Language & Speech, in press

The relationship between the perception and production of coarticulation during a sound change in progress.

Felicitas Kleber, Jonathan Harrington, and Ulrich Reubold
Institute of Phonetics and Speech Processing (IPS), Ludwig-Maximilian-Universität, Munich, Germany

Running head: Perception and production of coarticulation and sound change in progress

ACKNOWLEDGEMENTS

We are very grateful for the constructive criticisms of an earlier draft of this paper by Pam Beddor, Reiko Kataoka, and Editor Keith Johnson. We thank Aysen Demir, Laura Folk, Regina Kaiser, and Ann-Kathrin Killguss for help with labelling, Sarah Hawkins of the Department of Linguistics, University of Cambridge for her assistance in recruiting subjects and Janine Lilienthal for her contribution to running the experiments. We are grateful to all the subjects for participating in this study. 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."

Corresponding author

Felicitas Kleber
Institute of Phonetics and Speech Processing (IPS)
Ludwig-Maximilian-Universität
Schellingstr. 3/II
80799 Munich
Germany
E-Mail: kleber@phonetik.uni-muenchen.de

ABSTRACT

The present study is concerned with lax /ʊ/-fronting in Standard British English and in particular with whether this sound change in progress can be attributed to a waning of the perceptual compensation for the coarticulatory effects of context. Younger and older speakers produced various monosyllables in which /ʊ/ occurred in different symmetrical consonantal contexts. The same speakers participated in a forced-choice perception experiment in which they categorized a synthetic /ɪ-ʊ/ continuum embedded in fronting /s_t/ and non-fronting /w_l/ contexts. /ʊ/ was shown to be fronted for the younger age group both in production and perception. Although there was no conclusive evidence that younger listeners compensated less for coarticulation than did older listeners, the size of the coarticulatory influence of consonantal context on /ʊ/ in perception was found to be smaller than in production for the younger than for the older group. The findings are consistent with a model of sound change in which the perceptual compensation for coarticulation wanes ahead of changes that take place to coarticulatory relationships in speech production. As a result, the perception and production of coarticulation may be unusually misaligned with respect to each other for some speaker-listeners participating in a sound change in progress.

KEYWORDS

Sound change in progress, compensation for coarticulation, /ʊ/-fronting

INTRODUCTION

In the last 20-30 years, Ohala (1993a, 2005) has developed a model of sound change that is founded on the idea that there can be a mis-association between the speaker's production and the listener's perception of coarticulation. In Ohala's model, the production of coarticulation can lead to a so-called parsing error in perception that is, in turn, the driving force for sound change. As an example of the way that coarticulatory relationships can be misperceived, consider the first vowel in the German noun *Nutzen* (*benefit*) which often sounds like [ʏ] to the second author of this paper whose L1 is English. But it does not to the other two authors: this is not only because it is part of their L1-German knowledge that the vowel is /ʊ/ (which contrasts with /ʏ/ in German), but also because they have learned the coarticulatory patterns in German that cause these kinds of fronting effects due to the coronal context in production. Thus, L1-speakers of German do not hear the vowel as fronted because they compensate for this fronting effect in perception (Mann & Repp, 1980): that is, they factor out the proportion of F2-raising that they assume to be attributable to context and consequently hear a back, not a front, vowel.

For the L1-German listeners in this anecdotal example, the perception and production of coarticulation can be said to be aligned. This is because there are divergent variants in production between ones in which the tongue dorsum is produced as a back vowel in a labial context such as *Puppe* (*doll*) but advanced in a coronal context like *Nutzen*. If listeners compensate for coarticulation, then the perceptual boundaries for these words are analogously divergent: that is, listeners accept a higher F2 in a fronting than in a backing context as evidence for /ʊ/. A consequence of such correspondences between the production and perception of coarticulation is that listeners are often oblivious to marked acoustic differences in phonetic variants, if such differences can be attributed to context: thus the phonetic differences between initial schwas in /əCV/ sequences that arise due to anticipatory coarticulation of the transconsonantal vowel are largely imperceptible because, as various cross-splicing experiments have shown (Fowler, 2005), listeners subtract out the influences due to context, with the consequence that the same context-independent /ə/ (and hence the same percept) remains.

For Fowler, studies such as these demonstrate the common currency of speech perception and production (Fowler, Brown, Sabadini, & Weihing, 2003): listeners' percepts are context-independent because they parse in accordance with coarticulation the same gestures that are non-destructively layered in speech production. Nevertheless, listeners do not always manage to attribute accurately coarticulatory perturbations in production to context as perceptual compensation studies of nasal (Fowler & Brown, 2000; Beddor, Brasher, & Narayan, 2007) and VCV (Beddor, Krakow, & Lindemann, 2001; Beddor, Harnsberger, & Lindemann, 2002) coarticulation have shown.

One of the reasons why listeners may fail to reverse coarticulation accurately is that coarticulatory perturbations can make the acoustic signal inherently ambiguous. As Hombert, Ohala, and Ewan (1979) have shown, the diachronic development of phonological tone and loss of stop voicing contrast in many East and Southeast Asian languages may have come about because intonation and microprosody compete on the same acoustic parameter, fundamental frequency: consequently, listeners may attribute the microprosodic perturbations caused by the [±voice] distinction not to its source, the consonant, but incorrectly to the vowel. Yet another may be that if coarticulation is language-specific, as a number of studies (Öhman, 1966; Clumeck, 1976; Magen, 1984; Manuel & Krakow, 1984; Recasens, 1987; Huffman, 1988; Keating & Cohn, 1988; Choi & Keating, 1990; Recasens, Pallarès & Fontdevila, 1998; Manuel, 1999; Beddor, Harnsberger & Lindemann, 2002; Oh, 2002) have shown, then listeners of different dialects or linguistic backgrounds may not all undo the coarticulatory effects in production in the same way (as the earlier anecdotal example also suggests). Although there is much evidence to show that the temporal extent and magnitude of coarticulation are lawfully determined by the capabilities and limitations of the vocal tract in production (e.g., Mann, 1986; Fowler, Best, & McRoberts, 1990; Fowler, 2005), coarticulation has also been shown to be speaker-specific even when speakers are of the same language and variety (van den Heuvel, Cranen, & Rietveld, 1996; Magen, 1997; Grosvald, 2009). Such variation across speakers may come about because coarticulation is affected by speaking style (Krull, 1989) and because phonetic detail has been shown to be idiosyncratic (Johnson, 2006). If, following exemplar theory, phonological generalizations over phonetic detail depend statistically on learning experience (Pierrehumbert, 2003, 2006), then, given that no two speakers are ever exposed exactly to the same speaking situations, phonetic detail and therefore also the coarticulatory relationships are likely to vary slightly from speaker to speaker: therefore if within the same speaker the production and perception of coarticulation are closely tied, then listeners may not all compensate for coarticulation in the same way.

The question of whether different listeners compensate differently for coarticulation was investigated for two age-groups by Harrington, Kleber and Reubold (2008) in a study of the influence of consonantal context on the diachronic fronting of /u/ (lexical set GOOSE), a sound change that has been taking place in the Standard accent of England (henceforth Standard Southern British, SSB) in the last fifty years (Henton, 1983; Bauer, 1985; Hawkins & Midgley, 2005; de Jong, McDougall, Hudson, & Nolan, 2007; McDougall & Nolan, 2007). Harrington et al. (2008) showed that /u/ was phonetically more advanced and that the coarticulatory influences of context on /u/ were less marked for younger than for older subjects. Thus the older subjects differed from younger speakers both because they compensated more for coarticulatory influences of context on /u/ in perception, and because there was a greater divergence between the phonetic variants of /u/ due to the coarticulatory influences in fronting (e.g. *few*) vs. non-fronting (e.g. *swoop*) contexts than for younger subjects. Harrington et al. (2008) sought an explanation for their results in terms of two different models. In one, in which we extended Ohala's (1993a) model to a sound change in progress, the extent to which younger listeners compensate for coarticulation was presumed to have waned leading to a consequent realignment of the non-fronting/back variants of /u/ to the front: thus is this model, giving up compensation for coarticulation was considered to be the cause of diachronic /u/-fronting in SSB. In the other, which was inspired by an exemplar model of speech perception, we reasoned that the higher statistical frequency with which /u/ occurs in a fronting context in Standard Southern British (Harrington, 2007) had produced a realignment of the non-fronting variants towards the front in production and an associated realignment in perception. Thus in this model, the changed coarticulatory relationships observed between the two age groups was a by-product, rather than a determiner of diachronic /u/-fronting. The conclusion in Harrington et al. (2008) was that the observed age-dependent differences in the production and perception of speech were consistent with either of these interpretations.

In this paper, the aim is to test further such age-related differences in the perceptual compensation for coarticulation by analysing the fronting of lax /ʊ/ in the lexical set FOOT which, like /u/, has also been subject to diachronic fronting in SSB in the last 50 years (Hawkins & Midgley, 2005; Fabricius, 2007). Our aim in extending our analysis in Harrington et al. (2008) to /ʊ/ was threefold. Firstly, we sought to establish whether for the younger group /ʊ/ was fronted in both production and perception and secondly whether there were similar age-related differences in the perceptual compensation for coarticulation. We also sought to test whether any group differences in compensating for coarticulation perceptually were matched by their coarticulatory patterns in production, as we had found to be for tense /u/ in Harrington et al. (2008). One of the reasons why tense /u/ and lax /ʊ/ might be different in this respect is because diachronic fronting has been shown to be a more recent sound change for the lax than for the tense vowel (Hawkins & Midgley, 2005). Thus it is possible that the perceptual compensation for coarticulation may have begun to wane for younger listeners but that the coarticulatory perturbation due to consonantal context on /ʊ/ in production is still quite marked, or has, at least, not yet weakened to the extent that it was shown to be for /u/. Such a prediction would follow from our proposed extension of Ohala’s (1993a) model to a sound change in progress in which the waning of the compensation of coarticulation in perception precedes a realignment of the variants in productions.

Another reason why the production and perception characteristics of the diachronic change might be different in the tense and lax vowels is because /ʊ/ unlike /u/ does not occur with greater statistical frequency in a consonantal fronting context: therefore, the explanation summarized above that statistical frequency effects might have brought about the diachronic shift of the non-fronting variants towards the front in the production of /u/ does not readily carry over to the lax vowel.

EXPERIMENT I: SPEECH PRODUCTION

Method

Subjects. 33 SSB speakers participated in the following production experiment as well as the perception study described in the next section. The subjects were recruited from around the University of Cambridge. The subjects who had no known reading or hearing difficulties, were assigned to one of two age groups: 18 subjects (9 females, 9 males) aged between 19 and 21 belonged to the younger group (average age = 20.2 years) and 15 subjects (8 females, 7 males) aged between 54 and 89 to the older group (average age = 75.4 years). Seven from the younger and nine from the older group had also participated as subjects in the study on /u/-fronting in Harrington et al. (2008).

The subjects were recorded and tested in a quiet room at the University of Cambridge. However, some subjects from the older group were recorded and tested in a quiet room in their homes. No apparent differences in the quality of the recorded speech signal and no differences in the subjects' identification performance related to the recording and testing locations could be found. Two subjects from the older group participated only in the perception experiment.

Materials, digitization, labelling. The recording material consisted of 54 monosyllabic words which included eight target words containing the vowels from the lexical sets FOOT (*cook*, *hood*, *soot*, *wool*) and KIT (*kick, hid*, *sit*, *will*). We also analysed the vowel in *hard* which is an open back vowel (SSB is non-rhotic) in order to provide a reference point for determining the relative degree of fronting for any given speaker. The remaining 45 words were fillers and served as distracters. As far as the eight target words were concerned, we assumed that the coronal /s_t/ context would induce vowel fronting whereas the /w_l/ context would result in vowel backing (due both to the labial-velar /w/ and the velarised realisation of the lateral); the other two contexts /h_d/ and /k_k/ were expected to have a more marginal influence along a phonetic front-back dimension.

Recordings were made using the SpeechRecorder software (Draxler & Jänsch, 2004), a notebook (Toshiba Tecra) and a stereo headset (Sennheiser pc 165 USB). Ten repetitions each of the 54 words were presented in random order and in isolation to the subjects on a computer screen. Subjects were asked to read aloud each word within a time slot of 1.6 seconds. There was then an automatic pause of 1 second (that could be over-ridden by the subject if more time was needed) to the next items presented on the screen. Each subject produced a total of 540 words. The recordings were digitized at 44.1 kHz. Any mispronunciations were excluded from the analysis. Formant frequencies were calculated (LPC order of 10, a pre-emphasis of 0.95, 30 ms Blackman window with a frame shift of 5 ms) for the target words and manually corrected when necessary. All of the data were segmented and labelled automatically into phonetic segments using the Munich Automatic Segmentation System (MAuS, Schiel, 2004); manual readjustments were made subsequently whenever necessary to the target words. The formant frequencies in Hertz were converted to the Bark scale using the formula in Traunmüller (1990). The segmentation boundaries for the vowel extended between its acoustic onset and offset in *cook*, *hood*, *soot* and between the acoustic onset of /w/ and acoustic offset of /l/ in *wool*. Our dynamic parameterizations (described below) were based on formants extracted between the acoustic onset and offset of the vowel defined in this way. Our static analyses were taken from formants extracted at the temporal midpoint of the vowel relative to the acoustic vowel onset and offset for all contexts except /w_l/; for the latter, we extracted formant frequencies at the time point at which the energy reached a maximum value (which turned out to be very close to the temporal midpoint as defined above).

Parameterization. In order to quantify the extent of coarticulatory influence due to consonantal context, we calculated the distance of the vowels in *cook*, *soot*, and *wool* to those in *hood* (which is typically defined as a neutral context). This calculation was done separately for each speaker and based on parameterizing the dynamic F2 trajectory using the discrete-cosine-transformation (DCT) over the extent of the vowel (Watson & Harrington, 1999; Harrington et al., 2008; Marin, Pouplier, & Harrington, 2010). The application of the DCT to an F2-trajectory as a function of time results in a triplet of coefficients that are proportional respectively to the mean, linear slope, and curvature of F2 (see Appendix A for further details of the DCT). For each speaker separately, we calculated the centroid, i.e. the mean of *hood* in this three-dimensional space and then the Euclidean distance from each token of the other three words *soot*, *cook*, *wool* to this centroid. The greater the Euclidean distance, the further the deviation from the *hood* context and therefore by inference the greater the influence of the consonantal context on /ʊ/.

Results

We present in this section firstly an overview of the position of /ʊ/ in relation to the front vowel /ɪ/ and to the back vowel /ɑ/ for data extracted at the vowels' temporal midpoints. We then present the results in which the coarticulatory influence of consonantal context on /ʊ/ was quantified using the DCT-based, Euclidean distance metric described in the preceding section.

Figure 1 about here

For the first of these, the raised F2 and above all the relatively closer position of /ʊ/ to /ɪ/ than to /ɑ/ for the younger compared with the older speakers provides evidence consistent with other studies (Hawkins & Midgley, 2005; Fabricius, 2007) that /ʊ/-fronting is a sound change in progress in SSB. As Figure 1 shows, F2 is not only raised for both younger male and female speakers relative to F2 of older speakers, but in addition, context influences /ʊ/ differently in the age groups in two ways: firstly, whereas F2 of /ʊ/ in *soot* is evidently higher than F2 of /ʊ/ in *hood* for the older speakers (second row, Figure 1), *soot*, *hood* (and *cook*) cluster much more closely together on this parameter in the younger group; but secondly, whereas *hood* is closer to *wool* for the older speakers, the comparable distance for the younger speakers is much greater. In general, the data in Figure 1 is compatible with the following sound change: /ʊ/ has fronted and in doing so the differences due to neutral and fronting contexts have been reduced: thus the mean positions of *soot*, *hood* and *cook* are closer together in younger than in older speakers. Although /ʊ/ in *wool* has also fronted (the positions of *wool* are slightly more advanced in younger than in older speakers) it has not fronted to nearly the same degree: as Figure 1 shows, whereas *wool* lies within the ellipse for *hood* for older speakers, it is more peripheral for the younger speakers.

Figure 2 about here

A similar pattern of results and in particular these different effects of context on /ʊ/-fronting are also evident in Figure 2 which shows time-normalized F2 plots for older and younger speakers. The differences between *soot*, *hood*, *cook* at both the onsets and offsets of the trajectories (at proportional times 0 and 1 respectively) - that is, the points at which the influence of consonantal context is greatest - are much more pronounced in the older than in the younger group: thus the diachronic change of /ʊ/-fronting seems to be associated with a levelling of the contextual differences between these words. These contextual differences between the age groups are also in evidence in Figure 2 near the temporal midpoint of the trajectories (close to proportional time 0.5) at which the undershoot of *soot* relative to *hood* is much greater in the older than in the younger group. The same figure confirms what was evident in Figure 1: firstly, that the contextual separation of *wool* from the other three words is pronounced for younger speakers; and secondly, that while *wool* has fronted diachronically (as is evident in comparing the higher F2 peak in *wool* for the younger speakers at around 1000 Hz with the lower F2 at around 880 Hz in the same word for the older speakers), it has not fronted to nearly the same extent as in the other words from which it is therefore isolated in F2.

Figure 3 about here

The results of quantifying these contextual differences by calculating separately for each speaker the distance of each token to the speaker-centroid of *hood* in a space formed from the coefficients that were derived by applying the DCT to each F2 trajectory (in the manner described in 2.1.3) are shown in Figure 3. We made one further adjustment to this calculation in order to show the differences between the fronting and backing effects of context. If the mean F2-trajectory (calculated between the onset and offset of the vowel) for any given /ʊ/ token was less than the speaker's F2 mean for *hood*, then the Euclidean distance was multiplied by -1 to convert it to a negative value. Thus as Figure 3 shows, the median of *hood* is predictably distributed around zero (since these are the distances of the *hood* tokens to their own *hood* centroid); the median of *wool* is negative for both age groups because the mean F2 for *wool* was in almost all cases much lower than that of *hood*; and for the older speakers, the F2 median of *soot* is positive because, as Figure 2 shows, the F2 mean of *soot* is greater than that of *hood*. Other than this adjustment for sign, the y-axis of Figure 3 is the Euclidean distance (in the space of the first 3 DCT coefficients) to the centroid of *hood*. What the figure shows, then, is what was also evident in Figures 1 and 2: firstly, the influence of context on *cook* and *soot* was more marked in the older than in the younger speakers, as shown by the much greater differences for the older speakers between the boxplots for these words; and secondly, the effect of context on *wool* was probably greater for the younger than for the older speakers, as shown by the greater (negative) Euclidean distance for the younger speakers from *wool* to *hood*. Thus, once again it is evident that there is a closer approximation between all contexts except *wool* for younger compared with older speakers.

We quantified the influence of context on /ʊ/ with a repeated-measures MANOVA with dependent variable Euclidean distance (the values shown in Figure 3) and independent variables Word (three levels: *cook*, *soot*, *wool*) and Age (two levels: young vs. old). The results showed an effect for Age ($F[1,30] = 9.8$, $p < 0.01$). They also showed an effect for Word ($F[2,29] = 466.1$, $p < 0.001$) which, consistently with Figure 3, shows that the Euclidean distances to *hood* is different for the three word types. There was also a significant Word × Age interaction ($F[2,29] = 25.8$, $p < 0.001$) which means that the effect of context was different for the two age groups. Post-hoc *t*-tests with Bonferroni correction showed significant differences between the younger and older groups on all three word types. Thus, in combination with Figure 3, this last result shows that the Euclidean distances from *soot* to *hood* were significantly greater for the older than for the younger speakers ($t = 5.1$, $p < 0.001$) as were those from *cook* to *hood* ($t = 4.2$, $p < 0.01$) whereas the distances from *wool* to *hood* were significantly greater for the younger than for the older speakers ($t = 4.8$, $p < 0.001$).

We repeated the same analysis using a static measure of the raw F2 values extracted at the temporal midpoint of the vowel. The dependent variable in this case was calculated separately for each speaker as the absolute difference to the same speaker's F2 mean of *hood*. The results showed an effect for Age ($F[1,30] = 17.6$, $p < 0.001$), for Word ($F[2,29] = 165.7$, $p < 0.001$) and a significant Word × Age interaction ($F[2,29] = 16.8$, $p < 0.001$). Post-hoc *t*-tests with Bonferroni correction showed significant differences between the younger and older speakers on the distances from *soot* to *hood* ($t = 3.3$, $p < 0.05$) and on *wool* to *hood* ($t = 5.2$, $p < 0.001$) but no differences between the age groups on the *cook* to *hood* distances. Thus, with the exception of the *cook-hood* distances, the pattern of results was the same as for those based on the DCT transformation. The reason for the discrepancy between the two sets of results is likely to be that, whereas on the static analysis the difference in the distance between *cook* and *hood* was quite similar for both age groups (Figure 1), the difference in the entire shape of the F2 trajectory between *hood* and *cook* was much greater for the older speakers (because, as Figure 2 shows, of the steeply rising F2-transition towards the offset of *hood*) than for the younger speakers.

We also quantified the relative distance between *soot* and *wool* both in order to provide a more direct basis for comparison with the subsequent perception experiment and to test whether the inter-Euclidean distance between these words was greater for the younger group. To do this, we calculated E_{soot}, the Euclidean distance in the same DCT-parameterised F2 space from each separate *soot* and *wool* token to the centroid of *soot*; and E_{wool}, the Euclidean distance from each *soot* and *wool* token to the centroid of *wool*. The logarithm of these ratios, *Eratio*, was calculated from (1):

$$\mathrm{Eratio} = \log(\mathrm{E}_{\mathrm{wool}}/\mathrm{E}_{\mathrm{soot}}) \qquad (1)$$

Eratio was calculated in (1) separately for each speaker. When *Eratio* is zero, then a given token is equidistant between the same speaker's centroids of *soot* and *wool*. Tokens that are close to *soot* have positive values on (1) because the denominator (distance to the centroid of *soot*) is small in relation to the numerator (distance to the centroid of *wool*); analogously, tokens that are close to the centroid of *wool* have negative values on (1). The metric in (1) is therefore a way of quantifying the extent of overlap between the *soot* and *wool* spaces, while at the same time providing a form of speaker normalization (since for all speakers, a point equidistant between *soot* and *wool* is zero on (1)): thus, the more distant the distributions of *wool* and *soot* are from each other for any speaker, then the greater the separation between these words in opposite directions away from zero on (1). The hypothesis to be tested was that the *soot-wool* distances were greater for the younger than for the older group. The mean values for *Eratio* on *soot* and *wool* were 2.27 and -2.11 respectively for the older group, and 2.99 and -2.29 respectively for the younger group (Figure 4, left panel). The right panel of Figure 4 shows the mean difference on this parameter between *soot* and *wool*, with one value per speaker; the effect of age group (young vs. old) on these mean differences was significant ($t_{28.6} = 3.7$, $p < 0.001$). Thus the parameterized F2-distance between *soot* and *wool* was greater for the younger than for the older speakers.

Figure 4 about here

We repeated the above analysis using a static measure of the raw F2 values in Bark extracted at the temporal midpoint of the vowel. For this purpose, we calculated separately for each speaker the absolute distance from F2 at the vowel temporal midpoint of all *wool* tokens to the F2-mean of *soot*; and the absolute distance from F2 at the vowel temporal midpoint of all *soot* tokens to the F2-mean of *wool*. The means across all of these distances by age group were 1.54 Bark for the older and 2.70 Bark for the younger speakers. The results of a repeated measures MANOVA with F2 distance as the dependent variable and independent factors Age (two levels) and Type (two levels depending on whether the distances were calculated to the mean of *soot* or to the mean of *wool*, as outlined above) showed a significant effect for Age ($F([1,31] = 20.3$, $p < 0.001$), no effect for Type, and no interaction between Age and Type. Thus, consistently with the DCT analysis based on the entire trajectory, and irrespective of whether distances are measured from *soot* to the mean of *wool* or from *wool* to the mean of *soot* (the factor Type above), there was a significantly greater distance between *soot* and *wool* for the younger than for the older group based on an analysis of F2 at the vowels' temporal midpoints.

Discussion

/ʊ/ has fronted in younger speakers and in such a way that the differences between the variants in *cook*, *hood*, and *soot*, are close together, and much closer together than in older speakers. This implies that the coarticulatory influences of the consonant on /ʊ/ at least across these three contexts are less for the younger than for the older speakers. Secondly, the comparison across the age-groups suggests that the sound change has involved a fronting of /ʊ/ towards a position at which it is strongly coarticulated with the alveolar place of articulation. If this is so, then the F2-target of /ʊ/ for the younger speakers should be positioned at approximately the same frequency as the F2-onset of *soot* for the older speakers. Alternatively, since some of the /ʊ/-transition is likely to be masked by the frication of the /s/ (given that formant transitions extend into the aperiodic /s/, as Soli (1981) has shown), then the F2-offset of *hood* may be a better indicator of the extent to which /ʊ/ is displaced due to the alveolar context in the older speakers. As Figure 2 shows, the F2 values averaged across the older male speakers at the onset of *soot* and at the offset of *hood* are 10.1 Bark and 10.5 Bark respectively; the mean F2 at the temporal midpoint of *hood* for the younger male speakers is located at a similar frequency of 10.5 Bark. The corresponding values for the older female speakers were 11.2 Bark (onset of *soot*), 11.7 Bark (offset of *hood*), and for the younger female speakers 11.6 Bark (temporal midpoint of *hood*). These data are therefore consistent with the idea that the sound change has involved an approximation of /ʊ/ in *hood* (and *cook*, given that *hood* and *cook* have similar F2 values at the midpoint as Figure 2 shows) towards the point at which /ʊ/ coarticulates maximally with an alveolar consonant. On the other hand, diachronic /ʊ/-fronting hardly seems to have taken place at all in *wool*, or else the lip-rounding in [w] and tongue-dorsum retraction due to the final velarised [ɫ] combine to lower F2 so much in younger speakers' productions of /ʊ/ in this context that the age-dependent differences scarcely emerge. In the next experiment, we seek to shed light on how these age-dependent differences in /ʊ/-fronting and coarticulatory patterns are related both to perception and to the sound change that could have given rise to them.

The first prediction follows from our extension of Ohala’s (1993a) model in which the shift in variants in production *is a consequence of* a waning of the perceptual compensation of coarticulation. According to our extension of this model, the extent to which listeners attribute perceptually the perturbation of the vowel to the consonantal context diminishes as they begin to phonologize the variant in a fronting alveolar context. Since /ʊ/ would then be classified perceptually without attributing as much variation to a consonantal source, then the same acoustic token of /ʊ/ should be perceived to be similar even when embedded in different contexts: that is, the context-dependent separation between /ɪ-ʊ/ boundaries should be narrower for younger listeners (who have participated in this sound change) than for older listeners (who have not) if younger listeners no longer attribute as much variation in the vowel to a consonantal source as do older listeners. Such a finding would imply a mismatch for the younger listeners between the perception and production of coarticulation: they would have contextual boundaries that were narrowly spaced in perception (if they no longer compensate for coarticulation), even though their phonetic variants in fronting (*soot*) and non-fronting (*wool*) contexts were even more widely spaced than for the older subjects. In fact, this mismatch is just what might be expected from a model in which giving up perceptual compensation for coarticulation drives the realignment of variants in production. This is because in such a model, perceptual change (i.e., a closer approximation of the perceptual boundaries in the fronting and non-fronting contexts) should *precede* the corresponding changes in speech production.

Alternatively, perhaps listeners’ strategies for compensating for coarticulation are rather more directly tied to their coarticulatory patterns in production. Since the variants of *soot* and *wool* are more widely dispersed for the younger than for older speakers, then their perceptual boundaries in these words should also be differentiated to a greater extent. This is precisely the pattern of age-dependent relationships found in Harrington et al. (2008) for the tense vowel /u/ in which there was both a wider dispersion in production and a greater differentiation of the perceptual category boundaries between front and back variants of tense

/u/ for the older than for the younger group. This prediction of a close match between the production and perception of coarticulatory patterns is also consistent with the idea of a common currency of gestures (Fowler et al., 2003) as a result of which coarticulation is parsed in perception according to the way in which gestures are temporally layered in speech production.

In summary, there are two competing hypotheses. The first hypothesis is that the younger speakers no longer compensate for coarticulation perceptually, or do so less than the older speakers. Consequently, their phonetic-to-phoneme mapping is accomplished without taking account of the perturbing effect of consonantal context: this leads to the prediction that the perceptual /ɪ-ʊ/ boundaries in a fronting context like *sit-soot* and in a non-fronting context like *will-wool* should be quite close together, or certainly closer together than for older listeners, even though in production their variants in these words are further apart. The second hypothesis is that production and perception are closely matched: that is, since younger speakers have variants of *soot* and *wool* that are further apart than in older speakers, then their category boundaries in a fronting context like *sit-soot* and in a non-fronting context like *will-wool* will be further apart than for older listeners.

EXPERIMENT II: SPEECH PERCEPTION

Method

Participants. The same subjects as in Experiment I took part in the perception experiment.

Materials. We created two 13-step synthetic continua one each between the minimal word pairs *sit-soot*, and *will-wool*. The stimuli were created using the formant synthesiser HLSyn (High Level Parameter Speech Synthesis System, version 2.2).

Table I about here

The vowels of the stimuli were all synthesized with the same acoustic parameters and between the same endpoints (1100 Hz and 2100 Hz) that had been established in independent tests in which two trained L1-English phoneticians had been asked to rate the naturalness of various synthetic stimuli. The step-size between the 13 stimuli was fixed at 0.352 Bark (Table I). F1, F3 and F4 had fixed formant frequencies of 400 Hz, 2600 Hz and 3500 Hz respectively, and all vowels had a steady-state section of 80 ms. As summarized in Table II, only the F2-locus and the duration of the transition differed between the continua depending on the context.

Table II about here

The fundamental frequency across the vowel in all synthetic stimuli decreased linearly from 120 Hz to 95 Hz. The speaker's voice was that of a young to middle-aged male talker. This was confirmed in a subsequent perception test that we carried out with 19 first language German listeners (students at the IPS, Munich) who listened to four tokens (one from each of the four continua below) and were asked to circle one of six possible age ranges (10-15 years, 15-25 years, 25-35 years, 35-45 years, 45-55 years, 55-65 years, and over 65 years). The two most frequent categories chosen were 25-35 years (41% of responses) and 35-45 years (29% of responses); the mean and standard deviation of the estimated age of the synthesized voice were 36.2 years and 10.2 years.

Experimental procedures. Each stimulus was repeated 5 times and presented in randomized order (2 continua × 13 stimuli × 5 repetitions = 130 randomized stimuli per subject). In a two-alternative, forced choice identification task, subjects were asked to mark on an answer sheet – depending on the continuum – which word they had heard (*will* or *wool*; *sit* or *soot*). The 50% perceptual /ɪ-ʊ/ cross-over boundaries were calculated with logistic regression and a generalized linear mixed model using the stimulus response as the dependent variable, the stimulus number (1-13) as the independent variable, and with the listener as a random factor. The result of this operation was to fit a logistic function to the stimulus responses (separately by listener and by continuum) using the relationship

$$p = \frac{e^{(mx+k)}}{1+e^{(mx+k)}} \qquad (2)$$

where p was the predicted proportion of /ʊ/ responses ($0 < p < 1$), the coefficients m and k were calculated separately for each listener, and x was the stimulus number 1, 2, ...13. The 50% cross-over boundary (i.e. the value of x for which $p = 0.5$ in (2)) was given by:

$$-k/m \qquad (3)$$

33 (listeners) × 2 (contexts: *sit-soot*, *will-wool*) = 66 /ɪ-ʊ/ cross-over boundaries were obtained in this way from (2) and (3).

Results

Figure 5 about here

The results in Figure 5, in which the fitted response curves in (2) have been averaged by continuum and age group, show firstly that the *sit-soot* boundary was left-shifted (i.e. a greater proportion of /ʊ/ responses) relative to that of *will-wool*, secondly that (across both words) the /ɪ-ʊ/ boundaries were left-shifted for the younger compared with the older listeners, and thirdly that the difference in responses to the two continua was slightly less for the younger listeners. The first of these results is predictable from a number of studies demonstrating compensation for coarticulation (Lindblom & Studdert-Kennedy, 1967; Mann & Repp, 1980): the same acoustic stimulus was more likely to be heard as /ʊ/ when embedded in *sit-soot* than in *will-wool* because listeners attributed a certain amount of F2-raising to the coarticulatory effects of undershoot that is induced by the alveolar context. The second result is consistent with the finding from Experiment I in which /ʊ/ was shown to be phonetically more advanced in production for the younger than for the older subjects. The third result suggests that the younger listeners compensated less than the older listeners for the effects of context: that is, the effects of coarticulation in perception were not matched with those in production since the acoustic-phonetic difference between the /ʊ/-variants in *soot* and *wool* was greater when produced by the same younger compared with older subjects.

The results of a repeated measures MANOVA with dependent variable 50% cross-over boundary, independent variables Continuum (two levels: *sit-soot* vs. *will-wool*) and Age (two levels: young vs. old) showed a significant effect for Continuum ($F[1,31] = 176.3$, $p < 0.001$) and for Age ($F[1,31] = 22.4$, $p < 0.001$) and a not-quite significant Continuum × Age interaction ($F[1, 31] = 3.9$, $p = 0.06$). This last result is suggestive, although not entirely conclusive, that the effect of continuum type (context) on the two age groups is not the same.

Figure 6 about here

We explored this interaction further by testing for age effects in the difference in the cross-over boundaries between the two continua. In order to do this, we subtracted separately for each speaker the 50% cross-over boundaries of the *will-wool* from the *sit-soot* continuum. The results of this difference in Figure 6 are consistent with the average response-curves shown in Figure 5: thus as the boxplots in Figure 6 show, the cross-over boundary in *sit-soot* and *will-wool* were somewhat closer together for the younger than for the older listeners. The results of a *t*-test applied to these cross-over boundary differences showed a not quite significant effect for Age ($t = 1.99$, $df = 30.8$, $p = 0.06$). Thus, once again there is evidence only of a tendency that the younger listeners compensated less for the effects of coarticulation perceptually than older listeners.

Figure 7 about here

The second prediction from our extension to Ohala's (1993a) model was that if giving up compensating for coarticulation drives sound change in production, then the shift of the boundary in the non-fronting *will-wool* context towards that of the fronting *sit-soot* context should precede the approximation of the variants in analogous contexts in speech production.

Consequently, the influence of coarticulation in perception should be less than the influence of coarticulation in production for the younger compared with the older subjects.

In order to explore this further, we compared separately for each subject the extent of separation between their perceptual boundaries in the two contexts with the acoustic distance between two variants of /ʊ/ in the two contexts *soot* and *wool*. These data are shown in the left panel of Figure 7 in which each speaker-listener is represented by a pair of points. The horizontal axis in this figure is the same measure of logarithmic acoustic distance shown in the boxplots in Figure 4; the vertical axis shows the stimulus number of the 50% perceptual cross-over boundaries in *sit-soot* and *will-wool* (the data corresponding to the vertical lines from Figure 5 but expressed in terms of the stimulus number rather than in Bark). The main hypothesis to be tested here was that the same distance between the fronting and non-fronting contexts perceptually should correspond to a larger acoustic distance between the /ʊ/-variants in the two contexts for the younger compared with the older subjects. Equivalently, the gradient of the line that connects the pair of points for each speaker-listener in Figure 7 should be smaller for the younger than for the older subjects. In order to test this, the gradient, *g*, was calculated separately for each speaker from:

$$g = (\text{soot}_P - \text{wool}_P) / (\text{soot}_A - \text{wool}_A) \qquad (4)$$

where $soot_P$ and $wool_P$ are the values of the 50% cross-over boundaries from the *sit-soot* and *will-wool* continua respectively, and in which $soot_A$ and $wool_A$ are the relative acoustic distances in the same words (the data from Figure 4). The right panel of the same figure shows that the gradients calculated from (4) (one value per speaker) were smaller for the younger than for the older subjects and this difference was significant ($t_{28.1} = 2.68$, $p < 0.05$). Thus these data show that the same perceptual distance between the two contexts was matched by a larger acoustic difference between the variants of /ʊ/ in fronting and non-fronting contexts for the younger listeners. These data are therefore consistent with the hypothesis that during a sound change in progress, the shift in coarticulatory relationships in perception precede those in production: more generally, this pattern of results is compatible with the idea that the diachronic change in production is a consequence of a waning of compensation for coarticulation in perception.

GENERAL DISCUSSION

There were three main findings in this study. The first was that younger subjects' /ʊ/ was overall more fronted in both perception and production compared with that of older subjects. The second was that the differences between the age groups in the production of /ʊ/ of *wool* were marginal. The third was that the extent to which subjects compensated for coarticulation perceptually in relation to the influence of context on /ʊ/ in production was smaller for the younger than for the older subjects: that is, the same perceptual shift to the boundary of /ɪ-ʊ/ induced by context corresponded to a larger shift in production between the variants of /ʊ/ in fronting (*soot*) and non-fronting (*wool*) contexts for younger subjects. We now consider some interpretations of these data.

The first of these is our proposed extension to Ohala's (1993a) model in which younger listeners' perceptual compensation for /ʊ/ in a fronting context has waned bringing about a shift in production of the non-fronting variants to the front: this model was proposed as one of the possible interpretations for the coarticulatory differences between the older and younger subjects in the analysis of diachronic fronting of tense SSB /u/ in Harrington et al. (2008). According to this model, younger listeners give up compensating for the coarticulatory fronting effects of consonantal context and phonologize the fronted variant (i.e., extend it to other contexts). If the perceptual compensation for coarticulation has waned in younger listeners, then their perceptual boundaries in fronting and non-fronting contexts should be closer together than for older listeners who attribute a greater proportion of the variation in the vowel to consonantal context. However, a major difficulty in extending this model to the present lax vowel data is that, in contrast to the findings for the tense vowel in Harrington et al. (2008), the perceptual boundaries between the fronting (*sit-soot*) and non-fronting (*will-wool*) contexts were not located significantly closer together for the younger listeners: that is, there was no conclusive evidence (only a strong tendency) that younger listeners process /ʊ/ independently of consonantal context to a greater extent than do older listeners.

The second explanation that was considered in Harrington et al. (2008) for tense /u/ was couched in terms of an exemplar model and the more frequent occurrence of variants in a fronting (e.g. *few*, *soon*) than in a non-fronting context (*boom*, *move*) in SSB. However, this model cannot be extended to the lax vowel data for the reasons stated earlier: there is not a similar bias in the distribution of fronted vs. non-fronted variants for SSB lax vowels.

A third possible explanation for the pattern of findings in this study, which has been suggested to us by one of the reviewers, Patrice Beddor, is that the difference between the variants in contexts like *soot* and *wool* has evolved into a categorical distinction for the younger listeners. This idea is consistent with the production data in this study showing a tight clustering of younger speakers' /ʊ/ variants in the contexts of *cook*, *hood*, and *soot* and their collective marked separation from /ʊ/ in the context of *wool*. Independently of these data, the possibility that there are phonologized i.e., categorical allophonic variants has been shown for vowel nasalization in voiceless contexts in certain varieties of American English by Solé (1992), Ohala (1993b), and Beddor (2009); similarly, there is electropalatographic data showing a categorical difference in German between the fricative allophones [ç] and [x] that occur after front and back vowels respectively as opposed to a continuous distribution of of /k/ in the same post-vocalic context (Ambrazaitis & John, 2004; see also Harrington, 2010, Chapter 7). Such an interpretation, while entirely plausible, would be strengthened if it could also be shown that the /ʊ/-variants in *wool* had not shifted in younger relative to older speakers. However, the production data suggests that /ʊ/ in *wool* has also been subject to fronting: thus in Figure 2, the F2 peak of *wool* is higher for the younger group in both male and female speakers while in Figure 1, the mean F2 position of /ʊ/ in *wool* is located at a similar frequency as the (back) vowel in *hard* for younger speakers, whereas for older speakers it is somewhat lower. We quantified this difference between the vowels in *wool* and *hard* with the same methodology for computing the relative distance between vowels based on a DCT-transformation of the entire F2 trajectory (Figure 4). The results of this quantification showed that the relative distance between the vowels in *wool* and *hard* was indeed significantly different for the age groups ($t_{29.4} = 2.4$, $p < 0.05$): that is, /ʊ/ was significantly more advanced relative to /ɑ/ of *hard* in younger than older speakers. Moreover, the *will-wool* perception boundary was unequivocally left-shifted (a greater number of /ʊ/ responses) in younger than in older listeners. It therefore seems probable that, whatever mechanism has brought about /ʊ/-fronting in *cook*, *hood*, and *soot* has also caused diachronic /ʊ/-fronting in *wool*, but of course to a much lesser degree in production. Consequently, while we certainly would not want to rule out this explanation of our data, there does not yet seem to be sufficient evidence to support the idea that these are categorically different variants (in *cook*, *hood*, *soot* on the one hand and *wool* on the other) for the younger group.

The fourth possible explanation is that /ʊ/ has fronted in SSB by analogy to the fronting of tense /u/. While there is certainly evidence that analogy contributes at least to the spread (if not the origin) of sound change (Kiparsky, 1995), we note that such an explanation would be idiosyncratic to SSB, given that in other varieties such as Australian English, tense /u/ is especially front and has fronted diachronically, whereas /ʊ/ has retained a (markedly) back vowel quality (Harrington, Cox, & Evans, 1997; Cox, 1999; Cox & Palethorpe, 2001). Nevertheless, given the evidence that the diachronic fronting of tense /u/ seems to have begun earlier than that of lax /ʊ/ in SSB (Hawkins & Midgley, 2005), then the possibility that /ʊ/ has fronted by analogy to its tense counterpart is plausible. Under this interpretation, the fronting of /ʊ/ has not been driven by synchronic coarticulatory relationships whose change would instead be a consequence of the more fronted position of /ʊ/. Thus, as /ʊ/ shifted diachronically further forward in the vowel space, the perturbation caused to it by an alveolar context is likely to have diminished (see Harrington, Hoole, Kleber, & Reubold, 2011 for a recent study of tongue-dorsum movement and auditory confusions in German vowels in which alveolar context was shown to have a far more pronounced influence on high back than on high front vowels). According to this model of fronting by analogy, younger speakers have a fronted /ʊ/ (which we will henceforth denote by /ʏ/ to distinguish it from the older speakers' more retracted /ʊ/) and for this reason their variation between fronting (*soot*) and most non-fronting (*cook*, *wool*) variants is much less than for older speakers. However, this model of fronting by analogy does not have an explanation for why /ʊ/-fronting in *wool* is marginal (the second finding outlined in the first paragraph of this section). Nor is it easy to explain in terms of fronting by analogy why the extent to which subjects compensate for coarticulation in perception in relation to the influence of context in production should be different for the two age groups (the third finding outlined above).

Consider then once again our proposed extension of Ohala's (1993a) model which incorporates the independent finding that diachronic lax /ʊ/-fronting began some time after the diachronic fronting of its tense counterpart. For the latter, Harrington et al. (2008) showed that the difference between the perceptual boundaries for younger subjects in the fronting (*yeast-used*) and non-fronting (*sweep-swoop*) contexts was less than for older subjects, as was their difference between the variants of /u/ in fronting (*used*, past-tense) and non-fronting (*swoop*) contexts. Thus in both groups the perception and production of coarticulation were matched but differently: for the younger subjects, the influence of context in perception and production was small, whereas for the older subjects it was large.

The question to consider now is how one (large effects of context in production and perception) might have evolved diachronically into the other (small effects of context in production and perception). The prediction from our extension to Ohala's (1993a) model is that if a waning of listener compensation for coarticulation is responsible for sound change in production, then the changes to the coarticulatory relationships in perception should precede those in production: that is, close to the beginning of a sound change, listeners should compensate less perceptually for the effects of context than the extent to which they separate variants in analogous contexts in their own production. The evidence for this greater misalignment between the perception and production of coarticulation for the younger subjects is shown in Figure 7: note in particular that, whereas in perception the *will-wool* boundary for younger subjects is considerably ahead of that of older subjects (the vertical axis of the panel of Figure 7), the age-dependent differences in the position of /ʊ/ of *wool* in production (the horizontal axis of the same figure) are much less.

Thus these lax vowel data represent perhaps an earlier stage in the progression of the sound change that may now be complete for the tense vowel: that is, the evolution towards the diminished coarticulatory influence of context in both perception and production seen in

Harrington et al. (2008) for tense /u/ passes through an intermediary stage - seen for the lax vowel data in this study - in which the influence of coarticulation in perception is less than it is in production. Thus our suggestion is that diachronic /ʊ/-fronting is still progressing and will reach a stage in which the *will-wool* boundary will shift closer to the *sit-soot* boundary in perception and in which the /ʊ/-variant in the non-fronting *wool* context is further fronted in production such that ultimately - as for the tense vowel - the influences of coarticulation in perception and production are both equally diminished.

Finally, as a number of experiments have shown, the perception and production of coarticulation are typically in alignment because listeners have been shown to compensate for (Fowler, 2005) and to be sensitive to (Martin & Bunnell, 1982) just those coarticulatory effects in speech perception that are found in speech production. It is possible that this type of alignment between the modalities is the norm and is characteristic of a synchronically stable system. The outcome of sound change, we would propose, is a shift from one such stable system to another. During a sound change in progress, the association between the perception and production of coarticulation passes through an unstable state during which the two modalities are out of alignment (in those typically younger subjects who participate in the sound change) and in which changes to the coarticulatory relationships in perception lead those in production.

REFERENCES

Ambrazaitis, G. & John, T. (2004) On the allophonic behaviour of German /x/ vs. /k/ - an EPG investigation. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK), 36*, 1-14.

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

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

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

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

Beddor, P. S., Krakow, R. A., & Lindemann, S. (2001). Patterns of perceptual compensation and their phonological consequences. In E. Hume and K. Johnson (Eds.), *The role of perceptual phenomena in phonology* (pp. 55-78). San Diego: Academic Press.

Choi, J.-D.& Keating, P. (1990). Vowel-to-vowel coarticulation in Slavic languages. *The Journal of the Acoustical Society of America, 88(S1)*, S54.

Clumeck, H. (1976). Patterns of soft palate movements in six languages. *Journal of Phonetics*, *4*, 337-351.

Cox, F. M. (1999). Vowel change in Australian English, *Phonetica*, *56*, 1-27.

Cox, F. M. & Palethorpe, S. (2001). The Changing Face of Australian English Vowels. In D.B. Blair and P. Collins (Eds.), *Varieties of English around the world: English in Australia* (pp. 17-44). Amsterdam: John Benjamins Publishing.

Draxler, C. & Jänsch, K. (2004). SpeechRecorder - A universal platform independent multichannel audio recording software. *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, *Lisbon, Portugal*, 559–562.

Fabricius, A. (2007). Vowel Formants and angle measurements in diachronic sociophonetic Studies: FOOT-fronting in RP. *Proceedings of the 16th International Congress of Phonetic Sciences*, *Saarbrücken, Germany*, 1477-1480.

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

Fowler, Carol A., Best, C. T., & McRoberts, G. W. (1990). Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception & Psychophysics*, *48*, 559–570.

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

Fowler, C., Brown, J., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language*, *49*, 396–413.

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

Guzik, K. & Harrington, J. (2007) The quantification of place of articulation assimilation in electropalatographic data using the similarity index (SI). *Advances in Speech–Language Pathology*, *9*, 109 – 119.

Harrington, J. (2006). An acoustic analysis of ‘happy-tensing’ in the Queen’s Christmas broadcasts. *Journal of Phonetics*, *34(4)*, 439–457.

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

Harrington, J. (2010). *Phonetic Analysis of Speech Corpora*. Chichester: Wiley-Blackwell.

Harrington, J., Cox F., & Evans, Z. (1997). An acoustic phonetic study of broad, general, and cultivated Australian English vowels. *Australian Journal of Linguistics*, *17*, 155-184.

Harrington, J., Hoole, P., Kleber, F., & Reubold, U. (2011). The physiological, acoustic, and perceptual basis of high back vowel fronting: evidence from German tense and lax vowels. *Journal of Phonetics*. http://dx.doi.org/10.1016/j.wocn.2010.12.006

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

Hawkins, S. & Midgley, J. (2005). Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association*, *35*, 183-199.

Henton, C. (1983). Changes in the vowels of received pronunciation. *Journal of Phonetics*, *11*, 353-371.

van den Heuvel, H., Cranen, B., & Rietveld, T. (1996). Speaker variability in the coarticulation of /a, i, u/. *Speech Communication*, *18*, 113-130.

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

Huffman, M. K. (1988). Timing of contextual nasalization in two languages, *UCLA Working Papers in Phonetics*, *69*, 68-76.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*, 485-499.

de Jong, G., McDougall, K., Hudson, T., & Nolan, F. (2007). The speaker discriminating power of sounds undergoing historical change: A formant-based study. *Proceedings of the 16th International Congress of Phonetic Sciences*, *Saarbrücken, Germany*, 1813–1816.

Keating, P. & Cohn, A. (1988). Cross-language effects of vowels on consonant onsets. *Journal of the Acoustical Society of America*, *84*, S84.

Kiparsky, K. (1995). The phonological basis of sound change. In J. Goldsmith (Ed.), *Handbook of Phonological Theory* (pp. 640-670). Oxford: Blackwell.

Krull, D. (1989). Consonant-vowel coarticulation in spontaneous speech and in reference words. *Speech Transmission Laboratory Quarterly Progress Status Report*, *1*, 101-105.

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

Magen, H. (1984). Vowel-to-vowel coarticulation in English and Japanese. *Journal of the Acoustical Society of America*, *75*, S41

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

Mann, V. A. (1986): Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English "l" and "r". *Cognition*, 24, 169–196.

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

Manuel, S. (1999). Cross-language studies: relating language-particular coarticulation patterns to other language-particular facts. In W. J. Hardcastle and N. Hewlett (Eds.), *Coarticulation: Theory, Data and Techniques* (pp. 179-198). Cambridge: Cambridge University Press.

Manuel, S. & Krakow, R. A. (1984): Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories Status Reports on Speech Research*, SR-77/78, 69–78.

Marin, S., Pouplier, M., & Harrington, J. (2010). Acoustic consequences of articulatory variability during productions of /t/ and /k/ and its implications for speech error research. *Journal of the Acoustical Society of America*, *127*, 445-461.

Martin, J. G. & Bunnell, 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. & Nolan, F. (2007). Discrimination of speakers using the formant dynamics of /uː/ in British English. *Proceedings of the 16th International Congress of Phonetic Sciences*, *Saarbrücken, Germany*, 1825–1828.

Milner, B.P. & Shao, X. (2006). Clean speech reconstruction from MFCC vectors and fundamental frequency using an integrated front-end. *Journal of Speech Communication*, *48(6)*, 697-715

Nossair, Z.B. & Zahorian, S.A. (1991). Dynamic Spectral Shape Features as Acoustic Correlates for Initial Stop Consonants. *Journal of the Acoustical Society of America* , *89(6)*, 2978-2991.

Oh, E. (2002). Fronting of back vowels in coronal contexts: a cross-language study, *Studies in Phonetics, Phonology and Morphology*, *8*, 239-254.

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

Ohala, J. (1993b) Coarticulation and phonology. *Language & Speech*, *36*, 155-170.

Ohala, J. (2005). Phonetic explanations for sound patterns. Implications for grammars of competence. In W. J. Hardcastle and J. M. Beck (Eds.), *A Figure of Speech. A Festschrift for John Laver* (pp. 23-38). London: Erlbaum.

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

Pierrehumbert, J. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay, and S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 177–228). Cambridge: MIT Press.

Pierrehumbert, J. (2006). The next toolkit. *Journal of Phonetics*, *34*, 516-530.

Recasens, D. (1987). An acoustic analysis of V-to-C and V-to-V coarticulatory effects in Catalan and Spanish VCV sequences. *Journal of Phonetics*, *15*, 299–312.

Recasens, D., Pallarès, M. D., & Fontdevila, J. (1998). An electropalatographic and acoustic study of temporal coarticulation for Catalan dark /l/ and German clear /l/. *Phonetica*, *55*, 53–79.

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

Smith, S., Tayman, J., & Swanson, D.A. (2001). *State and Local Population Projections: Methodology and Analysis*. New York: Kluwer Academic/Plenum Publishers.

Soli, S. (1981). Second formants in fricatives: acoustic consequences of fricative-vowel coarticulation. *Journal of the Acoustical Society of America*, *70*, 976-984.

Solé, M-J. (1992) Phonetic and phonological processes: the case of nasalization. Language & Speech, *35*, 29-43.

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America*, *88*, 97–100.

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

APPENDIX A

The discrete cosine transformation DCT breaks down any signal into a sequence of ½ cycle cosine waves which, if summed, reconstruct the original signal (Watson & Harrington, 1999; Harrington, 2006; Guzik & Harrington, 2007; Harrington et al., 2008; see also Nossair & Zahorian, 1991, Milner & Shao, 2006, and Harrington, 2010 Ch. 8, for the relationship between DCT and cepstral coefficients). The version of the DCT used here (sometimes known as DCT-II) is included as part of the Emu package in R (Harrington, 2010). For an N-point F2-trajectory in the present study, $x(n)$, extending in time from $n = 0$ to $N-1$ points, the m^{th} DCT coefficient, C_m, ($m = 0, 1, 2$) was calculated with:

$$C_m = \frac{2k_m}{N} \sum_{n=0}^{N-1} x(n) \cos\left(\frac{(2n+1)m\pi}{2N} \right) \qquad (5)$$

$$k_m = \frac{1}{\sqrt{2}},\ m = 0;\ k_m = 1,\ m \neq 0 \qquad (6)$$

Thus the F2 trajectory of each vowel as a function of time was represented by a single point in a four-dimensional space whose axes were formed from the amplitudes of the cosine waves at frequencies $k = 0, 0.5, 1$, and 1.5 cycles. Since these DCT coefficients are proportional to the trajectory's mean, linear slope, curvature (parabola), and the extent to which the trajectory follows the shape of a cubic function respectively (Guzik & Harrington, 2007), then this form of data reduction encodes a significant amount of dynamic information of the F2 trajectory's changing shape in time. If all of the ½ cycle cosine waves into which the signal is decomposed by the DCT are summed (an inverse discrete cosine transformation), then the signal is reconstructed exactly (after division by N, the length of the signal). If only the first few of these ½ cycle cosine waves are summed, then the result is a smoothed version of the original signal (and the more cosine waves that are summed, then the closer the approximation to the original signal). Figure 8 shows the raw signal and its DCT-smoothed equivalent after summing the cosine waves at frequencies $k = 0, 0.5, 1$, and 1.5 cycles for one of the vowels in the database.

Figure 8 about here

We obtained a measure of the goodness of fit of the raw and smoothed F2-trajectories by calculating *mape* (Smith, Tayman, & Swanson, 2001), the mean absolute percentage error (6) for every F2 vowel trajectory

$$mape = \frac{100}{N}\sum_{i=0}^{N-1}\frac{|s(n)-x(n)|}{x(n)}\% \qquad (7)$$

where *x(n)* is the an F2 trajectory of length *N* at time values 0, 1, 2...*N*-1, and *s(n)* the smoothed DCT-signal into which it was decomposed by summing the ½ cycle cosine waves at frequencies *k* = 0, 0.5, 1, and 1.5 cycles. The error as computed from (6) across the entire sample of vowels used in this analysis was between 0.09% and 19.9% with a mean error of 1.00%. Thus since DCT-smoothing using coefficients *k* = 0, 0.5, 1, and 1.5 cycles provided an approximation to the original signal to a mean accuracy of 1%, then the DCT coefficients, i.e., the amplitudes of the cosine waves that were used to compute the relative distances between vowels in this study, can be assumed to have encoded accurately the shape of the F2 trajectory as a function of time.

TABLES

Table I

Stimulus number and corresponding F2 values on the four 13-point continua.

Stimulus No.	F2 value (Hz)
1	1100
2	1164
3	1231
4	1301
5	1374
6	1450
7	1530
8	1613
9	1701
10	1793
11	1890
12	1992
13	2100

Table II

Transition durations and F2-loci for the three continua.

	Onglide		Offglide	
Continuum	**Duration**	**F2-locus**	**Duration**	**F2-locus**
/sVt/	40 ms	1700	50ms	1700
/wVl/	50 ms	1000 Hz	50 ms	1600 Hz

FIGURE TITLES

Figure 1: 95% confidence ellipses in the F2 × F1 plane for the vowels in *hid* (dotted), *hood* (dashed), and *hard* (solid) plotted on the same scale for the younger (above) and older (below) groups and shown separately for male (left) and female (right) speakers with the centroids marked as the corresponding phonetic symbol. Also shown are the group averages for the vowels in *soot* (s), *cook* (c), and *wool* (w).

Figure 2: Linearly time-normalized and averaged F2 trajectories between the acoustic vowel onset and offset in *soot*, *cook*, and *hood* and between the acoustic onset and offset of the word in *wool* plotted on the same scale for the younger (above) and older (below) groups and shown separately for male (left) and female (right) speakers.

Figure 3: Boxplots for the older (grey) and younger (white) speakers in four word contexts showing the Euclidean distance between all /ʊ/ vowels and the centroid of *hood* calculated over the F2 trajectories in a space formed from the first three DCT-coefficients separately for each speaker. Positive/negative values denote that the F2 mean for a given vowel token was higher/lower than that of the *hood* centroid.

Figure 4: Boxplots (left) of the logarithmic Euclidean distance between *soot* and *wool* derived from a DCT-paramterization of the second formant frequency for the older and younger speakers. A value of zero corresponds to a point equidistant between *soot* and *wool* and the boxplots include one mean value per speaker. The difference between *soot* and *wool* on this parameter is shown separately for the two age groups on the right.

Figure 5: Fitted response curves to *sit-soot* (dashed) and *will-wool* (solid) for younger (black) and older (grey) listeners showing the proportion of /ʊ/ responses as a function of F2 of the stimulus. The locations of the 50% cross-over boundaries are shown for each group by vertical lines. Both the horizontal axis and the interval between stimuli are proportional to the

Bark scale. The corresponding Hz scale is shown at the top of the figure. The points shown are the proportional responses of the raw data averaged by speaker group and word type.

Figure 6: The difference in Bark in the 50% cross-over boundaries between *sit-soot* and *will-wool* shown separately for the older (left) and younger (right) listeners.

Figure. 7: Left: The *y*-axis shows the position on a 13-point continuum (higher numbers correspond to a higher F2) of the 50% cross-over boundary for *sit-soot* (circles) and *will-wool* (squares) in older (solid) and younger (open) listeners. (There are two data points per speaker, one for *sit-soot* and the other for *will-wool*). The *x*-axis is acoustic data of the relative logarithmic Euclidean distance between *soot* and *wool* (the data points on the *x*-axis are the same as those that make up the boxplots in the left panel of Figure 4). Right: the gradient calculated from the data on the left by connecting the same pair of points per speaker.

Figure 8: An F2 trajectory for a production by a female speaker from the older group of [ʊ] of *soot* showing the raw values (circles) and the DCT-smoothed trajectory obtained by summing the cosine waves at *k* = 0, 0.5, 1.0, and 1.5 cycles.

Fig. 1

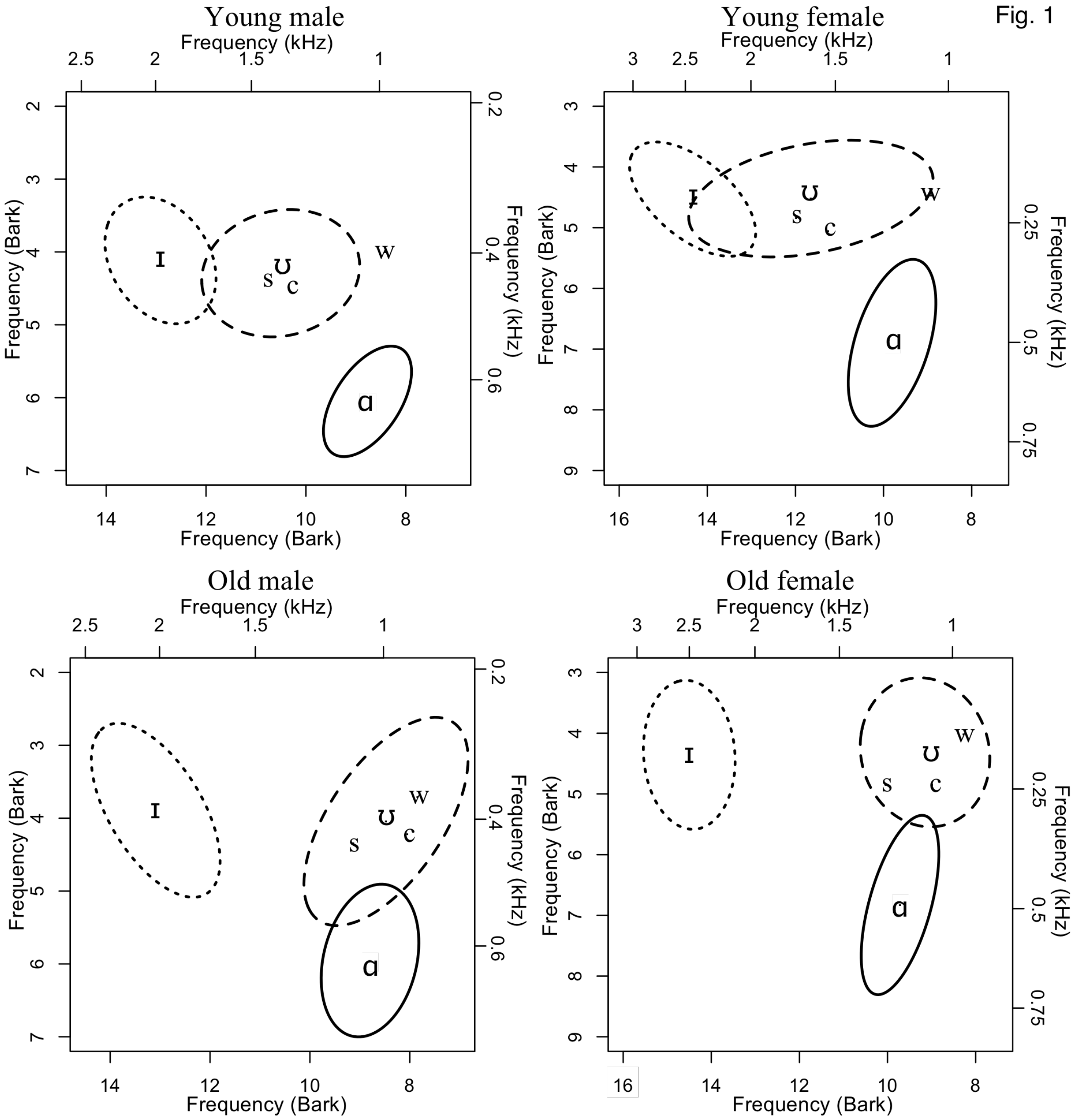

Young male
Frequency (kHz)
2.5
2
1.5
1
Frequency (Bark)
2
3
4
5
6
7
Frequency (kHz)
0.2
0.4
0.6
ɪ
s
ʊ
c
w
ɑ
14
12
10
8
Frequency (Bark)
Young female
Frequency (kHz)
3
2.5
2
1.5
1
Frequency (Bark)
3
4
5
6
7
8
9
Frequency (kHz)
0.25
0.5
0.75
ɪ
ʊ
s
c
w
ɑ
16
14
12
10
8
Frequency (Bark)
Old male
Frequency (kHz)
2.5
2
1.5
1
Frequency (Bark)
2
3
4
5
6
7
Frequency (kHz)
0.2
0.4
0.6
ɪ
w
ʊ
s
c
ɑ
14
12
10
8
Frequency (Bark)
Old female
Frequency (kHz)
3
2.5
2
1.5
1
Frequency (Bark)
3
4
5
6
7
8
9
Frequency (kHz)
0.25
0.5
0.75
ɪ
w
ʊ
s
c
ɑ
16
14
12
10
8
Frequency (Bark)

Fig. 2

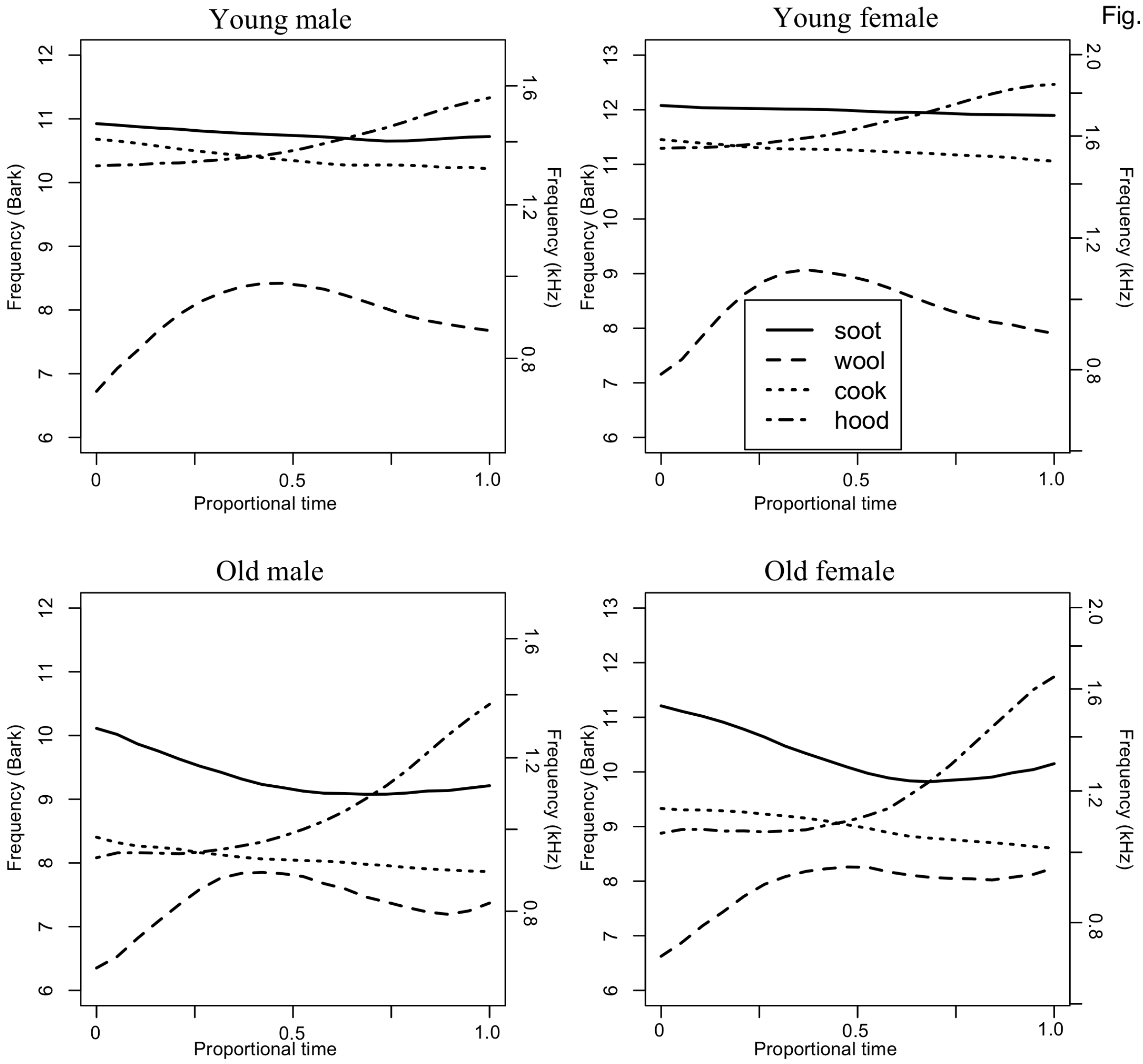

Young male
Young female
Old male
Old female
Frequency (Bark)
Frequency (kHz)
Proportional time
0
0.5
1.0
6
7
8
9
10
11
12
13
0.8
1.2
1.6
2.0
soot
wool
cook
hood

Fig. 3

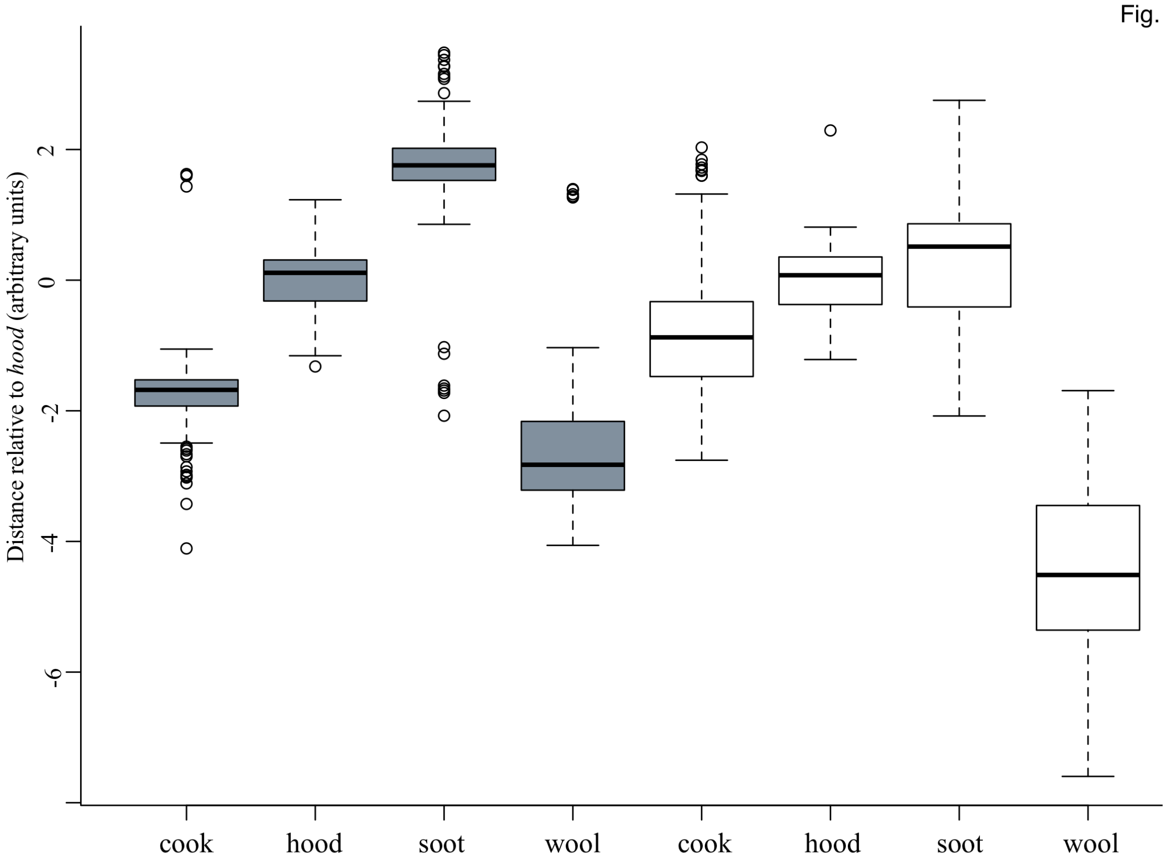

Distance relative to *hood* (arbitrary units)
2
0
-2
-4
-6
cook
hood
soot
wool
cook
hood
soot
wool

Fig. 4

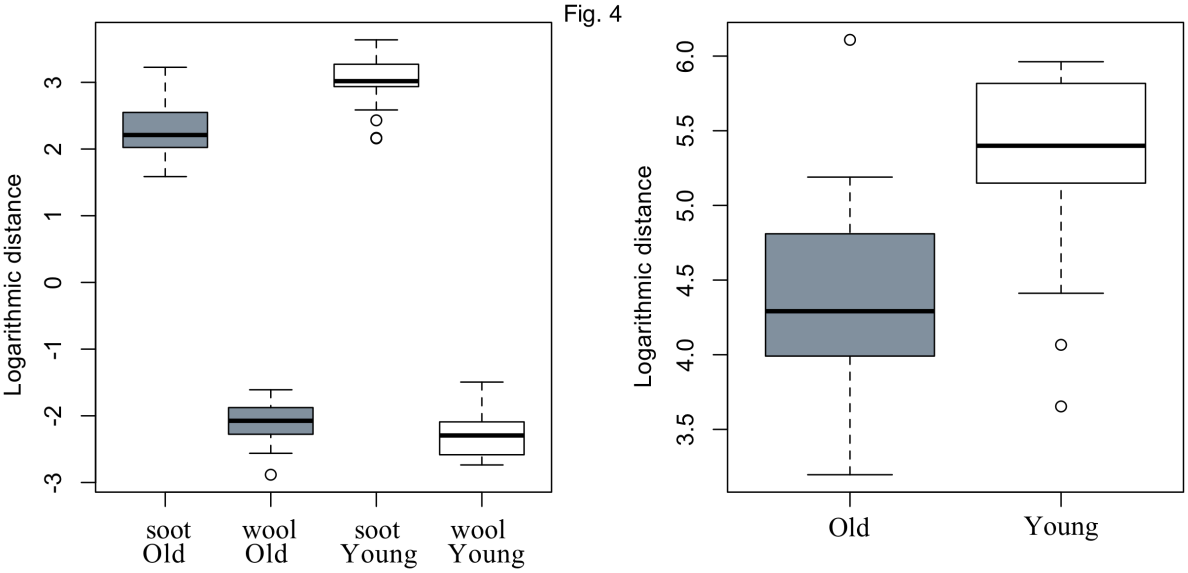

Logarithmic distance
3
2
1
0
-1
-2
-3
soot Old
wool Old
soot Young
wool Young
Logarithmic distance
6.0
5.5
5.0
4.5
4.0
3.5
Old
Young

Fig. 5

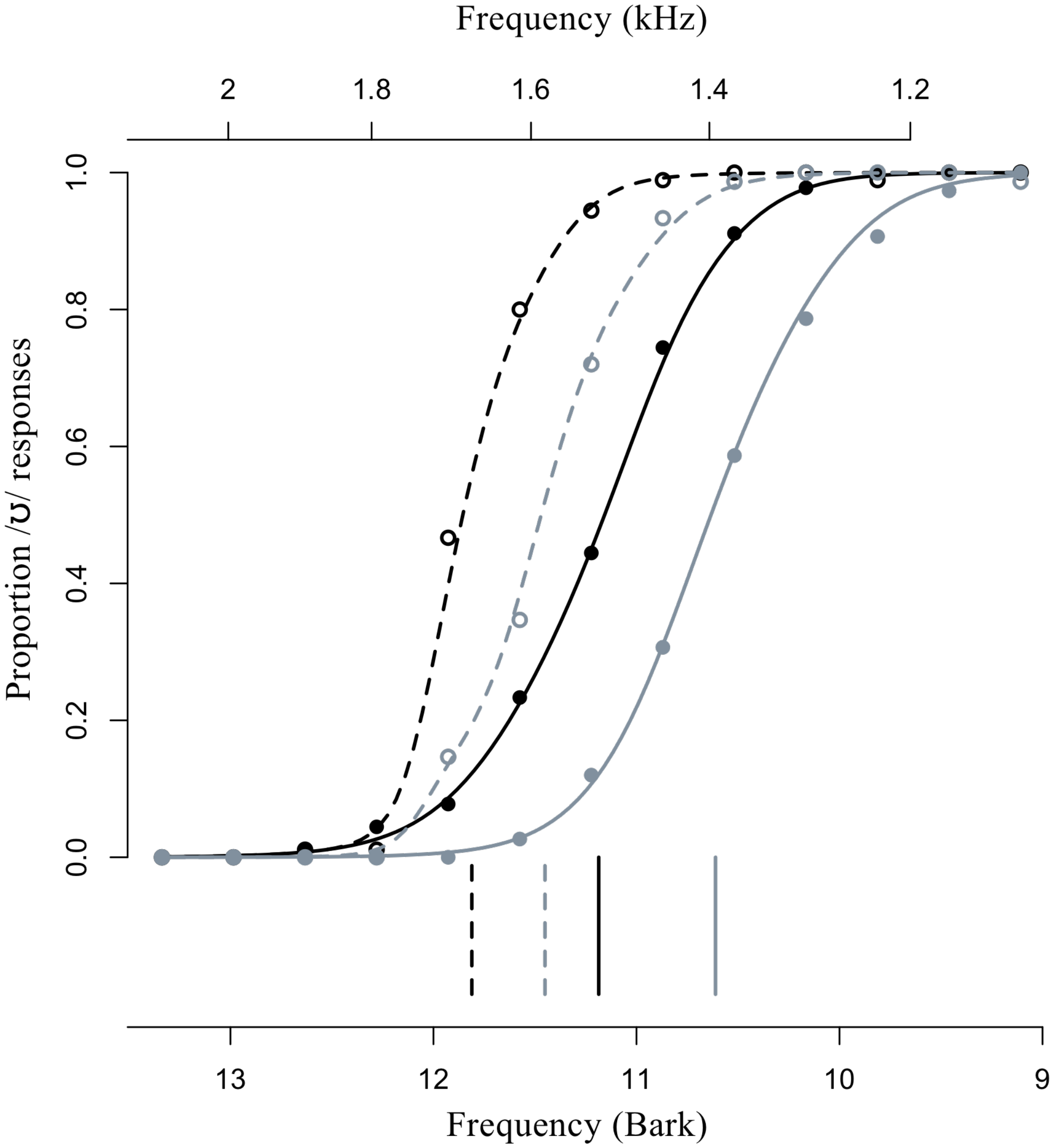

Frequency (kHz)
2
1.8
1.6
1.4
1.2
Proportion /ʊ/ responses
0.0
0.2
0.4
0.6
0.8
1.0
13
12
11
10
9
Frequency (Bark)

Fig. 6

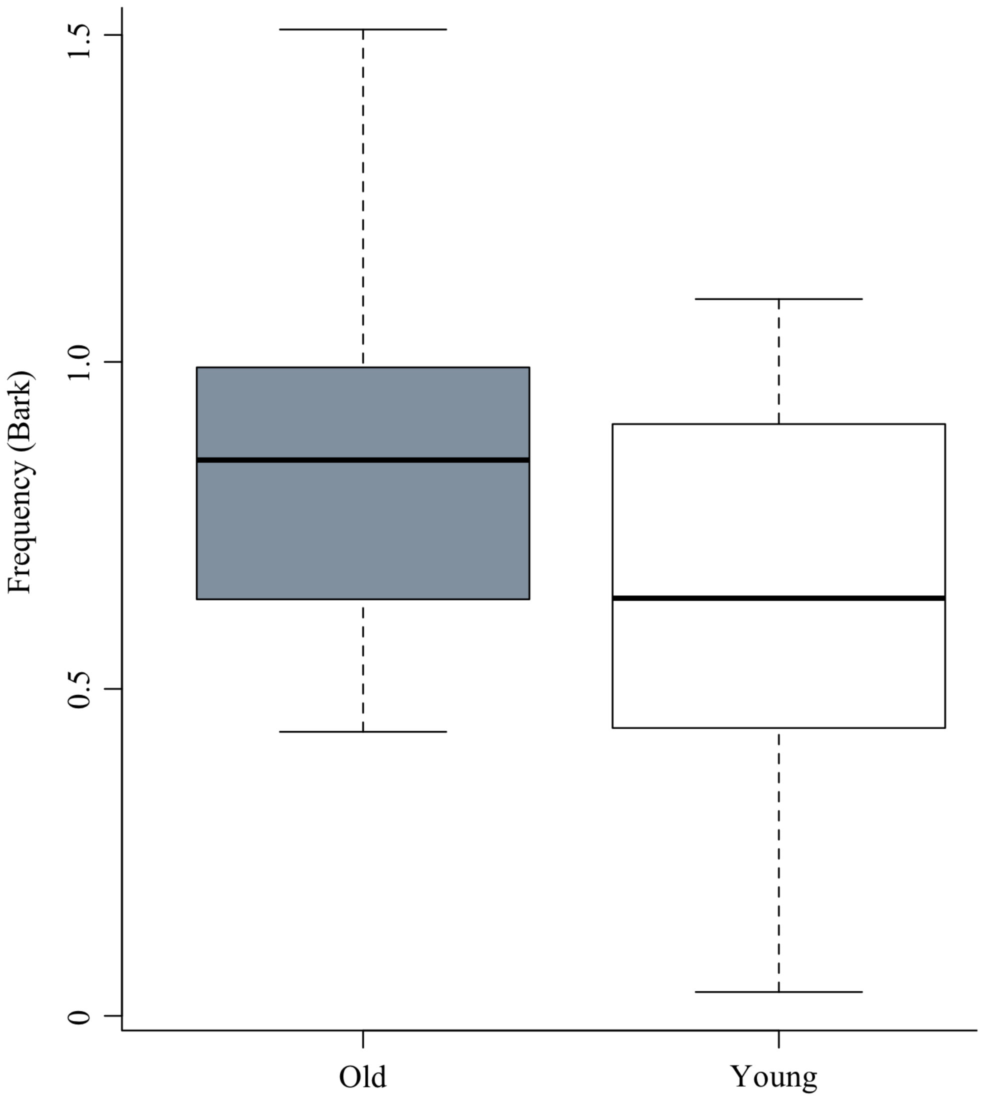


Fig. 7

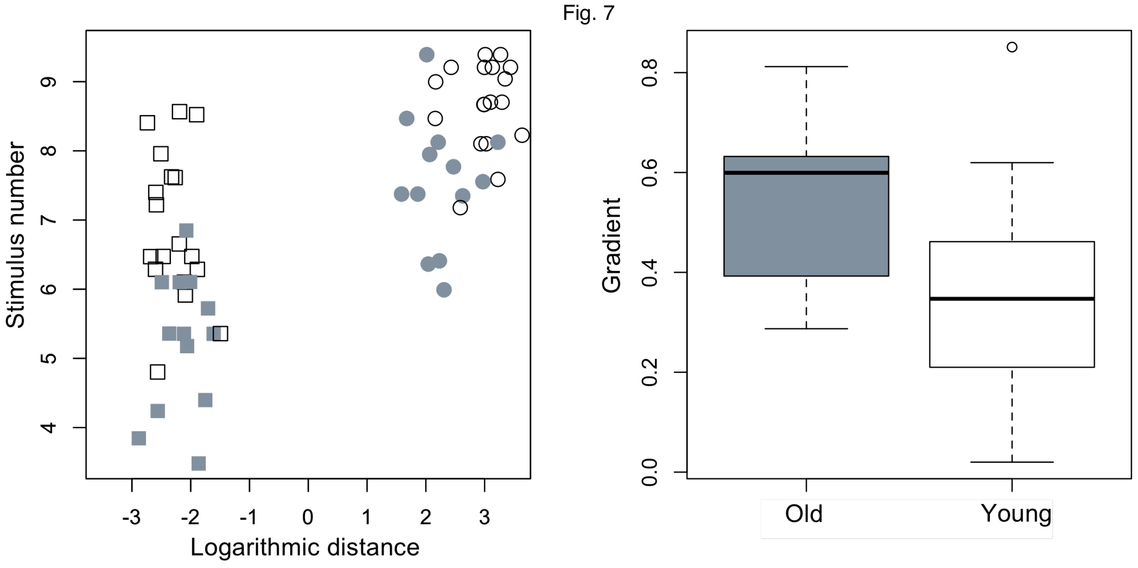
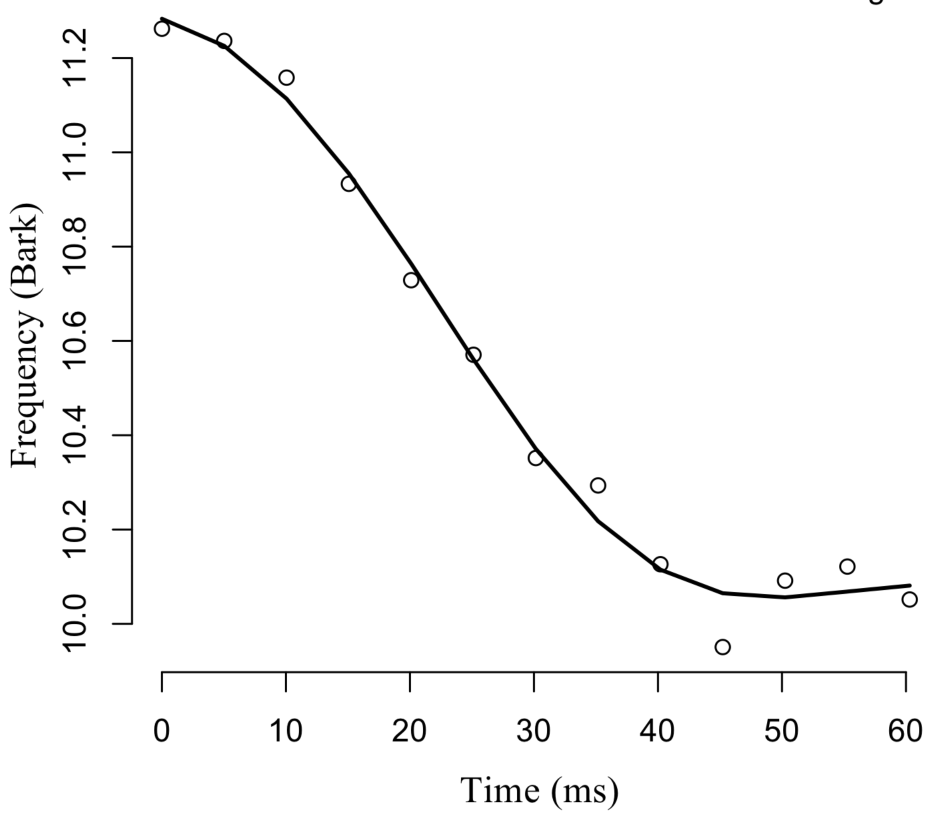

Stimulus number
4
5
6
7
8
9
-3
-2
-1
0
1
2
3
Logarithmic distance
Gradient
0.0
0.2
0.4
0.6
0.8
Old
Young

Fig. 8


Frequency (Bark)
11.2
11.0
10.8
10.6
10.4
10.2
10.0
0
10
20
30
40
50
60
Time (ms)