14

The phonetic basis of the origin and spread of sound change

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

Introduction

The city of Munich where the authors of this chapter all live was known after it was founded in the 10th–11th centuries as *apud Munichen*, which means "by/near the monks." Some of the sound changes that have led to the present-day form /mynçən/ (*München*) from the Old High German stem /munic/ "monk" (from Greek μοναχός: Grimm & Grimm, 1854) can be linked to well-known effects of phonetic variation. The high front rounded vowel in present-day *München* has evolved historically from umlaut that can be associated synchronically with trans-consonantal vowel coarticulation (Hoole & Pouplier, 2017; Öhman, 1966) which in this case caused /u/ to front to /y/ under the influence of /i/ in the following syllable of *Munichen*. The loss of this /i/ is likely to have come about because it occurs in a weak, unstressed syllable which is often prone to reduction or deletion in continuous speech.

The present-day German for "monk" is *Mönch*, /mœnç/ in which the vowel has been lowered relative to the High German form with /u/ or the /y/ of *München*. There are several examples in English of diachronic /u/-lowering in proximity to a nasal consonant such as "come" (Old English: *cuman*), "honey" (Old English: *hunig*), "some" (Old English: *sum*), "hound" (cf. Dutch *hond*: /ɦɔnt/, German *Hund*: /hʊnt/), and "storm" (cf. German, *Sturm*: /ʃtʊɐm/). Some (Denham & Lobeck, 2010, p. 447) have suggested that the origin of such alternations is a scribe's substitution of "u" for "o" in order to differentiate the vowel in the orthography more clearly from the following letters "m," "n," or "r." But perhaps there is instead a phonetic explanation. Vowels are often nasalized in nasal contexts. This leads acoustically to the introduction of a low frequency nasal formant which in turn has the perceptual effect of high vowel lowering, the same form of lowering that is responsible for the inflections in French for "one" between the masculine *un* /œ̃/ with a phonetically mid-low and nasalized vowel and feminine *une* /yn/, which has an oral high vowel (Kawasaki, 1986; Krakow et al., 1988).

The *München* example can begin to provide a framework for many of the issues to be discussed in this chapter. There are obviously categorical changes over time (e.g., the substitution of /u/ for /y/; the deletion of consonants and vowels) – possibly also involving the creation of new phonemic oppositions (development of front rounded /y, œ/ contrasting with

front unrounded /i, e/) that can be related to everyday synchronic continuous variation in how sounds overlap with and influence each other. The task at hand is not just to show associations between diachronic and synchronic observations as sketched above, but to explain sound change using a model of human speech processing. Thus, the central question is, "What is it about the assembly of words out of their constituent sounds in speech production and their transmission to a listener that can occasionally result in sound change?" The focus is on words, rather than on how words are built into larger phrases because, as argued in Ohala (1993), the domain of sound change is in almost all cases the word or clitic phrases (Kiparsky, 2015) that have been lexicalized.

The chapter is organized as follows: the first section explains how sound change originates in non-contrastive phonetic variation. The following section addresses phonologization, i.e., how such phonetic variation can come to carry the cues for a phonological contrast. Subsequently, a section examines production and perception dynamics at the level of the individual and the group as sound changes progress. The next section presents an overview of the relevance of variations in hyper- and hypoarticulated speech for understanding sound change. A final section outlines how cognitive-computational agent-based models of sound change that combine theories of system dynamics and exemplar models of speech might enable the origin and spread of sound change to be brought together. Throughout the chapter, and consistent with prominent sound change studies (e.g., Beddor, 2009), we use the term "effect" to refer to contextual influences on a speech sound and "source" to refer to the speech sound that gives rise to them.

Sound change and coarticulation

One of the very important insights from the framework of articulatory phonology (Browman & Goldstein, 1992) and its forerunner action theory (Fowler, 1983; Fowler, 1984) is that speech production can be modeled as a constellation of orchestrated gestures that overlap with each other in time. Contextual vowel nasalization that occurs in English words like *ban* comes about according to this theory because of overlap between the independently controlled gestures of tongue and soft-palate movement. Although this overlap causes dramatic changes to the acoustic signal's formant structure during the interval marked by vowel voicing, there is experimental evidence to show that listeners typically hear acoustically nasalized vowels as oral, when they occur in the context of nasal consonants (Kawasaki, 1986; Krakow et al., 1988). This is presumed to be so according to articulatory phonology because there is a parity between the modalities of production and perception: listeners hear not a blending of orality and nasality but instead the interleaving of the independently controlled gestures in speech production (Fowler & Smith, 1986; Fowler & Thompson, 2010). Thus, the onset of (coarticulatory) nasality is not blended in human speech processing with the vowel (as the acoustic record would suggest) but perceptually associated or parsed with the source of coarticulation, i.e., with the following nasal consonant. From another perspective, the gestures of the oral vowel and the nasal consonant with which it overlaps are both produced and perceived in parallel and independently of each other.

According to Ohala (1993, 2012), sound change is but a drop in the ocean of synchronic phonetic variation precisely because listeners have become so adept at normalizing for context (Fujisaki & Kunisaki, 1976; Mann & Repp, 1980). As long as listeners associate or parse the coarticulatory effect – such as vowel nasalization in English – with the source (the following consonant) that gave rise to it, then no sound change can occur. But if such parsing should fail, then (for this example) some of the nasalization will be perceptually

attached to the vowel. There is then a perceptual switch between the listener's interpretation of the vowel as oral (if coarticulatory nasalization is parsed with the following consonant) to an interpretation as inherently nasal (if it is not). The term "inherently nasal" means that nasalization has become a phonological distinctive property of the vowel and has the potential to become lexically contrastive (as in, e.g., French), just as a contact between the tip of the tongue and alveolar ridge is a distinctive property of /t/ and contrastive in most languages. For Ohala, such a switch in perceptual re-interpretation is both abrupt and categorical, and can be likened to the illusion created by a Necker cube. Thus, just as a Necker cube creates uncertainty for the viewer due to a repeated switch between two plausibly (and categorically) different ways of interpreting shape, so, too, do ambiguities in the acoustic signal suggest two categorically different ways to the listener of parsing the signal into the gestures that produced it. The sound change is therefore in speech perception: notice that no change to speech production has actually taken place. Indeed, such a change would only have the potential to occur if the perceptually confused listener subsequently produced a nasalized vowel in their own speech. It is because the probability of sound change actually occurring is so low – requiring the listener not to normalize for context and to carry this over to speech production and then for this behavior to spread throughout the community – that the typical, and far more probable, state in Ohala's model is one of extensive phonetic variation but with scarcely any sound change. Such scarcity raises the question of why phonetic variation should turn into sound change in any one language (or variety of a language) but not in another (e.g., why /k/ was lost around the 17th–18th centuries in English "knight" and "knot," but not in German *Knecht* and *Knoten* that still have initial /kn/). This intriguing question, known as sound change actuation (Weinreich et al., 1968) is in Ohala's model not due to phonetic principles but is instead simply a matter of chance (see also Sóskuthy, 2015; Stevens & Harrington, 2014 for recent reviews).

The well-known sound changes that have been found in many of the world's languages, such as vowel nasalization (Beddor, 2012), tonogenesis (Hombert et al., 1979), domain-final voicing neutralization (Charles-Luce, 1985; Kleber et al., 2010), velar palatalization (Guion, 1998), nasal deletion or insertion before obstruents (Ohala, 1975, 1997), and /u/-fronting (Ohala & Feder, 1994; Harrington et al., 2011) are all examples in which there can be perceptual ambiguity in the extent to which a coarticulatory effect can be attributed to the source that caused it. These sound changes have two main other commonalities. First, most show an asymmetry in that $x \rightarrow y$ does not imply $y \rightarrow x$: for example, while there are several instances in which velars have historically palatalized before front vowels, there are perhaps none in the other direction. Such asymmetries are typically matched by an asymmetry in phonetic variation. Thus, there is some evidence that listeners are much more likely to confuse /k/ with anterior consonants such as /t/ before front vowels than the other way round (Winitz et al., 1972), principally because when the mid-frequency peak that characterizes acoustically fronted velars is obscured, the signal bears a close acoustic resemblance to that of an alveolar (Chang et al., 2001). Because in spontaneous speech there is a much greater likelihood for a mid-frequency peak to be obscured than for it to be inserted, the perceptual confusion is asymmetric. Second, all these sound changes originate – at least in terms of Ohala's model – from the different ways that a speaker and a listener associate phonological categories with the acoustic signal. A listener must have inserted perceptually a /p/ into the surname "Thompson" (see also "glimpse" from Old English *glimsian* related to "gleam") which was originally derived from "the son of Thom" (and importantly, not from "the son of Thomp"). Thus, there is a mismatch between the modalities: in production the surname originally had the categories /ms/ and the acoustic silence [p] arose because the velum was

raised before the lips were opened for /m/; in perception, this acoustic silence was reinterpreted as /mps/.

All of the types of sound change sketched so far come about in terms of Ohala's (1993) model because a listener does not parse coarticulation with its source. A much rarer type of sound change is dissimilation in which in Ohala's (1993) model listeners over-normalize for the contextual effects of coarticulation. Grassman's law is an example of dissimilation in which an initial aspirated consonant came to be deleted in ancient Greek preceding another aspirated consonant in the same word. Thus, the nominative and genitive singular for "hair" in ancient Greek are/t^{h}riks/ and /trik^{h}os/, respectively, suggesting that the latter was originally derived from /t^{h}rik^{h}os/ in which the first aspirated segment came to be deleted under the influence of the second. The interpretation in Ohala's model is that aspiration spreads throughout the entire first syllable due to the coarticulatory effect of the second stop. Listeners then factor out this coarticulatory influence but erroneously also factor out the initial aspirated segment. Thus, dissimilation happens because coarticulation perceptually camouflages the initial aspiration: since listeners are unable to distinguish between the two, all the aspiration including that which is an inherent distinctive property of the first segment is filtered out perceptually. This filtering uses the same mechanisms that listeners routinely apply to filtering out the effects of coarticulation, such as coarticulatory vowel nasalization as discussed above. One of the differences in the output between sound change due to under-normalizing (e.g., tonogenesis, development of phonological vowel nasalization) and over-normalizing (dissimilation) for coarticulation is that the former can create new sets of phonological contrasts that were not previously in the language (e.g., tonal contrasts in vowels as a consequence of tonogenesis). Another is that dissimilation typically applies to speech sounds that have a long time window (Alderete & Frisch, 2006), i.e., whose influence extends through several consonants and vowels (e.g., West, 2000 for liquids). Although Ohala's theory is elegant in linking a range of sound changes within the same mechanism of listener parsing errors, it has so far been difficult to demonstrate in the laboratory that dissimilation comes about as a consequence of over-normalization for coarticulation (Abrego-Collier, 2013; Harrington et al., 2016).

Metathesis is a sound change in which speech sounds swap serial position, e.g., English "burnt" versus German *brennt*. Blevins and Garrett (2004) suggest that the origin of perceptual metathesis is quite similar to that of dissimilation: in both cases the source is a sound that has a long time window. Whereas in dissimilation the long time window can mask another sound, in perceptual metathesis it causes confusion in their serial order. Thus, Blevins and Garrett (2004) note that a sound change in which a breathy voice and a vowel swap position i.e., CVfiC → CfiVC, comes about because the post-vocalic breathy voice causes breathiness throughout the vowel which obscures their serial order. Ruch and Harrington (2014) use this type of argument to explain the results of their apparent-time study of production data of a sound change in progress in Andalusian Spanish, in which older speakers have *pre*-aspirated stops derived originally from /s/ in words like pasta, /pa^{h}ta/, which younger speakers increasingly produce as *post*-aspirated /pat^{h}a/.

Sound change and phonologization

Phonologization in sound change is usually concerned with explaining how a phonological contrast is lost or attenuated from a source or conditioning environment and transposed to a coarticulatory effect. Thus, with regard to tonogenesis, the aim is to understand how a phonological stop voicing contrast /pa, ba/ (the source) is obliterated as higher and lower pitch

in the vowel following a voiceless and voiced stop, respectively (the coarticulatory effect) develops into the rising vs. falling tonal contrast in the vowel /pá, pà/.

In Kiparsky's (2015) model based on stratal optimality theory (Bermúdez-Otero & Hogg, 2003), new contrasts can be added to a language when constraints percolate upwards from the autonomous levels of post-lexical phonology, to word phonology, to stem phonology. The first stage of Kiparsky's three-stage model of phonologization involves a constraint operating at the post-lexical level. Post-lexical constraints can then become perceptually salient as they enter the superordinate level of word phonology as so-called quasi-phonemes, which differ from phonemes only in that they are not contrastive (i.e., they are allophonic). The third stage involves a change in phonological status from quasi-phonemes to phonemes, which comes about when the conditioning environment is reduced or deleted. In the development of umlaut in German (see chapter Introduction), the initial stages of sound change were in the earliest forms of Old High German in which, according to Kiparsky (2015), trans-consonantal vowel coarticulation was post-lexical. Umlaut could therefore apply across word boundaries to forms such as *mag iz* → *meg iz* ("may it") in which the /a/ → /e/ change was induced by trans-consonantal vowel coarticulation with /i/ in the next word. At a later stage in time, umlaut occurred only word-internally and not across word boundaries: this is the stage at which the phonetic variants due to umlaut, including presumably the front rounded vowels /ø, y/, would have entered the lexicon as quasi-phonemes. Their subsequent change in status from quasi-phonemes to phonemes comes about when the source is lost: that is when /i/ reduces to /ə/ to give present-day *Füße*, /fysə/ "feet" from Old High German /fotiz/.

Various studies by Solé (1992, 1995, 2007) have been concerned with establishing diagnostics to determine whether or not a coarticulatory effect has been phonologized. Solé reasons that the temporal proportion over which coarticulation extends should show fewer changes due to variation in speech rate if the coarticulation is (or is in the process of) being phonologized. In American English, there are reasons to suppose that vowel nasalization has been phonologized (Beddor et al., 2007) so that the main feature distinguishing pairs like "bent"/"bet" or "bend"/"bed" is not so much the presence of the nasal consonant in "bent" and "bend" but rather the temporal extent of nasalization in the vowel. In Spanish, there may well be coarticulatory nasalization in the vowel in VN sequences, but in contrast to American English nasalization has not been phonologized. Solé shows, using physiological techniques, that in American English, the temporal proportion of the vowel that is nasalized remains about the same, irrespective of whether the vowel shortens or lengthens due to rate changes. This is exactly what is to be expected if nasalization has become an inherent property of the vowel, i.e., has been phonologized. In Spanish, by contrast, the temporal extent of nasalization does not vary in proportion to vowel duration, as it does for American English, but instead begins at a fixed time prior to the onset of the nasal consonant irrespective of rate-induced changes to vowel duration: this suggests that nasalization is for Spanish (in contrast to American English) a property of the nasal consonant and not of the vowel.

Studies by Beddor (2009, 2012, 2015) confront more directly the association between the destruction of the source that causes the coarticulatory effect and the latter's phonologization. One of the main paths towards sound change for Beddor is that listeners typically do not parse all of the coarticulation with its source. Thus, if there is total normalization for the effects of context, then listeners should hear a nasalized vowel in an NVN context as oral (if they parse all of the coarticulatory nasalization in the vowel with the surrounding nasal consonants). However, the results in Beddor and Krakow (1999) suggest that for some listeners at least, normalization is only partial. If so, then only some (but not all)

of the nasalization will have been parsed with the N and the rest remains attached to the vowel: that is, listeners in their experiments associated *some* of the contextual nasalization with the vowel, even though the source (the N) was clearly present. If some of the coarticulatory nasalization is parsed with the vowel, then the vowel must sound partially nasal, and it is this partial compensation for a coarticulatory effect that makes sound change possible in Beddor's model (although sound change does not *necessarily* take place under such conditions).

One of the first stages of sound change for Beddor is when the coarticulatory effect and source enter into a trading-relationship, i.e., start to covary. When cues to a phonological contrast covary, then there is a coherent and systematic relationship between them, such that if one happens to be not especially salient in the speech signal, then listeners can direct their attention to the other cue with which it trades and so still identify the phonological contrast. For example, the more listeners rely on voice-onset-time as a cue to obstruent voicing, the less they rely on fundamental frequency (Repp, 1982). (Fundamental frequency is often higher following voiceless stops because of the greater laryngeal tension in order to suppress voicing: Löfqvist et al., 1989). Beddor showed that a perceptual trading relationship exists between the duration of anticipatory coarticulatory nasalization in the vowel and the duration of following nasal consonant in American English VNC_{voice} sequences in words like *send*. The empirical evidence for such a trading relationship was that listeners had difficulty hearing the difference between $\tilde{V}_S N_L$ and $\tilde{V}_L N_S$ where the subscript denotes Short or Long and where $\tilde{V}$ is the portion of the vowel that is nasalized: that is, a short nasalized vowel followed by a long nasal consonant was shown to be perceptually indistinguishable for many listeners from a long nasal vowel followed by a short nasal consonant. Just this is to be expected if the temporal extent of coarticulatory vowel nasalization and the duration of the final nasal consonant trade perceptually.

In American English $VNC_{voiceless}$ sequences in words like *sent*, the sound change by which nasalization is phonologized is at a more advanced stage than in VNC_{voice} sequences. This implies that vowel nasalization is likely to be more critical to listeners for distinguishing pairs like "set"/"sent" than "said"/"send," just as it is (but perhaps not quite to the same extent) for French listeners when they distinguish between *vais* vs. *vingt*, /vɛ, vɛ̃/ (1st pers. sing. "go" vs. "twenty"). Accordingly, listeners in Beddor et al. (2007) were shown to be especially sensitive to the beginning of nasalization in the vowel in "sent" irrespective of the duration of the following nasal consonant. Thus, the importance of Beddor's studies is that the path to phonologization may well initially develop out of a trading relationship between coarticulatory effect (vowel nasalization) and source (the following nasal consonant) which then gives way to perceptual attention being focused on the coarticulatory effect *irrespective of whether or not the source is present*.

Sound change and the association between perception and production

In Ohala's (1993) model, the conversion of phonetic variation to sound change happens via two separate stages. The first is perceptual, in which phonetic variation can be phonologized for the listener, often as a result of under- or (more exceptionally) over-normalizing for coarticulation. In the second stage, the listener may or may not carry over this perceptual shift to speech production. The implication of this model is therefore that, when a sound change develops, changes to the perceptual processing of coarticulation should in general precede those that take place in speech production.

Harrington et al. (2008) tested this idea by analyzing the production and perception of coarticulation for tense /u/-fronting (lexical set GOOSE) in Standard Southern British in which a sound change has been in progress since the 1960s. This was an apparent time study (Bailey et al. 1991) in which the extent of sound change was inferred by comparing younger and older speakers of this accent on the production and perception of coarticulation. The reason for analyzing coarticulation was that there is other evidence indicating that /u/-fronting in Standard Southern British was a coarticulation-induced sound change, i.e., that it originated in contexts in which /u/ is fronted synchronically (Harrington, 2007).

The results in Harrington et al. (2008) showed that coarticulation was matched across the modalities. This is because older speakers had a retracted /u/ and extensive coarticulation in production, and they also normalized i.e., compensated for these effects in perception. By contrast younger speakers' production was characterized by a fronted /u/ in all contexts (i.e., a minimal change due to coarticulation) and they normalized much less for coarticulation than older speakers.

The study in Harrington et al. (2008) does therefore not provide any evidence that changes to the perception of coarticulatory relationships precede those in production during an ongoing sound change in progress, as predicted by Ohala's (1993) model. But there is some evidence for this in Kleber et al.'s (2012) apparent-time study of the sound change in progress to lax /ʊ/-fronting (lexical set FOOT) in the same variety. Their results showed that, whereas the magnitude of coarticulation in /ʊ/ was about the same for both age groups in production, the degree to which listeners normalized for this coarticulatory effect was much less for younger than for older subjects. This finding can be interpreted in favor of a mismatch between the modalities in which perceptual normalization for coarticulation in younger subjects has waned ahead of a reduction in the size of coarticulation in production. This finding is therefore to a certain extent consistent with Ohala's (1993) model that changes to coarticulation in perception precede those in production. A more recent apparent-time study by Kuang and Cui (2016) provides evidence for a sound change in progress in the Tibeto-Burman language of Southern Yi in which tongue root differences are taking over from phonation as the primary cues to a tense/lax distinction. Compatibly with the results in Kleber et al. (2012), they also showed that this change is occurring in perception ahead of production.

Based on analyses of perception and production within the same individuals in two different types of sound change in American English and Afrikaans, Beddor's (2015) results are consistent with an alignment between the modalities during sound change. For nasalization in American English, Beddor shows that some individuals who were highly sensitive to vowel nasalization as a cue to the distinction between e.g., "sent"/"set" also showed extensive nasalization in "sent" words in their production. Similarly, for a sound change in progress in Afrikaans, in which fundamental frequency is taking over from voice onset time (VOT) as the primary cue to the stop voicing contrast, those older listeners who produced the contrast with voicing differences were also sensitive (and more so than younger listeners) to VOT as a cue in perception. Compatibly with Beddor (2015), a study of coarticulatory nasalization in 39 American English participants by Zellou (2017) showed a relationship between subjects' perceptual sensitivity to nasalization and the magnitude with which nasalization occurred in their own speech production. There is therefore no evidence from any of these studies on sound change in progress within the individual to suggest that changes to coarticulation in perception lead those in production during a sound change in progress.

The results of the un-merging of a phonological contrast in both Müller et al. (2011) and Bukmaier et al. (2014) are equivocal about whether production and perception are aligned

during a sound change in progress. Both of these studies were concerned with neutralizations that are common at least amongst older speakers in two German dialects: in Müller et al. (2011), the concern was with post-vocalic voicing neutralization such that standard German word pairs like *baden*/*baten* ("bathe"/"bid") are homophonous (both with lenis /d/) in Franconian; the focus of the analysis in Bukmaier et al. (2014) was on neutralization of the post-vocalic sibilant place of articulation such that the standard German /s, ʃ/ contrast (e.g., *wisst*/*wischt*, /wɪst, wɪʃt/, "knows"/"wipes") is neutralized as /ʃ/ in Swabian German. The apparent-time study by Müller et al. (2011) showed that younger Franconian participants produced and perceived the post-vocalic /t, d/ contrast to a much greater extent than did older Franconians but not as sharply as Standard German participants. Thus, the modalities for all three subject groups were matched (between no contrast, partial contrast, and complete contrast in older Franconians, younger Franconians, and Standard German participants respectively). In Bukmaier et al. (2014), there was some evidence that the magnitude of the distinction was greater in perception than in production for older participants. They showed that older Swabians < younger Swabians < standard German participants, where < denotes the degree to which participants distinguished between /s, ʃ/ in production. But concerning the magnitude of the /s, ʃ/ contrast in perception, the results showed older Swabians = younger Swabians < standard German participants. Thus, older Swabians perceived the /s, ʃ/ contrast at least as sharply as younger Swabians even though they scarcely distinguished between the sibilants in production.

A merger is a sound change in the other direction in which a phonological contrast collapses. There is some evidence that mergers take place in perception before production, i.e., that individuals report not being able to hear a contrast, even though they consistently show differences on the contrast in production (Kiparsky, 2016; Labov et al., 1991; Yu, 2007). On the other hand, the results of a series of studies on the near-merger in New Zealand English /iə, eə/ (lexical sets NEAR, SQUARE) toward /iə/ show a more complex picture. In one study, Warren et al. (2007) found a strong linear correlation between the extent to which New Zealanders distinguished these falling diphthongs in their own production and their ability to identify NEAR and SQUARE type words perceptually. But Hay et al. (2010) found that the degree to which such near-merged items were perceived to be distinct depended on a number of other factors including whether the perceptual task tapped lexical or phonological knowledge. The extent to which participants perceived the contrast in a speaker was also found to be influenced by their judgment of the speaker's age and social class (Hay et al., 2006): this finding also shows that phonological categorization in perception is influenced by memory and social information (see also Jannedy & Weirich, 2014). In a more recent study of the /e, a/ merger in "Ellen"/"Allen" words in New Zealand English and of the vowels in lexical sets LOT and THOUGHT in American English using both real and nonwords, Hay et al. (2013) note that they were not able to find any systematic relationship between an individual's production and perception.

Based on the findings reviewed in this section, it is not yet clear how perception changes with respect to production during a sound change in progress *within* the individual. As far as coarticulation is concerned, the equivocal results may derive from the considerable variation in processing coarticulation within and between individuals in both production and perception (Grosvald & Corina, 2012; Yu, 2010; Yu et al., 2013). The variable nature of the results concerning how the modalities are related within individuals may also be due to the nature of the experimental task that is conducted. Zellou (2017) suggests that rating the presence of a phonetic feature (e.g., the degree to which a vowel is nasalized) may tap into community-level phonetic norms to a greater extent than in discrimination tasks.

Regarding changes at the group level – which can be assessed by comparing the alignment of modalities between e.g., younger and older participants – the review of the previously mentioned studies provides some evidence that perception can lead production during a sound change in progress. This greater influence of sound change on perception may come about because of the increased need for the perceptual processing of speech to remain flexibly adaptable given the considerable variation across the different types of speakers, speaking styles, and accents to which a listener is constantly exposed (Beddor, 2015). Research is yet to identify the mechanisms by which speech production does eventually catch up with the faster changes in speech perception associated with a sound change in progress.

Finally, it is of course well-known from the quantal theory of speech (Stevens, 1972, 1989) that the relationship between speech production and perception is non-linear: incremental articulatory changes can lead to discrete differences in the acoustic output. A question that has not been sufficiently addressed is whether this nonlinear relationship is a factor that contributes to incremental phonetic variation becoming a categorical change. Perhaps during a sound change in progress, incremental articulatory change leads to an abrupt and marked change in acoustics and perception as a quantal boundary is crossed. This would be another instance in which perception and production are out of step with each other, at least during a sound change in progress, since perceptual changes at the boundary between quantal regions would be more marked than those in production. For example, it is well known that there is just such a quantal change between /u/ and /y/ produced respectively with back and front tongue dorsum positions (Stevens, 1972, 1989). Suppose a sound change in progress initially involves incremental tongue-dorsum fronting in a tongue-backed /u/. Then at some point in this gradual fronting, there will be a sudden acoustic (and perceptual) change at the quantal boundary (Figure 14.1): that is, as a quantal boundary is crossed, F_2 increases

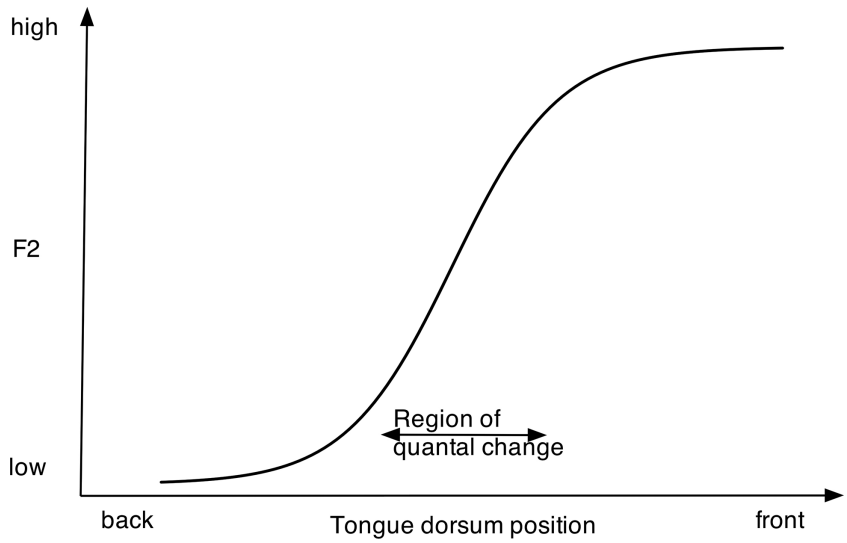

Figure 14.1 A schematic outline showing the quantal relationship (Stevens, 1972, 1989) between the tongue dorsum position (horizontal axis) and the second formant frequency (vertical axis) over the interval between a tongue back position for /u/ and tongue front position for /y/.

dramatically in relation to production. Sound change that is initially phonetic and gradient might become a category change (from /u/ to /ʉ/ or to /y/) once the incrementally gradual sound change has pushed the tongue dorsum sufficiently far forward into this quantal region so that the acoustic and perceptual change are suddenly quite marked.

This idea is largely untested. However, there is some evidence compatible with this hypothesis from the combined ultrasound and acoustic analysis of /l/-vocalization in American English in labial and lateral contexts in Lin et al. (2014). In /l/-vocalization, the tongue dorsum remains high but the tongue tip is lenited. But in this study, very small tongue-tip lenitions of just 1–2 mm were enough to cause quite a dramatic reduction in F_2 towards F_1. Perhaps it is this small articulatory but large acoustic change that is the key to understanding /l/-vocalization as a sound change resulting in e.g., French *autre* from Latin *alter* (cf. also English "false," French /fo/, *faux* < Latin *falsus*).

Sound change and hypoarticulation

Whereas Ohala sees the origin of many kinds of sound change in coarticulation, the main type of phonetic variation that causes sound change in Lindblom et al. (1995) is due to hyper- and hypoarticulation (H&H). According to the highly influential H&H theory (Lindblom, 1990), a speaker adapts speech production to the predicted needs of the listener for understanding what is being said. A word tends to be hypoarticulated – i.e., produced with minimal articulatory effort and lenited/reduced – if the speaker calculates that the listener will be able to predict the word from context; but it is hyperarticulated when the semantic and situational contexts provide only very limited clues to its meaning. In hypoarticulated speech, the attention of the listener is typically not directed at the signal both because it is an impoverished acoustic representation of its phonological structure and because there is no need to, if the word can be predicted from context. Sound change in Lindblom et al. (1995) exceptionally comes about if (for whatever reason) the listener's attention is focused on the signal, i.e., on the phonetic content during hypoarticulated speech. In this case, the hypoarticulated form can be added to the listener's lexicon. The model of Lindblom et al. (1995) is therefore a forerunner to some of the important ideas in episodic models of speech (Pierrehumbert, 2003, 2006), namely that a lexical item can be associated with multiple phonetic forms. It seems as if the theory of Lindblom et al. (1995) is restricted to predicting that only weak constituents of lexical items would be prone to change. However, this is not so since any word including its prominent syllables with full vowels can be hypoarticulated (if they are semantically predictable) and thereby subject to sound change in this model.

Both Lindblom and Ohala consider therefore that sound change comes about when the listener *decontextualizes* speech. In Ohala, the decontextualization is because coarticulation is interpreted independently of its source; in Lindblom et al. (1995), the decontextualization is because a mode of listening is engaged in hypoarticulated speech that is usually reserved for perceiving semantically unpredictable aspects of pronunciation. A prediction of both models is that less experienced listeners – perhaps children or second language learners – should be amongst the primary drivers of sound change if they have less ability to normalize perceptually for the effects of coarticulation and/or to vary listening strategies in response to the semantic predictability of the utterance. Both models also have answers to the following paradox raised by Kiparsky (2003). On the one hand, sound change according to the Neogrammarians (Osthoff & Brugman, 1878; Paul, 1886) proceeds largely imperceptibly, incrementally, and without the guidance of top-down processing from contrasts in the lexicon; but on the other hand, phonological systems across the known languages of the world

nevertheless fall into patterns (such that, as far as is known, most languages have some form of /i, u, a/ before complexifying the vowel system further). The paradox is that if the sounds of languages change imperceptibly and without regard to linguistic knowledge, then the types of phonological contrasts should be much more varied and unpredictable than the ones that are actually observed. That they are not is because the pool of phonetic variation is not infinite: in Lindblom's model there is a compromise between factors such as articulatory effort and perceptual distinctiveness, while in Ohala's model the possible sound changes are constrained by the types of coarticulatory overlap that the vocal tract is able to produce and that the listener is most likely to misperceive (i.e., phonological contrasts are unlikely to develop if they could not have developed from the synchronic way in which speech sounds overlap and influence each other).

The two models also differ in important ways that are illuminating for understanding sound change. For Ohala, the mismatch between the speaker and listener is in how phonological categories are associated with signal dynamics, whereas for Lindblom the mismatch relates to speaking style (hypoarticulated for the speaker, but hyperarticulated for the listener). Another difference is that Ohala's model is not concerned with how sound change spreads beyond an individual's grammar, whereas for Lindblom (1998) "it may be unnecessary to limit the phonetic contribution to sound change to the initiation stage" (p. 245). Yet another difference is that in Ohala's model, just as in the Neogrammarian view, sound change applies across the board to the words that are affected by it. In Lindblom on the other hand, sound change is word-specific because hypoarticulated forms of a particular word may be added to the lexicon. Lindblom et al. (1995) are somewhat vague about how sound change might then carry over to other words, but suggest that this will be a compromise between articulatory efficiency, perceptual distinctiveness, and evaluation by society for its social acceptability. Compatibly with some other studies (Bybee, 2002; Chen & Wang, 1975; Hay & Foulkes, 2016; Phillips, 2006), Lindblom et al.'s model contains the idea that sound change should initially take place in lexically frequent words. This is because lexically frequent words tend to be more predictable from context and therefore more likely to be hypoarticulated (which provides the conditions for sound change to occur in Lindblom et al., 1995). On the other hand, there is no association between sound change and lexical statistics in the model of Ohala (1993). Finally, Lindblom (1988) reasons that there are parallels between sound change and biological evolution. This is because in this model, sound change arises out of the variation caused by the flexible adaptations of the speaker to the needs of the listener or audience. On the other hand, Ohala (1988) provides several compelling reasons why sound change is not the same as biological evolution. One fundamental difference is that in Darwin's theory of evolution it is the individuals that are optimally adapted to their environment that are most likely to survive, whereas there is no evidence that there is competition between different pronunciation forms of words, nor that those variants that survive are the ones that are optimally adapted to their communication environment (Ohala, 1988, p. 177).

There is an interesting overlap between some of the predictions made from sociolinguistic typology in Trudgill (2011) and those of Lindblom's model. For Trudgill (2011), sound change that is internal to a community – i.e., which develops without any external contact from other varieties – is likely over a very long timespan to lead to greater complexity in phonological inventories in remote (i.e., with low contact), small, and socially tightly knit societies. This is because interlocutors are likely to be known to each other in such societies and because the range of topics is also to a large extent predictable. As a consequence, speakers should be able to deploy hypoarticulated speech extensively which, according to the model

of Lindblom et al. (1995), would then provide ample opportunity for radically hypoarticulated forms – that is, pronunciations which, if presented in isolation, would in all likelihood be unintelligible – to be added to the lexicon. Phonological complexification – which can take various forms, including the incorporation of typologically unusual contrasts – is in Trudgill's (2011) framework of sociolinguistic typology a direct consequence of the uptake over a long timespan of hypoarticulated speech in such remote, socially tightly knit communities. A challenge for future research will be, perhaps with the aid of computational modeling (see next section), to delimit more precisely the cognitive mechanisms by which phonological complexification emerges from hypoarticulated speech in low-contact, remote communities.

An unresolved issue is how hypoarticulation in semantically predictable parts of utterances might be related to phonologization. Recall that one of the major puzzles in phonologization is to explain how a coarticulatory effect becomes disassociated from its source: that is how vowel nasalization can become phonologized (i.e., function to distinguish between word meanings) with the loss of the following nasal consonant in some languages; or how umlaut develops from VCV coarticulation in words like *Gäste* ("guests") in present-day standard German from Old High German /gasti/ when the final /i/ (the source) that causes the vowel to be raised from /a/ to /ɛ/ (the effect) is bleached, resulting in present-day standard German /gɛstə/. In Kiparsky's (2015) model of sound change based on stratal OT mentioned earlier, phonologization comes about when the coarticulatory effect is promoted to quasi-phonemic and lexical status and is in a different stratum from the coarticulatory source that is post-lexical. Reductive processes (such as due to hypoarticulation) might then apply predominantly *post-lexically*, as a result of which the source would be eroded while the quasi-phonemic status of the coarticulatory effect at the superordinate *lexical* stratum would be unaffected.

A problem for this analysis is, however, more recent research showing that the attrition of the source interacts with lexical frequency. In their apparent-time study of coarticulatory nasalization in American English, Zellou and Tamminga (2014) found that lexically more frequent words tended to show the pattern like "sent" in Beddor's studies in which vowels were strongly nasalized coupled with an attrition of the source, the following nasal consonant. Similarly, the study by Lin et al. (2014) referred to earlier found that lenition of the tongue tip but maintenance of the tongue dorsum raising gesture as a path towards the sound change of /l/-vocalization is more likely in lexically frequent words such as "milk" than infrequent ones like "whelp." But lexical frequency is inherently lexical: consequently, the disassociation of the coarticulatory effect (which is preserved or enhanced) from the coarticulatory source (which is eroded) must refer to lexical information and cannot be explained by a post-lexical operation which has no access to lexical information, as in stratal OT phonology.

The explanation for the preservation or even enhancement (Hyman, 2013) of the coarticulatory effect but attrition of the source is likely to be phonetic rather than couched in terms of relegating the two to different lexical strata. More specifically, the phonologization of sound change may come about when hypoarticulation breaks the integrity of a consonant or vowel by leaving unaffected or even enhancing one set of cues and attenuating others. In many cases, the cues that are unaffected/enhanced are those due to the coarticulatory effect and the ones that are attenuated are associated with the coarticulatory source. This is so for the development of umlaut in /gasti/ → /gɛstə/ mentioned earlier. Trans-consonantal vowel coarticulation has the effect of shifting the first vowel (V_1) from /a/ to /ɛ/. Hypoarticulation also tends to shift /a/ in the direction of /ɛ/. This is because in a hypoarticulated

speaking-style – such as when the word is in an unaccented position or produced in a faster speaking style – vowel shortening is often accompanied by jaw raising (Beckman et al., 1992; Harrington et al., 1995) i.e., the vowel quality shifts in the direction of a phonetically raised vowel. *Thus, both coarticulation and hypoarticulation are additive* because they both cause phonetic V_1 raising. They are also additive in V_1 in Old High German / futiz/ → Standard German /fysə/ (*Füße*, "feet") because they both cause phonetic fronting: coarticulation because V_2 = /i/; and hypoarticulation because the high back vowel position is typically undershot in more casual speaking styles, which is manifested acoustically as F_2-raising (Moon & Lindblom, 1994; Harrington et al., 2011). Finally, while enhancing coarticulation in V_1, hypoarticulation simultaneously causes reduction and centralization of V_2 (= /i/) in both these cases: this is because V_2 is unstressed, i.e., occurs in a prosodically weak constituent.

Hypoarticulation also dismantles the integrity of the coda alveolar in the sound changes leading to vowel nasalization in *send* and to *l*-vocalization in *milk*. This is because once again there is one set of cues that, while not necessarily enhanced, is largely unaffected by hypoarticulation; and another set of cues whose effectiveness is compromised in a hypoarticulated speaking style. In *send*, coarticulatory vowel nasalization is unlikely to be affected by hypoarticulation. As Zellou and Tamminga (2014) show, the vowels in lexically more frequent words, which are more often hypoarticulated than their less frequent counterparts (Aylett & Turk, 2004; Wright, 2003), were found to be just as nasalized as vowels in words of low lexical frequency. In *l*-vocalization, the degree of tongue dorsum lowering is also largely unaffected by lexical frequency (and by extension hypoarticulation), as the previously mentioned study by Lin et al. (2014) demonstrates. On the other hand, as numerous studies have shown (Guy, 1980; Mitterer & Ernestus, 2006; Raymond et al., 2006; Zimmerer et al., 2014) alveolar consonants in coda position in words like *send* or *milk* are very often lenited or reduced.

The commonality across these different types of sound change is therefore that phonologization is brought about by the forces of hypoarticulation which enhances or leaves unaffected one set of cues while simultaneously destroying others, therefore dismantling the integrity of a phonological unit. The more general conclusion is that pragmatic variation drives sound change given that variations along an H&H continuum are very often made in response to the degree to which an utterance is semantically predictable from the dialogue and situational context in which it occurs. This idea therefore brings together pragmatics, phonologization, coarticulation, and hypoarticulation within a model of sound change.

The final question to be considered is whether sound change is more likely to be driven by normalizing or compensating for coarticulation in perception as in Ohala (1981, 1993) or – as suggested here and based on Lindblom et al. (1995) – by the forces of hypoarticulation. The different predictions of the two models can be assessed with respect to Figure 14. 2 which shows schematic psychometric curves when listeners provide categorical responses of either (high front) /y/ or (high back) /u/ to a synthetic continuum between these two vowels that has been embedded in a fronting t_t and non-fronting p_p contexts. As various studies have shown (e.g., Ohala & Feder, 1994; Lindblom & Studdert-Kennedy, 1967), there is a greater probability of hearing /u/ in a fronting consonantal context because listeners attribute some of the vowel fronting to the coarticulatory influence of the anterior consonant(s) and factor this out, i.e., they bias their responses towards /u/. According to Ohala (1981, 1993), the path to sound change is when listeners exceptionally do not compensate for coarticulation. In terms of the model in Figure 14.2, this would mean the following change to perception in a t_t context within the interval marked "normalize" in Figure 14.2: instead

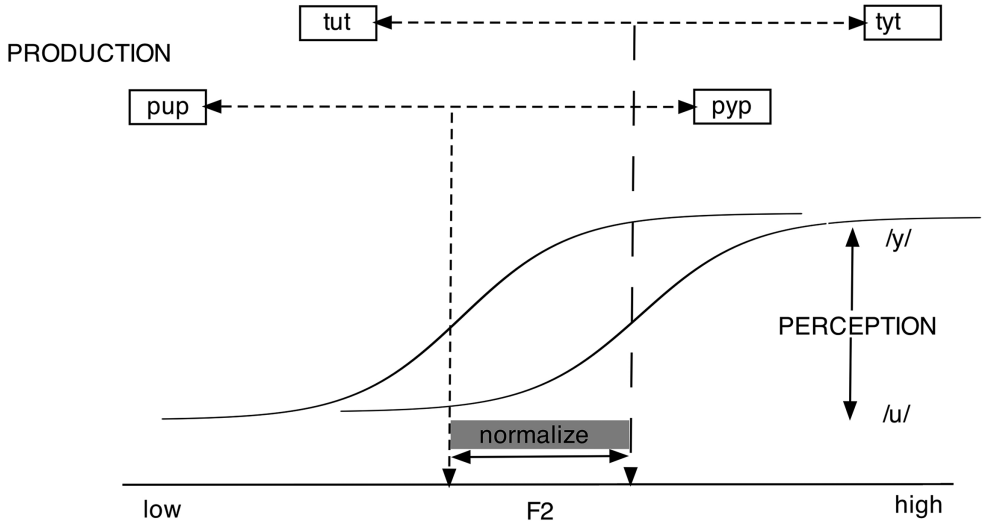

Figure 14.2 A schematic outline of the relationship between the production and perception of coarticulation. The upper part of the display illustrates the hypothetical positions of /pup, tut/ with low F_2 and with /pyp, tyt/ with high F_2. Because of the fronting effects of context, the positions in the t_t context are shifted towards higher F_2 values than for p_p. The lower panel shows the distribution of the corresponding perceptual responses under the assumption that the production and perception of coarticulation are exactly aligned. The degree to which listeners normalize for coarticulation in this model is shown by the gray shaded area marked "normalize" which extends between the two sigmoids' cross-over boundaries at which the probability of perceiving /u/ or /y/ are both 50%. Within this interval of ambiguity, a vowel is perceived as /y/ in a p_p context and as /u/ in a t_t context.

Adapted from Harrington et al. (2016).

of hearing /u/ within this interval in a t_t context, they would perceive /y/ if they no longer compensated for coarticulation. This is because perceptual responses would no longer be adjusted for (and biased towards /u/) to compensate for the coarticulatory fronting effects of the consonantal context.

Notice that in terms of this model, giving up compensating for coarticulation implies that the psychometric curve in the t_t context must shift to the left, i.e., towards the psychometric curve in the p_p context, if listeners ignore the fronting influence of the t_t context in making their judgments about vowel quality.

If, on the other hand, the sound change is driven by hypoarticulation, then listeners would notice high F_2 values in hypoarticulated speech but would not attribute them to the effects of this particular speaking style on vowel quality. Thus, high F_2 values, which may have been interpreted as /y/ are instead interpreted as /u/: that is, the range of what is considered to be an acceptable /u/ extends progressively towards higher F_2 values in a sound change led by hypoarticulation. In this case, it is the left boundary in Figure 14.2 that is shifted to the right as variants in non-fronting contexts that originally had low F_2 values catch up with the higher F_2 values of their coarticulated variants.

The issue that must be considered, then, is which of these alternatives is more plausible: whether it is the right boundary in Figure 14.1 that shifts to the left as listeners increasingly

judge vowels with a low F_2 in a t_t context to be /y/; or whether it is the left boundary that shifts to the right as listeners increasingly judge vowels with a high F_2 in a p_p context to be /u/. Two studies shed some light on this issue. In Standard Southern British, /u/ has fronted in the last 50 years or so (Henton, 1983; Hawkins & Midgley, 2005). Compatibly, the apparent-time study in Harrington et al. (2008) showed that participants from a younger age group typically had a fronted /u/ in production and that their cross-over boundary in perception between /u/ and /i/ was shifted toward the front so that on an /u–i/ continuum, younger listeners heard more /u/ than older listeners. But these age differences were confined to the non-fronting context. That is, the main difference in perception between the age groups was in the non-fronting *sweep-swoop* context, which was much more front for younger listeners. This result suggests therefore that in the last 50 years, the /u-i/ cross-over boundary in the non-fronting context has shifted from the back (as for older listeners) towards the front (as for younger listeners), just as would be expected in a model of diachronic /u/-fronting driven synchronically by hypoarticulation. In Harrington et al. (2016), first language child and adult speakers of German labeled /u-y/ continua (in which the difference in vowel fronting is phonologically contrastive) in precisely the contexts shown in Figure 14.2. The results showed that children's psychometric curves were closer together than for adults. If the children had normalized less for coarticulation, then this should have been manifested as response differences between the age groups in the fronting t_t context. Consistent with the apparent-time analysis in Harrington et al. (2008), the difference was present in the p_p context which was substantially fronted (towards t_t) for children compared with adults. This suggests that if children's lack of phonetic experience is in some way related to sound change, as Ohala (1993) suggests, then it is not because they have greater difficulty in compensating for coarticulation, but instead because they might over-estimate the extent of hypoarticulation in citation-form contexts of the kind presented in the experiment in Harrington et al. (2016). This over-estimation may come about because children and adults alike typically hear words in spontaneous speech in a hypoarticulated speaking style and rarely in this laboratory-style isolated word presentation; and so, children may not yet have learned sufficiently how to associate shifts in hyper- and hypoarticulation with speaking-style variation. The conclusion based on these results is the same as earlier: if children drive sound change then it is more likely to be because they have not yet had sufficient experience of the mapping between pragmatic meaning and spontaneous speech (e.g., Redford, 2009) that allows them to identify how a particular speaking style is positioned along the H&H continuum. Once again this result (based this time on language acquisition) points to the importance of the mapping between pragmatics and spontaneous speech as one of the drivers of sound change.

Sound change and agent-based modeling

In the last 50–60 years, there has been a fairly clear division between models concerned with the phonetic conditions that give rise to sound change (e.g., Ohala, 1993) as opposed to the social factors that cause the spread of sound change around a community of speakers (e.g., Eckert, 2012; Labov, 2001; Milroy, 1992). The general consensus has been that while phonetic factors in particular due to coarticulation and reduction provide the conditions by which sound change may take place, the spread of sound change is determined by social factors (Janda & Joseph, 2003). Thus, speakers might have knowledge of a social factor such as class as well as of certain spoken attributes that might characterize it (e.g., that London Cockney English has more *l*-vocalization than the standard accent of England). According

to this view of sound change, speakers preferentially copy the speaking style of the social category that they want to belong to. This view is characteristic of models of sound change in both Baker et al. (2011) and Garrett and Johnson (2013). Thus, Baker et al. (2011) in their analysis of /s/-retraction note that sound change may be started by speakers with extreme i.e., outlier forms of coarticulation; but that sound change becomes possible if the variation is conditioned by "social factors" in which there is a leader of linguistic change such as from upwardly mobile female speakers of the highest status social group. One of the computer simulations in Garrett and Johnson (2013) is built on the idea that "imitation is socially constrained" (p. 89) and that "a group that is aware of some social distance from another group may attend to phonetic deviations from the norm as marks of social differentiation" (p. 94).

There is, however, no reason to presume that the spread of sound change must be socially conditioned in this way. An alternative is that the propagation of sound change around a community of speakers derives from the principle of communication density (Bloomfield, 1933) and depends on which speakers talk to each other and how often (see also Labov, 2001, p. 24 for a similar view). Communication density is considered by Trudgill (2008a, 2008b) to be the main factor (at least in the earlier stages) that shaped the phonetic characteristics of New Zealand English, which bears a predictable relationship to the relative proportions of the different spoken accents of the first settlers to New Zealand in the 19th century. Similarly, the results of the longitudinal analysis of the annual Christmas broadcasts produced by Queen Elizabeth II in Harrington et al. (2000) suggest that the shift from her aristocratic accent in the direction of (but without attaining) a more middle-class accent came about, not because the Queen preferentially wanted to sound like one of the people, but instead because the Queen increasingly came into contact with persons of a middle-class accent during the decades (1960s, 1970s) in which a social revolution was taking place in England (Cannadine, 1998).

In other approaches derived rather more directly from communication density than either Baker et al. (2011) or Garrett and Johnson (2013), the spread of sound change at least in its initial stages is not a consequence of social factors but emerges instead from the propagation of different types of phonetic variation around a population of speakers and how these are modified during communication. According to a view based on communication density, sound change emerges from the often very slightly different ways in which speakers put their vocal organs and speech perception apparatus to use during speech communication. Thus, speakers constantly update their pronunciation during conversation without necessarily being socially selective about which aspects to imitate and which to ignore. Compatibly, while there is some evidence that phonetic imitation is socially selective (Babel, 2012; Babel et al., 2014), there are also studies showing that this need not be so (Delvaux & Soquet, 2007; Nielsen, 2011, 2014; Pardo et al., 2012) and that phonetic imitation may derive from the same mechanisms that cause non-speech alignments in posture, body movements, and sway (Fowler et al., 2008; Sebanz et al., 2006). Seen from this perspective, the issue to be modeled is how these types of inevitable phonetic modifications that arise at a microscopic level when individuals converse are related at a macroscopic level to community level, categorical change.

The problem of how microscopic phonetic variation and macroscopic sound change are connected cannot be solved from analyses of speech production or perception alone because it would require identifying sound changes before they have occurred as well as sampling from a large cross-section of the community longitudinally over a long time period. Partly for this reason, an alternative approach is to view language and language change in terms of the theory of systems dynamics: a theory often applied using computational simulation

in a variety of fields – biology, the economy, the environment (Meadows, 2015) to mention but a few. In systems dynamics, a system, as the famous parable of the Blind Men and the Elephant shows, cannot be understood just from the components out of which it is made: it requires a coherent model of the elements, their interconnections and how these are related to the overall purpose of the system. A characteristic feature of a system is that it is self-organizing (and often self-repairing) which means that an organizing structure emerges flexibly as a consequence of interacting elements (Oudeyer, 2006; Shockley et al., 2009). This idea is central to theories of emergent phonology in which macroscopic phonological categories emerge but are variably sustained as a consequence of microscopic interactions between the elements, in this case between the speakers of the community (Blevins & Wedel, 2009; de Boer, 2001; Lindblom et al., 1984; Studdert-Kennedy, 1998). Given that the speakers, the conversations that they have and how frequently they interact necessarily vary unpredictably, it follows that the association between phonological categories and the speech signals that sustain them is both stochastic and in constant flux. These stochastic variations occur because the system is inherently bi-directional with feedback (Wedel, 2007). It is bi-directional because, from a top-down point of view, phonology obviously shapes speech production output as well as judgments in speech perception (e.g., Hay et al., 2004); and because from a bottom-up point of view, speech output and perceived speech in this type of model shape phonological categories.

The phonetic variation might be small and category-internal, but under certain conditions the system self (re)organizes such that there is a phonological category change, i.e., sound change occurs at the level of the community. Establishing the conditions under which this type of change takes place – how for example top-down processing interacts with bottom-up changes to phonetic variation that might be caused by variations in the population – is of fundamental concern to research in this area.

The computational approach for solving this problem is typically agent-based modeling which is used to understand how the interaction between individuals that are represented by agents connected together in a network can bring about global (community) changes (Castellano et al., 2009). In research on sound change, the agent-based models draw upon many of the insights into human speech processing from usage (Bybee, 2002) and experienced-based (exemplar or episodic) models of speech (Pierrehumbert, 2003, 2006), precisely because such models are bi-directional in which phonological categories are updated by individuals' experiences in speech communication (Pierrehumbert, 2001; Blevins, 2004; Wedel, 2007).

In many agent-based models of sound change, the population might consist of a number of agents representing individuals. Each agent is typically equipped with a rudimentary lexicon, phonology, and parameterizations (e.g., formant frequencies, fundamental frequency etc.) of several stored speech signals per word. There is also usually some form of statistical, often Gaussian association between categories and parametric representations of the stored signals. Thus, a phonological category might be defined as a multidimensional Gaussian distribution (whose axes are the parameters) over the stored signals with which it is associated. Communication is often between an agent-talker and an agent-listener. One of the ways in which the agent talks is by generating a random sample from the statistical distributions of whichever categories are to be produced. The agent-listener may or may not add the perceived item (consisting of the generated sample and associated category labels) to memory depending on certain filtering conditions: in some models (Blevins & Wedel, 2009; Harrington & Schiel, 2017; Sóskuthy, 2015; Wedel, 2006), the item is not added if it is potentially confusable with another of the agent listener's categories (if for example, an agent-talker's /i/ is probabilistically closest to the agent listener's /u/). Models sometimes

also make use of some parameterized form of memory decay in order to remove items from the listener's memory. This can be important not just to offset the increase of items in memory that occur following a large number of interactions, but also to counteract the potentially infinite broadening, i.e., increase in the variance in the signals that make up a category with an increasing number of interactions.

The preceding is intended as a generic overview of agent-based modeling in sound change. The details vary quite considerably between studies. Thus, whereas in Blevins and Wedel (2009) and Sóskuthy (2015) there is only one agent that talks to itself, in Harrington and Schiel (2017) there are multiple agents based on real talkers, and in Kirby (2014) communication is from 100 learner to 100 teacher agents. Most models use artificially generated, static acoustic data as starting conditions (analogous to, e.g., formant values obtained at a single time slice from a theoretically derived, lossless model of the vocal tract); in Harrington and Schiel (2017) the starting conditions are dynamically changing parameters from real speakers. A Gaussian model is used in most studies to define the association between categories and signals, but in Ettlinger (2007) the statistical model is based on exemplar strength. In Harrington and Schiel (2017), the oldest exemplar is removed from memory each time an agent-listener adds a new one; in Ettlinger (2007), memory decay comes about by decrementing the strength of each exemplar exponentially over time (see also Pierrehumbert, 2001).

Agent-based models are typically designed to analyze specific aspects of sound change. In Ettlinger (2007), the main aim is to show that vowel chain shifting emerges as a natural consequence of stored and updated exemplars. The agent-based model in Kirby (2014) simulates a sound change by which fundamental frequency has taken over from duration as the main cue in distinguishing initial singleton stops vs. stop clusters with /r/ in the Phnom Penh variety of Khmer. Kirby's (2014) agent-based model is used to show that this type of change is driven by functional considerations, i.e., by a combination of the acoustic effectiveness of the cue for distinguishing between phonological categories combined with the extent to which the category distinguishes between lexical items. The purpose of the computational model in Blevins and Wedel (2009) is to explain why sound change often does not create homophones, especially if a pair of words that is about to merge acoustically cannot be further disambiguated by pragmatic information. They show how two phonological categories that are closely positioned in an acoustic space repel each other on the assumption that no update takes place from exemplars that are acoustically ambiguous between the categories. The agent-based model in Harrington and Schiel (2017) tested whether the phonetic approximation between two groups of speakers of Standard Southern British with retracted /u/ and fronted /ʉ/ was influenced by how these variants were oriented with respect to each in an acoustic space. They showed that, because older speakers' retracted /u/ was oriented toward that of younger speakers' fronted /ʉ/ (but not the other way round), the influence was correspondingly asymmetric, with a large shift following interaction in the older speaker's retracted variant towards the front of the vowel space. Stanford and Kenny (2013) used their agent-based model to test various aspects of Labov's (2007) theory of sound change by transmission, incrementation (brought about when children increment sound change from one generation to the next), and diffusion (brought about principally by contact between adults). Their model includes agents that represent adults (with stored speech knowledge) and children (without such knowledge) from two North American cities: from Chicago, in which a sound change, the Northern Cities vowel shift, is taking place; and from St. Louis where it is not. The model was set up to simulate travel (and therefore contact) between agents from the two cities. Only agents that were in close proximity could converse with

each other (and therefore influence each other's speech characteristics). Incrementation came about in their simulations because the agent children had fewer exemplars and so were less resistant to change. Diffusion arose because the St. Louis agent listeners learned the vowel chain shift from Chicago speakers imperfectly. Contrary to Labov (2007), the conclusion in Stanford and Kenny (2013) is that incrementation and diffusion are not due to different kinds of language learning, but instead both derive from exemplar learning (as outlined previously) and communication density, i.e., the frequency with which talkers represented by agents happen to communicate with each other.

Finally, there is the potential in an agent-based model to test the influence on sound change of different types of social network structures. Such research takes up the idea, explored predominantly within the sociolinguistics literature, that sound change might be differently affected depending on whether individuals are centrally or peripherally connected in the community (Borgatti et al., 2013; Mühlenbernd & Quinley, 2013; Stoessel, 2002; Wasserman & Faust, 1994). The computational simulation in Fagyal et al. (2010) suggests that leaders, i.e., those connected to many others, are drivers of sound change. On the other hand, Milroy and Milroy's (1985) study of Belfast English showed that sound change was less likely to occur in centrally connected members of a community (i.e., those with connections to many others); sound change was instead often caused by individuals with weaker ties to the community who introduced innovations from other communities with which they were associated. Compatibly, the computational model of Pierrehumbert et al. (2014) showed that highly connected individuals tended not to be instigators of sound change because their output was modulated by the very large number of connections to others (who might resist change). They suggest instead that linguistic change originates in tightly knit communities amongst individuals with only average connections but who tend to share innovations.

Concluding remarks

The sharp separation between the origin and the spread of sound change that was mentioned in the preceding section is to a certain extent a consequence of excluding social factors from cognitive models of speech processing, an approach that has typified speech research in much of the 20th century (see also Docherty & Mendoza-Denton, 2012; Docherty & Foulkes, 2014). The architecture suggested by episodic models in which there is a bi-directional, probabilistic association between phonological categories and speech signals that is incrementally updated through interaction provides the cognitive and computational architecture for bringing these strands of research together. This type of architecture can also begin to provide testable hypotheses concerning the actuation of sound change (Weinreich et al., 1968) that was mentioned in the section on sound change and coarticulation. This mercurial aspect of sound change is to be expected given a language model in which categories and signals are stochastically related and mutually updated by random interactions between individuals who, because they increment their phonological knowledge through experience, are also phonetically necessarily idiosyncratic (Laver, 1994, p. 66).

As Trudgill (2012) notes, sound change can cause what were once mutually intelligible spoken accents of a single language to evolve over a long time scale into separate languages. A somewhat neglected area of research lies in explaining quite how speech processing is incremented to produce this divergence in spoken accents that originally had a shared or at least quite similar phonology and marginally different phonetic characteristics. There is a gap in this area of research because most studies model the relationship between synchronic variation and diachronic change in terms of quite broad phonological features: how phonetic

variation leads to categorical changes in e.g., lenition, voicing, nasalization, palatalization, vowel height – and typically with an emphasis on patterns of change that are found across languages. But this type of analysis – mapping signals directly to distinctive phonological features – is too coarse-grained to explain spoken accent diversification. This is because spoken accents differ from one another at a much finer level of phonetic detail, especially when comparing sociolects of the same dialect (e.g., Docherty & Mendoza-Denton, 2012; Mendoza-Denton, 2008). Because this remarkable level of fine phonetic detail is nearly impossible to imitate in adulthood, spoken accent may in evolutionary terms have functioned as a tag for identifying imposters (Cohen, 2012). With databases of sufficiently large numbers of speakers and word items, it may be possible to model the steps by which sound change causes accents to diversify by using a cognitive-computational architecture based on episodes, incrementation, and feedback of the kind reviewed earlier. This is because this type of model provides and indeed predicts a stochastic link between phonological categories and precisely this very fine and nuanced level of phonetic detail that characterizes spoken accent and language differences (Pierrehumbert et al., 2000).

This review of the literature on sound change also suggests that this type of architecture needs to be extended in two ways. The first is by incorporating a model of hyper- and hypoarticulation in relation to pragmatic meaning, given the arguments presented in the section on sound change and hypoarticulation that it may well be this type of mapping that can shed new light on phonologization. The second is to incorporate associations between perception and production based more directly on the nonlinear mapping between speech physiology and acoustics (Stevens, 1989) possibly by incorporating the mathematics of nonlinear dynamics in associating phonological categories with speech output (e.g., Roon & Gafos, 2016). This will make it easier to test connections between the emergence of sound change from phonetic variation on the one hand and quantal jumps in acoustics due to incrementation in speech production on the other.

Acknowledgments

This research was supported by the European Research Council Advanced Grant no 742289 "Human interaction and the evolution of spoken accent" (2017–2022). Our thanks to Pam Beddor and to the editors for very many helpful comments on an earlier draft of this paper.

References

Abrego-Collier, C. (2013). Liquid dissimilation as listener hypocorrection. *Proceedings of the 37th annual meeting of the Berkeley linguistics society*, pp. 3–17.

Alderete, J., & Frisch, S. (2006). Dissimilation in grammar and the lexicon. In P. de Lacy (Ed.) *The Cambridge Handbook of Phonology*. Cambridge University Press: Cambridge.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47, 31–56.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, 40, 177–189.

Babel, M., McGuire, G., Walters, S., & Nicholls, A. (2014). Novelty and social preference in phonetic accommodation. *Laboratory Phonology*, 5, 123–150.

Bailey, G., Wikle, T., Tillery, J., & Sand, L. (1991). The apparent time construct. *Language Variation and Change*, 3, 241–264.

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.

Beckman, M., Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In G. Docherty & R. Ladd (Eds.) *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge University Press: Cambridge, pp. 68–86.

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

Beddor, P. (2012). Perception grammars and sound change. In M. J. Solé & D. Recasens (Eds.) *The Initiation of Sound Change: Perception, Production, and Social factors*. John Benjamin: Amsterdam, pp. 37–55.

Beddor, P. (2015). The relation between language users' perception and production repertoires. *Proceedings of the 18th international congress of phonetic sciences*, Glasgow, UK.

Beddor, P., Brasher, A., & Narayan, C. (2007). Applying perceptual methodsto phonetic variation and sound change. In M. J. Solé, et al. (Eds.) *Experimental Approaches to Phonology*. Oxford University Press: Oxford, pp. 127–143.

Beddor, P., & Krakow, R. A. (1999). Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *The Journal of the Acoustical Society of America*, 106, 2868–2887.

Bermúdez-Otero, R., & Hogg, R. (2003). The actuation problem in optimality theory: Phonologization, rule inversion, and rule loss. In D. Eric Holt (Ed.) *Optimality Theory and Language Change*. Kluwer: Dordrecht.

Blevins, J. (2004). *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge University Press: Cambridge.

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

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

Bloomfield, L. (1933). *Language*. Holt: New York.

Borgatti, S., Everett, M., & Johnson, J. (2013). *Analyzing Social Networks*. Sage Publications: London.

Browman, C., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49, 155–180.

Bukmaier, V., Harrington, J., & Kleber, F. (2014). An analysis of post-vocalic /s-ʃ/ neutralization in Augsburg German: Evidence for a gradient sound change. *Frontiers in Psychology*, 5, 1–12.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14, 261–290.

Cannadine, D. (1998). *Class in Britain*. Yale University Press: New Haven.

Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics*, 81, 591–646.

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*. Academic Press: San Diego, CA, pp. 79–101.

Charles-Luce, J. (1985). Word-final devoicing in German: Effects of phonetic and sentential contexts. *Journal of Phonetics*, 13, 309–324.

Chen, M., & Wang, W. (1975). Sound change: Actuation and implementation. *Language*, 51, 255–281.

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

de Boer, B. (2001). *The Origins of Vowel Systems*. Oxford University Press: Oxford.

Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica*, 64, 145–173.

Denham, K., & Lobeck, A. (2010). *Linguistics for Everyone: An Introduction* (2nd edition). Wadsworth: Boston.

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

Docherty, G., & Mendoza-Denton, N. (2012). Speaker-related variation – sociophonetic factors. In A. Cohn, C. Fougeron, & M. Huffman (Eds.) *The Oxford Handbook of Laboratory Phonology*. Oxford University Press: Oxford, pp. 44–60.

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.

Ettlinger, M. (2007). An exemplar-based model of chain shifts. *Proceedings of the 16th international congress of the phonetic science*, pp. 685–688.

Fagyal, Z., Escobar, A., Swarup, S., Gasser, L., & Lakkaraju, K. (2010). Centers and peripheries: Network roles in language change. *Lingua*, 120, 2061–2079.

Fowler, C. (1983). Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet. *Journal of Experimental Psychology General*, 112, 386–412.

Fowler, C. (1984). Segmentation of coarticulated speech in perception. *Perception and Psychophysics*, 36, 359–368.

Fowler, C., Richardson, M., Marsh, K., & Shockley, K. (2008). Language use, coordination, and the emergence of cooperative action. In A. Fuchs & V. Jirsa (Eds.) *Understanding Complex Systems*. Springer: Berlin, pp. 261–279.

Fowler, C., & Smith, M. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. Perkell & D. Klatt (Eds.) *Invariance and Variability in Speech Processes*. Erlbaum: Hillsdale, NJ, pp. 123–139.

Fowler, C., & Thompson, J. (2010). Listeners' perception of compensatory shortening. *Attention, Perception & Psychophysics*, 72, 481–491.

Fujisaki, H., & Kunisaki, O. (1976). Analysis, recognition and perception of voiceless fricative consonants in Japanese. *Annual Bulletin Research Institute of Logopedics and Phoniatrics*, 10, 145–156.

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

Grimm, J., & Grimm, W. (1854). *Das Deutsche Wörterbuch*. Online dictionary at: http://dwb.uni-trier.de/de/

Grosvald, M., & Corina, D. (2012). The production and perception of sub-phonemic vowel contrasts and the role of the listener in sound change. In M. J. Solé & D. Recasens (Eds.) *The initiation of sound change: Production, perception, and social factors*. John Benjamins: Amsterdam, pp. 77–100.

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

Guy, G. (1980). Variation in the group and the individual: The case of final stop deletion. In: W. Labov (Ed.) *Locating Language in Time and Space*. Academic Press: New York, pp. 1–36.

Harrington, J. (2007). Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts. In: J. Cole and J. Hualde (Eds.) *Laboratory Phonology 9*. Mouton: Berlin, pp. 125–143.

Harrington, J., Fletcher, J., & Roberts, C. (1995). Coarticulation and the accented/unaccented distinction: Evidence from jaw movement data. *Journal of Phonetics*, 23, 305–322.

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. *The Journal of the Acoustical Society of America*, 123, 2825–2835.

Harrington, J., Kleber, F., & Stevens, M. (2016). The relationship between the (mis)-parsing of coarticulation in perception and sound change: Evidence from dissimilation and language acquisition. In A. Esposito & M. Faundez-Zany (Eds.) *Recent Advances in Nonlinear Speech Processing*. Springer Verlag: Berlin, pp. 15–34.

Harrington, J., Palethorpe, S., & Watson, C. (2000). Does the queen speak the queen's English? *Nature*, 408, 927–928.

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.
Hawkins, S., & Midgley, J. (2005). Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association*, 35, 183–199.
Hay, J., Drager, K., & Thomas, B. (2013). Using nonsense words to investigate vowel merger. *English Language and Linguistics*, 17, 241–269.
Hay, J., Drager, K., & Warren, P. (2010). Short-term exposure to one dialect affects processing of another. *Language & Speech*, 53, 447–471.
Hay, J., & Foulkes, P. (2016). The evolution of medial /t/ over real and remembered time. *Language*, 92(2), 298–330.
Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review*, 23, 351–379.
Hay, J., Pierrehumbert, J., & Beckman, M. (2004). Speech perception, well-formedness, and the statistics of the lexicon. In *Papers in Laboratory Phonology VI*. Cambridge University Press: Cambridge, pp. 58–74.
Henton, C. (1983). Changes in the vowels of received pronunciation. *Journal of Phonetics*, 11, 353–371.
Hombert, J. M., Ohala, J., & Ewan, G. (1979). Phonetic explanations for the development of tones. *Language*, 55, 37–58.
Hoole, P., & Pouplier, M. (2017). Öhman returns: New horizons in the collection and analysis of articulatory imaging data. *Computer Speech and Language*, 45, 253–277.
Hyman, L. M. (2013). Enlarging the scope of phonologization. In A. Yu (Ed.) *Origins of Sound Change: Approaches to Phonologization*. Oxford University Press: Oxford, pp. 3–28.
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*. John Benjamins: Amsterdam, pp. 205–219.
Jannedy, S., & Weirich, M. (2014). Sound change in an urban setting: Category instability of the palatal fricative in Berlin. *Laboratory Phonology*, 5, 91–122.
Kawasaki, H. (1986). Phonetic explanation for phonological universals: The case of distinctive vowel nasalization. In J. Ohala & J. Jaeger (Eds.) *Experimental Phonology*. Academic Press: Orlando, FL, pp. 81–103.
Kiparsky, P. (2003). The phonological basis of sound change. In B. Joseph & R. Janda (Eds.) *The Handbook of Historical Linguistics*. Blackwell: Oxford.
Kiparsky, P. (2015). Phonologization. In P. Honeybone & J. Salmons (Eds.) *Handbook of Historical Phonology*. Oxford University Press: Oxford.
Kiparsky, P. (2016). Labov, sound change, and phonological theory. *Journal of Sociolinguistics*, 20, 464–488.
Kirby, J. (2014). Incipient tonogenesis in Phnom Penh Khmer: Computational studies. *Laboratory Phonology*, 5, 195–230.
Kleber, F., John, T., & Harrington, J. (2010). The implications for speech perception of incomplete neutralization of final devoicing in German. *Journal of Phonetics*, 38, 185–196.
Kleber, F., Harrington, J., & Reubold, U. (2012). The relationship between the perception and production of coarticulation during a sound change in progress. *Language & Speech*, 55, 383–405.
Krakow, R., Beddor, P., Goldstein, L., & Fowler, C. (1988). Coarticulatory influences on the perceived height of nasal vowels. *The Journal of the Acoustical Society of America*, 83, 1146–1158.
Kuang, J., & Cui, A. (2016). Relative cue weighting in perception and production of a sound change in progress. In *Proceedings of the 15th laboratory phonology conference*, Ithaca. Online dictionary at: https://labphon.org/labphon15/long_abstracts/LabPhon15_Revised_abstract_237.pdf
Labov, W. (2001). *Principles of Linguistic Change. Vol. 2: Social Factors*. Blackwell: Oxford.
Labov, W. (2007). Transmission and diffusion. *Language*, 83, 344–387.
Labov, W., Mark, K., & Corey, M. (1991). Near-mergers and the suspension of phonemic contrast. *Language Variation and Change*, 3, 33–74.
Laver, J. (1994). *Principles of Phonetics*. Cambridge University Press: Cambridge.

Lin, S., Beddor, P., & Coetzee, A. (2014). Gestural reduction, lexical frequency, and sound change: A study of post-vocalic /l/. *Laboratory Phonology*, 5, 9–36.
Lindblom, B. (1988). Phonetic invariance and the adaptive nature of speech. In B. A. G. Elsendoorn & H. Bouma (Eds.) *Working Models of Human Perception*. Academic Press: London, pp. 139–173.
Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H & H theory. In W. Hardcastle & A. Marchal (Eds.) *Speech Production and Speech Modeling*. Kluwer: Dordrecht, pp. 403–439.
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*. Cambridge University Press: Cambridge, pp. 242–264.
Lindblom, B., Guion, S., Hura, S., Moon, S. J., & Willerman, R. (1995). Is sound change adaptive? *Rivista di Linguistica*, 7, 5–36.
Lindblom, B., MacNeilage P., & Studdert-Kennedy, M. (1984). Self-organizing processes and the explanation of phonological universals. In B. Butterworth, B. Comrie, & O. Dahl (Eds.) *Explanation for Language Universals*. Mouton: Berlin, pp. 181–203.
Lindblom, B., & Studdert-Kennedy, M. (1967). On the role of formant transitions in vowel recognition. *The Journal of the Acoustical Society of America*, 42, 830–843.
Löfqvist, A., Baer, T., McGarr, N., & Story, R. (1989). The cricothyroid muscle in voicing control. *The Journal of the Acoustical Society of America*, 85, 1314–1321.
Mann, V., & Repp, B. (1980). Influence of vocalic context on the perception of [ʃ]-[s] distinction: I. Temporal factors. *Perception & Psychophysics*, 28, 213–228.
Meadows, D. (2015). *Thinking in Systems: A Primer*. Earthscan: London.
Mendoza-Denton, N. (2008). *Homegirls: Language and Cultural Practice Among Latina Youth Gangs*. Blackwell: Malden, MA.
Milroy, J. (1992). Social network and prestige arguments in sociolinguistics. In K. Bolton & H. Kwok (Eds.) *Sociolinguistics Today: International Pers*pectives. Routledge: London, pp. 146–162.
Milroy, J., & Milroy, L. (1985). Linguistic change, social network, and speaker innovation. *Journal of Linguistics*, 21, 339–384.
Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, 34, 73–103.
Moon, S. J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical Society of America*, 96, 40–55.
Mühlenbernd, R., & Quinley, J. (2013). Signaling and simulations in sociolinguistics. *University of Pennsylvania Working Papers in Linguistics*, Vol. 19, Article 16. Online dictionary at: http://repository.upenn.edu/pwpl/vol19/iss1/16
Müller, V., Harrington, J., Kleber, F., & Reubold, U. (2011). *Age-Dependent Differences in the Neutralization of the Intervocalic Voicing Contrast: Evidence from an Apparent-Time Study on East Franconian*. Interspeech: Florence.
Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, 39, 132–142.
Nielsen, K. (2014). Phonetic imitation by young children and its developmental changes. *Journal of Speech Language and Hearing Research*, 57, 2065–2075.
Ohala, J. (1975). Phonetic explanations for nasal sound patterns. In C. Ferguson, L. Hyman, & J. Ohala (Eds.) *Nasálfest: Papers from a Symposium on Nasals and Nasalization*. Language Universals Project: Stanford, pp. 289–316.
Ohala, J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, & M. F. Miller (Eds.) *Papers from the Parasession on Language and Behavior*. Chicago Linguistic Society: Chicago, pp. 178–203.
Ohala, J. (1988). Discussion of Lindblom's 'Phonetic invariance and the adaptive nature of speech'. In B. A. G. Elsendoorn & H. Bouma (Eds.) *Working Models of Human Perception*. Academic Press: London, pp. 175–183.
Ohala, J. (1993). The phonetics of sound change. In C. Jones (Ed.) *Historical Linguistics: Problems and Perspectives*. Longman: London, pp. 237–278.

Ohala, J. (1997). Aerodynamics of phonology. In *Proceedings of the 4th Seoul International Conference on Linguistics [SICOL]*, pp. 92–97.
Ohala, J. (2012). The listener as a source of sound change: An update. In In M. J. Solé & D. Recasens (Eds.) *The Initiation of Sound Change. Perception, Production, and Social factors*. John Benjamins: Amsterdam, pp. 21–36.
Ohala, J., & Feder, D. (1994). Listeners' identification of speech sounds is influenced by adjacent "restored" phonemes. *Phonetica*, 51, 111–118.
Öhman, S. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America*, 39, 151–168.
Osthoff, H., & Brugmann, K. (1878). *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen, Band I*. Verlag von S. Hirzel: Leipzig.
Oudeyer, P. Y. (2006). *Self-Organization in the Evolution of Speech*. Oxford University Press: Oxford.
Pardo, J., Gibbons, R., Suppes, A., & Krauss, R. (2012). Phonetic convergence in college roommates. *Journal of Phonetics*, 40, 190–197.
Paul, H. (1886). *Prinzipien der Sprachgeschichte* (2nd edition). Niemeyer: Halle.
Phillips, B. (2006). *Word Frequency and Lexical Diffusion*. Palgrave Macmillan: Basingstoke.
Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Eds.) *Frequency Effects and the Emergence of Lexical Structure*. John Benjamins: Amsterdam, pp. 137–157.
Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, 46, 115–154.
Pierrehumbert, J. (2006). The next toolkit. *Journal of Phonetics*, 34, 516–530.
Pierrehumbert, J., Beckman, M., & Ladd, D. R. (2000). Conceptual foundations of phonology as a laboratory science. In N. Burton-Roberts, P. Carr, & G. Docherty (Eds.) *Phonological Knowledge*. Oxford University Press: Oxford, pp. 273–303.
Pierrehumbert, J., Stonedahl, F., & Dalaud, R. (2014). A model of grassroots changes in linguistic systems. arXiv:1408.1985v1.
Raymond, W., Dautricourt, R., & Hume, E. (2006). Word internal /t, d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change*, 18, 55–97.
Redford, M. A. (2009). The development of distinct speaking styles in preschool children. *Journal of Speech, Language, and Hearing Research*, 52, 1434–1448.
Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, 92, 81–110.
Roon, K., & Gafos, A. (2016). Perceiving while producing: Modeling the dynamics of phonological planning. *Journal of Memory and Language*, 89, 222–243.
Ruch, H., & Harrington, J. (2014). Synchronic and diachronic factors in the change from pre-aspiration to post-aspiration in Andalusian Spanish. *Journal of Phonetics*, 45, 12–25.
Shockley, K., Richardson, D., & Dale, R. (2009). Conversation and coordinative structures. *Topics in Cognitive Science*, 1, 305–319.
Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, 10, 70–76.
Solé, M. (1992). Phonetic and phonological processes: The case of nasalization. *Language and Speech*, 35, 29–43.
Solé, M. (1995). Spatio-temporal patterns of velo-pharyngeal action in phonetic and phonological nasalization. *Language and Speech*, 38, 1–23.
Solé, M. (2007). Controlled and mechanical properties in speech: A review of the literature. In M. J. Solé, P. Beddor, & M. Ohala (Eds.) *Experimental Approaches to Phonology*. Oxford University Press: Oxford, pp. 302–321.
Sóskuthy, M. (2015). Understanding change through stability: A computational study of sound change actuation. *Lingua*, 163, 40–60.

Stanford, J., & Kenny, L. (2013). Revisiting transmission and diffusion: An agent-based model of vowel chain shifts across large communities. *Language Variation and Change*, 25, 119–153.
Stevens, K. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. David & P. Denes (Eds.) *Human Communication: A Unified View*. McGraw-Hill: New York, pp. 51–66.
Stevens, K. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17, 3–45.
Stevens, M., & Harrington, J. (2014). The individual and the actuation of sound change. *Loquens* 1(1), e003. http://dx.doi.org/10.3989/loquens.2014.003
Stoessel, S. (2002) Investigating the role of social networks in language maintenance and shift. *International Journal of the Sociology of Language*, 153, 93–131.
Studdert-Kennedy, M. (1998). Introduction: The emergence of phonology. In J. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.) *Approaches to the Evolution of Language*. Cambridge University Press: Cambridge, pp. 169–176.
Trudgill, P. (2008a). 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. (2008b). On the role of children, and the mechanical view: A rejoinder. *Language in Society*, 37, 277–280.
Trudgill, P. (2011). *Sociolinguistic Typology*. Oxford University Press: Oxford.
Trudgill, P. (2012). On the functionality of linguistic change. *Current Anthropology*, 53, 609–610.
Warren, P., Hay, J., & Thomas, B. (2007). The loci of sound change effects in recognition and perception. In J. Cole & J. I. Hualde (Eds.) *Laboratory phonology 9*. Mouton de Gruyter: Berlin, pp. 87–112.
Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press: Cambridge.
Wedel, A. (2006). Exemplar models, evolution and language change. *The Linguistic Review*, 23, 247–274.
Wedel, A. (2007). Feedback and regularity in the lexicon. *Phonology*, 24, 147–185.
Weinreich, U., Labov, W., & Herzog, M. (1968). Empirical foundations for a theory of language change. In W. Lehmann & Y. Malkiel (Ed.) *Directions for Historical Linguistics*, University of Texas Press: Austin, pp. 95–195.
West, P. (2000). Long-distance coarticulatory effects of British English /l/ and /r/: an EMA, EPG, and acoustic study. In *Proceedings of the 5th seminar on speech production: Models and data*. Online dictionary at: www.phonetik.uni-muenchen.de/institut/veranstaltungen/SPS5/abstracts/68_abs.html
Winitz, H., Scheib, M., & Reeds, J. (1972). Identification of stops and vowels for the burst portion of /p, t, k/ isolated from conversational speech. *The Journal of the Acoustical Society of America*, 51, 1309–1317.
Wright, R. (2003). Factors of lexical competition in vowel articulation. In J. Local, R. Ogden, & R. Temple (Eds.) *Papers in Laboratory Phonology VI*. Cambridge University Press: Cambridge, pp. 75–87.
Yu, A. (2007). Understanding near mergers: The case of morphological tone in Cantonese. *Phonology*, 24, 187–214.
Yu, A. (2010). Perceptual compensation is correlated with individuals' "autistic" traits: Implications for models of sound change. *PLoS One*, 5(8), e11950.
Yu, A., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality, and 'autistic' traits. *PLoS One*, 8(9), e74746.
Zellou, G. (2017). Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics*, 61, 13–29.
Zellou, G., & Tamminga, M. (2014). Nasal coarticulation changes over time in Philadelphia English. *Journal of Phonetics*, 47, 18–35.
Zimmerer, F., Scharinger, M., & Reetz, H. (2014). Phonological and morphological constraints on German /t/-deletions. *Journal of Phonetics*, 45, 64–75.