Subject-independent Classification of Japanese Spoken Sentences by Multiple Frequency Bands Phase Pattern of EEG Response during Speech Perception

Hiroki WATANABE, Hiroki TANAKA, Sakriani SAKTI, Satoshi NAKAMURA
Graduate School of Information Science,
Nara Institute of Science and Technology,
Japan
Developing a brain computer interface (BCI) for communication

- Communication prosthesis for locked-in syndrome patients
- For real-time speech communication without body movement

**BCI for online speech communication**

- Electrocorticography (ECoG)-based real-time vowel synthesizer (Guenther et al., 2009)
  - Based on invasive method
  - Higher accuracy

**Our goal**

- EEG (Electroencephalography)-based BCI to convey speech in real time
  - EEG-based:
    - Relatively low cost + Compact
    - Non-invasive method
  - But, difficult task
Our focus in the current study

- Non-invasive neural decoding of speech (EEG/MEG)
  - Imagined speech recognition
    - 2 English words (Salama et al., 2014)
    - 2 English syllables (D’Zmura et al., 2009)
    - 2 Japanese vowels (DaSalla et al., 2009)
  - Heard speech recognition
    - 3 English sentences (Luo & Poeppel, 2007, etc)
    - 5 English words (Chan et al., 2011), 2 English words (Correia et al., 2015)
    - 32 English syllables (Wang et al., 2012)
    - 8 English consonants (Wang et al., 2012)

Our current purpose: improvement of spoken sentence classification for BCI
Improvement points

Previous research (Luo & Poeppel, 2007, etc)

- **Language**
  - 3 English spoken sentences

- **Apparatus**
  - Magnetoencephalography (MEG)
    - Higher spatial resolution
    - Large + relatively high cost

- **Model training**
  - Construct models for each participant
    - New users have to collect their data
    - Time-consuming to collect a large amount of brain data from one person

- **Feature + classifier**
  - classifier: template matching
  - features: theta phase patterns

Current research

- 3 Japanese spoken sentences

- **EEG**
  - Poor spatial resolution
  - Compact + low cost

- **Subject-independent model**
  - Data collection is unnecessary before use
  - Large data set by combining all user’s data

- classifier: support vector machine
  - suitable for fewer + high dim. dataset
- features: phase patterns in various frequency bands
Phase-locked responses

- Neural oscillation tracks speech rhythm
  - Phase-locked responses
    - Phase-matching between external rhythm & brain oscillation
    - Extract linguistic information from speech

- Left hemisphere
  - phonetic rhythm: ~25ms
  - low-γ oscillation: ~40Hz (~25ms)

- Right hemisphere
  - syllabic rhythm: ~125-250ms
  - θ oscillation: ~4-8Hz (~125-250ms)

- Phase patterns inducing by phase-locked responses:
  - less individual differences (Kerlin et al., 2010)

(From Peelle and Davis, 2012, p.7)
Flowchart of classification

**Speech input**

Phase-locking brain response

**Feature extraction**

cross trial phase coherence (Luo & Poeppel, 2007)

Channel selection

- short-time Fourier transformation

- combination of phase patterns in all freq. bands

**Model training**

Template matching

SVM

Subject-independent model

Subject-dependent model

**Evaluation**

Leave-one-subject-out cross-validation

Leave-one-out cross-validation

© Hiroki Watanabe, AHC Lab, NAIST
EEG data collection

- EEG data collection
  - Participants
    - L1 10 Japanese speakers (age: $\mu=24.3$; 1 participant was excluded)
  - Speech stimuli
    - 3 Japanese spoken sentences
    - Average duration: 3,146 ms
  - Stimuli presentation
    - 24 times/ speech = total 72 presentation
    - Total # of trials = 605 trials (after artifact rejection + preprocessing)
Results
Phase-locked responses

**Crosstrial phase coherence**
(Cphase; Luo & Poeppel, 2007)

- Calculated per channel
- How consistent phase values are in each frequency band among trials

\[ \theta > \alpha, \beta, \text{low-}\gamma \]

(\( p < .05 \), paired t-tests with Holm’s p-value adjustment)
Cphase distribution map

- Cphase distribution map
  - theta
    - Right hemisphere lateralization
  - low-gamma
    - Tendency of left lateralization

θ (4-8Hz)  low-γ (38-42Hz)
Subject-dependent models

- **Accuracy improvement**
  - Template matching by $\theta$: 46.6%
  - SVM by ‘all’: 55.2%
  - + 8.6% ($p<.05$)

- **Differences b/w classifiers**
  - No differences in ‘All’ feature
Subject-independent models

Accuracy improvement
- Template by $\theta$: 38.5% ($p<.05$)
- Template by ‘all’: 44.0% ($p<.05$)
- + 5.5% ($p<.054$)

Differences b/w classifiers
- No differences in ‘All’ feature

Accuracies in subject-independent models

- SVM
- Template matching
- Baseline
- Best

Accuracy (%)

Discussion + Future plan

Phase-locked responses to Japanese spoken sentences
- syllable tracking in theta
  - tracking acoustic feature (Howard & Poeppel, 2010)
  - syllable timing: consistent across languages
- phoneme tracking in low-gamma: Not conclusive
  - might be partially due to low S/N ratio in higher frequency band

Classification performances
- Accuracy improvement in ‘All’ feature
  - above chance-level in subject-independent classification
  - based on neurophysiological speech perception model (Poeppel, 2003, etc)
- SVM: no better performances than template matching

Future plan
- Other classification algorithm + more amount of data
- application to imagined speech recognition
References

References


Thank you for listening
How to calculate Cphase

Average in time

Cphase_{stim1} ~ Cphase_{stim2} ~ Cphase_{stim3}

Average in stimulus type

Cphase

Average in each frequency band

Cphase_{\theta}