Directly Synthesizing Speech from Invasive Neural Recordings

M. Verwoert[1], M.C. Ottenhoff[1], S. Goulis[1], H. van Dijk[2], A. Colon[2], S. Tousseyn[2], L. Wagner[2], P. Kubben[1,3], and C. Herff[1]

[1]Department of Neurosurgery, School for Mental Health and Neuroscience, Maastricht University, Maastricht, The Netherlands
[2]Academic Center for Epileptology, Kempenhaeghe, Heeze, The Netherlands
[3]Maastricht University Medical Center, Maastricht, The Netherlands
contact: m.verwoert@maastrichtuniversity.nl

Abstract

Reconstructing speech directly from neural signals has the potential to be an intuitive tool for restoring communication in individuals who have lost the ability to speak. Recent advances have been made to reconstruct speech with techniques that are fast enough to be applied in real-time for a closed-loop brain-computer interface (BCI) system. However, the reconstructed speech has yet to reach high intelligibility. The invasive recording technique stereo-electroencephalography (sEEG) provides the opportunity to acquire signals from both cortical and subcortical structures distributed across the brain, yet the feasibility of reconstructing speech with sEEG has not been extensively studied thus far. Here, we evaluate if sEEG can be used for speech synthesis and how different pre-processing techniques impact the quality of the reconstructed speech.

We simultaneously recorded sEEG and acoustic data in epileptic patients undergoing assessment for resective surgery, while they articulated 100 uniquely prompted Dutch words continuously for five minutes. The sEEG data samples were divided into overlapping windows which were time-aligned to the corresponding acoustic data. We considered a timeframe of 200ms of neural data prior to and 50ms after the onset of the acoustic data for the reconstruction. We applied a unit selection approach to reconstruct the complete audio waveform from the neural data alone. We further investigated the influence of several pre-processing techniques on the reconstruction and on signal quality metrics.

We show that the speech audio waveform can be reconstructed using sEEG. We furthermore found that local referencing methods are beneficial compared to broader referencing in terms of apparent signal quality, however, this does not necessarily translate to improvements in the reconstruction of speech. Features were best selected based on their correlation with the speech audio in the training data. Future analyses will focus on contributing regions. This work encourages the potential use of sEEG electrodes for a speech neuroprosthesis and contributes to the knowledge on sEEG pre-processing methods.