The production and perception of coarticulation in two types of sound changes in progress

JONATHAN HARRINGTON
FELICITAS KLEBER
ULRICH REUBOLD

In Fuchs, Susanne, Melanie Weirich, Daniel Pape & Pascal Perrier (eds.). 2012. Speech planning and dynamics (Speech production and perception). Frankfurt/M.: Peter Lang: Internationaler Verlag der Wissenschaften.

Abstract: This paper presents two studies that are concerned with the mechanisms by which historical sound change develops from synchronic coarticulatory variation with reference to the diachronic fronting of high back vowels in Standard Southern British and the development of a post-vocalic voicing contrast in the East Franconian variety under the influence of Standard German. The direction and extent of the sound changes in progress were inferred through comparisons between older and younger speakers on production and perception tasks. Both studies suggest that diachronic change has developed out of synchronic coarticulatory variation. We then differentiate this from other aspects of the sound change, including its spread to other contexts unrelated to the direction of the diachronic change and, for the East Franconian data, the development of a trading relationship between coarticulatory source and effect. Taken together, the results show that the complex of sound change is composed of many different parts that are activated in either production or perception at different stages in its development.

1 Introduction

Synchronic variation is both ubiquitous in languages and infinite: as much empirical research has shown, post-lexical variation including the apparent deletion of the final /t/ in phrases such as 'perfect memory' (Browman and Goldstein, 1990, 1995) or the vocalisation of /l/ in English varieties in certain prosodic positions (Scobbie and Pouplier, 2010) arises out of continuous processes of spatial reduction and temporal overlap. Synchronic processes such as the effects of consonant voicing on intrinsic pitch in the vowel (Löfqvist et al., 1989) or anticipatory vowel nasalisation due to a following nasal consonant have been

shown to be closely related to historical sound change such as tonogenesis (Hombert et al., 1979) or the development of phonological oral-nasal vowel contrasts (Beddor, 2007). But how exactly does continuous synchronic variation turn into categorical diachronic change? This issue is central both to developing a phonetic 'path' to sound change (Beddor, 2009) and more generally to investigations in laboratory phonology of how phonetic variability and phonological categories are inter-related.
Phonetic models of sound change are often based on the idea that listeners occasionally de-contextualise synchronic variation in speech production. According to Ohala (1993), sound change can originate when a listener fails to factor out or compensate for the effects of context: in this model, a fronted realisation of the vowel due to consonant-on-vowel coarticulatory influences in contexts such as /tut/ can be perceived as /y/ and not as /u/, should the listener fail to attribute the fronting to the coarticulatory source from which it originates, the alveolar consonants. In Lindblom et al. (1995), the listener's occasional decontextualisation of hypoarticulated variants that tend to occur at semantically redundant points in the speech signal can cause novel phonetic forms of words to be remembered: usually, the listener engages extensive top-down processing in reconstructing hypoarticulated speech production, and it is exceptionally when this top-down processing is disengaged that the hypoarticulated production of the word can be added to the listener's lexicon. One of the reasons why listeners might sometimes not interpret the speech signal in relation to the context in which it was produced is if they are inexperienced users of the language. As Ohala (1993) emphasises, only a small fraction of synchronic variation is converted to sound change, precisely because adult listeners are so adept at normalising for context, as studies of the perceptual compensation for coarticulation (Mann and Repp, 1980) suggest. Perhaps then inexperienced listeners like children might be frequent initiators of sound change because they have not fully acquired the knowledge for interpreting the speech signal in its context (Ohala, 1981), just as in vision they may take until the age of ten years to reach adult-like competency in the visual normalisation for size (Doherty et al., 2010).
Another reason why a listener's and speaker's interpretation of context may be divergent is because adults may have learned context-effects slightly differently. This possibility is suggested both by numerous pro-

duction studies showing that there is a great deal of variability even amongst speakers of the same variety in the magnitude and extent of coarticulation (van den Heuvel et al., 1996; Magen, 1997; Grosvald, 2009) and also by various perception studies showing that listeners can differ in the extent to which they normalise or compensate for coarticulation (Beddor, 2009). This variability in contextual normalisation is likely to be exacerbated when two varieties of a language or two different languages come into contact with each other, given the additional influences of variety (Scobbie and Pouplier, 2010) and language (Mok, 2010) on coarticulation and phonetic detail in both production (Öhman, 1966) and in perception (Beddor and Krakow, 1999; Beddor et al., 2002).

Some further mechanisms may be needed in order to explain how speaker- and listener-dependent variations in coarticulation actually develop into sound change. One of these may be that tracking (and by extension mis-tracking) fine phonetic detail and coarticulatory relationships is an integral part of perceiving speech (Alfonso and Baer, 1982; Fowler and Smith, 1986; Martin and Bunnel, 1982). Another is imitation of speech production that may derive from mimicry in human neonates (Meltzoff and Moore, 1997) and which may also be implicated in language acquisition and the development of the lexicon (Studdert-Kennedy, 2005). Adult imitation has been demonstrated in shadowing tasks (Goldinger, 1997; Shockley et al., 2004), in the production of isolated words (Babel, 2012; Delvaux and Soquet, 2007; Nielsen, 2007, 2011) and in conversational exchanges (Pardo, 2006; Pardo et al., 2012). Imitation in speech may be part of a more general tendency of entrainment in which postural sway is matched during a conversation even in the absence of any intention to entrain (Fowler et al., 2008; Shockley et al., 2009). Imitation may also be responsible for the shift in adults' pronunciation in the direction of diachronic changes over a long time-scale (Harrington et al., 2000). Perhaps diachronic change is propagated as adults imitate each others' slightly different coarticulatory relationships.

The general aim in this study is to explore whether two types of sound changes in progress can be modelled in terms of some of the mechanisms outlined above such as the decontextualisation of speech and imitation. The analysis for doing so is based on apparent time studies (Weinreich et al., 1968) in which the magnitude and direction of diachronic change are inferred through phonetic comparisons of two age groups that share

the same linguistic background. The first study in the following analysis was concerned with a sound change in progress in the standard accent of English in which high back vowels have fronted diachronically: the main aim here was to determine whether the diachronic change took place in equal measure in contexts to which it was, and was not, synchronically related and whether these changes have occurred in both speech production and perception. The focus of the second study is the development of a phonological voicing contrast in a variety of German under the influence of Standard German. Here the concern was to explore whether the changes take place in both perception and production and how such changes are related to synchronic processes of coarticulation in two age groups of the same variety.

2 Fronting of high back vowels in Standard Southern British

The starting point for comparing the shift in the production and perception of speech due to a sound-change in progress is the diachronic fronting of tense /u/(*who'd*) and lax /ʊ/ (*hood*) in the standard accent of English, Standard Southern British (SSB) which, some 40-50 years ago were mid-back vowels with phonetically fronted variants in the context of alveolar (e.g. *noon, soot*) and following initial palatal (e.g. *cue, few*) consonants (Gimson, 1966; Wells, 1982). Various acoustic and auditory analyses have shown that these vowels have fronted diachronically (Bauer, 1985; Hawkins and Midgley, 2005; Henton, 1983; de Jong et al., 2007; McDougall and Nolan, 2007) in the last 40-50 years. A recent physiological analysis has also shown that the movement of the tongue dorsum in present-day SSB /u, ʊ/ patterns much more closely with the front vowels /i, ɪ/ than with the back vowels /ɔ/ (*saw*) and /ɒ/ (*pod*) and that the extent of lip-rounding in /ɔ, ɒ, u, ʊ/ is quite similar (Harrington et al., 2011): taken together, these studies suggest that, whereas 40-50 years ago both a retracted tongue-dorsum position and greater lip-rounding distinguished /i, ɪ/ from /u, ʊ/ respectively, the main basis for the distinction between these pairs of vowels in present-day SSB is lip-rounding and not tongue-fronting. Apparent time studies are typically based on acoustic analyses of speech production of the kind shown

in Figure 1: these data, which are taken from Harrington et al. (2008) and Kleber et al. (2012) show the mean positions of the second formant frequency (F2) at the vowel target of tense /u/ and lax /ʊ/ in older and younger speakers of the same variety.

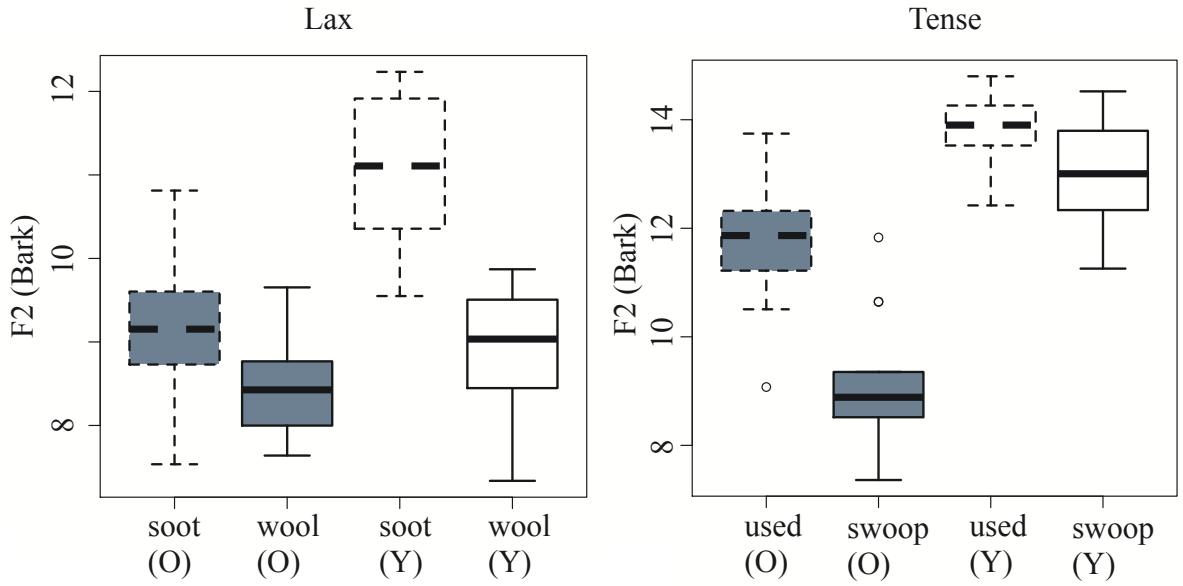

Figure 1: Boxplots (interquartile ranges) showing the distributions of the second formant frequency for younger and older speakers of a Standard Southern British variety. The data for lax vowels (left), which were taken from Kleber et al. (2012) include 18 younger subjects (9 females, 9 males) aged between 19 and 21 years and with an average age of 20.2 years and 15 older subjects (8 females, 7 males) aged between 54 and 89 years and with an average age of 75.4 years. The data for tense vowels (right), which were taken from Harrington et al. (2008), included 14 younger (Y) subjects (3 male, 11 female) aged between 18 and 20 years and with an average age of 18.9 years and 17 older (O) subjects (10 male, 7 female) aged over 50 years and with an average age of 69.2 years. Seven younger and nine older subjects who produced the lax vowels (left) had also participated in the study of tense vowels (right). The boxplots include one data point per subject that is a subject-mean at the point in the voiced vowel corresponding most closely to the voiced vowel target. The words whose distributions are shown in this figure were produced in isolation. The F2 values were extracted at the point in the vowel corresponding most closely to the acoustic vowel target.

.

These data are shown for two contexts: one of these is a fronting context and includes the tense vowel in *used* (past tense, /just/) and the lax vowel in *soot* (/sʊt/). In both cases, /u, ʊ/ are expected to front (synchronically) due to the coarticulatory influences of the flanking alveo-

lar or palatal consonants. The other is a non-fronting context that includes the tense vowel in *swoop* and the lax vowel in *wool*. Coarticulatory influences of the consonants on these words' vowels induce F2 lowering, either because of the effects of lip-rounding in the labial consonants and/or because of the tongue-dorsum retraction in /w/ and in the velarised /l/ that occurs syllable-finally in this variety.

Two trends are apparent in the data in Figure 1. The first is that, as expected, the variants in *used* and *soot* were evidently fronted relative to those in *swoop* and *wool*, as judged from the higher F2 values in the fronting context. The second is that F2 was lower for the older than for the younger speakers which suggests diachronic fronting in the tense and lax vowels. There was, however, also an interaction between the age groups, vowel tensity, and context. For the lax vowel data, the extent of diachronic fronting as inferred from comparisons across the age groups was greater in the fronting context than in the non-fronting context: the evidence for this is that older and younger speakers differed on F2 to a much greater extent in the variant of *soot* compared with that of *wool*. However, for the tense vowels, there is the opposite pattern: the older and younger speakers differed on F2 much more in the non-fronting variant *swoop* than they did in the fronting variant of *used*.

Figure 2 shows a model of the diachronic fronting in a hypothetical language that has /u/ in /tut/ (synchronically fronting) and /pup/ (non-fronting) contexts that will be used to account for the diachronic changes in these SSB vowels. In this model, the relationship between /tut/ and /pup/ is the same both before and after the sound change has taken place: thus, the only difference between the two contexts both before and after the diachronic change is due to synchronic coarticulation so that the vowel in /tut/ is synchronically phonetically fronted relative to that of /pup/. As the sound change takes hold, it affects /tut/ before it does /pup/ and, as a result, these variants come to be widely separated (middle panel). The completion of the sound change (between the middle and right panels) comes about when /u/ in non-fronting /pup/ 'catches up' diachronically with /tut/ so that the relationship between the two contexts is once again the same before (left panel) and after (right panel) the sound change. Figure 2 also shows how the sound change is presumed to be manifested across three generations relative to a hypothetical time-line between t=0 years when the sound change be-

gins and *t*=30 years when it is nearing completion. The sound change is initially manifested through contextual differences between the two generations *A* (aged 50 years) and *B* (aged 20 years) at *t*=0 years: the difference between /u/ in /tut/ and /pup/ is due only to synchronic coarticulation for generation *A* which is unaffected by the sound change; but for generation *B*, the difference between /u/ across the two contexts is greater because /u/ in /tut/ has additionally fronted diachronically. Thus according to this model, the variants between the fronting and non-fronting contexts are predicted to be widely spaced for the younger generation *B*, in comparison with those of the older generation *A* in the initial stages of a sound change.

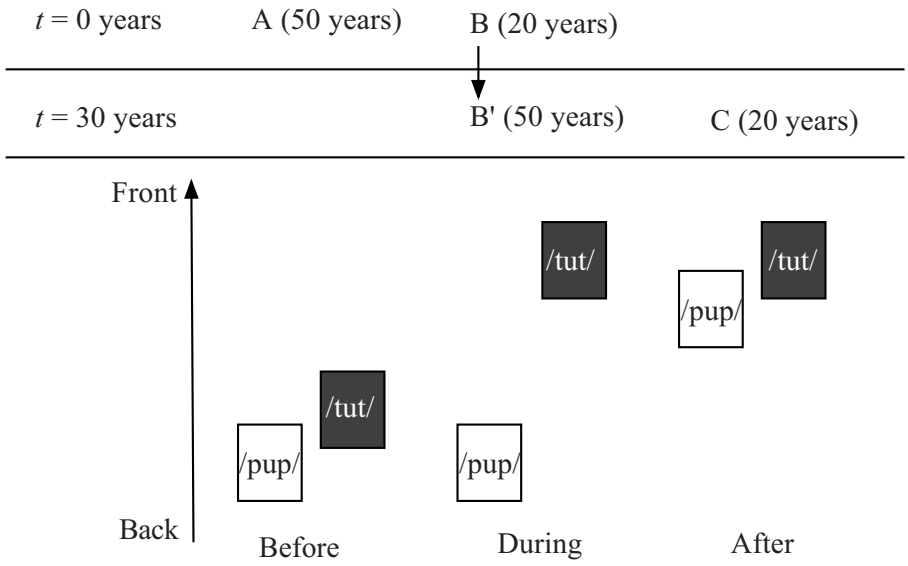

Figure 2: A model of the changes to /u/ in /tut/ (synchronically fronting) and /pup/ (synchronically non-fronting) contexts in a hypothetical language in which /u/ fronts diachronically from phonetically back to front. The stages shown are before the sound changes take place (left), during the sound change (middle), and close to the completion of the sound change (right). The top part of the figure shows how the sound change is expected to be manifested across three generations, *A, B, C* during its early stages when the sound change begins at *t*=0 years and close to its completion at *t*=30 years. The hypothetical ages of these three generations at these two timelines are shown (*B* and *B'* refer to the same speakers aged 20 years and 50 years respectively).

Just this relationship is presumed to characterise the difference across the generations in SSB /ʊ/ (left panel, Figure 1) in which the divergence between *soot* and *wool* was found to be much greater for the younger

(analogous to generation *B* in Figure 2) than it was for the older speakers (analogous to generation *A* in Figure 2). As the sound change nears completion (which is modelled to take place over a 30 year period), generation *B* now aged 50 years (shown as *B'* in Figure 2) is assumed to show similarly widely spaced /u/-variants in /tut/ and /pup/. But for the next generation *C*, the sound change /u/ in non-fronting /pup/ has shifted diachronically towards /u/ in /tut/: thus these two generations differ primarily in the relative position of /u/ in non-fronting /pup/ which is further back for the older generation *B'* relative to that of the younger generation *C* at time *t*=30 years. Such differences are characteristic of the data in the right panel of Figure 1 for the tense vowel /u/ which is a sound change in SSB that, according to Hawkins and Midgley (2005), took hold well before diachronic /ʊ/-fronting began (and is therefore closer to completion than for its lax counterpart): thus the variants in fronting *used* and non-fronting *swoop* in Figure 1 were found to be widely spaced for the older generation (analogous to generation *B'* of Figure 2) but narrowly spaced for the younger generation (analogous to generation *C* of Figure 2). The same speakers whose production data are shown in Figure 1 also participated in forced-choice perception experiments in which they labelled synthetic continua that differed only in the second formant frequency in minimal-pair fronting and non-fronting contexts (see Harrington et al., 2008; Kleber et al., 2012, for further details). For the tense vowel, the two synthetic contexts that were presented in the forced-choice perception experiment were *yeast-used* (fronting) and *sweep-swoop* (non-fronting); and for the lax vowel they were *sit-soot* (fronting) and *will-wool* (non-fronting). The results for psychometric curves and 50% cross-over boundaries averaged by age group show the well-known effect of perceptual compensation for coarticulation (Mann and Repp, 1980): evidently, the curves and boundaries were shifted to the left, towards the /i, ɪ/ end of the continua in the fronting context in both age groups, compatibly with their speech production data in which F2 was also raised in the fronting relative to that of the non-fronting context.

There is also a parity between the perception and production data for the tense vowel because both the interval separating the 50% cross-over boundaries in the two contexts (*yeast-used* and *sweep-swoop*) as well as the F2-distances between the corresponding variants in production (i.e.,

between those in *used* and *swoop*) were less for younger than they were for the older subjects. But there is no analogous parity between the modalities for lax vowels because, whereas the F2-distance between fronting (*soot*) and non-fronting (*wool*) contexts in production was evidently much greater for younger than for older speakers (Figure 1), younger listeners' psychometric curves and cross-over boundaries were marginally *closer* together than they were for older listeners (Figure 3).

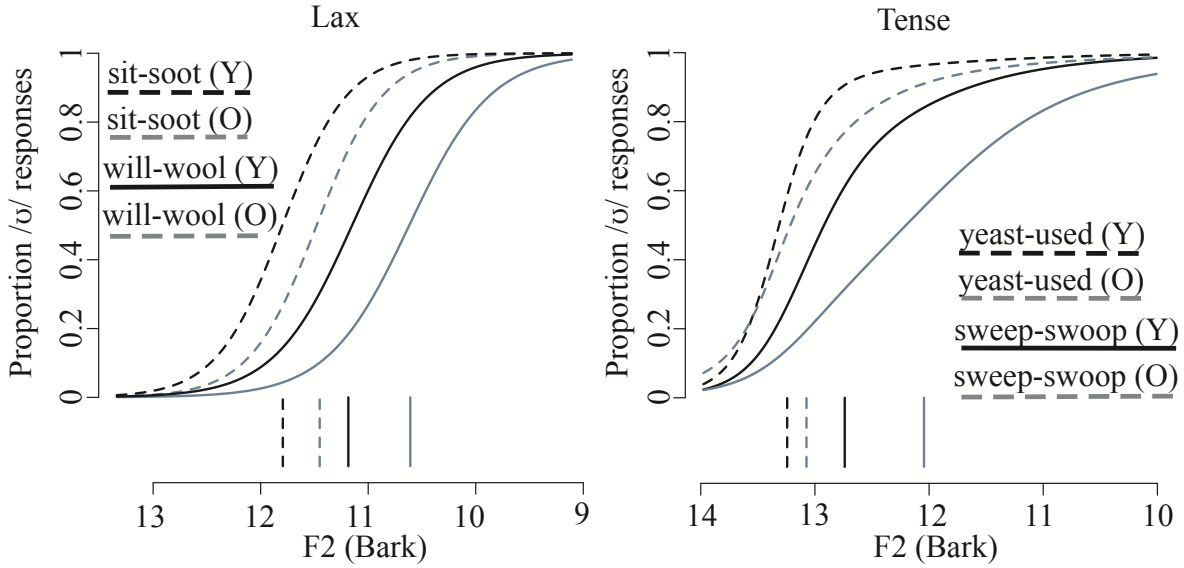

Figure 3: Averaged psychometric curves showing the proportion of /ʊ/ responses for the lax vowels (left) and the proportion of /u/ responses for the tense vowels (right) with the 50% decision boundaries marked as vertical lines for the same speaker groups whose production data are shown in Figure 1. The forced-choice responses were to stimuli created in the fronting and non-fronting contexts shown in the panels of the figures by reducing F2 in equal Bark steps over the range of values shown on the x-axis.

We argued earlier that diachronic fronting in production has applied scarcely at all to younger speakers' *wool*. But on the other hand, diachronic fronting must have applied perceptually in this non-fronting context because the 50% cross-over boundary in younger listeners' *will-wool* is positioned nearer to the /ɪ/-end of the continuum relative to that of older listeners and significantly so (t_{31}=4.6, p<0.001). From this it would follow that younger subjects' diachronic /ʊ/-fronting in contexts like *wool* is at a more advanced stage in perception than it is in production (see Kleber et al., 2012, for further details).

The present results, showing a production-perception asymmetry for the lax vowel data, are potentially inconsistent with two types of findings: firstly with compensation experiments in which speakers adjust their acoustic output due to artificially induced perturbations to auditory feedback (Munhall et al., 2009; Reubold, 2012); and secondly, with various studies suggesting that there is necessarily parity between the production and perception of coarticulation (Fowler, 2005). As far as the first of these is concerned, the results from compensation in speech production to altered auditory feedback and to the effects induced by bite-blocks (Lindblom et al., 1979) may be only marginally relevant to the present results given that they are concerned with short-term perturbations to on-line feedback whereas the data for the presented investigation are based instead on considerations of longer-term diachronic changes to the way in which coarticulatory relationships are perceived in relation to production. Of potentially greater relevance to the present results is the issue of whether parity between produced and perceived coarticulation can ever be violated. Our results, which suggest that it can be, are commensurate with other studies showing that listeners do not always compensate sufficiently in perception for coarticulation that occurs in production (Beddor, 2009; Fowler and Brown, 2000). Moreover, it is just this type of mismatch that can create an unstable relationship between the production and perception of speech that could lead to sound change (Kleber et al., 2012).

Finally, there is a third set of studies on near-mergers that are both similar to those found for the present SSB lax vowel study but also different. They are similar in that near mergers represent a stage in the development of sound change in which a measurable difference in speech production is not perceptible (Di Paolo and Faber, 1990; Faber and Di Paolo, 1995; Labov et al., 1972). Thus as Labov et al. (1991, p.46) note: "The existence of near-mergers puts into question the symmetry of production and perception. From the productive viewpoint, there are two categories; from a perceptual one, only one". At the same time, there are structural differences between the allophonic variation found in the present investigation on the one hand, and the variation between two phonemic categories in near-mergers on the other. Whether the same mechanism can account for these perception-production asymmetries in coarticulatory phonetic (the present study) and phonemic (near-mergers) re-

lationships will require further investigation.

3 Post-vocalic stop voicing in Franconian

The study in this section is concerned with a very different type of sound-change in the East Franconian variety of German which is spoken in a region of Northern Bavaria encompassing the cities of Nürnberg and Würzburg and extending northwards into Thuringia. In the standard variety of German, the ratio of the duration of the vowel to that of the following closure (henceforth the voicing ratio) is critical for the post-vocalic voicing distinction in minimal pairs such as *leiten* ('to lead') and *leiden* ('to suffer'): thus in the standard variety (and indeed in many other languages including English), post-vocalic voiceless stops are associated with a shorter vowel duration and longer closure duration (i.e., a lower voicing ratio) compared with those found in voiced stops; perceptually, the voicing ratio has been shown to be one of the most effective cues for the post-vocalic voicing distinction (Kohler, 1979). Some auditory studies have suggested that the voicing contrast is neutralized towards lenis, i.e. towards the voiced stop, in East Franconian so that words like *leiten/leiden* that are distinct in Standard German may be homophones in this variety (Barbour and Stevenson, 1990; Rowley, 1990).
The issue to be explored in the present paper is whether a post-vocalic voicing contrast may be developing diachronically as a result of an increasing influence of the Standard on the East Franconian variety. The analysis was based on an extension of a production and perception study by Müller et al. (2011) of 16 younger (3 male, 13 female; age range 15-25 years) and 16 older (4 male, 12 female; age range 51-74 years) subjects of an East Franconian variety of German. In addition, data were obtained from five speakers (4 male, 1 female; age range 25-52 years) and 21 (different) listeners of a standard variety (10 male, 11 female, 20-32 years) in order to assess whether any age-related changes in East Franconian were in the direction of the standard variety in either production or perception.
For the production data, the younger and older East Franconians as well as the five Standard speakers produced a number of isolated disyllabic words with a trochaic stress pattern and a tense vowel in the first sylla-

ble: these included both distracters and three sets of target word minimal pairs differing only in post-vocalic /t, d/ that occurs in the standard variety: these included *mieten/mieden* ('to rent'/'to avoid'), *baten/baden* ('asked for'/'to bathe') and *leiten/leiden* ('to lead'/'to suffer') which, in the standard variety, are often produced with nasal plosion as /miːtn, miːdn, laɪtn, laɪdn, baːtn, baːdn/. The same older and younger East Franconian subjects and the different set of 21 Standard subjects referred to above participated as listeners in two-alternative forced-choice labelling experiments in which they identified one of *leiten/leiden* or *mieten/mieden* from two continua that had been resynthesised by manipulating the voicing ratio in 11 equal steps between voiceless *leiten/mieten* (voicing ratios: *leiten*=0.665, *mieten*=0.478) and voiced *leiden/mieden* (voicing ratios: *leiden*=0.749, *mieden*=0.740) ends of the continuum. Only the voicing ratio was changed (through shortening and lengthening of the relative duration of the voiced vowel and consonant closure, respectively) between the stimuli: this ensured that the listeners depended on a single cue for the voicing distinction. The test words contained no schwa and all phonetic voicing (i.e. vocal fold vibration) during the closure was set to zero. The duration of the vowel+closure dyad was identical in all stimuli.

These data were used to address the following two issues: firstly, whether there was a trend for a greater differentiation between voiced and voiceless stops on the voicing ratio from older East Franconian to younger East Franconian to Standard speakers; and secondly, whether this trend was evident in equal measure in both the production and perception of speech. (In the presentation of the results below, we will henceforth refer to these three subject groups as *younger*, *older*, and *Standard* respectively). The results showed a significant trend in production (X^2_2=33.2, p<0.001) such that there was a progressively greater difference in voicing ratios between voiced and voiceless stops from older to younger to Standard speakers (Figure 4, left panel). The results from the forced-choice classification experiments (Figure 4, right panel) were to a certain extent compatible with those from production: there were progressively more pronounced S-shapes in the psychometric curves from older to younger to Standard listeners (X^2_2=25.8, p<0.001). Thus whereas the Standard listeners exhibited the most categorical-like response with an almost 100% identification of voiced and voiceless at either end of the continuum, the

older listeners showed continuous responses and labelled most stimuli with /d/ which indicates that this group perceptually neutralized the contrast towards lenis (Barbour and Stevenson, 1990).

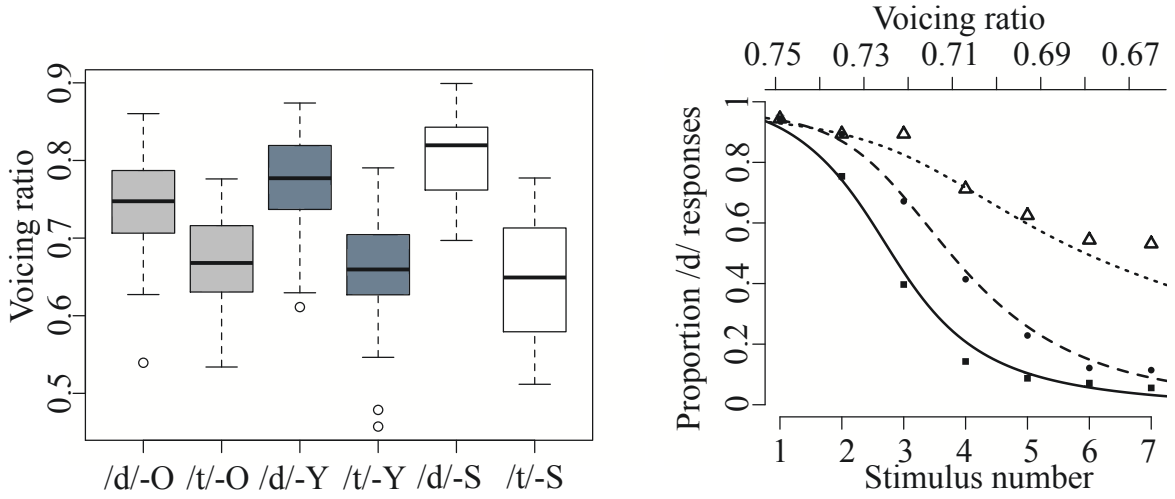

Figure 4: Boxplots (interquartile ranges) showing the voicing ratio calculated from V/(V+C) where V and C are the durations of the voiced vowel and consonant closure respectively for /t/ in isolated word productions of *baten*, *leiten*, and *mieten*, and for /d/ in *baden*, *leiden*, and *mieden* in older East Franconian (n = 16), younger East Franconian (n = 16), and Standard German (n = 5) speakers. The boxplots include three aggregated values (one per word type) per speaker. Right: Averaged psychometric curves showing the proportion of /d/ responses for the same speaker groups (dotted/triangles: older East Franconian; dashed/circles: younger East Franconian; solid/squares: Standard) whose production data are shown in Figure 1. The forced-choice responses were to stimuli created by changing the voicing-ratio in *leiten/leiden* and *mieten/mieden* in seven equal steps.

As far as the responses from the younger listeners were concerned, these were intermediate between those of the other two groups: that is, although their responses were categorical (as shown by the clear S-shape in the right panel of Figure 4), the categorical change was not as marked as it was for Standard listeners. The results from these production and perception experiments are compatible with a sound change in progress by which there has been a progressively greater differentiation in East Franconian between post-vocalic voiced and voiceless stops on the voicing ratio under the influence of the standard variety in which the distinction on this parameter is categorical.

The data also show that there was a clear difference between the older subjects on the one hand and the other two groups on the other in the extent to which the production and perception data were matched: whereas for the younger and Standard subjects the psychometric curves' 50% cross-over boundaries occurred close to voicing ratio values that separate their /t, d/ distributions in production (Figure 4, left panel), there was evidently no such correspondence between production and perception for the older subjects. Thus, on the one hand the older subjects' /t, d/ were separated at a voicing ratio value (of just over 0.7) that was quite similar to those of the other two groups in production; but on the other hand, older listeners had no 50% cross-over boundary within the voicing ratios presented in the perception experiment because, even at the 'voiceless' end of the continuum (stimulus number 7 in Figure 4, right panel) the older listeners' preferred response was /d/ and not /t/. Older subjects' difference between /t, d/ on the voicing ratio in production as opposed to their performance in perception in which they labelled almost the entire continuum as /d/ might suggest that the categorical development of the /t, d/ contrast in East Franconian has taken place in production before it has done so in perception. However, when voicing ratios were obtained from read speech in which the same words were embedded in longer passages that were constructed in such a way that the research objective (to measure the voicing distinction in minimal word pairs) would not be obvious to the reader, then, as Figure 5 shows, the distinction of post-vocalic /t, d/ on the voicing ratio collapsed (and was not significant) for the older speakers, but it was maintained for the other two groups. Thus, the apparent /t, d/ distinction in older speakers' isolated word productions is unlikely to have a phonological origin, given that the distinction collapsed in continuous speech and that they showed no categorical /t, d/ response in perception.

A possible explanation for older speakers' /t, d/ distinction on the voicing ratio in isolated word productions is that it was brought about as an indirect consequence of a strong oral release of /t/. It was evident to us that post-vocalic /t/ for many of the older East Franconian speakers was often produced with a strongly audible oral release (and they also exploited this cue in perception as subsequent tests that are not reported in this paper have indicated), whereas for the other two groups, the post-vocalic /t/ was often both weakly and nasally released. Perhaps then a

strongly and orally released /t/ was responsible for reducing its voicing ratio because of truncation by /t/ of the preceding vowel (which would produce a shorter vowel, thereby also decreasing the voicing ratio).

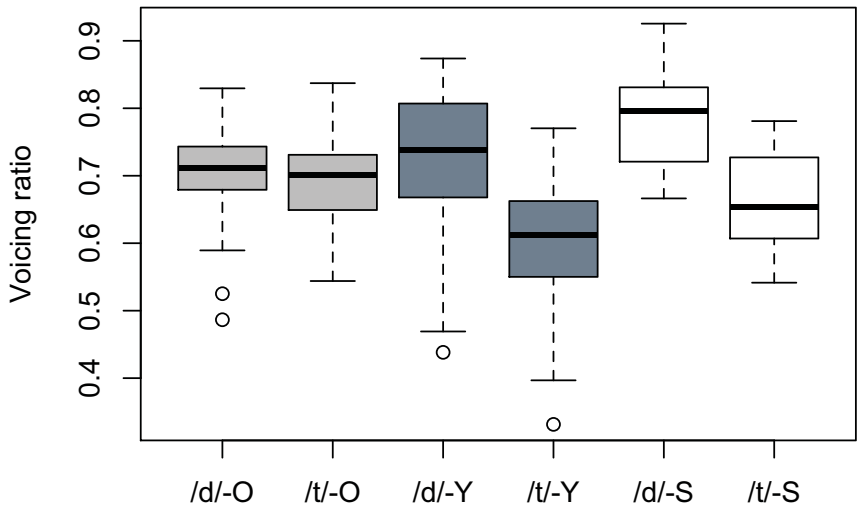

Figure 5: Boxplots (interquartile ranges) showing the voicing ratio for /t/ and /d/ produced in read speech for the same speaker groups and word types as shown in Figure 4.

There is independently of these considerations evidence from both physiological studies (Hoole and Mooshammer, 2002) and from studies of intonation (Grabe, 1998) that vowels in German are often truncated by following voiceless consonants and there is also some electromyographic evidence (cited in Lisker, 1974) to show that voiceless stops begin earlier in the vowel than do voiced stops. If the strength of the release of /t/ produced a shortening of the voicing ratio for older Franconian speakers, then these two variables should be negatively correlated: that is, stronger releases should be associated with shorter voicing ratios. Some evidence in support of this is shown in the left panel of Figure 6 in which voice onset time, measured from the release of the stop to the onset of periodicity in the second weak syllable of *baten, leiten*, and *mieten* was plotted as a function of the voicing ratio: for the older speakers, there was a general trend for longer /t/ releases to be associated with shorter voicing ratios. The relationship between these parameters was quite different for younger East Franconian (middle panel, Figure 6) and Standard (right panel, Figure 6) speakers who produced the postvocalic /t/

either with short voicing ratios or with longer releases of the stop: these two cues co-varied so that if the stop was unreleased (i.e. VOT is zero), then /t/ was distinguished from /d/ in production by correspondingly shorter voicing ratios.

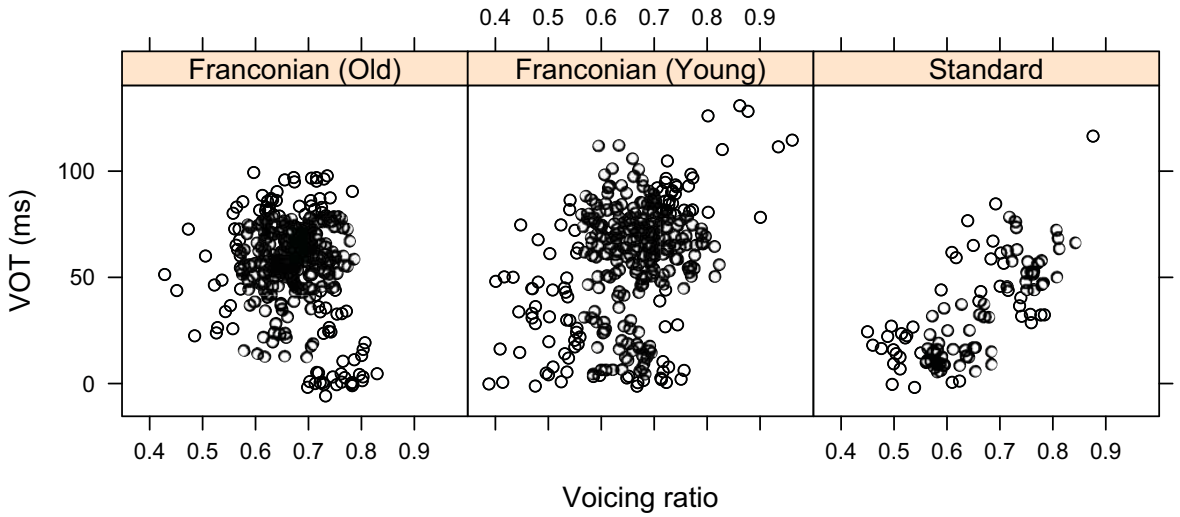

Figure 6: The voicing ratio as a function of voice onset time calculated from the stop release to the onset of periodicity in the following weak vowel for the same word types and speakers shown in Figures 3 and 4. Each point is an individual token (seven tokens per word type per speaker).

One of the reasons for the positive relationship between these parameters for younger and Standard speakers is that the voicing ratio and VOT may have entered into a trading relationship for these groups: of relevance here are findings by Jessen (1998) showing a perceptual trading relationship between stop closure duration and voice onset time in forced choice judgements by 20 listeners of Standard German to a post-vocalic /t-d/ continuum. It also seems clear from Figure 6 that there was a stronger positive relationship between VOT and voicing ratio (and thus closure duration) for the Standard than for the younger East Franconian speakers (although this may be an artefact of the greater number of speakers in the latter group) suggesting once again a developmental diachronic change from older to younger to Standard speakers.

The general conclusion from these data is that a phonetic distinction that is manifested in hyperarticulated isolated word productions in East Franconian has been phonologized under the influence of the standard

variety and that as part of this phonologization, the voicing ratio and the phonetic source that can give rise to voicing ratio differences in older East Franconians' isolated word productions have, as in the standard variety, entered into a trading relationship for younger East Franconian speakers.

4 General discussion

The two types of sound change, while very different, share the following characteristic: in both cases, existing synchronic variation due to coarticulation seems to have become magnified during a sound change in progress. In Standard Southern British, /u, ʊ/ are synchronically fronted due to coarticulation in certain contexts such as when they follow alveolar or palatal consonants. There is some evidence that the diachronic change has taken place in these fronting contexts ahead of that in non-fronting contexts: this was the explanation given both for the large acoustic separation in production between fronting and non-fronting /ʊ/ variants in younger speakers' variants and for the acoustic proximity between the age groups in the /ʊ/-variant of *wool* (whereas older and younger speakers' /ʊ/-variants in *soot* were far apart). As far as East Franconian is concerned, the existing synchronic, coarticulatory variation that has come to be magnified is the phonetic influence of the postvocalic /t, d/ contrast on the voicing ratio, that is on the ratio of the vowel duration to that of the vowel and following stop closure. This phonetic variation is likely to have existed in East Franconian because older speakers showed a voicing ratio distinction in (possibly hyper-articulated) isolated word productions, even though they were not as listeners able to make use of the voicing ratio perceptually for distinguishing post-vocalic voiced from voiceless stops. It is the exaggeration of this coarticulatory variation in production that leads diachronically to a categorical, phonological distinction between post-vocalic stops both in perception and in production.

Phonetic imitation of similar but not identical coarticulatory patterns may have propagated these sound changes in East Franconian. For example, both East Franconian and speakers of Standard German showed similar patterns of coarticulation because for both groups /t/ encroached

upon the vowel to a greater extent than it did on /d/. But they also differed: the extent to which /t, d/ affected the preceding vowel duration was greater for Standard German speakers than for older East Franconians. We would propose that it is the broad similarity between the groups in coarticulatory patterns that might provide the basis for the East Franconians to entrain to the Standard speakers: the outcome of this entrainment has been a shift in the direction of the Standard variety, as shown by the intermediate extent of voicing ratio distinctions in production and perception for younger East Franconians.

The spread to other contexts that are synchronically unrelated to the direction of the diachronic change may be another separate stage in the development of the sound change. Our evidence for this is that diachronic /u, ʊ/ fronting in labialized and velarized contexts (which is a change unrelated to the synchronic coarticulatory influences of these consonants on the vowel) in SSB may take place after it does in those contexts (alveolar, palatal) that influence the vowel synchronically in the same direction as the diachronic change. Our data also suggest that this extension to other contexts may come about in perception before it does in production: the evidence for this is that the perceptual boundaries between the fronting and non-fronting context for lax /ʊ/ were closer together for younger subjects than the coarticulatory differences between their fronting and non-fronting variants in production would suggest. Diachronic change may be slowed in production relative to perception in contexts such as these if it is opposed to coarticulatory influences of context for physiological reasons. In the present case, the requirements for the tongue dorsum to be back in the contexts of /w/ and syllable-final /l/ may extinguish the diachronic tongue-dorsum fronting in the vowel. This is especially likely in *wool* given firstly that the velarised /l/ that is typical in syllable-final position in this variety is resistant to coarticulation (Recasens and Espinosa, 2005) and given other evidence showing that segments that resist coarticulation also tend to exert strong coarticulatory influences in others (Fowler and Saltzman, 1993).

A further and quite separate component of sound change is the development of trading relationships between the coarticulatory source and effect as the diachronic change takes hold. Based on a number of perception experiments, Beddor (2009) has shown how trading relationships between anticipatory coarticulatory nasalization in the vowel and the

duration of the following nasal consonant (that is the source for these coarticulatory effects) may be one of the main mechanisms leading to the diachronic development of phonological vowel nasalization combined with the subsequent loss of the nasal consonant. Compatibly, our apparent-time study has suggested that diachronic change is accompanied by the development of a trading relationship between the effect (the voicing ratio) and the source that gives rise to it (the strength of the voiceless stop release). The reason why the development of such a trading-relationship is likely to lead to the waning and possible extinction of the source that produced the coarticulatory effect is because, whereas in older speakers the cues of the source and effect were additive in the sense that strongly released voiceless stops were accompanied by short voicing ratios that both cued /t/, they had instead a complementary association for younger East Franconian and Standard speakers such that short voicing ratios that provide cues for /t/ were matched with weak releases that do not. Consequently, the diachronic development of the voicing ratio as the main basis for the /t, d/ distinction necessarily results in the waning and possibly destruction of the source cue (the strong oral release of the /t/) that originally gave rise to voicing ratio differences, if the coarticulatory source and effect enter a trading relationship as the latter is phonologised during the sound change in progress.

The general conclusion from this analysis is that the complex of what is called sound change is made up of several different components. These include the phonetic conditions that give rise to the sound change, the development of the change through the entrainment of different speaker groups in contexts to which the diachronic change is related synchronically, its spread to other contexts possibly in perception ahead of changes in production, and the development in certain cases of trading relationships. Many more analyses of the production and perception of coarticulatory relationships for various types of sound changes in progress in different languages are needed to determine how these and other components that contribute to the complex of diachronic change are interconnected.

5 Acknowledgements

We are very grateful for the constructive criticisms of an earlier draft of this paper by two anonymous reviewers. This research was supported by German Research Council Grant No. HA 3512/3-2, "Sound change, lexical frequency, and variability: an experimental study of Southern British English, Received Pronunciation."

References

Alfonso, P. J. and Baer, T. (1982). Dynamics of vowel articulation. *Language and Speech*, 25:151–173.

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

Barbour, S. and Stevenson, P. (1990). *Variation in German: A Critical Approach to German Sociolinguistics.* Cambridge University Press, Cambridge.

Bauer, L. (1985). Tracing phonetic change in the received pronunciation of British English. *Journal of Phonetics*, 13:61–81.

Beddor, P., Harnsberger, J., and Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics*, 30(4):591–627.

Beddor, P. S. (2007). Nasals and nasalization: The relation between segmental and coarticulatory timing. In Trouvain, J. and Barry, W., editors, *Proceedings of the 16th International Congress of Phonetic Sciences*, pages 249–254.

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, 85(4):785–821.

Beddor, P. S. and Krakow, R. A. (1999). Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *Journal of the Acoustical Society of America*, 106(5):2868–2887.

Browman, C. and Goldstein, L. M. (1995). Gestural syllable position effects in American English. In Bell-Berti, F. and Raphael, L. J., editors, *Producing Speech: Contemporary Issues*, pages 19–33. AIP Press, Woodbury NY.

Browman, C. P. and Goldstein, L. M. (1990). Gestural specification using dynamically defined articulatory structures. *Journal of Phonetics*, 18:299–320.

de Jong, G., McDougall, K., Hudson, T., and Nolan, F. (2007). The speaker discriminating power of sounds undergoing historical change: A formant-based study. In

Trouvain, J. and Barry, W., editors, *Proceedings of the 16th International Congress of Phonetic Sciences*, pages 1813–1816.

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

Di Paolo, M. and Faber, A. (1990). Phonation differences and the phonetic content of the tense-lax contrast in Utah English. *Language Variation and Change*, 2:155–204.

Doherty, M. J., Campbell, N. M., Tsuji, H., and Phillips, W. (2010). The Ebbinghaus illusion deceives adults but not young children. *Developmental Science*, 13(5):714–721.

Faber, A. and Di Paolo, M. (1995). The discriminability of nearly merged sounds. *Language Variation and Change*, 7:35–78.

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

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

Fowler, C. and Smith, M. (1986). Speech perception as 'vector analysis': An approach to the problems of invariance and segmentation. In Perkell, J. S. and Klatt, D., editors, *Invariance and Variability in Speech Processes*, pages 123–139. Lawrence Erlbaum Associates Inc., Hillsdale NJ.

Fowler, C. A. (2005). Parsing coarticulated speech in perception: Effects of coarticulation resistance. *Journal of Phonetics*, 33:199–213.

Fowler, C. A. and Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech*, 36:171–195.

Gimson, A. C. (1966). *An Introduction to the Pronunciation of English.* E. Arnold, London.

Goldinger, S. (1997). Speech perception and production in an episodic lexicon. In Johnson, K. and Mullennix, J. W., editors, *Talker Variability in Speech Processing*, pages 33–66. Academic Press, New York.

Grabe, E. (1998). Pitch accent realization in English and German. *Journal of Phonetics*, 26(2):129–143.

Grosvald, M. (2009). Interspeaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation. *Journal of Phonetics*, 37(2):173–188.

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

Harrington, J., Kleber, F., and Reubold, U. (2011). The contributions of the lips and the tongue to the diachronic fronting of high back vowels in Standard Southern British English. *Journal of the International Phonetic Association*, 41(2):137–156.

Harrington, J., Palethorpe, S., and Watson, C. (2000). Monophthongal vowel changes in Received Pronunciation: An acoustic analysis of the Queen's christmas broadcasts. *Journal of the International Phonetic Association*, 30:63–78.

Hawkins, S. and Midgley, J. (2005). Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetics Association*, 35:183–199.

Henton, C. G. (1983). Changes in the vowels of Received Pronunciation. *Journal of Phonetics*, 11:353–371.

Hombert, J.-M., Ohala, J. J., and Ewan, W. (1979). Phonetic explanations for the development of tones. *Language*, 55(1):37–58.

Hoole, P. and Mooshammer, C. (2002). Articulatory analysis of the German vowel system. In Auer, P., Gilles, P., and Spiekermann, H., editors, *Silbenschnitt und Tonakzente*, pages 129–152. Niemeyer, Tübingen.

Jessen, M. (1998). *Phonetics and Phonology of Tense and Lax Obstruents in German.* John Benjamins, Amsterdam and Philadelphia.

Kleber, F., Harrington, J., and Reubold, U. (2012). The relationship between the perception and production of coarticulation during a sound change in progress. *Language and Speech*.

Kohler, K. J. (1979). Dimensions in the perception of fortis and lenis plosives. *Phonetica*, 36:332–343.

Labov, W., Karen, M., and Miller, C. (1991). Near-mergers and the suspension of phonemic contrast. *Language Variation and Change*, 3:33–74.

Labov, W., Yaeger, M., and Steiner, R. (1972). *A Quantitative Study of Sound Change in Progress.* Philadelphia, U.S. Regional Survey.

Lindblom, B., Guion, S., Hura, S., Moon, S.-J., and Willerman, R. (1995). Is sound change adaptive? *Rivista Di Linguistica*, 7:5–37.

Lindblom, B., Lubker, J., and Gay, T. (1979). Formant frequencies of some fixed-mandible vowels and a model of speech motor programmming by predictive simulation. *Journal of Phonetics*, 7:147–161.

Lisker, L. (1974). On time and timing in speech. In Sebeok, T. A., editor, *Current Trends in Linguistics*, volume 12, pages 2387–2418. Mouton, The Hague.

Löfqvist, A., Baer, T., McGarr, N., and Story, R. (1989). The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America*, 85:1314–1321.

Magen, H. S. (1997). The extent of vowel-to-vowel coarticulation in English. *Journal of Phonetics*, 25(2):187–205.

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

Martin, J. G. and Bunnel, H. T. (1982). Perception of anticipatory coarticulation effects in vowel-stop-vowel sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 8:473–488.

McDougall, K. and Nolan, F. (2007). Discrimination of speakers using the formant dynamics of /u:/ in British English. In *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, Germany*, pages 1825–1828.

Meltzoff, A. N. and Moore, M. K. (1997). Explaining facial imitation: A theoretical model. *Early Development and Parenting*, 6:179–192.

Mok, P. K. P. (2010). Language-specific realizations of syllable structure and vowel-to-vowel coarticulation. *Journal of the Acoustical Society of America*, 128(3):1346–1356.

Müller, V., Harrington, J., Kleber, F., and Reubold, U. (2011). Age-dependent differences in the neutralization of the intervocalic voicing contrast: Evidence from an apparent-time study on East Franconian. In *Proceedings of Interspeech, Florence*.

Munhall, K. G., MacDonald, E. N., Byrne, S. K., and Johnsrude, I. (2009). Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. *Journal of the Acoustical Society of America*, 125:384–390.

Nielsen, K. (2007). Implicit phonetic imitation is constrained by phonemic contrast. In Trouvain, J. and Barry, W., editors, *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbruecken*, pages 1961–1964.

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, 39:132–142.

Ohala, J. J. (1981). The listener as a source of sound change. In Masek, C. S., Hendrick, R. A., and Miller, M. F., editors, *Papers from the Parasession on Language and Behavior*, pages 178–203. Chicago: Chicago Ling. Soc.

Ohala, J. J. (1993). Sound change as nature's speech perception experiment. *Speech Communication*, 13:155–161.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, 39:151–168.