 DOI: 10.3217/978-3-99161-050-2-091

A Transfer Learning Framework for Across-speaker Articulatory Movement Decoding in Sensorimotor Cortex

R. Wu[1]*, J. Berezutskaya[1], Z. V. Freudenburg[1], E. C. Offenberg[1], N. F. Ramsey[1,2]

[1]*Dept. of Neurology and Neurosurgery, University Medical Centre Utrecht Brain Centre, The Netherlands*

[2]*Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands*

* Heidelberglaan 100, 3584 CX Utrecht, Netherlands. E-mail: rwu@umcutrecht.nl

Introduction: Severe paralysis resulting from medical conditions like neurodegenerative diseases and stroke can lead to speech loss. Recent advances in brain-computer interfaces demonstrate the possibility to restore speech by decoding articulatory movements from brain activity for participants who cannot speak intelligibly [1]. However, for individuals with vocal-tract paralysis, it is difficult to learn a mapping of brain data to articulatory movements without transferring from models trained on healthy individuals. Here, we propose a novel framework for 1) extracting articulatory features of spoken words shared across healthy speakers and 2) mapping them to the brain activity of individuals for whom we do not have articulatory movement data. Our findings show promising results for developing word decoding models for individuals with vocal-tract paralysis using group-level articulatory features derived from healthy speakers.

Materials and Methods: Articulatory data. We used articulatory movement data during word production from a publicly available electromagnetic articulography (EMA) dataset [2]. In the dataset, EMA were recorded in eight healthy participants. They were required to speak 97 Dutch words. Each word was repeated twice. In our study, we included seven EMA sensors: three on lips (upper lip, lower lip, and right side of the upper lip), three on the tongue (dorsal part, body, and tongue tip), and one on the chin. We analyzed articulatory movements in the front-back and the up-down directions. We applied a tensor component analysis (TCA) [3] to extract group-level low-dimensional articulatory features per word from EMA. We used dimensionality [4] as a metric to evaluate each participant's contribution to each articulatory feature. If all participants contribute equally, dimensionality is close to the number of participants. *Brain data.* High-density electrocorticographic (ECoG) recordings were obtained over the ventral sensorimotor cortex (vSMC) of the left hemisphere in three participants (P1, P2 and P3). Each participant spoke the same set of words as in the EMA dataset. TCA was applied to the ECoG recordings to extract neural features for each subject separately. We used a gradient boosting regression model to decode across-speaker articulatory features from neural features of each ECoG participant. The decoding performance was measured as the Pearson's correlation between decoded and target group-level articulatory features. Reported correlation values are cross-validated across all individual words using a leave-one-out procedure.

Results: The mean dimensionality of the extracted articulatory features was 7.61±0.14, which means the extracted articulatory features captured articulatory patterns shared across EMA participants. The mean correlation of the articulatory features between repetition 1 and 2 was 0.95±0.01, while the correlation of extracted neural features between repetition 1 and 2 was 0.70±0.07, 0.37±0.04 and 0.49±0.06 for ECoG participant P1, P2 and P3 respectively. These results demonstrate the robustness of the articulatory and neural features across repetitions. We decoded generalized articulatory features with the mean correlation of 0.78, 0.54, and 0.51 for ECoG participant P1, P2 and P3, respectively.

Conclusion: The proposed framework can extract articulatory features shared across healthy speakers and decode them from brain activity of unseen speakers. Our framework may provide a new way to develop speech BCI applications for people unable to make mouth movements.

Acknowledgments and Disclosures: This work is supported by Dutch Brain Interface Initiative (DBI2), project number 024.005.022 of the research programmed Gravitation, which is financed by the Dutch Ministry of Education, Culture, and Science (OCW) via the Dutch Research Council (NWO)

References:

[1] Metzger, Sean L., et al. "A high-performance neuroprosthesis for speech decoding and avatar control." Nature 620.7976 (2023): 1037-1046.

[2] Wieling, Martijn, et al. "Investigating dialectal differences using articulography." Journal of Phonetics 59 (2016):122-143.

[3] Williams, Alex H., et al. "Unsupervised discovery of demixed, low-dimensional neural dynamics across multiple timescales through tensor component analysis." Neuron 98.6 (2018): 1099-1115.

[4] Mazzucato, Alfredo Fontanini., et al. "Stimuli reduce the dimensionality of cortical activity." Frontiers in systems neuroscience 10 (2016): 11.

Published by Verlag der Technischen Universität Graz