The University of Texas at Austin
Speech, Language, and Hearing Sciences
Moody College of Communication

Methods for Investigating Continuous Speech Production and Perception with EEG

The University of Texas at Austin
Dell Medical School

Garret Kurteff[1], Amanda Martinez[1], Nicole Currens[1], Jade Holder[1], Valerie Mercado[1], Christopher Truong[1], Claire Huber[1], Paranjaya Pokharel[1], and Liberty S Hamilton[1,2]

[1]Department of Speech, Language, and Hearing Sciences, Moody College of Communication; [2]Department of Neurology, Dell Medical School, The University of Texas at Austin

Introduction

EEG is a method that historically has lent itself to the study of speech perception when compared to speech production due in large part to electromyographic (EMG) activity associated with speech articulators contaminating the signal [1]. The EEG speech production studies that do exist are restricted to single word utterances and are often not epoched to the onset of articulation.

Here we use canonical correlation analysis (CCA) [2] to remove EMG artifacts from a dual perception-production experiment. Using Event-Related Potential (ERP) analysis and Multivariate Temporal Receptive Field (mTRF) modeling, we demonstrate removal of EMG and preservation of auditory responses. We also demonstrate suppression of phonological feature encoding in continuous speech during production compared to perception.

Discussion

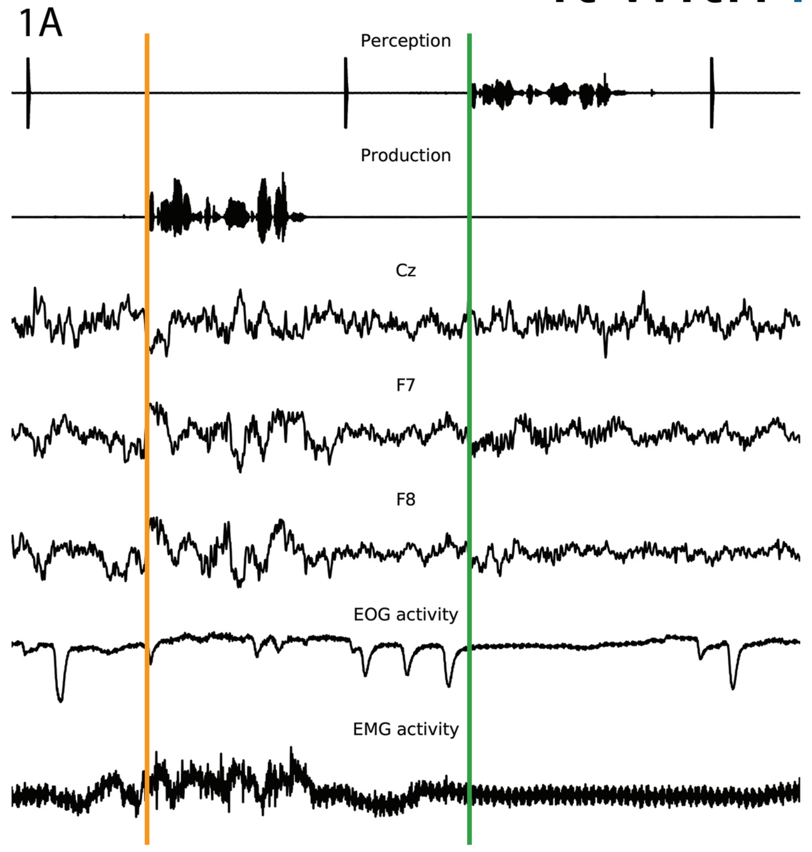
- It is feasible to study naturalistic speech production using EEG by correcting for EMG artifact using CCA.
- Proper correction of EMG artifact allows for future study of decoding speech from EEG for noninvasive BCI applications.
- mTRF modeling shows it is possible to predict EEG responses to overtly produced speech.
- Differences between speech perception and production using identical auditory stimuli may be due to feedback-based speech motor control suppressing phonological features during production.

References & Acknowledgments

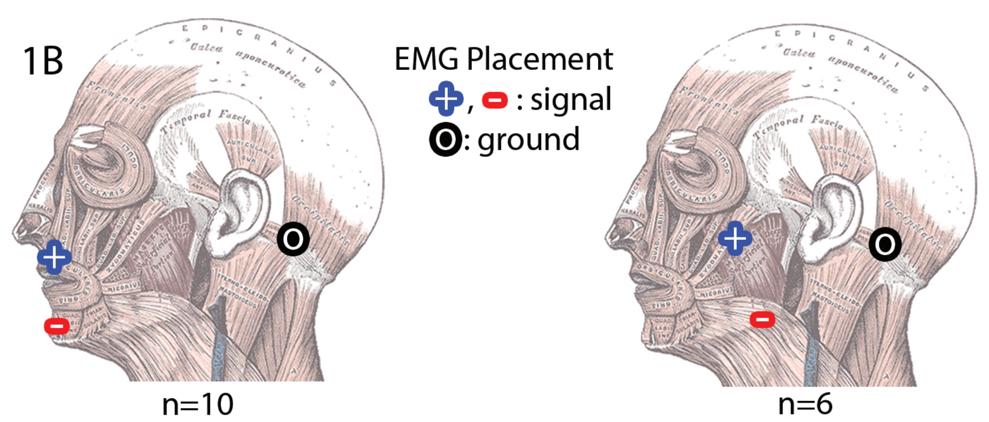
We thank the participants for their participation in this research. We thank Dr. Rosemary Lester-Smith for her assistance in recording facial EMG activity. We thank Stephanie Riès for her assistance in conducting CCA. We thank Maansi Desai and Ian Griffith for their assistance in preprocessing the data.

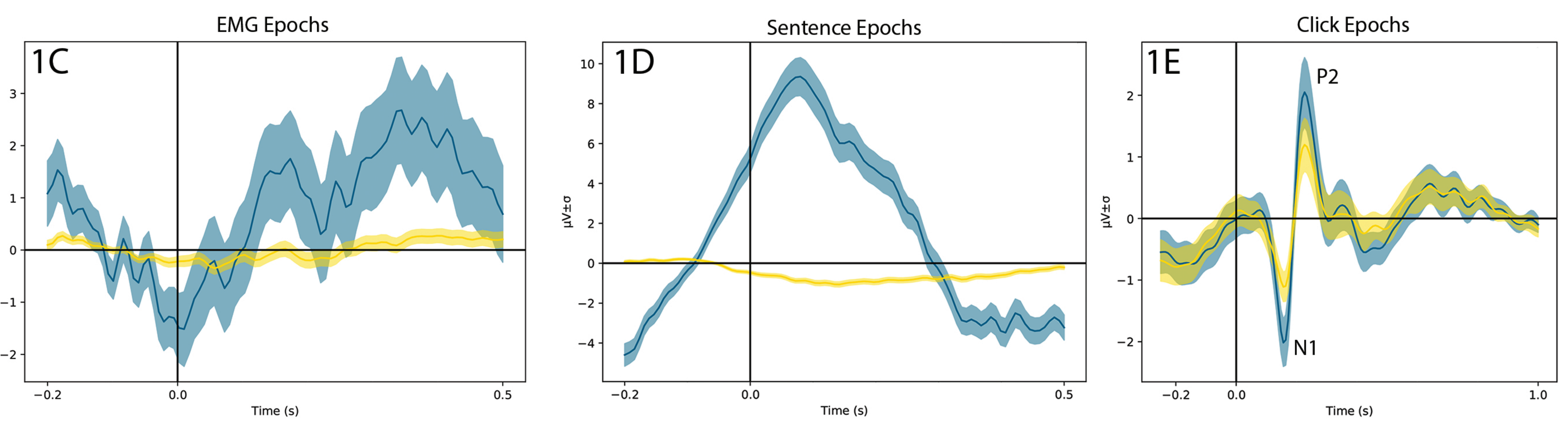
[1] Chen, X., Xu, X., Liu, A., Lee, S., Chen, X., Zhang, X., McKeown, M.J., & Wang, Z.J. (2019). Removal of Muscle Artifacts From the EEG: A Review and Recommendations. *IEEE Sensors Journal* 19(14).
[2] De Vos, M., Riès, S., Vanderperren, K., Vanrumste, B., Alario, F., Van Huffel, S., & Burle, B. (2010). Removal of muscle artifacts from EEG recordings of spoken language production. *Neuroinformatics* 8(2).
[3] Wrench, A. (1999). The MOCHA-TIMIT articulatory database.
[4] Di Liberto, G.M., O'Sullivan, J.A., & Lalor, E.C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biology* 25(19).

Evaluation of CCA artifact correction by comparing it with raw EEG data



1A: Data schematic. Data consists of two audio channels, 64 EEG, 1 vEOG and 1 EMG. Orange line indicates start of production trial; green line indicates start of perception trial.



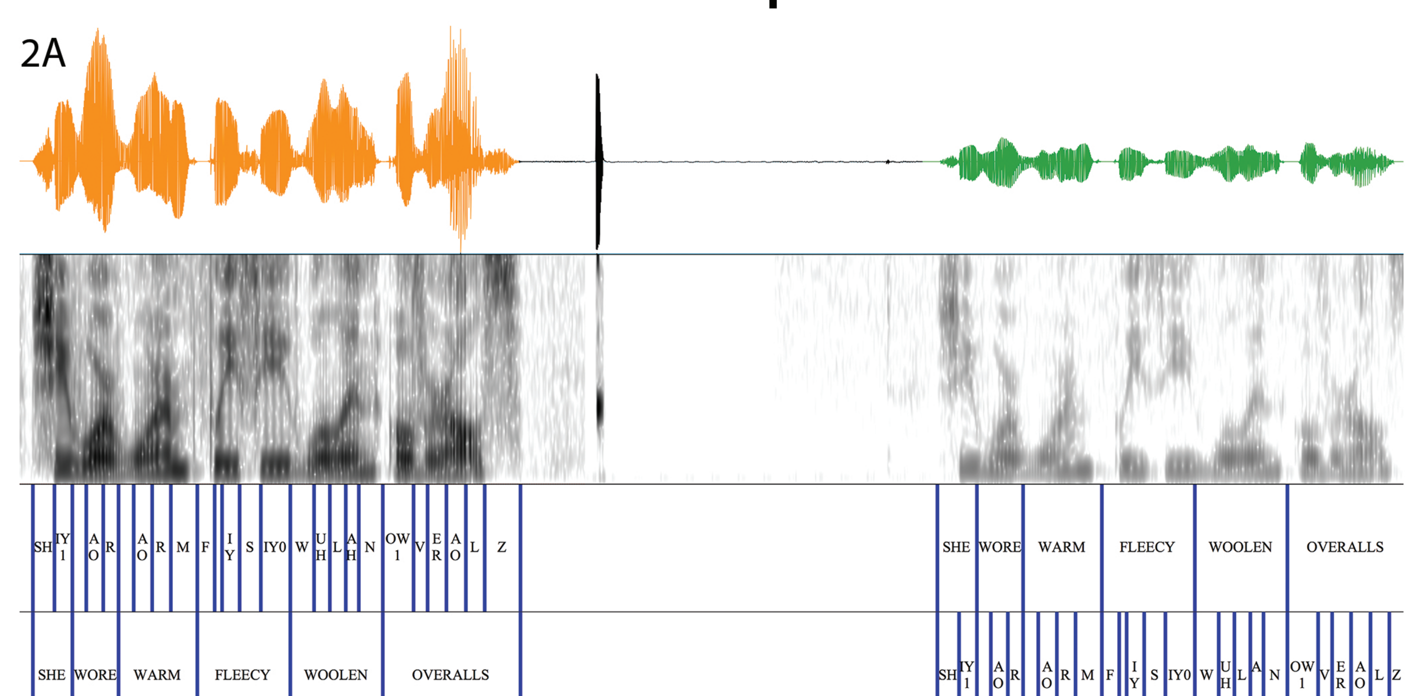


1C: Epochs generated from EMG electrodes confirm removal of EMG activity in CCA-corrected data when compared with raw data.

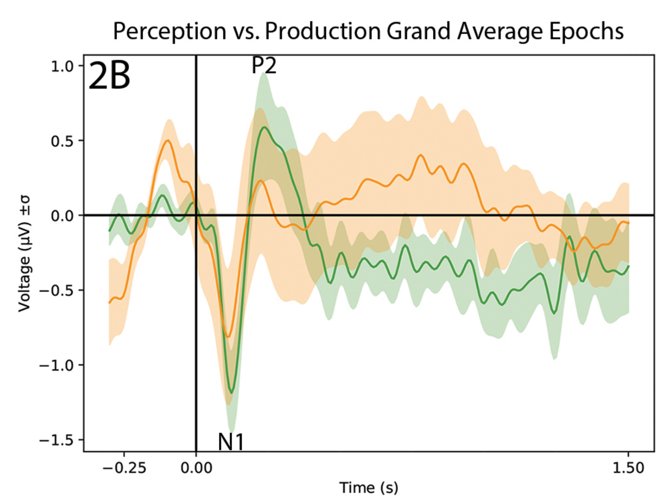
1D: Responses epoched to onset of sentence production show a reduction in artifact for CCA-corrected data when compared to raw data.

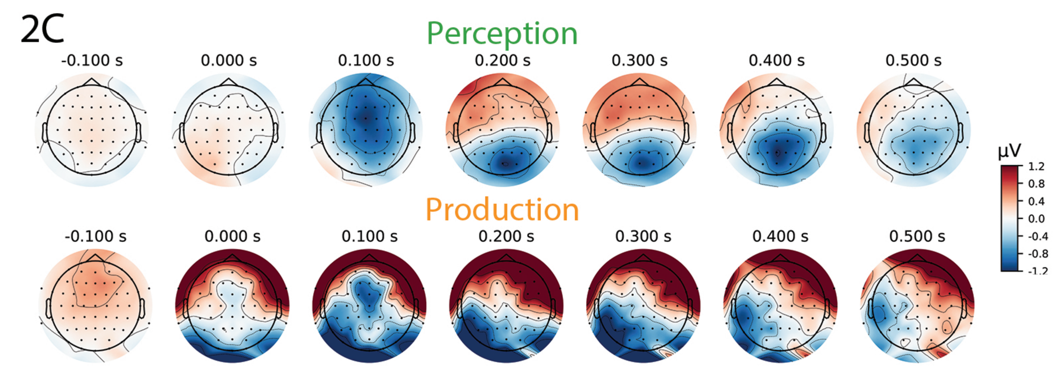
1E: Grand average response to inter-trial broadbrand click shows preservation of N1/P2 auditory responses in CCA-corrected data when compared to ICA-corrected data.

Differences between perception and production event-related potentials



2A: In the task, 19 participants overtly produced sentences from the MOCHA-TIMIT corpus [3]. After a broadband click, participants listened to playback of their production. Word, phoneme, and sentence level timing information was obtained from semi-automatically generated Praat TextGrids.

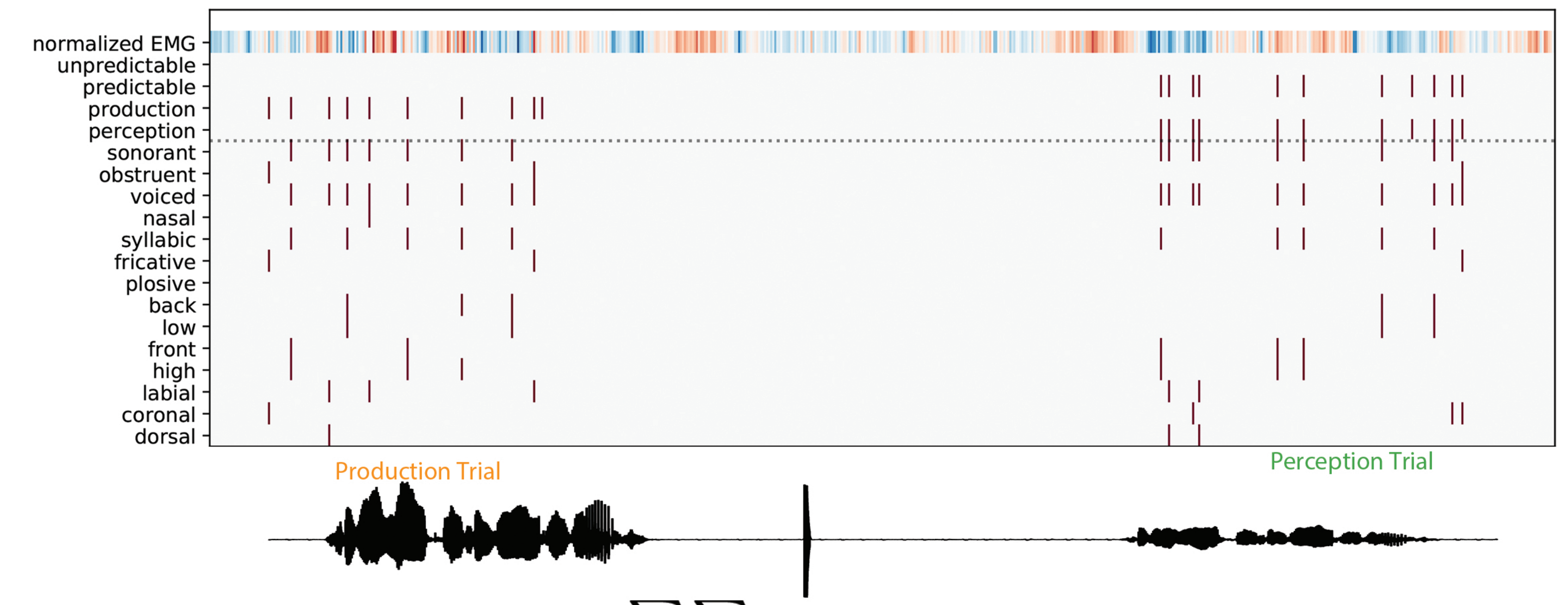




2B: Grand average sentence-level epochs comparing production and perception.

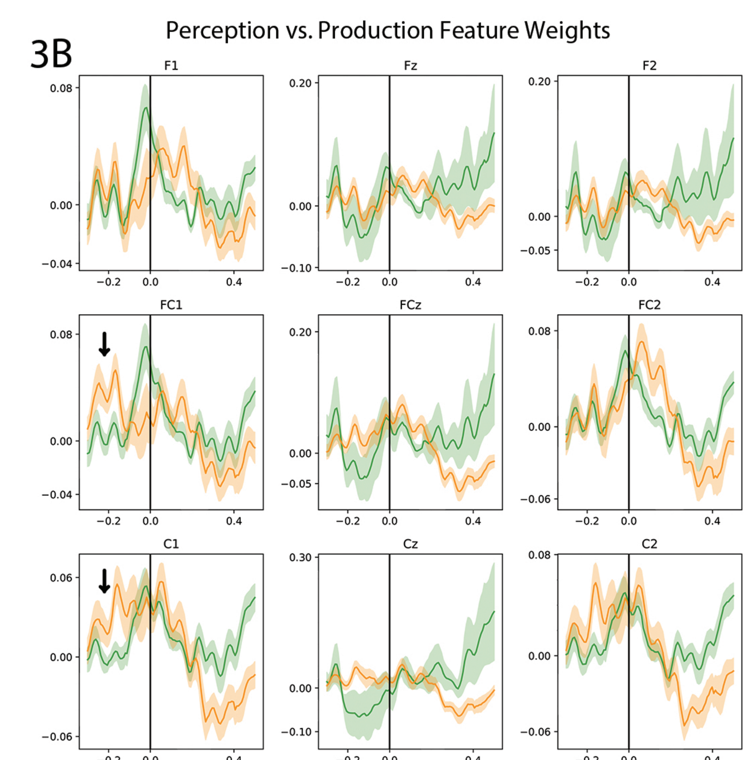
2C: Comparing topographic maps between perception and production reveals N100 activity in both conditions, with frontal/central activity at t=-0.1s present in production only.

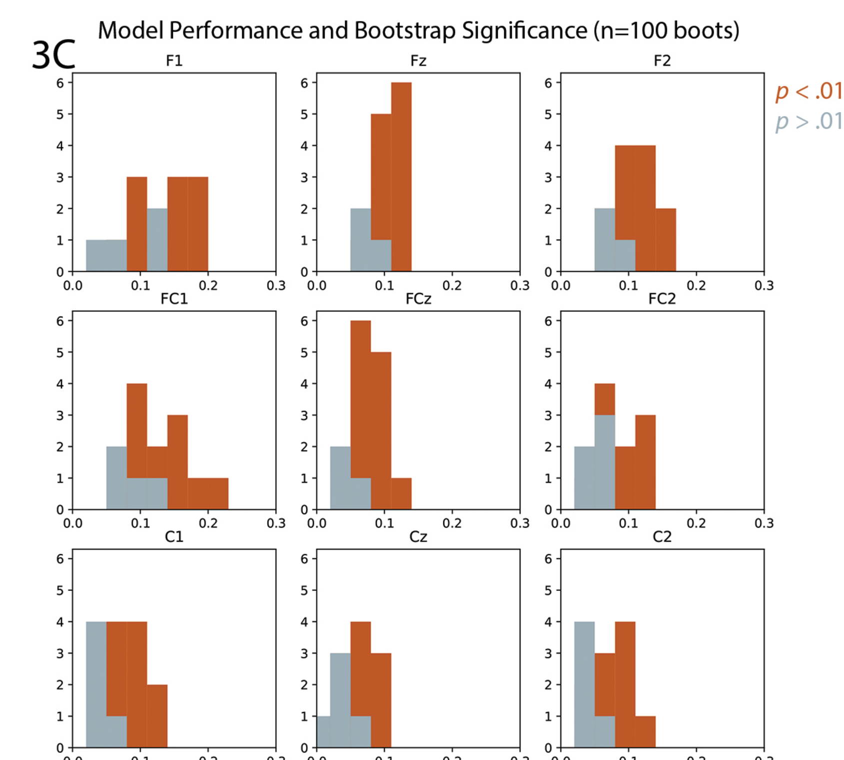
mTRF modeling of phonological features, experimental conditions, and EMG activity



$$\mathrm{EEG}(t,n) = \sum_f \sum_\tau w(f,\tau,n)s(f,t-\tau) + \epsilon(t,n)$$

3A: mTRF model features: 14 phonological features, 4 experimental conditions, and normalized EMG activity recorded from facial electrodes (1B).





3B: Model weights for perception and production epoched to onset of neural activity reveal a pre-articulatory increase in production weights.

3C: Model performance correlation values with p<.01 bootstrap significance (n=100 boots).