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

August 23, 2017, Interspeech 2017 | Stockholm, Sweden

Research motivation

- **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

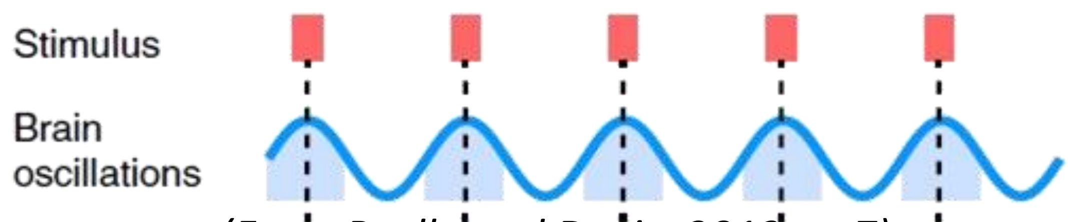
Current research

- 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



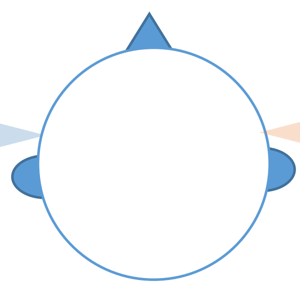
(From Peelle and Davis, 2012 , p.7)

Left hemisphere

phonetic rhythm: ~ 25ms

phase-matching

low-γ oscillation: ~40Hz
(~ 25ms)



Right hemisphere

syllabic rhythm: ~ 125-250ms

phase-matching

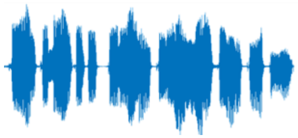
θ oscillation: ~4-8Hz
(~ 125-250ms)

(Poeppel, 2003; Luo & Poeppel, 2012)

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

Flowchart of classification

Speech input



Phase-locking

brain response

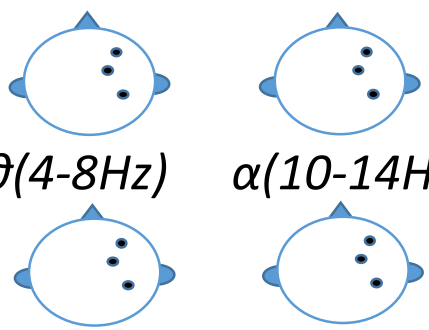
Feature extraction

crosstrial phase coherence (Luo & Poeppel, 2007)

extract phase patterns

- *short-time Fourier transformation*

Channel selection



ϑ(4-8Hz) *α(10-14Hz)*

β(16-20Hz) *low-γ(38-42Hz)*

\+ *combination of phase patterns in all freq. bands*

Model training

Template matching → Subject-independent model

SVM → Subject-dependent model

Evaluation

Leave-one-subject-out cross-validation

Leave-one-out cross-validation

EEG data collection

■ EEG data collection

- Participants
  - L1 10 Japanese speakers (age: μ=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

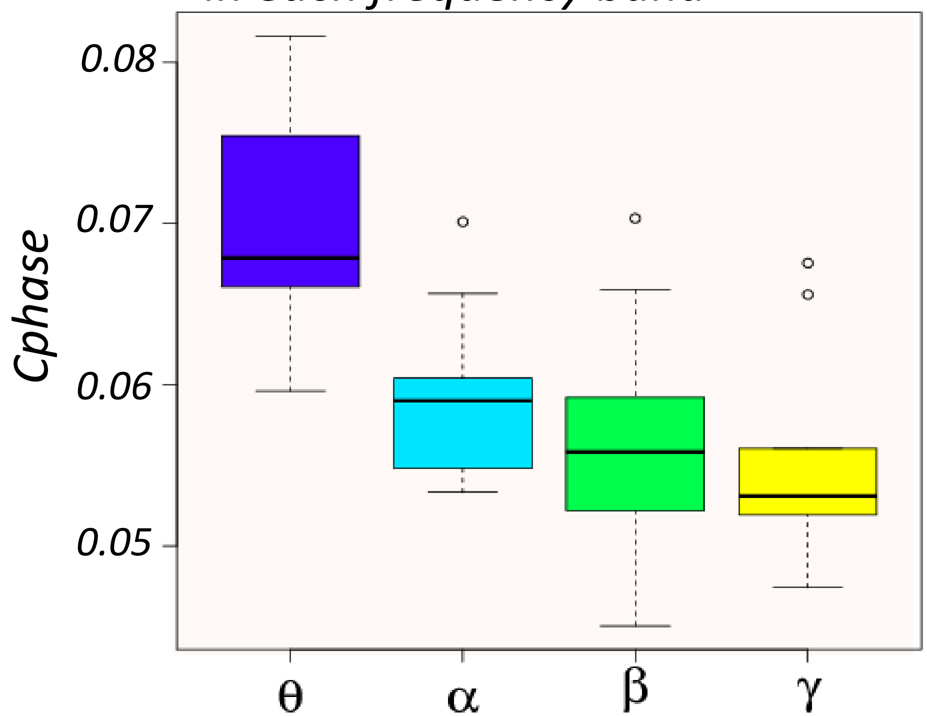
Phase-locked responses

- **Crosstrial phase coherence**
  (Cphase; Luo & Poeppel, 2007)
  - Calculated per channel
  - how consistent phase values are in each frequency band among trials

θ > α, β, low-γ

(*p*<.05, paired t-tests with Holm's p-value adjustment)

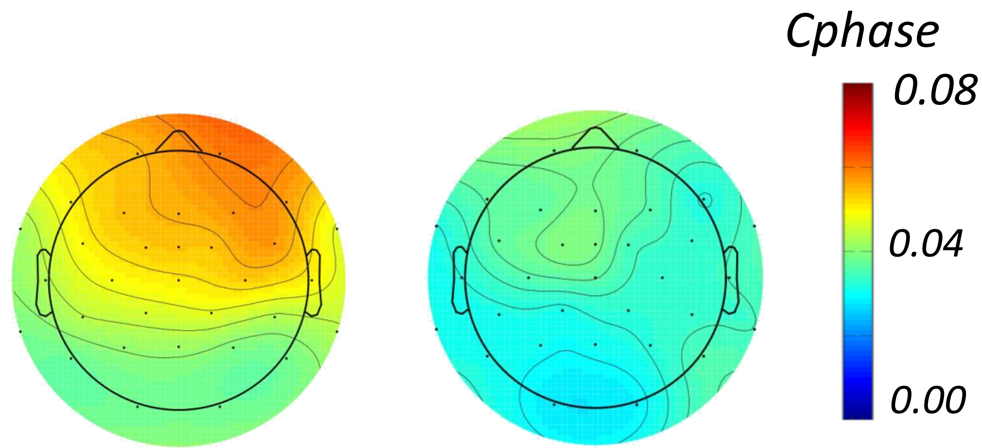
Boxplot in all subjects' Cphase in each frequency band



Cphase distribution map

Cphase distribution map

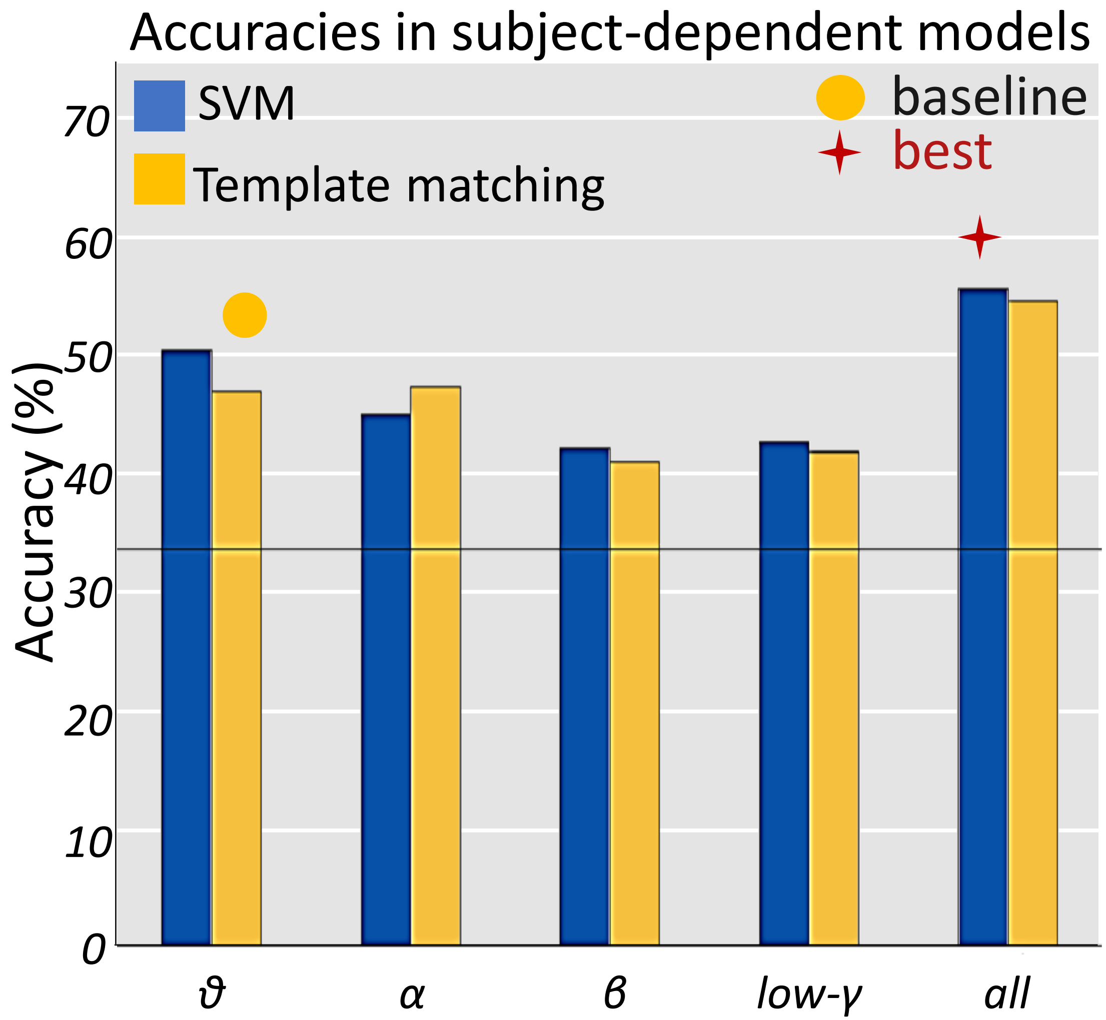
- theta
  - Right hemisphere lateralization
- low-gamma
  - Tendency of left lateralization



ϑ (4-8Hz) low-γ (38-42Hz)

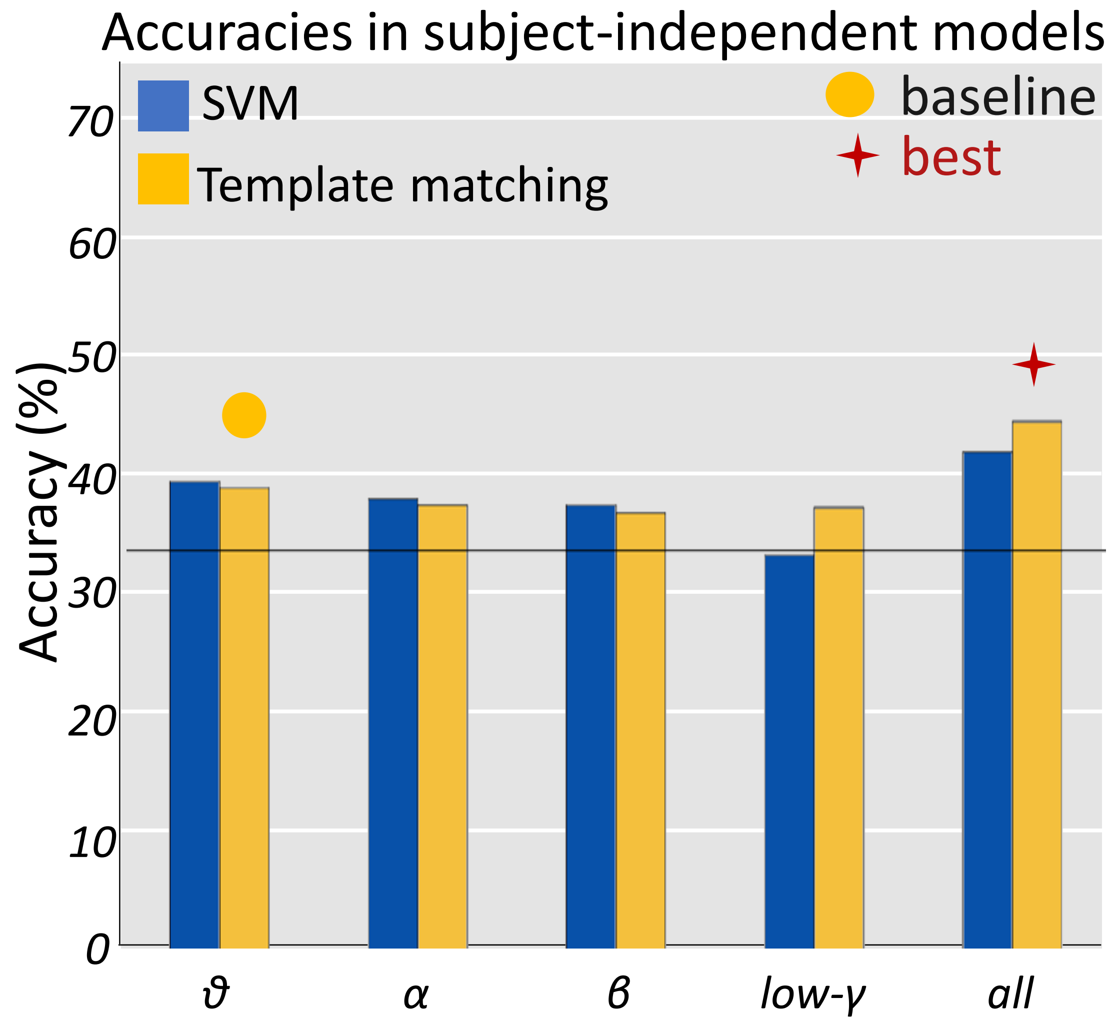
Subject-dependent models

- Accuracy improvement
  - Template matching by θ: 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 θ: 38.5% ($p<.05$)
  - Template by 'all': 44.0% ($p<.05$)
  - + 5.5% (p<.054)
- Differences b/w classifiers
  - No differences in 'All' feature



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

- Chan, A. M., Halgren, E., Marinkovic, K., & Cash, S. S. (2011). Decoding word and category-specific spatiotemporal representations from MEG and EEG. *Neuroimage, 54(4),* 3028-3039.
- Correia, J. M., Jansma, B., Hausfeld, L., Kikkert, S., Bonte, M. (2015) EEG decoding of spoken words in bilingual listeners from words to language invariant semantic-conceptual representations. *Frontiers in Psychology,* 6:1–10.
- DaSalla, C. S., Kambara, H., Sato, M., & Koike, Y. (2009). Single-trial classification of vowel speech imagery using common spatial patterns. *Neural Networks*, *22*(9), 1334-1339.
- D'Zmura, M., Deng, S., Lappas, T., Thorpe, S., & Srinivasan, R. (2009). Toward EEG sensing of imagined speech. *Human-Computer Interaction. New Trends*, 40-48.
- Guenther, F. H., Brumberg, J. S., Wright, E. J., Nieto-Castanon, A., Tourville, J. A., Panko, M., Law, R., Siebert, S. A., Bartels, J. L., Andreasen, D. S., Ehirim, P, Mao, H., and Kennedy, P. R. (2009). A wireless brain-machine interface for real-time speech synthesis. *PloS one*, *4*(12), e8218.
- Howard, M. F., & Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *Journal of Neurophysiology*, *104*(5), 2500-2511.
- Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party". *Journal of Neuroscience, 30(2),* 620-628.
- Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron, 54(6),* 1001-1010.
- Luo, H., & Poeppel, D. (2012). Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex. *Frontiers in Psychology*, *3*, 170:1-10.
- Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3, 320:1-17*.

References

- Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication*, *41*(1), 245-255.
- Salama, M., ElSherif, L., Lashin, H., and Gamal. (2014). Recognition of Unspoken Words Using Electrode Electroencephalographic Signals, *International Conference on Advanced Cognitive Technologies and Applications, 51-55.*
- Wang, R., Perreau-Guimaraes, M., Carvalhaes, C., & Suppes, P. (2012). Using phase to recognize English phonemes and their distinctive features in the brain. *Proceedings of the National Academy of Sciences, 109(50),* 20685-20690.

Thank you for listening

How to calculate Cphase