DOI: 10.3217/978-3-85125-962-9-58

Continuous speech synthesis and articulatory kinematics decoding from intracortical neural activity

M. Wairagkar[1*], L. R. Hochberg[2], D. M. Brandman[1+], S. D. Stavisky[1+]

[1]University of California, Davis, CA, USA; [2]Massachusetts General Hospital/ Brown University/ Harvard Medical School/ Providence VAMC, Boston, USA

*Davis, CA, USA. E-mail: mwairagkar@ucdavis.edu + These authors contributed equally

Introduction: Brain-computer interfaces (BCIs) have the potential to restore speech in individuals who have lost the ability to speak due to ALS, stroke, or brain injury [1,2]. Intracortical BCIs have shown promise for high accuracy communication through attempted handwriting [3] and point-and-click typing [4], but these communication speeds are still slower than natural speech. Intelligible speech synthesis from a BCI has not yet been demonstrated. Here, we present ongoing progress in developing a neural decoder for speech synthesis using intracortical signals.

Material, Methods and Results: Neural activity was recorded from BrainGate2 clinical trial participant 'T5' (65 year-old male who has tetraplegia and intact speech) by implanting two Utah microelectrode arrays in dorsal (hand) motor cortex (originally for hand BCI studies) with a total of 192 electrodes (Fig 1a). We recorded neural activity and speech audio from 100 open-loop BCI trials (87 mins of speech) while T5 read out loud long passages.

We developed two multi-layered recurrent neural network decoders to (1) synthesize speech continuously and (2) decode articulatory kinematics from intracortical activity. To synthesize speech, we trained the decoder to estimate low-dimensional spectral and pitch features of speech from the corresponding spike band power of intracortical activity every 10 ms. We used the LPCNet vocoder to reconstruct audible speech from these speech features [5]. For estimating articulatory kinematics, we first used an acoustic-to-articulatory inversion model [6] to estimate kinematics trajectories of 18 articulator parameters during speech and then trained a separate RNN decoder to predict these kinematics from neural activity.

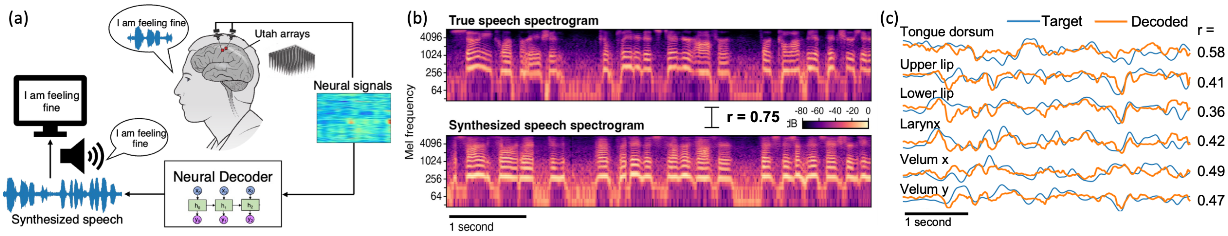

Figure 1. *(a) Speech BCI framework. (b) Speech synthesis from dorsal motor cortex. (c) Kinematics decoding for 6 example articulators.*

Offline speech synthesis from dorsal motor cortex activity yielded correlations of r = 0.75 ± 0.02 between the true speech and synthesized speech in 40 Mel spectrum bands (Fig 1b). We obtained a correlation of r = 0.42 ± 0.02 between the 18 target (estimated from audio) and decoded articulatory kinematics degrees-of-freedom (Fig 1c).

Discussion: Our neural decoder was able to synthesize speech offline with state-of-the-art accuracy. However, the reconstructed speech was not reliably intelligible, presumably due to low SNR for speech in the hand motor cortex. We view this work as a step towards developing intracortical speech BCIs recording from canonical speech brain areas, which may prove beneficial due to its potentially high SNR. We explored how a population of single neurons can predict articulatory kinematics, which could also be used for synthesizing speech using a vocal tract simulator.

Significance: We present a novel decoder framework towards developing closed loop intracortical speech BCIs for enabling high-speed natural and emotive communication. Future work will adapt the decoder to synthesize speech for people who cannot speak by generating synthetic voice data aligned to their neural activity for training. Our approach is language-agnostic and does not require a restricted vocabulary, facilitating general-purpose communication and multilanguage adaptation.

References

[1] G K Anumanchipalli, J Chartier, and E F Chang, "Speech synthesis from neural decoding of spoken sentences," Nature, 568(7753), 2019.
[2] M Angrick, et al., "Towards Closed-Loop Speech Synthesis from Stereotactic EEG: A Unit Selection Approach," in IEEE ICASSP, 2022.
[3] F R Willett, D T Avansino, L R Hochberg, et al., "High-performance brain-to-text communication via handwriting," Nature, 593(7858), 2021.
[4] C Pandarinath, et al., "High performance communication by people with paralysis using an intracortical BCI," eLife, 6(e18554), 2017.
[5] J M Valin and J Skoglund, "LPCNet: Improving Neural Speech Synthesis Through Linear Prediction," 2018. arXiv:1810.11846
[6] M Parrot, J Millet, E Dunbar, "Independent and automatic evaluation of acoustic-to-articulatory inversion models", 2019. arXiv:1911.06573

Published by Verlag der Technischen Universität Graz