AN ULTRASOUND PROTOCOL FOR COMPARING TONGUE CONTOURS: UPRIGHT VS SUPINE

Alan Wrench[a], *Joanne Cleland*[b] & *James M. Scobbie*[b]

[a]Articulate Instruments Ltd., UK; [b]Queen Margaret University, UK
awrench@articulateinstruments.com; jcleland@qmu.ac.uk; jscobbie@qmu.ac.uk

ABSTRACT

A study is described that employs ultrasound to measure the effects of gravity on production of vowels. The materials are designed to encourage consistent production over repetitions. A recording and analysis protocol is described which allows for correction for probe movement or rejection of data where correction is not possible. Results indicate a slight superior and posterior displacement of the tongue root in supine posture, consistent with a shift in the support structure of the tongue.

Keywords: tongue, ultrasound, upright, supine

1. INTRODUCTION

Magnetic Resonance Imaging (MRI) can provide a clear and detailed 2D image of the vocal tract articulators in the midsagittal plane. The more powerful 1.5T and 3T MRI systems require speakers to be recorded in a supine position. This is known to affect the speech production mechanism due to a change in the direction of gravitational force on the articulators [2] and it is important to establish how data recorded in this way might differ if it was acquired in the upright position.

A number of studies have sought to identify the effects of gravity on tongue shape. Engwall et al. studied MRI images of vowels from a single speaker in supine and prone posture, finding greater pharyngeal constriction in the supine condition.

Kitamura, et al. [3] record two speakers producing steady state vowels in isolation. MRI images show a general retraction in supine position. However, since there was no repetition of vowels within condition, it is not possible to be certain whether variation in shape was due to orientation or inconsistency of production. Inconsistency in production may occur when speakers are asked to produce vowels out of context.

Tiede, et al. [5] used x-ray microbeam (tongue tip, mid and dorsum coils) to study two speakers producing vowels, /bV/ syllables, and phrases, in upright and supine positions. One speaker had a slightly upward and sometimes anterior tongue when supine. The second speaker had a generally posterior tongue position when supine. However, the tongue root could not be observed.

Stone, et al. [4] studied real words “bang”, “golly” “dash” repeated 5 times in upright and supine conditions. They recorded 13 speakers. Posterior displacement of the tongue was observed in 7 of the speakers but the other speakers did not show this pattern. One possibility is that the speakers may have been compensating or overcompensating for the orientation change. It is also possible that some variance in the data could have been due to the protocol used to collect it. Correction for probe movement between conditions was based on a single palate trace. It is possible that there could have been probe shift between the time the palate trace was taken and the words recorded. The use of 30Hz video ultrasound signal could also introduce distortions into the image data due to discontinuities and motion blur. Finally, the use of kriging to extrapolate tongue contours at the root could introduce error. In view of these potential sources of variance, there is reason to revisit the question.

1.2. Aim

In this paper, ultrasound is again used to investigate the effect of an upright (U) versus supine (S) orientation on tongue shape and position. A protocol is used that attempts to control for production variability and measurement error.

2. METHOD

Data presented here is a subset of a larger corpus designed to look at the effects of gravity, and sustention on speech production in a wide range of consonants and vowels. Data for gravity, replication and the vowels /ɛ/ and /ɔ/ are presented here.

2.1. Speakers

Data was acquired from 6 female adult speakers. All participants were native speakers of Scottish or Irish-accented British English.

2.2. Materials

Two target words were used: pop /pɔp/ and pep /pɛp/. To limit coarticulatory effects from a lingual consonant, a pVp structure was used and the target words chosen because they are monophthongal in most accents of English and in all the speakers reported here, and sample a front and a back location. Distracters were peep, babe, pap, pope, poop, pip giving a range of vowels. Five repetitions of the two target words appeared in each condition: supine (S) and upright (U). They were randomized with one instance of each of the other words in each condition. The conditions appeared in two blocks, resulting in 10 supine and 10 upright productions of each target word.

2.3. Procedure

Ultrasound data was acquired using an Ultrasonix SonixRP machine remotely controlled via Ethernet from a PC running Articulate Assistant Advanced software™ [1]. The echo return data was recorded at 100fps with 76 beam-formed echo pulses evenly spread over a 112.5 degree field of view (FOV). A hardware pulse was generated by the SonixRP at the instant that each complete set of 76 echo pulses had been recorded. This synchronization pulse sequence was recorded on a multichannel analogue acquisition system at 22050kHZ along with the acoustic speech signal. The pulses were then detected in a post processing operation allowing each ultrasound frame to be accurately time tagged. A standard graphical interpolation is performed on the raw data to convert it to an image for analysis in AAA, similar to the image processing that is normally carried out within the ultrasound scanner (Fig. 1). The depth setting was 80mm and the echo return vectors had 412 discrete samples (providing approximately 5 pixels per mm). The transducer frequency was 5MHz providing an axial resolution of approximately 0.9mm.

Recordings were made in a sound-treated studio. Speakers were fitted with a headset (Fig. 2) to stabilize the ultrasound probe. Order of data acquisition was counterbalanced (U-S-U-S or S-U-S-U) between speakers. To determine whether there had been any movement of the headset and probe both within and across conditions, palate traces were obtained by asking speakers to press their tongues against the hard palate before and after each word. These were later overlaid on the tongue data enabling any rotation or translation of the probe-headset equipment relative to the head to be adjusted for.

Figure 1: Image reconstruction.

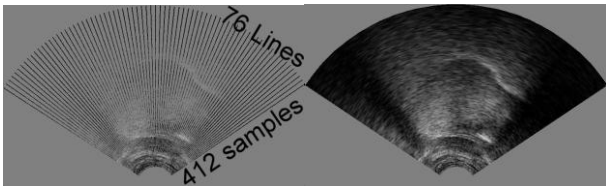


Figure 2: Speaker wearing headset in upright and supine orientations.

In addition, synchronous 60Hz de-interlaced NTSC video from a headset-mounted micro-camera, imaging a profile of the nose was used to verify that there was no movement of the probe relative to the head during speech.

2.4. Annotation

Vowels were annotated at their acoustic midpoint. For each vowel, the nearest ultrasound frame to the midpoint was selected and a spline indicating the tongue surface fitted to the image using the automatic function in AAA software [1]. Palate traces were also identified for each utterance and a spline fitted automatically.

The spline is defined by 42 control points, one on each of 42 equally spaced radial axes. An edge detection algorithm [1] is applied independently along each axis to determine the control point i.e. the point where the tongue contour crosses the axis. The algorithm generates a confidence level based on brightness and contrast of the detected edge on each of the 42 axes. Confidence is quantified, and indicated visually by fading the tongue contour line where confidence is low.

Figure 3: Speaker 1, pep.

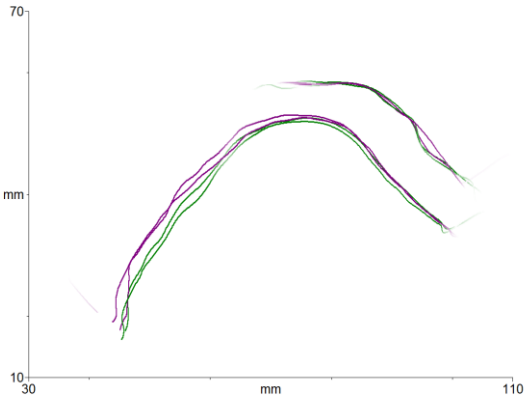


Figure 4: Speaker 2, pep.

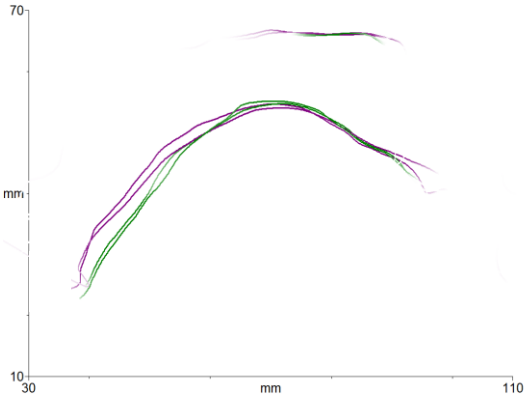


Figure 5: Speaker 3, pep

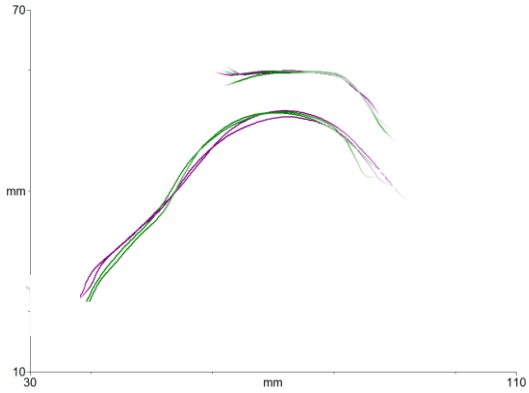


Figure 6: Speaker 4, pep.

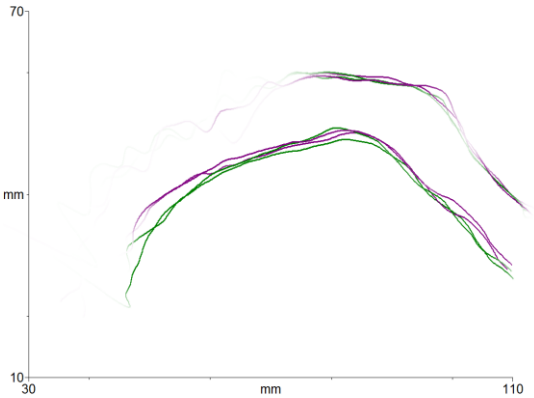


Figure 7: Speaker 1, pop.

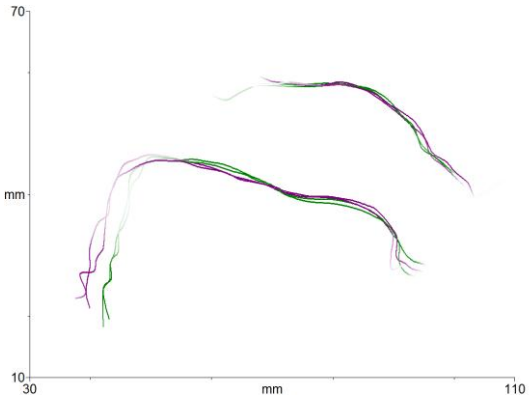


Figure 8: Speaker 2, pop.

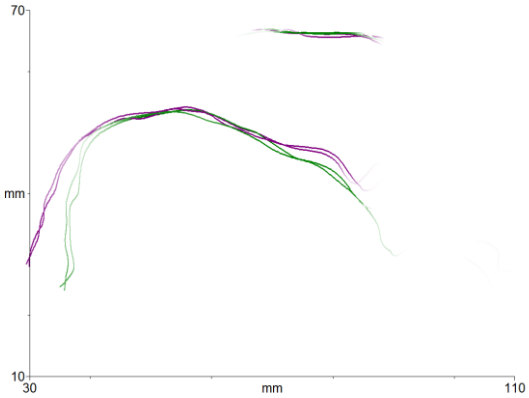


Figure 9: Speaker 3, pop.

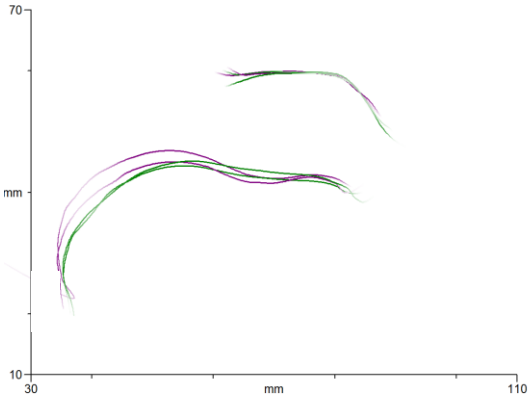
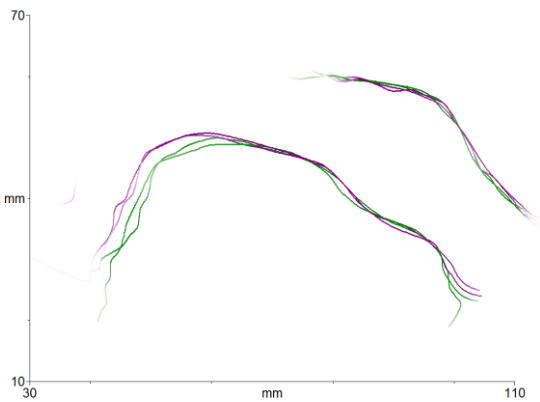


Figure 10: Speaker 4 pop.



3. RESULTS & DISCUSSION

Results for each vowel are presented separately. Figures 3 to 10 show mean tongue contours and palate traces for each set of five repetitions, with upright contours in green and supine contours in purple for each individual speaker. The figures therefore allow comparison both within conditions (i.e. across the U-U and S-S blocks) and across (U-S) conditions. Two of the six speakers did not manage to follow the instruction to place their tongue against their palate and so correction between conditions could not be performed and their data is not included here. A method of correcting for movement during speech, based on synchronous video images of the bridge of nose profile is being implemented to avoid this problem in future.

Examination of within-condition (same-coloured) splines shows minimal differences in tongue contours, suggesting that speakers were consistent across both blocks, for upright and supine conditions, with little vowel variation.

Across conditions, differences are seen in the posterior portion of the tongue in supine position, consistent with a gravitational effect.

All four speakers[i] exhibited displacement of tongue root between conditions for both vowels.

4. CONCLUSIONS

The protocol allowed for detection and rejection of data that exhibited movement of the probe relative to the head during speech by means of frequent palate traces and by synchronous video of the bridge of nose, referenced to the probe stabilization headset. The protocol additionally employed edge detection of the whole contour (no hand drawing and no extrapolation of contours at the root or tip). The images upon which the analysis was based were complete ultrasound scans at a high frame rate, minimizing motion blur and eliminating discontinuities that can appear in video port ultrasound. These factors improved consistency and fidelity of the measured contours.

The results presented here indicate that all four speakers have a slight superior and posterior displacement of the tongue root in supine position. This is possibly due to a change in the setting of the jaw, hyoid and larynx which must also be affected by the change in orientation and posture. A speaker then has to compensate not only for a new posterior force acting on the tongue mass but also a shift in the whole support structure of the tongue. One strategy to cope with these conditions could be to contract the geniohyoid. This could explain the better match for the vowel /**ɛ**/ where this muscle is invoked in any case to raise and elevate the tongue. A second strategy could be for the speaker to accept a more posterior position of the tongue body. This seems to be the strategy preferred for the vowel /**ɔ**/ where contracting the geniohyoid would make it difficult to maintain the tongue shape in the palatal region. Although, speaker 3 opted to preserve tongue body position in 3 out of 10 repetitions. Protrusion of the lips and lowering of the jaw could be used to compensate for a reduced anterior cavity when this strategy is adopted and it may be possible to use video footage to investigate this further. A raised velum is another possible compensatory strategy but this cannot be confirmed by observation with this protocol.

These findings in large measure, do not contradict Stone et al. but there are subtle differences. Results are more consistent between speakers and there was no evidence of over-compensation, but rather different strategies invoked to achieve the acoustic target.

5. ACKNOWLEDGEMENTS

This work was partly supported by an EPSRC grant (EP/I027696/1). Thanks to Steve Cowen for technical assistance.

6. REFERENCES

[1] *Articulate Assistant Advanced Ultrasound Module User Manual*, 2010. Revision 212, Articulate Instruments Ltd.

[2] Engwall, O. 2006. Assessing magnetic resonance imaging measurements: Effects of sustenation, gravitation, and coarticulation. In Harrington, J., Tabain, M., (eds.), *Speech Production: Models, Phonetic Processes, and Techniques*. Hove: Psychology Press, 301-314.

[3] Kitamura, T., et al. 2005. Difference in vocal tract shape between upright and supine postures: Observations by an open-type MRI scanner. *Acoustical Science and Technology* 26(5), 465-468.

[4] Stone, M., et al., 2007. Comparison of speech production in upright and supine position. *Journal of the Acoustical Society of America* 122(1), 532-541.

[5] Tiede, M., Masaki, S., Vatikiotis-Bateson, E. 2000. Contrasts in speech articulation observed in sitting and supine conditions. *Proc. 5th Seminar on Speech Production* Kloster Seeon, 25-28.

[i] Speaker 3 was re-recorded after analysis indicated that the probe moved during speech.