Second Biennial Boston Speech Motor Control Symposium, Boston, MA, USA, June 2019

Somatosensory Inputs from the Vocal Tract Enhance Perception of Concordant Visible Speech Movements

Matthew Masapollo[1] & Frank H. Guenther[1,2]

[1]Department of Speech, Language & Hearing Sciences, Boston University

[2]Department of Biomedical Engineering, Boston University

BACKGROUND

- It is well-established that somatosensory inputs from the vocal tract play an important role in speech production and motor control (e.g., Tremblay, Shiller & Ostry, 2003; see also, Perkell, 2012; Guenther, 2016, for reviews).
- Recently, studies have provided evidence that **orofacial somatosensory inputs also influence concurrent *perception* of speech sounds** with both adult speakers (Ito, Tiede & Ostry, 2009; Ito & Ostry, 2012) and pre-babbling infants (Yeung & Werker, 2013; Bruderer, Danielson, Kandhadai, & Werker, 2015).
- Whereas these and other psychophysical experiments have demonstrated complex *somatosensory-auditory interactions* during speech processing at a behavioral level, neuroimaging studies indicate that *visual* speech cues in talking faces influence activity in somatosensory (and motor) cortex above and beyond its response to auditory speech cues alone (e.g., Matchin, Groulx & Hickok, 2014).
  - **Thus, understanding the contribution of potential *somatosensory-visual interactions* during speech processing may yield additional insights into perception-action linkages for speech.**

Research Question:

Does engaging the speech articulators influence concurrent unimodal visual speech perception?

To address this question, we examined vocalic viseme discrimination under three experimental conditions: (1) normal (baseline) and while holding either (2) a bite block or (3) a lip tube in their mouths.

To test the specificity of somatosensory-visual interactions during speech perception, we assessed discrimination of vowel contrasts that are optically distinguished based on either their mandibular (English /ɛ/-/æ/) or labial (English /u/-French /u/) postures. In addition, we assessed perception of each contrast using *dynamically-articulating* videos and *static* (single-frame) pictorial images of each gesture (at vowel midpoint).

Predictions:

- If there are somatosensory-visual interactions during phonetic perception, then manipulating the posture of the jaw should selectively influence perception of the /ɛ/-/æ/ contrast, whereas manipulating the posture of the lips should selectively influence perception of English /u/-French /u/ contrast. In addition, if engaging the articulators affects how perceivers track speech movements, rather than changes in vocal tract configuration, then there should only be an effect of condition using the video stimuli.
- Alternatively, simultaneously engaging the articulators during concurrent perception may increase attentional processing load, which in turn, will lead to a decline in overall discrimination performance, regardless of which articulator is engaged.

METHODS

Stimuli: Vocalic visemes – silent visual-only articulation (dynamic and stilled speech)

- Native within- and between-category contrasts that are optically distinct
  - English /ɛ/-/æ/; English /u/-French /u/) – /æ/ was produced with a lower mandibular position than /ɛ/; the variants of /u/ were produced with different degrees of visible lip compression and protrusion
- Each contrast was naturally spoken by a female speaker
- Productions were audio-visually recorded from a straight, face-on view in a sound-treated booth (30 frames/sec and 1,400 × 1,000 pixels; audio at 44.1 kHz)
- Using Adobe Premiere (San Jose, CA), the video-only stimuli were created by removing the audio track from the AV videos.

Procedure & Design: Categorical AX discrimination task

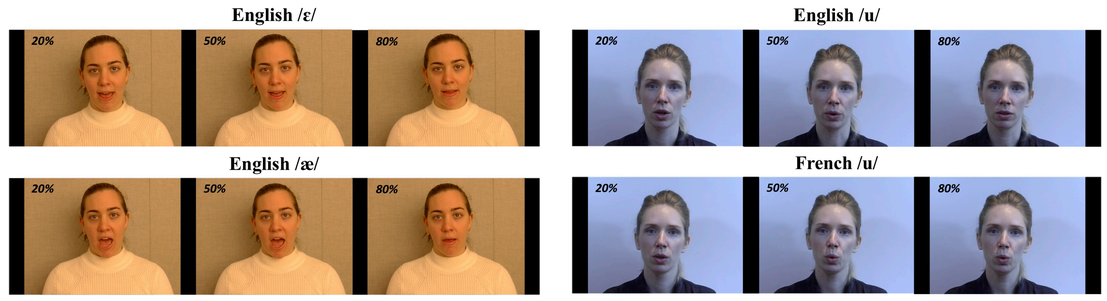
- Subjects watched silent video or image sequences of the model speaker articulating the vocalic gestures, and then judged whether they were the same or different by pressing a button on a response pad (1,500-ms inter-stimulus interval).
- Subjects saw every possible pairing of the stimuli in both presentation orders (blocked by vowel contrast; 90 trials/block; 360 trials total)
- The task was performed with either a tube (20-mm diameter) between the lips or a bite block between the upper and lower teeth. A baseline group performed the task with no oral-motor manipulation.

Subjects: Native, monolingual American English speaking adults

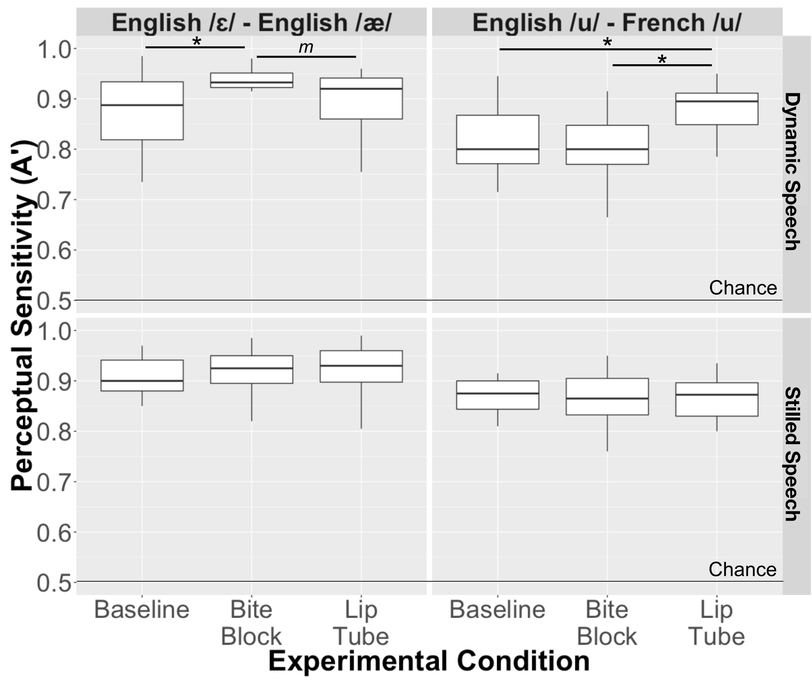
RESULTS

Analyses

- The dependent variable was ***A-prime*** (Grier, 1971); separate *A'* scores were computed for each subject for each vowel contrast in each of the three experimental conditions (lip tube vs. bite block vs. baseline).



Sample images of the model speakers' visible vocal tract configuration during the production of each vocalic gesture at 20%, 50% and 80% of the articulatory trajectory. Note that, in stilled speech condition, the stimuli only consisted of static (single-frame) images of the talking faces taken at 50% of the vocalic trajectory (as shown in the center panels). As the images show, French /u/ is executed with a greater degree of visible lip compression and protrusion, and English /æ/ is implemented with a lower mandibular position than English /ɛ/.



Dynamic Speech

Engaging the jaw selectively facilitated perception of vocalic gestures optically distinct in terms of jaw height, whereas engaging the lips selectively facilitated perception of vocalic gestures optically distinct in terms of their degree of lip rounding. Thus, subjects perceived visible speech movements in relation to the configuration of their own vocal tract (and possibly their ability to produce covert speech-like movements).

Stilled Speech

We failed to replicate the facilitation effect when the dynamically articulating faces were substituted with stilled facial speech images, suggesting that the manipulations affected perception of *time-varying* kinematic information rather than changes in "target" (i.e., movement end-point) mouth shapes.

CONCLUSIONS

- The present findings raise the intriguing possibility that **engaging the speech articulators *may* facilitate perceivers' ability to detect *time-varying movements* (rather than changes in movement end-points) of that articulator during concurrent visual speech perception.**
- Collectively, we interpret these findings as evidence that **somatosensory feedback from the articulators may "prime" premotor and somatosensory brain regions involved in the sensorimotor control of speech, thereby facilitating perception of concurrent speech movements.**

SELECTED REFERENCES

Guenther, F.H. (2016). Neural Control of Speech. MIT Press, Cambridge, MA.

Ito, T., Tiede, M., & Ostry, D.J. (2009). Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences*, 106 (4), 1245-1248.

Matchin, W., Groulx, K. & Hickok, G. (2014). Audiovisual speech integration does not rely on the motor system: evidence from articulatory suppression, the McGurk effect, and fMRI. *Journal of Cognitive Neuroscience*, 26 (3), 606-620.

Yeung, H.H. & Werker, J.F. (2009). Lip movements affect infants' audiovisual speech perception. *Psychological Science,* 24 (5), 603-612.

Funding: The research reported here was supported by a grant from the National Institute on Deafness and other Communication Disorders (NIH R01 DC002852; F.H. Guenther, P.I.).