Original Paper

Phonetica 2007;64:105–121
DOI: 10.1159/000107912

Received: June 10, 2005
Accepted: January 11, 2007

The 'Trough Effect': an Ultrasound Study

Yolanda Vazquez-Alvarez Nigel Hewlett

Speech and Hearing Sciences/Speech Science Research Centre (SSRC), Queen Margaret University, Corstorphine Campus, Edinburgh, UK

Abstract

Bilabial stops often show a lowering of the tongue in symmetrical VCV sequences. The causes of this phenomenon, sometimes called the 'trough effect', are unknown. However, it could have important implications for the study of time-varying aspects of speech events. Ultrasound is a non-invasive technique that has allowed us to image the shape of the tongue in real time and measure the actual tongue displacement that occurs in the C of a VCV sequence. Five repetitions of symmetrical V_1CV_2 sequences with the bilabial stops /b, p/ were obtained from 10 British English speakers. Results showed not only differences in the direction and degree of the tongue displacement but also differences in the tongue contour configuration between subjects. This study demonstrates the effectiveness of ultrasound as a technique in phonetic research, making possible the analysis of tongue surface movement for large amounts of data from multiple subjects.

Introduction

The tongue does not need to move to create a constriction in the oral cavity to produce a bilabial plosive. The crucial articulatory gesture is the closure of the lips. However, Houde [1968] was the first to find what would come to be known as the 'trough effect' from cineradiographic movement records. He found what he called 'closure-related tongue perturbations' in the form of tongue lowering during the closure phase of the bilabial plosive when in a symmetrical vowel context. For example, in the sequence /ipi/, a lowering of the tongue during the consonant is likely to occur.

A number of different techniques have been used for measuring the trough phenomenon. Using electropalatography (EPG), McAllister and Engstrand [1992] showed substantial reductions in the number of EPG contact points, indicating a lowering of the tongue, during the oral closure in both voiced (unaspirated) and voiceless (aspirated) plosives. The EPG technique gives real-time information on the place of articulation and the linguo-palatal contact pattern during continuous speech. However, this contact information is binary, i.e. it only indicates whether the tongue touches the electrode or not.

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0031–8388/07/0643–0105
$23.50/0
Accessible online at:
www.karger.com/pho

Yolanda Vazquez-Alvarez
The Centre for Speech Technology Research
University of Edinburgh, 2 Buccleuch Place
Edinburgh EH8 9LW (UK)
Tel. +44 131 651 3283, Fax +44 131 650 6626
E-Mail uva_es@yahoo.com

Lindblom et al. [2002] also investigated the nature of this phenomenon using acoustic F2 traces and X-ray data from symmetrical VCV sequences with voiced/voiceless plosives. However, there are problems associated with estimating tongue displacement using acoustic information. Due to the difficulty in teasing apart lip and tongue movement influencing F2 trajectories, they recommended a cautious approach to interpreting this kind of data. They concluded that in order to quantify and model VCV coarticulation the analysis must rely directly on movement records. They presented X-ray data, but only for 1 Swedish speaker. Unfortunately, X-ray databases are not abundant and X-ray is seldom used these days because of the radiation hazard.

Other studies of bilabial plosives using electromyographic (EMG) records by Bell-Berti and Harris [1974] and Gay [1975] reported a deactivation pattern in genioglossus activity, which could be related to the tongue lowering during the closure phase of the consonant. More recently, Fuchs et al. [2004] reported tongue displacement data (using electromagnetic articulography (EMA), and EPG) and posterior genioglossus (GGP) deactivation patterns (using EMG) from 1 native German speaker. The EMA technique has the advantage of being able to trace the movement of several fleshpoints at known locations along the tongue midline but suffers from a variety of limitations when it comes to data collection and positioning of pellets. These limitations include the possible rotational misalignments (e.g. movement of the pellets out of the midsagittal plane on the tongue surface [Hoole and Nguyen, 1999]), the invasive nature of the technique and the fact that the articulation data is point-wise, i.e. the data only provides information on the position and movement of the articulator fleshpoints where the sensors are attached. EMG data, on the other hand, are difficult to collect and interpret due to substantial muscle interdigitation, cross-tack, diffuse fiber distribution of the muscle of interest [Perlman et al., 1989] and fatigue [Faber and Raphael, 1989].

The existence of a tongue lowering/deactivation during the bilabial consonant closure in a symmetrical vowel context has important implications for coarticulation modelling and speech motor control.

A motor control explanation will favour the view of target-driven articulations with a segment-based neural input signal [Lindblom et al., 2002]. For example, the tongue body contour of the vowels [i, a, u] has been reported to move to a neutral position during the stop closure period [Lindblom and Sussman, 2002]. The neutral position for the tongue is usually assumed to be the position it takes during the production of a schwa.

On the other hand, there is a possible aerodynamic explanation, rejected by Lindblom et al. [2002] but supported by Houde [1968], Svirsky et al. [1997] and Fuchs et al. [2004], by which the trough effect could be explained by the rise in intraoral air pressure (P_o) forcing the tongue to move downwards during the stop closure. In support of this explanation, Houde [1968] found smaller tongue displacements when /b/ was spoken with an oral pressure leak tube than without a tube, and zero tongue displacement during the nasal consonant /m/, which could be caused by the intraoral pressure build-up being precluded by the open velum. These results led Houde [1968] to suggest that the tongue displacement during /b/ was the result of a passive reaction to oral pressure. Similar results obtained by Fuchs et al. [2004] also support this explanation. Hoole et al. [1998, p. 145] also support a possible aerodynamic explanation of this phenomenon and go a step further, suggesting a combined approach of both aerodynamic and muscle forces that would influence tongue movement in a way that 'the motor planning system may be anticipating the aerodynamic forces and planning movement trajectories to take advantage of the direction and magnitude of the force vector'.

So, there are at present radically different views on the nature of the trough effect: either it is an active target-driven tongue movement with a segment-based neural input signal or it is a passive tongue movement explained by the rise in intraoral air pressure. However, we felt that a lot more basic data on the phenomenon was needed before progress could be made in deciding among opposing theories. Previous studies have involved a comparatively small amount of experimental data from few subjects.

Inter-Subject Variation

Articulatory results on the trough effect reported up till now could have been misleading because when the tongue image or tongue fleshpoints were available, only 1 or 2 subjects were studied [EMA plus a Glottal Enterprises differential pressure transducer: Svirsky et al., 1997; X-ray recordings: Lindblom et al., 2002]. Otherwise, researchers estimated tongue position and movement from indirect measurements using techniques such as acoustic analysis (formant frequencies), EPG (tongue-palate contacts) and EMG (muscle activity). This raises the issue of what would be the most advantageous method to reveal this phenomenon across a larger population of speakers.

In comparison with other techniques previously used to investigate the 'trough effect', ultrasound is able to image most of the tongue contour [Stone and Davis, 1995; Stone, 2005; Whalen et al., 2005] instead of tracking individual points. The non-invasive and non-intrusive nature of this technique allows the production of fairly unconstrained speech when compared to other techniques such as EMA, EMG or EPG. While ultrasound only requires a transducer positioned under the chin to enable data collection, EMA and EPG require the insertion of pellets or an artificial palate inside the oral cavity and EMG requires a needle electrode inserted in the muscle. In this study, it was possible to acquire large data sets from a large number of speakers. Tongue displacement was measured in multiple repetitions of /p/ and /b/ in three different vowel contexts across 10 subjects.

The aims of this study were to test the application of ultrasound to the phonetic study of tongue contour movement; to gain evidence of the consistency or otherwise of the trough effect across speakers of the same language, and to gain evidence of intra-speaker variation with respect to the trough effect. To our knowledge there are no previous ultrasound studies of the trough phenomenon.

Materials and Methods

Procedure

Subjects

Ten subjects participated in this study. They were all native speakers of British English of various accents and they were all undergraduate students of speech and language therapy.

Speech Material

The speech sample consisted of symmetrical VCV sequences where the consonant (C) was /b/ or /p/ and the vowel (V) was a monophthong, namely /i/, /u/ or /a/. The experimental items comprised both nonsense words and real words. The nonsense words were: <aba> /aba/, <apa> /apa/, <ooboo> /ubu/, <oopoo> /upu/, <eebee> /ibi/ and <eepee> /ipi/. The real words were: <kebab at dawn> /kebab at dɔn/, <nap at noon> /nap at nun/, <freebie> /fribi/, <tepee> /tipi/, <booboo>

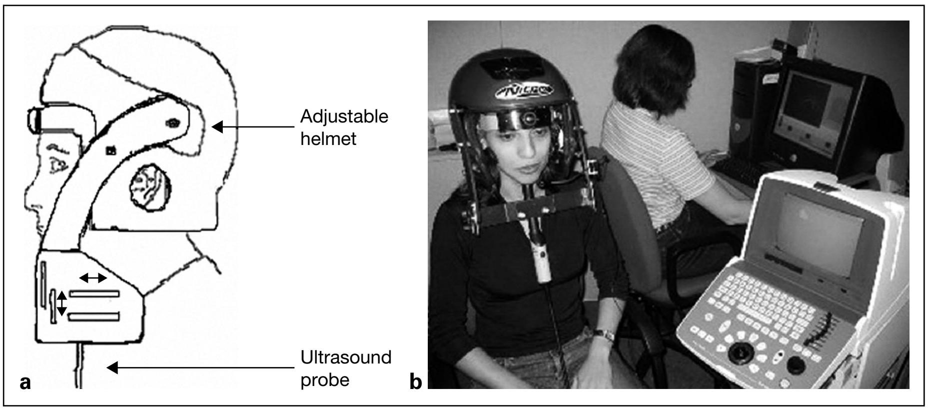

Fig. 1. Ultrasound probe placement (**a**). The acoustic standoff was placed between the probe and the subject's chin, but it is not visible in this illustration because of the size of the helmet. Illustration based on the schematic figure in Keller [1987]. Adjustable helmet set-up (**b**) shows the ultrasound probe being held by the helmet under the subject's chin.

/bubu/ and <two pools> /tu pulz/. Each word was spoken in the carrier sentence 'I said ___ too'. We did not attempt to control precisely for accent because we were investigating an intra-speaker effect. Our speakers were mostly Scottish English with 1 or 2 who had a more southern (towards Southern British Standard) pronunciation. In both these accents, and especially in Scottish English, the vowel in 'nap', for example, would tend to be lower and further back than in American English and so the phonemic symbol /a/ has been used for this vowel. Likewise, the pronunciation of /u/ tends to be more fronted, especially for Scottish English speakers [see Wells, 1982; Hewlett and Beck, 2006, pp. 156–160].

The speakers read five repetitions of each speech item in a randomised order and always started with the real words first. The 60 speech items (30 VCV nonsense sequences and 30 real words) were collected in a first set of data. After a short break of 20 min a second set of the same data was collected. The dynamic image of the tongue surface, together with the acoustic signal, was acquired for both sets of data using the QMU ultrasound system. This yielded a grand total of 60 items $\times$ 2 sessions $\times$ 10 subjects $=$ 1,200 tokens for analysis.

Recording Procedures: Instrumentation and Experimental Set-Up

The ultrasound probe placement and the experimental set-up are illustrated in figures 1 and 2. Tongue movement was recorded by means of ultrasound (Merlin Ultrasound Scanner; Type 1101, B-K Medical A/S, Herlev, Denmark), together with the acoustic signal. The probe (Endovaginal End-fire Transducer; Type 8561, B-K Medical A/S, Herlev, Denmark) was fixed underneath the subject's chin by means of a device attached to the helmet (fig. 1), designed and developed by the Department of Engineering at Heriot-Watt University, Edinburgh. A 7.5-mm acoustic standoff (Sonokit soft, Sonogel, Germany) was placed between the probe and the chin, to reduce tissue compression when the jaw lowers, as recommended by Stone and Davis [1995]. The centre frequency of the probe was set to 6.5 MHz with an image field of sector 160°. The ultrasound images were obtained at a frame rate of 25 frames per second (fps). Axial resolution was 0.5 mm and penetration depth was 95 mm, when measured in water for the specific frequency used for this experiment.

Bone shows up as a dark shadow on an ultrasound image. The mandible and hyoid bone create an 'acoustic shadow' (black region) at both edges of the image that can obscure parts of the tongue tip or root [Stone, 2005]. The probe was positioned so that the jaw shadow and the hyoid shadow were visible at either edge of the display when the vocal tract was in a resting state. This ensured a comparable orientation of the tongue for the first and second sets of data for each participant.

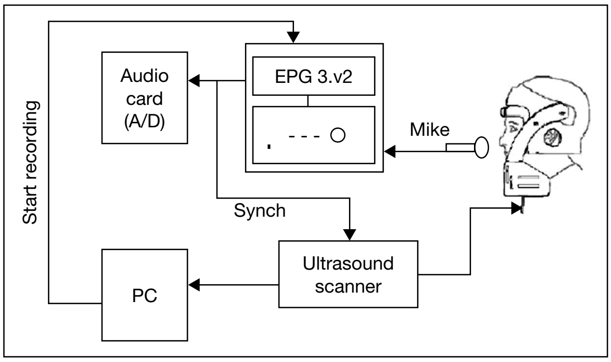

Fig. 2. The experimental set-up at QMU ultrasound lab. Acoustic and ultrasound data were acquired via a microphone attached to the helmet and an ultrasound scanner, respectively, and stored in a computer. The labelling, tongue edge tracking and statistical analysis were performed on a computer in the ultrasound data analysis area.

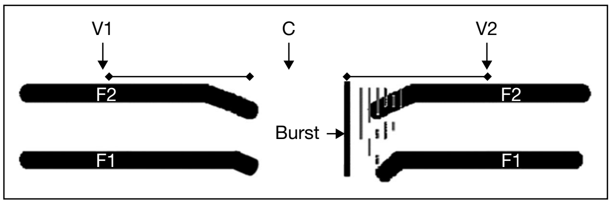

Fig. 3. Schematic illustration of data annotation. Three different annotation points were selected by looking at the acoustic signal and the spectrogram information: (1) V1 = centre of V_1, (2) C = centre of the closure phase of the bilabial plosive and (3) V2 = a point during V_2 that was calculated by first measuring the distance from the centre point of V_1 to the vowel offset. Then, that distance from the burst of the consonant was used to define the third annotation point in V_2. In this way, we established a symmetrical distance from each side of the consonant closure into the flanking vowels.

Data Analysis (Processing of Lingual Movement)

Ultrasound images were gathered in the form of mid-sagittal sections of the tongue and synchronised with the acoustic signal during the acquisition of the data. The data was later analysed using Articulate Assistant software (copyright 2005, Articulate Instruments Ltd.).

Temporal Annotation

First, the acoustic signal was annotated using the following criteria. Three different points in time were annotated (fig. 3): (1) centre of V_1, (2) centre of the closure phase of the bilabial plosive, (3) a point during V_2. This latter point was calculated by first measuring the duration from the centre point of V_1 to the vowel offset. Then, that duration (viz., V1 mid to V1 offset) was measured from the release of the consonant in order to define the third annotation point in V_2. In this way, we established a symmetrical distance from each side of the consonant closure into the flanking vowels.

The tongue contour was identified by a bright white band on the ultrasound frame (fig. 4a) that is created by the ultrasound waves travelling upwards from the probe through the tongue body until they reach the upper surface of the tongue (typically bounded by air or the palate bone) and reflects back to the probe. Thus, the bright white band is the air reflection at the upper surface of the tongue. Next, a spline (i.e. a smooth curve that connects discrete data points called knots) was manually fitted to the tongue contour on the ultrasound frame corresponding to each of the annotation points. The spline was fitted consistently to the lower edge of the bright white band, as this is the upper surface of the tongue. The upper edge of the band has no physical interpretation.

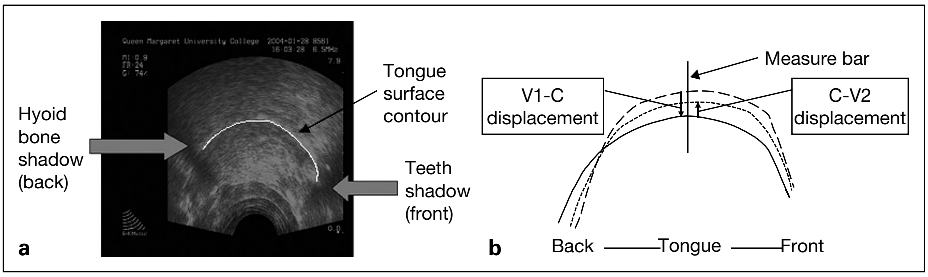

Fig. 4. **a** Example of spline fitting on the ultrasound image. **b** Schematic illustration of the method followed to measure tongue displacement. A measure bar, shown as a vertical line, was used to measure the distance between tongue contours. The dashed line represents the tongue contour during V_1. The solid line represents the tongue contour during closure phase of the bilabial plosive. The dotted line represents the tongue contour during V_2.

The 25-fps frame rate provided by the ultrasound scanner does not allow a fine time resolution with respect to the acoustics. Because different ultrasound machines have different delays between scanning the image and outputting it through the video port, an experiment was carried out to ascertain how likely the ultrasound frame would be reflecting the tongue movement at the time of the annotation. Results from this experiment showed that the variation error was +40 or −40 ms from the annotation point. Since the length of the bilabial closure was longer than this error [bilabial closure means across the two sessions: /b/ = mean: 75 ms, standard deviation (SD): 14; /p/ = mean: 86 ms, SD: 15], we can be sure that the ultrasound image selected to represent tongue position during the closure phase could rarely have occurred outside the consonant closure.

In addition, given that the data collected comprises 50 repetitions of each VCV, many different points during the closure would be represented in the data, distributing the error equally.

Qualitative and Quantitative Analysis of Tongue Displacement

In order to give a clear impression of overall tongue contour and its movement, averaged tongue contours will be presented in the 'Results' section when dealing with the inter-subject variation of the trough effect. To be able to compare the tongue shapes during the vowels with that during the bilabial closure, we exported the x and y coordinates from a tongue contour like the one illustrated in figure 4a and plotted them on a Cartesian grid using Matlab. For quantitative analysis, however, a measurement was made from the V1 contour to the C contour and from the C contour to the V2 contour, along a vertical line (labelled 'measure bar' in fig. 4b), which intersected the highest point of the C contour.

Figure 5 shows an example of how the exported data representing the tongue movement will be reported in the 'Results' section in this study. This is an example of the /ipi/ sequence and the scale is in millimetres (mm). A centimetre (cm) scale on the right hand side of the ultrasound image (fig. 4a), ranging from 0.0 to 7.9 cm, was used to calibrate the conversion of pixels into millimetres. The left bar shows the mean vertical distance from the highest point of the C contour to the V_1 contour (V1-C). The difference is displayed as negative, symbolising that, in this case, there was a downward movement of the (highest point of the) tongue between the V_1 and C contours. The right bar shows the mean vertical distance from the highest point of the C contour to the V_2 contour (C-V2). The difference is displayed as positive, symbolising that there was an upward movement of the tongue between the C and the V_2 contours. These two measurements will be referred to as measures of tongue displacement (MTD). The zero point on the y axis corresponds to the tongue height of the first segment in each MTD sequence. We regard a trough as occurring when the first MTD is negative and the second is positive. The error bars show ±1.0 SD.

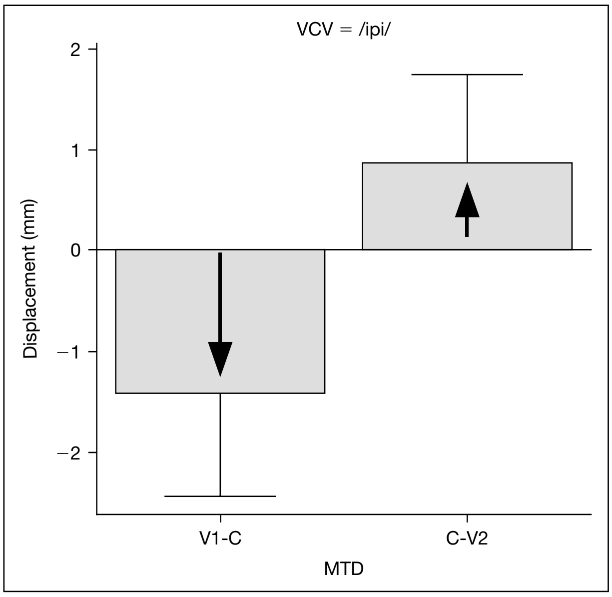

Fig. 5. Example of the representation of tongue movement from vowel 1 to closure (V1-C) and from closure to vowel 2 (C-V2) for /ipi/ in millimetres. These two measurements will be referred to as measures of tongue displacement (MTD).

Results

Quantitative Analysis of Tongue Displacement

Nonsense Words

Measurements of tongue displacement from V_1 to C and from C to V_2 were made for the nonsense words set, for the two sessions. A by-subjects four-way repeated measures ANOVA was performed on the means across the five repetitions of each speech item, crossed by the tongue displacement and grouped by session, vowel quality and voicing type. Greenhouse-Geisser correction, an adjustment used in univariate repeated measures, was applied when the sphericity assumption was violated. This correction attempts to adjust the degrees of freedom in the ANOVA test in order to produce a more accurate significance (p) value. Figure 6 shows the means for the tongue displacement per vowel quality, bilabial consonant (/b/ vs. /p/) and measurement (V1 to C vs. C to V2) across sessions. As can be observed from the means presented in figure 6, the nonsense words containing the vowel /a/ behave very differently from the nonsense words containing /i/ or /u/.

The repeated measures ANOVA results are summarized in table 1. As the nonsense words containing the /a/ vowel were behaving differently to the other two vowels, we removed this vowel from the statistical model illustrated in table 1 in an attempt to see how much of an influence the /a/ vowel was having on the overall results. Results showed that the interactions bilabial × MTD, vowel × MTD and bilabial × vowels × MTD stopped being significant, suggesting that the interactions were purely based on the different behaviour of the tokens containing the /a/ vowel.

There was a significant effect for the interaction sessions × MTD [$F(1, 9) = 9.136$, $p < 0.05$]. However, after plotting the data, it could be observed (fig. 7) that the interaction was caused by a reduction in the amount of tongue displacement for both types of

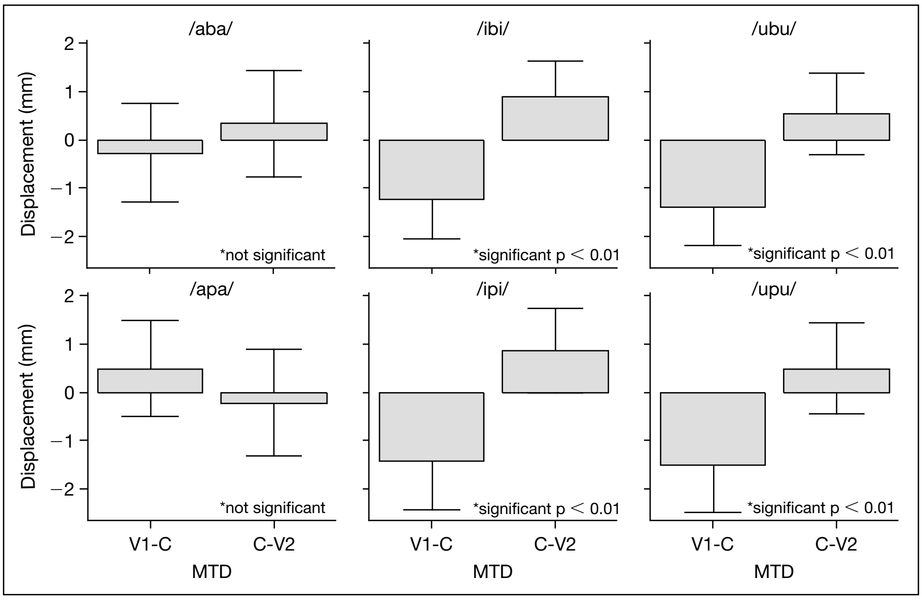

Fig. 6. Overall means for tongue displacement per VCV across sessions for the nonsense words. The bars show MTD means and the error bars show mean ± 1.0 SD. Significance is shown for post hoc paired t-test with Bonferroni correction.

Table 1. Table illustrating repeated measures ANOVA results for the nonsense words

Factor	F	d.f.	p value	Significance
Sessions	1.030	(1, 9)	0.337	
Bilabial	0.251	(1, 9)	0.628	
Vowels	10.676	(2, 18)	0.006[a]	**
MTD	28.202	(1, 9)	0.000	***
Sessions × bilabial	0.186	(1, 9)	0.667	
Sessions × vowels	3.340	(2, 18)	0.058	
Bilabial × vowels	2.873	(2, 18)	0.083	
Sessions × bilabial × vowels	1.279	(2, 18)	0.303	
Sessions × MTD	9.136	(1, 9)	0.014	*
Bilabial × MTD	6.007	(1, 9)	0.037	*
Sessions × bilabial × MTD	0.220	(1, 9)	0.650	
Vowels × MTD	25.087	(2, 18)	0.000	***
Sessions × vowels × MTD	1.102	(2, 18)	0.353	
Bilabial × vowels × MTD	16.974	(2, 18)	0.000	***
Sessions × bilabial × vowels × MTD	0.874	(2, 18)	0.434	

[a]Greenhouse-Geisser correction. Significance: ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.

measurements from session 1 to session 2. This result might be due to the subjects' familiarity with the experimental materials resulting in less carefully articulated second vowel.

There was also a significant effect for MTD [$F(1, 9) = 28.202$, $p < 0.001$] and for all three vowels [$F(1.22, 11.00) = 10.676$, $p < 0.01$] with Greenhouse-Geisser correction

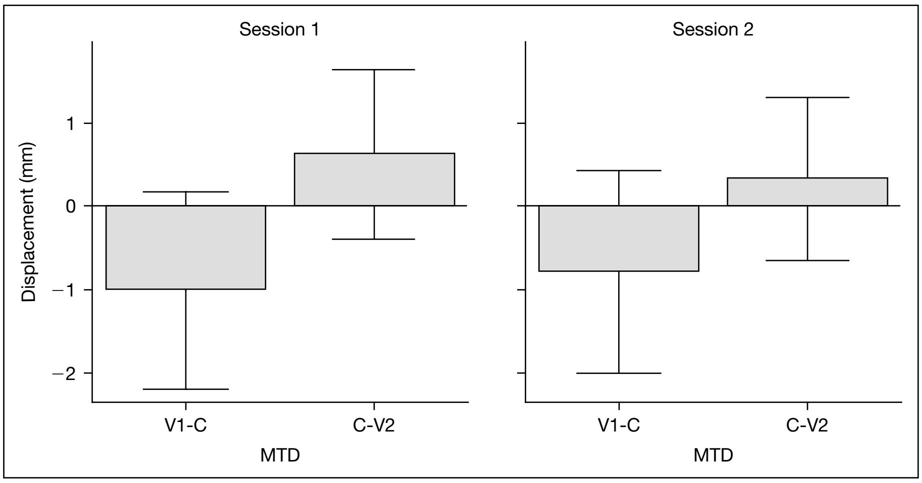

Fig. 7. Overall means for tongue displacement per measure and sessions for the nonsense words. The bars show MTD means and the error bars show mean ± 1.0 SD.

for the latter. Post hoc (paired samples t-test) tests for MTD showed no significant difference between the two tongue displacement measurements for the /aba/ and /apa/ VCVs. So, no trough effect was present in the /a/ VCVs. On the other hand, a significant difference between the two MTDs was found for /ipi/ [t(9) = −8.295, p < 0.010], /ibi/ [t(9) = −9.774, p < 0.010], /upu/ [t(9) = −6.059, p < 0.010] and /ubu/ [t(9) = −8.415, p < 0.010] with Bonferroni correction, where the first measurement mean was negative and the second was positive. So, a strong trough effect was present for both /i/ and /u/ VCVs. For vowel quality, paired samples t-tests revealed a significant difference between the vowels /i/ and /u/ for the closure-to-V_2 measurement in both /p/ and /b/ context [t(9) = 2.263, p = 0.050] and [t(9) = 2.361, p < 0.050, respectively]. However, this effect disappears once the Bonferroni correction is applied. There was no significant effect of voicing type.

Real Word Data

Measurements of the tongue displacement from V_1 to C and from C to V_2 were also made for the real words and compared with the findings for the nonsense words. Unfortunately, despite our intention to elicit a full /a/ vowel, too many subjects produced the second vowel of the relevant VCV sequence with a schwa-like quality. Therefore, the real word materials containing the /a/ vowel were not included in the analysis for this section. However, we do present results for /i/ and /u/ VCVs.

A by-subjects three-way repeated measures ANOVA was performed on the means across the five repetitions of each speech item, crossed by the tongue displacement and grouped by vowel and voicing type. Only the data from the first session was analysed for the real words. The real word data was used as a control for repeatability and given that the results from the first session showed that the trough was not only present in nonsense words, it was decided not to analyse the real word data from the second session. Figure 8 shows the means for the tongue displacement per vowel quality, bilabial consonant and measurement.

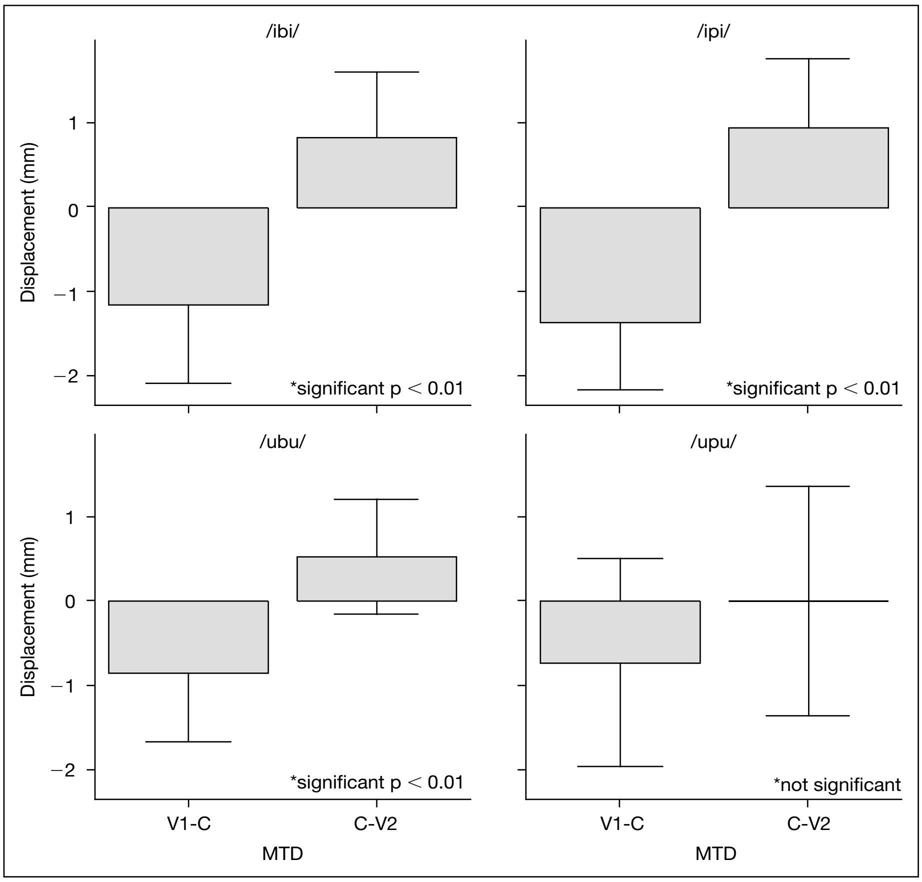

Fig. 8. Overall means for tongue displacement per VCV for the real words. The bars show MTD means and the error bars show mean ± 1.0 SD. Significance is shown for post hoc paired t-test with Bonferroni correction.

The repeated measures ANOVA results are summarized in table 2. Results show a significant main effect for MTD [$F(1, 9) = 77.368$, $p < 0.001$]. Post hoc (paired samples t-test) tests revealed a significant difference between the two MTDs, where the first measurement mean was negative and the second was positive, for /ipi/ [$t(9) = -7.558$, $p < 0.010$], /ibi/ [$t(9) = -5.485$, $p < 0.010$] and /ubu/ [$t(9) = -5.618$, $p < 0.010$] with Bonferroni correction. So, a strong trough effect was present for both /i/ and /u/ VCVs in real words, confirming the results for the nonsense words. However, paired t-test did not show a significant difference between measurements in the /upu/ VCVs in 'I said two pools too'. This result could be explained by the presence of a syllable boundary between V_1 and the bilabial plosive. Modarresi [2002] has shown that a syllable boundary has an effect on the trough effect. Also, unlike the other examples in the real word data set, both 'two' and 'pools' bear full lexical stress, which could have had an effect on vowel tongue height. A lexical stress effect together with the coarticulation of the back vowel /u/ with the lateral /l/ at the CV boundary could have prevented the

Table 2. Table illustrating repeated measures ANOVA results for the real words

Factor	F	d.f.	p value	Significance
Vowels	0.388	(1, 9)	0.549	
Bilabial	0.937	(1, 9)	0.358	
MTD	77.368	(1, 9)	0.000	***
Vowels × bilabial	0.263	(1, 9)	0.620	
Vowels × MTD	9.574	(1, 9)	0.013	*
Bilabial × MTD	0.187	(1, 9)	0.676	
Vowels × bilabial × MTD	7.284	(1, 9)	0.024	*

Significance: ***$p < 0.001$; *$p < 0.05$.

tongue from moving upwards following the consonant release. However, it is out of the scope of this study to examine this result in the real word data in more depth.

There was also a significant interaction of the type vowels × MTD [$F(1, 9) = 9.574$, $p < 0.050$] and a significant three-way interaction of the type vowels × bilabial × MTD [$F(1, 9) = 7.284$, $p < 0.050$] similar to the effects found in the nonsense word data. Post hoc (paired samples t-test) tests showed a significant difference between vowels /i/ and /u/ for the closure-to-V_2 measurement only for the /p/ context [$t(9) = 3.014$, $p = 0.015$]. However, this effect disappears once the Bonferroni correction is applied. Still, these interactions might be caused by the word boundary effects in the 'two pools' sequence. There was no significant effect of voicing type, confirming the results found in the nonsense words.

In summary, the results obtained from the real word data confirm the findings in the nonsense word data except for the /upu/ case, which we believe was caused by word boundary effects. A statistically significant difference between the two tongue displacement measurements showing a trough-like tongue movement in /ipi/, /ibi/, /upu/ and /ubu/ VCVs that behaved the same for the two different vowels and the two different bilabial plosives.

Qualitative Analysis of Tongue Displacement: Inter-Subject Variation of the Trough Effect

The statistical analysis based on one point on the tongue contour presented in the previous section confirmed a statistically significant difference between the two tongue displacement measurements, showing a trough-like tongue movement both in the nonsense words and the real word data. In this section of the paper, we will provide a qualitative analysis of the different realisations of this tongue trough between subjects (see fig. 9 for an illustration of the tongue displacement per subject and nonsense word across sessions).

Cluster Analysis

A hierarchical cluster analysis was applied to the nonsense data across sessions in order to find out how subjects were grouped based on the similarity of their MTDs. This allowed us to observe similarities and differences between subjects. Dendrograms

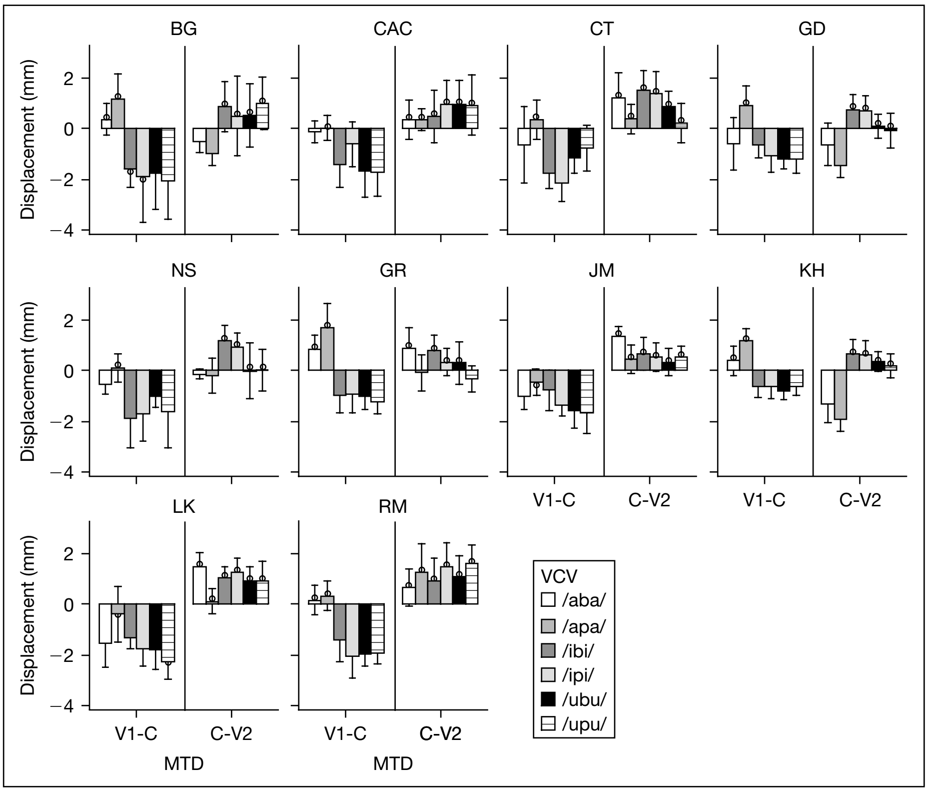

Fig. 9. Tongue displacement for all VCVs per subject across sessions. The bars show MTD means per subject and the error bars show mean ± 1.0 SD. For example, subject KH showed a rising tongue movement for /aba/ and /apa/ but a consistent lowering tongue movement for the V1-C MTD for /ibi/, /ipi/, /ubu/ and /upu/ VCVs. On the other hand, for the C-V2 MTD, we observed a consistent tongue lowering movement for the /aCa/ VCVs but a rising tongue movement from the middle of the consonant closure to V_2 for the /iCi/ and /uCu/ VCVs. This subject provides a good example of a trough-like tongue movement for the /iCi/ and /uCu/ VCVs and also an anti-trough tongue movement for the /aCa/ VCVs.

were generated with SPSS using Ward's cluster method (Euclidean distance) with standardized z scores, which is a type of analysis of variance to evaluate the differences between clusters. The resulting dendrograms showed subject groupings for the /a/, /i/ and /u/ VCVs separately.

Results for the /aCa/ VCVs are shown in figure 10. Given that there is no agreement on how to identify the number of clusters represented in a dendrogram, we selected the groups we were interested in by looking at the lower partitions of the forks showing homogeneous groups. For these data we identified two different clusters (fig. 10). If we look at the /aCa/ VCVs MTD averages for these subjects in figure 9, the first cluster differed from the second in that it grouped subjects that showed a trough-like tongue movement downwards from V_1 into the consonant closure and a following rising from the closure into V_2. On the other hand, the second cluster grouped subjects that showed

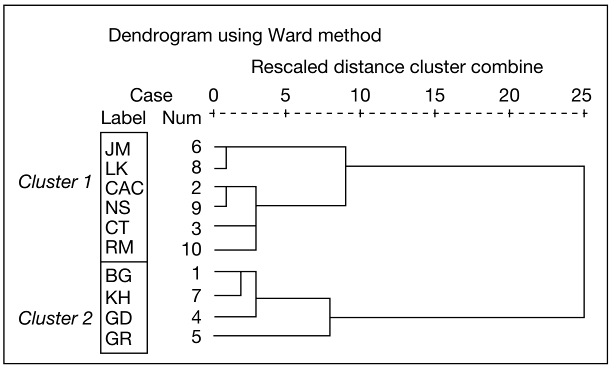

Fig. 10. Dendrogram from Ward's method (Euclidean distance) cluster analysis of /aCa/ VCVs (dimensions: MTD = V1-C, C-V2). The rectangular boxes on the dendrogram identify two different clusters. Cluster 1 includes subjects JM, LK, CAC, NS, CT and RM and cluster 2 includes subjects BG, KH, GD and GR

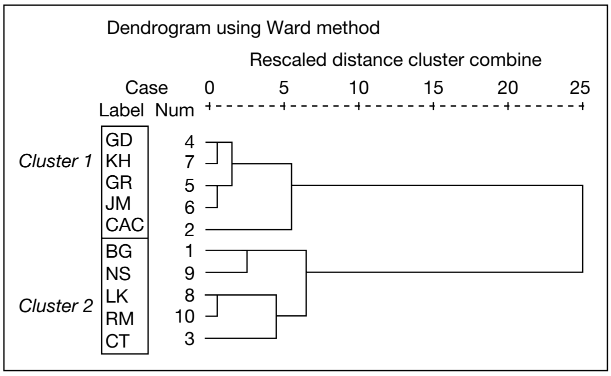

Fig. 11. Dendrogram from Ward's method (Euclidean distance) cluster analysis of /iCi/ VCVs (dimensions: MTD = V1-C, C-V2). The rectangular boxes on the dendrogram identify two different clusters. Cluster 1 includes subjects GD, KH, GR, JM and CAC and cluster 2 includes subjects BG, NS, LK, RM and CT

an anti-trough type of tongue movement where the tongue was higher during the consonant closure than during both V_1 and V_2.

Results for the /iCi/ VCVs are shown in figure 11. For these data, two different clusters were identified (fig. 11). All 10 subjects showed trough-like tongue movements for the /iCi/ VCVs, but they also showed differences in its realisation. If we look at the /iCi/ VCVs MTD averages in figure 9, the first cluster grouped subjects that showed a trough-like tongue movement, but the tongue displacement for both MTDs was smaller when compared to the tongue displacement observed for the subjects in the second cluster.

Results for the /uCu/ VCVs are shown in figure 12. For these data, two different groups were also identified (fig. 12). The first cluster grouped subjects that showed a tongue movement downwards from V_1 into the consonant closure but much variation in the direction of the tongue movement from the consonant closure into V_2. On the other hand, subjects in the second cluster showed a robust trough-like tongue movement with no variance in the tongue direction for MTDs.

Averaged Tongue Contours

In order to illustrate the whole tongue contour during the different symmetrical VCVs productions, we present in figure 13 a number of representative averaged tongue

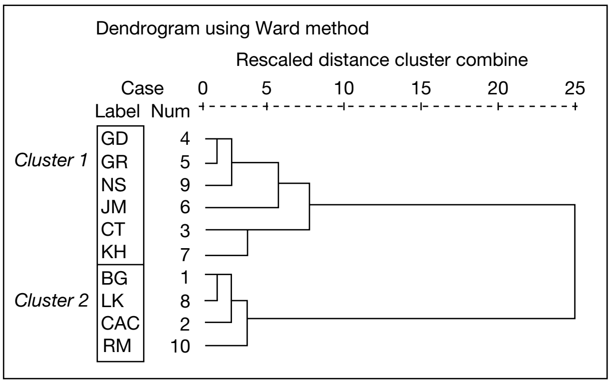

Fig. 12. Dendrogram from Ward's method (Euclidean distance) cluster analysis of /uCu/ VCVs (dimensions: MTD = V1-C, C-V2). The rectangular boxes on the dendrogram identify two different clusters. Cluster 1 includes subjects GD, GR, NS, JM, CT and KH and cluster 2 includes subjects BG, LK, CAC and RM

contour plots. A Matlab routine was used to plot tongue contours averaged across five repetitions for each vowel and for the voiceless bilabial plosive /p/ only. The tongue tip is always on the right. The subjects chosen for these illustrations were selected on the basis of the cluster groupings reported in the previous section. We chose representative subjects from the two clusters identified for each vowel.

By looking at the plots in figure 13, we can observe not only a big difference in the direction and degree of the tongue displacement, but also a big difference in the tongue contour configuration between subjects. For example, if we take subjects CT and BG for the /ipi/ VCV, we can observe that the direction and degree of the tongue displacement downwards during the consonant closure is very similar for both subjects, but the overall tongue configuration is very distinct. Whereas, if we compare subject CT with subject GR for the /ipi/ VCV, we can observe that the tongue configuration is very similar for both subjects, but the degree of the tongue displacement varies. Given the large sector width in the ultrasound image it is more likely that differences in tongue configuration were caused by the speakers using their tongues in different ways to produce the vowel than by anatomical differences between speakers.

Discussion and Conclusions

Ultrasound has proven to be an extremely useful technique in the investigation of the 'trough effect' carried out in this study. In combination with acoustic measurements, ultrasound has provided us with a large quantity of articulatory imaging data from a variety of different speakers that have been previously unavailable on this phenomenon. The non-invasive nature of this technique, the ability to capture tongue surface movement data synchronised with the acoustic signal and the relative simplicity of the experimental set-up makes ultrasound a very promising tool for speech research.

The results from this experiment support the existence of a trough effect, over a significant population sample, in symmetrical VCV sequences where the intervocalic consonant is a bilabial stop. However, the fact that in the VCVs where V = /a/ the tongue did not show a consistent raising movement during the consonant closure suggests that an explanation of the trough effect as being a tongue movement towards a neutral position, schwa-like, is unlikely.

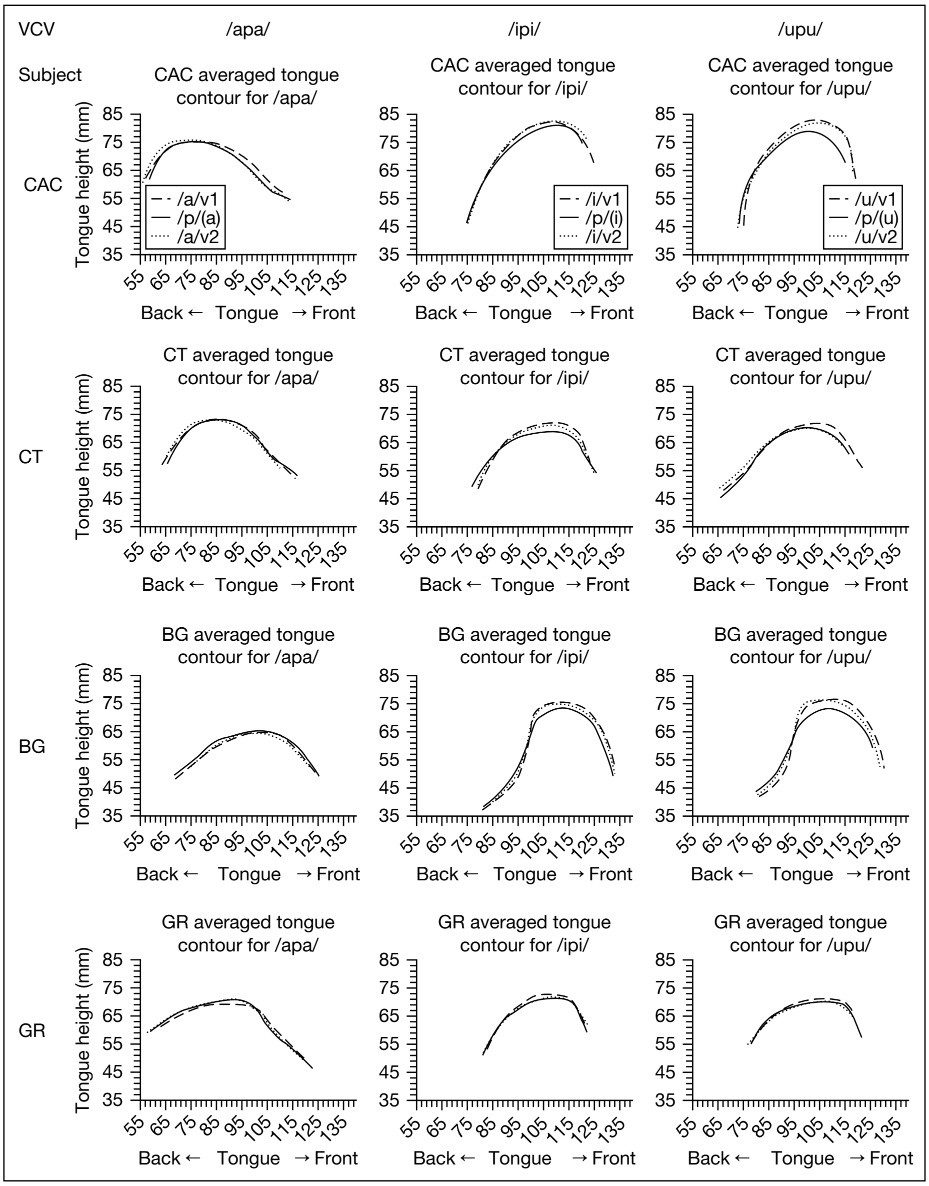

Fig. 13. Averaged tongue contours from subjects CAC, CT, BG and GR (all females) during the production of V_1, bilabial plosive and V_2 across five repetitions for /apa/, /ipi/ and /upu/ sequences. (Although the averaged tongue contours might give the appearance that they extend less far towards the tongue root for the /a/ contour than for the /i/ and the /u/, this is just a misapprehension caused by the much flatter tongue shape for /a/).

The data in this study have also shown quantitative and qualitative evidence of how much tongue displacement is present in a trough-like tongue movement. The extent of the tongue movement varied greatly both within and between subjects and per VCV sequence. Lindblom et al. [2002] reported a tongue displacement of 3 mm for only 1 subject, whereas Houde [1968] found a tongue height perturbation of 2.5 mm, again for only 1 subject. In this study, an average of 1.5 mm tongue displacement during the bilabial closure could be observed for the /iCi/ and /uCu/ sequences over a total of 10 subjects. However, 7 subjects out of 10 showed tongue displacements of more than 3 mm for /iCi/ and /uCu/ sequences. On the other hand, a few subjects did not show a trough-like tongue movement at all for some of the VCV sequences.

Averaged tongue contours showed not only a big difference in the direction and degree of the tongue displacement but also a big difference in the tongue contour configuration between subjects. This by-subject variation might have been caused by speaking styles or by physiological differences inherent to the subject.

Future work will investigate to what extent aerodynamic constraints are connected to the trough effect. We will look at whether the differences in intraoral pressure could account for the subject variation and if the motor planning system is anticipating the aerodynamic forces and planning movement trajectories to take advantage of the direction and magnitude of this aerodynamic force as suggested by Hoole et al. [1998]. This further work will help us ascertain whether there is room for an aerodynamic explanation of this phenomenon.

Acknowledgements

We are grateful to Alan Wrench, of Articulate Instruments, who designed the analysis software and to Bruce Davies, of Heriot-Watt University, who designed and adapted the helmet used in this investigation. We would also like to thank Harvey Sussman, Philip Hoole, John Ohala, Heriverto Avelino and Matthew Aylett for their many suggestions for improving this paper. The instrumentation was purchased and developed under a grant from the Science Research Investment Fund.

References

Bell-Berti, F.; Harris, K.: More on the motor organization of speech gestures. Haskins Lab. Status Rep. Speech Res., SR 37/38, pp. 73–77 (Haskins Laboratories, New Haven 1974).

Faber, A., Raphael, L.: Relationship of recorded EMG signals to within- and cross-utterance acoustic variations. J. acoust. Soc. Am. *85:* suppl. 1,p. S58 (1989).

Fuchs, S.; Hoole, P.; Brunner, J.; Inoue, M.: The trough effect – an aerodynamic phenomenon? In From sound to sense: MIT Meeting, 2004.

Gay, T.: Some electromyographic measures of coarticulation in VCV-utterances. Haskins Lab. Status Rep. Speech Res., SR 44, pp. 137–145 (Haskins Laboratories, New Haven 1975).

Hewlett, N., Beck, J.: An introduction to the science of phonetics (Erlbaum, Mahwah 2006).

Hoole, P.; Munhall, K.; Mooshammer, C.: Do airstream mechanisms influence tongue movement paths? Phonetica *55:* 131–146 (1998).

Hoole, P.; Nyguen, N.: Electromagnetic articulography in coarticulation research; in Hardcastle, Hewlett, Coarticulation: theory, data and techniques, pp. 260–269 (Cambridge University Press, Cambridge 1999).

Houde, R.A.: A study of tongue body motion during selected speech sounds. Speech Communications Res. Lab. Monogr., No. 2 (Santa Barbara, 1968).

Keller, E.: Ultrasound measurement of tongue dorsum movements in articulatory speech impairments; in Ryalls, Phonetic approaches to speech production in aphasia and related disorders, pp. 93–112 (College-Hill Press, San Diego 1987).

Lindblom, B.; Sussman, H.M.: Principal components analysis of tongue shapes in symmetrical VCV utterances. Q. Prog. Status Rep., Speech Transm. Lab., R. Inst. Technol., Stockh., Fonetik, vol. 44 (2002).

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

McAllister, R.; Engstrand, O.: Interpretations of tongue movement patterns in VCV sequences; in Huber, Papers from the Sixth Swedish Phonetics Conference held in Gothenburgh, May 20–22. Technical report No. 10, pp. 115–120 (Department of Information Theory, School of Electrical and Computer Engineering, Chalmers University of Technology, Gothenburg 1992).

Modarresi, G.: The effects of syllable boundary, stop consonant closure duration, and VOT on VCV coarticulation; doct. diss. University of Texas, Austin (2002).
Perlman, A.; Luschei, E.; Dumond, C.: Electrical activity from the superior pharyngeal constrictor during reflexive and nonreflexive tasks. J. Speech Hear. Res. *32:* 749–754 (1989).
Stone, M.: A guide to analyzing tongue motion from ultrasound images. Clin. Linguist. Phonet. *19:* 455–502 (2005).
Stone, M.; Davis, E.P.: A head and transducer support system for making ultrasound images of tongue/jaw movement. J. acoust. Soc. Am. *98:* 3107–3112 (1995).
Svirsky, M.; Stevens, K.; Matthies, M.; Manzella, J.; Perkell, J.; Willms-Tricarico, R.: Tongue surface displacement during bilabial stops. J. acoust. Soc. Am. *102:* 562–571 (1997).
Wells, J.C.: Accents of English 2: The British Isles (Cambridge University Press, Cambridge 1982).
Whalen, D.H.; Iskarous, K.; Tiede, M.K.; Ostry, D.J.; Lehnert-Lehouillier, H.; Vatikiotis-Bateson, E.; Hailey, D.S.: The Haskins optically corrected ultrasound system (HOCUS). J. Speech Lang. Hear. Res. *48:* 543–553 (2005).