Zharkova, Natalia and Hewlett, Nigel (2009) *Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using English /t/ and / /.* Journal of Phonetics, 37 (2). pp. 248-256. ISSN 00954470

This article has been published in the Journal of Phonetics. Below is a revised personal version of the article, reflecting the changes made in the peer review process.

The complete citation is as follows:

Zharkova, N. & Hewlett, N. (2009). Measuring lingual coarticulation from midsagittal tongue contours: description and example calculations using English /t/ and /ɑ/. *Journal of Phonetics*, **37**, 248-256.

DOI of the article: doi:10.1016/j.wocn.2008.10.005

Measuring lingual coarticulation from midsagittal tongue contours: description and example calculations using English /t/ and /ɑ/

Natalia Zharkova[*], Nigel Hewlett

Speech Science Research Centre, Queen Margaret University,

Queen Margaret University Drive, Musselburgh EH21 6UU, UK

nzharkova@qmu.ac.uk; nhewlett@qmu.ac.uk

[*] Corresponding author. Speech Science Research Centre, Queen Margaret University, Queen Margaret University Drive, Musselburgh EH21 6UU, UK. Tel.: +44 131 474 0000; fax: +44 131 474 0001.

E-mail address: nzharkova@qmu.ac.uk (N. Zharkova).

Abstract

The purpose of this study was to test the feasibility of a protocol for measuring coarticulation using tongue surface outline data derived from ultrasound imaging. Ultrasound and acoustic data were collected from three speakers of Southern British English while they repeated a list of three sentences 15 times. Tongue surface outlines for the consonant /t/ in /ɑtɑ/ (in "… Ma tasked …") were compared with those for the /t/ in /iti/ (in "… Leigh teased …") and tongue surface outlines for the vowel /ɑ/ in /ɑkɑ/ (in "… Ma cast …") were compared with those in /ɑtɑ/. Nearest neighbour distance calculations were used for the comparison of tongue surface outlines. Mean distance in midsagittal tongue surface outline between tokens of the same phoneme across two different environments was taken as a measure of the phoneme's susceptibility to environment influence. The calculations show that the tongue contour during /t/ adapts to the influence of the neighbouring vowels approximately three times more than the tongue contour during /ɑ/ adapts to the influence of the neighbouring consonants. The applicability of the measure proposed in this paper to future speech research using ultrasound and other articulatory techniques is discussed.

Keywords: coarticulation; ultrasound

1. Introduction

This study was a trial in the use of data from ultrasound imaging to quantify lingual coarticulation. The experimental segments were two English phonemes, one consonant and one vowel, namely /t/ and /ɑ/. Ultrasound was used to compare the tongue surface contours of the same phoneme in two different phonetic environments. The extent to which the tongue posture differed, between one environment and the other, was taken as a measure of that phoneme's susceptibility to coarticulation.

Lingual coarticulation has been studied using acoustic analysis as well as a number of articulatory techniques (see Recasens, 1999, for a review). Quantitative indices of lingual coarticulation based on articulatory results have been mostly designed for electropalatographic (EPG) data. Existing EPG-based indices are targeted at measuring, for example, the difference between overall tongue contact of a given phonetic segment in two different environments (Farnetani, 1990), the difference between two segments in location of the tongue-palate contact (Hardcastle et al. 1991; Gibbon et al. 1993) or the difference in behaviour of parts of the tongue (Recasens et al. 1993; Fontdevila et al. 1994; Recasens et al. 1997).

EPG-based measures of coarticulation are of course based only on information about the parts of the tongue that contact the palate. Thus, in a given comparison of two segments in different contexts, the tongue root is not represented, nor are any parts of the tongue that are not contacting the palate in either context. Electromagnetic articulography (EMA), unlike EPG, provides direct data on lingual position, as it registers movements of coils attached to the tongue. However, EMA analysis involves a comparatively elaborate and invasive data collection procedure. Also, EMA does not capture movements of the whole tongue contour, because there is no information on the behaviour of other parts of the tongue than the flesh points to which the coils are attached. In EMA studies of coarticulation, the data are collected from a limited number of flesh points, and the results are sometimes reported for only one flesh point. For example, Fowler & Brancazio (2000) and Nijland et al. (2004) reported the data from one coil, representing tongue body and tongue tip displacement, respectively. Recasens (2002) reported the results for three coils, located on the tongue tip, blade and dorsum.

Midsagittal tongue contour data have the potential of filling the gaps that exist in EPG and EMA analyses. X-ray studies by Ericsdotter et al. (1998; 1999) have shown that whole tongue contour data can be used to predict patterns of lingual coarticulation in V_1CV_2 sequences with coronal consonants. They also attempted to account for their results using a lingual coarticulation model based on acoustic and X-ray data, described in Öhman (1967). Ericsdotter et al. found a stronger vowel dependence in coronal consonants than could be predicted by Öhman’s model. They claimed that more articulatory data was needed to establish whether degree of consonantal coarticulation was indeed non-uniform across different vowel contexts.

Dynamic imaging of the tongue surface, synchronised with the audio signal, surely provides the best data from which to make measurements of lingual coarticulation. Ultrasound cannot yet rival X-ray or Magnetic Resonance Imaging in clarity of image or richness of detail. Nor is there, at the time of writing, an ultrasound system synchronised with audio, which can rival EMA in its time resolution. However, ultrasound has proved a convenient and effective technique for imaging a large, continuous portion of the tongue surface (see, e.g., Stone 1999; Iskarous 2005; Whalen et al. 2005; Stone 2005). It is comparatively non-invasive, inexpensive and undemanding of participants. The audio signal can be isolated from instrument noise, allowing the collection of excellent simultaneous acoustic data. Systems of ultrasound for speech analysis are currently under development which will yield much better frame rates than that in the data collected for the experiment reported here. Therefore, ultrasound offers the potential for providing large volumes of speech data on which to test claims about coarticulatory properties of different phonemes.

The phonemes used for testing and illustrating the quantification of coarticulation in this work were chosen to represent the two fundamentally opposed classes of sounds, vowels and consonants. A radical distinction between consonants and vowels is common to several theories of speech production (e.g., Öhman 1967; Perkell 1969; Fowler 1980), which claim that consonants are superimposed on vowels in speech and that vocalic effects on neighbouring segments are greater in size than consonantal effects. Some support for these theories comes from a study of jaw height variation in consonants and vowels by Keating et al. (1994). They reported that consonants in vowel-consonant-vowel (VCV) sequences varied more, as a function of vowel environment, than did the vowels as a function of consonant environment.

An alveolar consonant, /t/, and a back low vowel, /ɑ/, were used in this study. With regard to tongue posture, the only requirement for triggering identification of alveolarity by a listener is the creation of a stricture at (and only at) the dentoalveolar region. The rest of the tongue can adjust to neighbouring sounds, according to articulatory convenience (for example, Recasens, 2002, demonstrated considerable adjustments to tongue posture, according to environment, in the production of /n/). The vowel /ɑ/ requires a narrow pharynx and a wide oral cavity, both of fairly uniform cross-section (see, e.g., Hewlett & Beck, 2006, pp. 249-250). Raising of the back of the tongue could compromise contrast with other vowels and raising of the forward part of the tongue could introduce r-colouring. Thus, the acoustic-auditory properties which trigger successful identification of an /ɑ/ in a listener depend not just on the posture of one critical region of the tongue, but rather on its whole surface outline. Similar constraints, with respect to lingual coarticulation, may well distinguish vowels in general from consonants in general.

The aim of this experiment was to quantify the extent (if any) to which the tongue surface outline during the consonant /t/ differed according to the identity (/ɑ/ or /i/) of the flanking vowels of a symmetrical VtV sequence and to similarly quantify the difference in /ɑ/ according to the identity (/t/ or /k/) of the medial consonant in an ɑCɑ sequence. This study used the same paradigm to measure the vowel-on-/t/ (V-on-t) effect as the consonant-on-/ɑ/ (C-on-ɑ) effect, allowing us to quantitatively compare differences in lingual coarticulatory behaviour between the two phonemes.

Existing ultrasound studies of coarticulation have used several different approaches for quantifying coarticulation extent in consonants. Wodzinski & Frisch (2006) analysed coarticulation in /k/ by measuring the angle from the centre of the ultrasound transducer to the centre of the velar closure. Vazquez Alvarez & Hewlett (2007) used a vertical line superimposed on the tongue contour, in order to quantify differences in the tongue dorsum position between an intervocalic bilabial consonant and surrounding symmetrical vowels. Several measures that could be applied to coarticulation analysis were developed by Bressmann et al. (2005). Their study did not focus on coarticulation. Bressmann et al. obtained three-dimensional ultrasound scans from sustained speech sounds and proposed a method for comparing tongue shapes. The indices were based on tongue height values computed using a concentric measurement grid with measurement lines spaced out in equal intervals, superimposed on the ultrasound scan. For examples of using measurement lines to quantify differences in tongue regions in other ultrasound phonetic studies not specifically focussed on analysing coarticulation, see Gick et al. (2006) and Pouplier (2008).

Gordon et al. (2007) compared whole curves for several consonants, using the smoothing spline ANOVA technique (see Davidson, 2006, for more details on the technique); however, the technique allows for determining “whether the shapes of multiple curves are significantly different from one another” (Davidson 2006, p. 411), but not for quantifying the extent of difference between the curves. Davidson (2007) used the programme SURFACES (Parthasarathy et al. 2006), in order to calculate the mean difference in millimetres along the length of the tongue curve for the averaged tongue contours of pairs of phonemes. The approach taken in this work has some technical similarities with that used in Davidson (2007), in that we carried out a numerical comparison of whole tongue curves for pairs of phonemes.

In this experiment, mean nearest neighbour distances were used to represent average distances between pairs of individual curves. The nearest neighbour measure has been widely used in object modelling and manipulation in computer graphics (e.g., see Sankaranarayanan et al., 2007, for a description of an algorithm used to find the shortest distance between each point and all the other points in a collection of points). Since in our data each curve had been captured as a sequence of x-y values, calculations of nearest neighbour distances were easy to implement in Matlab. For descriptions of other methods for quantitative comparison of whole curves derived from tongue outlines used in current ultrasound speech research see Davidson & Stone (2003), Stone (2005) and Davidson (2006).

2. Method

2.1. Data collection

There were three subjects, all native speakers of Southern British English, two female and one male. The stimuli were the sequences /ɑtɑ/, /iti/, and /ɑkɑ/ (with a word boundary before the consonant), in the following sentences:

- At 4 pm M**a ta**sked Janet to paint the roof
- Little L**eigh tea**sed Janet
- After that M**a ca**st an angry look at Leigh.

15 tokens of each sentence were collected. The total number of tokens of the VCV sequences recorded was therefore 135. The participants were given a printout of the sentences for some pre-recording practice. During the recording, the subjects read the sentences as they appeared, one by one, on the computer screen in front of them. The sentences appeared in sequence in the order in which they are listed above and the sequence was repeated 15 times. The subjects were asked to produce the sentences at a comfortable speaking rate.

The ultrasound scans were collected at Queen Margaret University. The hardware consisted of a Merlin Ultrasound Scanner Type 1101 and an Endovaginal End-fire Transducer Type 8561 (manufactured by B-K Medical A/S, Herlev, Denmark). The field of view of the transducer was 160 degrees. The transducer frequency employed in this study was 6.5 MHz. Scanning was performed at a frame rate of 24 frames per second. A helmet, with the transducer attached, was used for immobilising the head in relation to the ultrasound transducer. The helmet was designed and developed by the

Department of Mechanical Engineering at Heriot-Watt University, Edinburgh, UK. In order to reduce across-subject differences in transducer location, that could potentially have an impact on the coarticulation measure derived from tongue surface outlines, we explicitly aimed at positioning the transducer so that the tongue contour from the tip to the shadow of the hyoid bone could be captured and analysed.

The experimental setup, including a diagram and a photograph of the ultrasound scanner and a subject wearing the helmet, is described in more detail in Vazquez Alvarez & Hewlett (2007). The software used for data recording and analysis ("Articulate Assistant"; copyright 2005, Articulate Instruments Ltd), allowed ultrasound video to be recorded, and splines to be fitted by hand. The hardware and software were configured to capture the dynamic ultrasound image from the video output at the PAL video rate of 25 frames per second and synchronise it with the acoustic signal. In this configuration the accuracy of synchronisation with the acoustic signal was limited by the frame rate of the video output and by the ultrasound frame rate, which itself was slightly slower than that of the video. The synchronisation accuracy was estimated at ±40 ms. While such accuracy may be a challenge in speech research, as some sounds can be less than 40 ms, this did not present problems for the present study, because all tokens of all three segments of interest, V1, C and V2, lasted more than 80 ms, and each of them was represented by one annotation point, corresponding to one ultrasound frame. Following video capture, signs of interlacing (two ultrasound images overlapping in the same video frame) are sometimes visible in some frames of the video. Relevant frames were examined for this possibility. In the event, no signs of interlacing were found in the relevant frames.

The acoustic signal was recorded using an Audiotechnica ATM10a microphone. All the recordings were made in a sound-treated studio. The ultrasound scanner and the computer running the Articulate Assistant software were located in an adjacent room, to reduce background noise on the acoustic recording.

2.2. Analysis

2.2.1. Frame selection and spline fitting

Three ultrasound frames were selected from each VCV sequence, one from each vowel and one from the consonant. The three time points at which the frames were selected were determined as follows: the middle of the first vowel (V1), the middle of the consonant closure (C), and a point within the second vowel (V2). The latter was identified using the acoustic record, by measuring the interval from the middle of V1 to the middle of C, and then measuring the same interval from the middle of C into V2. This ensured that the sampling points of the two vowels were equally remote from that of the consonant with which the vowels were to be compared.

For each of the three frames, a cubic spline was manually fitted to the tongue surface contour. The researcher’s input consisted in specifying the number and positions of several (between five and ten) points along the tongue surface, called “hand-labelled points” henceforth in this section. Each stretch of the spline between two adjacent hand-labelled points was defined by a polynomial. Fig. 1 shows the ultrasound frame corresponding to the middle of V1 (in this case, ɑ1) in a token of /ɑkɑ/. The tissue-air interface at the tongue surface is revealed by a line of brightness in the ultrasound image. A spline, represented by the dashed line, has been fitted to the tongue surface contour. It should be noted that the measure of the distance between tongue curves described below is based on the image of the tongue visible on the screen, which may not necessarily be the entire tongue contour. This limitation of the measure should be taken into account when, in midsagittal images, the tongue tip or part of the tongue root drop out from the image; for example, when the tongue tip or blade is extended, as in an interdental consonant, or when the hyoid bone is raised, and its shadow obscures part of the tongue root.

<Fig. 1 about here>

The final contribution of the Articulate Assistant software was to define each spline in terms of x-y values, for storage in a text file. The x-y values were determined by dividing each interval between the hand-labelled points according to the polynomial function parameter (t). Intervals between pairs of hand-labelled points were split in half, then the resulting smaller intervals were again split in half, and so on. The splitting process stopped when the next interval would be less than one pixel. The mean number of x-y values per spline was 100, and the standard deviation was 14. Intervals between pairs of hand-labelled points were not equal, and curve shapes between pairs of hand-labelled points differed, therefore adjacent x-y values were not equally spaced. The mean Euclidean distance between pairs of adjacent x-y values was 0.9 mm, and the standard deviation was 0.28 mm. At the moment, it is not possible to control for the spacing of the x-y values along the curve within the QMU ultrasound system. However, we do not think that these factors affected the validity of the analysis, given the great number of x-y values representing each tongue contour.

2.2.2. Comparison of curve sets

Each text file was imported into Matlab for plotting and analysis. The shortest distance between each two curves was computed, as follows. Taking any two curves, A and B, for any given x-y value (“point” henceforth in the article) on curve A, its nearest neighbour on curve B is that point on curve B which, among all the points on curve B, provides the shortest distance to the given point on curve A; the nearest neighbour on curve A of any given point on curve B is defined in similar fashion. In this experiment, for each relevant pair of curves, A and B, nearest neighbour distances were calculated for each point on curve A and for each point on curve B and the mean of all these nearest neighbour distances was used as a measure of the distance between the two curves.

In order to quantify the difference in tongue position between two realisations of a phoneme, across-environment distance was calculated, by taking each curve for the phoneme in one environment (e.g., /t/ in /ɑtɑ/) and measuring its distance from each curve, in turn, of that phoneme in the other environment (e.g., /t/ in /iti/). For example, distances were calculated between each one of the 15 curves of /t/ in /ɑtɑ/ and each of the 15 curves, in turn, of /t/ in /iti/, producing 225 distances in all. The mean distance value was obtained for each subject separately, for three segments: /t/, /ɑ1/ and /ɑ2/.

To check whether two sets of curves (for example, the set of curves for /t/ in /ɑtɑ/ and the set of curves for /t/ in /iti/) were significantly different from each other, the across-environment distance was compared with the within-environment distance (e.g., the distance between different tokens of /t/ in /ɑtɑ/). Within-environment distance was calculated for each subject separately, by taking each one of the curves for the phoneme in one environment and measuring its distance from each of the curves, in turn, of the other tokens of the phoneme in that same environment. For example, the distance was measured between each one of the 15 curves of /t/ in /ɑtɑ/ and each of the other 14 curves, in turn, producing a total of 105 distances.

For ascertaining the significance level of a difference between across-environment distance and within-environment distance, the set of across-environment distances (225 values) was compared with each set of within-environment distances (105 values). A Univariate ANOVA was conducted in SPSS, separately for /t/, /ɑ1/ and /ɑ2/. The data from all subjects were pooled. The independent variable had three levels: one across-environment distance and two within-environment distances (the two within-environment distances were not collapsed). A probability of less than 0.05 was required in the Post Hoc test between the across-environment distance and each of the within-environment distances, for a difference to be deemed significant. If that occurred, it was concluded that the coarticulatory effect was significant. In all ANOVAs run in this study, Levene's test for equality of population variances produced significant results. We used the Games-Howell Post Hoc test, which is generally used when variances are unequal and sample sizes are unequal (e.g., Field, 2002).

Since one objective of the investigation was to compare the vocalic coarticulatory effect on /t/ with the consonantal coarticulatory effect on /ɑ/, the distance value for /t/ was compared with those for /ɑ1/ and /ɑ2/, using a Univariate ANOVA. This analysis also enabled us to address the issue of directionality of coarticulation: the distance values for /ɑ1/ and /ɑ2/ were compared with each other.

3. Results

3.1. Tongue contours for /t/ and /ɑ/

Fig. 2 (a-c) shows all 15 tongue curves for /t/ in the two vowel environments, for each subject. Fig. 2 (d-f) shows the tongue curves for the first /ɑ/, for each subject, in the two consonant environments. Table 1 shows the mean distance, in mm, between the /t/ in /iti/ and the /t/ in /ɑtɑ/ ($t_i – t_{ɑ}$), between the first /ɑ/ in /ɑtɑ/ and the first /ɑ/ in /ɑkɑ/ ($ɑ1_t – ɑ1_k$) and between the second /ɑ/ in /ɑtɑ/ and the second /ɑ/ in /ɑkɑ/ ($ɑ2_t – ɑ2_k$), for each subject. Within-environment distances are also reported in Table 1. Plots featuring all of the frames of the VCV (for example, 15 curves for /ɑ1/, /t/ and /ɑ2/ from /ɑtɑ/ on the same graph) for all of the tested sequences, for each subject separately, are presented in Appendix A.

<Fig. 2 about here>

<Table 1 about here>

3.2. Comparison of the two realisations of /t/ for significant differences

The second column of Table 1 shows that for all subjects the mean $t_i – t_{ɑ}$ distance was well over 4 mm. The ANOVA showed a significant effect: $F(2,1302) = 7176.06$, $p < 0.001$. The Post Hoc test demonstrated that across-environment distance was significantly greater than both within-environment distances, at the 0.001 level; the difference in within-environment distances (t_i and $t_ɑ$) was also significant.

3.3. Comparison of the two realisations of /ɑ/ for significant differences

The third and the fourth columns of Table 1 show mean $ɑ_t – ɑ_k$ distances of between 1 mm and 2 mm, for all subjects. ANOVAs revealed a significant effect in both ɑ1 and ɑ2: in ɑ1, $F(2,1302) = 117.07$, $p < 0.001$; in ɑ2, $F(2,1302) = 420.82$, $p < 0.001$. Post Hoc tests demonstrated that in both ɑ1 and ɑ2 across-environment distance was significantly greater than both within-environment distances, at the 0.001 level; the difference in within-environment $ɑ1_t$ and $ɑ1_k$ distances was not significant; the difference in within-environment $ɑ2_t$ and $ɑ2_k$ distances was also not significant.

3.4. Comparison of distances between two realisations of /ɑ/ with the distance between two realisations of /t/

Fig. 2 shows that the distance between consonant curves in different vowel environments was greater than the distances between vowel curves in different consonant environments. An ANOVA, which included the distances $t_i – t_ɑ$, $ɑ1_t – ɑ1_k$, and $ɑ2_t – ɑ2_k$, confirmed the presence of a significant difference: $F(2,2022) = 8717.28$, $p < 0.001$. The Post Hoc test results demonstrated a significant difference, at the 0.001 level, between the $t_i – t_ɑ$ distances and each of the $ɑ_t – ɑ_k$ distances.

3.5. Directionality of coarticulation

The ANOVA described in the preceding section also produced results that were used to compare the C-on-ɑ1 effect and the C-on-ɑ2 effect. The Post Hoc test results showed that the $ɑ1_t – ɑ1_k$ distances were significantly smaller than the $ɑ2_t – ɑ2_k$ distances ($p < 0.001$).

4. Discussion

This study presented a measurement technique based on midsagittal tongue contour data derived from ultrasound imaging. The technique consists in quantifying the distance between two sets of curves. It also allows for detecting a significant difference between two sets of curves, from multiple repetitions (in other words, it establishes whether two sets of curves belong to different populations). The distance measure proved robust and internally consistent. Data collection was comparatively convenient and quick. The processes of spline fitting and analysis were rather laborious; however, developments currently underway hold out the prospect of considerably increased automation of these processes. In this event, fairly large scale studies could be attempted with this methodology and it could be applied to the comparison of large classes of different sounds.

The extent of coarticulatory effect on /t/ was greater than the extent of coarticulatory effect on /ɑ/, as expected. It is interesting to contrast our results with those of a study of jaw coarticulation by Keating et al. (1994). They found that some variation in vowel jaw height was due to consonantal influence. However, they observed only a trend towards significance for an effect of consonants on vowels, as opposed to a significant effect of vowels on consonants. In our study, however, the C-on-V effect was also significant. An explanation of this difference in outcome may be that the tongue is more sensitive than the jaw to consonantal coarticulatory effects during vowel production. So our results would appear to confirm the advantage of obtaining direct evidence of tongue posture through a midsagittal view of a large portion of the tongue surface.

As separate calculations were conducted in this study for measuring coarticulatory effects of the consonant on the first vowel and on the second vowel, we could compare the two effects. The tongue contour for the second vowel adapted to the consonant more than the tongue contour for the first vowel, as evidenced by the fact that the distance between /ɑ2/ from /ɑkɑ/ and /ɑ2/ from /ɑtɑ/ was significantly greater than the distance between /ɑ1/ from /ɑkɑ/ and /ɑ1/ from /ɑtɑ/. These results are in agreement with many studies which claim that coarticulation within a CV syllable is stronger than coarticulation within a VC sequence whose segments are separated by a syllable boundary (e.g., Kozhevnikov & Chistovich 1965; Gay 1978; Perkell 1986; Browman & Goldstein 1988; Byrd 1995; Lindblom et al. 2002). We should mention here that the surrounding consonant phonemes were not identical for /ɑkɑ/ and /ɑtɑ/ sequences, therefore a differential coarticulatory influence from these consonants on the vowels cannot be ruled out. For example, /t/ from the word “that” (in “After that Ma cast…”) could conceivably have contributed to the smaller distance between /ɑ1/ from the two sequences than between /ɑ2/ from the two sequences.

The method for quantifying coarticulation based on the ultrasound technique introduced in this study has an advantage over EPG-based methods in that it allows for establishing a C-on-V effect at the middle of a low vowel, in which there is no tongue-palate contact. The measures used by Recasens et al. (1997) allowed them to quantify consonantal influence on the vowel in symmetrical VCV sequences, by comparing the tongue-palate contact pattern at the middle of the vowel with the patterns in each EPG frame during the transition into the neighbouring consonant. The rationale behind this method of measurement was that C-on-V effects occur “during a period when a significant consonant-related acoustic or articulatory difference extends into the vowel” (Recasens et al. 1997, p. 547). This reasoning, however, does not take into account the fact that the steady state of the vowel may have also been affected by the consonant. The present study has demonstrated that there occurs a small but unambiguous effect on /ɑ/ according to consonant environment.

The ability of ultrasound to capture distances between the target segment and its context, though not specifically addressed in this work, provides us with some insight into the question about non-uniformity of V-on-C coarticulation (see the reference to Ericsdotter et al. in the Introduction, above). Visual inspection of the data presented in Appendix A suggests that the distance between /t/ and /ɑ/ was greater than the distance between /t/ and /i/; this observation is confirmed by distance calculations (the mean distance between /t/ and /ɑ/, averaged across subjects, is 4.58 mm, while the mean /t/ – /i/ distance averaged across subjects is 2.02 mm). This supports the claim made by Ericsdotter et al. about non-uniformity of vocalic coarticulatory effects on consonants. The distance between neighbouring segments can also inform us of the degree of gestural compatibility of these two segments, which was claimed by Recasens et al. (1997) to affect the size of coarticulatory effects in speech. Recasens et al. (1997) write that “a more accurate formulation of the articulatory constraints for consonants and vowels” (p. 545) could be used to improve the Degree of Articulatory Constraint model presented in their paper.

The distance measure developed in this paper can be applied to analysing coarticulatory effects that can also be studied using other methodologies. EMA data could be analysed using the distance measure proposed here – for example, if a difference between two locations of the same flesh point was measured for assessing the degree of coarticulation of a particular speech segment. The ultrasound distance measure results can be compared with EPG-based measures of differences in the amount of EPG contact (e.g., the Coarticulation Index proposed in Farnetani 1990). A study of this kind was recently reported in Zharkova (2008); it used synchronised EPG, ultrasound and acoustic data for comparison of EPG and ultrasound measures of several V-on-C coarticulatory effects, in order to assess to what extent these techniques capture the same information, and to describe the effect on the measurement results due to differences in the nature of the data produced by ultrasound and EPG.

The protocol of detecting the presence of a significant difference between two curves developed in this study could be applied to any tongue curve comparison, using midsagittal or coronal ultrasound scans. This approach makes it possible to establish whether the two tongue postures (for example, two realisations of the same segment with different stress or spoken at different speaking rates) differ significantly from each other.

It is worth emphasising that the points on individual curves, that are used for distance calculations in an ultrasound analysis of the sort made in this study, are not actual flesh-points, unlike, for example, the points obtained in EMA data. Rather, the points define a tongue surface outline. However, the shape of the tongue-surface outline is what matters for the acoustic result, rather than the destination of any particular piece of flesh after a movement. The fact that ultrasound does not track flesh-points is therefore immaterial for most purposes. More serious is the difficulty in determining the location of the palate with respect to the tongue surface, because this of course is highly significant from an acoustic point of view. We did not attempt to determine the position of the hard palate in the present study although other researchers have attempted to develop techniques for doing so (e.g., Stone 2005).

The use of the nearest neighbour technique to measure mean distance between curves is not an essential element of ultrasound-derived tongue surface outline analysis and it may not be the best technique in all cases. For example, the same amount of tongue displacement in different directions may yield different mean distances using this measure (e.g., a shift of the tongue 1 mm forwards, which results in a crossover of the two tongue surface outlines under comparison, can be expected to produce a smaller distance value than a shift 1 mm upwards, because distance values are reduced at the crossover).

Finally, with regard to the measure of coarticulation proposed here, note that it has an advantage of not having to posit an ‘intrinsic’ or ‘ideal’ target for the segment under consideration. Coarticulation is measured purely relatively, as a difference between two actual realisations of a phoneme.

Acknowledgements

We would like to express our gratitude to Alan Wrench and Stephen Cowen for their help with the instrumentation and to Ivan Yuen and Robert Rush for their statistical advice. We are also grateful to the editor and three anonymous reviewers for helpful comments. The work was partly supported by an ESRC postdoctoral fellowship (no. PTA-026-27-1268) to Natalia Zharkova.

Appendix A

Plots featuring three sets of tongue curves from 15 repetitions for all VCV sequences, for each subject separately, are shown in Fig. A1. The top row is /ɑkɑ/: a) S1; b) S2; c) S3. The middle row is /ɑtɑ/: d) S1; e) S2; f) S3. The bottom row is /iti/: g) S1; h) S2; i) S3. Red curves – the first vowel, dashed black curves – the consonant, solid black curves – the second vowel.

<Fig. A1 about here>

References

Bressmann, T., Thind, P., Uy, C., Bollig, C., Gilbert, R. W., & Irish, J. C. (2005). Quantitative three-dimensional ultrasound analysis of tongue protrusion, grooving and symmetry: data from 12 normal speakers and a partial glossectomee. *Clinical Linguistics and Phonetics*, *19*, 573-588.

Browman, C. P. & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, *45*, 140-155.

Byrd, D. M. (1995). C-Centers revisited. *Phonetica*, *52*, 263-282.

Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America*, *120*, 407-415.

Davidson, L. (2007). Coarticulation in contrastive Russian stop sequences. In *Proceedings of the 16th International Congress of Phonetic Sciences*. 417-420.

Davidson, L. & Stone, M. (2003). Epenthesis versus gestural mistiming in consonant cluster production: An ultrasound study. In G. Garding and M. Tsujimura (Eds.), *WCCFL 22 Proceedings* (pp. 165-178). Somerville, MA: Cascadilla Press.

Ericsdotter, C., Lindblom, B., & Stark, J. (1999). Articulatory coordination in coronal stops: implications for theories of coarticulation. *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS)*, San Francisco, *3,* 1885-1888.

Ericsdotter, C., Lindblom, B., Stark, J., & Sundberg, J. (1998). The coarticulation of coronal stops: a modeling study. *Journal of the Acoustical Society of America*, *104*, 1804.

Farnetani, E. (1990). V-C-V lingual coarticulation and its spatio-temporal domain. In W. J. Hardcastle and A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 93-110). Netherlands: Kluwer Academic.

Field, A. (2002). *Discovering Statistics: Using SPSS for Windows*. London: SAGE Publications.

Fontdevila, J., Pallarès, M. D., & Recasens, D. (1994). The contact index method of electropalatographic data reduction. *Journal of Phonetics*, *22*, 141-154.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, *8*, 113-133.

Fowler, C. A. & Brancazio, L. (2000). Coarticulation resistance of American consonants and its effects on transconsonantal vowel-to-vowel coarticulation. *Language and Speech*, *43*, 1-41.

Gay, T. (1978). Articulatory units: segments or syllables? In A. Bell & J. B. Hooper (Eds.), *Syllables and Segments* (pp. 121-131). Amsterdam: North Holland.

Gibbon, F. E., Hardcastle, W. J., & Nicolaidis, K. (1993). Temporal and spatial aspects of lingual coarticulation in /kl/ sequences: a cross-linguistic investigation. *Language and Speech*, *36*, 261-277.

Gick, B., Campbell, F., Oh, S., & Tamburri-Watt, L. (2006). Toward universals in the gestural organization of syllables: a cross-linguistic study of liquids. *Journal of Phonetics*, *34*, 49-72.

Gordon, M., Kennedy, R., Archangeli, D., & Baker, A. Distributed effects in coarticulation: an ultrasound study. [Oral presentation at Ultrafest IV, Department of Linguistics, New York University, New York, USA, 28-29 September 2007.]

Hardcastle, W., Gibbon, F., & Nicolaidis, K. (1991). EPG data reduction methods and their implications for studies of lingual coarticulation. *Journal of Phonetics*, *19*, 251-266.

Hewlett, N. & Beck, J. (2006). *An Introduction to the Science of Phonetics.* Mahwah, NJ: Lawrence Erlbaum Associates, (Chapter 17).

Iskarous, K. (2005). Patterns of tongue movement. *Journal of Phonetics*, **33**, 363-381.

Keating, P., Lindblom, B., Lubker, J., & Kreiman, J. (1994). Variability in jaw height for segments in English and Swedish VCVs. *Journal of Phonetics*, *22*, 407-422.

Kozhevnikov, V. A. & Chistovich, L. A. (Eds.) (1965). *Speech: Articulation and Perception*. Washington: Joint Publications Research Service.

Lindblom, B., Sussman, H. M., Modarresi, G., & Burlingame, E. (2002). The trough effect: implications for speech motor programming. *Phonetica*, *59*, 245-262.

Nijland, L., Maassen, B., Hulstijn, W., & Peters, H. (2004). Speech motor coordination in Dutch-speaking children with DAS studied with EMMA. *Journal of multilingual communication disorders, 2*, 50-60.

Öhman, S. E. G. (1967). Numerical model of coarticulation. *Journal of the Acoustical Society of America*, *41*, 310-319.

Parthasarathy, V., Stone, M., & Prince, J. L. (2005). Spatiotemporal Visualization of the Tongue Surface using Ultrasound and Kriging (SURFACES). *Clinical Linguistics and Phonetics*, *19*, 529-544.

Perkell, J. S. (1969). *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study*. Cambridge, MA: MIT Press.

Perkell, J. S. (1986). Coarticulation strategies: preliminary implications of a detailed analysis of lower lip protrusion movements. *Speech Communication*, *5*, 47-68.

Pouplier, M. (2008). The role of a coda consonant as error trigger in repetition tasks. *Journal of Phonetics*, *36*, 114-140.

Recasens, D. (1999). Lingual coarticulation. In W. Hardcastle & N. Hewlett (Eds.), *Coarticulation: Theory, Data and Techniques* (pp. 80-104). Cambridge: Cambridge University Press.

Recasens, D. (2002). An EMA study of VCV coarticulatory direction. *Journal of the Acoustical Society of America*, *111*, 2828-2841.

Recasens, D., Fontdevila, J., Pallarès, M. D., & Solanas, A. (1993). An electropalatographic study of stop consonant clusters. *Speech Communication*, *12*, 335-355.

Recasens, D., Pallarès, M. D., & Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America, 102*, 544-561.

Sankaranarayanan, J., Samet, H., & Varshney, A. (2007). A fast all nearest neighbour algorithm for applications involving large point-clouds. *Computers & Graphics*, *31*, 157-174.

Stone, M. (1999). Laboratory techniques for investigating speech articulation. In W. J. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences* (pp. 11-32). Oxford: Blackwell.

Stone, M. (2005). A guide to analyzing tongue motion from ultrasound images. *Clinical Linguistics and Phonetics, 19*, 455-502.

Vazquez Alvarez, Y. & Hewlett, N. (2007). The trough effect: an ultrasound study. *Phonetica*, *65*, 105-121.

Whalen, D. H., Iskarous, K., Tiede, M., Ostry, D., Lehnert-LeHouillier, H., Vatikiotis-Bateson, E., & Hailey, D. S. (2005). The Haskins Optically Corrected Ultrasound System (HOCUS). *Journal of Speech, Language, and Hearing Research*, *48*, 543-553.

Wodzinski, S. & Frisch, S. (2006). Ultrasound study of velar-vowel coarticulation. *Journal of the Acoustical Society of America*, *120*, 3373-3374.

Zharkova, N. (2008). Using synchronised EPG and ultrasound data for analysing coarticulation. [Oral presentation at the 5th EPG Symposium, Speech Science Research Centre, Queen Margaret University, Edinburgh, UK, 1-2 September 2008.]

FIGURE CAPTIONS

Fig. 1. Ultrasound frame at the middle of V1 in a token of /ɑkɑ/. Distance is in cm and the origin is at bottom right. The tongue tip is on the right. The line of brightness in the ultrasound image represents the tissue-air interface at the tongue surface. The dashed line is the cubic spline that has been fitted to the tongue surface contour.

Fig. 2. Left: tongue contours for 15 repetitions of /t/ in two vowel environments in the three subjects: a) S1; b) S2; c) S3. Black curves are of /t/ in the environment of /i/; red curves are of /t/ in the environment of /ɑ/. Right: tongue contours for 15 repetitions of the first /ɑ/ in two consonant environments in the three subjects: d) S1; e) S2; f) S3. Black curves are of /ɑ/ in the environment of /t/; red curves are of /ɑ/ in the environment of /k/.

Fig. A1. Plots featuring three sets of tongue curves from 15 repetitions for all VCV sequences, for each subject separately. The top row is /ɑkɑ/: a) S1; b) S2; c) S3. The middle row is /ɑtɑ/: d) S1; e) S2; f) S3. The bottom row is /iti/: g) S1; h) S2; i) S3. Red curves – the first vowel, dashed black curves – the consonant, solid black curves – the second vowel.

Table 1

Across-environment distances, in mm, for t_i – $t_ɑ$, ɑ1_t – ɑ1_k and ɑ2_t – ɑ2_k; within-environment distances, in mm, for each target segment. Standard deviations are in italics, below mean values. See subsection 3.1 for further explanation.

Subject	Across-environment distances			Within-environment distances					
	t_i – $t_ɑ$	ɑ1_t – ɑ1_k	ɑ2_t – ɑ2_k	t_i	$t_ɑ$	ɑ1_t	ɑ1_k	ɑ2_t	ɑ2_k
S1	4.72	1.47	1.52	1.15	0.85	1.08	1.11	0.84	1.15
	0.64	*0.46*	*0.39*	*0.42*	*0.26*	*0.38*	*0.40*	*0.21*	*0.40*
S2	4.70	1.17	1.33	1.18	0.85	0.72	0.79	0.85	0.68
	0.78	*0.25*	*0.30*	*0.48*	*0.24*	*0.21*	*0.21*	*0.33*	*0.17*
S3	4.47	1.13	1.91	0.73	1.02	0.95	0.95	1.13	0.93
	0.56	*0.37*	*0.38*	*0.19*	*0.39*	*0.44*	*0.30*	*0.52*	*0.31*

Fig. 1

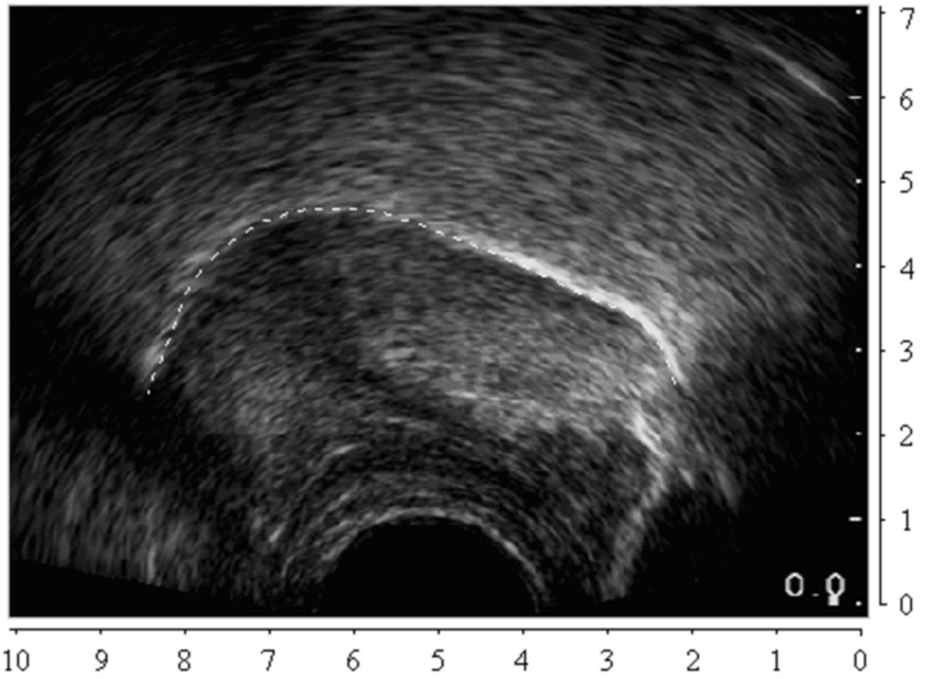

Fig. 2

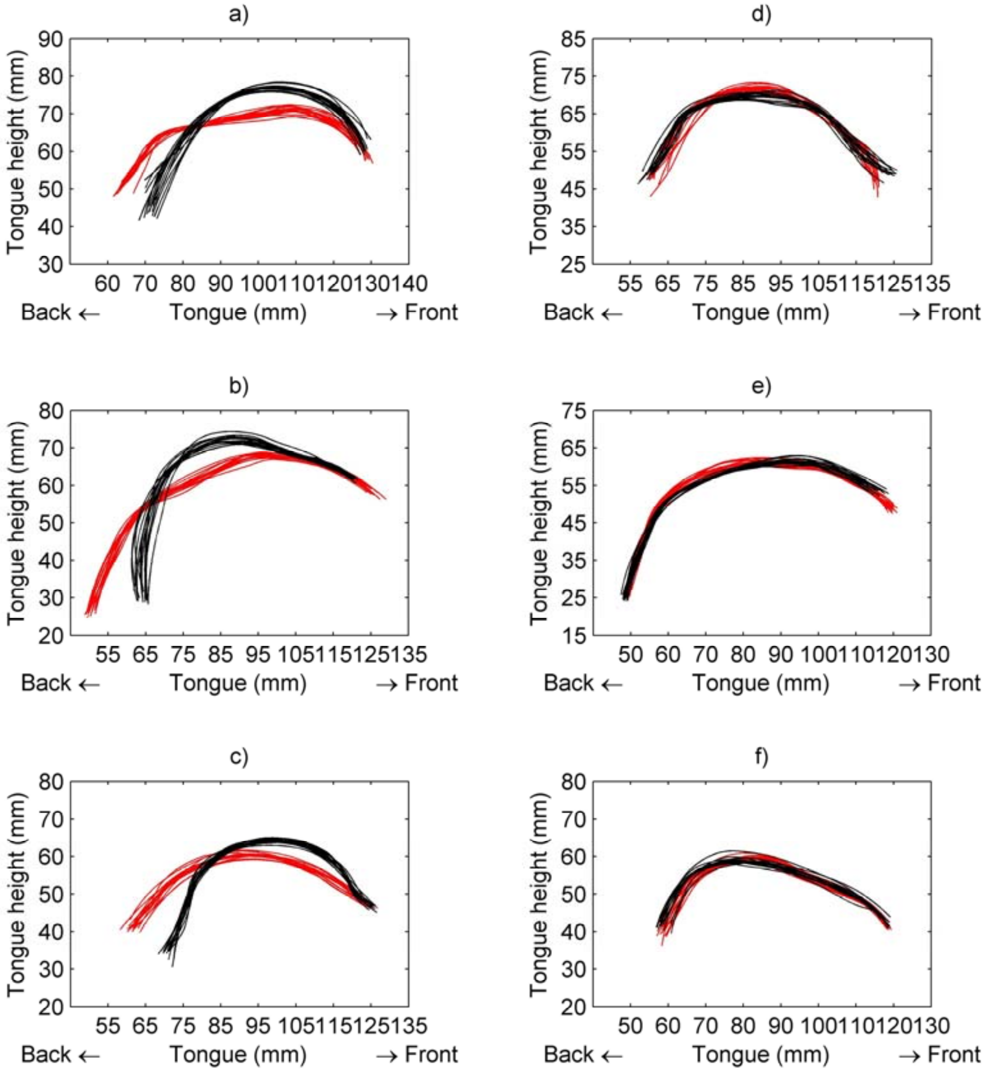

Fig. A1