Jonathan Harrington*, Felicitas Kleber, Ulrich Reubold
and Jessica Siddins

The relationship between prosodic weakening and sound change: evidence from the German tense/lax vowel contrast

Abstract: The study tests a model of sound change based on how prosodic weakening affects shortening in polysyllabic words. Twenty-nine L1-German speakers produced minimal pairs differing in vowel tensity in both monosyllables /zakt, zaːkt/ and disyllables /zaktə, zaːktə/. The target words were produced in accented and deaccented contexts. The duration ratio between the vowel and the following /kt/ cluster was less for lax than tense vowels and less for disyllables than monosyllables. Under deaccentuation, there was an approximation of tense and lax vowels towards each other but no influence due to the mono- vs. disyllabic difference. On the other hand, Gaussian /a/ vs. /aː/ classifications of these data showed a lesser influence due to the syllable count in deaccented words. Compatibly, when the same speakers as listeners classified synthetic *sackt-sagt* and *sackte-sagte* continua, they were shown to compensate for the syllable count differences, but to a lesser extent in a deaccented context. Deaccentuation may therefore provide the conditions for sound change to take place by which /aː/ shortens in polysyllabic words; it may do so because the association between coarticulation and the source that gives rise to it is hidden to a greater extent than in accented contexts.

1 Introduction

The paper is concerned with the mechanisms by which coarticulatory variation due to so-called polysyllabic shortening can give rise to sound change, but taking into

Corresponding author: Jonathan Harrington, Institute for Phonetics and Speech Processing, Ludwig-Maximilians University, Munich, Germany, E-mail: jmh@phonetik.uni-muenchen.de
Felicitas Kleber, Institute for Phonetics and Speech Processing, Ludwig-Maximilians University, Munich, Germany, E-mail: kleber@phonetik.uni-muenchen.de
Ulrich Reubold, Institute for Phonetics and Speech Processing, Ludwig-Maximilians University, Munich, Germany, E-mail: reubold@phonetik.uni-muenchen.de
Jessica Siddins, Institute for Phonetics and Speech Processing, Ludwig-Maximilians University, Munich, Germany, E-mail: jessica@phonetik.uni-muenchen.de

account to a greater extent than before the roles of both phonological categorisation and prosodic variation. The starting point for the analysis is that the conditions for sound change can be met when the acoustic signal provides insufficient information for distinguishing between phonetic timing changes due to polysyllabic compression on the one hand and a phonological shortening of the long vowel on the other.

Following Ohala (1981, 1993), a sound change can occur when speakers and listeners parse coarticulatory relationships in different ways as a result of ambiguities in how speech production is associated with the acoustic speech signal. Thus, the presumed trigger for tonogenesis (Hombert et al. 1979) is that listeners parse the high fundamental frequency not with the source that gives rise to it, the increased tension in the vocal folds in voiceless stop production (Löfqvist et al. 1989), but instead with the following vowel, ultimately leading to a phonologization of this mis-parsed variation as tonal differences. Sound change is, according to Ohala's model, infrequent in relation to the ubiquitous phonetic variation because listeners undo the contextual modification to pho-nological categories in speech production through perceptual normalisation; for example, listeners factor out the anticipatory lip-rounding that is often produced in consonants preceding rounded vowels (Bell-Berti and Harris 1979; Perkell 1990) and so reverse or undo coarticulation in speech production (see Lindblom and Studdert-Kennedy 1967; Fujisaki and Kunisaki 1976; Mann and Repp 1980 for compatible perceptual evidence).

The further development to Ohala's model in this paper is to explore the extent to which sound change is conditioned by prosody and its association with the variation in speech production between hyper- and hypoarticulated speech. While there is extensive evidence that sound change depends on prosodic factors such as prominence and phrasing (e.g., Beckman et al. 1992; Cole and Hualde 2013), prosodic processing is rarely incorporated into models of sound change. One of the few exceptions to this is Lindblom et al. (1995), who suggest that sound change emerges from the interaction between an economy of effort in speech production combined with the need for the speaker to make the signal sufficiently clear in order that the listener's understanding is not compromised. Lindblom et al. (1995) reason that phonetic innovation can emerge from the high degree of variation in semantically predictable parts of the speech signal. However, listeners do not typically absorb these innovations into the lexicon because they reconstruct the meaning predominantly from top-down and not bottom-up processing. It is when top-down processing becomes disengaged in such hypoarticulation contexts that phonetically innovative forms can become more salient and may be added to the lexicon.

In both Lindblom et al. (1995) and Ohala's (1993) models, sound change can be brought about when listeners exceptionally decontextualize the speech signal

(from top-down processing and from coarticulation, respectively). The models are also similar in explaining sound change as arising as a consequence of how speech production is perceptually processed. The main differences are that Lindblom et al. (1995), like Bybee (2002), give greater emphasis to the mechanisms that give rise to undershoot and hypoarticulation, and that in contrast to Ohala (1993), sound change does not come about in Lindblom et al. (1995) model because of a listener error, but emerges instead as a consequence of how speech production is adapted to the needs of the listener. Harrington et al.'s (2013) draw upon the commonalities of both models in testing whether the perception of coarticulation might be compromised in prosodically weak constituents. They showed that listeners' compensation for coarticulation in relation to the size of anticipatory V_2-on-V_1 coarticulation in the production of German/pV_1CV_2l/non-words was less in unstressed than in stressed syllables. According to their model, sound change may be more likely in prosodically weak constituents because in such contexts listeners are less likely to attribute coarticulation to the source from which it originated than they are in prosodically strong contexts.

The present study extends that of Harrington et al. (2013) to sentence stress and to the sound change that can arise from the well-established finding of vowel shortening in polysyllabic words, which has been extensively documented for Germanic languages (e.g., Lindblom and Rapp 1973 for Swedish; Lehiste 1970; Klatt 1973; Port 1981 for English; Nooteboom 1972 for Dutch). The focus here is on the changes to segmental timing in German trochaic (*sackte*/*sagte*; 'sagged'/'said') words in relation to their corresponding monosyllables (*sackt*, *sagt*; 'sags'/'says').

Synchronic changes due to polysyllabic shortening can be related to two types of sound change. The first is the shortening of long vowels before two heterosyllabic consonants (Luick 1964; Bermúdez-Otero 1998), a process sometimes referred to as closed syllable shortening. This has resulted in changes such as late Old English *sōfte*, *cēpte* that were produced with long vowels in the stem to Middle English and present-day forms with an initial short vowel *soft*, *kept* (see also Hickey [1986] for an analysis based on presumed heterosyllabic geminates in late Old English *fōder* leading to present-day *fodder* with a short vowel in the stem via an intermediate diachronic stage in which the word was produced with a short vowel and geminate consonant). The second is trisyllabic shortening by which long vowels are shortened in antepenultimate syllables (Hogg 1992; Bermúdez-Otero 1998; Lahiri and Fikkert 1999) leading to vowel shortening in initial syllables between late Old English and present-day forms in words such as h_o_liday and s_o_utherner.

There are two main hypotheses to be tested in the present study. The first is that, consistent with other studies (e.g., Fowler and Thompson 2010), there will

be vowel shortening in the initial syllable of trochaic words and that listeners will compensate for this shortening. The second is that listeners will compensate less for the shortening when the words are in a prosodically weak, deaccented position. The first part of the paper is concerned with a production study designed to assess the influence of disyllables vs. monosyllables both on timing relationships in the initial syllable and on lax /a/ vs. tense /aː/ classification. The purpose of the second part of the production study was to relate the amount of information in the signal for classifying /a/ and /aː/ to listeners' responses on a similar task in the subsequent speech perception study.

For the standard variety of German analysed in this study, there is a phonological opposition between lax and tense vowels. The opposition between lax /a/ and tense /aː/ in German is characterised predominantly by length (Heike 1972) and to a lesser extent quality differences (Mooshammer and Geng 2008; Harrington et al. 2011). There is also some evidence to show that German lax vowels can be modelled as similar to tense vowels but truncated by an earlier timing of the closing consonant (Vennemann 1991; Hoole and Mooshammer 2002). Domain-final stops are neutralised in German (Vennemann 1972; Wiese 1996), possibly incompletely (Port and O'Dell 1985; Piroth and Janker 2004; Kleber et al. 2010); thus the vowel tensity distinction for the present materials is within monosyllabic /zakt, zaːkt/ (*sackt*/*sagt*) and disyllabic /zaktə, zaːktə/ (*sackte*/*sagte*) words.

2 Speech production

2.1 Method

The subjects included 29 speakers of Standard German who produced the target words *sackt*, *sagt*, *sackte*, *sagte* in the carrier sentence *Anna hatte _ verstanden* ('Anna had understood _'). For the accented context, the sentence was produced with a nuclear pitch accent (typically L + H⋆) on the first syllable of the target word. For the deaccented context, the L + H⋆ nuclear accent was on the first syllable of utterance-initial *Anna*; since all the following words were deaccented, the pitch was low and level following the L + H⋆ pitch peak.

Ten repetitions of the target sentences and an equal number of fillers were read in randomized order one at a time from a computer monitor at the subject's self-selected pace. Each sentence was preceded by a question to elicit the appropriate accentual pattern: for the accented context in which the target word was nuclear accented, the sentence was preceded by the question *WAS hatte Anna verstanden*? ('WHAT had Anna understood?'); for the context in which the target word was deaccented, the preceding sentence was *WER*

hatte_ verstanden? ('WHO had understood _?'). Three of the subjects were unable to complete the task: two of these because they consistently misread the sentences and one because of difficulty in producing the accented/deaccented contrast. Of the final 26 speakers, all were Standard German speakers with only slight regional colouring: 16 were from southern Germany (Bavarian and Baden-Württemberg) and the remaining 10 from central and northern regions of Germany. The group consisted of 17 females and 9 males aged 15 to 34 years (mean age 24 years). The final total number of analysed tokens was 10 (repetitions) × 2 (/a, aː/) × 2 (monosyllabic, disyllabic) × 2 (accented, deaccented) × 26 (speakers) = 2,080 target word tokens.

The entire corpus was automatically segmented into phoneme-sized units using the Munich automatic forced-alignment system (Schiel 2004). The boundaries of each segment were checked and manually corrected. The segments that formed part of the present analysis included the /a/ or/ aː/ vowel of the target word defined acoustically as the interval between the acoustic onset and offset of periodicity, and the consonant cluster /kt/ which marked the interval between the /a, aː/ offset and the release of the /t/-closure with no further sub-segmentation of the cluster into /k/ and /t/ components (Figure 1). The /t/-release was excluded in the analysis from the consonant cluster because of the general
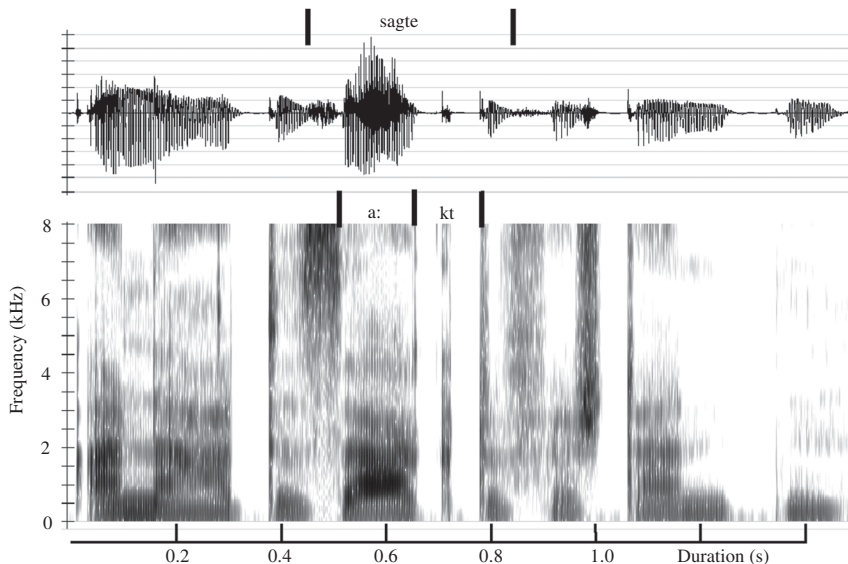
Figure 1: Waveform and spectrogram of an utterance containing nuclear accented *sagte* produced by one of the speakers from the corpus of this study, showing word and $V = $ /aː/and $C = $ /kt/segmentations.

difficulty of segmenting reliably its offset either from the /f/ of the following word *verstanden* or from the final weak vowel of *sackte*/*sagte* which was often partially devoiced. We also decided not to carry out a further sub-segmentation of /kt/ because it was not possible to determine for a large proportion of the tokens whether or not the /k/ had been released (as a result of which there were often unreliable acoustic landmarks for separating the two plosives). The dependent variables were (1) V, the log of the vowel duration; (2) C, the log of the duration of the /kt/ cluster as defined above; and (3) V/C, the log of the ratio between the vowel duration and the duration of the following /kt/.

The V/C values were classified as lax /a/ or tense /aː/ separately in the monosyllabic and disyllabic and separately in the accented and deaccented contexts following the procedure in Harrington et al. (2013). These (Gaussian) classifications were also carried out separately for each speaker (thus giving four separate classifications per speaker). The classification included a training stage in which normal distributions were fitted separately to these four contexts from which the posterior probabilities $p(a|(V/C))$ and $p(aː|(V/C))$ were derived that a given V/C value is a member of the category/a/or/aː/. The posterior probabilities were then used to derive/a-aː/classification functions (separately per speaker and for each of the four contexts). In order to do so, linear regression was used to estimate the slope, m, and intercept, k, in (1):

$$\log\left(\frac{p[aː|(\frac{V}{C})]}{p[a|(\frac{V}{C})]}\right) = m\left(\frac{V}{C}\right) + k \qquad [1]$$

from which the corresponding sigmoid classification function was derived in eq. [2]:

$$p\left[aː\middle|\left(\frac{V}{C}\right)\right] = \frac{e^{[m(\frac{V}{C})+k]}}{1 + e^{[m(\frac{V}{C})+k]}} \qquad [2]$$

using the relationship $p[aː|(V/C)] = 1 - p[a|(V/C)]$. The decision boundary was given by $-k/m$ which is the V/C value in eq. [2] for which $p[aː|(V/C)] = p[aː|(V/C)] = 0.5$. Four such decision boundaries were derived separately per speaker based on a total of 20 (10 /a/ and 10 /aː/) V/C values for each of the four contexts.

The influence of the syllable count on vowel tensity classifications was determined by calculating the difference between the monosyllabic and disyllabic classification functions on two measures: the 50% decision boundaries at which classifications between /a/ and /aː/ are equiprobable, and the area enclosed within the classification functions extending between these two decision boundaries in disyllabic and monosyllabic contexts (this latter measure corresponds to the shaded area between the fitted sigmoids shown in the lower

panel of Figure 5). The second of these measures also takes account of the slope of the classification function which is not incorporated just by subtracting the location of the decision boundaries.

As far as classification was concerned, the hypothesis to be tested was that /a-a:/ classifications would be influenced less by the syllable count (monosyllabic vs. disyllabic) in deaccented than in accented words. If so, then the difference between the decision boundaries and the area between the classification functions should be smaller in the deaccented than in the accented context.

2.2 Results

The first part of the results (Figures 2–4) is concerned with the effect of tensity, syllable count, and sentence stress on V, C, and V/C durations; the second part is concerned with classifications of V/C. Further descriptive statistics on the former are given in the Appendix.

As far as vowel duration is concerned, Figure 2 shows the expected clear distinction between lax /a/ and tense /a:/ in both accented and deaccented words. In the deaccented context, there was a greater overlap between /a, a:/
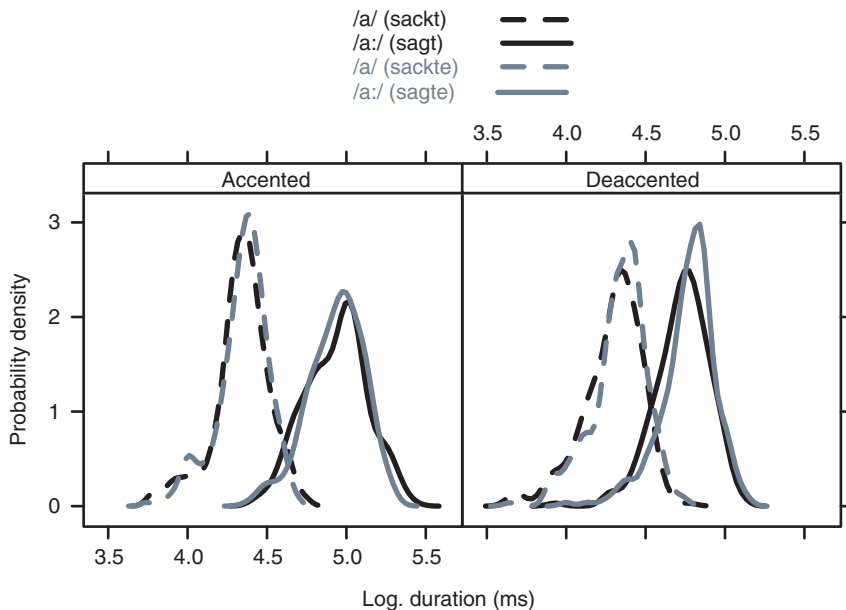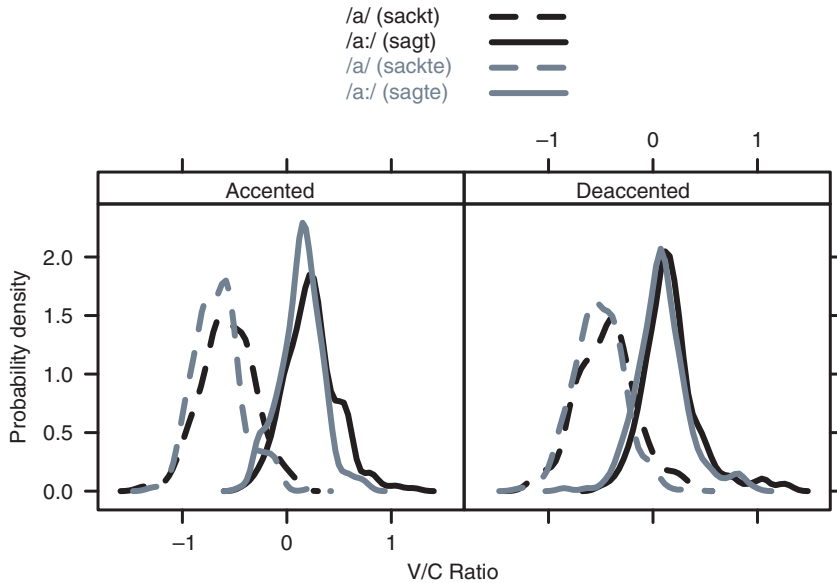

Figure 2: Probability density plots of the log vowel duration for all samples of the corpus in monosyllabic (black) and disyllabic (grey) words in accented (left) and deaccented (right) contexts for tense (solid) and lax (dashed) categories.
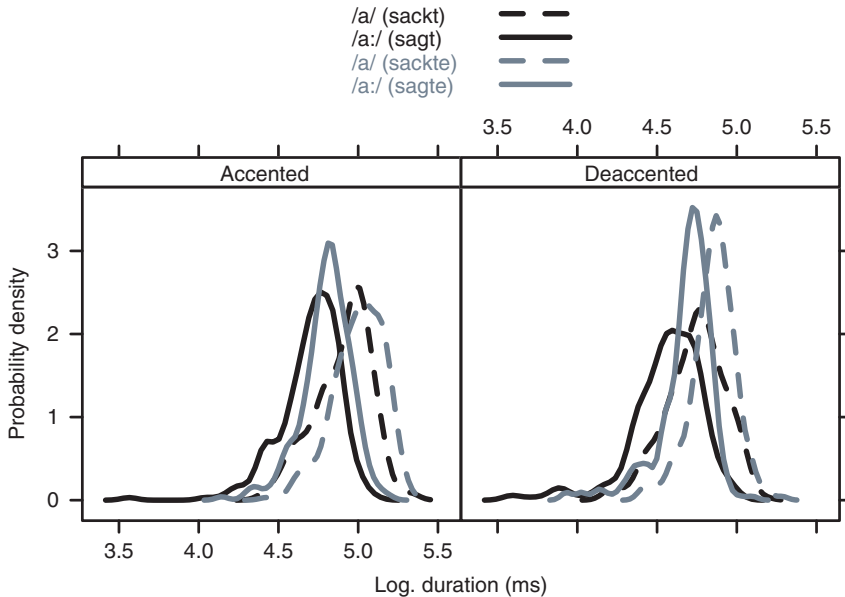
Figure 3: Probability density plots of the log C = /kt/duration for all samples of the corpus in monosyllabic (black) and disyllabic (grey) words in accented (left) and deaccented (right) contexts for tense (solid) and lax (dashed) categories.

that was primarily caused by a shift of the /aː/ distribution towards /a/. This is in line with findings from Mooshammer and Fuchs (2002) and Mooshammer and Geng (2008) that prosodic weakening shortens tense but not lax vowels. There was (to our surprise) no evidence that vowel duration was shorter in the disyllabic vs. monosyllabic contexts. The results of a repeated measures ANOVA with dependent variable V (log vowel duration) and within-subject factors Tensity (two levels /a, aː/), Syllable Count (monosyllabic, disyllabic), and Stress (accented, deaccented) showed a predictable main effect on V of Tensity ($F[1, 25]$ = 616.3, $p < 0.001$), and of Stress ($F[1,25] = 44.5$, $p < 0.001$) but no significant effect of Syllable Count. There was also a significant interaction between Stress and Tensity ($F[1, 25] = 92.6$, $p < 0.001$). Post-hoc Bonferroni corrected t-tests showed, consistent with the evidence in Figure 2, a significant difference between accented and deaccented vowel duration for tense /aː/ ($t[24] = 8.3$, $p_{adj} < 0.001$) but not for lax /a/. There was also a significant interaction between Stress and Syllable Count ($F[1,25] = 9.0$, $p < 0.01$). Post-hoc tests showed, consistent with Figure 2, that vowel duration in disyllables was actually significantly greater than in monosyllables but only in deaccented ($t[25] = 3.1$, $p_{adj} < 0.05$) and not in accented words.

Figure 4: Probability density plots of the V/C ratio for all samples of the corpus in monosyllabic (black) and disyllabic (grey) words in accented (left) and deaccented (right) contexts for tense (solid) and lax (dashed) categories.

The results for C (log duration of the /kt/ cluster as defined in 2.1 above) in Figure 3 showed a greater cluster duration for lax /a/ vs. tense /a:/ and also for disyllables vs. monosyllables. The results of a repeated measures ANOVA with cluster duration as the dependent variable and with the same independent factors as above showed main effects for Tensity ($F[1,25] = 178.6$, $p < 0.001$), Stress ($F[1,25] = 109.8$, $p < 0.001$), and Syllable Count ($F[1,25] = 26.4$, $p < 0.001$) and no interaction between these factors. Thus these results show that cluster duration was greater in lax vs. tense vowels, in disyllables vs. monosyllables, and in accented vs. deaccented words.

The third parameter to be considered was the ratio of the vowel to the /kt/ cluster duration: this measure (V/C) should magnify the difference between lax and tense vowels given that, as shown above, lax vowels had both a shorter vowel duration and longer cluster duration compared with their tense counterparts. Figure 4 shows smaller V/C ratios for lax vs. tense vowels and to some extent also for disyllables vs. monosyllables. The results of a repeated measures ANOVA with V/C as the dependent variable and with the same independent factors as above showed a significant main effect for Tensity ($F[1,25] = 759.5$, $p < 0.001$), and for Syllable Count ($F[1,25] = 17.1$, $p < 0.001$), but no main effect of

Stress. There was a significant interaction between Tensity and Stress ($F[1,25] = 55.7$, $p < 0.001$). Post-hoc tests showed a larger V/C ratio for deaccented vs. accented words, but only in lax vowels ($t[25] = 5.7$, $p < 0.001$).

Finally, Figure 5 shows the same data as in Figure 4 together with the corresponding classification functions calculated by fitting sigmoids to the data in the manner described in Section 2.1. Three observations can be made about these averaged data. First, the decision boundaries between /a/ and /aː/ were located as expected roughly halfway between their corresponding V/C distributions. Secondly, and again compatibly with the acoustic data, the /a-aː/ classification function in the disyllabic context was left-shifted towards lower V/C values relative to that in the monosyllabic context (i.e., the vertical grey dashed lines are to the left of the vertical black dashed lines in both accented and deaccented contexts). Thirdly, the classification functions for the monosyllabic and disyllabic contexts were closer together for deaccented than for accented words. This may come about because, as the top right panel of Figure 5 shows, the distributions of lax and tense vowels on V/C were closer together. Importantly, the closer
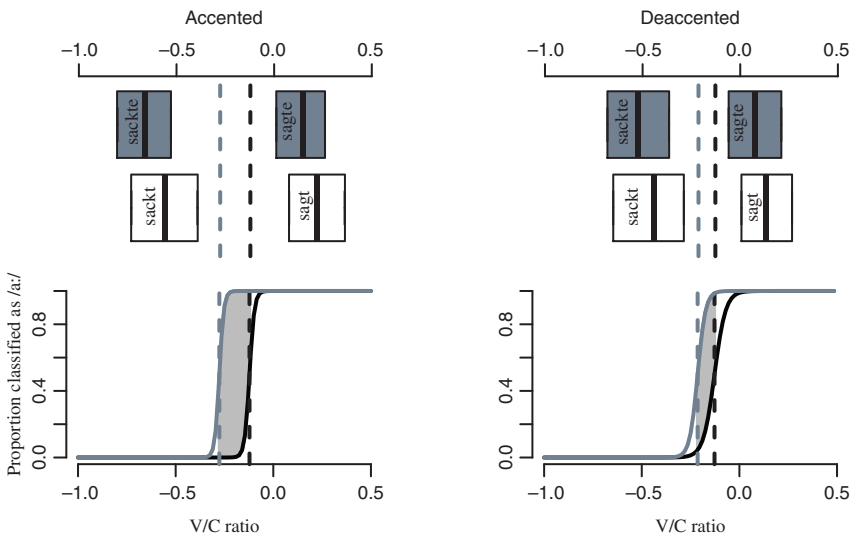
Figure 5: Boxplots of the data in Figure 4 showing the interquartile range of the V/C ratio (above) for disyllables (grey boxes) and monosyllables (white boxes) in accented (left) and deaccented (right) contexts. The corresponding /a/ vs. /aː/ classification functions averaged across all speakers are shown below for *sackte* vs. *sagte* (grey) and *sackt* vs. *sagt* (black) with 50% decision boundaries superimposed as the vertical dashed lines. The shaded region is the area between the disyllabic and monosyllabic sigmoids within their respective decision boundaries.

proximity of the classification functions in deaccented words did not result from the weaker influence of the syllable count. As the top row of Figure 5 shows, the difference between the monosyllabic (white) and disyllabic (grey) distributions was about the same in accented and deaccented words, which is compatible with the results in Figure 4 showing that stress had no influence on the V/C ratio and did not interact with the syllable count. In deaccented words, the closer proximity of the monosyllabic and disyllabic classification functions was instead an indirect consequence of the smaller separation between lax and tense vowels. The shallower slopes in the classification functions for deaccented words in Figure 5 are also consistent with the greater overlap and therefore greater classification uncertainty in the deaccented compared with the accented context.

The above trends were further quantified by analysing the speaker-specific decision boundaries, areas, and slopes. As far as the decision boundaries are concerned, there were two trends evident in Figure 6: first, the decision boundaries were lower at lower V/C ratios in the disyllabic than in the monosyllabic context; secondly, the separation between monosyllabic and disyllabic decision boundaries was slightly greater in accented than in deaccented words. A repeated measures ANOVA with the Decision Boundary as the dependent variable and with independent factors Syllable Count (monosyllabic/disyllabic) and Stress (accented/deaccented) showed a main effect for Syllable Count ($F[1,25] = 15.5$, $p < 0.001$) but no effect for Stress and no interaction between these factors. The significant effect for Syllable Count supports what is evident in Figure 5: the decision boundaries were shifted towards lower values in disyllables than in
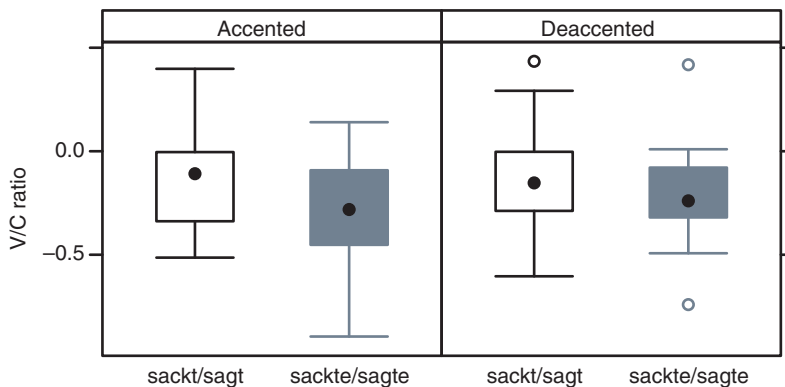
Figure 6: Distribution of the 50% decision boundaries resulting from classifying the V/C ratio as /a/ or /aː/ in accented monosyllables (white boxes) and disyllables (grey boxes) in accented (left) and deaccented (right) contexts. The extent of the box is the interquartile range. Each of the four distributions consists of one decision boundary per speaker.

monosyllables. However, the lack of a significant interaction between the independent factors does not support the averaged data in Figure 5 showing that the decision boundaries were closer together in the deaccented context.

Figure 7 shows that the areas between the mono- and disyllabic sigmoids (corresponding to the grey shaded area in Figure 5) were larger in the accented than in the deaccented context. Consistent with the observations in Figures 5 and 7, a paired (one pair per speaker) signed Wilcoxon rank test showed larger areas in an accented compared with a deaccented context ($V = 266$, $p < 0.05$).
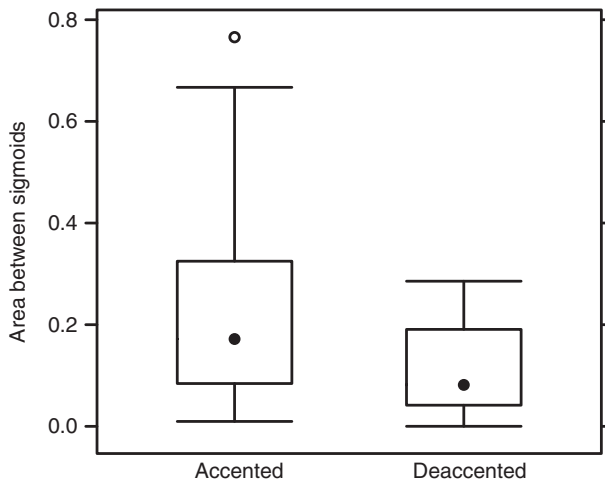
Figure 7: Distribution resulting from calculating the area between the sigmoids in the mono- and disyllabic contexts (the shaded area in Figure 5) in accented (left) and deaccented (right) words separately per speaker. The extent of the box is the interquartile range. Each of the two distributions consists of one area per speaker.

Figure 8 shows that the slopes of the classification functions were steeper in accented than in deaccented words. These trends were confirmed by a repeated measures ANOVA with Slope as the dependent variable which showed a main effect for Stress ($F[1, 25] = 22.8$, $p < 0.001$), a non-significant effect for Syllable Count, and no interaction between these independent factors. Thus the likely reason for the discrepancy between the decision boundaries (showing a non-significant effect for stress in Figure 6) and area differences (in which the effect for stress was significant in Figure 7) is that the area between the sigmoids is determined not just by decision boundaries but also by slopes: that is, the shallower slopes are likely to have been a contributory factor in the smaller area between the mono- and disyllabic sigmoids in the deaccented context.
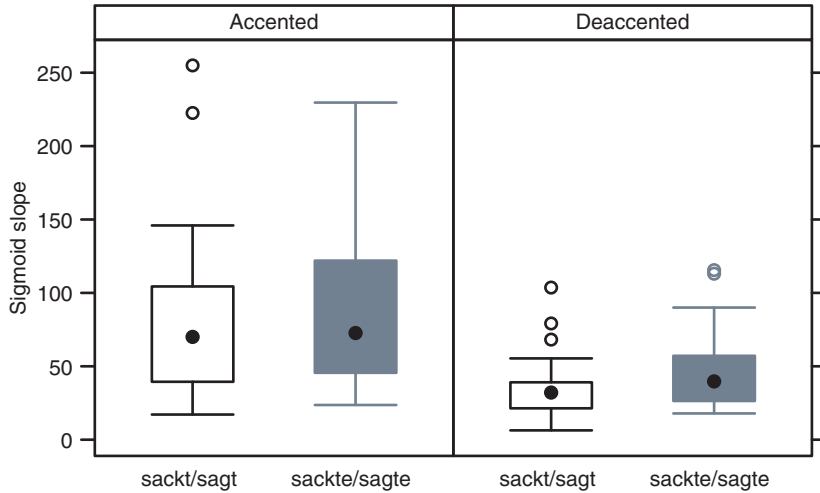
Figure 8: Distribution of the sigmoid slopes resulting from classifying V/C ratio as /a/ or /aː/ in monosyllables (white boxes) and disyllables (grey boxes) and in accented (left) and deaccented (right) words separately per speaker. The extent of the box is the interquartile range. Each of the four distributions consists of one slope per speaker.

2.3 Discussion

The first part of the analysis was concerned with the way in which the syllable count influenced vowel and cluster duration in accented and deaccented words. Syllable count was found (somewhat surprisingly) not to affect vowel duration (V). On the other hand, the duration of the post-vocalic cluster (C) was greater; as a result, the V/C ratio decreased in a disyllabic vs. monosyllabic context. Stress predictably caused a lengthening of both the vowel and of the following cluster. However, it did not affect the V/C ratio. Consequently, the syllable count's influence was not different in accented and deaccented words, at least on the V/C ratio parameter.

The second part of the analysis was concerned with the influence of the syllable count on tensity (/a/ vs. /aː/) classifications. The results of this analysis showed that the classification functions were shifted towards lower values on the V/C parameter (resulting in a greater likelihood of /aː/ classifications) in a disyllabic compared with a monosyllabic context. However, this shift from a disyllabic to a monosyllabic context was not as great for deaccented compared with accented words; the likely cause of this diminished shift in the deaccented context (in which the vertical dashed lines are closer together for the deaccented than in the accented context in Figure 5) was that the distributions of lax and

tense vowels on V/C overlapped to a greater extent in the deaccented than in the accented context.

Thus the overall conclusion from all of these findings is that, whereas the syllable count's influence on the V/C parameter was much the same in an accented and deaccented context, its influence on /a/ vs. /aː/ categorisations was greater in accented than in deaccented words.

From the point of view of speech perception, the implication of these results is that listeners may find it more difficult to factor out the influence of the syllable count from /a-aː/ classifications in a deaccented context. Suppose, for example, that the sigmoids in Figure 5 had been psychometric functions resulting from forced-choice classification of /a/ vs. /aː/. In this case, the evidence in Figure 5 would show that listeners compensate for coarticulation because they would be shifting their categorisations towards lower values in the disyllabic context, commensurate with the lowering effect of disyllabic words on V/C seen in the top panel of the same figure. But the same data would also show that listeners compensate less for the coarticulatory effect of syllable count in deaccented words; this is because the monosyllabic and disyllabic classification functions are evidently closer together in deaccented words, even though the size of the monosyllabic vs. disyllabic effect on V/C in speech production (Figure 5, top panel) was much the same in accented vs. deaccented words.

The question of whether listeners do compensate less for the effects of syllable count in a deaccented context is the main concern of the next section.

3 Speech perception

3.1 Method

Four continua were created by embedding the same 11-step continuum between tense /aː/ and lax /a/ in monosyllabic /z_kt/ (*sagt*/*sackt*) and disyllabic /z_ktə/ (*sagte*/*sackte*) contexts in the carrier *Anna hatte _ verstanden* (the same sentence that had formed part of the materials for the speech production experiment). Just as in speech production, the target word was either nuclear accented or deaccented, with the nuclear accent occurring on the initial word in the second case.

In order to create these continua, a trained phonetician and female speaker of Standard German with slight South German regional characteristics produced 10 repetitions each of the target sentences in the two sentence-stress contexts. The mean durations of /z/ and /kt/ of the target word were calculated separately in the accented and deaccented context; then, two productions of /zVkt/ were

selected, one from each of the speaker's accented and deaccented productions which were closest in duration to these means. The vowel was spliced out leaving an accented /z_kt/ and a deaccented /z_kt/. In both cases, the /k/ was produced with a release.

An /aː/ vowel from one of the speaker's accented productions of *sagte* which had F1 values closest to the F1 median of this speaker's /a/ and /aː/ tokens was manipulated in duration to create 11 equidistant steps using Praat's (Boersma and Weenink 2012) implementation of the PSOLA algorithm (Moulines and Charpentier 1990). The 11 steps ranged from 40 ms to 112 ms; these selected durations were based on the model speaker's range of target vowel durations in two small pilot studies. The vowel duration and quality were identical in both accented and deaccented contexts, but the deaccented vowels were subsequently manipulated to contain a slightly falling F_O contour from 215 to 195 Hz and an intensity of 7 dB lower than that of the accented vowels. These vowel continua were then spliced into the aforementioned two (accented, deaccented) /z_kt/ contexts. The closure duration of the /kt/ cluster between the acoustic vowel offset and onset of the /t/ release was 81 ms for all accented stimuli and 62 ms for all deaccented stimuli. Finally, the *sackte-sagte* continua were derived by post-pending one of the speaker's /ə/ tokens to accented /zVkt/ and to deaccented /zVkt/ (where V denotes the 11-step continuum).

The four (monosyllabic/disyllabic × accented/deaccented) continua created in this way were spliced into the carrier sentence *Anna hatte _ verstanden*, with one carrier taken from an accented utterance and the other from a deaccented utterance. As such, the final stimuli contained natural intonation contours without the need for pitch and intensity resynthesis (see Figure 9).

The 44 synthetic utterances were each repeated 10 times, and presented in the same randomised order over headphones in a quiet room in a two-alternative forced-choice identification test to 32 subjects, 29 of whom were the same subjects as in the speech production experiment. The listeners, who were paid for their participation, classified each word by clicking on one of four possible orthographic forms (*sackt, sagt, sackte, sagte*) displayed on a computer monitor. The two alternatives were presented counter-balanced so that half of the time the orthographic representation of the tense vowel was on the right-hand side of the screen and half of the time on the left. Participants were requested to respond as quickly as possible. Repeated listening of the same stimulus was not allowed, but listeners could change their choice before clicking the "OK" button to confirm their response. There were 11 (steps) × 2 (mono/disyllabic) × 2 (accented/deaccented) × 10 (repetitions) × 32 (listeners) = 14,080 responses in total.

Classification (sigmoid) functions explaining the relationship between the response (/a/ or /aː/) and the V/C ratio were obtained using binary logistic
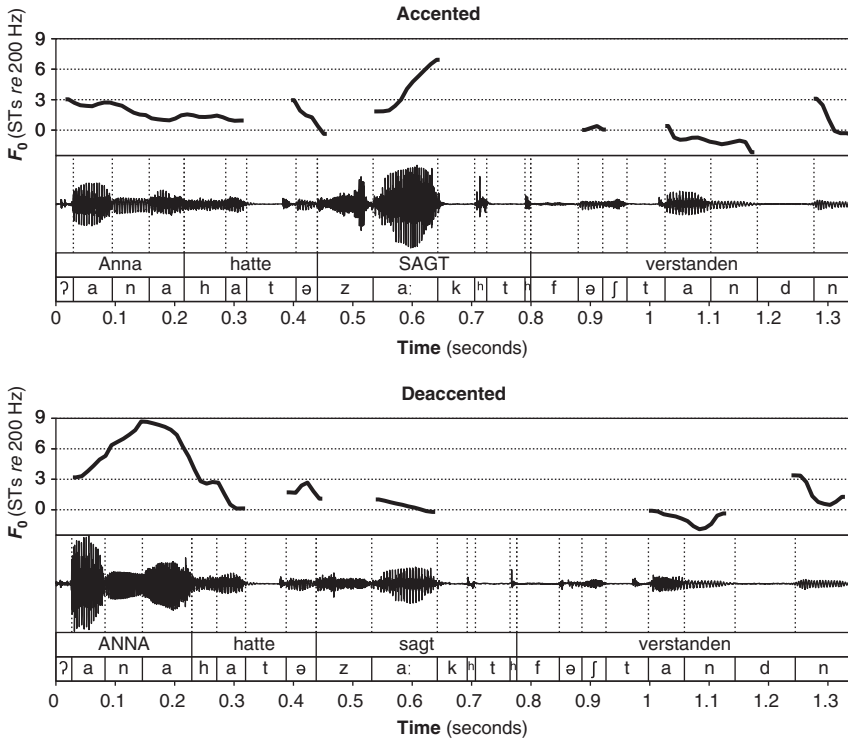
Figure 9: Fundamental frequency contour, waveform, and segmentation of two of the synthetic utterances in the accented and deaccented contexts. The target word is shown for *sagt* at the /aː/ endpoint of the *sackt-sagt* continuum.

regression separately per listener in each of the four contexts (monosyllabic/ disyllabic × accented/deaccented) using the relationship:

$$p_{aː} = \frac{e^{[m\left(\frac{V}{C}\right)+k]}}{[1 + e^{[m\left(\frac{V}{C}\right)+k]}]} \tag{3}$$

in which $p_{aː}$ was the proportion of /aː/ responses ($p_{aː} + p_a = 1$), V/C the values on the V/C ratio in the synthetic continuum, and m and k respectively the listener-specific slope and intercept that were fitted using logistic regression. Decision boundaries at 50% cross-over points were obtained, one for each of the four contexts (i.e., four per listener) from $-k/m$ in eq. [3]. In addition, and analogously to the procedure for classifying the production data, the areas between the classification functions in the disyllabic and monosyllabic contexts were calculated (two areas per listener, one for the accented, one for the deaccented context).

The hypothesis to be tested was that the decision boundaries would be closer together and that the area between the classification functions would be smaller in the deaccented than in the accented context.

3.2 Results

The aggregated results in Figure 10 suggest that listeners' decisions were influenced both by syllable count and by stress. Consistent with the production data (Figure 5), listeners' decision boundaries were, on average, shifted towards lower values of the V/C ratio in disyllabic compared with monosyllabic words. The much greater positive V/C values in the deaccented context are, however, not at all consistent with the production data which showed no main effect of stress, neither in duration (Figure 4 and Figure 5, top panel) nor classification (Figure 5, lower panel). The rightwards shift in the deaccented context observed in these perception data is likely to have come about because the range of V/C
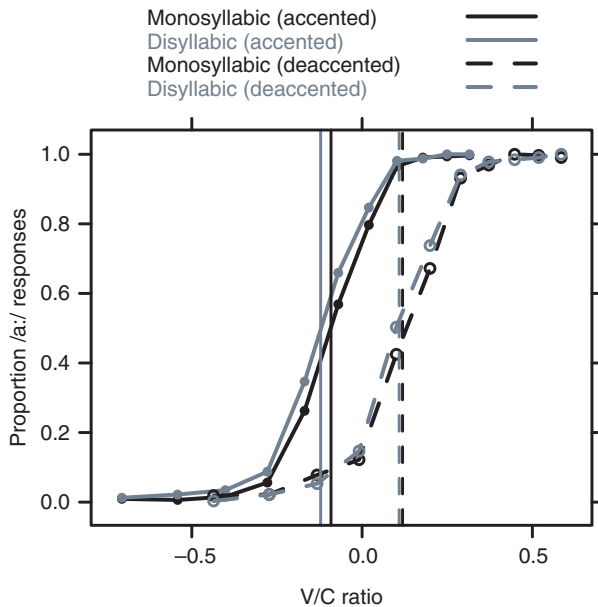
Figure 10: The points are proportional /a:/ responses averaged across listeners responding to the 11-step monosyllabic (black) and disyllabic (grey) continua in the accented (solid) and deaccented (dashed) contexts. The decision boundaries averaged across listeners in these four contexts are shown as vertical lines.

values of the 11-step continuum on which responses were obtained was itself right-shifted for deaccented (between −0.44 and 0.58) compared with accented (between −0.70 to 0.32) by an amount (approximately 0.25) which corresponds to the size of the rightward displacement for deaccented observed in Figure 10. That is, the unexpected rightwards displacement of the responses in the deaccented relative to the accented context is most likely an artefact of the different V/C values for the two stress contexts on which responses were obtained.

As far as the main hypothesis is concerned, there is some marginal evidence from Figure 10 that the difference between responses in a disyllabic and monosyllabic context was greater in the accented vs. deaccented context: this is shown both by the slightly greater separation between the decision boundaries derived from logistic regression and by the generally greater leftward displacement of the disyllabic responses in the accented context. The distribution of the listener-specific decision boundaries in Figure 11 shows marginally lower decision boundaries for disyllables vs. monosyllables in the accented context. A repeated measures ANOVA with the 50% decision boundary (the data in Figure 11) as the dependent variable and with independent factors Syllable Count (mono/disyllabic) and Stress (accented/deaccented) showed significant effects for Syllable Count ($F[1,31] = 12.4$, $p < 0.01$) and for Stress ($F[1,31] = 483.2$, $p < 0.001$) and a significant interaction between these factors ($F[1,31] = 5.9$, $p < 0.05$). The significant effect of Syllable Count shows that listeners compensated
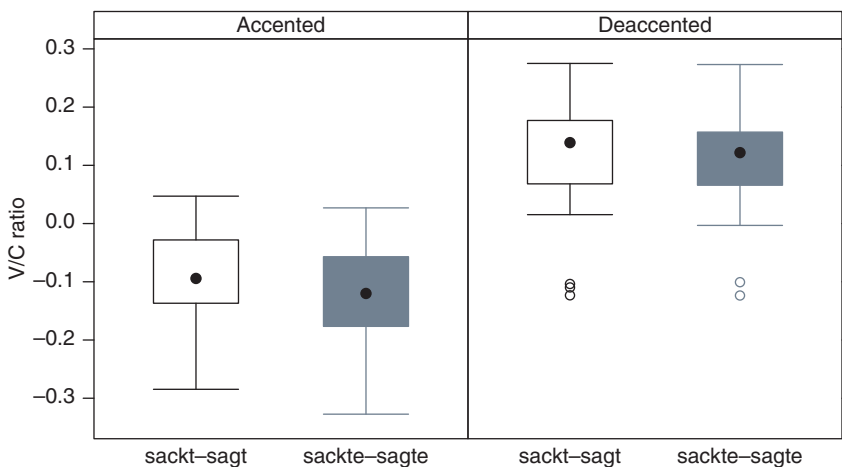

Figure 11: Distribution of the 50% decision boundaries resulting from perceptual classification of the continua in accented monosyllables (white boxes) and disyllables (grey boxes) in accented (left) and deaccented (right) contexts. The extent of the box is the interquartile range. Each of the four distributions consists of one decision boundary per speaker (compare with Figure 6).

for the coarticulatory effects of V/C reduction in a disyllabic context. The significant effect of stress is likely to be an artefact of the different range of V/C ratios on which responses were obtained in the two stress contexts, as discussed above. The significant interaction is consistent with the evidence in Figure 10 showing a greater separation between disyllabic and monosyllabic responses in the accented than in the deaccented context. Compatibly, post-hoc Bonferroni corrected *t*-tests showed a significant difference in the location of the decision boundaries between monosyllabic and disyllabic contexts in the accented ($t = 4.4$, $p < 0.001$) but not in the deaccented context.

The other dependent variable used to test the hypothesis was the area between the monosyllabic and disyllabic classification functions (between the decision boundaries) which was predicted to be greater for the accented than for the deaccented context. Although, as Figure 12 shows, there is a very general tendency for this to be so, the results of a paired Wilcoxon signed rank test with Area as the dependent variable (one pair per listener) and with Stress as the independent factor showed no significant differences between accented and deaccented words on this parameter.

Based on the results of the speech production data, the prediction about slopes in perception was that these should be flatter for the deaccented than for the accented context (Figure 8). The statistical quantification of slope for the perception data was, however, problematic because of data from six listeners
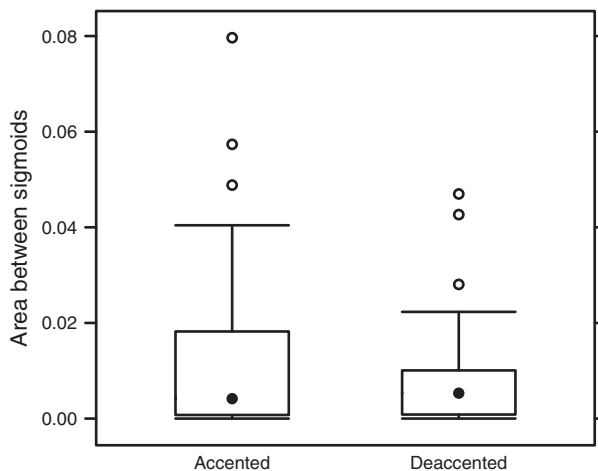
Figure 12: Distribution resulting from calculating the area between the sigmoids in the mono- and disyllabic contexts (corresponding to the shaded area in Figure 5) in accented (left) and deaccented (right) words. The extent of the box is the interquartile range. Each of the two distributions consists of one area per listener (compare with Figure 7).

who had almost infinite slopes, i.e., who showed no ambiguity in their responses at the 50% decision boundary. After removing the data from these listeners, a similar tendency (although not as marked as for speech production) emerged for steeper slopes in the accented than in the deaccented context (Figure 13). Consistent with this trend in Figure 13, the results of a repeated-measures ANOVA with Slope as the dependent variable and the same independent factors as above showed steeper Slopes for accented vs. deaccented ($F[1,25]$ = 8.1, $p < 0.01$), no effect of Syllable Count, and no interaction between these factors.[1] Thus, there is some evidence to show that, at least for those 26/32 listeners who showed some kind of ambiguity in the vicinity of the decision boundary, the slope was steeper for accented than deaccented words.
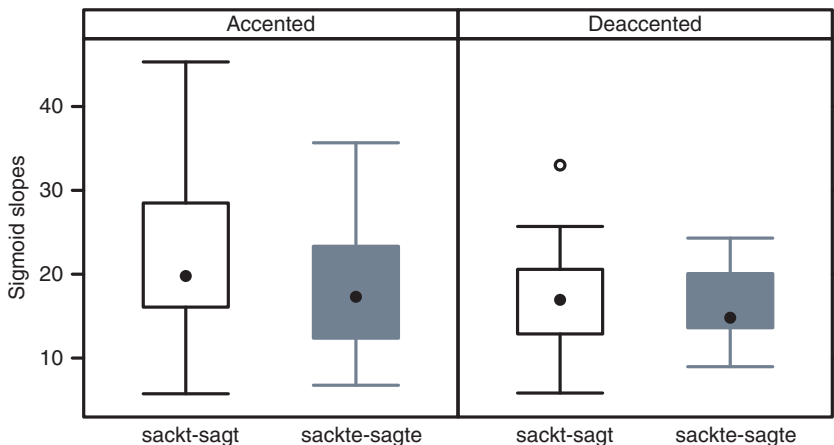
Figure 13: Distribution of the sigmoid slopes resulting from perceptual classifications in monosyllables (white boxes) and disyllables (grey boxes) in accented (left) and deaccented (right) contexts for 26/32 listeners. The extent of the box is the interquartile range. Each of the four distributions consists of one slope per listener (compare with Figure 8).

3.3 Discussion

There are two main findings from this perception experiment. The first is that listeners adjusted their responses due to the syllable count such that their /a/ vs. /a:/ classifications were shifted towards lower V/C values in disyllabic vs.

---

1 When the repeated measures ANOVA is re-run on these data with only 26 listeners to test for an effect of decision-boundaries, then the interaction between the independent factors of Syllable Count and Stress is still significant ($F[1,25]$ = 4.8, $p < 0.05$).

monosyllabic words. This result is compatible with their speech production data, which showed both that V/C ratios were smaller (Figure 4) and that the binary /a/ vs. /aː/ classification of the production data was shifted towards lower V/C ratios (Figure 5). Independently of this finding, the perception experiment also demonstrates that listeners must have responded to the V/C ratio (or at least to some combination of V and C) in making their responses. If they had been responding just to synthetic vowel duration, then their classifications should have been at about the same point on the V/C continuum in the accented vs. deaccented context; yet the results show that their responses in the deaccented context were shifted towards higher V/C values by an amount corresponding to the difference in the synthetic /kt/ cluster duration between the two stress contexts. Listeners must therefore have taken the cluster duration into account in making /a/ vs. /aː/ categorisations.

The second main finding is that the influence of the syllable count on classifications was less in the deaccented than in the accented context: that is, consistent with the main hypothesis that was tested, there was some evidence to suggest that listeners did not compensate as much for the phonetic influence of a disyllabic vs. monosyllabic context in deaccented words. The speech perception and production classifications were generally in the same direction, although not always significantly so: thus in deaccented compared with accented words, the decision boundaries were closer together (significantly for perception), the areas between the classification functions were smaller (significantly for production), and the slopes were flatter (significantly in production and perception, but only after removing 6 listeners from the latter due to near-infinite slopes).

4 General discussion

The results from speech production showed that the syllable count – whether the word was monosyllabic or disyllabic – had an influence on V/C ratios, where V is the vowel duration and C the duration of the following cluster. The size of this phonetic effect was no different in accented and deaccented words. When the same speech production data were classified for tensity as /a/ or /aː/ on the same V/C parameter, then categorisations were affected by the syllable count such that, compatibly with this phonetic effect, classifications were shifted towards lower V/C values in disyllabic words: that is, the probability of classifying the signal as /aː/ was slightly greater in disyllabic words. However, the extent of the syllable count's influence on classification was

less in the deaccented context, even though the size of the V/C lowering in disyllables was no different in deaccented than in accented words. Thus, although deaccentuation did not diminish the influence of the syllable count on speech production, it did weaken its influence on /a/ vs. /a:/ categorisation. The weakening in categorization was an indirect consequence of the closer proximity (less differentiation) of lax vs. tense vowels and the greater speech production variability in deaccented speech. From another point of view: the contextual effect of the syllable count was not obliterated in speech production but was instead hidden to a greater extent from tensity categorisations due to the greater /a/ vs. /a:/ overlap and variability in deaccented words. That is, when speech production was categorised for tensity, there was less information available in a deaccented context about the confounding influence of the syllable count.

When the same speakers classified synthetic stimuli that varied only in vowel duration (and therefore also in the V/C ratio), their responses were related to the speech production classifications in two ways. First, their responses were affected by the syllable count: just as in classifications of the speech production data, they were more likely to classify a given V/C value as /a:/ in a disyllabic context. Secondly, and also entirely compatibly with the speech production classifications, their responses were affected by the syllable count to a lesser degree in a deaccented context, even though the (synthetic) changes to the V/C ratio were the same in mono- and disyllables. The speech perception results showed that listeners parsed the V/C ratio into at least two components: the component that is responsible for differences in vowel tensity and the component that is due to the influence of the syllable count. This perceptual factoring (Fowler and Smith 1986; Fowler 2005) of the signal is manifested as a shift in the classification function in the disyllabic relative to the monosyllabic context. But listeners were less successful at factoring the signal perceptually in a deaccented context, even though acoustically the same information was there for them to do so, and even though in speech production the shift in the V/C ratio due to the syllable count was significant and largely unaffected by whether the words were accented or not. The question as to why listeners should have compensated less for coarticulation in deaccented speech is more difficult to answer. But one possibility is suggested by the results from classifying the speech production data in which the greater proximity of the /a, a:/ distributions in deaccented words also pushed the decision boundaries for monosyllabic and disyllabic words closer together. Perhaps for this reason, and also because there is less certainty in /a, a:/ decisions in deaccented words, as shown by the shallower classification

slopes, the information about how the signal is factored into contributions from tensity on the one hand and syllable count on the other is noisier and therefore not as learnable as in accented speech.

Deaccented speech may be a source of sound change precisely because the effect of context on categorisation – i.e., what has often been termed compensation for coarticulation – is weakened. If sentence stress has no effect on the size of the phonetic effect in distinguishing *sackt* from *sackte* or *sagt* from *sagte* but diminishes compensation for coarticulation, then there is likely to be variation in V/C that is not parsed with the syllable count. The conditions for sound change to occur by which /aː/ changes to /a/ would then be met if the decrease in the V/C ratio towards/a/caused by disyllables were instead parsed with vowel tensity. This type of parsing error is more likely in any condition such as deaccentuation in which the parsing or factoring of the signal is itself compromised.

The further issue raised by the present study is how sound change interacts with prosodic prominence. On the one hand, the results are consistent with the finding of an association between sound change and prosodic weakening (e.g., Beckman et al. 1992), but on the other hand, they do not seem to be compatible with the evidence that, as recently discussed in Cole and Hualde (2013), there are numerous sound changes such as umlaut and diphthongization that are often confined to stressed syllables. There is, however, no contradiction between our results and the occurrence of sound change in lexically stressed syllables because we are not proposing an association between word stress and sound change; our model suggests instead that it is a hypoarticulation context brought about by semantic redundancy (of which deaccenting is an example) that can provide the conditions for sound change to occur. These conditions are met in lexically stressed syllables when they are hypoarticulated as they have been shown to be in deaccented words (de Jong 1995; Harrington et al. 2000) or perhaps more generally in semantically redundant contexts (Lindblom 1988, Lindblom 1998).

From another point of view, stressed syllables that are accented at the level of the utterance (i.e., marked for a pitch-accent in Germanic languages) are processed more rapidly by listeners, because prominent syllables direct the listener's attention towards semantically salient parts of the utterance (Cutler 1976; Mehta and Cutler 1988) and/or because metrical expectations cause listeners to focus their attention on metrically prominent syllables (Zheng and Pierrehumbert 2010). Analogously, speech signals that are low in prominence are processed more slowly, perhaps because listeners can deploy semantic context to a greater extent in such contexts, which are often hypoarticulated

(Lindblom 1988). Our suggestion is that this processing advantage in prominent syllables also extends to the listener's compensation for coarticulation. That is, in hyperarticulated speech in which the listener's attention is closely focussed on the signal content (Lindblom et al. 1995), listeners also normalise or compensate for coarticulation. But they may do so to a lesser extent in hypoarticulated and weakly prominent speech in which speech processing is less rapid and in which, following Lindblom et al.'s (1995) model, they shift their attention away from the detailed composition of the signal as they engage their 'what-mode' of perception (i.e., top-down processing) to a greater extent.

The more general conclusion is that it is hypoarticulation which is more likely to cause an error in parsing coarticulation. Although this proposed generalisation to hypoarticulation is not directly inferable from the results of the present study, hypoarticulation may be the common cause linking the findings of diminished perceptual compensation for coarticulation in deaccentuation (this paper) and in rhythmically weak syllables (Harrington et al. 2013). Both these studies suggest that the conditions for sound change to take place are likely to be met when, first, category boundaries become more blurred, either through greater variation and/or because they are shifted closer together; and, secondly, when the magnitude of the coarticulatory effect that a source gives rise to is maintained or increased. In Harrington et al. (2013), this argument was applied to anticipatory V_1CV_2 coarticulation in which the high vowels in V_1 position were found to shift closer together when V_1 was prosodically weak, but the coarticulatory influence of V_2 on V_1 was maintained.

Our assumption has been that the shortening effect from *sackt* to *sackte* or *sagt* to *sagte* is attributable to polysyllabic shortening of the kind in which there is an association between a compression to segment duration and the number of syllables in the word (Klatt 1973; Lindblom and Rapp 1973), leading to a progressive stressed vowel shortening in, e.g., triplets such as *speed*, *speedy*, and *speedily* (Lehiste 1972; Port 1981; see Turk and Shattuck-Hufnagel 2000; White and Turk 2010 for more recent studies). However, there are at least two reasons why polysyllabic shortening may not be the mechanism responsible for the observed duration differences in the present data. First, in contrast to those studies, we found no evidence of a shortening of the lexically stressed vowel in disyllables compared with monosyllables, which, for accented and deaccented words in our study, were instead associated with differences in the ratio of the lexically stressed vowel to that of the following consonant cluster. Second, the greater duration of the /kt/ cluster in disyllables than in their monosyllabic counterparts may instead be due to the

presence of a syllable boundary within the /kt/ cluster in *sack.te* and *sag.te* (as a result of which, according to some accounts, an underlying voiced /g/ in the stem of *sag* becomes voiceless by a rule of syllable or word-final devoicing – see, e.g., Röttger et al. [2014] for a recent review and analysis). Whether the mechanism is polysyllabic shortening or presence vs. absence of a syllable boundary within the cluster or some form of both, the conclusion is the same: classifications of lax /a/ vs. tense /aː/ are influenced by the syllable count (monosyllables vs. disyllables) but less so in deaccented compared with accented words.

The prediction from these results is that the conditions for sound change to take place are likely to be met when, due to a hypoarticulated speaking style, coarticulation is maintained (or increased) but categorisation strength is weakened. For example, there is some, albeit weak, empirical evidence that the well-documented sound change of velar palatalization by which sequences of /k g/ + /i j e/ change diachronically into alveolar or palatal affricates (Grammont 1933; Bhat 1978; Guion 1998; see Ohala and Solé 2008 for numerous examples) arises from a perceptually conditioned re-analysis of faster speech (Guion 1998). The explanation for such a change in terms of the model derived from the present set of results is first that the magnitude of the coarticulatory influence of front vowels on velar stops would be the same or even increase in a hypoarticulated speaking style (the hypoarticulation in the case of Guion's [1998] analysis being due to fast speech); and second, that /k/ and /t/ would shift closer together in this context, possibly as a result of the degradation of the mid-frequency spectral burst peak in /ki/ (Winitz et al. 1972; Chang et al. 2001) in a hypoarticulated speaking style. That is, the maintenance or increase of anticipatory coarticulation coupled with a weakening of the /k-t/ boundary could diminish the available information in a hypoarticulated speaking style for factoring the fricative noise into a /k/-release on the one hand and the coarticulatory influence due to the high vowel on the other.

The model that is being proposed here is inspired by Ohala's (1993) idea that coarticulatory misparsing contributes to sound change, but with the difference that speaker-hearer relationships are not quite as prominent and also that, following Lindblom et al. (1995) and Bybee (2002), there is a much greater role for the contributory effects of reduction and undershoot in sound change. Thus, whereas in Ohala's (1993) model there is a mismatch between how the speaker interleaves and how the listener parses coarticulation, the mismatch in the model being presented here is instead between coarticulation and categorisation. More specifically, context affects speech production (coarticulation), but it also affects categorisation and not necessarily always in the

same way. It is especially in hypoarticulated speech that the contextual influences that give rise to coarticulation in speech production on the one hand and to shifts in category boundaries on the other may diverge, providing a mismatch that can provide the conditions for sound change to take place. The listener is of course likely to have a role in this divergence, but the mismatch between categories and coarticulation should be in evidence in both speech production (as in Figure 5 of the present study) and in speech perception (Figure 10); thus, what is being proposed here is that under hypoarticulation there can be a change in the relationship between categorisation and coarticulation in both modalities. This proposed model is also consistent with earlier (Phillips 1984; Pierrehumbert 2001, Pierrehumbert 2003; Bybee 2002) and more recent findings (Lin et al. 2014) that sound change may be more likely in words of high frequency because such words, being prone to reduction (Wright 2004), should make the association of coarticulation with the source that gives rise to it more opaque.

In summary, the main conclusion from the study is that there is greater noise in hypoarticulated speech such that the information for separating coarticulation from categorisation is weakened, which in turn creates an ambiguity that can provide the conditions for sound change to take place.

Appendix

Some descriptive statistics of the three duration parameters analysed in Section 2 are included in Table 1 and Figure 14 below.

Table 1: Mean (m) and standard deviation (s) calculated across all speakers for the duration of the stressed vowel (V), the consonant cluster C = /kt/, and the ratio V/C between the two shown separately for stress (accented vs. deaccented), vowel tensity /a, a:/, and mono/disyllables. The V and C durations are in milliseconds.

	Accented								Deaccented							
	/a/				/a:/				/a/				/a:/			
	sackt		sackte		sagt		sagte		sackt		sackte		sagt		sagte	
	m	s	m	s	m	s	m	s	m	s	m	s	m	s	m	s
V	77.7	12.9	78.2	11.8	141.8	27.3	141.1	23.8	74.2	12.9	77.4	12.6	115.1	18.6	119.1	19.1
C	136.7	23.4	150.4	22.8	112.0	19.1	122.9	18.2	116.3	20.9	129.3	17.5	98.0	19.5	109.0	16.8
V/C	0.588	0.152	0.534	0.134	1.307	0.365	1.168	0.244	0.665	0.196	0.613	0.152	1.239	0.437	1.130	0.316
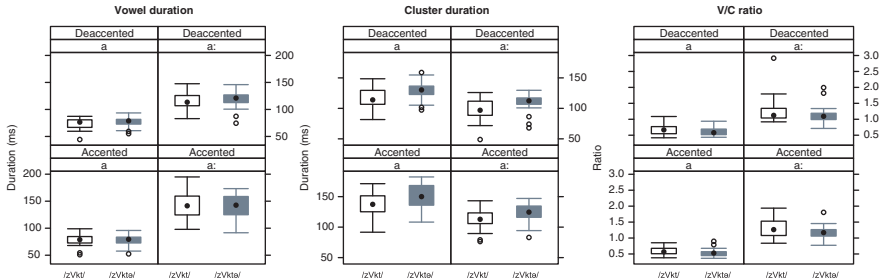
Figure 14: Mean durations of the vowel (left), of the following /kt/ cluster (center), and the ratio between the two (right) shown separately for stress (accented/deaccented), tensity (/a, a:/), and syllable count (white/grey rectangles for monosyllabic/disyllabic). The data were aggregated by speaker so that each boxplot includes one data point per speaker. The extent of the rectangle is the interquartile range.

References

Beckman, Mary E., Kenneth De Jong, Sun-Ah Jun & Sook-Hyang Lee. 1992. The interaction of coarticulation and prosody in sound change. *Language and Speech* 35. 45–58.

Bell-Berti, Fredericka & Katherine Harris. 1979. Anticipatory coarticulation: Some implications from a study of lip-rounding. *Journal of the Acoustical Society of America* 65. 1268–1270.

Bermúdez-Otero, Ricardo. 1998. Prosodic optimisation: The middle English length adjustment. *English Language and Linguistics* 2. 169–197.

Bhat, D. 1978. A general study of palatalization. In Joseph Greenberg (ed.), *Universals of human language*, Vol. 2, 47–91. Stanford: Stanford University Press.

Boersma, Paul & David Weenink. 2012. Praat: doing phonetics by computer [Computer program]. Version 5.3.35. http://www.praat.org/ (accessed 10 December 2012).

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

Chang, Steve, Madelaine Plauché & John J. Ohala. 2001. Markedness and consonant confusion asymmetries. In Elizabeth Hume & Keith Johnson (eds.), *The role of speech perception in phonology*, 79–101. San Diego, CA: Academic Press.

Cole, Jennifer & José Hualde. 2013. Prosodic structure in sound change. In Shu-Fen Chen & Benjamin Slade (eds.), *Studies in South Asian, historical, and Indo-European linguistics: A Festschrift in honor of Hans Henrich Hock on the occasion of his 75th birthday*, 28–45. Ann Arbor, MI: Beech Stave Press.

Cutler, Anne. 1976. Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics* 20. 55–60.

de Jong, Kenneth. 1995. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America* 97. 491–504.

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

Fowler, Carol A. & Mary R. Smith. 1986. Speech perception as 'vector analysis': An approach to the problems of invariance and segmentation. In Joseph Perkell & Dennis Klatt (eds.), *Invariance and variability in speech processes*, 123–139. Hillsdale, NJ: Lawrence Erlbaum Associates.

Fowler, Carol & Jaqueline Thompson. 2010. Listeners' perception of compensatory shortening. *Attention, Perception & Psychophysics* 72. 481–491.

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

Grammont, Maurice. 1933. *Traité de phonétique*. Paris: Delagrave.

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

Harrington, Jonathan, Janet Fletcher & Mary E Beckman. 2000. Manner and place conflicts in the articulation of accent in Australian English. In Michael Broe (ed.), *Papers in laboratory phonology V*, 40–55. Cambridge: Cambridge University Press.

Harrington, Jonathan, Phil Hoole, Felicitas Kleber & Ulrich Reubold. 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, Jonathan, Felicitas Kleber & Ulrich Reubold. 2013. The effect of prosodic weakening on the production and perception of trans-consonantal vowel coarticulation in German. *Journal of the Acoustical Society of America* 134. 551–561.

Heike, Georg. 1972. Quantitative und qualitative Differenzen von/a:/-Realisationen im Deutschen. *7th International Congress of Phonetic Sciences*, Montreal, 725–729.

Hickey, Raymond. 1986. Remarks on syllable quantity in late old English and early middle English. *Neuphilologische Mitteilungen* 87. 1–7.

Hogg, Richard M. 1992. *A grammar of old English, vol. 1: Phonology*. Oxford: Blackwell Publishers.

Hombert, Jean-Marie, John J. Ohala & William Ewan. 1979. Phonetic explanations for the development of tones. *Language* 55. 37–58.

Hoole, Phil & Christine Mooshammer. 2002. Articulatory analysis of the German vowel system. In Peter Auer, Peter Gilles, & Helmut Spiekermann (eds.), *Silbenschnitt und Tonakzente*, 129–152. Tübingen: Niemeyer.

Klatt, D. H. 1973. Interaction between two factors that influence vowel duration. *Journal of the Acoustical Society of America* 54. 1102–1104.

Kleber, Felicitas, Tina John & Jonathan Harrington. 2010. The implications for speech perception of incomplete neutralization of final devoicing in German. *Journal of Phonetics* 38. 185–196.

Lahiri, Aditi & Paula Fikkert. 1999. Trisyllabic shortening in English: Past and present. *English Language and Linguistics* 3. 229–267.

Lehiste, Isle. 1970. *Suprasegmentals*. Cambridge, MA: MIT Press.

Lehiste, Isle. 1972. The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America* 51. 2018–2024.

Lin, Susan, Patrice Speeter Beddor & Andries Coetzee. 2014. Gestural reduction, lexical frequency, and sound change: A study of post-vocalic/l/. *Journal of Laboratory Phonology* 5. 9–36.

Lindblom, Björn. 1988. Phonetic invariance and the adaptive nature of speech. In Ben Elsendoorn & Herman Bouma (eds.), *Working models of human perception*, 139–173. London: Academic Press.

Lindblom, Björn. 1998. Systemic constraints and adaptive change in the formation of sound structure. In James Hurford, Michael Studdert-Kennedy & Chris Knight (eds.), *Approaches to the evolution of language*, 242–264. Cambridge: Cambridge University Press.

Lindblom, Björn, Susan Guion, Susan Hura, Seung-Jae Moon & Raquel Willerman. 1995. Is sound change adaptive? *Rivista Di Linguistica* 7. 5–36.

Lindblom, Björn & Karin Rapp. 1973. Some temporal regularities of spoken Swedish. *Papers in Linguistics from the University of Stockholm* 21. 1–59.

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

Luick, Karl. 1964. (1914–1940). *Historische Grammatik der englischen Sprache*. Stuttgart: Bernhard Tauchnitz/Oxford: Basil Blackwell.

Löfqvist, Anders, Thomas Baer, Nancy McGarr & Robin Story. 1989. The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America* 85. 1314–1321.

Mann, Virginia & Bruno Repp. 1980. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception and Psychophysics* 28. 213–228.

Mehta, Gita & Anne Cutler. 1988. Detection of target phonemes in spontaneous and read speech. *Language & Speech* 31. 135–156.

Mooshammer, Christine & Susanne Fuchs. 2002. Stress distinction in German: Simulating kinematic parameters of tongue-tip gestures. *Journal of Phonetics* 30(3). 337–355.

Mooshammer, Christine & Christian Geng. 2008. Acoustic and articulatory manifestations of vowel reduction in German. *Journal of the International Phonetic Association* 38. 117–136.

Moulines, Eric & Francis Charpentier. 1990. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication* 9. 453–467.

Nooteboom, Sieb. 1972. Production and perception of vowel duration: a study of durational properties of vowels in Dutch. University of Utrecht Ph.D. dissertation.

Ohala, John. 1981. The listener as a source of sound change. In Carris Masek, Roberta Hendrick, & Mary Miller (eds.), *Papers from the parasession on language and behavior*, 178–203. Chicago, IL: Chicago Linguistic Society.

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

Ohala, John & Maria-Josep Solé. 2008. Turbulence and phonology. *UC Berkeley Phonology Lab Annual Report*. 297–355.

Perkell, Joseph. 1990. Testing theories of speech production: Implications of some detailed analyses of variable articulatory data. In W. Hardcastle & Alain Marchal (eds.), *Speech production and speech modelling*, 263–288. Dordrecht: Kluwer.

Phillips, Betty. 1984. Word frequency and the actuation of sound change. *Language* 60. 320–342.

Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Joan L. Bybee & Paul Hopper (eds.), *Frequency and emergence in grammar*, 137–157. Amsterdam: John Benjamins.

Pierrehumbert, Janet B. 2003. Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech* 46. 115–154.

Piroth, Hans & Peter Janker. 2004. Speaker-dependent differences in voicing and devoicing of German obstruents. *Journal of Phonetics* 32. 81–109.

Port, Robert. 1981. Linguistic timing factors in combination. *Journal of the Acoustical Society of America* 69. 262–274.

Port, Robert & Michael O'Dell. 1985. Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13. 455–471.

Röttger, Timo, Bodo Winter, Sven Grawunder, James Kirby & Martine Grice. 2014. Assessing incomplete neutralization of final devoicing. *Journal of Phonetics* 43. 11–25.

Schiel, Florian. 2004. MAuS goes iterative. *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, Lisbon, Portugal (European Language Resources Association, Paris, France), 1015–1018.

Turk, Alice & Stefanie Shattuck-Hufnagel. 2000. Word-boundary-related duration patterns in English. *Journal of Phonetics* 28. 397–440.

Vennemann, Theo. 1972. On the theory of syllabic phonology. *Linguistische Berichte* 18. 1–18.

Vennemann, Theo. 1991. Syllable structure and syllable cut prosodies in modern standard German. In Pier Marco Bertinetto, Michael Kenstowicz, & Michele Loporcaro (eds.), *Certamen phonologicum II: Papers from the 1990 cortona phonology meeting*, 211–243. Turin: Rosenberg & Sellier.

White, Laurence & Alice Turk. 2010. English words on the procrustean bed: Polysyllabic shortening reconsidered. *Journal of Phonetics* 38. 459–471.

Wiese, Richard. 1996. *The phonology of German*. Oxford: Clarendon Press.

Winitz, H, M Scheib & J Reeds. 1972. Identification of stops and vowels for the burst portion of /p, t, k/ isolated from conversational speech. *Journal of the Acoustical Society of America* 51. 1309–1317.

Wright, Richard. 2004. Factors of lexical competition in vowel articulation. In John Local, Richard Ogden, & Rosalind Temple (eds.), *Papers in laboratory phonology VI*, 75–87. Cambridge: Cambridge University Press.

Zheng, Xiaoju & Janet Pierrehumbert. 2010. The effects of prosodic prominence and serial position on duration perception. *Journal of the Acoustical Society of America* 128. 851–859.