VELOCITY MEASURES IN ULTRASOUND DATA. GESTURAL TIMING OF POST-VOCALIC /l/ IN ENGLISH

Patrycja Strycharczuk and James M. Scobbie

CASL, Queen Margaret University
pstrycharczuk@gmail.com, jscobbie@qmu.ac.uk

ABSTRACT

We propose a new method for extracting dynamic information from midsagittal ultrasound images of tongue shape and location. The method is based on analysing parts of the tongue contour as they appear to travel up and down fan lines at key articulatory locations, representing velar and alveolar constrictions. The tongue displacement in these dimensions serves as the basis for calculating instantaneous velocity, whose maxima and minima help to define articulatory events. We validate the proposed method using data on /l/-darkening in Standard Southern British English. Our analysis extracts systematic information about the relative timing of gestures involved in the articulation of post-vocalic /l/, and it provides the basis for selecting representative images of consonantal constriction.

Keywords: ultrasound; tongue movement; kinematics; /l/-darkening; apparent time

1. INTRODUCTION

1.1. Ultrasound and articulatory dynamics

Ultrasound Tongue Imaging is a relatively new tool in the study of articulation, and the methodology of analysing ultrasound data is a very active area of research. Most ultrasound analyses so far have focused on comparing the tongue contour shape and location at selected time points. However, this static approach is limited in focusing on only one specific aspect of articulation; it provides potential insights into individual articulatory states, but not into dynamic characteristics of articulation. In an attempt to address this issue, a body of work is developing in which dynamic information is extracted from changes in entire ultrasound images, using pixel difference [14, 16], optical flow [15], or Principal Components Analysis of pixel intensity [5].

In this paper, we focus instead on methods inspired by traditional dynamic analysis of flesh points on the tongue surface familiar from studies using X-ray microbeam or electromagnetic articulography (EMA). If a mid-sagittal ultrasound image plane is processed on a frame-by-frame basis by tracking the tongue surface, and if linear dimensions are placed so that they lie orthogonal to major paths of constriction, movement of parts of the tongue surface along those dimensions can be quantified. We therefore track the displacement of the tongue surface along a line in a 2D image, located in a specified region of interest. Velocity is calculated based on this displacement, and it serves as the basis of dynamic analysis. The nature of tongue displacement as employed here is different from EMA, as we are not tracking any individual point on the tongue surface, but rather we are tracking movement in a vector to represent the dynamics of consonantal constriction. The aim of the present paper is to provide an evaluation of the approach we have developed. We test the method using gestural coordination data in the articulation of /l/ by speakers of Standard Southern British English (SSBE).

1.2. /l/-darkening in English

Traditional descriptions distinguish between two broadly defined /l/-variants in English dialects, i.e. light and dark /l/ [6, 21]. Previous articulatory research on English /l/ identifies two distinct gestures involved in /l/ production: an apical gesture (raising of the tongue tip towards the alveolar ridge), and a dorsal gesture (raising of the tongue dorsum towards the velum) [18, 19]. The darkening of /l/ has various articulatory correlates, including increased magnitude of the dorsal gesture [13, 18, 20], reduced apical gesture [8], and the relative timing between the apical the dorsal gesture. [19], using X-ray microbeam, find that in light /l/, the apical gesture generally precedes the dorsal gesture, whereas in dark /l/, the reverse is true. They furthermore observe that, in dark /l/, the degree of tip delay (time lag between the dorsal and the apical gesture) ranges from about 50 to above 100 ms, and that the tip delay is correlated with the strength of the boundary following /l/. For instance, more darkening is found phrase-finally (*Beel*) than word-finally (*Beel equates*), and there is more darkening word-finally than morpheme-finally (*beel-ing*), or morpheme-internally (*Mr Beelik*). Morphosyntactic effects on the degree of /l/-darkening have also been found in different dialects of English by two ultrasound studies [13, 20], based on the comparison of tongue shape at selected time points.

In this paper, we attempt to replicate previous findings on gestural phasing of /l/, using the timing of defined articulatory events. In addition, we use the articulatory targets defined by the dynamic landmarks as representative time points for the static analysis of tongue shape and location. We expect that the relative gestural timing and gestural magnitude will be correlated with the strength of the morphosyntactic boundary following /l/: the stronger the boundary, the more /l/-darkening. We also consider the degree of darkening in apparent time. Under the hypothesis that /l/-darkening is a change in progress in Southern English dialects, we expect to see more darkening in younger speakers, compared to older ones. Finally, we analyse the influence of the preceding segment on the degree of /l/-darkening, by comparing /l/ in two contexts, preceded by the /uː/ vowel (e.g. *fool#five*) and the /ʊ/ vowel (e.g. *pull#five*). We further discuss our rationale for including these two vowels in Section 4.

2. MATERIALS AND METHOD

2.1. Stimuli

The experimental stimuli included /l/ preceded by a high-back vowel (/uː/ or /ʊ/) in four different conditions: 1) morpheme-internal, e.g. *hula*, *bully*; 2) morpheme-final, e.g. *fool-ing*, *pull-ing*; 3) word-final pre-vocalic, e.g. *fool#it*, *pull#it*; 4) word-final pre-consonantal, e.g. *fool#five*, *pull#five*. Three different lexical items were used per every combination of vowel and condition. Non-lingual consonants, such as labials or /h/, were preferred preceding the /uːl/ or /ʊl/ sequence. If, due to lexical restrictions, lingual consonants had to be used, they were counterbalanced across the set. Lexical items with yod-insertion before /uː/, such as *mule*, were avoided. The test items were embedded within a fixed carrier phrase: *Say X five times.*

2.2. Speakers

The speakers were 3 older females (45, 55, 60) and 3 younger females (22, 23, 25). They had all been born and grown up in the South of England or Midlands. They were not aware of the purpose of the experiment. They were paid £10 for participation.

2.3. Procedure

Time-synchronised articulatory and audio data were collected in the experiment. Tongue movement data were captured using a high-speed Sonix RP ultrasound system (Frame Rate = 121.5 fps, Scanlines = 63, Pixels per Scanline = 412, Field of Vision = 134.9, Pixel offset = 51, Depth = 80 mm). The ultrasonic probe was positioned under the participant's chin and stabilised using a headset [1]. The audio data were captured using a lavalier Audio-Technica AT803 condenser microphone connected to a synchronisation unit [2]. The participants read four repetitions of the experimental material (96 test items altogether). In addition, each participant was recorded swallowing water, in order to image the hard palate, and biting on a piece of plastic (a bite plate), in order to image the occlusal plane [17].

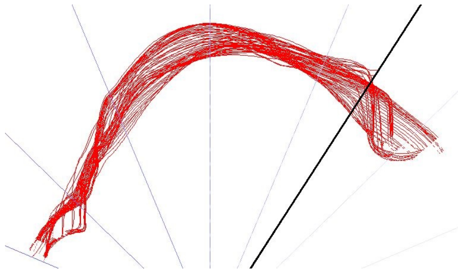
2.4. Analysis

The ultrasound data were analysed using Articulate Assistant Advanced software (AAA), version 2.16 [3]. For each test item, the tongue contour was traced using the tracking function embedded in AAA. The analysis was semi-automatic: the tongue contour tracing was manually initialised by a researcher, and subsequently tracked by the software under real time supervision, with a possibility for manual corrections. The ultrasonic frames included in the analysis were the frames corresponding to the acoustic region delimited by the onset of the consonant preceding /uː/ or /ʊ/ and the offset of the segment following /l/. For instance, in a word like *hula*, we traced the tongue contour from the acoustic onset of /h/ to the acoustic offset of the final vowel.

For each speaker, we analysed example tokens by plotting all the individual tongue contours in a single coordinate frame. The coordinate frame was a fan, centred at the origin of the ultrasound signal (the probe) and consisting of 42 equidistant radials, with a maximal angle of 67° relative to the centre of the probe. Tongue displacement during production of the vowel + /l/ sequences in representative tokens led us to identify two fan lines along which the relatively greatest tongue displacement occurred. This is illustrated in Figure 3, which shows the anterior fan line selected for speaker YF1.

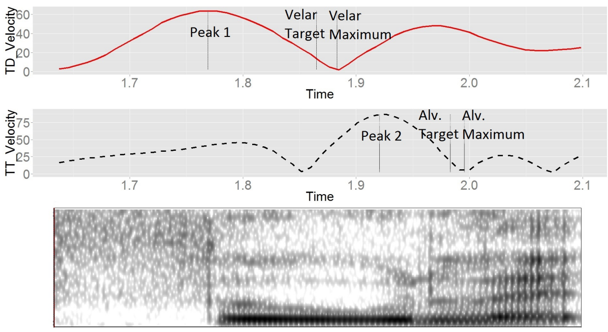
In AAA, we calculated tongue displacement and absolute velocity along the selected fan lines (Figure 2). We could identify two articulatory gestures involved in the production of /l/. We refer to them as 'velar gesture' and 'alveolar gesture' to highlight that they involve movement towards velar or alveolar constriction. The velar gesture occurs during

Figure 1: Selected fan line (11) for the analysis of apical movement for Speaker YF1



what is acoustically the preceding vowel. The articulatory maximum for this gesture occurs where the tongue dorsum velocity reaches its minimum. The trough in the velocity contour is preceded by a peak (Peak 1), a key identifier of movement towards the target. We defined the articulatory target as the point where the velocity reached 20% of the peak velocity. Similarly, the alveolar velocity contour shows a minimum where the tongue tip reached its maximal displacement, preceded by a velocity peak (Peak 2). We defined the alveolar target as the point where the alveolar velocity reached 20% of the peak velocity. We then analysed the timing of the velar gesture and the alveolar gesture based on an adapted version of the tip delay measure [19]. We define tip delay as the time lag between the achievement of the dorsal target and the apical target. We also extracted the coordinates of the tongue contour at velar and alveolar maxima for static analysis.

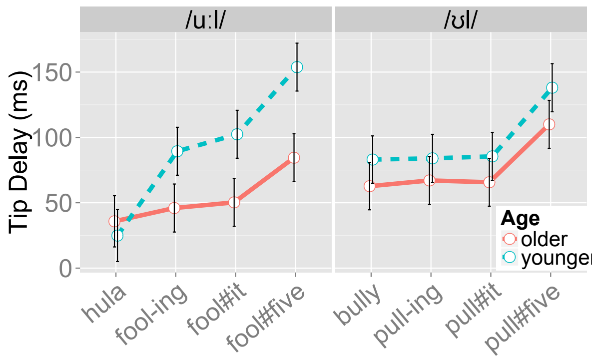
Figure 2: Example displacement and velocity data for Speaker YF1 pronouncing *fool#it*



3. RESULTS

We analysed the tip delay data using linear mixed effects regression in the R Package lme4 [4]. We used the following effect structure in the modelling: speaker and item were the random effects, whereas the fixed effects included speaker age (older vs. younger), preceding vowel (/uː/ vs. /ʊ/), and the morphosyntactic boundary following the /l/ (morpheme-internal, morpheme-final, word-final pre-vocalic, word-final). We calculated the *p*-values using the lmer test package [12]. The results show a significant ($p < 0.01$) three-way interaction between the three fixed predictors in the model, as illustrated in Figure 3. Generally, tip delay is greater (i.e. the /l/ is darker) for younger speakers, compared to older speakers. However, there was no age difference within morpheme-internal /l/ preceded by /uː/ (e.g. *hula*): no increased /l/-darkening was observed in apparent time for the *hula*-type context. We also observe more /l/-darkening preceding relatively stronger morphosyntactic boundaries, where /l/ was preceded by /uː/ (*hula<fool-ing<fool#it<fool#five*). This effect is more prominent in younger speakers, with relatively greater differences due to the morphosyntactic environment. This is consistent with the generalisation that morphosyntactic effects in /l/-darkening are a relatively recent innovation and they may be absent in more conservative speakers or dialects [20]. Finally, we observe that the morphosyntactic effect on the degree of /l/-darkening varies, depending on the vowel preceding /l/. Instances of /l/ preceded by /ʊ/ do not show the morphosyntactic gradience we observed for /l/ preceded by /uː/. While we do see increased tip delay in *pull#five*, there is no significant difference between any of the remaining conditions (*bully*, *pull-ing*, *pull#it*). For /l/ preceded by /ʊ/, we also see darkening in apparent time, with greater degree of tip delay for younger speakers, compared to older speakers, across all morphosyntactic conditions.

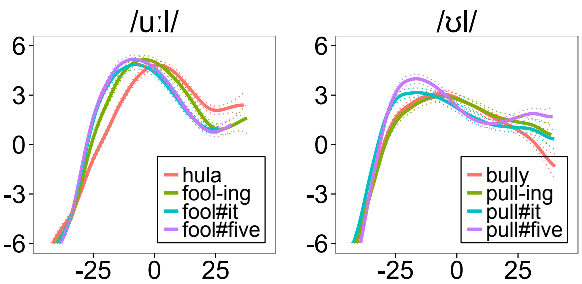
Figure 3: Results of a linear mixed-effects regression model of tip delay



We also find a morphological effect in a comparison of tongue contour location at constriction targets. As an example, consider the data from

Speaker YF1. Figure 4 shows the results of Smoothing Splines ANOVA [7], examining the morphological effect within this speaker in the two vocalic contexts. The tongue contours were extracted at the dorsal maximum for the /l/ and rotated in the occlusal plane [17]. In /l/ preceded by /uː/, we find a subset of the morphological contrasts previously identified based on gestural timing. The dorsal retraction increases in the direction: *hula*<*fool-ing*<*fool#it*, *fool#five*, with no significant difference in darkening between the last two conditions. In /l/ preceded by /ʊ/, there is no *bully*∼*pull-ing* contrast, but we find significantly more darkening in *pull#it* and *pull*. These results confirm the general morphosyntactic tendencies previously seen in the dynamic analysis, although they also show that individual speakers may have different levels of contrast than the population mean.

Figure 4: Results of SS-ANOVA on tongue contour at the dorsal maximum for speaker YF1, depending on the vocalic context. Non-overlapping confidence intervals indicate a significant difference.



4. DISCUSSION AND CONCLUSIONS

To a large extent, our findings are as predicted based on existing literature, which helps validate our method. The degree of tip delay increases in the word-final condition where /l/-darkening is expected (*fool#five*, *pull#five*), and the size of the effect is around 100 ms, similar to [19]. We see intermediate /l/-darkening, depending on the morphological environment: tip delay is smallest in word-medial intervocalic /l/ (*hula*), larger in stem-final and word-final intervocalic /l/ (*fool-ing* and *fool#it*) and largest in word-final pre-consonantal /l/ (*fool#five*). We can conclude that our measure of tip delay provides a feasible way of analysing the inter-gestural timing in /l/-production which corresponds to darkening. However, we must also consider the less expected results in our data, i.e. the interaction between the preceding vowel and the morphological boundary.

Our data show a morphological effect on tip delay when /l/ is preceded by /uː/, but not when /l/ is preceded by /ʊ/. This generalisation is also supported by the tongue contour data from speaker YF1, who shows a contrast in darkening between *hula* and *fool-ing*, but not between *bully* and *pull-ing*. Such an interaction is not expected under the hypothesis that /l/-darkening is directly conditioned by the presence of a specific morphosyntactic boundary, as proposed by [13] and [20]. However, if we assume that /l/-darkening can be modulated by coarticulation to the preceding vowel, the observed vowel effect can be attributed to ongoing sound change in SSBE /uː/ and /ʊ/. Both /uː/ and /ʊ/ are currently undergoing fronting in SSE [9, 10, 11], but a back variant of the vowel is retained before /l/ (*fool* and *pull*). Note that this contextual realisation of darkening can be interpreted in two ways: 1) the back vowel variant occurs in the context of a following dark [ɫ], or 2) the back vowel variant occurs in the context of /l/ in the same syllable rime. The latter interpretation allows the vowel fronting/backing to apply in different morphosyntactic domains, with the possibility that /uː/-fronting may be blocked before /l/ in the context of a morpheme boundary (*fool-ing*), whereas no such similar blocking occurs for /ʊ/, as in *pull-ing*. A difference in this direction is also consistent with the hypothesis that /ʊ/ fronting is a relatively more recent innovation, where the front and back allophones have not yet become separate allophones [11].

Gestural timing (Figure 3) and gestural magnitude (Figure 4) may cue different information, so combined evidence from the two analyses provides a more complete picture of significant factors and contrasts involved in /l/-darkening. In the dynamic analysis, it is easier to aggregate data from different speakers, allowing for across-speaker comparisons. The tongue contour data, on the other hand, provide additional information on factors which vary within speaker. One of the advantages of our method is that it links the two types of analysis in a principled way, using dynamic information to select a representative set of frames for static comparison.

Unlike automated ways of analysing dynamics in ultrasound data [5, 14, 15, 16], our method relies on extracted tongue contour data. The success of the semi-automatic contour tracker we used depends on good quality ultrasound image, which is a potential limitation. An advantage of preserving tongue contour data, however, is that velocity measurements can be readily linked to specific regions on the tongue, allowing for a relatively straightforward interpretation of the data.

5. ACKNOWLEDGMENTS

This study was supported by a British Academy Postdoctoral Fellowship (PDF/pf130029) to the first author. We wish to thank the experiment participants, and our colleagues at QMU, especially Steve Cowen, Conny Heyde, Eleanor Lawson and Alan Wrench.

6. REFERENCES

[1] Articulate Instruments Ltd, 2008. Ultrasound stabilisation headset users manual, revision 1.4.

[2] Articulate Instruments Ltd, 2010. SyncBrightUp users manual: Revision 1.10.

[3] Articulate Instruments Ltd, 2014. Articulate Assistant Advanced ultrasound module user manual, revision 2.16.

[4] Bates, D., Maechler, M. 2009. *lme4: Linear mixed-effects models using S4 classes*. Accessed 13 August 2010. R package version 0.999375-32.

[5] Carignan, C. 2014. TRACTUS (Temporally Resolved Articulatory Configuration Tracking of Ultrasound) software suite.

[6] Cruttenden, A. 2001. *Gimson's pronunciation of English*. London: Arnold.

[7] Davidson, L. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *The Journal of the Acoustical Society of America* 120, 407–415.

[8] Giles, S. B., Moll, K. L. 1975. Cinefluorographic study of selected allophones of English /l/. *Phonetica* 31, 206–227.

[9] Harrington, J., Kleber, F., Reubold, U. 2008. Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *The Journal of the Acoustical Society of America* 123, 2825–2835.

[10] Harrington, J., Kleber, F., Reubold, U. 2011. The contributions of the lips and the tongue to the diachronic fronting of high back vowels in Standard Southern British English. *Journal of the International Phonetic Association* 41, 137–156.

[11] Kleber, F., Harrington, J., Reubold, U. 2011. The relationship between the perception and production of coarticulation during a sound change in progress. *Language and Speech* 55, 383–405.

[12] Kuznetsova, A., Bruun Brockhoff, P., Haubo Bojesen Christensen, R. 2013. lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R package version 2.0-0.

[13] Lee-Kim, S.-I., Davidson, L., Hwang, S. 2013. Morphological effects on the darkness of English intervocalic /l/. *Laboratory Phonology* 4, 475–511.

[14] McMillan, C. T., Corley, M. 2010. Cascading influences on the production of speech: Evidence from articulation. *Cognition* 117, 243–260.

[15] Moisik, S. R., Lin, H., Esling, J. H. 2014. A study of laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound (SLLUS). *Journal of the International Phonetic Association* 44, 21–58.

[16] Palo, P., Schaeffler, S., Scobbie, J. M. 2014. Pre-speech tongue movements recorded with ultrasound. *Proceedings of the 10th International Seminar of Speech Production (ISSP 10)* 300–303.

[17] Scobbie, J. M., Lawson, E., Cowen, S., Cleland, J., Wrench, A. A. 2011. A common co-ordinate system for mid-sagittal articulatory measurement. QMU CASL Working Papers WP-20.

[18] Scobbie, J. M., Pouplier, M. 2010. The role of syllable structure in external sandhi: An EPG study of vocalisation and retraction in word-final English /l/. *Journal of Phonetics* 38, 240–259.

[19] Sproat, R., Fujimura, O. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21, 291–311.

[20] Turton, D. 2014. *Variation in English /l/: Synchronic reflections of the life cycle of phonological processes*. PhD thesis University of Manchester.

[21] Wells, J. 1982. *Accents of English. 3 vols*. Cambridge: Cambridge University Press.