

Somatosensory-based Compensation to Mechanical Perturbations of the Larynx during Speech

408.05/LL33

Dante J. Smith¹, Andrés F. Salazar-Gómez², Cara E. Stepp^{1,3}, Frank H. Guenther^{1,3}

¹Graduate Program for Neuroscience, Boston University, Boston MA; ²Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge MA; ³Speech Language & Hearing Sciences, Boston University, Boston MA

Background and Motivation

Disorders that involve impaired control of vocal folds, such as vocal hyperfunction, spasmodic dysphonia, muscle tension dysphonia, and Parkinson's disease, are to date poorly understood, yet have a significant impact on the lives of over 20 million people across the country. In order to better understand and ultimately treat these disorders, we aim to better characterize the neural circuit involved in producing healthy voice. The **Directions into Velocities of Articulators (DIVA)** model is a neural mode of speech and voice production detailing the feedforward control mechanisms required for fluent production, as well as the auditory and somatosensory feedback (FB) controllers that correct errors in production. Behavioral experiments studying changes in **fundamental frequency (f0, perceived as pitch)** allow direct study of the pathways involved in voice motor control.

Previous work has shown that participants compensate for auditory perturbations to f0. When f0 was artificially decreased in headphones, participants compensated by increasing their f0. Compensation was incomplete, likely because of a competition between auditory and somatosensory FB controllers arising from the fact that participants experienced mismatching sensory information: They hear an f0 that is too low, but they feel the vocal folds vibrating at the proper frequency.

Loucks et al. (2005) found that a mechanical perturbation to the larynx also elicits a compensatory response. The perturbation caused a speaker's f0 to decrease, to which they compensated by increasing their f0, with complete or near-complete compensation within about 200 ms of the perturbation. However, auditory feedback was available to the participants during the mechanical perturbation, thereby invoking the auditory FB controller in addition to the somatosensory FB controller in a cooperative attempt to correct the perceived errors. **The individual contributions of the two FB controllers to the observed compensations thus cannot be determined.**

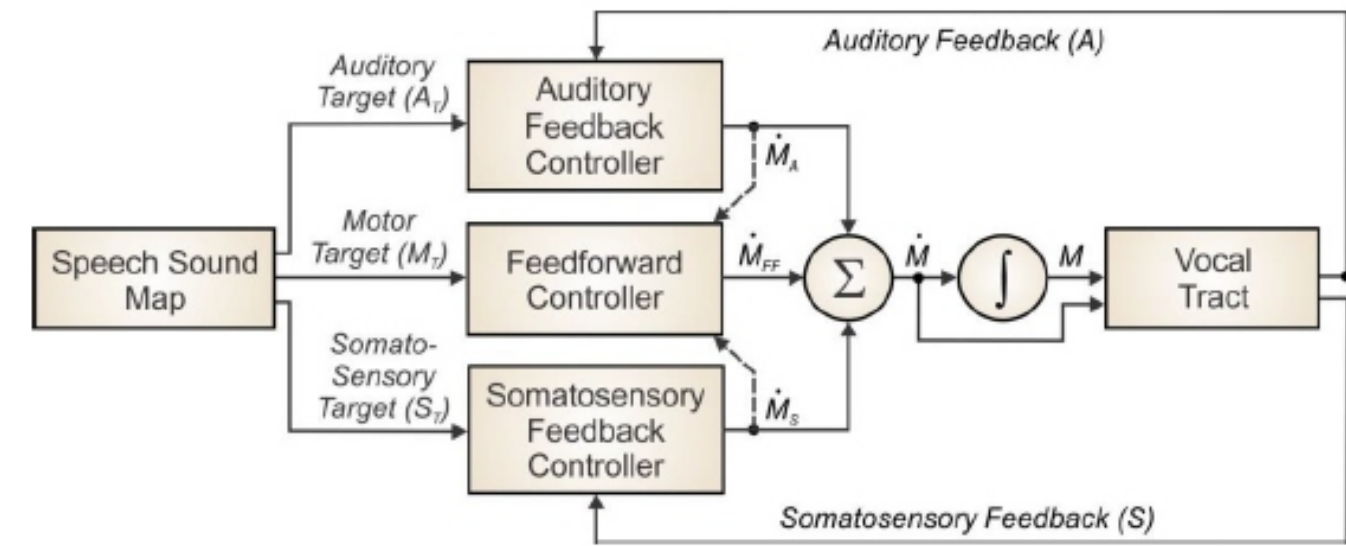


Figure 1: DIVA Model Control Scheme for speech-sound production.

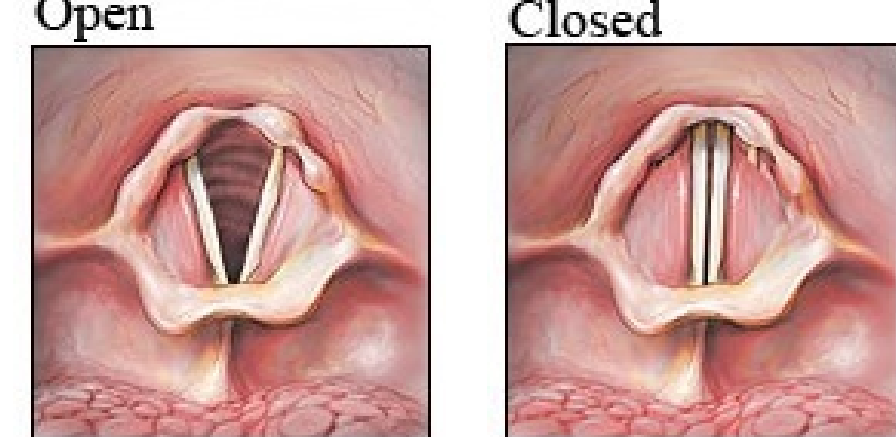


Figure 2: Illustration of superior view of the larynx; the source of voice.

The purpose of the current study was to isolate the somatosensory FB controller by applying a mechanical perturbation to the larynx while auditory feedback was masked with acoustic noise. Based on the DIVA model, we made the following predictions.

- Subjects will compensate for the perturbations by increasing f0 even in absence of auditory feedback since the somatosensory feedback will attempt to correct for somatosensory feedback errors indicating a decrease in vocal fold vibration due to the perturbation.
- Because somatosensory feedback is faster and more reliable than auditory feedback, masking auditory feedback will only cause a small decrease in the magnitude of compensation compared to compensation when auditory feedback is available.

Methods

Participants: 4 (1 female) native English speakers

No history of a speech, language, or hearing disorder, nor any hearing loss above 25dB HL
No formal experience (1 year or more) speaking a tonal language or singing instruction

Task: Vocalize the /i/ vowel while receiving a perturbation to the larynx (Figure 3)

- Performed 4 runs of the task, 40 trials each, in a sound-proof booth facing a computer monitor
- 25% of trials received a non-invasive laryngeal displacement (Figure 4)
- Instructed to maintain constant pitch and loudness through each trial
- At the end of each trial, received visual feedback about voice volume
- On 2 runs received auditory feedback in the form of their own voice (+5dB above mic)
- Other 2 runs received speech-shaped masking noise (90dB SPL) to block auditory feedback
- Microphone:** Audio-Technica Omnidirectional Condenser Lavalier head-set microphone
- Headphones:** Etymotic ER-1 insert headphones
- Audio Card:** MOTU Microbook II
- Digital Signal I/O:** National Instruments USB-6212 (BNC)

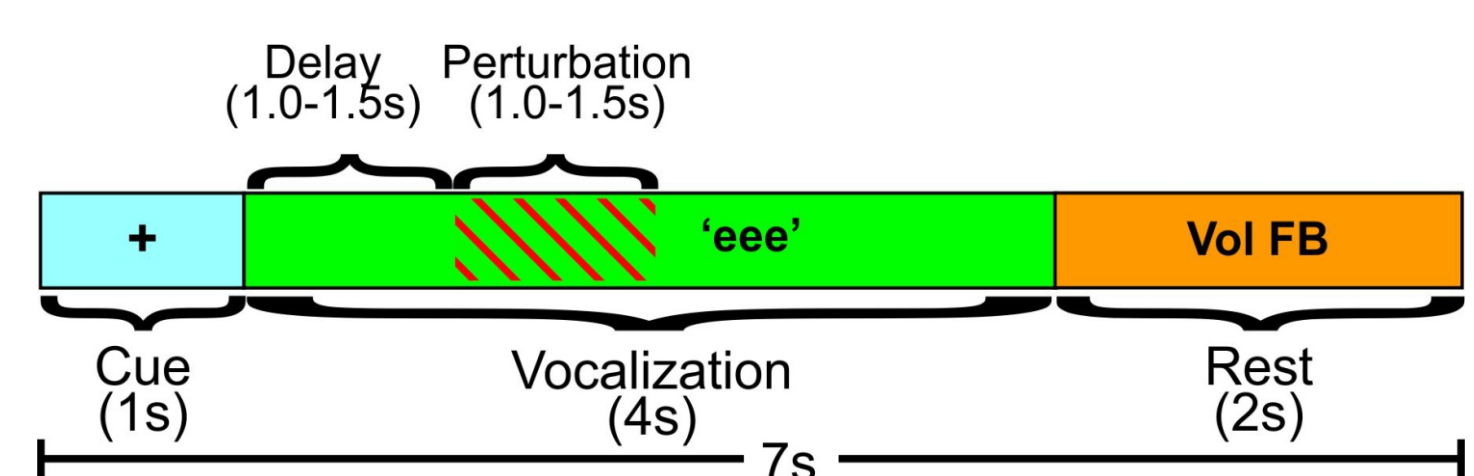
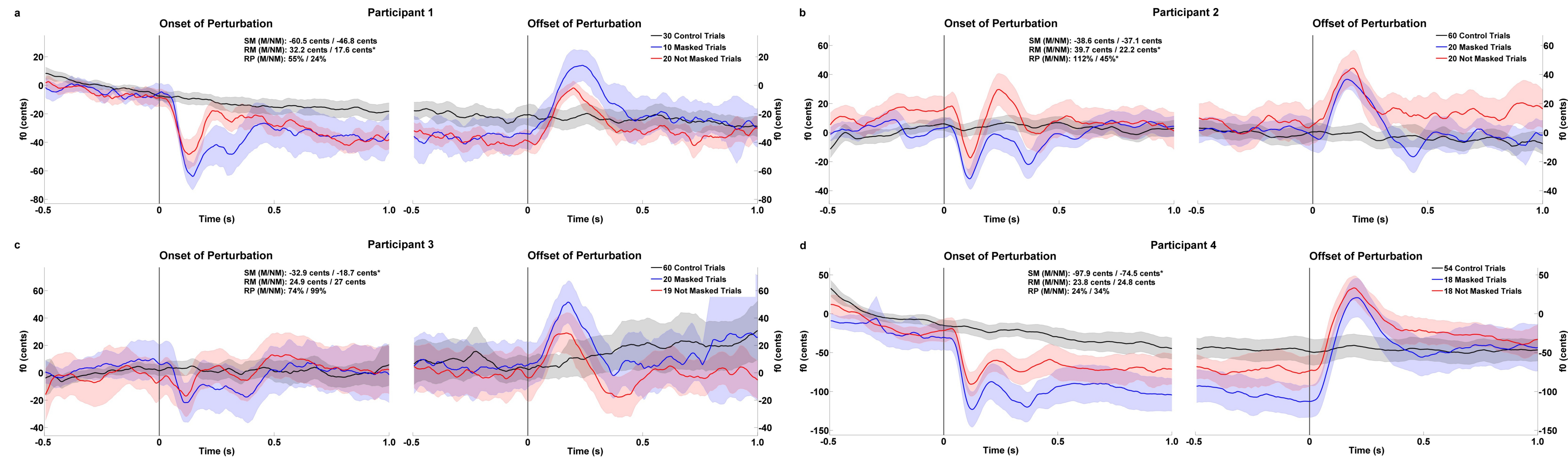


Figure 4: Laryngeal displacement device. A rigid plastic collar is fitted to a participant's neck using an elastic band. A balloon embedded in the collar is placed just under the cricoid cartilage of the larynx. During perturbation, the balloon inflates by an air tube, and pushes the cricoid cartilage dorsally and superiorly. Deflating the balloon removes the perturbation.

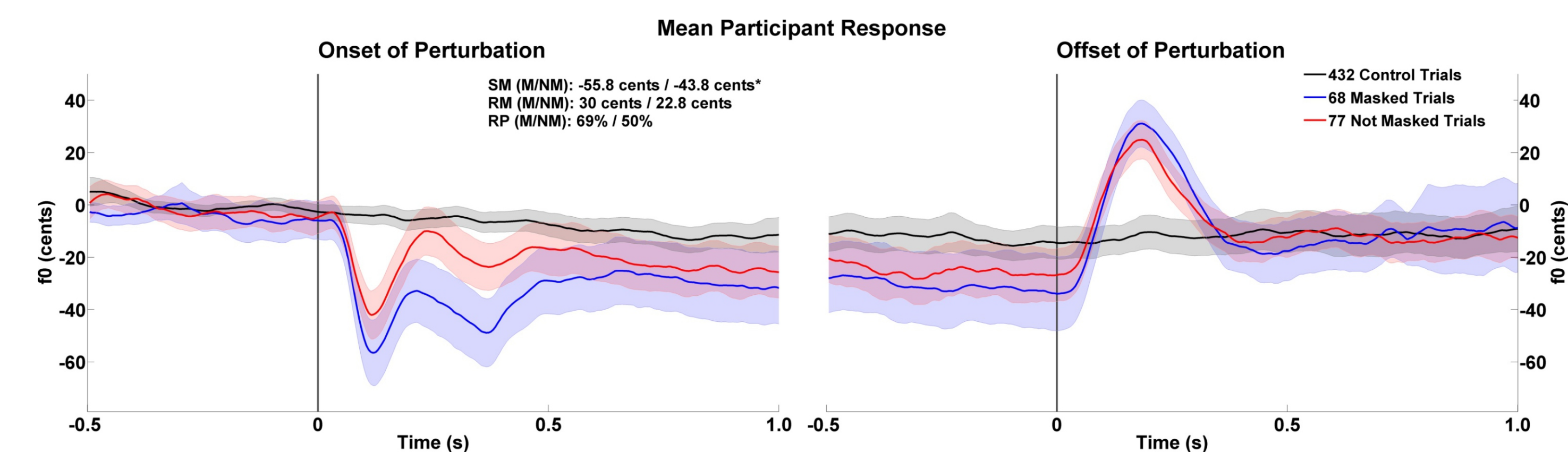
Figure 3: Epoch of an experimental trial. Participants are cued to begin phonating with a cross-hair on a computer screen. When the cross-hair is replaced by the letters 'eee' they are instructed to begin vocalizing the /i/ vowel for 4 seconds. On 25% of trials, after a semi-random time between 1.0 and 1.5s following voice onset, they receive a perturbation to their somatosensory feedback that lasts for a semi-random time between 1.0 and 1.5s. They are then given a 2s rest, during which they receive visual feedback on the computer screen about the volume of their voice on the last trial, and if it was within the bounds of their baseline speaking volume.

Pilot Results

- Fundamental frequency was calculated by auto-correlation of the recorded microphone signal with 5ms window
- Individual trials were smoothed and normalized by the baseline f0 in the 0.5s preceding perturbation onset
- Trials were aligned on perturbation onset and offset, and meaned over matching experimental conditions



Figures 5a, 5b, 5c, 5d (Above): Average run results for participants. **Figure 6 (Below):** Average participant response. Shaded areas are a 95% confidence interval of the mean calculation of f0. The annotation M/NM signifies the trials where masking noise was present and when it was not present in the headphones. The **stimulation magnitude (SM)** is the mean distance from the f0 value at perturbation onset to the minimum value following perturbation onset across trials. The **response magnitude (RM)** is mean distance between the minimum f0 following onset, and the f0 value calculated at 1 second following perturbation onset across trials. The **response percent (RP)** is the mean percent return to baseline in response to the stimulation (SM/RM) across trials. Higher percents signify a more complete return to baseline. An unpaired two-sample t-test is performed on each set of trial values to check if the mean values are equal. If any set of means, between the masked and not masked conditions, were found to be significantly different ($p < 0.05$) an asterisk marks the significantly different means.



Discussion

- In the absence of auditory feedback, participants still compensate to displacement of their larynx by increasing their fundamental frequency.
- Compensation for perturbation onset was not complete regardless of whether auditory feedback was available or not, indicating that somatosensory feedback control is at least partly responsible for responses noted in prior studies.
- Compensation for the perturbation onset was greater when auditory feedback was absent (69%) than when it was present (50%), suggesting that auditory feedback control may have an effect on the response when feedback is available.
- Compensation for the f0 rebound upon perturbation removal was near-complete and did not differ between the two auditory feedback conditions.
- Intersubject differences may be the result of individual preference in using somatosensory feedback to correct errors in voice.

References

- Ramig, L. O., & Verdolini, K. (1998). Treatment Efficacy: Voice Disorders. *Journal of Speech, Language, and Hearing Research*, 41(1), S101-S116. <http://doi.org/10.1044/jslhr.4101.s101>
- Guenther, F. H. (2016). *Neural Control of Speech*. Cambridge, MA: MIT Press.
- Burnett, T. A., Senner, J. E., & Larson, C. R. (1997). Voice F0 responses to pitch-shifted auditory feedback: A preliminary study. *Journal of Voice*, 11(2), 202-211. [http://doi.org/10.1016/S0892-1997\(97\)80079-3](http://doi.org/10.1016/S0892-1997(97)80079-3)
- Loucks, T. M. J., Poletto, C. J., Saxon, K. G., & Ludlow, C. L. (2005). Laryngeal muscle responses to mechanical displacement of the thyroid cartilage in humans. *Journal of Applied Physiology*, 99(3), 922-930. <http://doi.org/10.1152/jappphysiol.00402.2004>
- Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *The Journal of Neuroscience*, 32(27), 9351-9358. <http://doi.org/10.1523/JNEUROSCI.0404-12.2012>

This study was supported by NIH grant R01 DC002852 (P.I. Frank Guenther) & NIH grant R01 DC015570 (P.I. Cara Stepp)

Contact: Dante Smith (djsmith@bu.edu)