Towards direct speech synthesis from ECoG: A pilot study

C.Herff[1], G. Johnson[2], L. Diener[1], J. Shih[3], D. Krusienski[2], T. Schultz[1]

[1]University of Bremen, Germany; [2]Old Dominion University, Norfolk, VA, USA; [3]Mayo Clinic, Jacksonville, FL, USA

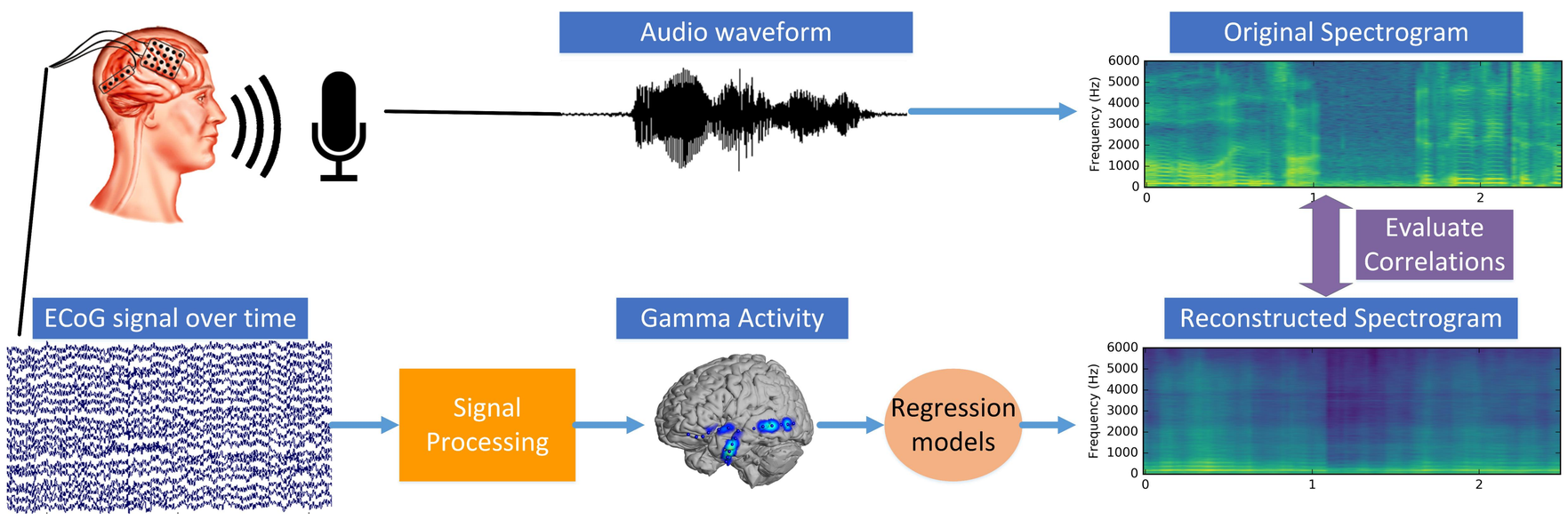
christian.herff@uni-bremen.de

Motivation

- Speech promising paradigm for BCI
- Current BCI lack opportunity to convey additional information like **accentuation**, **prosody** and **accent**
- Ultimate Goal: Direct synthesis of speech from neural activity
- Speech synthesis is fast and provides almost complete expressive power of human interaction

Experiment

- Pilot study participant repeats prompted sentences
- Sentences are presented both aurally and visually
- ECoG activity is recorded from 16 subdural ECoG electrodes
- ECoG activity is time-aligned to recorded audio

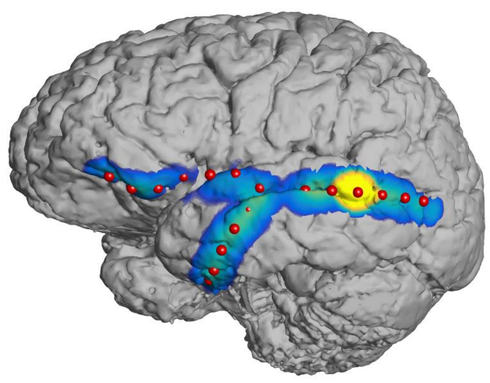


Data processing

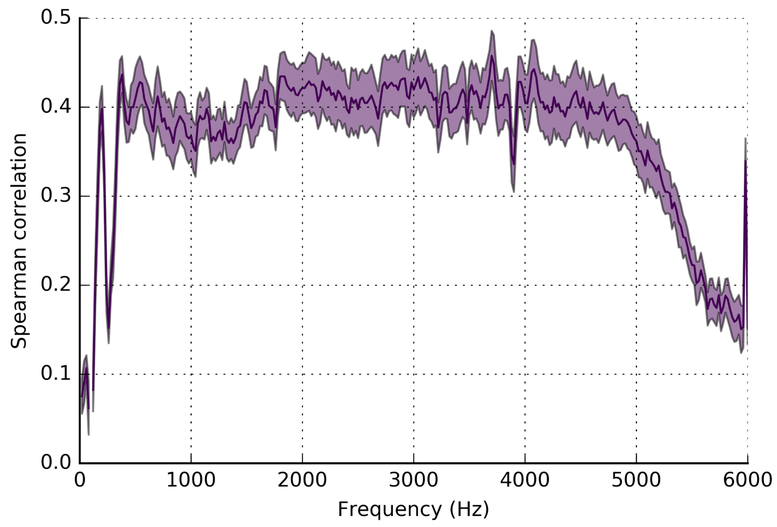
- Extract broadband-gamma activity (70-170 Hz) in 50 ms intervals, 25 ms overlap
- Add context information by stacking feature vectors up to 200 ms prior/after
- Extract audio magnitude spectrogram
- Lasso Regression model to reconstruct each spectral coefficient from neural activity
- 10-fold cross-validation
- Reconstruct waveform from reconstructed magnitude spectrogram without need for phase information using method by Griffin & Lim (1984)

Results

- Successful reconstruction of all relevant frequency ranges
- Envelope of original and reconstructed waveform correlate significantly
- Interpretation of models show usage of perisylvian areas



- All processing steps can be done in real-time
- Resulting audio not comprehensible for this electrode montage



Universität Bremen