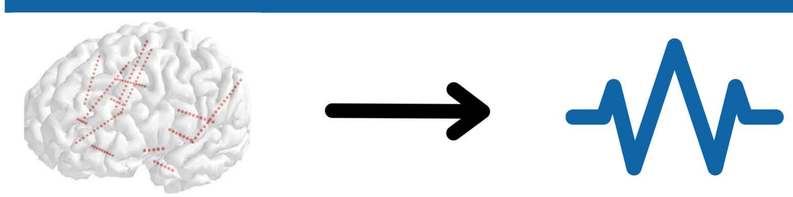
universität wien

053631 (2023W)
Data Analysis Project

Synthesizing Speech from Invasive Neural Data Using an Encoder-Decoder Framework

Damian Bednarz & Alina Behrens supervised by Anja Meunier

Motivation



We want to help people suffering from speech disorders by reconstructing speech from their neural data only.

Research Question

Can we translate stereoelectroencephalography (sEEG) data and data collected with Multi-Electrode Arrays (MEA) to speech using an encoder-decoder framework?

- Reproduce the existing work of Kohler et al. (2022)
- Adapt the framework to MEA data for a patient with speech impairment at TUM
- Use Riemannian features of covariance matrices instead of plain signals

Data

Data from Kohler et al.

- 3 patients with implanted electrodes (locations and number based on clinical necessity)
- Read out 100 sentences from the Mozilla Common, between 5 and 7 words long
- Neural data and audio data were recorded and synchronized

MEA data from TUM

- 1 patient with implanted Multi-Electrode Arrays (MEA) suffering from Broca's Aphasia
- Read out single words - 165 trials
- Neural data and audio data were recorded and synchronized

Methods

Data Preprocessing

sEEG data

- IIR bandpass filter to extract high-gamma band between 70 and 170 Hz
- Attenuation of first two harmonics of line noise using elliptic IIR notch filters
- Estimation of the signal envelope using Hilbert transform

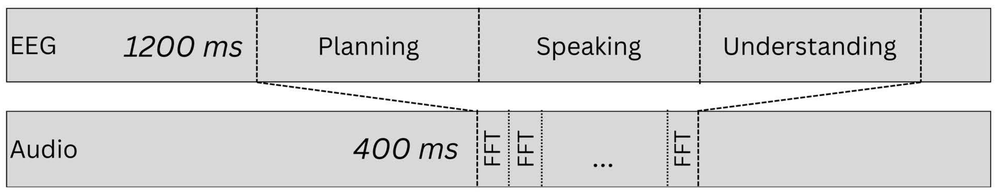
Audio data

- Transformation to mel-spectrograms
  for each 12.5 ms block:
  - short-time Fourier transformation
  - spectrally normalize using dynamic range compression

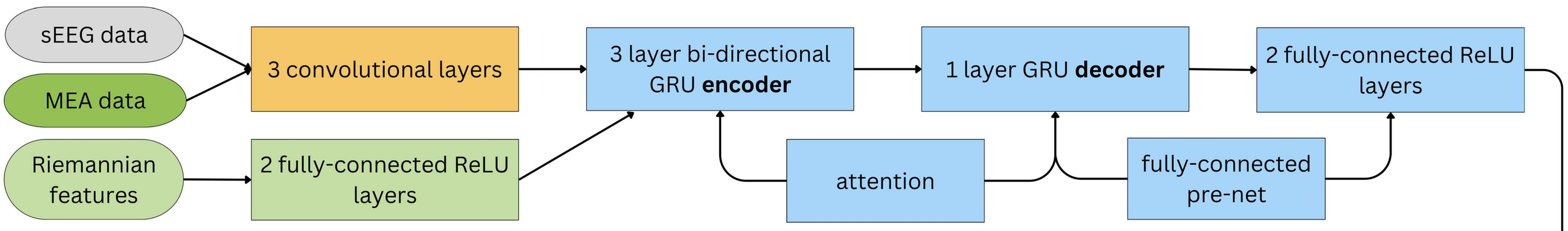
Input of the Neural Network

Training samples

Each training sample consists of 1200ms preprocessed multidimensional neural signals as X and corresponding 400ms of audio data converted to mel-spectograms as y.



Network Architecture

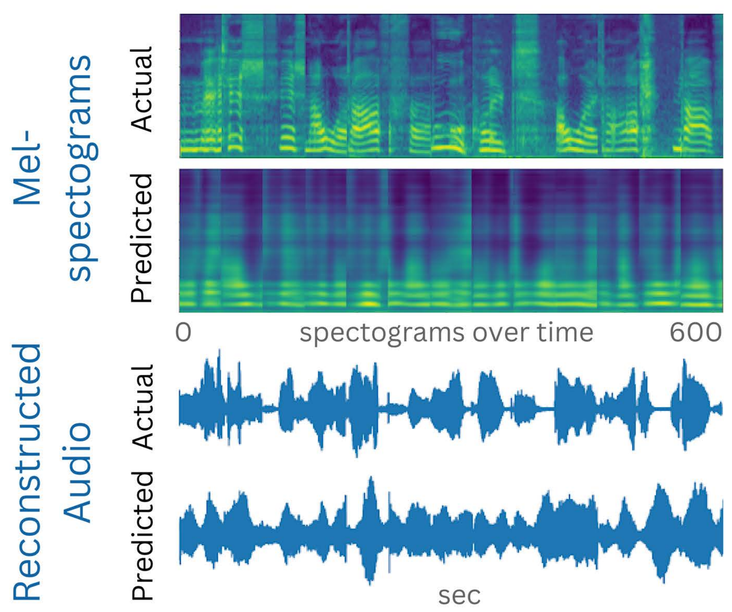


Results

- Successfully reproduced the results of Kohler et al. (2022)
- Riemannian features yield the highest correlation (0.86)
- Despite the high correlations, the reconstructed spectograms and audio signals are visually substantially different from the true data, as is visible in the two exemplary plots on the right.

Average Pearson's Correlation on the test sets of MEA data

Results from non-silent MEA data



Discussion

- Averaging over Pearson's correlations computed on single mel-spectograms is not a measure indicating good reconstruction abilities of a given model.
- Correlations are difficult to compare since data are systematically different, e.g., silence removal leads to a dataset characterized by relevant variability only.
- Riemannian methods lead to better results in terms of correlations, however cannot reproduce the original audio for the MEA data. Since the model architecture requires the training of even more parameters, the method has a lot of potential to work best on a larger dataset.
- Future research is needed to construct models that generalize over time. This is especially important for patients progressively losing their ability to communicate.

Kohler, J., Ottenhoff, M. C., Goulis, S., Angrick, M., Colon, A. J., Wagner, L., ... & Herff, C. (2022). Synthesizing speech from intracranial depth electrodes using an encoder-decoder framework. *arXiv*. https://doi.org/10.48550/arXiv.2111.01457v4.