

The contributions of the lips and the tongue to the diachronic fronting of high back vowels in Standard Southern British English

Jonathan Harrington, Felicitas Kleber & Ulrich Reubold

Institute of Phonetics and Speech Processing,
Ludwig-Maximilians-Universität München, Germany
jmh@phonetik.uni-muenchen.de
kleber@phonetik.uni-muenchen.de
reubold@phonetik.uni-muenchen.de

Recent acoustic studies have provided evidence that /u/ (GOOSE) and /ʊ/ (FOOT) have fronted in the standard accent of England in the last fifty years, but what is less clear is whether this fronting is due entirely to a repositioning of the tongue or whether it has been accompanied by an unrounding of the lips. Four experiments were carried out to shed light on this issue. An acoustic study of anticipatory coarticulation in /s/ in the first of these suggested a similar degree of lip-protrusion for young speakers whose F2 of /u/ was raised compared with that of older speakers. Compatibly, judgments of lip-rounding elicited from cross-dubbed auditory-visual stimuli and an analysis of lip movement showed young speakers' /u/ to be produced with rounded lips. Their tongue positions and movements in the final experiment were found to be almost as advanced for /u/ as for /i/ (FLEECE) and nearer to a central position for lax /ʊ/ (FOOT). Taken together, these results confirm firstly, that the diachronic shift in /u/ has involved a realignment of the tongue, but not of the lips; and secondly, that the diachronic shift in /ʊ/ is likely to be a more recent innovation than that of its tense counterpart.

1 Introduction

The aim of the present study is to contribute to our understanding of the phonetic basis of diachronic back vowel fronting that has been documented in Australian (Cox 1999, Cox & Palethorpe 2001), American (Labov, Ash & Boberg 2006, Fridland 2008), Southern British (Henton 1983, Hawkins & Midgley 2005, Fabricius 2007, Harrington 2007, Harrington, Kleber & Reubold 2008, McDougall & Nolan 2007) and New Zealand (Gordon et al. 2004) varieties of English in recent years. The idea that a phonetic explanation seems plausible is consistent with the evidence that diachronic back vowel fronting has occurred in many (often unrelated) languages which has led Labov (1994) to incorporate this historical change as one of the major forces acting upon vowel chain shifts.

Harrington et al. (2008) proposed a phonetic basis for diachronic /u/-fronting in the standard accent of England, Standard Southern British English (SSBE), and more specifically

that it could be related to coarticulation-induced, synchronic /u/-fronting.¹ As Lindblom (1963) and Stevens & House (1963) have shown, the tongue-body of /u/ is pulled forward synchronically in consonantal fronting contexts such as /tut/ and there is also evidence that these coarticulatory effects are compensated for perceptually (Lindblom & Studdert-Kennedy 1967, Ohala 1981, Ohala & Feder 1994). In Harrington et al.'s (2008) apparent-time study, the vowel formants of young and old speakers were found to differ only marginally when /u/ followed consonants that induce fronting in words such as *soup* and *used* (past tense), whereas the age-group differences in /u/ were far more marked in the context of a preceding neutral (e.g. *who'd*) or backing (e.g. *swoop*) consonant. Based on this evidence, they reasoned that diachronic /u/-fronting in SSBE was context-sensitive involving primarily a shift of the neutral and back allophones of /u/ to the front. They also showed that young listeners compensated much less for the coarticulatory effects of context on /u/ than did older listeners. Thus, whereas old speakers had both distributed allophones of /u/ in production and compensated for these context effects in perception, young speakers' /u/ variants were closer together in production and they compensated much less for contextual influences on /u/ perceptually. Harrington et al. (2008) argued that these findings were compatible with an extension of Ohala's (1981, 1993) model in which compensating for the coarticulatory fronting effects of a preceding consonant had waned in SSBE in recent years. Alternatively, their results were also shown to be compatible with an episodic model of speech (Pierrehumbert 2003) in which diachronic /u/-fronting was related to the high statistical frequency reported in Harrington (2007) with which /u/ follows alveolar consonants (e.g. *noon, soon*) or /j/ (*dune, few, view*), i.e. consonants that induce synchronic fronting of /u/ in SSBE.

A difficulty with these arguments in which coarticulatory pressures acting on the tongue are advanced as an explanation for diachronic back vowel fronting is that they are inferred from acoustic changes and in particular from group differences in the second formant frequency. However, as is well known, F2-raising in back vowels can be brought about not only by tongue-fronting but also by lip-unrounding in which formants are raised due largely to the shorter vocal tract length (Lindblom & Sundberg 1971). Thus, an unresolved issue – and one which is also the main concern of the present study – is whether the F2-raising that is taken as evidence for back vowel fronting in many apparent-time studies really is brought about by a repositioning of the tongue, or whether there are also changes to lip-rounding. At least for SSBE, Wells (1997) raises the latter as a definite possibility based on his auditory impressions. He comments:

Traditionally classified as back and rounded, these vowels [tense /u/ and lax /ʊ/] are not only losing their lip-rounding but also ceasing to be very back. Thus *spoon*, conservatively [spu:n], may now range to a loosely rounded [spʊʌn] or even [spɪɪn], while *good* /gʊd/ is often pronounced with a schwa-like quality.

The idea that lax /ʊ/ may have become unrounded is also suggested by Fabricius (2007). On the other hand, a recent study by Majors & Gordon (2008), using a video analysis of the lips of two speakers producing a variety of American English that has been shown to be affected by the Northern Cities Shift and diachronic back vowel fronting, showed no evidence for an association between lip-unrounding and vowel fronting in /u/. However, as the authors acknowledge, their video technique may be insufficiently precise to quantify the dynamics of lip-protrusion. In any case, their findings would be neutral with regard to Wells' (1997)

¹ The variety of English in this study, Standard Southern British English, is non-rhotic. The vowels analysed in this variety include those from /hVd/ words *heed, hid, who'd, hood, hoard, and hard*, for which (following the usual notational conventions for SSBE) the transcriptions /i ɪ u ʊ ɔ ɒ ɑ/, respectively, will be used. According to traditional accounts, /u ʊ ɔ ɒ/ are rounded in this variety, the others unrounded. /i u/ are high, /ɪ ʊ ɔ/ are mid-high, and /ɒ ɑ/ are low. /i ɪ/ are front, /ɔ ɒ ɑ/ are back, and /u ʊ/ are between back and central. /i ʊ/ are phonologically lax and phonetically short, the others phonologically tense and phonetically long; /i u/ are more peripheral than /ɪ ʊ/.

suggestions concerning lip-unrounding in /u/ in a different variety, Standard Southern British English.

A second motivation for the present study is to find out more about the extent of tongue-fronting in Standard Southern British English /u ʊ/. If these present-day SSBE vowels really are produced with a degree of lip-unrounding, then the tongue may not have advanced very much beyond a central-back position. On the other hand, it could be that lip-rounding has been fully maintained and that the extent of tongue-fronting in present-day SSBE /u/ has been somewhat underestimated. One of the goals of the present study is to carry out a physiological investigation to disentangle this confound between these articulators and more specifically to begin to resolve the question of the relative importance of lip-rounding and tongue-fronting for the distinction between /u/ and /i/ and between their lax counterparts /ʊ/ and /ɪ/ in this variety.

A third goal is to shed light on the extent of tongue-fronting in tense /u/ compared with lax /ʊ/ in order to establish how their patterns of diachronic change are related to each other. A study by Hawkins & Midgley (2005) suggests that /ʊ/-fronting may be a sound change that started some time after diachronic /u/-fronting. Based on auditory impressions at least (Cruttenden 1994), tense and lax /u ʊ/ have been judged to be similar in phonetic backness at some stage in the last 50 years: prior to or at an early stage of this sound change, both vowels were presumed to occupy about the same position between central and back. Consequently, if diachronic fronting of tense /u/ has preceded that of lax /ʊ/, then the tongue position for /u/ should be phonetically advanced even in non-fronting contexts relative to /ʊ/ in young SSBE speakers, assuming that /u/ and /ʊ/ have fronted diachronically at the same rate.

These goals were tested using three different kinds of materials and corpora described in further detail in four experiments. For the first (Section 2), comparisons were made on the spectral centre of gravity of /s/ before unrounded (*seep*) and rounded (*soup*) vowels in two age groups of SSBE speakers. The hypothesis to be tested here is that if lip-rounding has diminished in /u/ diachronically, then the /s/ in *seep* and *soup* should be acoustically closer together on this measure for younger than older speakers. The test of lip-rounding was extended in a subsequent study in which first-language German speakers classified /u/ from video signals alone or with an audio signal of SSBE /i/ dubbed onto it. If young SSBE speakers produce /u/ with lip-rounding, then they should be classified preferentially as a rounded vowel by first-language German speakers under either of these two experimental conditions. For the final two experiments, movement data using electromagnetic articulometry (EMA) were obtained from young SSBE speakers: these data were analysed in order to determine the extent of lip-rounding and the positions of three points on the surface of the tongue for /u/ and /ʊ/ in relation to some other SSBE vowels.

2 Anticipatory lip-rounding in /s/

2.1 Method

The speakers and materials were taken from the same database that had formed part of the acoustic study of /u/-fronting in Harrington et al. (2008), recorded in 2007. Speech production data were available from 27 speakers including 14 (11 female, 3 male) from a young group (aged between 18 and 20 years) and 17 (7 female, 10 male) from an old group (over the age of 50 years). Recordings were obtained using a Sennheiser stereo headset pc165 USB and a Toshiba Tecra notebook in the Phonetics Laboratory of the University of Cambridge from the majority of speakers; some of the older subjects were recorded in a quiet room at their homes. There were no apparent differences in the quality of the recorded speech signal and no differences in the subjects' performance related to the testing location.

Subjects produced a number of isolated monosyllabic words with different vowel nuclei to cover most of the SSBE vowel space. The words were displayed individually on a notebook

computer screen and recordings were made with the SpeechRecorder software that is routinely used at the University of Munich for speech recording outside the laboratory (Draxler & Jänsch 2004). A total of 540 words were produced per speaker in this way from a randomized list of 10 repetitions of 54 words. Any words that were mispronounced were excluded. The words that were analysed here include only a subset of these: specifically the productions of *seep* and *soup* (27 speakers \times 2 words \times 10 repetitions = 540 tokens).

The words were digitized at 44.1 kHz and 2048 point spectra (spectral frequency interval of 21.5 Hz) were calculated using the Emu speech database system (Harrington 2010) every 5 ms. Acoustic boundaries of /s/ in *seep* and *soup* were marked at the acoustic onset of fricative energy and at the periodic onset of the following vowel. For the present investigation, spectra were extracted at the temporal midpoint in /s/ between these two boundaries. The spectra were then parameterized firstly by converting the frequency axis in Hz to the Bark scale and then by obtaining cepstrally smoothed spectra based on summing the first 10 coefficients after the application of the discrete cosine transformation to these Bark scaled spectra (Watson & Harrington 1999). Finally, a procedure was applied to each such spectrum in turn to obtain the frequency between 2500 Hz and 10000 Hz at which a spectral peak occurred. Since the vocal tract tends to be lengthened in /s/ due to anticipatory lip-rounding before a rounded vowel, then the spectrum should be shifted down the frequency axis compared with that of /s/ before a vowel with a spread lip position (Lindblom & Sundberg 1971). Consequently, the frequency at which the spectral peak occurred was expected to be lower for /s/ in *soup* than in *seep*. The argument for attributing the lowering in frequency of the spectral peaks in /s/ to lip-rounding is based on two kinds of experimental evidence: firstly, perceptual studies showing that listeners factor from /s/ the proportion of spectral centre of gravity lowering that they attribute to anticipatory coarticulatory lip-rounding induced by a following rounding vowel (Mann & Repp 1980); and secondly, the evidence from articulatory-to-acoustic modeling (Fant 1960) which not only predicts such an effect but also shows that the contribution of the cavity behind the constriction in /s/ makes a negligible contribution to the spectrum (as a result of which spectral centre of gravity lowering is unlikely to be due to increased tongue-body raising and backing due to secondary velarization in /s/, for example). Thus, the hypothesis to be tested was that the difference between the locations of these peak frequencies in these rounded and unrounded contexts should be less for the young than for the old, if they produced *soup* with a lesser degree of lip-protrusion than did the older group.

2.2 Results

The acoustic difference between the two phonetic variants of /s/ is apparent in the ensemble-averaged spectra shown separately for the two age groups in Figure 1. Thus, for both male and female speakers and in both age groups, the peak frequency for /s/ above 3000 Hz was some 1–2 kHz lower for *soup* than *seep*. Indeed, there is evidence that the entire spectrum was shifted to the left in *soup*, consistent with the interpretation that it was produced with a longer vocal tract due to lip-protrusion.

At the same time, Figure 1 shows no evidence that the difference in the spectra between the two words was less for the young than for the old, as it might have been, if the young had produced /u/ with less protruded lips. The further quantification of the data in Figure 2, showing distributions of the frequency at which the spectra attained a peak, is consistent with this interpretation: firstly, there is clear evidence that these maxima were between 0.5 and 1 Bark lower for *soup* than *seep*; and secondly, that the difference between *seep* and *soup* on this parameter was about the same for both age groups. The frequency locations across both words were also evidently higher for the young compared with the old. In an acoustic study comparing male and female speakers in two age groups (21–30 years and 61–73 years), Schötz (2003: 2587) also found that ‘the typical energy platform on higher frequencies began around 4 kHz for younger-sounding speakers, but already between 3.5–4 kHz for older-sounding speakers’. Thus, it seems possible that this overall lower centre of gravity

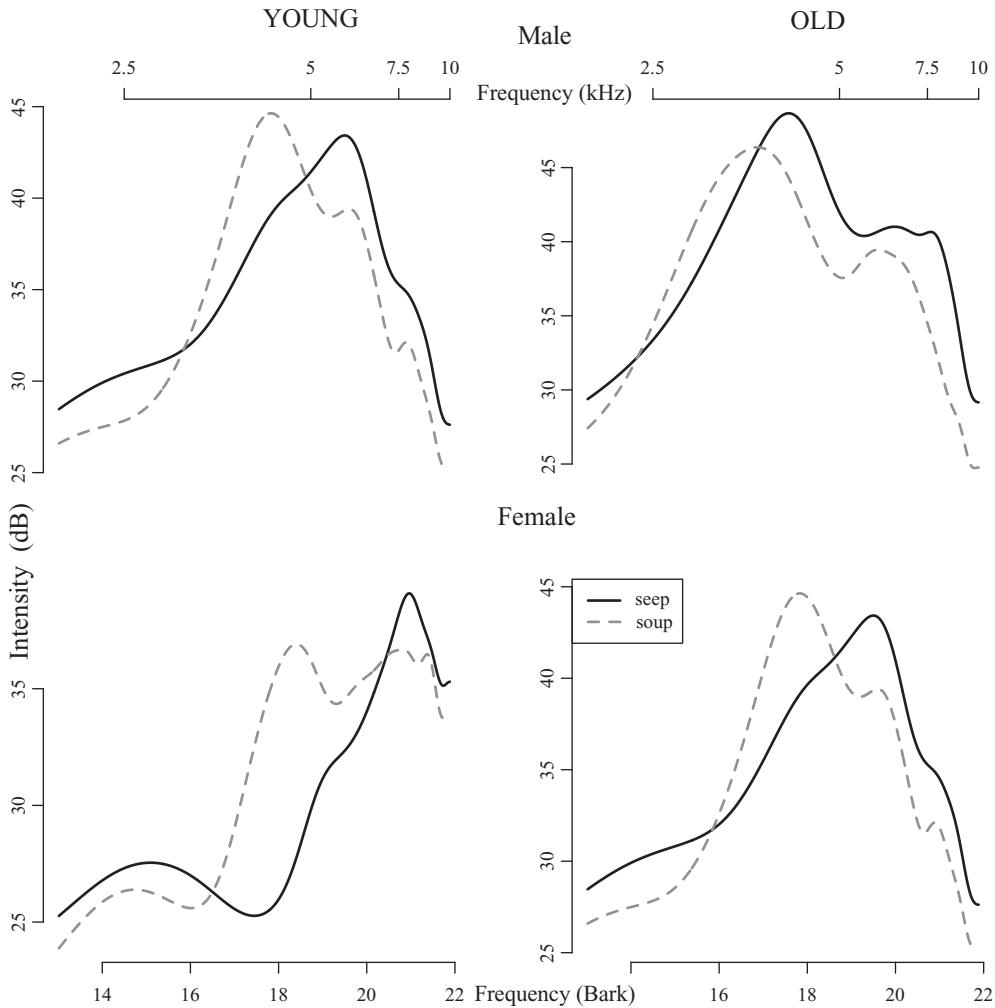


Figure 1 Ensemble-averaged spectra at the temporal midpoint of /s/ in *seep* (solid) and *soup* (grey, dashed) for the young (left) and old (right) groups of speakers and shown separately for male (above) and female (below) speakers.

of the spectrum in older compared with younger speakers may be a non-phonetic effect due to age-related changes in the morphology of the vocal tract. For example, Xue & Hao (2003) have shown that the volume and length of the oral cavity are greater in older than in younger speakers. However, in the absence of any age-related physiological studies of fricatives, it is currently difficult to assess whether such differences can provide a sufficient explanation for the lowering of the fricative's centre of gravity.

The results of a repeated measures MANOVA with the parameter in Figure 2 as the dependent variable and Age (two levels: young, old), Word (two levels: *seep*, *soup*), and Gender as fixed factors showed significant effects for Age ($F(1,23) = 10.7$, $p < .01$), Gender ($F(1,23) = 35.1$, $p < .001$), and Word ($F(1,23) = 39.1$, $p < .001$). These results show that *seep* and *soup* differed from each other on this parameter (the Word effect), that young speakers had higher locations of spectral peaks than did old speakers (possibly for non-phonetic, physiological changes due to age as suggested above) and that there were differences on this

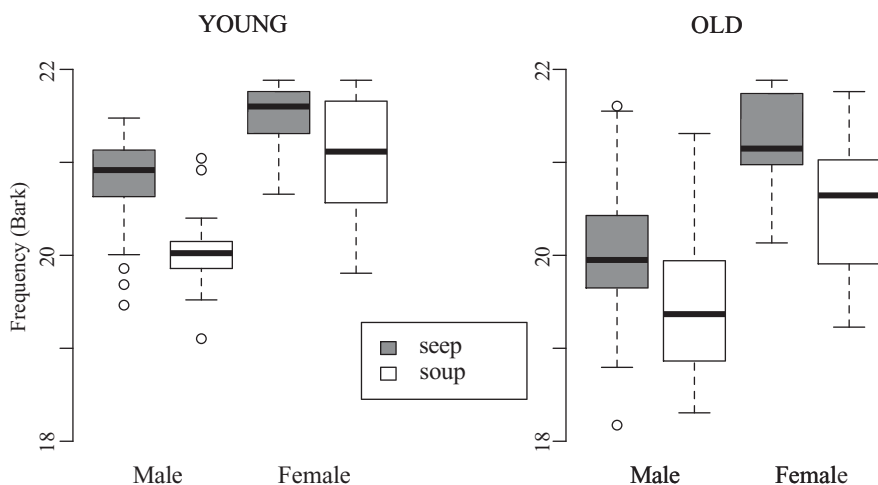


Figure 2 Distribution of the frequency in the spectrum at which the peak frequency occurs in *seep* (grey) and *soup* (white) for young (left) and old (right) speakers. The spectral data were obtained at the temporal midpoint of /s/.

parameter between male and female speakers (female speakers have shorter vocal tracts and consequently the major peak in the /s/ spectra are shifted to the right on the frequency axis compared with those of male speakers). But importantly, there was no significant Age \times Word interaction ($F(1,23) = 0.5$, NS), which confirms the evidence in Figures 1 and 2 that the parameterized spectral differences between *seep* and *soup* were about the same for the two age groups on this parameter. (No other interactions were significant.)

3 Audio-visual perception

The most plausible interpretation of these results together with the evidence in Figures 1 and 2 is firstly, that /s/ was produced with more rounded lips in *soup* than in *seep*; but secondly, that the extent of rounding was about the same in young as in old speakers. The following audio-visual study was designed to test the extent to which the lip-rounding differences inferred from the preceding acoustic analysis could be detected. For this purpose, and also in order to begin to assess the degree of fronting of /u/ (Section 4), first-language German listeners made judgments of rounding of auditory and visual stimuli from one of the young speakers who had participated in the preceding experiment. German has a three-way phonemic contrast in high vowels /i y u/ (high front unrounded, high front rounded, high back rounded, as in German *Biene*, *Bühne*, *Bühne*) in the tense vowel series; in addition, German /o/ (mid-high, back, rounded, as in *Bohne*) was presented as a classificatory choice. Phonetically, these four German vowels are monophthongal and peripheral and, based on our impressionistic auditory judgments, very close in quality to the corresponding cardinal vowels (CV1, CV9, CV8, CV7, respectively).

The hypothesis to be tested was whether or not lip-rounding in SSBE /u/ was detectable: if so, then German listeners could be expected to classify SSBE /u/ as a rounded vowel when presented with video images of lip movement in the absence of an auditory signal. The second test of the strength of the lip-rounding cue was inferred from cross-spliced stimuli in which an audio recording of SSBE /i/ was grafted onto a video-signal of SSBE /u/. If /u/ is unrounded, then this combined signal should be perceived as /i/; but if on the other hand /u/ is rounded, then listener judgments should be more equivocal in perceiving the signal as an unrounded or rounded vowel.

The motivation for eliciting judgments from first-language German speakers was partly a pragmatic one (availability of subjects) but also to provide some preliminary data for the subsequent physiology experiment concerned with the degree of tongue-fronting in SSBE /u/ (Section 5 below): that is, if present-day SSBE /u/ is phonetically closer to front than back, and if SSBE /u/ is produced with lip-rounding, the German speakers could be expected to classify either of the stimuli described above as /y/ rather than /u/ or /o/.

3.1 Method

3.1.1 Preparation of stimuli

The stimuli for this experiment were obtained from one of the young female speakers of SSBE who had participated in the preceding experiment (she was therefore also one of the speakers in Harrington et al. 2008) and who was 21 years of age at the time of recording the materials for the present experiment. The main criterion for selecting this speaker was that she had produced the phonetically most fronted productions of /u/, as judged by the acoustic analysis of the second formant frequency in Harrington et al. (2008). Video recordings were made in a sound-treated room using a camera at a distance of 1.5 m from the speaker. The zoom was adjusted so that the head filled the display. The speaker's lips were additionally highlighted with a blue marker. The speaker produced eight repetitions of six isolated /hVd/ words (*heed*, *hid*, *hod*, *hoard*, *hood*, *who'd*; see footnote 1 above). The words were randomized and presented individually to the speaker on a computer screen using the same software (Draxler & Jänsch 2004) as in the previous experiment. The synchronized video and audio signals were captured at a rate of 25 frames per second and 44.1 kHz, respectively, using a Sony NP-F330 digital video camera with a Neumann TLM 103 condenser microphone.

Two *heed*, three *who'd* and one *hoard* token were selected for further analysis and presentation to the German subjects. These included the two *heed* tokens with the highest (2880 Hz) and the lowest (2700 Hz) second formant frequency; three *who'd* tokens with the highest (2250 Hz), lowest (2050 Hz) and intermediate (2150 Hz) F2; and one *hoard* token whose F2 was closest to the mean F2 for that vowel (1060 Hz). All of these tokens were scaled to have the same mean dB level.

Audio-visual mismatched stimuli were created by grafting the audio signals from the two *heed* tokens onto the video signals of three *who'd* tokens and onto the video signal of the single *hoard* token. Thus, there were six stimuli with an audio signal for *heed* and synchronized video signal for *who'd* (henceforth, 6 $i_{\text{aud}}u_{\text{vid}}$ stimuli) and two stimuli with an audio signal for *heed* and a synchronized video signal for *hoard* (2 $i_{\text{aud}}o_{\text{vid}}$ stimuli). The mismatched stimuli were created using Adobe Premiere Pro 2.0 by synchronizing the video and audio signals relative to the acoustic onset of [h] (see Traunmüller & Öhrström 2007 for a similar procedure).

3.1.2 Experimental conditions

There were two experimental conditions: video-only (VO) and audio-visual (AV). For the VO experiment, the audio signal was removed from the six original words. (Henceforth, i_{vid} , u_{vid} and o_{vid} denote video-only signals from the two *heed*, three *who'd* and single *hoard* tokens, respectively.) For the AV experiment, there were eight mismatched stimuli (6 $i_{\text{aud}}u_{\text{vid}}$ + 2 $i_{\text{aud}}o_{\text{vid}}$) and two matched stimuli ($i_{\text{aud}}i_{\text{vid}}$), the latter being the original *heed* tokens with the audio signal intact. Thus, for all eight stimuli in the AV condition, the audio signal was from *heed*.

For both experimental conditions, the stimuli were repeated six times (thus 36 stimuli for the VO condition and 48 stimuli for the AV condition) and randomized together with a number of other VO or AV stimuli that do not form part of the analysis in the present experiment.

3.1.3 Presentation of stimuli

In both VO and AV conditions, the video signal of the speaker was presented in a 4:3 aspect ratio from a 17" color monitor. For the AV condition, the synchronized audio signal was

additionally presented via two loudspeakers situated near the video screen and at a distance of 75 cm from the subject. Two further loudspeakers were positioned at right angles to the subject and also at 75 cm from the subject: from these, noise was presented at a signal-to-noise ratio of -15 dB and synchronized with the speech signal from the other two loudspeakers. We included noise in this way because prior pilot experiments had shown that subjects' judgments would otherwise be swayed to too great an extent by the audio signal (i.e. subjects tended to ignore the information from the video signal at full presentation volume of the acoustic speech signal).

3.1.4 Subjects and procedure

Sixteen first-language German speakers with no known speech or hearing disorders and with no phonetic training recruited from around Munich participated in both VO and AV experiments. The subjects all produced a standard variety of German in which rounding was contrastive in the front vowels. The subjects were not told anything about the nature of the speaker – so the fact that the speaker was not a first-language speaker of German was not known to them. The subjects' task upon presentation of a stimulus was to depress one of four computer keys (the numbers 1–4) that had been overlaid with the letters I, Ü, U, O (corresponding to the German phonemes /i y u o/) that they believed to be the closest match to the stimulus. (We did not give any real word examples corresponding to these four choices since the relationship between these graphemes and phonemes is quite clear in most cases.) There was no time pressure in producing a response. The software Eprime was used for presentation of the stimuli and for recording the answers.

3.2 Results

The distribution of subjects' classifications of the video-only stimuli in Figure 3 shows unequivocally that u_{vid} patterns with \varnothing_{vid} and not with i_{vid} : that is, 6% of i_{vid} as opposed to 97.5% of u_{vid} and 100% of \varnothing_{vid} were classified as one of the three rounded vowel choices available to the German participants. This experiment shows that this SSBE speaker's /i/ and /u/ were distinguished on facial and presumably lip gestures to about the same extent as were her /i/ and /ɔ/. A further inspection of Figure 3 shows that there was some further difference in how the German subjects classified u_{vid} and \varnothing_{vid} : whereas they were fairly equivocal in classifying the former as /y/ or /u/ (but much less so as /o/), they classified \varnothing_{vid} as /u/ or /o/ (but much less so as /y/). Visible differences in jaw height may explain this pattern of differences: thus, German subjects may have preferentially matched u_{vid} to /y/ or /u/ because of the high jaw position for both German /y u/ and SSBE /u/; and their greater classification of \varnothing_{vid} as /o/ can probably be attributed to the mid-high jaw position in both SSBE /ɔ/ and German /o/. However, irrespective of these differences, the results in Figure 3 strongly support the conclusion that present-day SSBE /u/ and /ɔ/ are both similarly lip-rounded.

The pattern of results from the cross-spliced, auditory-visual stimuli in Figure 4 points to a similar conclusion. German subjects expectedly (mis)classified only 2% of the matched auditory-visual $i_{\text{aud}i_{\text{vid}}}$ stimuli as one of the three rounded vowels. But when the same audio stimuli were combined with the video signals of the other vowels, then German subjects' judgments were much more equivocal: they classified 58% of $i_{\text{aud}u_{\text{vid}}}$ and also 58% of $i_{\text{aud}\varnothing_{\text{vid}}}$ as a rounded vowel (Figure 4). Evidently, then, the visual cues from u_{vid} and \varnothing_{vid} were about equally potent in shifting subjects' judgments of auditory *heed* towards the perception of a rounded vowel.

We used a generalized linear mixed model to compare whether there were differences between the three types of stimuli in their classifications as /i/ as opposed to a rounded vowel (i.e. as opposed to the pooled classifications from /y u o/). The results of this test with classification as the dependent variable (two levels: /i/ vs. a rounded vowel), Stimulus as a fixed factor (three levels: $i_{\text{aud}i_{\text{vid}}}$, $i_{\text{aud}u_{\text{vid}}}$, $i_{\text{aud}\varnothing_{\text{vid}}}$), and Subject (16 levels) as a random factor showed, commensurately with Figure 4, that $i_{\text{aud}i_{\text{vid}}}$ was more likely to be classified as /i/

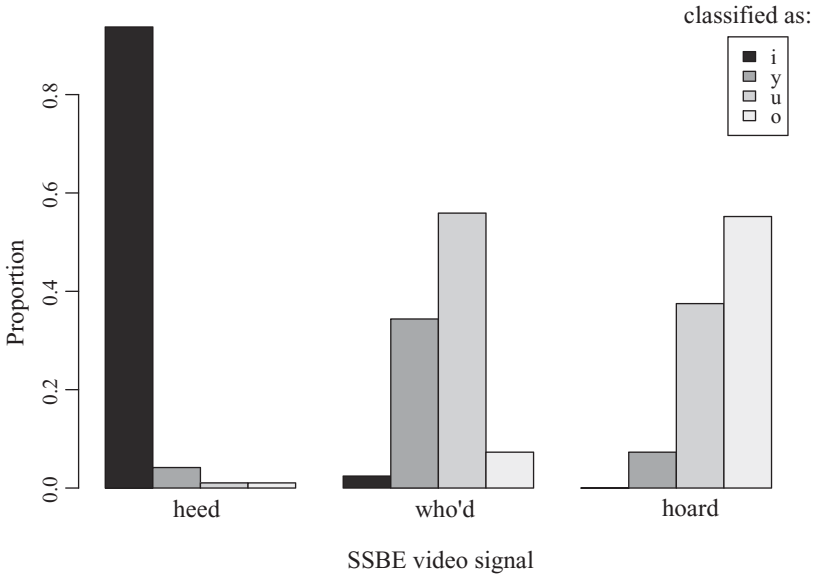


Figure 3 Distribution (proportion) of the forced choice classification as /i y u o/ by 16 German subjects of the video signals of three words produced by a young SSBE speaker.

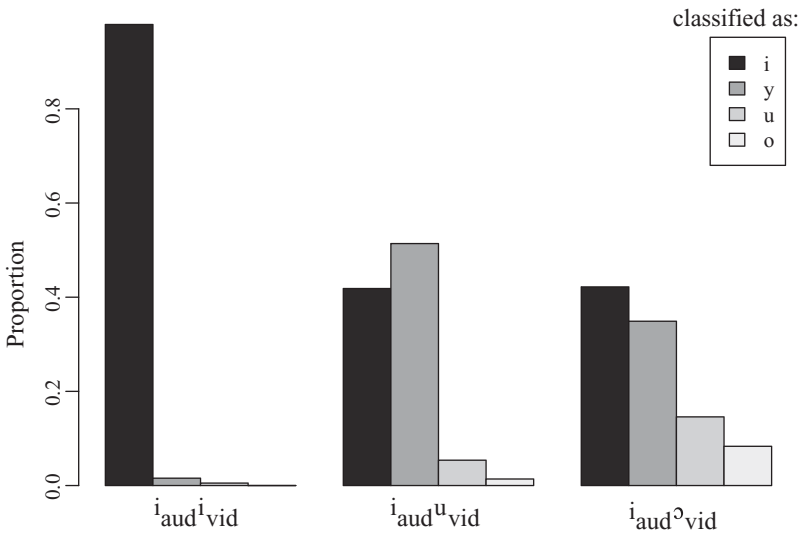


Figure 4 Distribution (proportion) of the forced choice classification as /i y u o/ by 16 German subjects of the audio signals of *heed* simultaneously presented with the video signal of *heed* ($i_{aud} i_{vid}$), *who'd* ($i_{aud} u_{vid}$), and *hoard* ($i_{aud} o_{vid}$).

than were either $i_{aud} u_{vid}$ ($z = 9.4, p < .001$) or $i_{aud} o_{vid}$ ($z = 9.1, p < .001$); there were also no significant differences between $i_{aud} u_{vid}$ and $i_{aud} o_{vid}$ on this measure ($z = 0.1, NS$). Thus, the German subjects were more likely to perceive i_{aud} as a rounded vowel in about equal measure when it was grafted onto u_{vid} or o_{vid} compared with matched $i_{aud} i_{vid}$ stimuli (which were responded to with /i/ almost unequivocally).

4 Physiological analysis of lip movement

So far, the results from the first acoustic experiment have suggested that the differences in lip-rounding between /i/ and /u/ were about the same for old and young SSBE speakers while the second auditory-visual experiment has shown that German listeners interpreted a young SSBE speaker's /u/ and /ɔ/ as rounded vowels in approximately equal measure. The purpose of the following physiological experiment was to analyse the lip position and movement of young SSBE speakers' /u/ more directly and it also included an analysis of lax /ʊ/. The more specific aim was to build on the results of the previous experiment in order to determine whether the lip configuration for /u/ and /ʊ/ was closer to that of /i/ than to that of /ɔ/. If SSBE /u/ and /ʊ/ have shown a tendency to become unrounded, then they should show some differentiation in their lip configuration towards /i/ and away from /ɔ/. This was tested in the next experiment.

4.1 Method

4.1.1 Subjects and materials

Five subjects (three female, two male), including the female subject who had participated in the audio-visual experiment, took part in this study. They were all students from the University of Cambridge of which four had been included in the young SSBE speaker group in the acoustic study of Harrington et al. (2008). Their age range at the time of recording the experiment of the present study was between 21 and 22 years.

The materials were constructed in part based on a physiology study of vowel production and anticipatory lip-rounding of /u/ in Perkell, Matthies, Svirsky & Jordan (1993) and were of the form in which the target /hVd/ word was either medial in a /ma hVd S/ context in which S = /hi hɔ ma/ (*he, haw, ma*) or else final in /ma ma hVd/ context. The same syllable /ma/ was included to ensure a relatively constant starting position for the articulators before each target word. There were six target words (*heed, hid, who'd, hood, hod, hoard*; see footnote 1 above) and therefore 18 medial + 6 final = 24 items. These 24 items were repeated eight times (thus 192 analysed items per subject) and randomized together with a number of fillers (also syllable triplets) whose analysis was not included as part of the present study. The randomization order was the same for each subject.

Each item was presented individually on a computer screen for production in the corresponding orthography (e.g. MA HEAD HE, MA MA HOOD, etc.). In order to minimize the effects of either tempo or variations in prosody and stress, the speakers produced each syllable of each item in time to a metronomic pulse at a rate of 120 pulses per minute (as a result of which each syllable was produced with sentence-stress i.e. was pitch-accented). Before the experiment began, there were practice trials in order for the subjects to familiarize themselves with this procedure. The productions were monitored by one of the authors of this study and any syllables deemed to be substantially out of time with the metronome or otherwise misarticulated were repeated.

4.1.2 Physiological recordings

Physiological data of the movement of the lips, jaw, and tongue were acquired using the system for 5D electromagnetic articulometry at the IPS, University of Munich (see Hoole & Zierdt 2010 for details) allowing the horizontal, vertical, and lateral positions of the articulators to be measured. The movement data were recorded from sensors fixed in the mid-sagittal plane to three points on the tongue, to the upper and lower lips, and to the jaw (Figure 5). The tongue-tip (TT) sensor was attached approximately 1 cm behind the tip of the tongue; the tongue-body (TB) sensor was positioned as far back as the subject could tolerate; the tongue-mid (TM) sensor was equidistant between the two with the tongue protruded. The jaw sensor was positioned in front of the lower incisors on the tissue just below the teeth. The

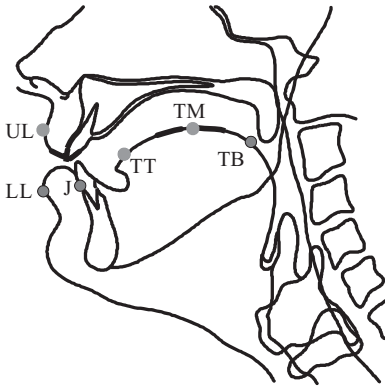


Figure 5 The position of the sensors in the sagittal plane for the upper lip (UL), lower lip (LL), jaw (J), tongue tip (TT), tongue mid (TM), and tongue body (TB). The reference sensors are not shown.

upper-lip (UL) and lower-lip (LL) sensors were positioned on the skin just above and below the lips respectively. In addition, there were four reference sensors which were used to correct for head movements: one each on the left and right mastoid process, one high up on the bridge of the nose, and one in front of the upper incisors on the tissue just above the teeth.

The physiological data were sampled at a frequency of 200 Hz and band-pass filtered with a FIR filter (Kaiser window design, 60 dB at 40–50 Hz for the tongue tip, at 20–30 Hz for all other articulators, at 5–15 Hz for the reference sensors). The data were rotated so that they were parallel to the occlusal plane that was estimated by having a subject bite onto a bite-plate. The synchronized acoustic waveform was digitized at 16 kHz. These procedures were carried out in Matlab and the output stored in self-documented Matlab files. All of the data were converted into an Emu compatible format and analysed in the R programming language (Harrington 2010, Chapter 5).

4.1.3 Parameterization of lip-protrusion

The horizontal movement of the lower-lip sensor, LL_X , has been found in previous studies to provide quite a good indication of lip-rounding (e.g. Perkell et al. 1993) and this was so for the present data in all five subjects. Accordingly, LL_X data were extracted between the acoustic onset and acoustic offset of each vowel in the /hVd/ target words.

The next stage was to reduce each time-varying LL_X trajectory between the vowel's acoustic onset and offset to the same small number of points that encode the trajectory's dynamic shape. For this purpose, the discrete cosine transformation (DCT) was used whose output is a set of $\frac{1}{2}$ -cycle cosine waves which, when summed, reconstruct the original signal. The first three amplitudes of these cosine waves at 0, $\frac{1}{2}$, and 1 cycle that result from the DCT-decomposition are proportional to the signal's mean, linear-slope and curvature (see Watson & Harrington 1999, Harrington et al. 2008 for further details). Thus, after applying the DCT, each dynamic LL_X trajectory between the acoustic onset and offset of the vowel was reduced to a point in a three-dimensional space.

The logarithm of the Euclidean distance ratio, d_V , was used in the following formula to quantify lip-protrusion of /u/ and /ʊ/ relative to that of (spread) /i/ and (rounded) /ɔ/:

$$d_V = \log(E_i/E_o)$$

E_i and E_o above are the Euclidean distances of a given vowel, V , to the same speaker's means of /i/ and of /ɔ/, respectively, in this three-parameter DCT space (see also Harrington et al. 2008 for an implementation of this formula acoustically). Thus, when d_V is zero, any given vowel token is equidistant in this three-parameter DCT space between /i/ and /ɔ/; when d_V

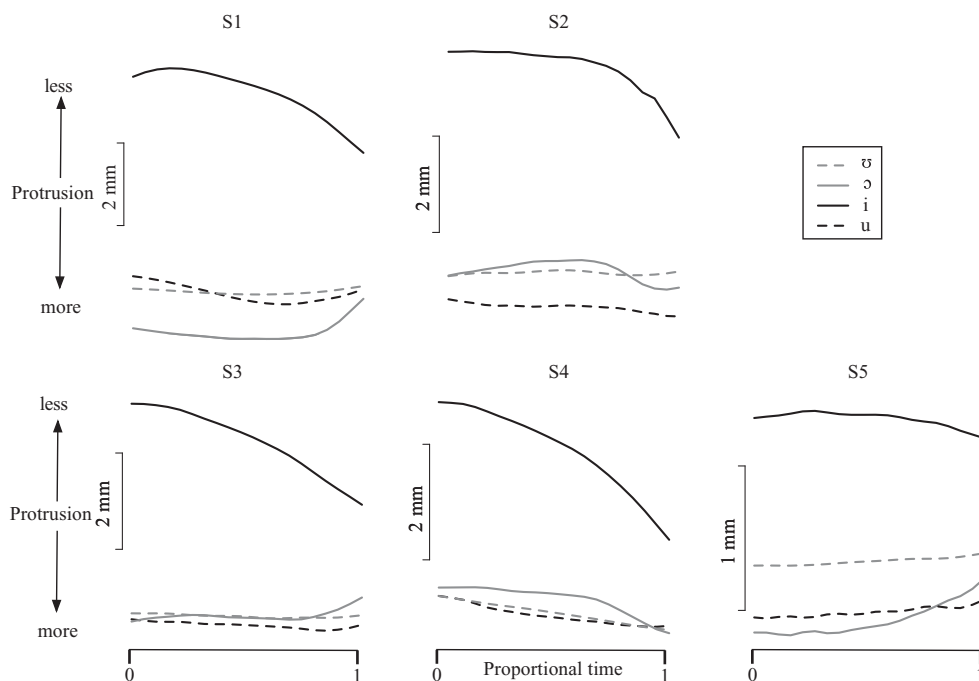


Figure 6 Linearly time-normalized and ensemble-averaged trajectories between the acoustic vowel onset and offset of the horizontal movement of the lower-lip (LL_x) trajectory shown separately for five SSBE speakers.

is negative, then the vowel is closer to (the mean of) /i/ (this is because the numerator in the above formula is small in relation to the denominator); and when it is positive, then it is closer to the mean of /ɔ/. The quantity d_V was obtained separately for each vowel category and separately by subject: in this way, a speaker-specific distribution was obtained of the proximity of vowels to /i/ and to /ɔ/ in this dynamic DCT-parameterized space of lip movement between the acoustic vowel onset and offset.

If /u/ and /ʊ/ were produced with spread lips, then they should be closer to /i/ on this measure whereas if they were produced with lip-rounding, then they should be closer to /ɔ/. These were the hypotheses to be tested.

4.2 Results

The results in Figure 6 of the horizontal movement of the lower lip between the acoustic onset and offset of the vowel show for all five subjects that /u ʊ/ patterned with /ɔ/ and not with /i/: given that SSBE /ɔ/ is unequivocally produced with rounded lips, then so too were /u ʊ/ for these five speakers.

This association of /u ʊ/ with /ɔ/ was confirmed by the distribution of the vowels on the log. Euclidean distance ratio (Figure 7) in which /u ʊ/ were evidently much closer to /ɔ/ than to /i/. The results of a repeated measures MANOVA with the log. Euclidean distance ratio (Figure 7) as the dependent variable showed firstly, that there were significant differences between /i/ and /ɔ/ ($F(1,4) = 128.4, p < .001$): this simply confirms that there was an effective distinction between these two vowels on this parameter; secondly, there was no effect of Vowel on this measure when this factor included only /u ʊ ɔ/ ($F(1,3) = 2.1, NS$), that is, the effect of these three vowels on this parameter was about the same.

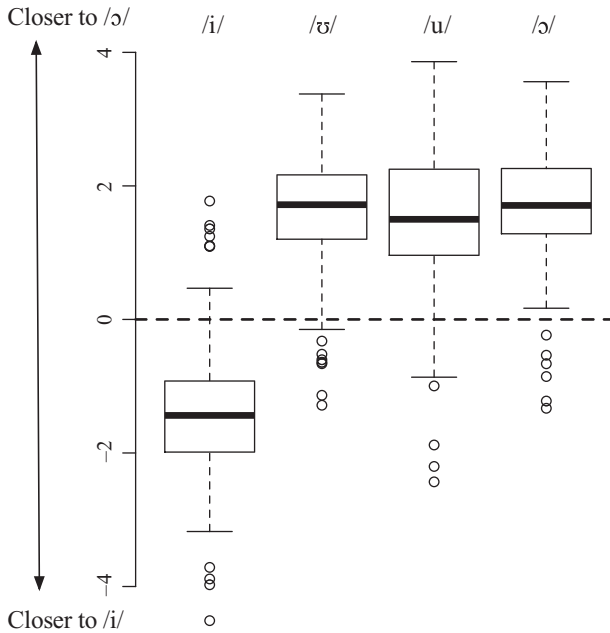


Figure 7 Distribution (arbitrary units) pooled across five SSBE speakers of the log Euclidean distance ratio of a DCT-parameterization of the lower-lip trajectory between the acoustic vowel onset and offset for four vowels. Zero denotes a value that is equidistant on this parameter between /i/ and /ɔ/.

Thus, the conclusion from these results is that the lips in /u ʊ ɔ/ were all protruded to about the same degree.

5 Physiological analysis of tongue movement

The results so far from three separate acoustic, perceptual, and physiological analyses suggest that young SSBE speakers produced /u ʊ/ with rounded lips; in addition, the study on /s/ spectra in this paper is consistent with the view that old and young speakers have about the same degree of rounding in these vowels. Consequently, the raised F2 in /u ʊ/ that has been found for young SSBE speakers in various studies must have been brought about by quite considerable tongue-fronting (as has often been presumed). It is the nature and extent of this tongue-fronting in relation to that of other vowels that was the subject of the investigation described below. Another motivation for the following study was to assess whether there was any evidence that tongue-fronting was less advanced in /ʊ/ than in /u/, as suggested by recent studies (e.g. Hawkins & Midgley 2005) showing that diachronic /ʊ/-fronting had begun some time after that of its tense counterpart.

5.1 Parameterization of tongue movement and position

The subjects, materials, and details of the physiological recordings were as described in the preceding sections (4.1.1, 4.1.2). In addition, the six-dimensional space formed from the vertical and horizontal positions of three sensors on the surface of the tongue tip, mid, and body was rotated using principal components analysis (PCA) separately for each subject producing a new set of orthogonal dimensions that accounted for the greatest variance in the original distribution. PCA was applied separately for each subject in this way to the

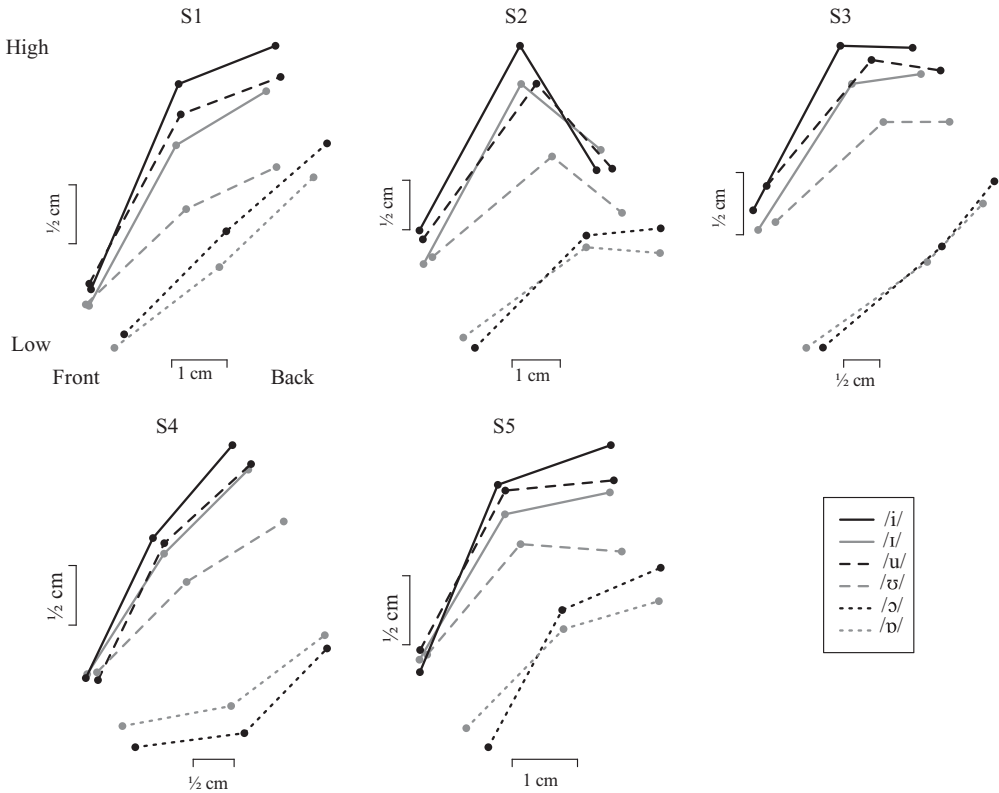


Figure 8 Averaged horizontal and vertical positions of three tongue sensors shown separately for five SSBE speakers and six vowel categories. The three points per vowel from left to right in each subject's data mark the positions of the sensors on the tongue tip, mid, and body, respectively.

tongue data at the temporal midpoint of the vowels in *heed*, *hid*, *who'd*, *hood*, *hod*, *hoard*. The eigenvectors that were thus derived from this procedure were then used to rotate the remaining tongue data between the onset and offset of these words (again, separately for each subject).

5.2 Results

The mean shape of the tongue surface at the temporal midpoint of the vowel is shown separately by subject and by vowel in Figure 8. These data show that there was a greater similarity between /u ʊ/ (dashed) and the front vowels /i ɪ/ (solid) than between /u ʊ/ and the back vowels /ɔ ɒ/ (dotted): this is so for all five speakers both on the horizontal positions of the tongue-body and tongue-mid sensors, as well as on the overall shape of the tongue, especially for subjects S1–S4.

The distribution of the vowels on the first two dimensions that resulted from applying PCA to the six-dimensional space formed from the horizontal and vertical positions of the tongue-tip, tongue-mid, and tongue-back sensors at the temporal midpoint of the vowel is shown in Figure 9. These data show fairly clearly that the first transformed dimension, PCA-1 provided an effective separation between /i ɪ/ (front) and /ɔ ɒ/ (back) vowels for all five subjects. Although somewhat less convincing (in part because only one low vowel was included in the PCA analysis), there is also some evidence that the second rotated dimension, PCA-2, was

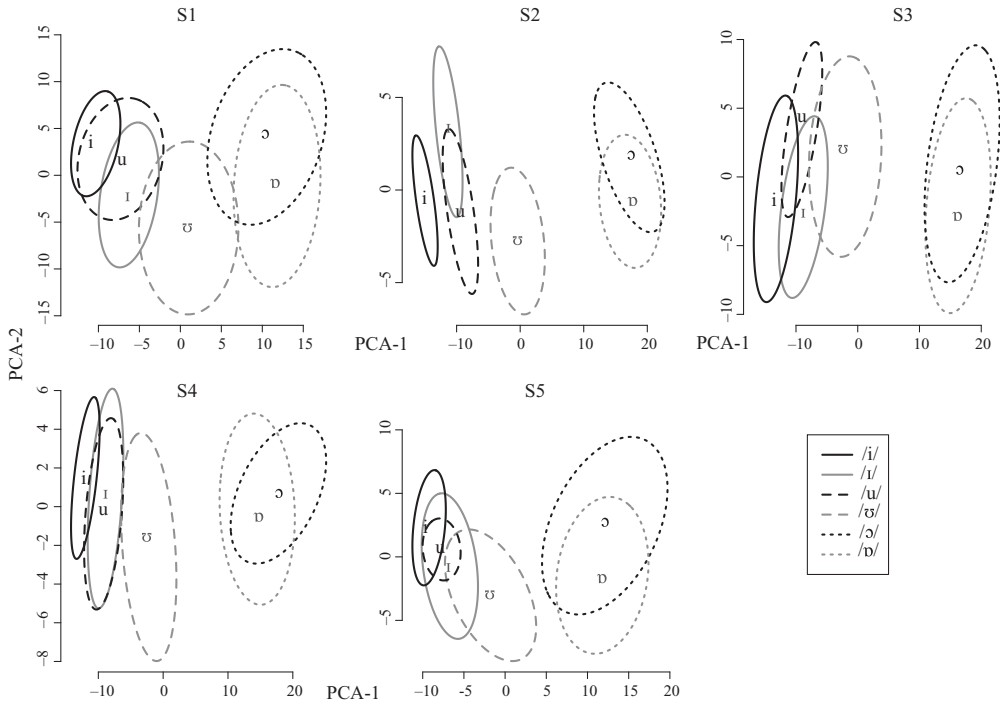


Figure 9 95% confidence ellipses of the first two dimensions (arbitrary units) that resulted from applying a PCA analysis to the combined horizontal and vertical positions of the tongue tip, mid, and body data at the vowels' acoustic temporal midpoints separately for five SSBE speakers. The vowel symbols are positioned close to the centroid of each ellipse.

a phonetic height separator, as shown, for example, by the higher positions of the mid-high vowel /ɔ/ than the low vowel /ɒ/, and (for four of the five subjects) by the higher positions on PCA-2 of /i u/ than /ɪ ʊ/, respectively. As far as the transformed tongue positions of /u ʊ/ are concerned, these were closer to the front than to the back vowels; this is especially so for /u/ which showed overlap in this space with /i/ for 80% of the speakers.

The data in Figure 10 show averaged trajectories of PCA-1 (which, as shown in Figure 9, is a phonetic front–back separator) as a function of time across the entire /hVd/ words. The distribution of these dynamic PCA-1 trajectories were in accordance with many expected phonetic differences. Thus, not only was there a front–back separation as in Figure 9, but also a divergence and convergence between the trajectories on the left and the right of the displays for all speakers respectively: these patterns come about because of the strong coarticulatory influence of the vowels on /h/ at the beginning of the word (thus divergence) and because of the approximation towards a broadly similar tongue configuration (i.e. convergence towards a locus) for /d/ at the end of the /hVd/ word. From the same figure, it is also clear that the tongue position for /u/ was very close to that of /i/ while /ʊ/ occupied a more central position for all speakers.

In order to quantify further the relative position of these vowels, a comparison was made on the inter-vowel distance between /i/ and /u/ with the inter-vowel distance between /ɪ/ and /ʊ/. If the /i–u/ inter-vowel distance was less than that of /ɪ–ʊ/ then, assuming similar positions of phonetic backness in /u ʊ/ some 40–50 years ago, such a finding would be consistent with the view either that /u/ fronted diachronically before /ʊ/, or that the rate of change in fronting across the years had been greater for the tense than for the lax vowel. In order to test these differences, each PCA-1 trajectory as a function of time over the extent of the entire /hVd/

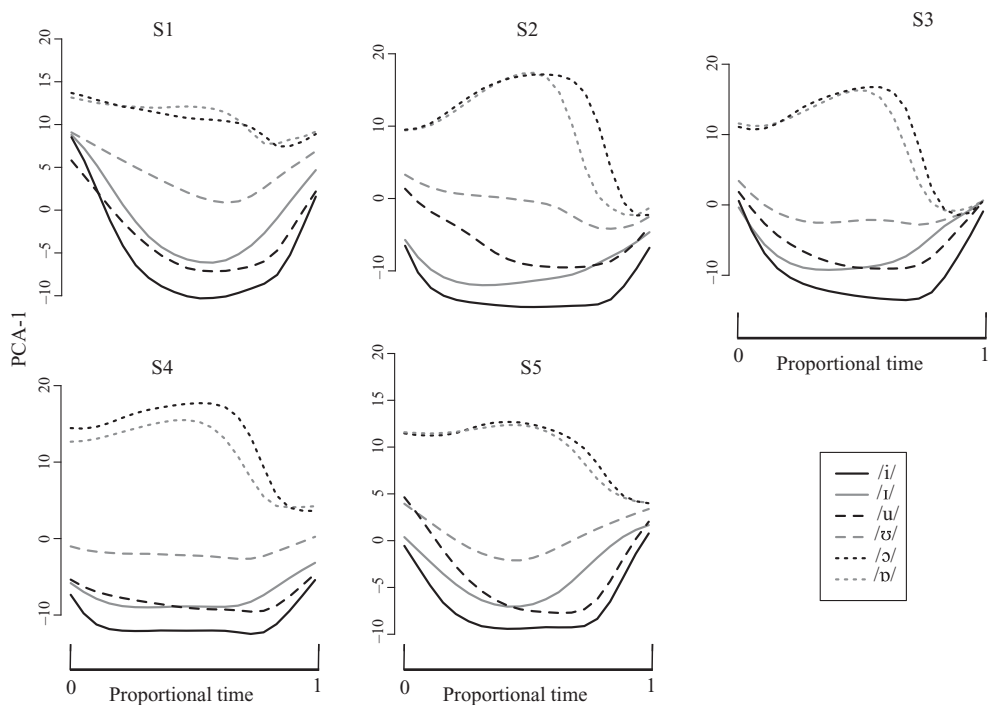


Figure 10 Linearly time-normalized and ensemble-averaged trajectories of PCA-1 tongue data (the first dimension resulting from PCA) between the acoustic onset and offset in six /hVd/ words (coded by vowel nucleus) for five SSBE speakers.

word (i.e. over the extent shown in Figure 10) was reduced to a point in a three-dimensional space using the DCT-transformation described in Section 4.1.3. Then the Euclidean distances firstly of /i–u/ and then of /ɪ–ʊ/ to each others' centroids were calculated separately for each speaker. Thus, for /i–u/, two centroids, m_i and m_u were calculated in this three-dimensional DCT-space for /i/ and /u/, respectively (and for each speaker separately); then the distances of all /u/-tokens in the same speaker's space were measured to m_i as were the distances of all /i/ tokens to m_u . Analogous calculations were made for /ɪ–ʊ/ distances based on m_i and m_u . The hypothesis was that these inter-vowel distances for tense /i–u/ were less than those for lax /ɪ–ʊ/. The distributions of these distances in Figure 11 suggests that this was so for each speaker. The results of a repeated measures MANOVA with inter-vowel distance as the dependent variable and vowel tenseness (two levels: tense /i–u/ vs. lax /ɪ–ʊ/) as the independent factor showed that the distance was significantly less for the tense than for the lax vowel pair ($F(1,4) = 23.1, p < .01$). Consequently, /u/ was indeed closer to /i/ on this dynamic parameterization of tongue-fronting than /ʊ/ was to /ɪ/. These data are therefore consistent with the hypothesis that the fronting of tense /u/ started earlier than that of lax /ʊ/ and/or that the rate of diachronic change has been greater in the tense than in the lax vowel.

6 General discussion

The experiments in this study have provided converging evidence that present-day SSBE /u/ ʊ/ are produced with rounded lips. The results of the first experiment based on an acoustic analysis of anticipatory coarticulation showed that /s/ was produced with rounded lips before

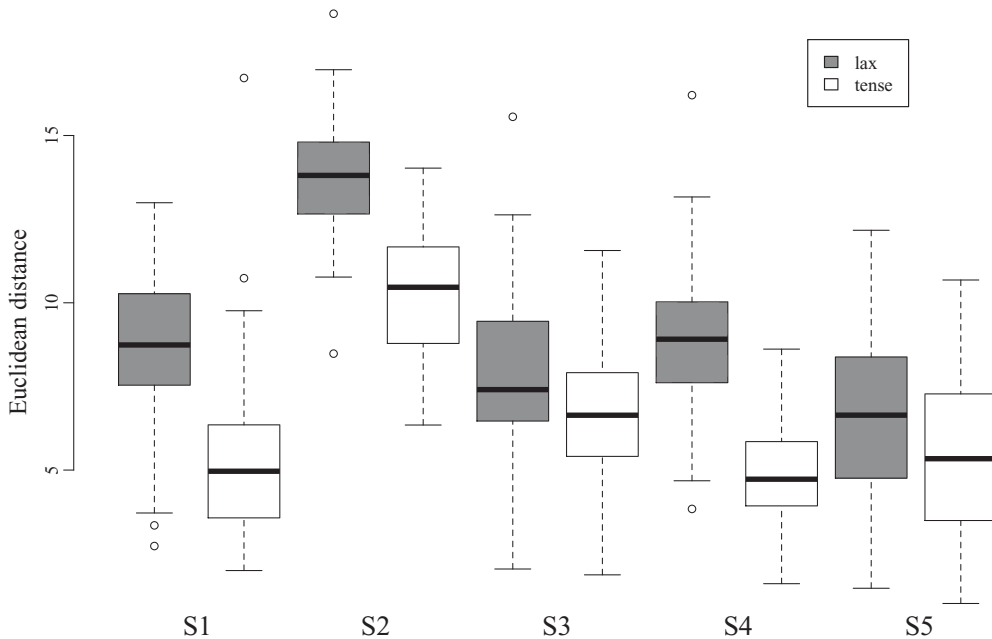


Figure 11 Distribution for five SSBE speakers of the inter-vowel Euclidean distance between *hid* and *hood* (grey) and between *heed* and *who'd* (white) calculated in a three-dimensional space obtained by DCT-parameterizing PCA-1 tongue data between the acoustic onset and offset of these words.

/u/ and also that the extent of lip-rounding inferred from the acoustic signal for young and old speakers was about the same. In the subsequent audio-visual experiment, /u/ produced by a young SSBE speaker was classified in forced choice experiments based both on the video signal only and on cross-dubbed audio and video signals as a rounded vowel by German subjects. The third experiment showed that the lips in SSBE /u/ were as rounded as (the unequivocally rounded) SSBE /ɔ/. Given that the speakers analysed in the present study were also shown to have a higher F2 in /u/ than older SSBE speakers in Harrington et al. (2008), a result which has been found in other acoustic studies as well (Hawkins & Midgley 2005; see also Harrington 2007 for a longitudinal analysis), then such F2-differences must almost certainly be due to tongue-fronting and not lip-unrounding. Just this result was confirmed in the analysis of tongue movement and position in the final part of the study in which the lingual space for /u/ overlapped with that of /i/ and in which /u/ occupied a central position between /i/ and /ɔ/. The results of this part of the study were also shown to be consistent with the hypothesis that diachronic fronting began earlier in SSBE /u/ than in /ɔ/.

The above findings suggest that the lingual position of present-day SSBE is now so front that lip-protrusion is the principal feature for its differentiation from /i/: thus, the sound change has involved a shift in which these vowels used to be differentiated based on both lingual and labial features some 50 years ago to one in which the importance for the /i–u/ distinction of the lingual feature has waned and the labial feature has strengthened. Such a shift in the relative importance of lingual and labial features is less in evidence for lax /i–u/ probably for the reasons stated above that this sound change is a more recent one: but we would speculate that the relative importance of linguality and labiality for the /i–u/ distinction will continue to shift analogously to the way that it has done in the last 50 years for /i–u/.

An immediate objection to the idea that present-day SSBE /i–u/ are distinguished primarily on the basis of labiality is that young SSBE speakers' /u/ (in e.g. *mood*) is auditorily

different from German /y/ (in e.g. the first vowel of *müde* ‘tired’) in which there is a phonemic lip-rounding contrast with /i/. However, there are two points that are relevant here. Firstly, German /i/ and /y/ do not necessarily have the same tongue positions. For example, the factor analysis of tongue data in Hoole (1999) showed that the tongue is not as raised and (given the slight differences on one of the factors) not quite as advanced for front rounded (e.g. /y/) as front unrounded (e.g. /i/) vowels. It is therefore possible – although not yet physiologically investigated – that the tongue-fronting differences between SSBE /i u/ are of a magnitude similar to those of German /i y/. The reason for the perceived difference between present-day SSBE /u/ and German /y/ may instead be because German phonologically tense vowels are themselves more peripheral than their tense counterparts in most varieties of English. Thus, German /lif/ (*lief* ‘ran’) has a more peripheral vowel quality (i.e. one that is closer to CV1) than English /lif/ (*leaf*) and it is this greater peripherality instead of a greater difference in tongue position between German /i y/ than between present-day SSBE /i u/ that may underlie the very clear auditory differences between German /y/ and present-day SSBE /u/. (Differences in the extent of lip-rounding between present-day SSBE /u/ and German /y/ may also contribute to the perceived differences between these vowels, as one of the reviewers has suggested.)

We would also propose that lingual coarticulation of this kind underlies one of Labov’s general principles of chain shifting that back vowels shift to the front diachronically or that, at the very least, diachronic back vowel fronting is more likely than front-vowel backing (there are of course exceptions to this principle, as the diachronic retraction of /ɪ/ in New Zealand English in the last fifty years has shown – see Watson, Maclagan & Harrington 2000, Maclagan & Hay 2007). As Ohala (1993) has argued, it can be shown that many sound changes that tend to recur in the world’s languages have a coarticulatory basis to them – either one of hypo-correction in which listeners underestimate the degree of coarticulatory perturbation or (less probably) one of hyper-correction in which coarticulatory influences are perceptually overestimated and incorrectly factored out. Thus, it is appropriate to seek an explanation for a sound change such as back vowel fronting that has been reported to occur in several structurally unrelated languages (see Labov 1994 for further details) in terms of the coarticulatory fronting effects of coronal consonants on vowels.

The results of a recent study in Harrington, Hoole, Kleber & Reubold (2011) provide further evidence for the coarticulatory basis to back vowel fronting. They showed that the articulatory lingual maneuver – parameterized as the magnitude and peak velocity of the movement of the back of the tongue dorsum over the interval from an initial consonant to a following German /u/ (which is close to CV8) – was significantly more extensive than the analogous movement towards German /i/ (which is close to CV1). Thus, /u/ in languages like German, in which it really is a phonetically back vowel, seems to make more demands on the tongue in CV transitions than do either /i/ or /y/. In addition, listeners were shown in the same study to be much more prone to misidentifying German /u/ as /ʏ/ (whose difference is phonemic, as in e.g. *mussten* vs. *müsten*) due to coarticulatory perturbation than the other way round. In the same study, it was shown that both /u/ and /ʊ/ were much more likely to stray into the /y ʏ/ space based on a parameterization of the movement of the back of the tongue dorsum. Thus, there seems to be not only articulatory but also perceptual pressure on high back vowels to front because of the influence of lingual coarticulation. Finally, if diachronic /u/-fronting is phonetically grounded, then the languages of the world should show a slight bias against /u/. On the one hand, /i u a/ are, of course, the most frequently occurring vowels in the languages of the world. At the same time, /u/ occurs in 28 fewer of the 451 languages of the UPSID database than does /i/ (Maddieson 1984); compatibly, Schwartz, Boë, Vallée & Abry’s (1997) study of UPSID showed that when languages have a non-symmetrical distribution of vowels along a front–back dimension, then they were more likely to be left-dominant (i.e. with a greater number of front vowels) than right-dominant.

Thus, our general conclusion is that it is lingual coarticulation which is likely to be at the core of the general principle of diachronic back vowel fronting, of which these data from

SSBE are one example. There is no evidence that the SSBE diachronic shift in /u ʊ/ has involved an unrounding of the lips.

Acknowledgements

We thank Adrian Simpson and three anonymous reviewers for very helpful comments on an earlier version of this paper. We also thank Lia Saki Bucar, Phil Hoole, and Susanne Waltl for help with recording and data processing of the physiological data, and Manfred Pastätter for help with running the audio-visual experiment. This research was supported by German Research Council grant HA 3512/3-2.

References

- Cox, Felicity. 1999. Vowel change in Australian English. *Phonetica* 56, 1–27.
- Cox, Felicity & Sallyanne Palethorpe. 2001. The changing face of Australian English vowels. In David B. Blair & Peter Collins (eds.), *Varieties of English around the world: English in Australia*, 17–44. Amsterdam: John Benjamins.
- Cruttenden, Alan. 1994. *Gimson's pronunciation of English*, 5th edn. London: Arnold.
- Draxler, Christoph & Klaus Jänsch. 2004. SpeechRecorder – A universal platform independent multichannel audio recording software. *The Fourth International Conference on Language Resources and Evaluation, Lisbon, Portugal*, 559–562.
- Fabricius, Anne. 2007. Vowel formants and angle measurements in diachronic sociophonetic studies: FOOT-fronting in RP. *16th International Congress of Phonetic Sciences (ICPhS 16), Saarbrücken, Germany*, 1477–1480.
- Fant, Gunnar. 1960. *The acoustic theory of speech production*. The Hague: Mouton.
- Fridland, Valerie. 2008. Patterns of /uw/, /ʊ/, and /ow/ fronting in Reno, Nevada. *American Speech* 83, 432–454.
- Gordon, Elizabeth, Lyle Campbell, Jennifer Hay, Margaret MacLagan, Andrea Sudbury & Peter Trudgill. 2004. *New Zealand English: Its origins and evolution*. Cambridge: Cambridge University Press.
- Harrington, Jonathan. 2007. Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts. In Jennifer Cole & José Ignacio Hualde (eds.), *Laboratory phonology 9*, 125–143. Berlin: Mouton de Gruyter.
- Harrington, Jonathan. 2010. *The phonetic analysis of speech corpora*. Chichester: Wiley-Blackwell.
- Harrington, Jonathan, Philip 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(2), 121–131.
- Harrington, Jonathan, Felicitas Kleber & Ulrich Reubold. 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, Sarah & Jonathan Midgley. 2005. Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association* 35, 183–199.
- Henton, Caroline G. 1983. Changes in the vowels of Received Pronunciation. *Journal of Phonetics* 11, 353–371.
- Hoole, Philip. 1999. On the lingual organization of the German vowel system. *Journal of the Acoustical Society of America* 106, 1020–1032.
- Hoole, Philip & Andreas Zierdt. 2010. Five-dimensional articulatory. In Ben Maassen & Pascal H. H. M. van Lieshout (eds.), *Speech motor control: New developments in basic and applied research*, 331–349. Oxford & New York: Oxford University Press.
- Labov, William. 1994. *Principles of linguistic change: Internal factors*. Oxford: Blackwell.
- Labov, William, Sharon Ash & Charles Boberg. 2006. *The atlas of North American English: Phonetics, phonology*. Berlin: Mouton de Gruyter.

- Lindblom, Björn. 1963. Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35, 1773–1781.
- 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.
- Lindblom, Björn & Johan Sundberg. 1971. Acoustical consequences of lip, tongue, jaw, and larynx movement. *Journal of the Acoustical Society of America* 50, 1166–1179.
- Maclagan, Margaret & Jennifer Hay. 2007. Getting fed up with our feet: Contrast maintenance and the New Zealand English front vowel shift. *Language Variation and Change* 19(1), 1–25.
- Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.
- Majors, Tivoli & Matthew J. Gordon. 2008. The [+spread] of the Northern Cities Shift. *University of Pennsylvania Working Papers in Linguistics* 14(2), 110–120.
- Mann, Virginia A. & Bruno H. Repp. 1980. Influence of vocalic context on perception of the [ʃ]–[s] distinction. *Perception & Psychophysics* 28, 213–228.
- McDougall, Kirsty & Francis Nolan. 2007. Discrimination of speakers using the formant dynamics of /u:/ in British English. *16th International Congress of Phonetic Sciences (ICPhS 16)*, Saarbrücken, Germany, 1825–1828.
- Ohala, John. 1981. The listener as a source of sound change. In Carrie S. Masek, Roberta A. Hendrick & Mary Frances Miller (eds.), *Parasession on Language and Behavior (CLS)*, 178–203. Chicago: Chicago Linguistic Society.
- Ohala, John. 1993. The phonetics of sound change. In Charles Jones (ed.), *Historical linguistics: Problems and perspectives*, 237–278. London: Longman.
- Ohala, John & Deborah Feder. 1994. Listeners' normalization of vowel quality is influenced by 'restored' consonantal context. *Phonetica* 51, 111–118.
- Perkell, Joseph S., Melanie L. Matthies, Mario A. Svirsky & Michael I. Jordan. 1993. Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot 'motor equivalence' study. *Journal of the Acoustical Society of America* 93, 2948–2961.
- Pierrehumbert, Janet. 2003. Phonetic diversity, statistical learning, and acquisition of phonology. *Language & Speech* 46, 115–154.
- Schötz, Susanne. 2003. Speaker age: A first step from analysis to synthesis. *15th International Congress of Phonetic Sciences (ICPhS 16)*, Barcelona, Spain, 2585–2588.
- Schwartz, Jean-Luc, Louis-Jean Boë, Nathalie Vallée & Christian Abry. 1997. Major trends in vowel system inventories. *Journal of Phonetics* 25, 233–253.
- Stevens, Kenneth & Arthur House. 1963. Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research* 6, 111–127.
- Trautmüller, Hartmut & Niklas Öhrström. 2007. Audiovisual perception of openness and lip rounding in front vowels. *Journal of Phonetics* 35, 244–258.
- Watson, Catherine I. & Jonathan Harrington. 1999. Acoustic evidence for dynamic formant trajectories in Australian English vowels. *Journal of the Acoustical Society of America* 106, 458–468.
- Watson, Catherine I., Margaret Maclagan & Jonathan Harrington. 2000. Acoustic evidence for vowel change in New Zealand English. *Language Variation and Change* 12, 51–68.
- Wells, John C. 1997. Whatever happened to Received Pronunciation? In Carmelo Medina Casado & Concepción Soto Palomo (eds.), *II Jornadas de Estudios Ingleses*, 2nd edn., 19–28. Universidad de Jaén, Spain. <http://www.phon.ucl.ac.uk/home/wells/rphappened.htm> (retrieved 8 November 2010).
- Xue, Steve An & Grace J. Hao. 2003. Changes in the human vocal tract due to aging and the acoustic correlates of speech production: A pilot study. *Journal of Speech, Language & Hearing Research* 46(3), 689–701.