This is an Author's Accepted Manuscript of an article published in *Clinical Linguistics & Phonetics* on 4/02/2015 (copyright Taylor & Francis), available online:
<http://www.tandfonline.com/doi/full/10.3109/02699206.2015.1007528>.

The complete citation is as follows:

Zharkova, N., Gibbon, F. E. & Hardcastle, W. J. (2015). Quantifying lingual coarticulation using ultrasound imaging data collected with and without head stabilisation. *Clinical Linguistics & Phonetics*, **29** (4), 249-265.

Quantifying lingual coarticulation using ultrasound imaging data collected with and without head stabilisation

Natalia Zharkova

*Clinical Audiology, Speech and Language Research Centre
Queen Margaret University, Edinburgh, UK*

Fiona E. Gibbon

*Department of Speech and Hearing Sciences
University College Cork, Cork, Ireland*

William J. Hardcastle

*Clinical Audiology, Speech and Language Research Centre
Queen Margaret University, Edinburgh, UK*

Correspondence to:

Natalia Zharkova

*Clinical Audiology, Speech and Language Research Centre
Queen Margaret University
Queen Margaret University Drive
Musselburgh EH21 6UU
East Lothian
United Kingdom*

Tel.: +44 (0) 131 4740000

Fax: +44 (0) 131 4740001

Email: nzharkova@qmu.ac.uk

Abstract

Previous studies reporting the use of ultrasound tongue imaging with clinical populations have generally provided qualitative information on tongue movement. Meaningful quantitative measures for use in the clinic typically require the speaker's head to be stabilised in relation to a transducer, which may be uncomfortable, and unsuitable for young children. The objective of this study was to explore the applicability of quantitative measurements of stabilisation-free tongue movement data, by comparing ultrasound data collected from ten adolescents, with and without head stabilisation. Several measures of tongue shape were used to quantify coarticulatory influence from two contrasting vowels on four different consonants. Only one of the measures was completely unaffected by the stabilisation condition for all the consonants. The study also reported cross-consonant differences in vowel-related coarticulatory effects. The implications of the findings for the theory of coarticulation and for potential applications of stabilisation-free tongue curve measurements in clinical studies are discussed.

Keywords that are not in the title: phonetics, articulation, tongue, speech production measurement, adolescent

Introduction

Ultrasound research on speech production in typical speakers has long used quantitative measurements (Stone 2005). By contrast, previous studies using ultrasound tongue imaging with clinical populations of all ages have generally reported qualitative information on tongue movement, or used this information for biofeedback (e.g., Bacsfalvi & Bernhardt, 2011; Bressmann et al., 2011; Preston et al., 2013; Hitchcock & McAllister Byun, Early Online; though see Klein et al., 2013). For these populations, as well as for very young typically developing children, the lack of quantitative measurements has limited progress in our knowledge of areas such as tongue motor control and its development in infancy. We know of only one published study that has used quantitative ultrasound measures with children under the age of four years old (Song et al., 2013). The problem in obtaining quantitative measures is the need to stabilise the ultrasound transducer in relation to the head of the speaker. Restricting head movement limits the use of ultrasound for clinical purposes, as it may be uncomfortable for certain clinical populations, while for very young children head stabilisation is not suitable at all. Quantitative measures producing similar results regardless of whether the head is stabilised or not would be useful in clinical research. The objective of the present study was to explore the applicability of quantitative measurements of stabilisation-free tongue movement data, by analysing ultrasound data collected from the same speakers with and without head stabilisation. The stabilisation device used in the study was a headset holding the transducer (Articulate Instruments Ltd, 2008).

When the speaker's head is stabilised in relation to the ultrasound transducer, statistical comparisons across segments can be based on assessing the degree of similarity in absolute tongue location for the different segments (although a certain amount of residual movement is allowed even in stabilisation devices such as the one used in the current study,

e.g., Scobbie et al., 2008). Without the stabilisation, post-processing of the data, which involves correction for head movement in relation to the transducer (e.g., Whalen et al., 2005; Mielke et al., 2005), can be applied, in order to bring the tongue curve data from different productions by the same speaker into the same coordinate plane (again, with a certain margin of error). When no such technique is used, comparing non-stabilised tongue curve data across segments and repetitions is most reliable when each tongue curve is analysed without reference to absolute coordinates of the other curves of interest. In these circumstances, measures based on a single tongue curve appear promising. They have been applied to non-stabilised data (e.g., Bressmann et al., 2005; Gick et al., 2008; Ménard & Noiray, 2011; Song et al., 2013; Klein et al., 2013), but also to stabilised data (e.g., Stone et al., 1987; Stone et al., 1988; Aubin & Ménard, 2006; Stolar & Gick, 2013; Zharkova, 2013b). We are not aware of any studies that have addressed the performance of tongue shape measures on the data collected in stabilised versus non-stabilised conditions. Ménard et al. (2012) carried out an analysis of several measures' sensitivity to changes in tongue location, using two measures of tongue shape called Curvature Degree and Curvature Position, as well as measures capturing aspects of the absolute location of the tongue (Aubin & Ménard, 2006). Their analysis used tongue curves generated by an articulatory model (Maeda, 1979), and they applied mathematical transformations to tongue curve data in order to approximate effects from transducer movement. Ménard et al. (2012) showed that horizontal and vertical displacement of the transducer, as well as its rotation, significantly affected measures based on the absolute tongue location, but not the measures based on tongue shape. In the present study, we collected ultrasound tongue data from ten speakers who were recorded with and without head-to-transducer stabilisation producing the same stimuli, and we analysed the performance of Curvature Degree and Curvature Position (Aubin & Ménard, 2006), as well as three other measures of tongue shape: Dorsum Excursion Index (DEI, Zharkova, 2013a), Tongue

Constraint Position Index (TCPI; Zharkova, 2013a), and a new measure described below in *Method*. Based on the findings from Ménard et al. (2012), and considering that the stabilised condition would likely restrict jaw movement compared with the handheld condition, we hypothesised that the absolute locations of the tongue curves in the stabilised condition would be less variable than in the non-stabilised condition. For all measures of tongue shape, we predicted the same results in both stabilised and non-stabilised conditions.

In this study, all measures of tongue shape were used to quantify coarticulatory influence from two contrasting vowels (/i/ versus /a/) on four different consonants. The consonants were chosen to test the predictions of the Degree of Articulatory Constraint (DAC) model of lingual coarticulation (Recasens et al., 1997). The DAC model postulates that the degree of tongue dorsum involvement in the segment production affects the segment's ability to resist lingual coarticulation from neighbouring segments. Consonants with a greater constraint on the dorsum (such as the postalveolar /ʃ/) undergo less vowel-related coarticulation than consonants which do not directly involve the dorsum in forming the closure or constriction. Bilabial consonants are among the least resistant consonants, because the tongue is generally not constrained for their production; alveolar consonants are more resistant than labials, but less than postalveolars or palatals (Recasens, 1999). The alveolar fricative /s/ has been shown to have more coarticulation resistance than /t/ (e.g., Stone et al., 1992; Recasens et al., 1997), due to an additional constraint on the lateral margins of the tongue for the fricative production. On the other hand, /ʃ/ has been demonstrated to be more resistant than /s/ (e.g., Tabain, 2001; Recasens & Espinosa, 2009; Niebuhr & Meunier, 2011; Pouplier et al., 2011; Zharkova et al., 2012; 2014a), likely due to the biomechanical properties of the tongue dorsum, specifically its mass and inertia (Recasens & Espinosa, 2010). In the present study, vowel-related coarticulation on /ʃ/, /s/, /t/

and /p/ was analysed, and we expected to observe consonant-specific differences according to the DAC model predictions.

Method

Participants, stimuli and data collection

The participants were ten adolescent speakers, six female and four male, aged between 13 years 0 months and 13 years 11 months, recruited from mainstream schools and with self-reported typically developing speech. The speakers produced CV syllables, with the consonants /p/, /t/, /s/ and /ʃ/, and the vowels /a/ and /i/, in the carrier phrase “It’s a ..., Pam”.

Each target was repeated six times, and the tokens were presented in random order.

Recordings were made in a sound-treated studio. The ultrasound scanner used for the recordings was Ultrasonix SonixRP, with a microconvex transducer C9-5/10, 112.5° field of view and 5 MHz frequency. The recordings were made using the software Articulate Assistant Advanced (Articulate Instruments Ltd, 2012). The acoustic signal, recorded with an Audiotecnica AT803d microphone, was synchronised with the ultrasound signal (the latter recorded at 100 Hz) by Articulate Assistant Advanced, which used hardware pulses generated for every frame by the scanner.

Each participant performed the task two times. The first time was recorded while the participant wore the headset, and the second time no headset was worn, instead the transducer was hand-held by the experimenter (see Figure 1). The former condition will henceforth be referred to as stabilised, and the latter as hand-held. The headset, described in more detail in Articulate Instruments Ltd (2008), was made of aluminium, and contained a polycarbonate inner shell designed to hug the contour of the head. The headset was held in place by a quick

release strap, and it had a number of knobs and screws allowing the adjustment of the headset for different head sizes.

The data collection without the headset was video recorded in two planes, in profile and *en-face*. A separate channel within the multichannel Articulate Assistant Advanced setup was used to record the data from both video cameras. The total number of tokens produced by the speakers was 960 (4 consonants x 2 vowel contexts x 6 repetitions x 10 speakers x 2 conditions).

Insert figure 1 about here

Token selection, annotation and tongue curve fitting

In order to calculate all measures of tongue shape in the present paper, two tongue curve ends of each individual tongue curve needed to be identified. Capturing information about the ends of the tongue curve with ultrasound is known to be difficult and unreliable, due to factors related to the transducer position and angle, and the tongue imageability (Stone, 2005). In order to achieve measurement consistency, during the recording it was ensured that the shadow of the hyoid bone and the shadow of the mandible were present in the midsagittal tongue image, with the imaged tongue curve fitting between the two shadows. For the data recorded without the headset, video recordings were used to verify that for the consonant tokens selected for analysis the ultrasound transducer was located in the midsagittal position. In the data without the headset, one token of /ʃi/ from speaker C4 was corrupted in the recording process and could not be used for analysis. In addition, in the recording process of the total of 59 other tokens from three speakers, due to technical malfunction, the signal from the video cameras was not saved (21 tokens for speaker C3, 28 tokens for speaker C4 and 10 tokens for speaker C8). The data from these recordings were only included in the analysis if

the two shadows were present in the ultrasound image, and the tongue shape at mid-consonant was visually similar to that for the same consonant in the same vowel context in the images that were accompanied by video recordings. The number of tokens without corresponding video recordings that were included in tongue shape analyses was 41 (13 from C3, 20 from C4, and 8 from C8). Given that the above selection procedures eliminated some tokens from the data recorded without the headset, and in order to have a consistent number of repetitions of each target across conditions and speakers, it was decided to include in the analyses only five repetitions out of six (in cases where all 6 repetitions were acceptable, the first five were selected). The total number of tokens for further analyses was 800 (4 consonants x 2 vowel contexts x 5 repetitions x 10 speakers x 2 conditions).

Annotations were made in Articulate Assistant Advanced, based on the acoustic signal. For every consonant token included in the analysis, an annotation was placed at mid-consonant (mid-closure for the stops). The onset of the consonant was defined as the end of the periodic signal associated with the preceding vowel. In several children, preaspirated tokens of the consonant were present (122 tokens in total, with 53% in the handheld condition; 45% of all preaspirated tokens were produced by the speaker C1; 20% by C4, 15% each by C5 and C8, and the remaining tokens by another three speakers). Preaspiration was not included in the consonant annotation; in such cases, the onset of the fricative was located at the abrupt increase of the high-frequency noise amplitude, and the onset of the stop was located at the end of the aperiodic noise associated with preaspiration. The consonant offset for the fricatives was identified as the start of the periodic signal for the following vowel, and the consonant offset for the stops was identified at the beginning of the burst noise. For each consonant token, the outline of the tongue curve was automatically traced in Articulate Assistant Advanced, with manual correction. All the curves were exported from the software

as text files with xy coordinates (the mean number of xy data points per curve was 75.48, and the standard deviation was 10.52).

Quantifying variability in tongue location

The Nearest Neighbour method (Zharkova & Hewlett 2009) was used to quantify variation in absolute tongue locations across repetitions. In this method, the mean Nearest Neighbour (NN) distance between two curves is obtained by calculating the shortest Euclidean distance between each xy point on one curve and the other curve, and then averaging all these shortest distances. In the present study, this method was used in order to establish mean NN distances for each target consonant and each condition, separately for the two vowel contexts. Mean NN distances were computed separately for each speaker, using Python (Lutz, 2008).

Measures of tongue shape

Calculations of all five measures of tongue shape were implemented by the first author in R (R Development Core Team, 2013). A detailed description of Curvature Degree and Curvature Position can be found in Aubin & Ménard (2006) and in Ménard et al. (2012).

More details on the DEI and the TCPI can be found in Zharkova (2013a; 2013b). LOC_{a-i} (a measure described further below) is a new measure introduced in the present study, so it is described in most detail below, together with the rationale for introducing the measure.

Calculations of Curvature Degree and Curvature Position are illustrated in Figure 2(a). The straight line AB is traced between two ends of the curve, followed by tracing the straight line CD, a perpendicular from AB to the point on the tongue furthest away from line CD.

Curvature Degree is the ratio of CD to AB. Higher values of this measure correspond to more bunched tongue shapes. Curvature Position is the ratio of AD to DB. Higher values of

Curvature Position correspond to tongue shapes in which the furthest excursed part of the tongue is located further forward along the tongue contour.

Calculations of the Dorsum Excursion Index (DEI) and the Tongue Constraint Position Index (TCPI) are illustrated in Figure 2(b). The straight line n is first traced between two ends of the curve. A perpendicular d is then traced from mid- n to the point on the tongue curve which is taken to represent the tongue dorsum. DEI is the ratio of d to n . Then, out of all possible perpendiculars from n to the tongue curve, the longest one is selected (line l in Figure 2b). The distance is measured between d and l , marked by a thick grey line s on the graph. TCPI is the ratio of s to $n/2$, and it represents the distance between the tongue dorsum (as defined above) and the furthest excursed part of the tongue. TCPI is assigned a positive value if l is further advanced than d , and a negative value if l is further back than d .

Calculations of the LOC_{a-i} are illustrated in Figure 2(c). Similarly to the other measures used in this study, a straight line is traced between two ends of the curve (line n in Figure 2c). Also like the other measures taken in this study, LOC_{a-i} is based on a ratio of two straight lines. The rationale for developing this measure was the need to quantify coarticulatory difference between consonant tongue curves across two different vowel contexts, /a/ and /i/. For producing these two vowels, the most excursed part of the tongue tends to be further advanced along the tongue curve for /i/ than for /a/ (see the top left panel of Figure 3, for an illustration of tongue curves for /p/ from /pa/ and /p/ from /pi/, which strongly resemble those for the respective vowels). In contrast to the other four measures used in this paper, calculating LOC_{a-i} does not directly involve the straight line between the curve ends. In order to have consistent reference points for comparing the extent of tongue front and tongue back excursion for the consonant in the two vowel contexts, the two reference points are, starting from the front of the tongue, at one third and two thirds of line n , respectively. In other words, line n is divided into three equal parts, and the two

intermediate points are taken to represent the tongue front and the tongue back.

Perpendiculars are traced from each of the two points to the tongue curve (lines f and b in Figure 2c, standing for “front” and “back”, respectively). LOC_{a-i} is the ratio of f to b . Higher values of LOC_{a-i} correspond to more /i/-like tongue shapes, and lower values of LOC_{a-i} correspond to more /a/-like tongue shapes. An additional benefit that we see in this measure is that it is not as strongly dependent on tongue curve ends as the other four measures, because the line n is not included in calculating the index.

Insert figure 2 about here

Difference in magnitude of effect across consonants

The difference across consonants in the magnitude of any vowel-related effect was measured in the following way. First, mean values were calculated for all indices over five repetitions for each speaker, consonant and vowel context. In the event of a significant effect of the vowel context (see the next subsection for the statistical models used to establish significant effects) upon more than one consonant in the same condition, the magnitude of the vowel effect was established for each of those consonants. To establish the magnitude of the vowel effect, for Curvature Degree, Curvature Position, DEI and LOC_{a-i} , ratios were calculated of the mean value in the context of /i/ to the mean value in the context of /a/, separately for each speaker, consonant and vowel context. Larger ratios signified greater vowel-on-consonant influence, therefore greater effect magnitudes. For the TCPI, which could have both positive and negative values, the magnitude of the vowel effect was established by calculating the absolute difference between /i/ context and /a/ context values, also separately for each

speaker, consonant and vowel context. Larger differences were taken to represent greater effect magnitudes.

Statistical analyses

All inferential statistical analyses were carried out in R (version 3.0.1), using the *lmer* software package (Baayen, 2008). Linear Mixed Models (LMMs) were run on the NN distance measure separately for each vowel context and consonant, as well as on each of the tongue shape measures separately for each consonant and condition, with uncorrelated random intercept and random slope for speaker included in each LMM. For the NN distances, the independent variable was condition, and for the tongue shape measures the independent variable was vowel context. Following Reubold et al. (2010), the denominator degrees of freedom were set at 60 for all LMMs in the study. The main effect was deemed significant at the 0.05 level if the F value in the ANOVA exceeded 7.2, and at 0.01 level if the F value exceeded 8.49. All consonants that had a significant vowel-related effect were compared to each other on the magnitude of effect, within condition, using LMMs with the significance criteria described above. To establish any significant differences between consonants, the *multcomp* software package (Hothorn et al., 2008) was used, with Tukey *post hoc* tests.

Results

Tongue curves from an example speaker are presented in Figure 3. The stabilised and hand-held conditions are markedly different from each other in terms of the across-token variability in absolute tongue location, with the curves in the stabilised condition being noticeably closer together within vowel context than the curves in the hand-held condition (cf. Figure 1 of Noiray et al., 2014, for a comparable difference between tongue curves from

handheld ultrasound data corrected to fit the head frame and tongue curves from uncorrected data). A difference in tongue shape depending on the vowel context is also visible on the graphs, but it is not the same across the four consonants. The bilabial stop, in both conditions, appears to show much more difference between the vowel contexts than the other consonants. For the alveolar consonants, in the stabilised condition there is a visible difference in the back of the tongue, with the root further advanced in the context of /i/. In the hand-held condition, the tongue curves for the alveolar consonants in the two vowel contexts are more overlapped than in the stabilised condition, and it is harder to visually determine if there is a difference in tongue shape across vowel contexts. For /ʃ/, in both conditions there does not seem to be a vowel-related difference in tongue shape.

Insert figure 3 about here

Figure 4 shows mean NN distances in the two conditions. For all consonants and for both vowel contexts, the distances in the hand-held condition were significantly greater than in the stabilised condition (see Table 1 for *F* values). Individual patterns matched the group results, in that the distances were larger in the hand-held condition for every speaker, and this applied across consonants and vowel contexts.

Insert figure 4 about here

Insert table 1 about here

Vowel-on-consonant coarticulation

The values of the five measures of tongue shape across condition, consonant and vowel context are presented in Figures 5 to 7. For all four consonants, every measure had a higher mean group value in the context of /i/ than in the context of /a/, in both conditions. The results of the statistical tests are reported in Table 2. The table shows that the only measure which had the same LMM results for all four consonants across conditions was LOC_{a-i} . For this measure, the values in the context of /i/ were significantly greater for /p/, /t/ and /s/, while for /ʃ/ there were no significant vowel-related differences in either condition. On Curvature Degree, three consonants had the same results across conditions: /t/, /s/, and /ʃ/. DEI had the same results across conditions for the two alveolar consonants, while for TCPI and Curvature Position there were the same results across conditions for /p/ and /ʃ/. Significant differences across vowel context in both conditions were observed on three different measures for /p/, /t/ and /s/ (LOC_{a-i} for all of them; TCPI and Curvature Position for /p/ ; DEI and Curvature Degree for the alveolars). For /ʃ/, only one measure produced significant results in both conditions, namely Curvature Degree.

Mean individual values from all speakers over five repetitions (4 consonants x 2 vowel contexts x 10 speakers x 2 conditions = 160 values per measure) generally followed the group pattern of significant results (i.e., the value in the context of /i/ larger than in the context of /a/), with 11 exceptions. The opposite pattern to that reported for the significant group results was observed in one case for Curvature Degree, in two cases for Curvature Position, DEI and TCPI, and in four cases for LOC_{a-i} . The split between stabilised and hand-held conditions on these exception cases was nearly equal, with six cases for the former and five cases for the latter. While the hand-held exceptions were relatively evenly split between speakers (two cases for C3, and one case each for C1, C6 and C10), five out of six stabilised exceptions came from one speaker, C6 (/t/ and /s/ for Curvature Position and for the TCPI,

and /t/ for LOC_{a-i}). The reason for this may be that for this speaker, despite the efforts to keep the imaged plane similar across the two conditions, different amounts of the tongue curve could have been imaged, which would have affected the values of the indices.

Insert figure 5 about here

Insert figure 6 about here

Insert figure 7 about here

Insert table 2 about here

Magnitude of coarticulatory effects across consonants

The consonants significantly affected by the contrasting vowel contexts were compared to each other on the magnitude of effect. In the event, eight comparisons were performed, which correspond to eight columns of Table 2 (all columns except the hand-held condition for Curvature Position and for the TCPI). Four LMMs had a significant effect of consonant on the effect magnitude: in the stabilised condition for Curvature Position ($F = 14.37$), TCPI ($F = 10.98$) and LOC_{a-i} ($F = 15.13$), and in the hand-held condition for LOC_{a-i} ($F = 10.31$). For all four models, applying the Tukey *post-hoc* test showed significant differences between /p/ and each of the alveolar consonants ($p < 0.001$ in every case), but no significant difference across the two alveolars.

Discussion

The results of this study demonstrate that it is possible to get reliable information on lingual coarticulation through applying quantitative measurements to stabilisation-free tongue movement data. Previous articulatory and acoustic studies have found substantial across-speaker variability in absolute values of various indices (e.g., Gibbon et al., 1993; Hardcastle & Gibbon, 1997; Gibbon et al., 2003; Gibbon et al., 2007; Liker & Gibbon, 2008; Koenig et al., 2013). In the present study, across-speaker variability in absolute values of all five indices, as shown in Figures 5-7, was in agreement with previous findings. Similarly to the conclusions from other studies, our results confirm that relative measurements, such as comparisons of index values across vowel contexts, can provide consistent results across speakers.

The results also show some of the limitations that would apply to a study intending to carry out quantitative comparisons of stabilised and hand-held tongue curve data. As predicted, absolute distances between curves for the same consonant and vowel context were affected by whether or not the speakers wore the headset. Measures based on single curves have previously been found to be unaffected by the changes resulting from computer-based simulations of transducer displacement (Ménard et al., 2012). Our findings were only partially consistent with this pattern, with measure-specific and consonant-specific differences, which are discussed below.

Vowel-on-consonant coarticulation as reflected by all measures

Vowel-related coarticulation was reflected, to an extent, in all measures, but the most robust results across consonants were observed for the new measure introduced in this study, LOC_{a-i} . Only this measure had the same statistical results across conditions for all four consonants. This finding means that LOC_{a-i} can be used to compare vowel-related coarticulation in these consonants across groups of speakers, where one group has head-to-transducer stabilisation,

and another group does not have the stabilisation. This would apply when comparing tongue behaviour in groups of speakers who can wear the headset and other groups of speakers who cannot wear it, for example, young children. The fact that, unlike the other four measures, LOC_{a-i} performed in the same way in the two conditions suggests that it is the most robust of all measures used in the study, with most consistent differences between consonant tongue curves in /a/ and /i/ contexts. A possible reason for this, in our view, is that LOC_{a-i} may not be as dependent on accurately locating the two endpoints of the tongue curve – known to be problematic and unreliable landmarks – as the other four measures. We could expect LOC_{a-i} to differentiate between contrasting vowel tongue shapes at least as well as between the tongue shapes for the less resistant consonants (such as bilabials) in the respective vowel contexts. While this study did not aim to carry out a comparison of the actual vowel tongue shapes, we did run LMMs on a subset of the data, with five speakers chosen at random, comparing ratios of LOC_{a-i} values at mid-/i/ to mid-/a/ with ratios of LOC_{a-i} values at mid-/p/ from /pi/ to mid-/p/ from /pa/. Conforming to the predictions, there were no significant differences between ratios for the vowels and ratios for the consonants.

Similarly to LOC_{a-i} , Curvature Position and TCPI quantify the location of the most bunched part of the tongue curve in relation to the rest of the curve. The latter two measures performed in a similar way to LOC_{a-i} in the two conditions for the bilabial (i.e., they showed vowel-related coarticulation) and the postalveolar (i.e. they did not show coarticulation). Both Curvature Position and TCPI, however, failed to show significant coarticulatory effects in the handheld condition for the alveolar consonants. Curvature Degree and DEI both quantify the extent of tongue bunching. Curvature Degree showed a coarticulatory difference in both conditions for three consonants (all but /p/), while DEI only showed a corresponding coarticulatory difference for the alveolars but not for /ʃ/. This suggests that Curvature Degree and DEI could be used for comparing tongue shapes for alveolar consonants in contrasting

vowel contexts across groups of speakers with different stabilisation conditions. Curvature Degree could also be used for similar analyses of /ʃ/, although we have to bear in mind that in the stabilised condition the vowel effect only just reached significance.

Consonant-specific differences

Following the DAC model predictions, and in agreement with previous studies of cross-consonant coarticulation, differences in vowel-related coarticulatory effects were found across consonants regardless of condition, with /ʃ/ least affected by the neighbouring vowels, and /p/ more affected than the alveolars. LOC_{a-i} was the only measure which had the same outcomes in both conditions on cross-consonant differences in magnitude of effect. This result means that out of all five measures used in this study, only LOC_{a-i} can be used for comparing magnitude of coarticulatory effects on these consonants across groups of speakers with different stabilisation conditions, such as older children who can wear the headset and younger children who cannot wear it.

The postalveolar fricative was the only consonant that showed almost no vowel effect in the stabilised condition. A significant coarticulatory effect on the postalveolar was only reported in the stabilised condition on Curvature Degree, and the F value was just above the significance threshold. The results on the bilabial stop are revealing both in terms of describing vowel-related changes in tongue shape for this consonant and in terms of the observed differences across the five measures. While we would expect that /p/ would be differentiated the most of the four consonants (e.g., Recasens, 1999), on Curvature Degree and DEI it had no significant difference between the two vowel contexts in the stabilised condition. The explanation for this is related to the nature of both those measures, and it is illustrated using Figure 3, which presents the data from speaker C2. For Curvature Degree, in the Stabilised condition C2 had the mean value of 0.36 for the context of /a/ and 0.39 for the

context of /i/. The corresponding mean values for the DEI were 0.35 and 0.37. For both measures the values across the two vowel contexts were very close together, whereas for the corresponding values were much further apart for the other three measures (0.73 and 1.57 for Curvature Position, -0.17 vs 0.21 for the TCPI, and 0.85 and 1.43 for LOC_{a-i}). The three latter measures assess the location of the bunched part of the tongue across the two vowel contexts, and the two sets of curves of /p/ in Figure 3 are clearly differentiated by its location, with it being further forward in the context of /i/. The two former measures capture the *extent* of “bunchiness” of the tongue shape, and visual observation of Figure 3 shows that it is quite similar across the two vowel contexts in the stabilised condition. In the hand-held condition, it is possible that the significant difference on Curvature Degree and the DEI was observed because less of the tongue curve could potentially have been imaged due to less pressure from the hand-held probe on the chin than from the stabilised probe. Indeed, in a comparison of the number of xy data points carried out for each consonant, /p/ was the only consonant with a significant effect of condition, and larger values in the stabilised condition. This difference in imaging across conditions shows that caution is needed when choosing measures for comparing coarticulation in bilabial consonants across groups of speakers recorded with versus without head stabilisation.

Methodological challenges and limitations

This study analysed productions by 13-year-old adolescents, who are likely to be more compliant as participants than younger children. In future studies of coarticulation in younger children that use measures of tongue shape on non-stabilised ultrasound data, interpretations of the results need to take this into account. Another limitation of this study is that we used only one technique for head-to-transducer stabilisation, and it would be interesting to find out if other stabilisation techniques (e.g., Stone & Davis, 1995; Bressmann et al., 2005) and other

ways of putting the tongue surface into head space (Whalen et al., 2005; Mielke et al., 2005) would produce similar results. Also, this study included a limited set of consonants, and measurements were carried out at consonant midpoint. It is possible that other consonants and/or other timepoints throughout the consonant may pattern differently with respect to vowel-related differences across conditions.

Due to the nature of the headset used for stabilising the head in this study, we cannot rule out the possibility that some of the stabilised data appeared more consistent than hand-held data because jaw movement was somewhat impeded in the stabilised compared to the hand-held recordings (specifically, TCPI and Curvature Position results for /t/ and /s/, where restricted jaw movement in the stabilised condition might have resulted in compensatory lingual articulations at mid-consonant anticipating the vowel configuration). The same restriction, in varying degrees, would apply to other stabilisation systems, even where the head itself is not fixed but rather the transducer is placed on a stand under the speaker's chin (e.g., Noiray et al., 2008). In our experimental setup, we could not control for any effect of jaw movement or its restriction on the measures in the handheld or stabilised data, respectively. Other factors that were not controlled for in the study are the relationship of the images to the palate and to the occlusal plane. Both these references could have been imaged in the stabilised condition, thus providing additional help in interpreting the results, but neither of them could be imaged in the handheld condition without postprocessing.

Given that some of the results that were expected in the study turned out to be null results (particularly the performance of most measures on /ʃ/ in both conditions), we cannot discount the possibility that an effect might have been observed if the number of speakers and/or the number of tokens had been larger. These unavoidable limitations were due to logistical challenges involved in both data collection and data analysis stages of the study. As far as our findings on /ʃ/ are concerned, however, increasing numbers of speakers or tokens

may not have been helpful, in view of another possible factor, namely the nature of single curve based measures, discussed next.

Measures of tongue shape based on ratios of two straight lines reduce the information from the tongue curve, and therefore comparisons based on such measures are not as powerful as comparisons that take into account information from the whole tongue curve (such as SSAnova, Davidson, 2006, or the NN distance technique, Zharkova & Hewlett, 2009). For example, in a recent study by Zharkova (2013c), lingual coarticulation was analysed using the NN distance technique on tongue curve data from the same recordings of 13-year-old speakers as those used in the current study, and there was a significant vowel-related difference for /ʃ/ across vowel contexts /a/ and /i/, based on comparing absolute tongue locations of the two curves in the two vowel contexts. It is important to bear in mind this difference between the two approaches when interpreting results of studies based on tongue shape measures. However, the increasing number of studies using quantitative ultrasound measurements on non-stabilised speech data from typically developing children and/or children with speech disorders (e.g., Song et al., 2013; Noiray et al., 2013; Klein et al., 2013; Zharkova et al., 2014b) suggests that in future applications it is likely that using non-stabilised data with young children will be a more frequent approach. Therefore investigating the reliability of single curve based measures is warranted, as well as developing more measures that do not depend on multiple curves being located in the same coordinate plane. The results from this study demonstrate that when data from head-stabilised recordings are compared with data from non-stabilised recordings, the choice of phonemes and measures can affect the results. This would also apply in studies comparing lingual articulation across speakers or across groups of speakers recorded exclusively without head stabilisation, as well as in studies comparing stabilised or non-stabilised recordings of the same speaker made on separate occasions.

Acknowledgements

We are grateful to Stephen Cowen for his help in conducting the recordings, to Alan Wrench for advice on instrumentation, and to two anonymous reviewers for very helpful comments.

Declaration of interest

This study was supported by a grant from the Economic and Social Research Council (ES/K002597/1). The authors report no conflict of interest.

References

- Articulate Instruments Ltd. (2008). *Ultrasound Stabilisation Headset Users Manual: Revision 1.4*. Edinburgh, UK: Articulate Instruments Ltd.
- Articulate Instruments Ltd. (2012). *Articulate Assistant Advanced Ultrasound Module User Guide: Version 2.14*. Edinburgh, UK: Articulate Instruments Ltd.
- Aubin, J., & Ménard, L. (2006). Compensation for a labial perturbation: An acoustic and articulatory study of child and adult French speakers. In H. C. Yehia, D. Demolin, & R. Laboissière (Eds.), *Proceedings of the 7th international seminar on speech production* (pp. 209–216), Ubatuba, Brazil.
- Baayen, R. H. (2008). *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge: Cambridge University Press, pp. 241–302.
- Bacsfalvi, P. & Bernhardt, B. M. (2011). Long-term outcomes of speech therapy for seven adolescents with visual feedback technologies: ultrasound and electropalatography. *Clinical Linguistics & Phonetics*, 25, 1034–1043.

- Bressmann, T., Ackloo, E., Heng, C.-L., & Irish, J. C. (2007). Quantitative three-dimensional ultrasound imaging of partially resected tongues. *Otolaryngology – Head and Neck Surgery*, 136, 799–805.
- Bressmann, T., Radovanovic, B., Kulkarni, G.V., Klaiman, P. & Fisher, D. (2011). An ultrasonographic investigation of cleft-type compensatory articulations of voiceless velar stops. *Clinical Linguistics & Phonetics*, 25, 1028–1033.
- Bressmann, T., Thind, P., Bollig, C. M., Uy, C., Gilbert, R. W., & Irish, J. C. (2005). Quantitative three-dimensional ultrasound analysis of tongue protrusion, grooving and symmetry: Data from twelve normal speakers and a partial glossectomee. *Clinical Linguistics & Phonetics*, 19, 573–588.
- 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.
- Gibbon, F. E., McNeill, A. M., Wood, S. E., & Watson, J. M. (2003). Changes in linguapalatal contact patterns during therapy for velar fronting in a 10-year-old with Down's syndrome. *International Journal of Language and Communication Disorders*, 38, 47–64.
- Gibbon, F., Hardcastle, W., & Nicolaidis, K. (1993). Temporal and spatial aspects of lingual coarticulation in /kl/ sequences: a cross-linguistic investigation. *Language and Speech*, 36, 261–277.
- Gibbon, F., Lee, A., & Yuen, I. (2007). Tongue-palate contact during bilabials in normal speech. *Cleft Palate-Craniofacial Journal*, 44, 87–91.
- Gick, B., Bacsfalvi, P., Bernhardt, B. M., Oh, S., Stolar, S., & Wilson, I. (2008). A motor differentiation model for liquid substitutions: English /r/ variants in normal and disordered acquisition. *Proceedings of Meetings on Acoustics*, 1(060003), 1–9.

- Hardcastle, W. J., & Gibbon, F. (1997). Electropalatography and its clinical applications. In M. J. Ball, & C. Code (Eds.), *Instrumental clinical phonetics* (pp. 149–193). London: Whurr.
- Hitchcock, E., & McAllister Byun, T. (Early Online). Enhancing generalisation in biofeedback intervention using the challenge point framework: a case study. *Clinical Linguistics & Phonetics*, 2014, Early Online, 1–17.
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50, 346–363.
- Klein, H. B., McAllister Byun, T., Davidson, L., & Grigos, M. I. (2013). A multidimensional investigation of children's /r/ productions: perceptual, ultrasound, and acoustic measures. *American Journal of Speech and Language Pathology*, 22, 540–553.
- Koenig, L. L., Shadle, C. H., Preston, J. L., & Mooshammer, C. R. (2013). Toward improved spectral measures of /s/: results from adolescents. *Journal of Speech, Language, and Hearing Research*, 56, 1175–1189.
- Liker, M. & Gibbon, F. (2008). Tongue palate contact patterns of velar stops in normal adult English speakers. *Clinical Linguistics & Phonetics*, 22, 137–148.
- Lutz, M. (2008). *Learning Python* (3rd ed.). Sebastopol, CA: O'Reilly Media.
- Maeda, S. (1979). An articulatory model of the tongue based on a statistical analysis. *Journal of the Acoustical Society of America*, 65, S22.
- Ménard, L., Aubin, J., Thibeault, M., & Richard, G. (2012). Comparing tongue shapes and positions with ultrasound imaging: a validation experiment using an articulatory model. *Folia Phoniatrica et Logopaedica*, 64, 64–72.
- Ménard, L., and Noiray, A. (2011). The development of lingual gestures in speech: Comparing synthesized vocal tracts with natural vowels, *Faits de langue*, 37, 189–202.

- Mielke, J., Baker, A., Archangeli, D., and Racy, S. (2005). Palatron: a technique for aligning ultrasound images of the tongue and palate. In Daniel Siddiqi and Benjamin V. Tucker, eds., *Coyote Papers*, 14, 97–108.
- Niebuhr, O., & Meunier, C. (2011). The phonetic manifestation of French /s#f/ and /ʃ#s/ sequences in different vowel contexts: On the occurrence and the domain of sibilant assimilation. *Phonetica*, 68, 133–160.
- Noiray, A., Iskarous, K., Bolaños, L., & Whalen, D. H. (2008). Tongue-jaw synergy in vowel height production: Evidence from American English. In R. Sock, S. Fuchs & Y. Laprie (Eds), *Proceedings of the 8th International Seminar on Speech Production (ISSP)*, 8 – 12 December 2008, Strasbourg, France. Pp. 81–84.
- Noiray, A., Iskarous, K., & Whalen, D. H. (2014). Variability in English vowels is comparable in articulation and acoustics. *Laboratory Phonology*, 5, 271–288.
- Noiray, A., Ménard, L., & Iskarous, K. (2013). The development of motor synergies in children: Ultrasound and acoustic measurements. *Journal of the Acoustical Society of America*, 133, 444–452.
- Poupier, M., Hoole, P., & Scobbie, J. M. (2011). Investigating the asymmetry of English sibilant assimilation: Acoustic and EPG data. *Journal of Laboratory Phonology*, 2, 1–33.
- Preston, J., Brick, N., & Landi, N. (2013). Ultrasound biofeedback treatment for persisting childhood apraxia of speech. *American Journal of Speech-Language Pathology*, 22, 627–643.
- R Development Core Team. (2013). R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. <http://www.R-project.org>.

- Recasens, D. (1999). Lingual coarticulation. In W. Hardcastle & N. Hewlett (Eds),
Coarticulation: theory, data and techniques. Cambridge: Cambridge University Press.
80–104.
- Recasens, D., & Espinosa, A. (2009). An articulatory investigation of lingual coarticulatory
resistance and aggressiveness for consonants and vowels in Catalan. *Journal of the
Acoustical Society of America*, 125, 2288–2298.
- Recasens, D., & Espinosa, A. (2010). Lingual kinematics and coarticulation for
alveolopalatal and velar consonants in Catalan. *Journal of the Acoustical Society of
America*, 102, 544–561.
- Recasens, D., Pallarès, M. D., & Fontdevila, J. (1997). A model of lingual coarticulation
based on articulatory constraints. *Journal of the Acoustical Society of America*, 127,
3154–3165.
- Reubold, U., Harrington, J., & Kleber, F. (2010). Vocal aging effects on F0 and the first
formant: a longitudinal analysis in adult speakers. *Speech Communication*, 52, 638–
651.
- Scobbie, J. M., Wrench, A. A., and van der Linden, M. L. (2008). Head-probe stabilisation in
ultrasound tongue imaging using a headset to permit natural head movement. In R.
Sock, S. Fuchs & Y. Laprie (Eds), *Proceedings of the 8th International Seminar on
Speech Production (ISSP)*, 8 – 12 December 2008, Strasbourg, France. Pp. 373–376.
- Song, J. Y., Demuth, K., Shattuck-Hufnagel, S., and Ménard, L. (2013). The effects of
coarticulation and morphological complexity on the production of English coda
clusters: Acoustic and articulatory evidence from 2-year-olds and adults using
ultrasound. *Journal of Phonetics*, 41, 281–295.
- Stolar, S., & Gick, B. (2013). An index for quantifying tongue curvature. *Canadian
Acoustics*, 41, 11–15.

- Stone, M. (2005). A guide to analyzing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics*, 19, 455–502.
- Stone, M., & Davis, E. P. (1995). A head and transducer support system for making ultrasound images of tongue/jaw movement. *Journal of Acoustical Society of America*, 6, 3107–3112.
- Stone, M., Faber, A., Raphael, L. J., & Shawker, T. H. (1992). Cross-sectional tongue shape and linguopalatal contact patterns in [s], [ʃ], and [l]. *Journal of Phonetics*, 20, 253–270.
- Stone, M., Morrish, K. A., Sonies, B. C., & Shawker, T. H. (1987). Tongue curvature: a model of shape during vowel production. *Folia phoniatrica*, 39, 302–315.
- Stone, M., Shawker, T. H., Talbot, T., & Rich, A. (1988). Cross-sectional tongue shape during the production of vowels. *Journal of the Acoustical Society of America*, 83, 1586–1596.
- Tabain, M. (2001). Variability in fricative production and spectra: implications for the Hyper- & Hypo- and Quantal theories of speech production. *Language and Speech*, 44, 58–93.
- Whalen, D. H., Iskarous, K., Tiede, M. K., Ostry, D. J., Lehnert-LeHouillier, H., Vatikiotis-Bateson, E., & Hailey, D. S. (2005). HOCUS, the Haskins Optically-Corrected Ultrasound System. *Journal of Speech, Language, and Hearing Research*, 48, 543–553.
- Zharkova, N. (2013a). Using ultrasound to quantify tongue shape and movement characteristics. *The Cleft Palate-Craniofacial Journal*, 50, 76–81.
- Zharkova, N. (2013b). A normative-speaker validation study of two indices developed to quantify tongue dorsum activity from midsagittal tongue shapes. *Clinical Linguistics & Phonetics*, 27, 484–496.

Zharkova, N. (2013c). Lingual coarticulation and articulatory constraints in adolescents.

Paper presented at Ultrafest VI (Ultrasound meeting), University of Edinburgh, Edinburgh, UK, 6-8 November 2013.

Zharkova, N., & Hewlett, N. (2009). Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using English /t/ and /a/.

Journal of Phonetics, 37, 248–256.

Zharkova, N., Hewlett, N., and Hardcastle, W. J. (2012). An ultrasound study of lingual coarticulation in /sV/ syllables produced by adults and typically developing children.

Journal of the International Phonetic Association, 42, 193–208.

Zharkova, N., Hewlett, N., Hardcastle, W. J., & Lickley, R. J. (2014a). Spatial and temporal lingual coarticulation and motor control in preadolescents. *Journal of Speech,*

Language, and Hearing Research, 57, 374–388.

Zharkova, N., Lickley, R. J. & Hardcastle, W. J. (2014b). Development of lingual coarticulation and articulatory constraints between childhood and adolescence: an

ultrasound study. In S. Fuchs, M. Grice, A. Hermes, L. Lancia & D. Mücke (Eds),

Proceedings of the 10th International Seminar on Speech Production (ISSP), 5 – 8

May 2014, Cologne, Germany. Pp. 472–475.

Figure captions

Figure 1. Illustrations of the recordings: participant wearing a headset (on the left); transducer hand-held by the experimenter (on the right).

Figure 2. An example midsagittal tongue curve illustrating calculations of the tongue shape indices: a) Curvature Degree and Curvature Position; b) DEI and TCPI; c) LOC_{a-i} . The front of the tongue is on the right in this figure and in Figure 3.

Figure 3. Tongue curves for speaker C2, showing all four consonants in the two vowel contexts, separately for the two conditions (the plots for the stabilised condition are on the left, and the plots for the hand-held condition are on the right). The units on both axes are millimeters. The tongue curves for all five repetitions are displayed on the graphs. The curves in the context of /a/ are represented by solid lines, and the curves in the context of /i/ are represented by dotted lines.

Figure 4. Group averages for mean Nearest Neighbour distances, in mm. The bars for the stabilised condition are white, and the bars for the hand-held condition are shaded. The panels representing the context of /a/ are on the left, and the panels representing the context of /i/ are on the right.

Figure 5. Group averages for Curvature Degree (a) and Curvature Position (b). In this figure and in Figures 6 and 7, the panels representing the stabilised condition are on the left, the panels representing the hand-held condition are on the right, the bars for the context of /a/ are white, and the bars for the context of /i/ are shaded.

Figure 6. Group averages for DEI (a) and TCPI (b).

Figure 7. Group averages for LOC_{a-i} .

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6

Figure 7