Audiovisual cue enhancement in the production and perception of the COT-CAUGHT contrast

Jonathan Havenhill

October 19, 2018

NWAV 47

香 港 大 學

THE UNIVERSITY OF HONG KONG

Introduction

- Most work on language variation and change focuses on the acoustic/auditory properites of speech:
  - In sociophonetics, variation typically described in acoustic terms
  - Ohala (1981, 1993); Blevins (2004): Sound change is the result of misperception of the acoustic signal
  - Lindblom (1990); Hayes et al. (2004): Speakers optimize their speech for auditory perceptibility

Introduction

- Yet, sounds can also vary in their articulation:
  - Bunched vs. Retroflex /ɹ/ (Delattre & Freeman, 1968)
  - Apical vs. Laminal /s/ (Bladon & Nolan, 1977)
  - Tensing vs. Nasalization of /æ/ (De Decker & Nycz, 2012)
- Acoustic study alone can't reveal this type of variation.

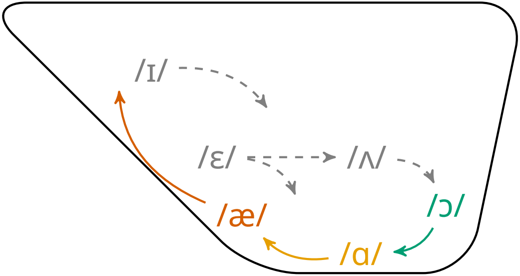
Introduction

- Speech perception is also influenced by a variety of non-auditory cues:
  - **Visual** (Sumby & Pollack, 1954; McGurk & MacDonald, 1976; Mayer et al., 2013)
  - **Haptic** (Fowler & Dekle, 1991)
  - **Aerotactile** (Gick & Derrick, 2009)
  - **Somatosensory** (Ito et al., 2009)

Overview

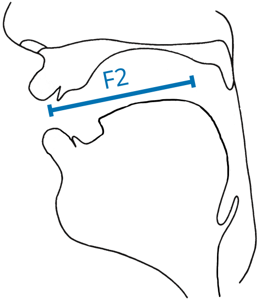
- **This talk**: Investigation of articulatory patterns underlying COT-CAUGHT contrast
- Two experiments: **production** and **perception**
- **Production**: Most speakers from Chicago preserve lip rounding distinction between LOT and THOUGHT
- **Perception**: Round variants of THOUGHT are perceptually more robust because they provide both auditory and visual contrast

The Northern Cities Shift



- NCS characterized by raised TRAP (/æ/), fronted LOT (/ɑ/), and fronted/lowered THOUGHT (/ɔ/).
- Fronting of LOT precedes fronting of THOUGHT, so these vowels remain distinct.
- Fronting described as an increase in F2.

Articulation of fronted LOT and THOUGHT



- An increase in F2 can be the result of any gesture that shortens the front cavity of the vocal tract.
- Fronting of THOUGHT may be achieved by tongue fronting, lip unrounding, or both.

Articulation of fronted LOT and THOUGHT

- Havenhill & Do (2018): All three types of contrast observed among speakers from Metro Detroit.
- Loss of rounding or backness distinction for THOUGHT results in weaker acoustic contrast with LOT.
- Unround variants of THOUGHT are perceptually weaker than round variants, due to loss of visual contrast.
  - Unrounding predicted to be dispreferred, but data are insufficient.

Research Questions

- How do Chicagoans maintain the LOT-THOUGHT contrast in terms of articulation?
  - Does THOUGHT retain or lose its rounding as it undergoes fronting?
- Do fronting/unrounding strategies differ in their perceptibility?
  - Can visual rounding cues enhance a weak acoustic contrast?

Experiment 1:
Production

Methods: Participants

- Sixteen (4 men, 12 women) Chicago natives, ages 20 to 77.
- Participants born and raised in Chicago through the age of 18.
  - Majority are lifelong Chicagoans.
- Data collected at the Northwestern University Phonetics Lab.

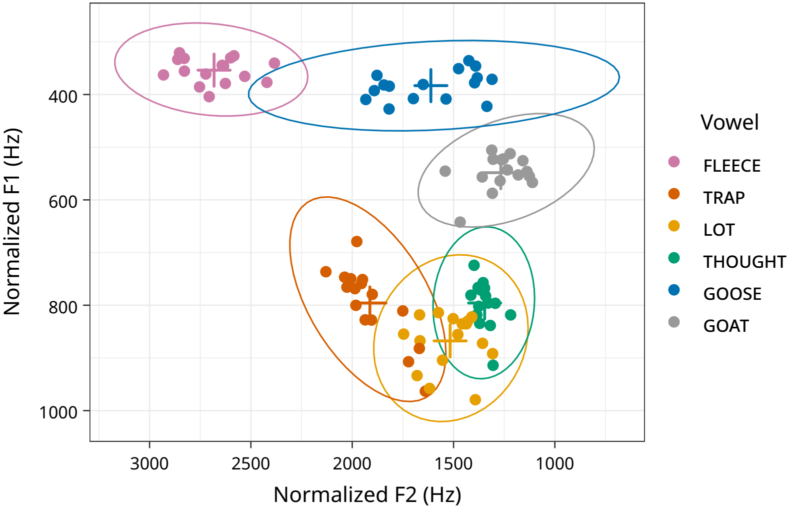
Methods: Recording

- Three repetitions of 112 words containing /æ ɑ ɔ i u o/
  - Embedded in carrier phrase “say _______ again”
- Simultaneous audio, video, and ultrasound recorded in Articulate Assistant Advanced (AAA)
- **Ultrasound**
  - 84 frames per second w/ SonoSpeech Micro system
  - Probe fixed to speaker’s head with Articulate Instruments stabilizing headset
- **Video**
  - Sagittal view @ 60 fps

Methods: Analysis

- **Audio**:
  - Formants measured at F1 maximum
  - Normalized with *ANAE* log-mean method (Labov et al., 2006)
- **Ultrasound**:
  - Tongue contours extracted at point of maximum constriction
  - Modeled with polar smoothing spline ANOVA (Gu, 2002; Davidson, 2006; Mielke, 2015) in R.
- **Video**:
  - Lip spread: Distance between lower lip and corner of mouth

Acoustic Results



Acoustic Results

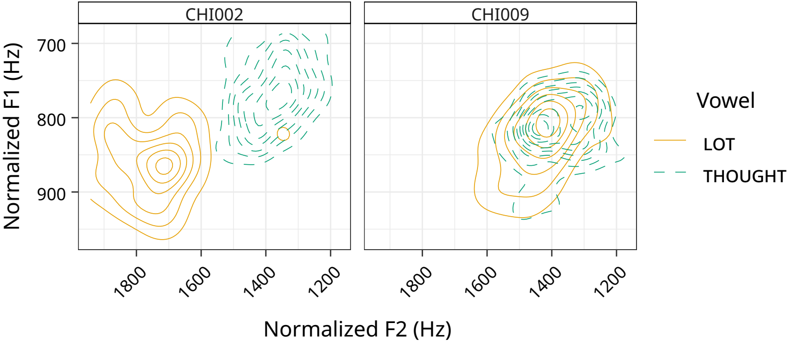


Figure 1: Kernel density estimation plot for LOT and THOUGHT.

- Distributions for LOT and THOUGHT range from almost no overlap to almost complete overlap.

Acoustic Results

Figure 2: Pillai score measure of vowel overlap

- Numeric measure of vowel distribution: **Pillai score** (Hay et al., 2006).
- Multivariate ANOVA, incorporates both distance between means and overlap.

Acoustic Results

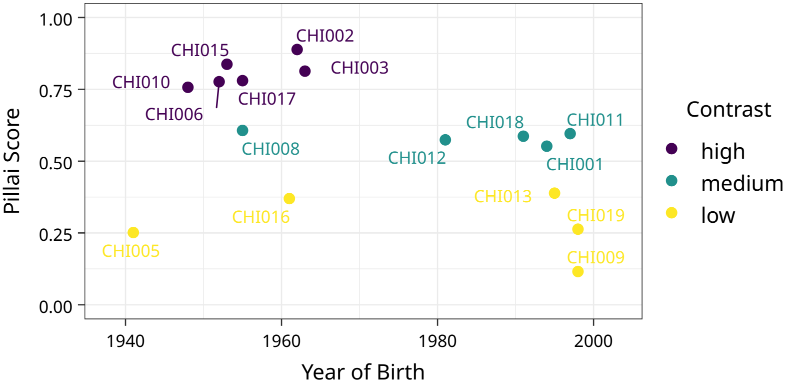


Figure 3: Pillai scores by year of birth.

- Roughly three degrees of contrast.
- Older speakers generally exhibit strongest LOT-THOUGHT contrast.

Articulatory Results

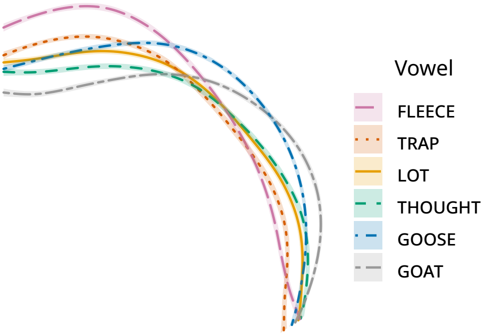


Figure 4: Polar smoothing spline estimates for CHI010, all vowels.

- SS ANOVA tongue contours with 95% confidence intervals.
- Tongue front is to the left.

Articulatory Results

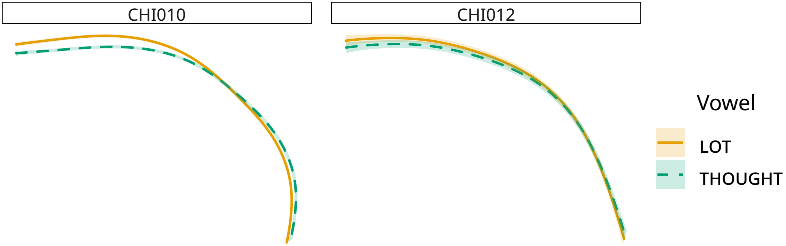


Figure 5: Polar smoothing spline estimates for LOT and THOUGHT.

- Speaker 10 exhibits significant difference in tongue position, Speaker 12 does not.

Articulatory Results

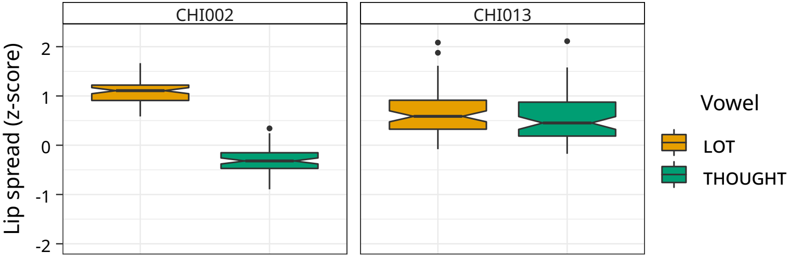


Figure 6: Lip spread measurements for LOT and THOUGHT.

- Speaker 2 exhibits significant difference in lip spread, Speaker 13 does not.

Articulatory/Acoustic Results

Table 1: Summary of acoustic and articulatory results.

Speaker ID	Pillai score	Tongue difference	Lip spread difference	Strategy
Speaker 2	0.889	Yes	Yes	Both
Speaker 15	0.837			
Speaker 3	0.813			
Speaker 6	0.776			
Speaker 10	0.757			
Speaker 8	0.607			
Speaker 18	0.587			
Speaker 17	0.780	No	Yes	Lip
Speaker 11	0.596			
Speaker 12	0.574			
Speaker 1	0.552			
Speaker 16	0.370			
Speaker 5	0.252			
Speaker 13	0.389	Yes	No	Tongue
Speaker 19	0.263	No	Yes	Near-merger
Speaker 9	0.116		No	Merged

- Four articulatory-acoustic patterns, including merger

Articulatory/Acoustic Results

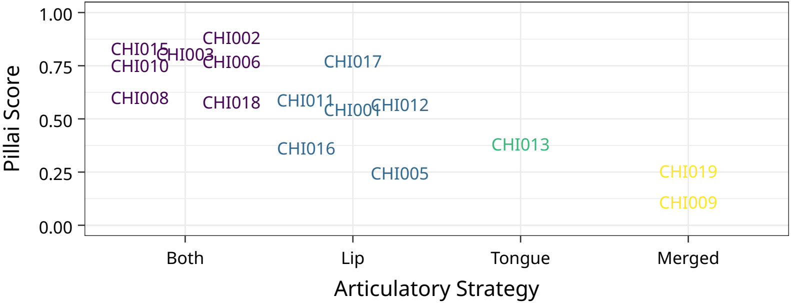


Figure 7: Pillai scores by articulatory strategy.

- Acoustic contrast is strongest when THOUGHT and LOT differ in both tongue position and lip rounding.

Summary: Production Experiment

- Wide range between speakers in terms of acoustic contrast.
    - Acoustic contrast generally weaker for speakers with only one type of articulatory distinction.
    - Articulatory data reveal patterns not observable in acoustic signal alone.
- Dispreference for THOUGHT-unrounding is supported; strategy observed for only one speaker.
    - 13 of 16 speakers maintain lip rounding for THOUGHT

Experiment 2:
Perception

Audiovisual perception experiment

- Acoustic contrast is weaker for THOUGHT variants with only one articulatory distinction; do they differ in how they're perceived?
  - **Prediction 1**: Unround variants of THOUGHT more likely to be (mis)perceived as LOT.
  - **Prediction 2**: Speakers who produce THOUGHT as unround will not rely on visual cues to perceive the contrast.

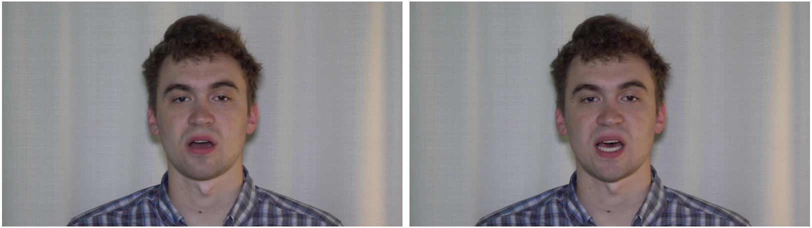
Methods: Participants

- Same participants as production experiment.
- One participant did not complete the perception task.

Methods: Stimuli

- 120 monosyllabic nonce words, presented in pink noise:
  - Target vowels: LOT and THOUGHT
  - Control vowels: FLEECE, GOOSE, FACE, GOAT
- Target items mismatched for visible lip rounding (à la McGurk Effect).
- Control items mismatched for height.
- Four NCS talkers (2 men, 2 women) who produce THOUGHT with lip rounding.
  - 4 blocks $\times$ 120 stimuli $=$ 480 total stimuli

Methods: Stimuli



(a) auditory THOUGHT, visual THOUGHT

(b) auditory THOUGHT, visual LOT

Figure 8: Audiovisual stimuli

- Stimulus: “say [skɔθ] again”
- Identical audio paired with congruous and incongruous video

Methods: Experimental Paradigm

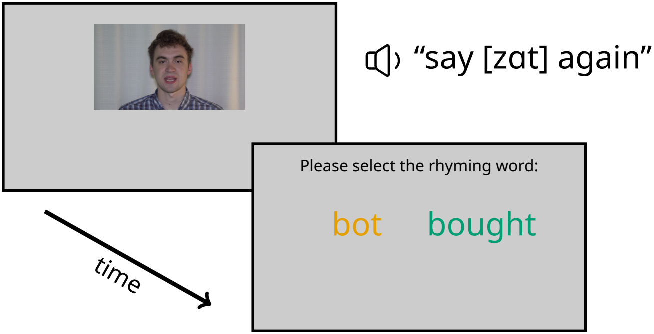


Figure 9: Experimental design.

- Participants asked to identify which (real) word rhymes with the stimulus.

Results: Target Items

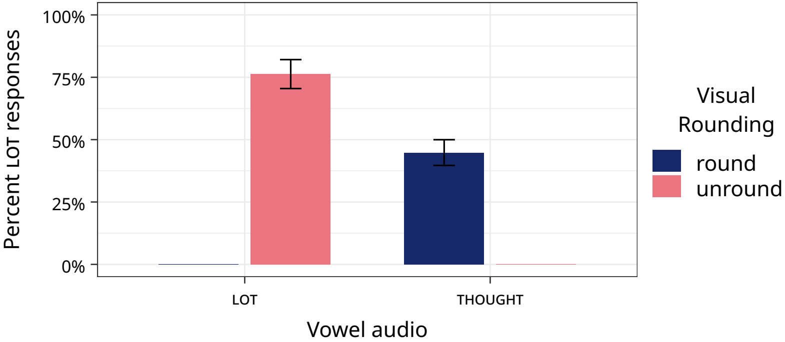


Figure 10: Perception results for participants (N = 13) who distinguish THOUGHT from LOT with lip rounding.

- Effect of incongruity for THOUGHT, but not for LOT.

Results: Target Items

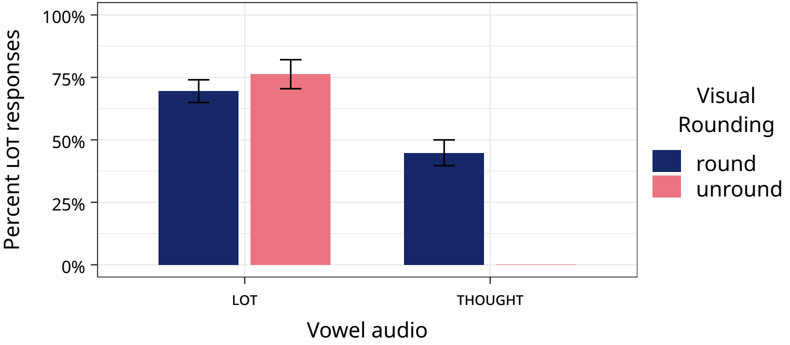


Figure 10: Perception results for participants (*N* = 13) who distinguish THOUGHT from LOT with lip rounding.

- Effect of incongruity for THOUGHT, but not for LOT.

Results: Target Items

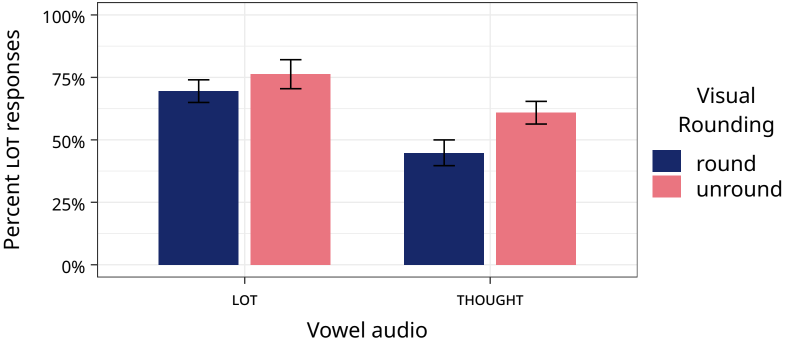


Figure 10: Perception results for participants (N = 13) who distinguish THOUGHT from LOT with lip rounding.

- Effect of incongruity for THOUGHT, but not for LOT.

Results: Target Items

Table 2: Mixed effects logistic regression model for participants (N = 13) who produce THOUGHT with lip rounding.

Predictor	Estimate	SE	z value	Pr(>|z|)	
Intercept (LOT, Congruous)	0.759	0.067	11.282	<0.001	***
Vowel Audio					
THOUGHT	-0.309	0.065	-4.786	<0.001	***
Visual Congruity					
Incongruous	-0.062	0.065	-0.952	>0.05	
Audio * Congruity					
THOUGHT * Incongruous	0.220	0.091	2.406	<0.05	*

- Significant interaction effect between auditory THOUGHT and visual incongruity (unrounding).

Results: Target Items

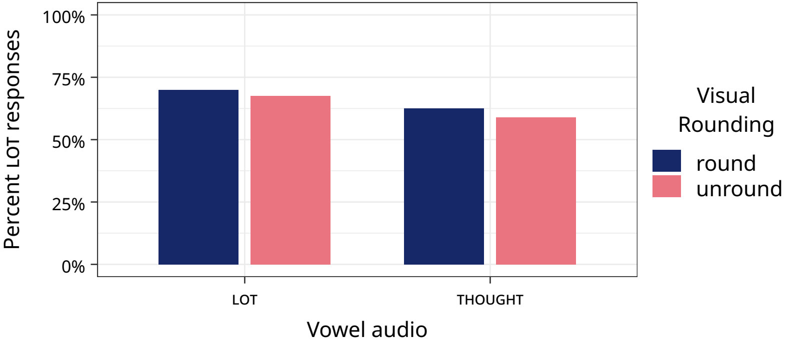


Figure 11: Perception results for participant (*N* = 1) who distinguishes THOUGHT from LOT with tongue position alone.

- No effect of visual incongruity.

Results: Target Items

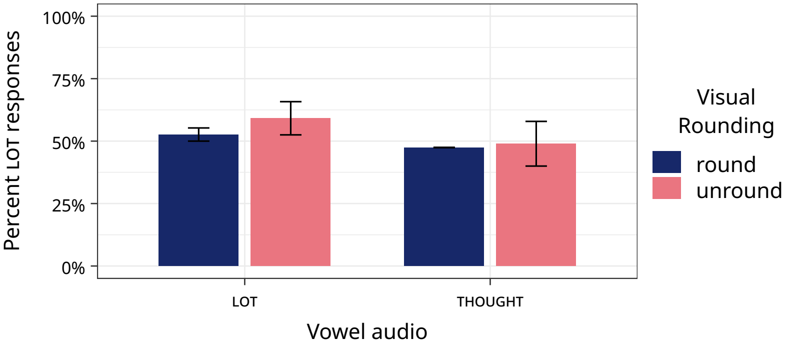


Figure 12: Perception results for participants (N = 2) who do not produce a contrast between LOT and THOUGHT.

- Perception of LOT and THOUGHT at chance; vowels are perceptually merged.

Summary: Perception Experiment

- Unround variants of THOUGHT significantly more likely to be perceived as LOT
  - Accounts for THOUGHT-unrounding dispreference.
- Participant(s) who do not rely on lip rounding in production of THOUGHT show no effect of visual incongruity.

Conclusions

- For speakers who maintain COT-CAUGHT contrast, visual perceptibility may drive preference for retaining lip rounding.
- Implications for NCS reversal?
- Consideration of articulatory factors is warranted in cases where acoustic changes can have competing articulatory strategies.

Thank you!

Contact

- jhavenhill@hku.hk
- https://jhavenhill.com
- @jhavenhill

Many thanks to:

Youngah Do, Lisa Zsiga, and Jen Nycz for comments and suggestions; Jennifer Cole, Annette D’Onofrio, and Chun Chan for assistance with data collection.

References I

Bladon, R. Anthony W. & Francis J. Nolan. 1977. A video-fluorographic investigation of tip and blade alveolars in English. **Journal of Phonetics** 5. 185–193.

Blevins, Juliette. 2004. **Evolutionary phonology: The emergence of sound patterns**. Cambridge: Cambridge University Press.

Davidson, Lisa. 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.

De Decker, Paul M. & Jennifer Nycz. 2012. Are tense [æ]s really tense? The mapping between articulation and acoustics. **Lingua** 122(7). 810–821. doi: 10.1016/j.lingua.2012.01.003.

Delattre, Pierre & Donald C. Freeman. 1968. A dialect study of American R's by X-ray motion picture. **Linguistics** 6(44). 29–68. doi: 10.1515/ling.1968.6.44.29.

Fowler, Carol A. & Dawn J. Dekle. 1991. Listening with eye and hand: Cross-modal contributions to speech perception. **Journal of Experimental Psychology: Human Perception and Performance** 17(3). 816–828. doi: 10.1037/0096-1523.17.3.816.

Gick, Bryan & Donald Derrick. 2009. Aero-tactile integration in speech perception. **Nature** 462(7272). 502–504. doi: 10.1038/nature08572.

Gu, Chong. 2002. **Smoothing spline ANOVA models**. New York: Springer.

References II

Havenhill, Jonathan & Youngah Do. 2018. Visual speech perception cues constrain patterns of articulatory variation and sound change. **Frontiers in Psychology** 9. 728. doi: 10.3389/fpsyg.2018.00728.

Hay, Jennifer, Paul Warren & Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. **Journal of Phonetics** 34(4). 458–484. doi: 10.1016/j.wocn.2005.10.001.

Hayes, Bruce, Robert Kirchner & Donca Steriade (eds.). 2004. **Phonetically based phonology**. Cambridge: Cambridge University Press.

Ito, Takayuki, Mark Tiede & David J. Ostry. 2009. Somatosensory function in speech perception. **Proceedings of the National Academy of Sciences of the United States of America** 106(4). 1245–8. doi: 10.1073/pnas.0810063106.

Labov, William, Sharon Ash & Charles Boberg. 2006. **The atlas of North American English**. Berlin: Walter de Gruyter. doi: 10.1515/9783110206838.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H theory. In William J. Hardcastle & Alain Marchal (eds.), **Speech production and speech modelling**, 403–439. Dordrecht: Springer. doi: 10.1007/978-94-009-2037-8_16.

Mayer, Connor, Bryan Gick, Weigel Tamra & Doug H. Whalen. 2013. Perceptual integration of visual evidence of the airstream from aspirated stops. **Canadian Acoustics** 41(3). 23–27. `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4474184/`.

References III

McGurk, Harry & John MacDonald. 1976. Hearing lips and seeing voices. **Nature** 264. 746–748. doi: 10.1038/264746a0.

Mielke, Jeff. 2015. An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons. **The Journal of the Acoustical Society of America** 137(5). 2858–2869. doi: 10.1121/1.4919346.

Ohala, John J. 1981. The listener as a source of sound change. In Carrie S. Masek, Roberta A. Hendrick & Mary F. Miller (eds.), **Papers from the parasession on language and behavior**, 178–203. Chicago: Chicago Linguistic Society.

Ohala, John J. 1993. The phonetics of sound change. In Charles Jones (ed.), **Historical linguistics: Problems and perspectives**, 237–278. London: Longman.

Sumby, W. H. & Irwin Pollack. 1954. Visual contribution to speech intelligibility in noise. **The Journal of the Acoustical Society of America** 26(2). 212–215. doi: 10.1121/1.1907309.