 DOI: 10.3217/978-3-85125-962-9-69

Decoding articulatory trajectories during speech production from intracranial EEG

Joaquín Amigó-Vega[1,2,*], Maxime Verwoert[2], Maarten C. Ottenhoff[2], Pieter L. Kubben[2,3], and Christian Herff[2]

[1]Department of Computer Science, Gran Sasso Science Institute, L'Aquila, Italy
[2]Department of Neurosurgery, School of Mental Health and Neurosciences, Maastricht University, Maastricht, The Netherlands
[3]Academic Center for Epileptology, Kempenhaeghe/Maastricht University Medical Center, Maastricht, The Netherlands
*contact: joaquin.amigo@gssi.it

Abstract

Introduction: Speech Brain-Computer Interfaces (BCIs) are a technology that can help restore the ability to communicate of people with neurological impairments, aiming at synthesizing speech from brain signals. Most studies have focused on directly decoding text or speech segments like phonemes or words. However, it is unclear if this is how the speech production process is represented in neural recordings. An interesting approach is to model the behavior of the vocal tract, which has been successfully decoded from several brain areas. The vocal tract is composed of different physiological structures called articulators (i.e., the jaw, velum, and lips). The combination of all articulators' position and movement define the sounds heard during speech production. Recent advances have made it possible to reconstruct speech from these articulators' time trajectories, making them a good candidate for the construction of speech BCIs. This study will investigate the possibility of decoding articulatory trajectories from minimally invasive electroencephalography.

Materials and Methods: With this work, we will systematically evaluate the decoding of articulatory trajectories from neural signals and, thus, the feasibility of constructing speech BCIs with articulatory trajectories as an intermediate representation. We plan to use the SingleWordProductionDutch (SWPD) dataset presented by Verwoert *et al.* [2] where 10 participants read out individual words while stereotactic electroencephalography (sEEG) and audio data was measured.

Results: We extract articulatory trajectories from the audio using the model presented by Gao *et al.* [1]. From the sEEG recordings, we extract the high-gamma power, which contains highly localized information about speech processes. Fig. 1 shows the location of the sEEG electrodes implanted in one of the subjects from the SWPD dataset, as well as the articulatory trajectories from one recording. We train a linear regression model to predict the articulatory trajectories directly from the neural data and evaluate the reconstruction through the correlation with the actual trajectory.

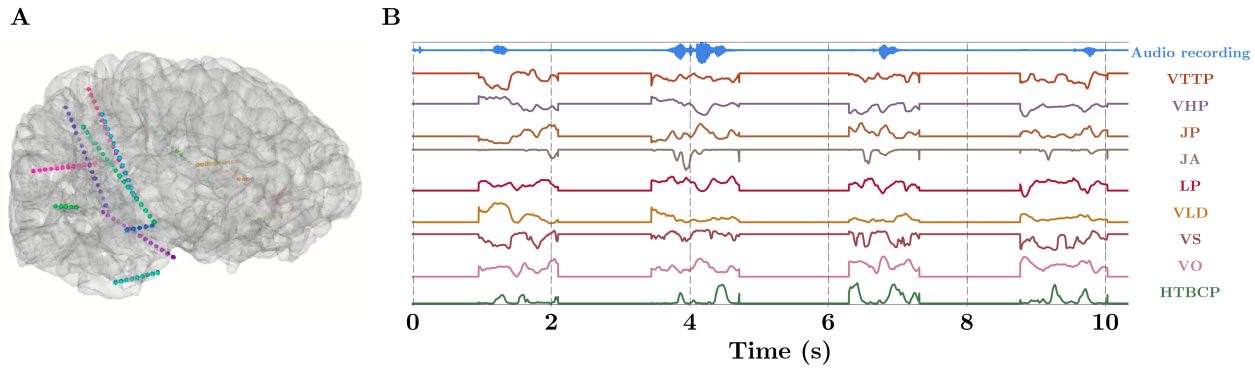

Figure 1: **A** Right lateral view of the electrode locations (subject 6, SWPD dataset). **B** Audio waveform and sample of the extracted articulatory trajectories for 10 seconds of data (subject 6, SWPD dataset). VTTP: Vertical tongue tip position, VHP: Vertical hyoid position, JP: Jaw position, JA: Jaw angle, LP: Lip protrusion, VLD: Vertical lip distance, VS: Velum shape, VO: Velum opening, HTBCP: Horizontal tongue body center position

References

[1] Y. Gao, P. Steiner, and P. Birkholz. Articulatory copy synthesis using long-short term memory networks. In A. Wendemuth, R. Böck, and I. Siegert, editors, *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2020*, pages 52–59. TUDpress, Dresden, 2020.

[2] M. Verwoert, M. C. Ottenhoff, S. Goulis, A. J. Colon, L. Wagner, S. Tousseyn, J. P. van Dijk, P. L. Kubben, and C. Herff. Dataset of Speech Production in intracranial Electroencephalography. *Sci Data*, 9(1):434, July 2022.