Visually-oriented enhancement of vowel contrast in the Northern Cities Shift

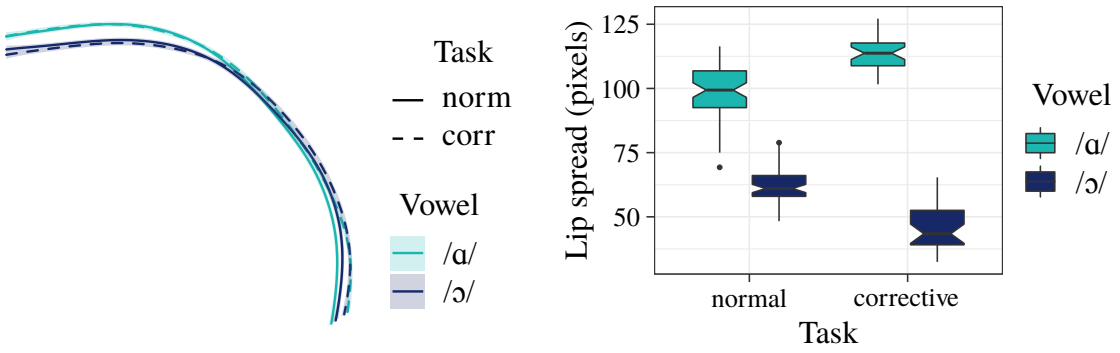
Jonathan Havenhill, The University of Hong Kong

Acoustic distance and auditory distinctiveness are known to play an important role in the organization of vowel systems [1–3] as well as in sound change [4, 5]. Although speech perception is also influenced by non-auditory cues, such as vision [6], it has not widely been considered whether non-auditory perception influences vowel inventories or patterns of sound change (but see [7, 8]). This study tests the hypothesis that vowel systems are organized around principles of both auditory and visual dispersion using articulatory and acoustic data from the Chicago variety of English.

The Chicago vowel system is characterized by the Northern Cities Shift, a chain shift involving the raising of /æ/ and the fronting of /ɑ/, followed by the fronting of /ɔ/ [9]. Although acoustic theories of vowel dispersion can explain why /ɔ/ undergoes fronting (to fill the gap in the vowel space left behind by /ɑ/), they offer no prediction as to how fronted /ɔ/ will be articulated. Both tongue fronting and lip unrounding shorten the front cavity of the vocal tract (thereby raising F2), so either strategy is predicted to be possible on a purely acoustic basis. On the other hand, tongue fronting is predicted to be favored on an audiovisual basis; if fronted /ɔ/ retains its lip rounding gesture, it will remain both auditorily and visually distinct from /ɑ/. Two specific research questions are addressed here. First, are round variants of fronted /ɔ/ more common than unround variants, given that they avoid the loss of visual contrastiveness? Second, do speakers actively enhance the /ɑ/-/ɔ/ contrast for visual perceptibility by hyperarticulating lip rounding gestures in a corrective speech task?

Fifteen adult Chicagoans (3 men, 12 women) participated in the study. Each participant completed two speech production tasks, which were recorded using synchronized audio, lip video, and ultrasound tongue data. Ultrasound data were captured at 84 frames per second (fps) with the transducer held in place by a stabilizing headset [10]. Sagittal lip video was recorded at 60 fps with a camera mounted to the headset. In the normal speech task, participants produced a list of 123 mostly monosyllabic words containing the vowels /i æ ɑ ɔ o u/, embedded in in the carrier phrase "say ____ again." Each phrase was repeated three times in succession. In the corrective focus task, participants produced minimally contrastive words containing the vowels /ɑ/ and /ɔ/ in the carrier phrase "I said $target_x$ and $target_y$, not $contrast_a$ and $contrast_b$," e.g., "I said *nod* and *sod*, not *gnawed* and *sawed*." Ultrasound tongue contours were analyzed using polar SS ANOVA [11], indicating for each speaker whether /ɑ/ and /ɔ/ exhibit significantly different tongue positions. Horizontal lip spread was taken as a measure of lip rounding and analyzed using one-way ANOVA with Tukey post hoc tests for pairwise comparison. Acoustic distance between /ɑ/ and /ɔ/ was quantified using Pillai scores [12].

In the normal speech task, 14 of the 15 speakers produced a significant ($p < 0.01$) difference between /ɑ/ and /ɔ/ in terms of lip spread. Of these 14, seven speakers also produced /ɑ/ and /ɔ/ with significantly different tongue positions (Fig. 1a), while seven contrasted /ɔ/ from /ɑ/ through lip rounding alone. This finding supports the prediction that round variants of /ɔ/ are preferred. Consistent with recent articulatory studies of Lombard speech [13–15], participants showed a range of strategies in the corrective focus task, but a large majority (11 of 15) showed a significant ($p < 0.05$) increase in the lip rounding distinction between /ɑ/ and /ɔ/ (Fig. 1b). Three of these speakers showed an increase in lip rounding with no accompanying increase in acoustic distance, suggesting that increased lip rounding is not necessarily a byproduct of auditory enhancement. These results demonstrate that articulatory strategies that preserve or enhance both auditory *and* visual contrast are favored over strategies that improve contrast only in the auditory domain. Considering visual perceptibility will strengthen predictions of how vowel systems develop as they undergo change.



(a) SS ANOVA tongue contours **(b)** Lip spread

Figure 1: Representative articulatory data for one Chicago speaker. For SS ANOVA, tongue front is to the left. Shading indicates 95% confidence interval; overlap indicates no significant difference in tongue contours. For lip spread, smaller values indicate increased rounding.

[1] J. Liljencrants and B. Lindblom. "Numerical simulation of vowel quality systems: The role of perceptual contrast". In: *Language* 48.4 (1972), pp. 839–862. **[2]** B. Lindblom. "Phonetic universals in vowel systems". In: *Experimental Phonology*. Ed. by J. J. Ohala and J. J. Jaeger. Orlando: Academic Press, 1986, pp. 13–44. **[3]** E. Flemming. "Contrast and Perceptual Distinctiveness". In: *Phonetically-Based Phonology*. Ed. by B. Hayes, R. Kirchner, and D. Steriade. Cambridge: Cambridge University Press, 2004. **[4]** J. J. Ohala. "The phonetics of sound change". In: *Historical Linguistics: Problems and Perspectives*. Ed. by C. Jones. London: Longman, 1993, pp. 237–278. **[5]** W. Labov. *Principles of Linguistic Change*. Malden, MA: Wiley-Blackwell, 1994. **[6]** H. McGurk and J. MacDonald. "Hearing lips and seeing voices". In: *Nature* 264 (1976), pp. 746–748. **[7]** G. McGuire and M. Babel. "A cross-modal account for synchronic and diachronic patterns of /f/ and /θ/ in English". In: *Laboratory Phonology* 3.2 (2012), pp. 1–41. **[8]** J. Havenhill and Y. Do. "Visual speech perception cues constrain patterns of articulatory variation and sound change". In: *Frontiers in Psychology* 9 (2018), p. 728. **[9]** W. Labov, S. Ash, and C. Boberg. *The Atlas of North American English*. Berlin: Walter de Gruyter, 2006. **[10]** Articulate Instruments Ltd. *Ultrasound Stabilisation Headset Users Manual: Revision 1.4*. Edinburgh, UK: Articulate Instruments Ltd., 2008. **[11]** J. Mielke. "An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons". In: *The Journal of the Acoustical Society of America* 137.5 (2015), pp. 2858–2869. **[12]** J. Hay, A. Nolan, and K. Drager. "From fush to feesh: Exemplar priming in speech perception". In: *The Linguistic Review* 23.3 (2006), pp. 351–379. **[13]** V. Hazan and J. Kim. "Acoustic and visual adaptations in speech produced to counter adverse listening conditions". In: *Proceedings of the 12th International Conference on Auditory-Visual Speech Processing*. Ed. by S. Ouni, F. Berthomier, and A. Jesse. Annecy, France: Inria, 2013, pp. 93–98. **[14]** M. Garnier, L. Ménard, and B. Alexandre. "Hyper-articulation in Lombard speech: An active communicative strategy to enhance visible speech cues?" In: *Journal of the Acoustical Society of America* 144.2 (2018), pp. 1059–1074. **[15]** J. M. Scobbie and J. Ma. "Say again? Individual articulatory strategies for producing a clearly-spoken minimal pair wordlist". In: *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*. Ed. by S. Calhoun et al. Canberra, Australia: Australasian Speech Science and Technology Association Inc., 2019, pp. 358–362.