

# Towards Closed-Loop Speech Synthesis from Stereotactic EEG using a Unit-Selection Approach

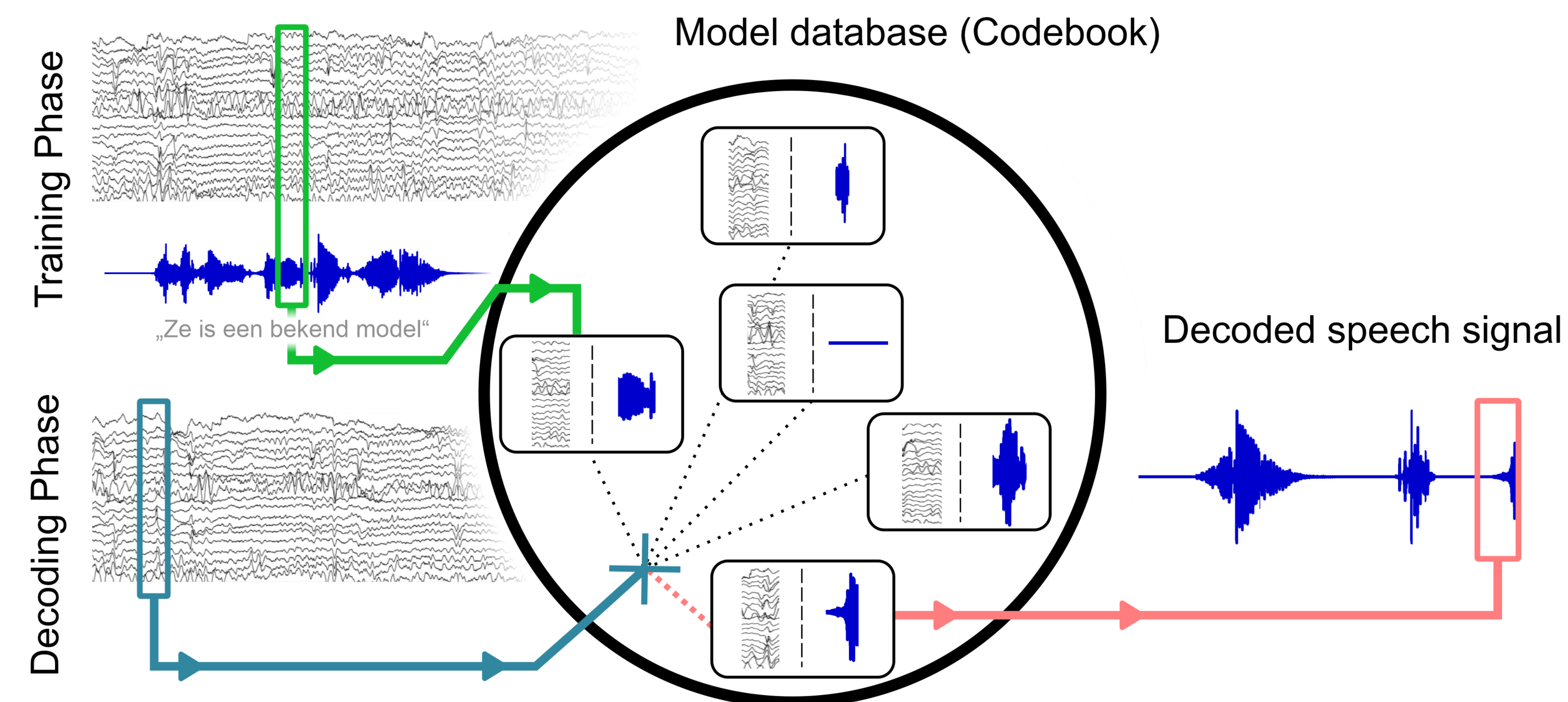
M. Angrick, M. Ottenhoff, L. Diener, D. Ivucic, G. Ivucic, S. Goulis, A. Colon, L. Wagner, D. Krusinski, P. Kubben, T. Schultz, C. Herff

miguel.angrick@uni-bremen.de

## Motivation

- Neurological diseases can lead to speech impairments
  - In severe cases even result in losing the ability to speak
- Invasive brain signals for decoding speech production processes
  - Complex spatio-temporal dynamics
  - Linear mapping not applicable
- Unit-Selection Approach
  - Enables direct speech synthesis
- Prior studies indicated sEEG recordings as suitable method
- Here, we build upon previous findings of Unit Selection approach with ECoG
  - Incorporate decoding approach in real-time framework
  - Decode speech from deeper brain structures

## Real-Time Unit-Selection Approach



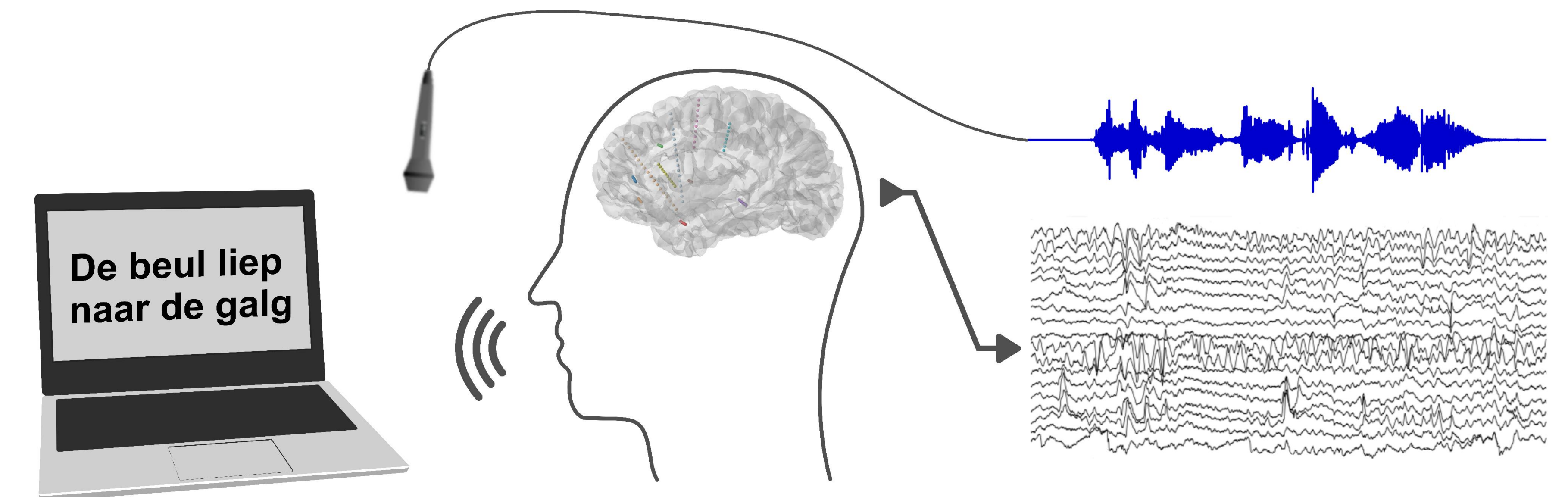
### Training Phase:

- Extraction of time-aligned frames
  - 50 ms of neural data
  - 150 ms of acoustic data
  - Both use a 10 ms frameshift
- Store pairs in model database

### Decoding Phase:

- Extract 50 ms frames of neural data
- Find closest pair in model database
  - Cosine similarity
- Waveform generation
  - Extract acoustics from closest pair
  - Hamming window for reweighting
  - Concatenate with 50 ms overlap

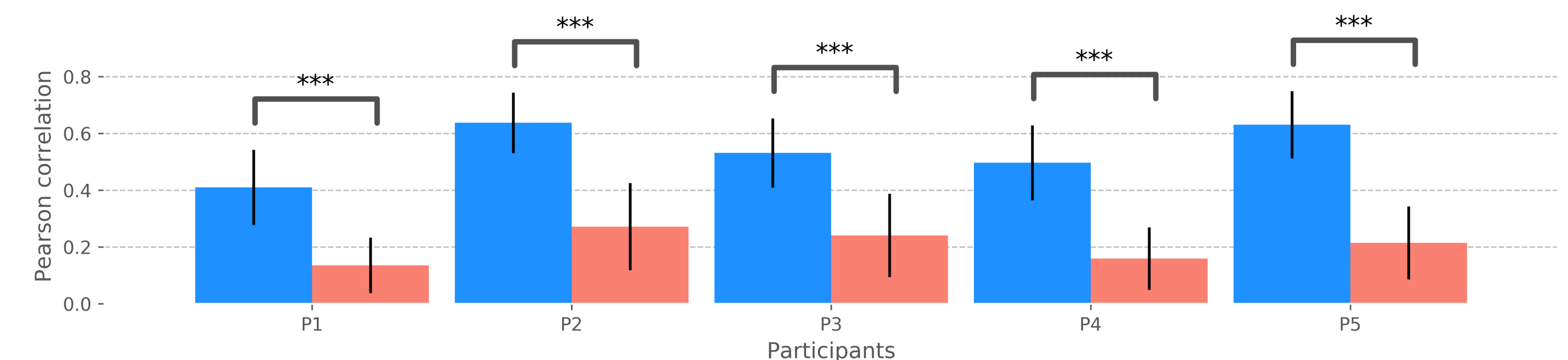
## Experiment Design



- Five native Dutch patients with implanted sEEG electrodes
- Read aloud utterances presented on screen
- Simultaneous recording of audio and sEEG data
- Acoustic data preprocessing
  - Transformed into 40 LogMels
- Neural data preprocessing
  - Extraction of high gamma activity
  - Temporal context of past 200 ms

## Results

- Reconstruct audio for each trial using 5-fold cross validation
- Compare speech spectrograms from original and reconstructed audio
  - Proposed method (blue) significantly outperforms chance (red)
  - Significance test: Mann-Whitney-U test, \*\*\*  $p < 0.001$



- For each 10 ms shifted frame, decoding approach needs 6.565 ms for computing
  - System is able to calculate output acoustic signal in real-time
- First step towards synthesizing audible speech in real-time with a non-linear decoding approach for closed-loop experiments.**