AN ULTRASOUND ANALYSIS OF LOW BACK VOWEL FRONTING IN THE NORTHERN CITIES VOWEL SHIFT

Jonathan Havenhill

Georgetown University
jeh241@georgetown.edu

ABSTRACT

While the acoustic characteristics of the Northern Cities Vowel Shift (NCVS) are well documented, research on the articulatory components of this shift is comparatively limited. This study combines acoustic, video, and ultrasound analysis to examine the productions of seven Metro Detroit speakers and determine the relative contributions of lip configuration and tongue position to the production of fronted /ɑ/ and /ɔ/. NCVS speakers are found to exhibit variation with regard to how this change is achieved articulatorily. While some speakers distinguish /ɔ/ from /ɑ/ with a combination of tongue position and lip rounding, others do so using either tongue position or lip rounding alone. For speakers who maintain the contrast with only one articulatory gesture, /ɑ/ and /ɔ/ are acoustically more similar than for speakers who use multiple gestures.

Keywords: ultrasound, video analysis, articulation, sociophonetics, Northern Cities Vowel Shift

1. INTRODUCTION

Like the majority of sociophonetic phenomena, the Northern Cities Vowel Shift (NCVS) is typically described in terms of its acoustic characteristics; the articulatory mechanisms underlying this shift have remained largely unstudied. The present study considers the NCVS in terms of both its acoustic and its articulatory components. Specifically, this study uses a combination of ultrasound, video, and acoustic measure to investigate the fronting of the low back vowels /ɑ/ and /ɔ/, a change which is described as an increase in the second formant. However, while an increase in F2 is strongly correlated with tongue fronting, other articulatory gestures, such as lip rounding, can also influence the realization of F2.

It is found that Metro Detroit speakers differ in their articulatory realizations of /ɑ/ and /ɔ/. While some speakers produce fronted /ɔ/ with a combination of tongue-fronting and lip-unrounding, others do so using either tongue-fronting or lip-unrounding alone. For speakers who maintain the contrast through only one articulatory gesture, the acoustic differences between /ɑ/ and /ɔ/ are smaller than for speakers who use multiple gestures.

2. PREVIOUS LITERATURE

Descriptions of the NCVS have been based almost exclusively on acoustic measurement, with the exception of a study by Plichta [24], who used nasal/oral airflow measurement to investigate the effect of nasalization on /æ/-raising in the NCVS. He suggests that acoustically-raised /æ/ may be an artifact of the high degree of nasal airflow found for Northern Cities speakers in both nasal and non-nasal environments.

In the past several years, sociophonetic inquiry has seen an increase in the number of studies incorporating articulatory analysis, several of which inform the present study [11, 7, 14]. Most directly related to the present investigation is a study by Majors and Gordon [18], who use video recording to perform an analysis of lip-unrounding in two speakers from St. Louis, where the NCVS is in effect to some extent. They find that /ɔ/ can be acoustically fronted while maintaining its rounding, suggesting that fronting and lowering of /ɔ/ may be accomplished through tongue position alone.

However, as Labov et al. [13] note, St. Louis is the least consistent of the Inland North cities in terms of the number of NCVS-related changes and the number of speakers exhibiting these shifts, while Detroit and Chicago are the most consistent. This provides strong motivation for conducting a similar study on speakers from other NCVS cities, such as Detroit. In addition, because video analysis only allows for measurement of labial articulation, any conclusions drawn by Majors and Gordon relating to lingual articulation are necessarily speculative. Augmenting video analysis with ultrasound tongue imaging allows for simultaneous consideration of both labial and lingual articulation.

The change in question here, that of /ɔ/ and /ɑ/-fronting, is typically described as an increase in the second formant. However, it is well known in the phonetics literature that the value of F2 is influenced by a number of articulatory gestures, including both lip rounding and tongue position [3]. It is reasonable to hypothesize that language learners might adopt differing articulatory configurations when acquiring this vowel system. Either the tongue position for /ɔ/ may move forward, approaching that of /ɑ/, while the lips remain round, or /ɔ/ may lose its lip rounding, both of which would increase the value of F2 for /ɔ/. Alternately, a combination of these two configurations may be employed.

The broadly-defined goal of this study is to investigate the strategies that Northern Cities speakers employ in low back vowel fronting. More specifically, this paper seeks to determine a) whether speakers of the Northern Cities dialect differ in the way they achieve fronting of the low back vowels, b) whether interspeaker articulatory variation is observable in the acoustic signal, and c) what effects this sort of variation may have on the progression of the Northern Cities Shift.

3. METHODOLOGY

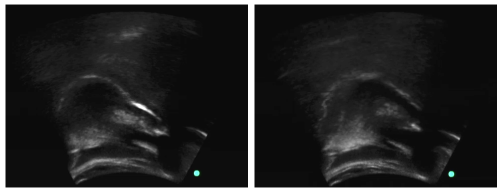
This study combines acoustic analysis with ultrasound imaging of the tongue surface and video recording of lip configuration. These combined methodologies facilitate analysis of the relative contributions of both lip rounding and tongue position to an increase in F2 for the low back vowels.

Data for this study were collected from seven speakers, comprising four males (ages 24–29) and three females (ages 22, 23, and 39). All speakers were born and raised in Metro Detroit, having lived there until at least age 18. For the purposes of this study, Metro Detroit is defined as the Detroit-Ann Arbor-Flint Combined Statistical Area. One speaker was raised outside this area in neighboring Jackson County.

Tokens were elicited from a wordlist containing 100 monosyllabic words, consisting of 20 words for each of the vowels /i/, /u/, /æ/, /ɑ/, and /ɔ/. /ɑ/ and /ɔ/ were the target vowels, while /i/, /u/, and /æ/ were included to serve as reference points for lip spread, lip protrusion, and participation in the NCVS, respectively. Words were embedded in the carrier phrase "say ____ again" and were presented pseudo-randomly in five blocks of 20 words.

Data collection took place in a sound-attenuated booth at Georgetown University. Audio was recorded using a Shure SM48S cardoid microphone and an Olympus LS-100 digital recorder. Video was recorded at a resolution of 1920×1080 pixels at 30 frames per second using a Canon XA10 camcorder. Ultrasound data were captured using a SonoSite M-Turbo ultrasound machine with a C60x 5–2 MHz transducer at a depth of 9.2 cm. Ultrasound data were recorded at a rate of 30 fps with synchronized audio using an Elgato Video Capture device.

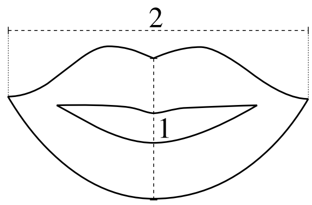
Figure 1: Extracted ultrasound frames for the tokens *caught* (left) and *odd* (right).



Vowel formants were measured in accordance with the techniques used in the *Atlas of North American English* (ANAE) [13]. Measurements were taken in Praat [4] at the point of F1 maximum, with the exception of /æ/, which was measured at the point of F2 maximum. Formant measurements were normalized using the *ANAE* log-mean normalization formula with the *ANAE* grand mean of 6.896874.

Ultrasound frames corresponding to the vowel measurement points were captured and imported into EdgeTrak [17], which was used to extract tongue contour data. Tongue contours were analyzed using smoothing spline analysis of variance (SS ANOVA), which is a statistical method for determining whether significant differences exist between best-fit smoothing splines for two or more sets of data. SS ANOVA is described by Gu [9] and has been used in linguistic research to analyze both ultrasound tongue contour data [6, 5, 7, 15, 16], and formant measurements [1, 20, 8].

Figure 2: Measurement points for vertical lip openness (1) and horizontal lip spread (2).

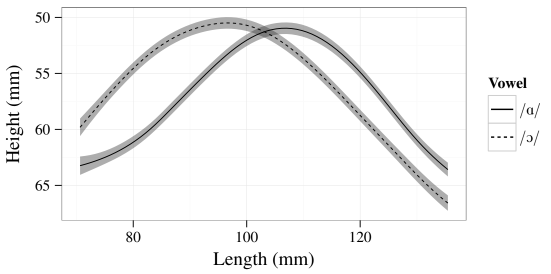


Extracted video frames (also corresponding to the vowel measurement points) were analyzed using the vector graphics editor Inkscape. Calibration frames containing a metric ruler were used to calculate the number of pixels per centimeter. Paths corresponding to the measurement points shown in Figure 2 were drawn. The vertical and horizontal measurements were extracted in pixels and converted to centimeters using the predetermined conversion ratio.

4. RESULTS

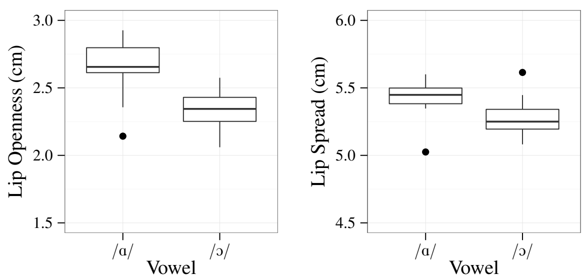
Three patterns of articulation are observed. First, /ɑ/ and /ɔ/ may differ in both lip rounding and tongue position. This strategy is found for Speaker 1, a 24 year-old male, Speaker 2, a 26 year-old male, Speaker 3, a 29 year-old male, and Speaker 4, a 39 year-old female. Smoothing splines for /ɑ/ and /ɔ/ as produced by Speaker 2 are presented with 95% confidence intervals in Figure 3. As in the ultrasound images in Figure 1, the right side of the image corresponds to the tongue blade, while the left side corresponds to the tongue root. Here, the constriction for /ɑ/ is higher than for /ɔ/, but /ɔ/ exhibits a greater degree of pharyngeal constriction. Where the confidence intervals overlap, the difference between the contours is not significant.

Figure 3: Tongue contours for /ɑ/ and /ɔ/, Speaker 2.



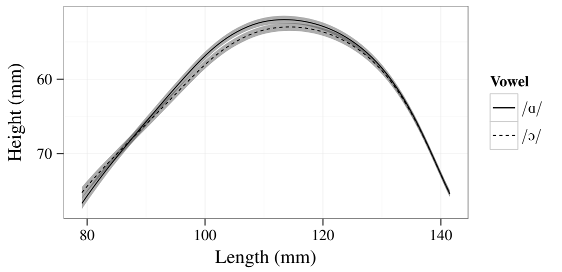
Lip rounding measurements for Speaker 2 are shown in Figure 4. For both vertical lip openness and horizontal lip spread, a smaller value indicates a greater degree of lip rounding. A one-way ANOVA test was performed for each measure. For this speaker, vowel class is a significant predictor of both lip openness ($F_{4,95}$ = 74.5, $p < 0.001$) and lip spread ($F_{4,95}$ = 63.4, $p < 0.001$). The difference between /ɑ/ and /ɔ/ in both lip spread and lip openness is significant, as revealed by a Tukey post hoc test. For Speakers 1, 3, and 4, /ɑ/ and /ɔ/ differ significantly in lip openness, but not in lip spread.

Figure 4: Lip measurements for Speaker 2.



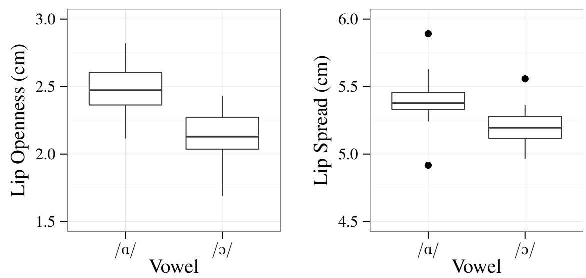
For Speaker 5, a 22 year-old female, and Speaker 6, a 26 year-old male, /ɑ/ and /ɔ/ differ in lip rounding, but not in tongue position. Smoothing spline estimates for /ɑ/ and /ɔ/, as produced by Speaker 5, are presented in Figure 5. Except for a small region near the tongue dorsum, the smoothing splines for /ɑ/ and /ɔ/ do not differ significantly.

Figure 5: Tongue contours for /ɑ/ and /ɔ/, Speaker 5.



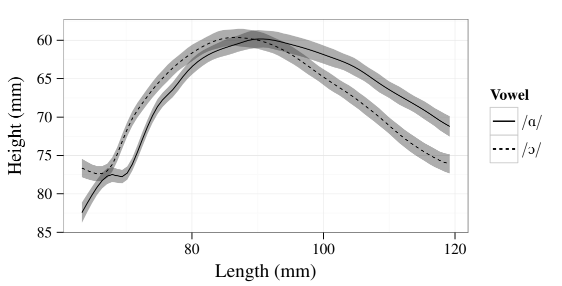
However, this speaker maintains a contrast between /ɑ/ and /ɔ/ in lip openness, as observed in Figure 6. Vowel class is a significant predictor of lip openness ($F_{4,95}$ = 35.93, $p < 0.001$) and lip spread ($F_{4,95}$ = 27.8, $p < 0.001$). Tukey post hoc test results show that /ɑ/ and /ɔ/ differ significantly in lip openness, but not in lip spread.

Figure 6: Lip measurements for Speaker 5.



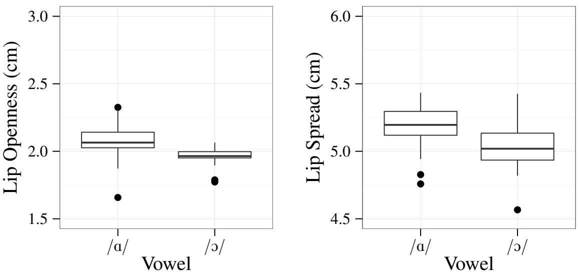
Speaker 7, a 21 year-old female, displays the opposite pattern, where /ɑ/ and /ɔ/ differ in tongue position but not in lip rounding. Smoothing splines for /ɑ/ and /ɔ/ as produced by Speaker 7 are presented in Figure 7. For this speaker, tongue contours for /ɑ/ and /ɔ/ differ significantly throughout the tongue root and body.

Figure 7: Tongue contours for /ɑ/ and /ɔ/, Speaker 7.



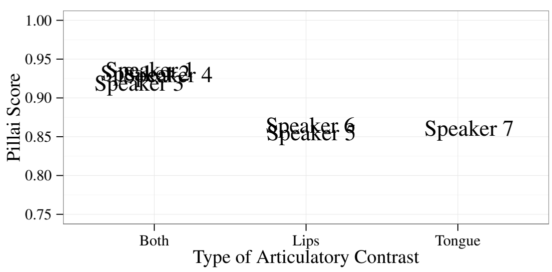
This speaker does not maintain a significant contrast between /ɑ/ and /ɔ/ in either lip spread or openness, although /ɔ/ is somewhat more round than /ɑ/. These results are presented in Figure 8. For Speaker 7, vowel class is a significant predictor of both lip openness ($F_{4,93}$ = 13.74, $p < 0.001$) and lip spread ($F_{4,93}$ = 35.24, $p < 0.001$). However, a Tukey post hoc test reveals that /ɑ/ and /ɔ/ do not differ significantly in either measure.

Figure 8: Lip measurements for Speaker 7.



To compare the acoustic patterns associated with these configurations, a Pillai-Bartlett trace ('Pillai score') was calculated for each speaker, taking into account preceding and following phonological environment [12, 10, 21]. The Pillai score method was first used for sociophonetic research by Hay et al. [12], and has since been highlighted by Hall-Lew [10] and Nycz and Hall-Lew [21]. A Pillai score is the output of a multivariate analysis of variance (MANOVA), which allows for statistical analysis of multiple dependent variables. This method returns a score between 0 and 1, where 0 indicates identical distributions and where 1 indicates no overlap at all.

Figure 9: Pillai score for each speaker, by articulatory configuration.



The results are presented in Figure 9, where the Pillai score for each speaker is plotted with speakers grouped by articulatory configuration. For speakers who use both lip rounding and tongue position to distinguish /ɔ/ from /ɑ/, the mean Pillai score is 0.929, indicating that these vowels are quite distinct. Among speakers who use a single gesture to distinguish between these vowels, the mean Pillai scores are 0.86 for those who use lip rounding, and 0.861 for Speaker 7, who uses tongue position to do so.

5. DISCUSSION

While the speakers in this study exhibit three different articulatory patterns, only two distinct acoustic patterns are observed. Speakers who produce a contrast between /ɑ/ and /ɔ/ through both lip rounding and tongue position have a higher Pillai score than speakers who produce a contrast along only one articulatory dimension, indicating that the acoustic difference between /ɑ/ and /ɔ/ is smaller for speakers who use only one articulatory gesture to distinguish between these vowels.

It therefore appears that the use of both gestures serves to enhance the acoustic contrast. However, the question of how these patterns are acquired remains. It is not immediately clear whether speakers with a smaller contrast intend to produce these vowels more similarly, and use an appropriate number of articulatory gestures to achieve this, or whether the smaller acoustic difference is a side effect of the articulatory configuration inferred by the learner during acquisition.

These findings also raise a number of questions with regard to their implications for language change. One area in which this relationship may be explored is in the domain of speech perception. It has long been argued that sound change can result from misperception of ambiguous speech signals [22, 23, 25, 2]. Similarly, visual cues and top-down processing can influence speech perception, as known, for example, from the McGurk effect [19, 26, 11]. In a situation where /ɑ/ and /ɔ/ are acoustically similar, and where the loss of lip rounding has caused these vowels to become visually similar, a merger of these vowels might occur as the result of misperception. However, where lip rounding is preserved, visual cues may help to preserve the contrast, despite acoustic similarity.

6. CONCLUSION

This study has demonstrated not only that Northern Cities speakers employ differing strategies to achieve fronting of /ɑ/ and /ɔ/, but also that some speakers produce a greater acoustic contrast between these vowels than others. Future questions to address include whether these differences in articulatory configuration influence perception of these vowels, whether these patterns are socially or geographically stratified, and how these patterns are acquired. Expanding upon this research will provide a worthwhile contribution to our understanding of the Northern Cities Vowel Shift and to language variation and change more generally.

7. REFERENCES

[1] Baker, A. 2006. Quantifying diphthongs: a statistical technique for distinguishing formant contours. *Presentation given at NWAV 35, Columbus, OH*.

[2] Blevins, J. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge, UK: Cambridge University Press.

[3] Blevins, J., Garrett, A. 2004. The evolution of metathesis. In: Hayes, B., Kirchner, R., Steriade, D., (eds), *Phonetically-Based Phonology*. Cambridge, UK: Cambridge University Press 117–156.

[4] Boersma, P., Weenink, D. 2015. Praat: doing phonetics by computer [Computer program]. Version 5.4.05, retrieved 6 March 2015 from http://www.praat.org/.

[5] Chen, Y., Lin, H. 2011. Analysing tongue shape and movement in vowel production using SS ANOVA in ultrasound imaging. *Proceedings of the 17th International Congress of Phonetic Sciences* 124–127.

[6] Davidson, L. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *The Journal of the Acoustical Society of America* 120(1), 407–415.

[7] De Decker, P. M., Nycz, J. 2012. Are tense [æ]s really tense? The mapping between articulation and acoustics. *Lingua* 122(7), 810–821.

[8] Fruehwald, J. 2010. SS ANOVA. Unpublished manuscript.

[9] Gu, C. 2002. *Smoothing Spline ANOVA Models*. New York, NY: Springer.

[10] Hall-Lew, L. 2010. Improved representation of variance in measures of vowel merger. *The Journal of the Acoustical Society of America* 127(3).

[11] 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(2), 137–156.

[12] Hay, J., Warren, P., Drager, K. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34, 458–484.

[13] Labov, W., Ash, S., Boberg, C. 2006. *The Atlas of North American English*. Berlin: Walter de Gruyter.

[14] Lawson, E., Scobbie, J. M., Stuart-Smith, J. 2011. The social stratification of tongue shape for postvocalic /r/ in Scottish English. *Journal of Sociolinguistics* 15(2), 256–268.

[15] Lee-Kim, S.-I., Davidson, L., Hwang, S. 2013. Morphological effects on the darkness of English intervocalic /l/. *Laboratory Phonology* 4(2), 475–511.

[16] Lee-Kim, S. I., Kawahara, S., Lee, S. J. 2014. The 'whistled' fricative in Xitsonga: Its articulation and acoustics. *Phonetica* 71(1), 50–81.

[17] Li, M., Kambhamettu, C., Stone, M. 2005. Automatic contour tracking in ultrasound images. *Clinical Linguistics and Phonetics* 19(6-7), 545–554.

[18] Majors, T., Gordon, M. J. 2008. The [+spread] of the Northern Cities Shift. *University of Pennsylvania Working Papers in Linguistics* 14(2), 111–120.

[19] McGurk, H., MacDonald, J. 1976. Hearing lips and seeing voices. *Nature* 264, 746 – 748.

[20] Nycz, J., De Decker, P. 2006. A new way of analyzing vowels: comparing formant contours using smoothing spline ANOVA. *Poster presented at NWAV* 35.

[21] Nycz, J., Hall-Lew, L. 2014. Best practices in measuring vowel merger. *Proceedings of the 166th Meeting of the Acoustical Society of America*.

[22] Ohala, J. J. 1981. The Listener as a Source of Sound Change. In: Masek, C. S., Hendrick, R. A., Miller, M. F., (eds), *Papers from the Parasession on Language and Behavior*. Chicago: Chicago Linguistic Society 178–203.

[23] Ohala, J. J. 1993. The phonetics of sound change. *Historical linguistics: Problems and perspectives* 237–278.

[24] Plichta, B. 2005. *Interdisciplinary perspectives on the Northern Cities Chain Shift*. PhD thesis Michigan State University.

[25] Steriade, D. 2001. The phonology of perceptibility effects: the P-map and its consequences for constraint organization.

[26] Traunmüller, H., Öhrström, N. 2007. Audiovisual perception of openness and lip rounding in front vowels. *Journal of Phonetics* 35(2), 244–258.