This article is part of the topic "Sociolinguistic Variation and Cognitive Science," Jean-Pierre Chevrot, Katie Drager and Paul Foulkes (Topic Editors). For a full listing of topic papers, see http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1756-8765/earlyview.

Linking Cognitive and Social Aspects of Sound Change Using Agent-Based Modeling

Jonathan Harrington, Felicitas Kleber, Ulrich Reubold, Florian Schiel, Mary Stevens

Institute for Phonetics and Speech Processing (IPS), Ludwig-Maximilians University of Munich

Abstract

The paper defines the core components of an interactive-phonetic (IP) sound change model. The starting point for the IP-model is that a phonological category is often skewed phonetically in a certain direction by the production and perception of speech. A prediction of the model is that sound change is likely to come about as a result of perceiving phonetic variants in the direction of the skew and at the probabilistic edge of the listener's phonological category. The results of agent-based computational simulations applied to the sound change in progress, /u/-fronting in Standard Southern British, were consistent with this hypothesis. The model was extended to sound changes involving splits and mergers by using the interaction between the agents to drive the phonological reclassification of perceived speech signals. The simulations showed no evidence of any acoustic change when this extended model was applied to Australian English data in which /s/ has been shown to retract due to coarticulation in /str/ clusters. Some agents nevertheless varied in their phonological categorizations during interaction between /str/ and /ʃtr/: This vacillation may represent the potential for sound change to occur. The general conclusion is that many types of sound change are the outcome of how phonetic distributions are oriented with respect to each other, their association to phonological classes, and how these types of information vary between speakers that happen to interact with each other.

1. Introduction

Sound change can provide a useful framework for considering the main theme of this special edition: how social variation and cognition are related. This is so because there are evidently social and cognitive sides to understanding sound change, but these have until fairly recently been pursued from their separate disciplinary perspectives (Harrington & Stevens, 2014), stemming in part from the artificial division that is often made between the conditions that give rise to sound change and those that are responsible for its spread throughout the community (e.g., Baker, Archangeli, & Mielke, 2011; Janda & Joseph, 2003; Ohala, 1993). Theories of the origin of sound change relate sound change to models of human speech processing. They often draw upon models of how coarticulation and timing relationships are transmitted between a speaker and hearer (Beddor, 2009; Solé & Ohala, 2010); and they are concerned with developing general principles of how sound change is related to the typological preferences or biases of sounds and their combinations in the world's languages (Solé, 2014). On the other hand, the focus of research in the sociolinguistic tradition has often been on establishing how social and dialect variation (Labov, 2001; Trudgill, 2011) as well as the type of community—whether, for example, tightly knit or loosely connected (Milroy, 1992)—influences sound change. This second research line is often just as concerned with establishing a general theory of how and why languages change, but it gives more prominence to the cultural and social conditions for how sound change spreads around and between communities from speaker to speaker and somewhat less to the mechanisms of human speech processing. One of the main tasks in this paper is to begin to provide the foundations for unifying both approaches within a cognitively inspired computational model of sound change.

We focus on sound change for two further reasons, both succinctly expressed by two leading researchers in the sociolinguistics tradition. First, as Trudgill (2012) has recently noted, "Perhaps the greatest puzzle that faces linguistic scientists is the phenomenon of linguistic change," which concurs with the view of Labov (2006) who comments that "For me, the largest questions that require the analysis of social variation are the problems of language change: how such changes are initiated, how they are transmitted across generations, what drives them forward, how they reach completion." In the same paper, Labov (2006) goes on to make the very important point that studying language change can provide direction in prioritizing analyses of social variation. As we now know, and as so very many experiments in sociophonetics in the last 10–15 years have shown (see, e.g., Docherty & Foulkes, 2014 for a recent review), sociophonetic variation is virtually limitless. But to cite once again Labov (2006): "it does not follow automatically that all such indexing should be described. ... Some further justification for the description of variation is required; otherwise there will be no stop to the enterprise and we will be

plunged into an endless pursuit of detail." Studying sound and language change provides a framework for establishing testable hypotheses. More specifically, we may have knowledge of a sound change that has or is taking place and we also know much about variation. So how then are the two connected? The question is simple and reduces to an input (variation) output (categorical change) problem. The solution of finding how the input and output are connected can, in turn, shed light on the architecture of human speech processing, how it is flexibly adapted to social variation in language, and which mechanisms within this architecture can give rise to change.

2. Phonetic bias, sound change, and phonological categories

Sound change is typically directional (Garrett & Johnson, 2013). This type of asymmetry can often have its origins in a corresponding synchronic, phonetic bias. Thus, there are several examples of an /s/ → /ʃ/ sound change in many languages (Kümmel, 2007; Rohlfs, 1966). There is also evidence that /s/ often assimilates synchronically to /ʃ/ (Pouplier, Hoole, & Scobbie, 2011; Recasens & Mira, 2013), that it retracts before post-alveolars (Baker et al., 2011; Stevens & Harrington, 2016) and that /s/ becomes acoustically more /ʃ/-like before rounded vowels (Mann & Repp, 1980). But on the other hand, whereas synchronic and diachronic /ʃ/ → /s/ changes may be possible, they are much less likely. Further examples of such asymmetries include the greater likelihood for fronted velars to be perceived as coronals (Chang, Plauché, & Ohala, 2001) than in the other direction, an asymmetry that has been associated with the sound change velar palatalization (Guion, 1998); and the greater tendency than in the other direction for domain-final voiced stops as in "tab" to become devoiced toward "tap" (Jongman, Sereno, Raaijmakers, & Lahiri, 1992; Pierrehumbert, 2001) which can be associated with diachronic domain-final neutralization toward the voiceless counterpart in various languages (Port & O'Dell, 1985; Warner, Jongman, Sereno, & Kemps, 2004).

We schematize this asymmetry in Fig. 1 by a difference in the direction of variation of the two phonological categories in an acoustic space. Thus, in Fig. 1, the phonological category that is represented by the ellipse encompassing the filled circles (=/s/) points more toward that of the filled triangles (=/ʃ/) than the other way round. Such differences in orientation affect the degree to which a category shifts depending on the uptake of outliers.

More specifically, Fig. 1 shows that there is a much greater likelihood of the /s/-category shifting due to the absorption of /s/-outliers (open circles, Fig. 1, row 1) than there is of a shift due to the /ʃ/-category as a result of absorbing /ʃ/-outliers (open triangles, Fig. 1, row 2). Three principles define whether an outlier is incorporated into a category.

1. An outlier in a lexical item *L* can only ever be absorbed into the phonological category with which *L* is associated.
   Suppose that the fricative in the lexical item "string" maps to the phonological category /s/. Then according to (1), a fricative outlier in "string"—even one that was in

Absorbption of /s/-outliers (open circles) into /s/ (filled circles)

Absorption of /ʃ/-outliers (open triangles) into /ʃ/ (filled triangles)

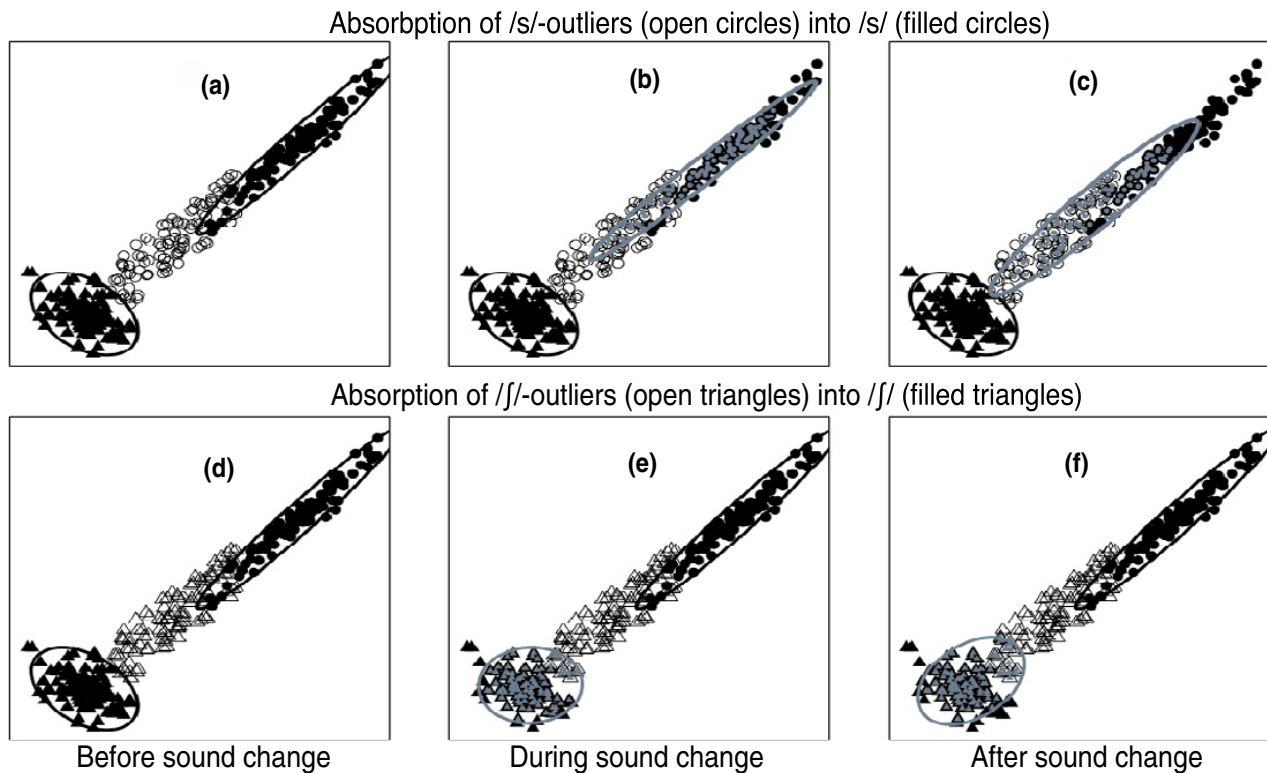Before sound change   During sound change   After sound change

Fig. 1. The two standard-deviation ellipses per panel are statistical generalizations across members of two separate phonological categories (filled data points) in a hypothetical two-dimensional space such that one phonological category (filled circles, e.g., /s/) is skewed toward the other (filled triangles, e.g., /ʃ/). The open data points are outliers of either the circle (top row 1a–c) or triangle (bottom row, 1d–f) phonological categories. The gray ellipses and associated filled gray data points show the change in the positions of the phonological categories after the first (middle panels, 1b, 1e) and final stages (right panels, 1c, 1f) of absorbing outliers and removing an equivalent number of the probabilistically most marginal category members.

the middle of the /ʃ/ space acoustically—could only ever be absorbed into /s/ and never into /ʃ/ (or into any other phonological category). Analogously, the open circles in row 1 (open triangles of row 2) of Fig. 1 are only ever considered for inclusion as part of the filled circle category in row 1 (filled triangle category in row 2). The further implication of (1) is that, while of course not denying that listeners sometimes confuse phonologically similar lexical items, this type of confusion is not considered to drive sound change: that is, there is no sense in which the mis-identification of "bet" for "bed" can contribute to sound change in the model being presented here.

2. The probability of category membership must be higher than to any other category. Probability is defined here as the posterior probability of category membership. That is, the listener has a number of phonological categories that are each defined by a parametric Gaussian distribution across the signals with which they are associated. A calculation is then made of the posterior probability that a perceived signal could be a member of any of the categories stored by the listener. According to (2) an /s/-outlier, O_s, may be absorbed into the /s/ category only if $p(O_s|s) > p(O_s|ʃ)$, that is, if the probability that it belongs to the /s/-category is greater than the probability of it

belonging to the /ʃ/ category (or indeed to any other category). Similarly, an /ʃ/-outlier $O_ʃ$ in row 2 is only absorbed if $p(O_ʃ|ʃ) > p(O_ʃ|s)$. This constraint is based on a similar principle adopted elsewhere (e.g., Blevins & Wedel, 2009; Garrett & Johnson, 2013; Hay, Pierrehumbert, Walker, & LaShell, 2015) of not incorporating ambiguous signals into memory. It is also the same constraint in Labov (2010) and in related variationist research (e.g., Dinkin & Dodsworth, 2017) of not absorbing an outlier into a phonological category if it is positioned too close to another category. The incorporation of outliers that are distant from a category (but closer than to any other category) is one of the mechanisms that variationist research has used to explain drag-chain shifts in vowels (see Labov, 2010, pp. 143–145, also for an extension of this principle to push-chains). Taken together, principles 1 and 2 guarantee a degree of stability in the computational model of sound change (see Kirby, 2014 for a further discussion); that is, they prevent mergers of different phonological categories and therefore of semantic loss.

3. Whenever an outlier is absorbed, then a member of the same category is removed either randomly or using a form of memory loss or probabilistically.

In Fig. 1 and following (3), whenever an outlier is absorbed, then the probabilistically most marginal member of the same category was removed (which can cause an outlier to be absorbed and then immediately removed if it happens to be the probabilistically least likely category member). Principle (3) is a form of memory loss that is designed to counteract an indefinite broadening of categories due to the uptake of outliers (see also Ettlinger, 2007; Pierrehumbert, 2001, for the incorporation of memory decay in an exemplar model). Once the outlier is removed, then it forms no further part of the analysis in the model.

The application of (1–3) initially to all outliers (middle panel) and then iteratively to any remaining outliers until there is no further change causes a much bigger category shift (gray ellipses) relative to the category starting position (black ellipses) when /s/-outliers are absorbed into the /s/-category (row 1) than when /ʃ/-outliers are absorbed into the /ʃ/-category (row 2).

Experiments from perceptual learning (Mitterer & Reinisch, 2013; Norris, McQueen, & Cutler, 2003; Samuel & Kraljic, 2009) in the last decade provide some support for such a cognitive mechanism by which the category boundary between two sounds varies depending on asymmetries between them in the direction of variance. In these experiments, a listener's decision boundary between two phoneme classes *x, y* is shifted toward *x* after exposure to a speaker whose productions of *y* are perceived to be skewed toward *x*, that is, whose *y*-productions are more likely to extend into the *x*-space than the other way round. The actual experiments typically make use of word-final fricatives /f, s/ and involve an exposure and a test. Listeners are firstly exposed to words with final /f/ containing an acoustically unambiguous [f], as well as to words with final /s/ that is acoustically ambiguous between /f, s/. The listeners subsequently categorized a continuum between /f/ and /s/. Tokens from this continuum that prior to exposure might be perceived to be ambiguous between the fricatives have a greater probability of being classified as /s/ following exposure; that is, the category boundary shifts toward /f/. Translated into Fig. 1, the circles and triangles represent /s/ and /f/,

respectively, and the /s/-variation points more in the direction of /f/ than the other way round (because the /s/-stimuli in the exposure phase had been acoustically skewed toward /f/). This result in perceptual learning experiments is bidirectional so that the category boundary shifts more in the other direction following exposure to words with final unambiguous /s/ and words with final ambiguous /f/. This bi-directionality in perceptual learning provides evidence for a general cognitive mechanism by which the mapping between words and ambiguous (or in terms of Fig. 1, skewed) speech signals can warp the boundaries between phonological categories.

There is another similarity between the results from perceptual learning and the model in Fig. 1 in that the lexicon is activated in both cases in categorizing outliers (or ambiguous tokens) and in shifting the boundary between categories. In perceptual learning, it is the activation of lexical items with ambiguous tokens that retunes the mapping between phonological categories and speech signals (McQueen, Tyler, & Cutler, 2012; Norris et al., 2003; Reinisch & Mitterer, 2016). Analogously, an outlier /b/ of "tab" in our model activates the lexical item "tab" by way of principle (1), even if it is an outlier that falls acoustically in the space of the /p/ category. Whether or not the /b-p/ category boundary is actually retuned by such an outlier in our model depends according to principle (2) on whether it is probabilistically closer to the /b/ than to the /p/ category; and according to principle (3) on whether it is then retained in the /b/ category.

There is further support for a mechanism such as in Fig. 1 but this time from research in sociolinguistics. Trudgill (2004, 2008) provides extensive evidence that many of the sound changes in new-world settlements such as New Zealand in the 19th century are a consequence of imitation especially among children. Trudgill's analysis suggests that the emerging and defining characteristics of the New Zealand English accent were a function of the type and strength (in terms of number of settlers) of dialects that came into contact with each other. Thus, New Zealand English shares many characteristics with the accents from the East of England at that time, because this is where the majority of settlers were from. A puzzle for Trudgill (2004) is, however, why New Zealand English has neutralized the /ɪ, ə/ contrast in unstressed syllables, so that, as in Australian English, *dances* and *dancers* have become homophones (but distinct in many British English varieties with final /ɪz, əz/, respectively). Trudgill (2004) reluctantly appeals to markedness as a possible explanation. However, Fig. 1 suggests an alternative solution. In spontaneous speech, there is a tendency for the distinctive vowel quality in unstressed syllables to be lost so that words with weak /ɪ/ are likely to become schwa-like; whereas there is no reason for /ə/ in *dancers* to shift toward /ɪ/. Consequently, there is once more an asymmetry in the direction of variance such that /ɪ/ points more toward /ə/ than the other way round. If schwa-like productions of /ɪ/ are increasingly incorporated into the distribution of words like *dances* (which is highly likely in the developing New Zealand accent, given that many speakers at the time of settlement had a final /ə/ and no /ɪ/ in such words), then /ɪ/ incrementally should shift toward /ə/ according to the model in Fig. 1. However, while the model can explain why /ɪ/ should shift toward /ə/ it has no explanation for why these vowels should merge. For this, a separate mechanisms of splitting and merging phonological categories is needed, as discussed in Section 4.

There is some but as yet only very limited evidence of a shift related to Fig. 1 when acquiring a new dialect through relocation. Evans and Iverson (2007) report for a Northern variety of English that high back lax [ʊ] in words like *luck* became more centralized when Northern students moved to the Southern English dialect region that has an open central [ʌ̈] in such words. In terms of Fig. 1, this comes about because contact might magnify an existing tendency for [ʊ] to centralize in spontaneous speech. Interestingly, there was much less shift of the Northern fronted [a] toward the Southern English back [ɑ] (in words like *bath*) and perhaps this is because the direction of the synchronic [a]-variation is "vertical," that is, toward the center and not horizontal toward the back of the vowel space.

There is, however, no reason according to Fig. 1 why variation should necessarily be drawn to the center of the vowel space or even in the direction of variation due to spontaneous speech, as is of course known from various patterns of vowel chain shifts (Labov, 2010). Compatibly, Bigham (2010) reported that the wide horizontal variation in female Southern Illinois high school students' [o, u] shifted toward the back of the vowel space after contact with Northern students who had a more compact and retracted [o, u]. Evidence from dialect contact suggests that children more readily acquire the features of the new dialect that tend to apply to all words with limited exceptions (Chambers, 1992; Nycz, 2015; Payne, 1980). Kerswill (1985, cited in Chambers, 1992) reports that immigrants to Bergen, Norway, nevertheless acquired a more complex rule of word-final schwa lowering that only applies in certain contexts. However, the migrant group already had a tendency to shift schwa in this direction in their variety. Thus, following the reasoning associated with Fig. 1, an existing phonetic tendency may have become magnified through contact with another group whose variants lie in the path of synchronic variation.

3. Testing the model: Variants of the same phonological category

In the model in Fig. 1, there are separate phonological categories that are differently oriented with respect to each other in an acoustic space.[1] The model can also be applied to modeling different orientations of two phonetic variants of the *same* phonological category. This version of the model is relevant for predicting the outcome when two dialects that have realizational differences of the same phoneme come into contact with each other. Although there is some evidence that second dialect acquirers shift their variety in the direction of the one that they are acquiring (Munro, Derwing, & Flege, 1999; Siegel, 2010) sometimes resulting in a form of phonetic averaging between the two (Trudgill, 1999, 2008), there are few studies concerned with how the outcome might be affected by the types of phonetic asymmetries considered in the previous section.

The single difference compared with the scenario in the preceding section centers around principle (2) which only allows a signal to be absorbed into a phonological category as long as it is not probabilistically closer to any other phonological category (as a result of which two phonological categories can never merge). This principle does not apply in the case of Fig. 2, which has different phonetic realizations of the same

phonological category. Thus, the prediction of the model is that a certain degree of approximation of two phonetic variants that are close to each other in an acoustic space (as in Fig. 2) is inevitable; moreover, their complete merger is possible given that principle (2) does not apply. The question to be considered is the following: Is one phonetic variant absorbed at a faster rate into the other?

We make the prediction that if the phonetic variant of a dialect x lies along the path of phonetic variation of another dialect y, then y will be drawn toward x (Fig. 2). This is because the probability that an outlier of x could belong to y is greater than in the other direction, that is, $p(x'|y) > p(y'|x)$ where x' and y' are outliers of their dialect groups and where x and y are different (dialect-dependent) variants of the same phoneme as in Fig. 2. The shift in y is predicted to be caused by the same type of directional asymmetry as in Fig. 1 in which there is an interaction between two separate phonological categories.

Harrington and Schiel (2017) provided a test of the model in Fig. 2 using an agent-based simulation applied to data from an earlier apparent-time study in Harrington, Kleber, and Reubold (2008).

This earlier study had provided evidence for /u/-fronting as a sound change in progress in both production and perception based on an apparent-time comparison of older and younger speakers of the Standard Southern British English (SSBE) variety. Of relevance to the present investigation is that there was also an asymmetry between the groups in the direction of the variation in the second formant frequency (F2) that indexes fronting.
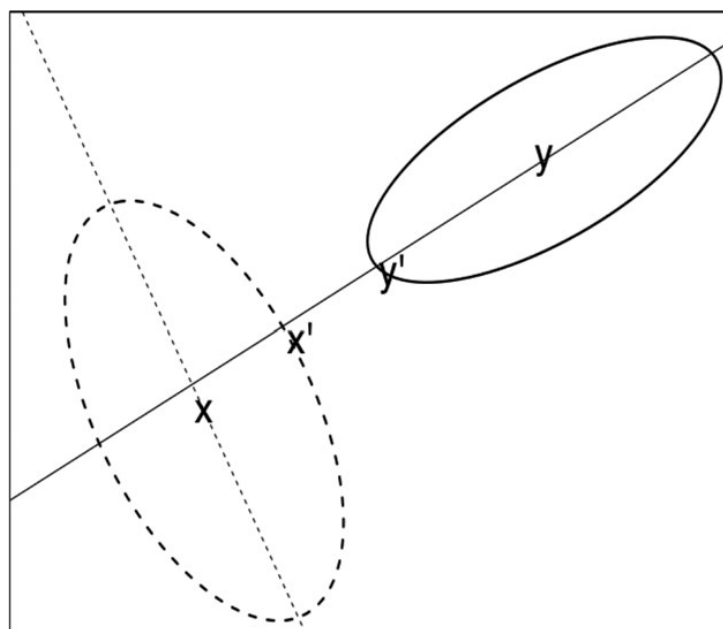

Fig. 2. Contours of equal probability at the same number of ellipse standard deviations around x and y that are two phonetic variants of the same phonological category in an arbitrary two-dimensional perceptual space. x' and y' are outliers in their respective distributions. The lines show the first principal component, that is, major axis of variation in x (dashed) and in y (solid). Because x falls along the path of the major axis of variation of y (but not the other way round), the probabilistic (Mahalanobis) distance of x' to y is smaller than that of y' to x. For this reason, outliers such as x' are more likely to be absorbed into the distribution of y, causing y to shift incrementally toward x.

Thus, the upper range of F2-variation for older speakers was quite close to the F2-mean for younger speakers; in contrast, younger speakers' lower range of F2-variation was considerably further from older speakers' F2-mean. This asymmetry may come about because some of the older speakers were already participating in the sound change in progress. As discussed in Harrington and Schiel (2017), the more likely reason is that older speakers' retracted /u/ is often drawn to the front of the vowel space by target undershoot and/or coarticulation with coronal consonants (Harrington, Hoole, Kleber, & Reubold, 2011; Lindblom & Studdert-Kennedy, 1967); whereas the corresponding backing of younger speakers' fronted /ʉ/ due to these synchronic forces is much smaller. Following Fig. 2, the older speakers' retracted /u/ should shift more toward younger speakers' fronted /ʉ/ following interaction and imitation across these groups.

To test such a prediction, the model in Harrington and Schiel (2017) included 22 agents, 11 representing older speakers with clearly retracted /u/ and 11 representing younger speakers with a fronted /u/. Each agent was initialized with a lexicon containing minimal-pair /i, ju, u/ triplets (e.g., *feed*, *feud*, *food*) together with the corresponding F2-trajectories of /i, ju, u/ that had been produced for these words by the original speakers in Harrington et al. (2008). There were typically 10 repetitions of each word and each trajectory was compressed using the discrete cosine transformation (DCT) to a point in a three-dimensional space whose axes are proportional to the F2-mean, linear-slope, and curvature (see Watson & Harrington, 1999, for further details of the DCT). Thus, consistently with episodic models of speech (Pierrehumbert, 2003a,b), each word-class was associated with a multidimensional cloud of points stored in memory. A Gaussian model was constructed for each word over the cloud of points and used to generate a random sample (a DCT-triplet) whenever an agent produced a word. In addition, and also following episodic models, phonological classes were defined as the union of the cloud of points across the corresponding word classes; thus, /i/ was a Gaussian model over the cloud of points associated with all words (*feed*, *heed, keyed, seep*) in which /i/ occurred.

An interaction was always between a pair of agents selected at random, one of which was the agent-speaker and the other the agent-listener. A word was randomly chosen from the agent-speaker's lexicon. A DCT-triplet, sampled from the word's Gaussian model of the agent-speaker, was transmitted together with the word label to the agent-listener. The probabilistic distance of the incoming DCT-triplet to the phoneme classes determined whether or not the agent-listener memorized the incoming signal. More specifically, and following principle (2) outlined earlier, if the word-class *heed* was transmitted, then the DCT-triplet had to be probabilistically closer to the agent listener's /i/ than to the distributions of the other phoneme classes (i.e., /ju, u/). Following principle (3), if the agent-listener memorized an item, then an item from the same word class was removed whose vowel was probabilistically most marginal (for this example, whichever *heed* item whose vowel was probabilistically most marginal in the agent-listener's /i/ distribution was removed from memory). Compatibly with the predictions based on Fig. 2, the shift after 50,000 interactions (beyond which there was scarcely any further change) was asymmetric such that there was a greater shift of older agents' retracted /ju, u/ toward the front of the vowel space than in the other direction (Fig. 3).
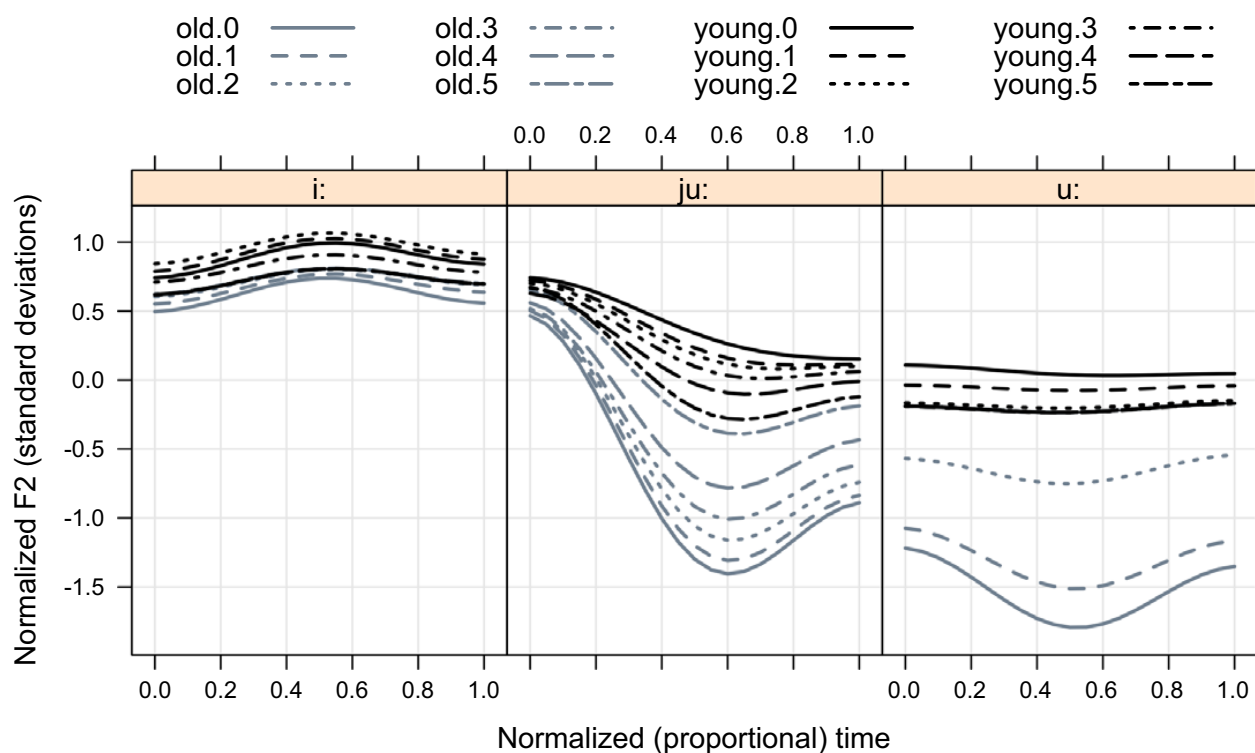
Fig. 3. Speaker and linearly time normalized F2 trajectories aggregated in the original data prior to any inter-action (0.0) and then at intervals of 10,000 (0.1) up to 50,000 (0.5) interactions in the older (old/gray) and younger (young/black) agents for the three vowel classes in /i/ (e.g., *feed*), /ju/ (e.g., *feud*), and /u/ (e.g., *food*) words (e.g., "old.1" is the trajectory aggregated across all older agents after 10,000 interactions). From Harrington and Schiel (2017).

It should be noted that this result excluded any effects of statistical frequency as far as the number of speakers was concerned (which was the same for both older and younger agents) and also as regards lexical or indeed phoneme frequency. The results would certainly have been different if there had been a different number of agents per dialect group or if word access had been differently weighted in relation to lexical statistics: This could be used to test the extent to which sound change is affected by lexical frequency (e.g., Bybee, 2000; Hay et al., 2015). Both of these effects of statistical frequency could be straightforwardly modeled using the existing architecture by increasing or decreasing the number of times different types of agents (older or younger) or lexical items (high or low frequency) are accessed.

4. Splits and mergers

A shortcoming of the model so far is that, while phonological categories might approx-imate each other in an acoustic space, they can never merge (Fig. 1). This is because of principle (2), according to which, for example, an outlier /b/ from "tab" could never be absorbed into the phonological category associated with "tap", even if the outlier falls well within the /p/ space acoustically. The model was therefore extended to incorporate mergers and category splits both of which are common types of sound change (Kiparsky,

2016; Labov, 1994, Ch. 12). The general idea behind this extension is to allow a phonological category to split such that any of the components derived from the split can merge with other categories. This procedure might allow, for example, a /u/ category to split into two categories, one of which might include /u/ variants following fronting ("Sue") and the other non-fronting ("woo") consonants. The synchronic variation leading to domain-final neutralization of voicing in obstruents in many languages might initially involve a shift of the voiced toward the voiceless category in the manner of Fig. 1 after which the voiced category splits into subcategories which are then merged with the voiceless category.

An algorithm for splitting and merging was therefore incorporated into the architecture of the agent-based model that had been applied to /u/-fronting described earlier.

For the splitting algorithm, a phonological class, P, defined as a statistical distribution over an acoustic space (in this case over the space formed by the three DCT-coefficients) was split into two maximally acoustically distant classes, p_1, and p_2, using k-means clustering (Hartigan & Wong, 1979). If the probability of category membership (based on the Mahalanobis distance: Duda, Hart, & Stork, 2001) of the cloud of points to P was less than to p_1, and p_2, then no split occurred; otherwise the split was made (Fig. 4).

Two further conditions were applied: Any phonological class had to be associated with more than one word (since there would otherwise be no phonological generalization across the lexicon); and no word could be associated with more than one phonological class. If a word class was split across two phonological classes after applying k-means clustering (if, e.g., following a split, *food* mapped onto both a front /ʉ/ and back /u/), then reassignment based on majority membership to one of the classes occurred. Merger was essentially the reverse and occurred if the probability of category membership of the cloud of points to the union of two classes, p_1, and p_2, was greater than the probabilities
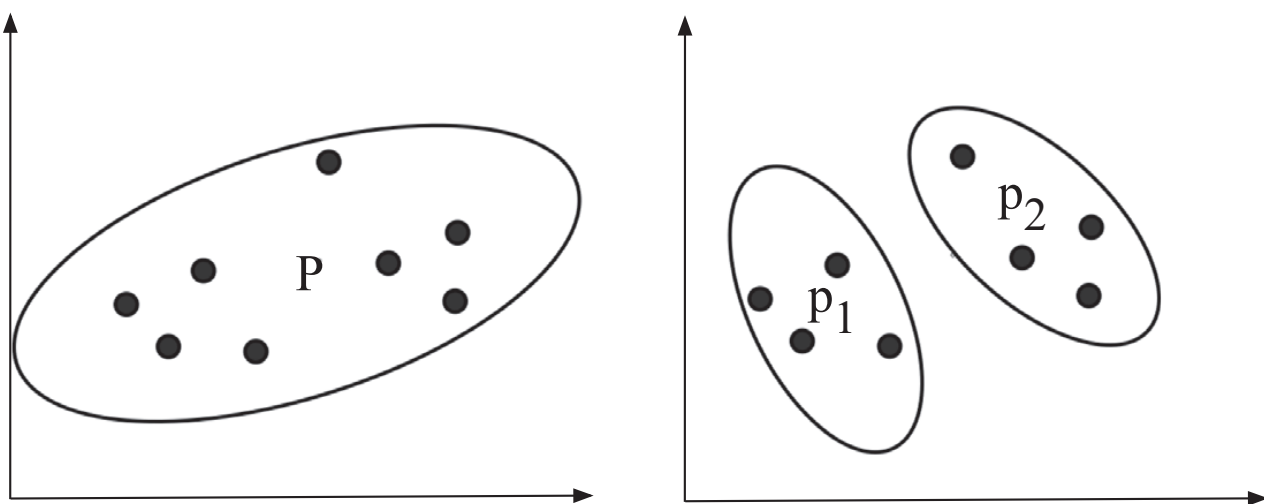

Fig. 4. A phonological category, P, containing points in a two-dimensional space and the two categories, p_1, and p_2, into which it is split following k-means clustering. The split was only made if the probability of category membership of the four points to p_1 and the four points to p_2 was greater than that of the eight points together to P.

to the two classes separately. A test for merger was applied iteratively to all existing pair-wise combinations of phonological classes in an agent-listener's memory.

A modification of the agent-based model that incorporated splitting and merging was applied in this study to speech data from 20 adult speakers to model /s/-retraction in Australian English (Stevens & Harrington, 2016). The aim of the investigation was to assess whether /s/ in /str/ clusters would shift toward /ʃ/ as a consequence of agent interaction, given the evidence for /s/-retraction to occur in such contexts in various varieties of English (Baker et al., 2011; Cox, 2012; Warren, 1996) as a consequence of the anticipatory coarticulatory influences of the post-alveolar approximant. Labov's (2010) merger-by-approximation (rather than by transfer or by expansion) is most closely related to the type of /s/ to /ʃ/ merger considered here because in merger-by-approximation and in these /s/-retraction data the merger takes place through the gradual shift of one category toward another in potentially all lexical items.

Each speaker was represented by an agent. The lexicon consisted of 10 monosyllabic words: two with word-initial singleton /s/, two with word-initial singleton /ʃ/, and the remainder with initial /str/. As discussed in further detail in Stevens and Harrington (2016), some of the speakers produced an audibly retracted /s/ in /str/-words, but others did not. Once again, each word was associated with a cloud of points that was in this case a DCT-parameterization of the first spectral moment (Jongman, Wayland, & Wong, 2000) calculated at 10 ms intervals between the fricative's acoustic onset and offset. The initial conditions included two phonological classes: one for the /s, str/ words combined and the other for the two /ʃ/-words. As before, each phonological class was associated with a cloud of DCT data points. We ran the agent-based model using the same architecture as for /u/-fronting described earlier, but incorporating splitting and merging as sketched above.

Overall, the results showed scarcely any acoustic change over 20,000 pairwise agent interactions. Perhaps this is unsurprising, given firstly the relative homogeneity of the speakers (who were all adults of the same Australian English variety and from the same community) and secondly because of the unrealistically small number of separate lexical items. There was, however, some evidence of a change in how words were associated with phonological classes as a result of splitting and merging. As Fig. 5 shows, after 1,000 interactions /str/ split from singleton /s/ for most agents, whereupon between 1 and 2 agents merged /str, ʃ/ into a single class. Typically, those agents who did merge /str, ʃ/ (dashed lines) vacillated in doing so over the cycle of iterations. Thus, as shown in Fig. 6, for one of the agents, /s, str, ʃ/ were in three separate classes at 10,000 interactions; then /str/ and /ʃ/ merged between 11,000 and 12,000 interactions; and at 13,000 interactions unmerged them again into two separate classes.

This instability in the category label but stability in the acoustic signal may represent the first stages of change, or the potential for change to take place. Vacillation in phonological affiliation may also be an appropriate model for the listener's uncertainty in whether the speech production had actually corresponded to /str/ or /ʃtr/. An acoustic bias that remains stable, combined with the potential for categorical change, is in some ways analogous to the
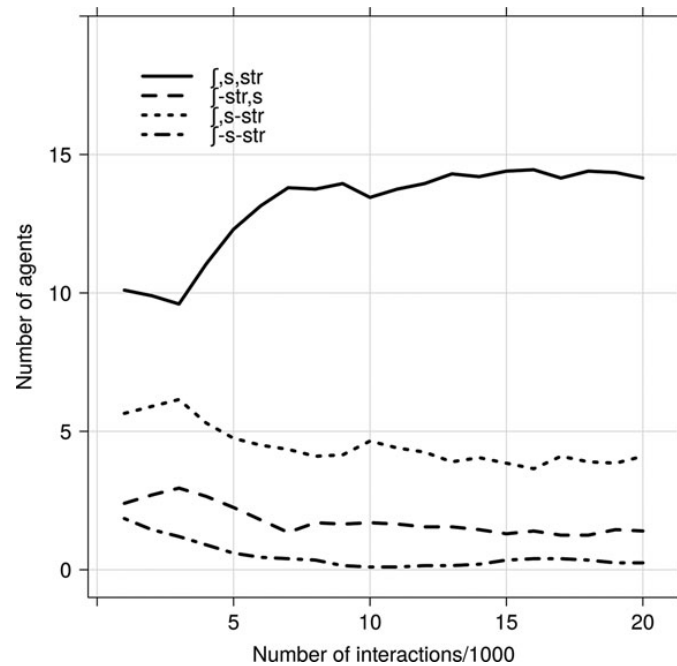
Fig. 5. The figure shows the number of agents (*y*-axis) per phonological class combination between 1,000 and 20,000 interactions averaged over 20 separate simulations. The lines representing the possible different combinations are as follows: *solid*, three classes, /s, str, ʃ/; *dashed*, two classes, merged /ʃ+str/ and /s/; *dotted*, two classes, merged /s+str/ and /ʃ/; *dashed-dotted*, one class, merged /s+str+ʃ/. For example, the top solid line at the far left shows that 10 agents had on average three separate (unmerged) classes /s, ʃ, str/ at 1,000 interactions.

first stages of sound change in Ohala's (1993) model in which there is an abrupt change of phonetic variants (e.g., from retracted /u/ to fronted /ʉ/) if the listener misparses coarticulation in perception. But this re-categorization happens without the listener necessarily duplicating this change (and/or applying it in other contexts) in speech production.

5. Discussion

A consequence of incorporating splitting and merging into the IP-model is that words map onto sub-phonemic classes (e.g., *food* and *move* may come to be associated with fronted /ʉ/ and retracted /u/, respectively). From the perspective of human speech processing, Pierrehumbert (2003b) notes that positional allophones may be a more viable form of abstraction than phonemes because phoneme classes are often too coarse for parametric distributions to be distinguished from each other. Sub-phonemic abstraction is compatible with the idea that non-contrastive phonetic variants can form part of the lexicon (Kiparsky, 2015, 2016) and with recent studies from perceptual learning showing that listeners adapt to units that are more fine-grained than the phoneme (Reinisch & Mitterer, 2016; Reinisch, Wozny, Mitterer, & Holt, 2014). There is also other evidence for sub-phonemic processing from second language (Polka, 1991) and new dialect learning (German, Carlson, & Pierrehumbert, 2013). Splitting and merging were also agent-specific in
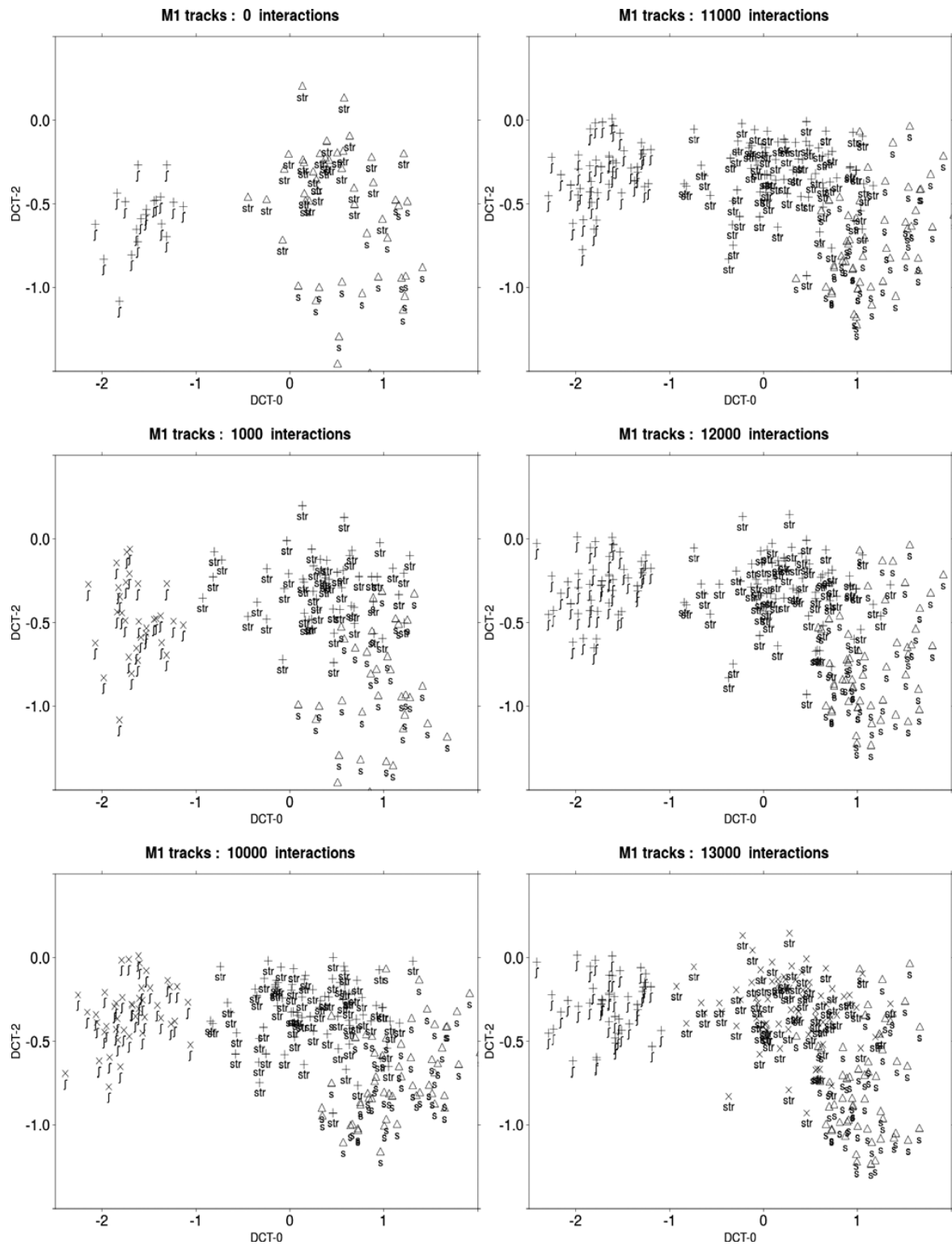
Fig. 6. Each panel shows the position of /s, str, ʃ/ word tokens in a two-dimensional acoustic DCT space between 0 and 13,000 interactions for one agent. The symbols (+, ×, Δ) show the affiliation to phonological classes. For example, at 0 interactions, that is, the starting condition (top left) /s, str/ word tokens are in the same class because they are both represented by Δ; at 11,000 and 12,000 interactions (bottom left), /ʃ, str/ word tokens are in the same class (both represented by +) but in separate classes at 13,000 interactions (bottom right).

the IP-model. The analogous idea that sub-phonemic abstraction is speaker-specific is also compatible with episodic models of speech. This is because phonological abstraction is a generalization across word classes that are themselves generalizations over listener-specific stored episodes of speech (Pierrehumbert, 2003a,b). The sub-phonemic classes have a stabilizing function that limits the uptake of episodes that stray into the phonetic space of another sub-phonemic class. This stability through sub-phonemic classification of signals in perception is one of the main reasons why phonetic variation need not (and typically does not) result in sound change in the IP-model (Kirby, 2014; Ohala, 1993).

Various studies have shown that categorization in perception is strongly influenced by different kinds of context. For example, the category boundary between /i-u/ is closer to /i/ in a fronting context such as /Vdə/ (Ohala & Feder, 1994) or /jist-just/, *yeast-used* (Harrington et al., 2008) than in a non-fronting context such as /Vbə/ or /swip-swup/, *sweep-swoop*. This effect of context on category boundaries is a consequence in the IP-model of sub-phonemic abstraction. That is, the category boundaries between *yeast-used* and *sweep-swoop* are necessarily different, if *used* and *swoop* map onto different sub-phonemic classes. The size of such differences also varies between agents. Analogously, there is also evidence from human speech processing that groups (Kataoka, 2011; Yu, 2013) or specific listeners (Beddor, 2009; Fowler & Brown, 2000) vary in the extent to which they normalize for these types of context effects.

Compensating for coarticulation insufficiently, which is one of the potential causes for certain types of sound change in Ohala's (1993) model, corresponds in the IP-model to a merger of sub-phonemic classes. A major difference between the two models is that whereas the listener's failure to normalize for context in Ohala (1993) is causally related to sound change, in the IP-model it may instead be an effect of the incrementally closer approximation and eventual merger of sub-phonemic classes brought about by interaction. This interpretation is then closer to incremental sound change in episodic models of speech (see also discussion in Harrington et al., 2008).

In both episodic models and in the IP-model, word-specific phonetic detail that has been found in many studies (e.g., Yaeger-Dror, 1996) is a consequence of associating word classes with their perceived speech signals or exemplars. Pierrehumbert (2002) reasons, however, that speech production does not involve sampling directly from such exemplars but is instead mediated by phonological abstraction. Among the reasons given are that many sound changes nevertheless apply with a great deal of consistency across the words that meet the context for the sound change. In the IP-model by contrast, agents sample from the word-specific space when producing a word. Phonological abstraction in our model acts as a brake in perception to prevent any exemplar from being absorbed into a phonological class if it is probabilistically closer to another class (principle (2) outlined earlier). For this reason, the statistical generalization across the exemplars of a word class is necessarily molded in the IP-model by both word-specific detail and phonological conformity (that is imposed indirectly in speech perception). Thus, sampling from the word space in production need not preclude across-the-board sound changes. Moreover, the dichotomy in the IP-model that has word-specific detail but which allows phonological categories to merge in perception may provide a way of accounting for near-mergers

(Labov, 1994, Ch. 12; Yu, 2007) and incomplete neutralizations (Warner et al., 2004) in which minimal-pairs are shown to have subtle production differences but that are not perceptible.

The machinery in the IP-model needs to be adapted to handle sound change with a much longer time window than the types of sound change considered so far, such as metathesis (Blevins & Garrett, 2004; Egurtzegi, 2014), dissimilation (Abrego-Collier, 2013; Alderete & Frisch, 2006), and phonologization (Hyman, 2013; Kiparsky, 2015; Kirby, 2013, 2014). We are currently exploring the extent to which it is possible in the IP-model to derive this type of change as a consequence of an interaction-driven incremental shift in shape trajectories.

The phonetic biases that are likely to lead to sound change in the model by Lindblom, Guion, Hura, Moon, and Willerman (1995) are those due to hypo- and hyperarticulation. Phonetic reduction is also central to Bybee's (2001) usage-based model of sound change. For Trudgill (2011), reduced forms are especially likely to be phonologized over long time periods, possibly leading to greater phonological complexity, in small, remote, socially tightly knit communities. This is because listeners can bring to bear more top-down knowledge in such societies in which interlocutors are likely to be known to each other and in which vocabulary items and topics are also often repeated. Modeling such relationships is theoretically possible using an extended architecture of the IP-model, but it will obviously require very many more speakers, lexical items, and variations in speaking style.

6. Conclusion

The purpose of the IP-model is to bridge the gap between two different models of sound change: those that are concerned with the phonetic conditions that can give rise to sound change (Ohala, 1993) and those whose primary focus is on how social factors constrain or propagate its spread around the community (Eckert, 2012; Labov, 2001; Milroy, 1992). Both provide crucial insights into the evolution of sound change and while recognizing that some variationist research has been concerned with how phonological change is related to the transmission of phonetic variation between individuals (Dinkin & Dodsworth, 2017; Raymond, Brown, & Healy, 2016; Rohena-Madrazo, 2015), there is not much common ground between them. Moreover, we agree with Lindblom (1998) that it is unnecessary to limit the phonetic contribution to sound change to the initiation stage.

To bridge this gap, the IP-model seeks an answer to the question: What conditions are required for the interaction between individuals to convert existing phonetic biases into sound change? This question is a computationally tractable way of reformulating the famous question of the actuation of sound change: Why do changes in a structural feature take place in a particular language at a given time, but not in other languages with the same feature, or in the same language at other times (Weinreich, Labov, & Herzog, 1968)?

This intermediary stage of sound change in the IP-model when a stable phonetic bias is pushed by interaction into unstable change is essentially non-social, that is, a more general

consequence of speech (Pardo, Gibbons, Suppes, & Krauss, 2012) and motoric (Shockley, Richardson, & Dale, 2009) accommodation across individuals who happen to come into contact with each other (Labov, 2001, pp. 19–20). We contrast this with a social view of sound change in which speakers preferentially copy the speaking style of the social category that they want to belong to (see, e.g., Baker et al., 2011; Garrett & Johnson, 2013). These two different perspectives are appropriately summarized by Siegel (2010), who distinguishes between a linguistic ambience effect of imitation in which the imitation is not conditioned by social factors but is "automatic" or "reflexive" as opposed to accommodation (Giles, Taylor, & Bourhis, 1973), which, according to Siegel (2010), "is socially motivated, arising from an unconscious desire for social approval from one's interlocutors." This nonsocial propagation that in the IP-model is consequence of a combination of phonetic asymmetries and population dynamics is also likely to be the initial force controlling the development of dialects in new world (Trudgill, 2008) or new town (Kerswill & Williams, 2000) settings. It is also likely to be the level of operation of what in sociolinguistics is referred to as sound change from below (Labov, 2007). Trudgill (2004) notes that children are the main drivers of this non-social change. Compatibly, Nielsen (2014) has shown that children are especially prone to imitation while a study by Nardy, Chevrot, and Barbu (2014) showed phonetic convergence due to peer interaction within a group of primary school children, but without any significant influence from role models (neither from teachers nor peers that children liked). In the agent-based model developed by Pierrehumbert, Stonedahl, and Dalaud (2014), language innovations are primarily spread between ordinary near neighbor individuals rather than from those who enjoy high prestige by being well connected. These are the types of interaction and change that are most relevant to the IP-model in linking the phonetic conditions that have the potential to bring about change to its initial spread through interaction among a community of speakers.

The cognitive architecture for effecting these initial stages of non-social change in the IP-model requires a probabilistic association between word classes and remembered dynamically changing speech signals that may be updated through interaction. This updating is in turn dependent upon classification in perception using a level of sub-phonemic abstraction that is both a further generalization across word-classes and also speaker-specific. Determinism (Trudgill, 2004, 2008) only has a very limited role to play in the IP-model and is a consequence of both commonality in the design of language and its use across individuals who have similar sets of vocal organs and hearing systems, as well as the potential to imitate and mould each other's mappings between words, sub-phonemic classes, and speech signals. The far bigger component of the IP-model is stochastic in that whether or not sound change actually comes about depends upon which speakers regularly speak to each other, whether a phonetic bias happens to be magnified by interaction, and whether or not sub-phonemic classes are fragmented and regrouped over time.

These probabilities in combination explain why spoken accents (Cohen, 2012) and languages are phonetically so idiosyncratic (Pierrehumbert, 2002). They also explain sound change's mercurial side, which goes to the very core of the actuation problem referred to earlier: why sound change may happen under one set of circumstances, but not another.

Acknowledgments

Note

1. The software for running the agent-based modeling can be downloaded from ftp:// ftp.bas.uni-muenchen.de/pub/BAS/ABM/DemoABM.1.zip

References

Abrego-Collier, C. (2013). Liquid dissimilation as listener hypocorrection. In C. Cathcart, I. Chen, G. Finley, S. Kang, C.S. Sandy, & E. Stickles (Eds.), *Proceedings of the 37th Annual Meeting of the Berkeley Linguistics Society* (pp. 3–17). Berkeley, CA: Berkeley Linguistics Society.

Alderete, J., & Frisch, S. (2006). Dissimilation in grammar and the lexicon. In P. de Lacy (Ed.), *The Cambridge handbook of phonology* (pp. 379–398). Cambridge, UK: Cambridge University Press.

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

Beddor, P. (2009). A coarticulatory path to sound change. *Language*, *85*, 785–821.

Bigham, D. (2010). Mechanisms of accommodation among emerging adults in a university setting. *Journal of English Linguistics*, *38*, 193–210.

Blevins, J., & Garrett, A. (2004). The evolution of metathesis. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically-based phonology* (pp. 117–156). Cambridge, UK: Cambridge University Press.

Blevins, J., & Wedel, A. (2009). Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica*, *26*, 143–183.

Bybee, J. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 65–85). Stanford, CA: CSLI.

Bybee, J. (2001). *Phonology and language use*. Cambridge, UK: Cambridge University Press.

Chambers, J. (1992). Dialect acquisition. *Language*, *68*, 673–705.

Chang, S., Plauché, M., & Ohala, J. (2001). Markedness and consonant confusion asymmetries. In E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 79–101). San Diego, CA: Academic Press.

Cohen, E. (2012). The evolution of tag-based cooperation in humans: The case for accent. *Current Anthropology*, *53*, 588–616.

Cox, F. (2012). *Australian English: Transcription and pronunciation*. Melbourne, Vic: Cambridge University Press.

Dinkin, A., & Dodsworth, R. (2017). Gradience, allophony, and chain shifts. *Language Variation and Change*, *29*, 101–127.

Docherty, G., & Foulkes, P. (2014). An evaluation of usage-based approaches to the modelling of sociophonetic variability. *Lingua*, *142*, 42–56.

Duda, Richard, Hart, P., & Stork, D. (2001). *Pattern classification* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*, *41*, 87–100.

Egurtzegi, A. (2014). *Towards a phonetically grounded diachronic phonology of Basque*. PhD dissertation, University of the Basque Country.

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

Evans, B., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *Journal of the Acoustical Society of America*, *121*, 3814–3826.

Fowler, C., & Brown, J. (2000). Perceptual parsing of acoustic consequences of velum lowering from information for vowels. *Perception & Psychophysics*, *62*, 21–32.

Garrett, A., & Johnson, K. (2013). Phonetic bias in sound change. In A. Yu (Ed.), *Origins of sound change* (pp. 51–97). Oxford, UK: Oxford University Press.

German, J., Carlson, K., & Pierrehumbert, J. (2013). Reassignment of consonant allophones in rapid dialect acquisition. *Journal of Phonetics*, *41*, 228–248.

Giles, H., Taylor, D., & Bourhis, R. (1973). Towards a theory of interpersonal accommodation through speech: Some Canadian data. *Language in Society*, *2*, 177–192.

Guion, S. (1998). The role of perception in the sound change of velar palatalization. *Phonetica*, *55*, 18–52.

Harrington, J., Hoole, P., Kleber, F., & Reubold, U. (2011). The physiological, acoustic, and perceptual basis of high back vowel fronting: Evidence from German tense and lax vowels. *Journal of Phonetics*, *39*, 121–131.

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

Harrington, J., & Schiel, F. (2017). /u/-fronting and agent-based modeling: The relationship between the origin and spread of sound change. *Language*, *93*(2), 414–445.

Harrington, J., & Stevens, M. (2014). Cognitive processing as a bridge between phonetic and social models of sound change. *Journal of Laboratory Phonology*, *5*, 1–8.

Hartigan, J. A., & Wong, M. A. (1979). A K-means clustering algorithm. *Applied Statistics*, *28*, 100–108.

Hay, J., Pierrehumbert, J., Walker, A., & LaShell, P. (2015). Tracking word frequency effects through 130 years of sound change. *Cognition*, *139*, 83–91.

Hyman, L. M. (2013). Enlarging the scope of phonologization. In A. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (pp. 3–28). Oxford, UK: Oxford University Press.

Janda, R., & Joseph, B. (2003). Reconsidering the canons of sound-change: Towards a "big bang" theory. In B. Blake & K. Burridge (Eds.), *Selected papers from the 15th International Conference on Historical Linguistics* (pp. 205–219). Amsterdam, the Netherlands: John Benjamins.

Jongman, A., Sereno, J. A., Raaijmakers, M., & Lahiri, A. (1992). The phonological representation of [voice] in speech perception. *Language and Speech*, *35*, 137–152.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, *108*(3), 1252–1263.

Kataoka, R. (2011). *Phonetic and cognitive bases of sound change*. PhD dissertation, University of California, Berkeley.

Kerswill, P. (1985). *A sociolinguistic study of rural immigrants in Bergen, Norway*. PhD dissertation, University of Cambridge, UK.

Kerswill, P., & Williams, A. (2000). Creating a new town koine: Children and language change in Milton Keynes. *Language in Society*, *29*(65–1), 15.

Kiparsky, P. (2015). Phonologization. In P. Honeybone & J. Salmons (Eds.), *The Oxford Handbook of historical phonology* (pp. 563–582) Oxford, UK: Oxford University Press.

Kiparsky, P. (2016). Labov, sound change, and phonological theory. *Journal of Sociolinguistics*, *20*, 464–488.

Kirby, J. (2013). The role of probabilistic enhancement in phonologization. In A. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (pp. 228–246). Oxford, UK: Oxford University Press.

Kirby, J. (2014). Incipient tonogenesis in Phnom Penh Khmer: Computational studies. *Laboratory Phonology*, *5*, 195–230.

Kümmel, M. J. (2007). *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen*. Wiesbaden, Germany: Reichert.

Labov, W. (1994). *Principles of linguistic change. Vol. 1: Internal factors*. Malden, MA: Blackwell.

Labov, W. (2001). *Principles of linguistic change. Vol. 2: Social factors*. Oxford, UK: Blackwell.

Labov, W. (2006). A sociolinguistic perspective on sociophonetic research. *Journal of Phonetics*, *34*, 500–515.

Labov, W. (2007). Transmission and diffusion. *Language*, *83*, 344–387.

Labov, W. (2010). *Principles of linguistic change: Cognitive and cultural factors*. Chichester, UK: Wiley-Blackwell.

Lindblom, B. (1998). Systemic constraints and adaptive change in the formation of sound structure. In J. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language* (pp. 242–264). Cambridge, UK: Cambridge University Press.

Lindblom, B., Guion, S., Hura, S., Moon, S.-J., & Willerman, R. (1995). Is sound change adaptive? *Rivista di Linguistica*, *7*, 5–36.

Lindblom, B., & Studdert-Kennedy, M. (1967). On the role of formant transitions in vowel recognition. *Journal of the Acoustical Society of America*, *42*, 830–843.

Mann, V., & Repp, B. (1980). Influence of vocalic context on the perception of [ʃ]-[s] distinction: I. Temporal factors. *Perception & Psychophysics*, *28*, 213–228.

McQueen, J. M., Tyler, M., & Cutler, A. (2012). Lexical retuning of children's speech perception: Evidence for knowledge about words' component sounds. *Language Learning and Development*, *8*, 317–339.

Milroy, J. (1992). Social network and prestige arguments in sociolinguistics. In K. Bolton & H. Kwok (Eds.), *Sociolinguistics today: International perspectives* (pp. 146–162). London, UK: Routledge.

Mitterer, H., & Reinisch, E. (2013). No delays in application of perceptual learning in speech recognition: Evidence from eye tracking. *Journal of Memory and Language*, *69*, 527–545.

Munro, M., Derwing, T., & Flege, J. (1999). Canadians in Alabama: A perceptual study of dialect acquisition in adults. *Journal of Phonetics*, *27*, 385–403.

Nardy, A., Chevrot, J.-P., & Barbu, S. (2014). Sociolinguistic convergence and social interactions within a group of pre-schoolers: A longitudinal study. *Language Variation and Change*, *26*, 273–301.

Nielsen, K. (2014). Phonetic imitation by young children and its developmental changes. *Journal of Speech Language and Hearing Research*, *57*, 2065–2075.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*, 204–238.

Nycz, J. (2015). Second dialect acquisition: A sociophonetic perspective. *Language and Linguistics Compass*, *9*(11), 469–482.

Ohala, J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical linguistics: Problems and perspectives* (pp. 237–278). London, UK: Longman.

Ohala, J., & Feder, D. (1994). Listeners' normalization of vowel quality is influenced by "restored" consonantal context. *Phonetica*, *51*, 111–118.

Pardo, J., Gibbons, R., Suppes, A., & Krauss, R. (2012). Phonetic convergence in college roommates. *Journal of Phonetics*, *40*, 190–197.

Payne, A. (1980). Factors controlling the acquisition of the Philadelphia dialect by out-of-state children. In W. Labov (Ed.), *Locating language in time and space* (pp. 143–178). New York: Academic Press.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure* (pp. 137–157). Amsterdam, the Netherlands: John Benjamins.

Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology VII* (pp. 101–139). Berlin, Germany: Mouton de Gruyter.

Pierrehumbert, J. (2003a). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*, 115–154.

Pierrehumbert, J. (2003b). Probabilistic Phonology: Discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probability theory in linguistics* (pp. 177–228). Cambridge, MA: MIT Press.

Pierrehumbert, J., Stonedahl, F., & Dalaud, R. (2014). A model of grassroots changes in linguistic systems. arXiv:1408.1985v1.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America*, *89*, 2961–2977.

Port, R., & O'Dell, M. (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics*, *13*, 455–471.

Pouplier, M., Hoole, P., & Scobbie, J. (2011). Investigating the asymmetry of English sibilant assimilation: Acoustic and EPG data. *Journal of Laboratory Phonology*, *2*, 1–33.

Raymond, W., Brown, E., & Healy, A. (2016). Cumulative context effects and variant lexical representations: Word use and English final t/d deletion. *Language Variation and Change*, *28*, 175–202.

Recasens, D., & Mira, M. (2013). An articulatory and acoustic study of the fricative clusters /sʃ/ and /ʃs/ in Catalan. *Phonetica*, *70*, 298–322.

Reinisch, E., & Mitterer, H. (2016). Exposure modality, input variability and the categories of perceptual recalibration. *Journal of Phonetics*, *55*, 96–108.

Reinisch, E., Wozny, D., Mitterer, H., & Holt, L. L. (2014). Phonetic category recalibration: What are the categories? *Journal of Phonetics*, *45*, 91–105.

Rohena-Madrazo, M. (2015). Diagnosing the completion of a sound change: Phonetic and phonological evidence for /ʃ/ in Buenos Aires Spanish. *Language Variation and Change*, *27*, 287–317.

Rohlfs, G. (1966). *Grammatica Storica della Lingua Italiana e dei suoi Dialetti: Fonetica*. Pisa, Italy: Einaudi.

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, *71*, 1207–1218.

Shockley, K., Richardson, D., & Dale, R. (2009). Conversation and coordinative structures. *Topics in Cognitive Science*, *1*, 305–319.

Siegel, J. (2010). *Second dialect acquisition*. Cambridge, UK: Cambridge University Press.

Solé, M. (2014). The perception of voice-initiating gestures. *Journal of Laboratory Phonology*, *5*, 37–68.

Solé, M-J., & Ohala, J. (2010) What is and what is not under the control of the speaker: Intrinsic vowel duration. In C. Fougeron, B. Kuehnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory phonology 10* (pp. 607–655). Berlin, Germany: de Gruyter.

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

Trudgill, P. (1999). Dialect contact, dialectology and sociolinguistics. *Cuadernos de Filología Inglesa*, *8*, 1–8.

Trudgill, P. (2004). *Dialect contact and new-dialect formation: The inevitability of colonial Englishes*. Edinburgh, UK: Edinburgh University Press.

Trudgill, P. (2008). Colonial dialect contact in the history of European languages: On the irrelevance of identity to new-dialect formation. *Language in Society*, *37*, 241–254.

Trudgill, P. (2011). *Sociolinguistic typology*. Oxford, UK: Oxford University Press.

Trudgill, P. (2012). On the functionality of linguistic change. *Current Anthropology*, *53*, 609–610.

Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics*, *32*, 251–276.

Warren, P. (1996). /s/-retraction, /t/-deletion and regional variation in New Zealand English /str/ and /stj/ clusters. In M Tabain (Ed.), *Proceedings of the 11th Australian International Conference on Speech Science & Technology* (pp. 466–471). Canberra: Australasian Speech Science and Technology Association.

Watson, C., & Harrington, J. (1999). Acoustic evidence for dynamic formant trajectories in Australian English vowels. *Journal of the Acoustical Society of America*, *106*, 458–468.

Weinreich, U., Labov, W., & Herzog, M. (1968). Empirical foundations for a theory of language change. In W. Lehmann & Y. Malkiel (Eds.), *Directions for historical linguistics* (pp. 95–195). Austin, TX: University of Texas Press.

Yaeger-Dror, M. (1996). Phonetic evidence for the evolution of lexical classes: The case of a Montreal French vowel shift. In G. Guy, C. Feagin, J. Baugh, & D. Schiffrin (Eds.), *Towards a social science of language* (pp. 263–287). Philadelphia, PA: Benjamins.

Yu, A. (2007). Understanding near mergers: The case of morphological tone in Cantonese. *Phonology*, *24*, 187–214.

Yu, A. (2013). Individual differences in socio-cognitive processing and the actuation of sound change. In A. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (pp. 201–227). Oxford, UK: Oxford University Press.