent to utterances produced by real speakers. The advantage of this approach is not only that the computational model bears a direct relationship to a sound change in progress (as analyzed within the apparent-time paradigm in Harrington et al. 2008 for this particular case), but also that the starting conditions are not, as in some other models, artificially generated, which in turn risks building into the model the desired outcome of the hypothesis to be tested. The second methodological advance of this study is that the agents in our model communicated dynamic parameters and not, as in almost all other computational studies on sound change, values at either a single time point or aggregated over a time window. Such an approach is more appropriate than one based on static snapshots of speech, not just because speech is an inherently dynamic activity, but also because so much of the synchronic variation due to coarticulation and reduction that is undeniably associated with sound change is itself a consequence of how layered articulatory movements unfold and are perceived in time (Beddor 2012, Browman & Goldstein 1992, Lin et al. 2014, Ohala 2012, Solé 2014).

The more specific finding of the present study is that the sound change that is the result of the interaction between two groups of different varieties depends not just on the groups' starting positions prior to contact, but also on the direction of variation in a phonetic space. This finding seems to be quite robust, given that the observed changes were replicated under two different sets of assumptions (in §§3 and 4) about how produced items were absorbed into an agent-listener's memory. Thus, for the present study, it is because the retracted /u/ of the older group was oriented toward the advanced /u/ of the younger group for reasons to do synchronically with coarticulation and undershoot that the outcome of intergroup interaction was an asymmetric shift of older toward younger speakers' /u/. More generally, we would suggest that the orientation of variation provides the conditions for one group to act as an attractor and pull the phonetic values of another group toward it. This, as argued earlier, can explain why, for example, in early New Zealand English there was a shift of unstressed /ɪ/ toward /ə/ in weak syllables in words like *dances*, even though the /ɪ/ may have had majority usage, at least as far as the early settlers were concerned.

The more general conclusion from this study is that, compatible with Bloomfield 1933, Labov 2001, and Trudgill 2004, 2008, certain types of sound change are not necessarily driven by social factors, such as prestige, that are more closely connected synchronically with style shifting, but that they may instead fall out from the system dynamics that defines a population of speakers, that is, out of the forces that are shaped by the position and orientation of the phonetic variation of the speakers that come into contact with each other. Future studies could show that factors such as gender and social class are less potent predictors of sound change and may instead be modeled as a deterministic consequence of speaker interaction; this idea remains largely untested, however, for the reason that sociolinguistic studies investigating the relationship between synchronic variation and diachronic change have not tended to base their analyses on the shape and orientation of dynamic phonetic variation in a multidimensional acoustic space.

An unresolved issue is whether there would also be a shift of the kind we have observed if the distribution of the older speakers were not necessarily skewed toward that of the younger speakers but just larger. If there are two distributions *A* and *B* analogous to those in Fig. 1 but such that neither distribution is skewed toward the other but *B* is larger, then *B* is likely to shift toward *A* in the computational model that we have implemented. Moreover, the distributions of both /ju/ and /u/ are indeed typically larger for older than for younger speakers in our data. This greater variation comes about because of coarticulation. For example, the differences in the first half of /u/ between *soup* and *food* are much greater for older than for younger speakers (because /s/ exerts a fronting influence on older speakers' otherwise retracted /u/). Furthermore, if an older speaker has started to participate in the sound change in progress by which /u/ has fronted diachronically, then such a speaker may produce /u/ with varying degrees of phonetic fronting (see e.g. Nolan et al. 2006 for relevant data on within-speaker variation and /u/-fronting in SSB). Whatever the source of variation, the sizes of the /ju, u/ distributions were typically larger for the older speakers in our study, and this greater size may also have been a contributory factor in their approximation toward younger speakers' /ju, u/ spaces in our computational simulations. A so-far-unresolved issue that we plan to address in the future is the relative potency of size vs. skew as far the influence of one distribution on another is concerned.

The overall conclusion from these simulations is that sound change comes about when synchronic phonetic variation is magnified by external group contact: this is the link we are proposing between the conditions that give rise to sound change and its spread through the community. The propagation of synchronic variation around a community of speakers need not (and usually does not) result in sound change (Baker et al. 2011, Kirby 2014, Sóskuthy 2015): this type of stability was demonstrated in the present study in the simulations carried out within each age group separately, in which the phonetic variation across the speakers was comparatively small.

Some have suggested (e.g. Baker et al. 2011, Garrett & Johnson 2013) that the conversion of synchronic variation into sound change requires innovative speakers who more frequently produce strongly coarticulated variants or who carry this variation over into other contexts beyond the one that induces coarticulation. This is entirely consistent with the simulations carried out in this study: sound change was shown to occur when speakers were included with phonetic variants that lie along the trajectory in which variation due to coarticulation/undershoot takes place. Moreover, although there was overall a significant shift in older agents' /ju, u/ when the model with all agents together was run 100 times (at 50,000 iterations each; see §4.2), there was also some vari-

ation in both the size and direction of shift between the separate runs. More specifically, while in 88/100 runs there was a very clear shift of older toward younger agents' /ju, u/ in the manner shown in Fig. 8, for the remaining twelve runs, there was either convergence toward a form intermediate—that is, at roughly an average position—between older and younger agents' starting positions, or else there was a greater shift of younger agents' /ju, u/ toward the positions of older speakers. This variation even within this very small number of agents, entirely balanced between those who do and do not front /ju, u/, demonstrates that sound change is not deterministic in the sense of being inevitable when two or more dialect groups come into contact, but is instead stochastically influenced by factors such as which speakers come into contact with each other, how often they do so, and whether upon contact a produced item is absorbed and then subsequently retained in the perceiver's memory.

The computational model presented in this study does not take account of numerous recent findings showing that phonological categorization is often adapted to social factors (see e.g. Docherty & Foulkes 2014 for a recent review and Munson et al. 2012 with respect to language acquisition). For example, studies showing that the same acoustic signal is differently categorized depending on the listener's beliefs about the speaker's social-indexical attributes (e.g. Hay & Drager 2010, Hay et al. 2006, Jannedy & Weirich 2014, Niedzielski 1999) are certainly strong evidence that phonological processing is adapted to speaker-specific attributes. In order to establish a direct relationship to sound change, however, it would have to be demonstrated that such effects of social beliefs on phonological categorization influence speech production; there is not a great deal of evidence that this is so (but see the discussion on speech production in Drager & Kirtley 2016). By contrast, there is much evidence that perception influences production as a consequence of interaction between speakers, as outlined in §1. Accordingly, just this perception-production association forms part of our computational model in which agent-listeners absorb perceived items (depending on various probabilistic and memory criteria) that subsequently indirectly influence their production output (see §3 'Agent-based modeling'). The cognitive analogue of such a computational implementation is that adaptation in production to a social-indexical variable is entirely a by-product of how often a person interacts with interlocutors who have just those social characteristics. Alternatively, attitudes toward groups or speakers that have certain social-indexical attributes (Drager 2011), combined with the tuning of so-called attentional weights (Nosofsky 1992) toward social representations that are salient to the listener, possibly because the listener judges such characteristics to be positive (Drager et al. 2010), may also have some influence on speech production and sound change. Such influences are not currently incorporated into our computational model.

A further question is whether contact between groups with different variants (or the inclusion of innovative speakers who frequently produce strongly coarticulated variants) is a PREREQUISITE for synchronic variation to become sound change. An alternative possibility is suggested by Ohala's (1993, 2012) model, in which the conditions for sound change are met when a listener parses coarticulation perceptually in a way that is inconsistent with coarticulation in production. As far as synchronic /u/ variation is concerned, experiments show that listeners make decisions about whether an ambiguous signal is /i/ or /u/ in a context-dependent way (Harrington et al. 2008, Lindblom & Studdert-Kennedy 1967, Ohala & Feder 1994): since in speech production, /u/ is fronted following alveolars (*do*) and palatals (*few*), then listeners accordingly shift their perceptual boundary toward /i/ and are more likely to categorize an acoustic signal as /u/ in these consonantally fronting contexts than in other nonfronting contexts (*food*). Thus

stable conditions—that is, when there is no sound change—are likely to be associated with parity (Fowler 2005) between how coarticulation is produced and perceived (Harrington, Kleber, & Stevens 2016). In the agent-based models of the present study, the decision about whether an agent-listener absorbs a perceived item into memory was context-dependent: this was because the signal of a perceived *queued* produced by another agent was only incorporated into memory if the signal was probabilistically close to the words in the lexicon with vowel class /ju/ (*queued*, *feud*, *hewed*) and different from the agent-listener's vowel class /u/ (*cooed*, *food*, *soup*, *who'd*). Although different in many respects, we can begin to approximate a scenario of not normalizing or compensating for coarticulation sufficiently in perception—which is argued to be one of the conditions for sound change to take place (Ohala 1981, 1993, 2012)—by having the agent-listeners base their decisions on a COMBINED /ju, u/ model. When in such a combined model an agent-listener perceives *queued*, the decision about whether the signal is absorbed into memory would be based on the probabilistic distance to the distribution of the combined /ju, u/ items in the agent's memory, rather than just on the probabilistic distance to the agent's /ju/ items, as in the context-dependent manner of the modeling in §§3 and 4.

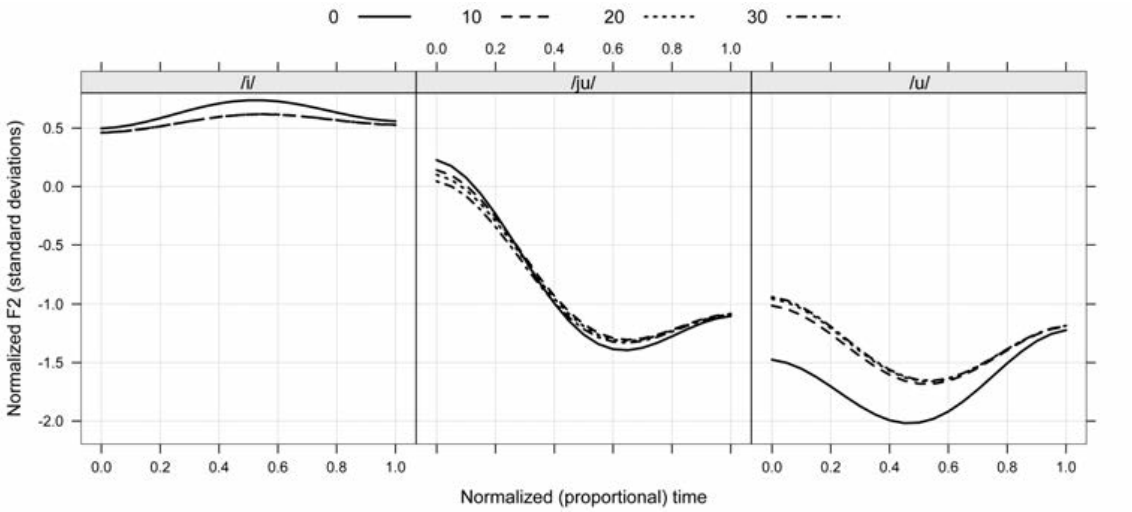

FIGURE 11. F2 trajectories aggregated in the original data (0) and then at intervals of 10,000 (10) iterations, up to 30,000 (30) total, in the older agents for the three vowel classes, based on collapsing /ju, u/ classification into a single class in agent perception.

We ran just such a combined model using only the eleven agents from the older group. The hypothesis was that /u/ (in *food*, *cooed*, *soup*, *who'd*) should shift more in the direction of /ju/ (*feud*, *queued*, *hewed*) than the other way around. The reason for this asymmetry is the same as before: there is an inherent synchronic bias for /u/-words to front phonetically due to coarticulation (especially in *soup*) and undershoot, rather than for the /ju/-words to retract. The results of this simulation for the eleven agents in the older group show some evidence in support of this hypothesis (Figure 11). In particular, although there was some downward flattening of the F2 trajectory of /ju/, there was no change in /ju/'s F2 minimum; and in addition there was a marked F2 raising of /u/ throughout the trajectory and particularly so in the first half of the vowel. The shift was small—much smaller than in the simulations based on group contact with the younger agents—but it does show that there is the potential for sound change to occur within a phonetically homogeneous group that might then come to be further magnified by external group contact (or by innovative speakers in the sense defined earlier). We

emphasize that this modification to our computational approach is by no means an exact implementation of Ohala's (1993) model, although there are parallels. First, there is an abrupt change in both Ohala's and our model: in Ohala's, the abrupt change is in category representation (if the listener fails to normalize for the effects of context), while in our model it is that two variants that were initially represented by separate probability models (that were used for categorization of perceived items) are merged. Second, Ohala (1993:266) suggests that the gradual spread of sound change is likely to come about INTER ALIA because of its spread from one speaker to another. Compatible with this idea of gradualness, the merger of what were initially separate probability models for front and retracted variants of /u/ subsequently has an incremental, gradual influence on speech production in our model that is brought about by the interaction between agents.

The simulations in this article represent only a very first step in building a computational model to link the conditions that give rise to sound change and the reinforcement of these conditions, which can lead to sound change through propagation by imitation around a community of speakers. We recognize that there are very many ways in which these simulations need to be augmented in order to develop a more comprehensive model of sound change. One of these is by changing the lexicon to include a more representative vocabulary size and one that takes into account factors such as lexical frequency and neighborhood density, which have been shown to be important in the development of sound change (e.g. Hay & Foulkes 2016, Hay et al. 2015, Lin et al. 2014). Another is by changing the network links between speakers to test hypotheses about whether the course of sound change is affected by varying the degree to which agents are connected with other agents (e.g. Fagyal et al. 2010, Pierrehumbert et al. 2014). A further limitation is that, since the agents do not make mistakes in the transmission of word or vowel classes, there is currently no possibility of modeling mergers when two phonological categories collapse into one (Hay et al. 2006, Milroy & Harris 1980). There is also no sense in which these simulations have provided a solution to the problem of how phonologization is associated with the waning of the source that gave rise to it, as in the development of umlaut in German in which the source, for example, the /i/ in Proto-Germanic /fotiz/, has centralized to a neutral vowel with the fronting (and subsequent raising) of the first vowel to give present-day German /fysə/ (*Füße* 'feet'). The model is, however, extendable to incorporate all of these issues. Moreover, a model such as the one proposed here that is grounded in speech dynamics is above all well suited to testing hypotheses about how phonologization and the neutralization of the source might be reinforced as a consequence of agents that exchange dynamic coarticulatory information.

APPENDIX

The Mahalanobis distance (Duda et al. 2001), which is used by the agent-listener in classifying a perceived item as one of three classes, is a warping of the Euclidean distance to take account of the distribution (spread and direction in a multidimensional space) of class members.

Consider, for example, two classes with members shown by the gray and black circles in the left panel of Figure A1 in a two-dimensional space. (With respect to the present study, the gray and black circles might be /i, u/, respectively). There is additionally an *X* in the figure, which is the position of an unknown item (which might be the item that is perceived by the agent-listener). The problem is to decide which of the two classes *X* is most likely to belong to. A simple metric might be based on the Euclidean (straight-line) distance to the center of the two distributions. On this basis, *X* would be assigned to the black class, since it has a much smaller Euclidean distance to the center of the black than to the gray circles (Fig. A1, left panel). However, this simple Euclidean metric does not take account of the pattern of distribution of the classes: in particular, *X* falls more readily along the direction of greatest variation for the gray (at roughly an angle of about -60° to

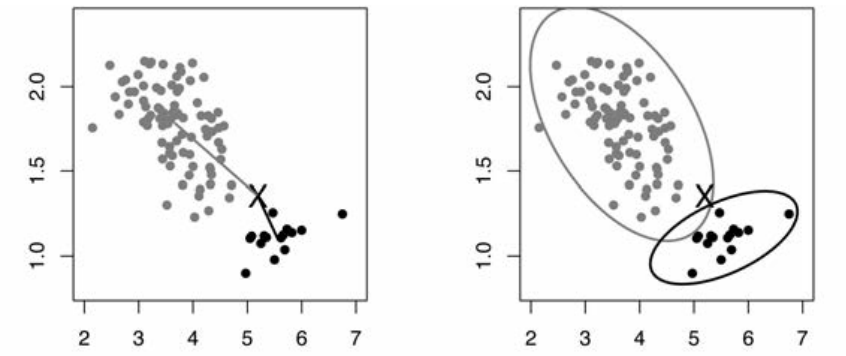

FIGURE A1. Two classes (gray, black) and an unknown item (*X*) in a two-dimensional space (with arbitrary values). The lines in the plot on the left are the Euclidean distances to the center (the mean on both dimensions) of these two classes. The right panel shows the same data but with 99% confidence ellipses (corresponding to 3.0348 ellipse standard deviations) superimposed on the data.

the horizontal) than for the black (an angle of around +30°) distributions. The Mahalanobis distance is designed to take account of such differences in the spread and orientation of variation of class membership.

A Mahalanobis distance is equivalent to the number of ellipse standard deviations. Any ellipse is, in turn, a horizontal slice through a two-dimensional Gaussian distribution.

In the right panel of Fig. A1, the gray and black ellipses have been constructed to have the same number of ellipse standard deviations corresponding in each case to a posterior probability of 0.99 (which also typically means that at least 99% of the actual data points should fall inside each ellipse). This means that any point positioned directly on either of the ellipses has a probability of belonging to that class of 0.99.

The same figure also shows that *X* falls within the gray but not the black ellipse. Therefore, the number of ellipse standard deviations—and hence the Mahalanobis distance—of *X* is less to the gray than to the black class. Therefore, *X* is classified as a member of the gray class based on the Mahalanobis distance, in contrast to the Euclidean distance.

The relationship between the Mahalanobis distance and posterior probability for a two-dimensional space is given by quantiles of the χ^2 distribution with two degrees of freedom. Thus the Mahalanobis distance of any point situated on either of the ellipses in Fig. A1 is given in the R programming environment by sqrt(qchisq(0.99, 2)), which calculates to 3.0348 (equivalently then, both ellipses in Fig. A1 are drawn at 3.0348 ellipse standard deviations).

The above example for a two-dimensional space is applicable to higher dimensions (such as the three-dimensional space used in this article, in which case the classification of an unknown item was based on three-dimensional ellipsoids rather than two-dimensional ellipses as in this example).

REFERENCES

BABEL, MOLLY. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* 40.177–89. DOI: 10.1016/j.wocn.2011.09.001.

BABEL, MOLLY; GRANT MCGUIRE; SOPHIA WALTERS; and ALICE NICHOLLS. 2014. Novelty and social preference in phonetic accommodation. *Laboratory Phonology* 5.123–50. DOI: 10.1515/lp-2014-0006.

BAILEY, GUY; TOM WIKLE; JAN TILLERY; and LORI SAND. 1991. The apparent time construct. *Language Variation and Change* 3.241–64. DOI: 10.1017/S0954394500000569.

BAKER, ADAM; DIANA ARCHANGELI; and JEFF MIELKE. 2011. Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change* 23.347–74. DOI: 10.1017/S0954394511000135.

BAXTER, GARETH J.; RICHARD A. BLYTHE; WILLIAM CROFT; and ALAN J. MCKANE. 2009. Modeling language change: An evaluation of Trudgill's theory of the emergence of New Zealand English. *Language Variation and Change* 21.257–96. DOI: 10.1017/S095439450999010X.

BECKMAN, MARY E.; FANGFANG LI; EUN JONG KONG; and JAN EDWARDS. 2014. Aligning the timelines of phonological acquisition and change. *Laboratory Phonology* 5.151–94. DOI: 10.1515/lp-2014-0007.

BEDDOR, PATRICE SPEETER. 2009. A coarticulatory path to sound change. *Language* 85. 785–821. DOI: 10.1353/lan.0.0165.

BEDDOR, PATRICE SPEETER. 2012. Perception grammars and sound change. *The initiation of sound change: Perception, production, and social factors*, ed. by Maria-Josep Solé and Daniel Recasens, 37–55. Amsterdam: John Benjamins.

BEDDOR, PATRICE SPEETER; JAMES D. HARNSBERGER; and STEPHANIE LINDEMANN. 2002. Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics* 30.591–627. DOI: 10.1006/jpho.2002.0177.

BLEVINS, JULIETTE, and ANDREW WEDEL. 2009. Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica* 26.143–83. DOI: 10.1075/dia.26.2.01ble.

BLOOMFIELD, LEONARD. 1933. *Language*. New York: Holt.

BROWMAN, CATHERINE, and LOUIS GOLDSTEIN. 1992. Articulatory phonology: An overview. *Phonetica* 49.155–80. DOI: 10.1159/000261913.

BYBEE, JOAN L., and CLAY BECKNER. 2010. Usage-based theory. *The Oxford handbook of linguistic analysis*, ed. by Bernd Heine and Heiko Narrog, 827–56. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199544004.013.0032.

CANNADINE, DAVID. 1998. *Class in Britain*. New Haven, CT: Yale University Press.

CHITORAN, IOANA, and JOSÉ IGNACIO HUALDE. 2007. From hiatus to diphthong: The evolution of vowel sequences in Romance. *Phonology* 24.37–55. DOI: 10.1017/S095267570700111X.

CLOPPER, CYNTHIA G. 2014. Sound change in the individual: Effects of exposure on cross-dialect speech processing. *Laboratory Phonology* 5.69–90. DOI: 10.1515/lp-2014-0004.

COX, FELICITY, and SALLYANNE PALETHORPE. 2001. The changing face of Australian English vowels. *English in Australia* (Varieties of English around the world G26), ed. by David Blair and Peter Collins, 17–44. Amsterdam: John Benjamins.

DELVAUX, VÉRONIQUE, and ALAIN SOQUET. 2007. The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica* 64.145–73. DOI: 10.1159/000107914.

DOCHERTY, GERARD J., and PAUL FOULKES. 2014. An evaluation of usage-based approaches to the modelling of sociophonetic variability. *Lingua* 142.42–56. DOI: 10.1016/j.lingua.2013.01.011.

DRAGER, KATIE. 2011. Sociophonetic variation and the lemma. *Journal of Phonetics* 39.694–707. DOI: 10.1016/j.wocn.2011.08.005.

DRAGER, KATIE; JENNIFER HAY; and ABBY WALKER. 2010. Pronounced rivalries: Attitudes and speech production. *Te Reo* 53.27–53.

DRAGER, KATIE, and JOELLE KIRTLEY. 2016. Awareness, salience, and stereotypes in exemplar-based models of speech production and perception. *Awareness and control in sociolinguistic research*, ed. by Anna M. Babel, 1–24 . Cambridge: Cambridge University Press.

DUDA, RICHARD O.; PETER E. HART; and DAVID G. STORK. 2001. *Pattern classification*. 2nd edn. New York: John Wiley & Sons.

ECKERT, PENELOPE. 1989. The whole woman: Sex and gender differences in variation. *Language Variation and Change* 1.245–67. DOI: 10.1017/S095439450000017X.

FAGYAL, ZSUZSANNA; SAMARTH SWARUP; ANNA MARIA ESCOBAR; LES GASSER; and KIRAN LAKKARAJU. 2010. Centers and peripheries: Network roles in language change. *Lingua* 120.2061–79. DOI: 10.1016/j.lingua.2010.02.001.

FOWLER, CAROL A. 2005. Parsing coarticulated speech in perception: Effects of coarticulation resistance. *Journal of Phonetics* 33.199–213. DOI: 10.1016/j.wocn.2004.10.003.

FRIDLAND, VALERIE. 2008. Patterns of /uw/, /ʊ/, and /ow/ fronting in Reno, Nevada. *American Speech* 83.432–54. DOI: 10.1215/00031283-2008-030.

GARRETT, ANDREW, and KEITH JOHNSON. 2013. Phonetic bias in sound change. *Origins of sound change: Approaches to phonologization*, ed. by Alan C. L. Yu, 51–97. Oxford: Oxford University Press.

GILBERT, NIGEL. 2007. *Agent-based models*. Los Angeles: Sage.

GORDON, ELIZABETH; LYLE CAMPBELL; JENNIFER HAY; MARGARET MACLAGAN; ANDREA SUDBURY; and PETER TRUDGILL. 2004. *New Zealand English: Its origins and evolution*. Cambridge: Cambridge University Press.

HARRINGTON, JONATHAN; PHIL HOOLE; FELICITAS KLEBER; and ULRICH REUBOLD. 2011. The physiological, acoustic, and perceptual basis of high back vowel fronting: Evi-

dence from German tense and lax vowels. *Journal of Phonetics* 39.121–31. DOI: 10.1016/j.wocn.2010.12.006.

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

Harrington, Jonathan; Felicitas Kleber; Ulrich Reubold; and Mary Stevens. 2016. The relevance of context and experience for the operation of historical sound change. *Toward robotic socially believable behaving systems, vol. 2: Modeling social signals*, ed. by Anna Esposito and Lakhmi C. Jain, 61–85. Basel: Springer.

Harrington, Jonathan; Felicitas Kleber; and Mary Stevens. 2016. The relationship between the (mis)-parsing of coarticulation in perception and sound change: Evidence from dissimilation and language acquisition. *Recent advances in nonlinear speech processing*, ed. by Anna Esposito, Marcos Faundez-Zanuy, Antonietta M. Esposito, Gennaro Cordasco, Thomas Drugman, Jordi Solé-Casals, and Francesco Carlo Morabito, 15–34. Basel: Springer.

Harrington, Jonathan; Sallyanne Palethorpe; and Catherine I. Watson. 2000. Does the Queen speak the Queen's English? *Nature* 408.927–28. DOI: 10.1038/35050160.

Hawkins, Sarah, and Jonathan Midgley. 2005. Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association* 35.183–99. DOI: 10.1017/S0025100305002124.

Hay, Jennifer, and Katie Drager. 2010. Stuffed toys and speech perception. *Linguistics* 48.865–92. DOI: 10.1515/ling.2010.027.

Hay, Jennifer, and Paul Foulkes. 2016. The evolution of medial /t/ over real and remembered time. *Language* 92.298–330. DOI: 10.1353/lan.2016.0036.

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

Hay, Jennifer; Paul Warren; and Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34.458–84. DOI: 10.1016/j.wocn.2005.10.001.

Henton, Caroline. 1983. Changes in the vowels of received pronunciation. *Journal of Phonetics* 11.353–71.

Hombert, Jean-Marie; John J. Ohala; and William G. Ewan. 1979. Phonetic explanations for the development of tones. *Language* 55.37–58. DOI: 10.2307/412518.

Jacewicz, Ewa, and Robert Allen Fox. 2012. The effects of cross-generational and cross-dialectal variation on vowel identification and classification. *The Journal of the Acoustical Society of America* 131.1413–33. DOI: 10.1121/1.3676603.

Jannedy, Stefanie, and Melanie Weirich. 2014. Sound change in an urban setting: Category instability of the palatal fricative in Berlin. *Laboratory Phonology* 5.91–122. DOI: 10.1515/lp-2014-0005.

Jespersen, Otto. 1909. *A modern English grammar on historical principles, vol. 1: Sounds and spelling*. London: Allen and Unwin.

Kerswill, Paul. 2003. Dialect levelling and geographical diffusion in British English. *Social dialectology: In honour of Peter Trudgill*, ed. by David Britain and Jenny Cheshire, 223–43. Amsterdam: John Benjamins.

Kirby, James P. 2013. The role of probabilistic enhancement in phonologization. *Origins of sound change: Approaches to phonologization*, ed. by Alan C. L. Yu, 228–46. Oxford: Oxford University Press.

Kirby, James P. 2014. Incipient tonogenesis in Phnom Penh Khmer: Computational studies. *Laboratory Phonology* 5.195–230. DOI: 10.1515/lp-2014-0008.

Kirby, James P., and Morgan Sonderegger. 2013. A model of population dynamics applied to phonetic change. *Proceedings of the 35th annual conference of the Cognitive Science Society (CogSci 2013)*, 776–81. Online: https://mindmodeling.org/cogsci2013/papers/0160/.

Labov, William. 1990. The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2.205–54. DOI: 10.1017/S0954394500000338.

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* 83.344–87. DOI: 10.1353/lan.2007.0082.
LIN, SUSAN; PATRICE SPEETER BEDDOR; and ANDRIES COETZEE. 2014. Gestural reduction, lexical frequency, and sound change: A study of post-vocalic /l/. *Laboratory Phonology* 5.9–36. DOI: 10.1515/lp-2014-0002.
LINDBLOM, BJÖRN; SUSAN GUION; SUSAN HURA; SEUNG-JAE MOON; and RAQUEL WILLERMAN. 1995. Is sound change adaptive? *Rivista di Linguistica* 7.5–36.
LINDBLOM, BJÖRN, and MICHAEL STUDDERT-KENNEDY. 1967. On the role of formant transitions in vowel recognition. *The Journal of the Acoustical Society of America* 42.830–43. DOI: 10.1121/1.1910655.
MCDOUGALL, KERSTY, and FRANCIS NOLAN. 2007. Discrimination of speakers using the formant dynamics of /u:/ in British English. *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)*, Saarbrücken, 1825–28. Online: http://www.icphs2007.de/conference/Papers/1567/.
MILROY, JAMES, and JOHN HARRIS. 1980. When is a merger not a merger? The MEAT/MATE problem in a present-day English vernacular. *English World-Wide* 1.199–210. DOI: 10.1075/eww.1.2.03mil.
MILROY, JAMES, and LESLEY MILROY. 1993. *Real English: The grammar of English dialects in the British Isles*. London: Longman.
MUNSON, BEN; MARY BECKMAN; and JAN EDWARDS. 2012. Abstraction and specificity in early lexical representations: Climbing the ladder of abstraction. *The Oxford handbook of laboratory phonology*, ed. by Abigail Cohn, Cécile Fougeron, and Marie Huffman, 288–309. Oxford: Oxford University Press.
NIEDZIELSKI, NANCY. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18.62–85. DOI: 10.1177/0261927X99018001005.
NIELSEN, KUNIKO. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39.132–42. DOI: 10.1016/j.wocn.2010.12.007.
NIELSEN, KUNIKO. 2014. Phonetic imitation by young children and its developmental changes. *Journal of Speech, Language, and Hearing Research* 57.2065–75. DOI: 10.1044/2014_JSLHR-S-13-0093.
NOLAN, FRANCIS; KIRSTY MCDOUGALL; GEA DE JONG; and TOBY HUDSON. 2006. A forensic phonetic study of 'dynamic' sources of variability in speech: The DyViS Project. *Proceedings of the 11th Australian International Conference on Speech Science and Technology*, 13–18.
NOSOFSKY, ROBERT. 1992. Exemplars, prototypes, and similarity rules. *From learning theory to connectionist theory: Essays in honor of William K. Estes*, ed. by Alice Healy, Stephen Kosslyn, and Richard Shiffrin, 149–67. Hillsdale, NJ: Lawrence Erlbaum.
OHALA, JOHN J. 1981. The listener as a source of sound change. *Chicago Linguistic Society (Parasession on language and behavior)* 17(2).178–203.
OHALA, JOHN J. 1993. The phonetics of sound change. *Historical linguistics: Problems and perspectives*, ed. by Charles Jones, 237–78. London: Longman.
OHALA, JOHN J. 2012. The listener as a source of sound change: An update. *The initiation of sound change: Perception, production, and social factors*, ed. by Maria-Josep Solé and Daniel Recasens, 21–36. Amsterdam: John Benjamins.
OHALA, JOHN J., and DEBORAH FEDER. 1994. Listeners' normalization of vowel quality is influenced by 'restored' consonantal context. *Phonetica* 51.111–18. DOI: 10.1159/000261963.
PARDO, JENNIFER S.; RACHEL GIBBONS; ALEXANDRA SUPPES; and ROBERT M. KRAUSS. 2012. Phonetic convergence in college roommates. *Journal of Phonetics* 40.190–97. DOI: 10.1016/j.wocn.2011.10.001.
PARDO, JENNIFER S.; ISABEL CAJORI JAY; and ROBERT M. KRAUSS. 2010. Conversational role influences speech imitation. *Attention, Perception, & Psychophysics* 72.2254–64. DOI: 10.3758/BF03196699.

PIERREHUMBERT, JANET B. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. *Frequency effects and the emergence of linguistic structure*, ed. by Joan L. Bybee and Paul J. Hopper, 137–57. Amsterdam: John Benjamins.
PIERREHUMBERT, JANET B. 2002. Word-specific phonetics. *Laboratory phonology 7*, ed. by Carlos Gussenhoven and Natasha Warner, 101–39. Berlin: Mouton de Gruyter.
PIERREHUMBERT, JANET B. 2003a. Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech* 46.115–54. DOI: 10.1177/00238309030460020501.
PIERREHUMBERT, JANET B. 2003b. Probabilistic phonology: Discrimination and robustness. *Probabilistic linguistics*, ed. by Rens Bod, Jennifer Hay, and Stefanie Jannedy, 177–228. Cambridge, MA: MIT Press.
PIERREHUMBERT, JANET B. 2006. The next toolkit. *Journal of Phonetics* 34.516–30. DOI: 10.1016/j.wocn.2006.06.003.
PIERREHUMBERT, JANET B.; FORREST STONEDAHL; and ROBERT DALAND. 2014. A model of grassroots changes in linguistic systems. arXiv:1408.1985v1 [cs.CL]. Online: https://arxiv.org/abs/1408.1985.
QUENÉ, HUGO. 2013. Longitudinal trends in speech tempo: The case of Queen Beatrix. *The Journal of the Acoustical Society of America* 133.452–57. DOI: 10.1121/1.4802892.
SANKOFF, GILLIAN, and HÉLÈNE BLONDEAU. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83.560–88. DOI: 10.1353/lan.2007.0106.
SEBANZ, NATALIE; HAROLD BEKKERING; and GÜNTHER KNOBLICH. 2006. Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences* 10.70–76. DOI: 10.1016/j.tics.2005.12.009.
SHOCKLEY, KEVIN; DANIEL C. RICHARDSON; and RICK DALE. 2009. Conversation and coordinative structures. *Topics in Cognitive Science* 1.305–19. DOI: 10.1111/j.1756-8765.2009.01021.x.
SILVERMAN, DANIEL. 2006. The diachrony of labiality in Trique, and the functional relevance of gradience and variation. *Laboratory phonology 8*, ed. by Louis Goldstein, D. H. Whalen, and Catherine T. Best, 133–54. Berlin: Mouton de Gruyter.
SOLÉ, MARIA-JOSEP. 2014. The perception of voice-initiating gestures. *Laboratory Phonology* 5.37–68. DOI: 10.1515/lp-2014-0003.
SÓSKUTHY, MÁRTON. 2015. Understanding change through stability: A computational study of sound change actuation. *Lingua* 163.40–60. DOI: 10.1016/j.lingua.2015.05.010.
STANFORD, JAMES N., and LAURENCE A. KENNY. 2013. Revisiting transmission and diffusion: An agent-based model of vowel chain shifts across large communities. *Language Variation and Change* 25.119–53. DOI: 10.1017/S0954394513000069.
TRAUNMÜLLER, HARTMUT. 1990. Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America* 88.97–100. DOI: 10.1121/1.399849.
TRUDGILL, PETER. 1999. Dialect contact, dialectology and sociolinguistics. *Cuadernos de Filología Inglesa* 8.1–8.
TRUDGILL, PETER. 2004. *Dialect contact and new-dialect formation: The inevitability of colonial Englishes*. Edinburgh: Edinburgh University Press.
TRUDGILL, PETER. 2008. Colonial dialect contact in the history of European languages: On the irrelevance of identity to new-dialect formation. *Language in Society* 37.241–54. DOI: 10.1017/S0047404508080287.
TRUDGILL. PETER. 2010. Contact and sociolinguistic typology. *The handbook of language contact*, ed. by Raymond Hickey, 299–319. Oxford: Wiley-Blackwell.
TRUDGILL, PETER; ELIZABETH GORDON; GILLIAN LEWIS; and MARGARET MACLAGAN. 2000. Determinism in new-dialect formation and the genesis of New Zealand English. *Journal of Linguistics* 36.299–318.
VAN SON, ROB J. J. H., and LOUIS C. W. POLS. 1990. Formant frequencies of Dutch vowels in a text, read at normal and fast rate. *The Journal of the Acoustical Society of America* 88.1683–93. DOI: 10.1121/1.400243.
WATSON, CATHERINE, and JONATHAN HARRINGTON. 1999. Acoustic evidence for dynamic formant trajectories in Australian English vowels. *The Journal of the Acoustical Society of America* 106.458–68. DOI: 10.1121/1.427069.
WEDEL, ANDREW B. 2006. Exemplar models, evolution and language change. *The Linguistic Review* 23.247–74. DOI: 10.1515/TLR.2006.010.
WEDEL, ANDREW B. 2007. Feedback and regularity in the lexicon. *Phonology* 24.147–85. DOI: 10.1017/S0952675707001145.

WEINREICH, URIEL; WILLIAM LABOV; and MARVIN HERZOG. 1968. Empirical foundations for a theory of language change. *Directions for historical linguistics*, ed. by Winfred Lehmann and Yakov Malkiel, 95–105. Austin: University of Texas Press.
WOLFRAM, WALT, and NATALIE SCHILLING-ESTES. 2003. Dialectology and diffusion. *The handbook of historical linguistics*, ed. by Brian D. Joseph and Richard D. Janda, 713–35. Oxford: Blackwell.

[jmh@phonetik.uni-muenchen.de] (Harrington)
[schiel@phonetik.uni-muenchen.de] (Schiel)

[Received 27 November 2015;
accepted pending revisions 22 June 2016;
revision received 2 August 2016;
accepted 11 August 2016]