THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER.

Decoding Speech Categorization using Microstate Cortical EEG Signals and Machine Learning

Md Sultan Mahmud, PhD
Div. of General Internal Medicine
UTHSC, Memphis, TN, USA

Outline

- Introduction and Motivation
- Methods and materials
- Results and discussion
- Conclusion

Introduction and motivation

- **CP:** The process of mapping an incredibly large number of stimulus features into a smaller set of abstract grouping.
- ***Why do we consider CP as an event in our study***?
- ‘l’ and ‘r’ (e.g., “lag” and “rag” speech) in the Japanese language

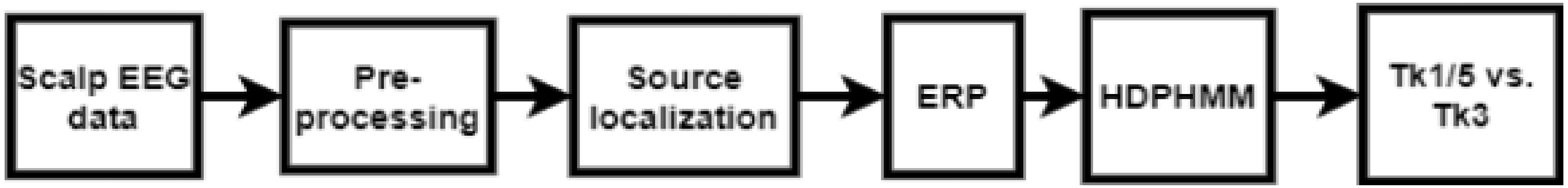

Figure 1. Working flowchart.

Methods and materials

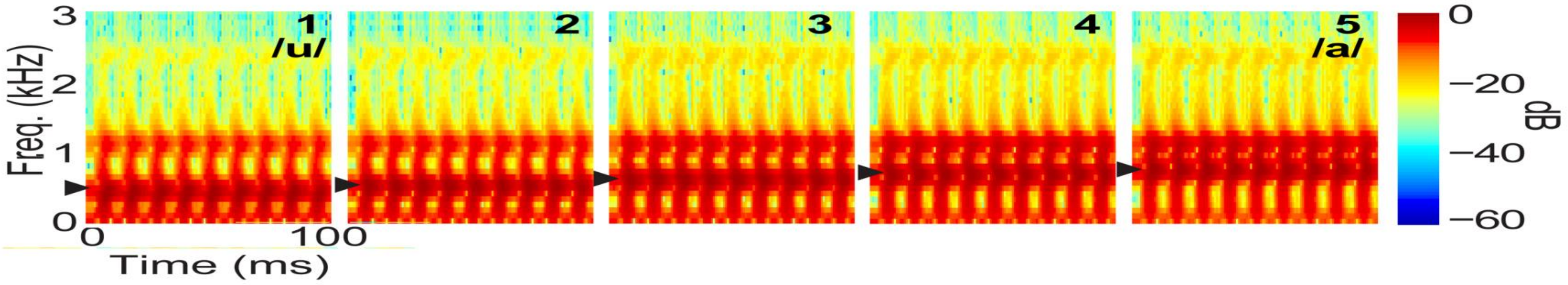

Fig 2: Speech stimuli. Acoustic spectrograms of the speech continuum from /u/ and /a/. Arrows = first formant frequency

Participants:

- 50 Participants (male 15, Female 35, age: 18~33 yrs).
- All had NH (<25 dB HL) and right-handed.
- None had a neurological or psychiatric illness.

Task, procedure, behavioral, & EEG recording:

- Stimulus were presented at 83 dB SPL
- Each token (e.g., 100 ms) presented 150~200 times
- Asked to response /u/ or /a/
- Response and RT were logged
- EEGs were recorded standard 10-10 with 64 channels
- Epoched (-200 to 800 ms) filtered 1-100 Hz (notched 60 Hz)
- BEM volume conductor and sLORETA for source localization
- Desikan-Killiany (DK) atlas has 68 ROIs

Methods and materials

- ERPs (i.e., mean activation within each ROI) averaged over randomly chosen 100 trails using Bootstrapping
- Used this ERPs to non-parametric HDP-HMM and ML classifiers
- Gaussian Mixture Model (GMM) with the Expectation Maximization (EM) algorithm
- We did the z-score normalize before feeding to the classifier.

Methods and materials

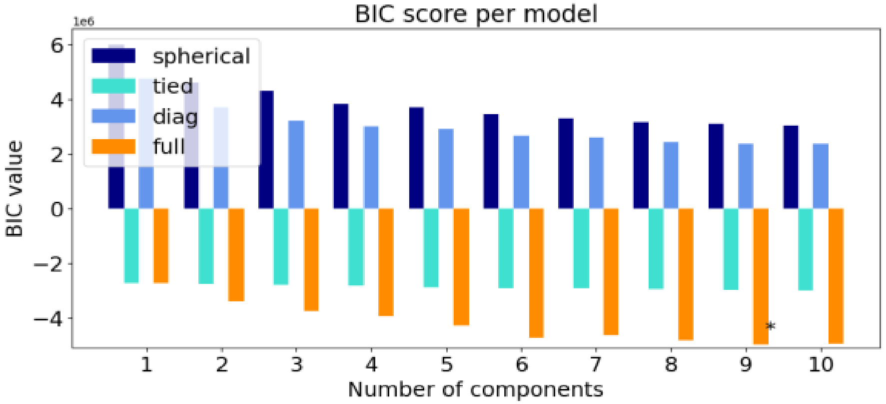

Figure 3. Clustering of ERPs data while categorizing the speech tokens. Bayesian Information Criterion (BIC) scores with different co-variance methods and number of clusters. “*" represents the lowest BIC scores, and a suitable number of clusters exist in the data.

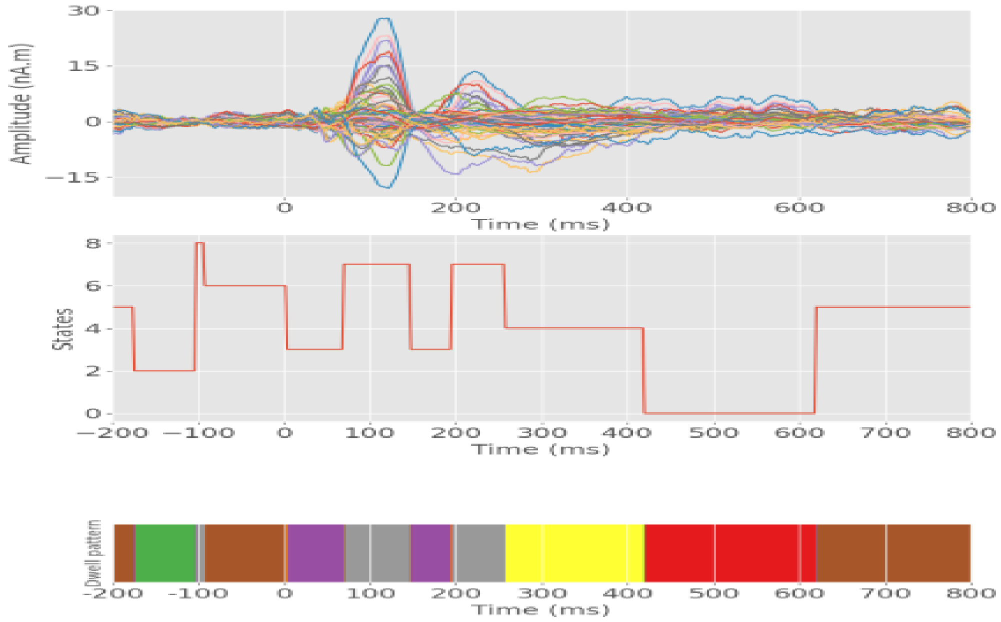

Figure 4. Representation of dwell time pattern. Top figure: Cortical ERPs signal; Middle figure: dwell-time pattern; bottom: state transition with color-coded (each color code represents a distinct state). Brown: state 5; green: state 2; light gray: state 8; light brown: state 6; purple: state 3; gray: state 7; yellow: state 4; red: state 0.

Results and discussion

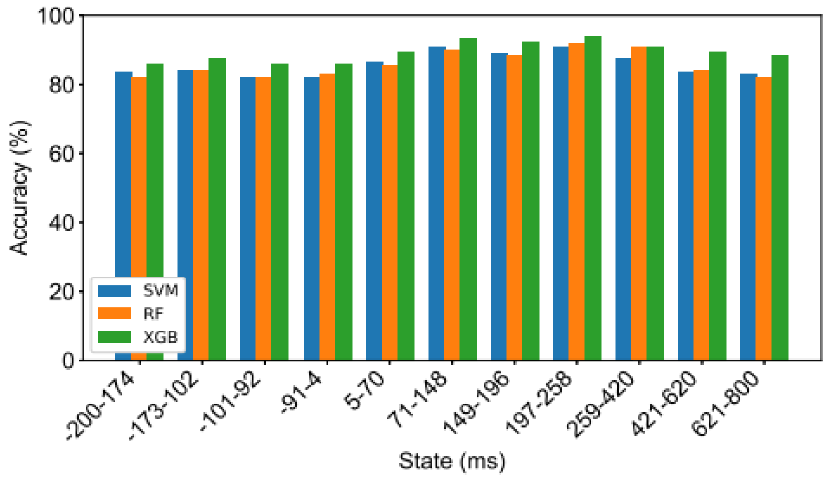

Figure 5. Prototypical speech token vs. ambiguous speech token classification accuracy over the epoch on different microstates using SVM, RF, and XGBoost classifiers. Epoch 1000 ms (-200 ms to 800 ms; -200 ms to -1 ms pre-stimulus; At 0 ms stimulus onset. Green: XGB; orange: RF; blue: SVM.

Table 1. XGBoost classifiers' performance metrics (%) on encoding window (0–260 ms) for distinguishing prototypical vowel vs. ambiguous speech tokens.

Microstate	Time Window (ms)	Average measure(%)	Whole-brain data	Top 15 ROIs
3	5-70	Accuracy	89.82	82.21
		AUC	89.13	82.20
		Precision	89.00	82.00
		Recall	89.00	82.00
		F1-score	89.00	82.00
7	71-148	Accuracy	93.61	87.82
		AUC	93.11	87.27
		Precision	93.00	88.00
		Recall	93.00	88.00
		F1-score	93.00	88.00
3	149-196	Accuracy	92.27	89.15
		AUC	92.10	89.20
		Precision	93.00	89.00
		Recall	92.00	89.00
		F1-score	92.00	89.00
7	197-258	Accuracy	94.12	90.28
		AUC	94.13	90.17
		Precision	94.00	90.00
		Recall	94.00	90.00
		F1-score	94.00	90.00

Results and discussion

- We classified the Tk1/5 vs. Tk3 based on the top 15 selected ROIs' ERPs in the encoding time window
- Lowest accuracy in the time window 5-70 ms
- Highest accuracy in the time window 197-258 ms

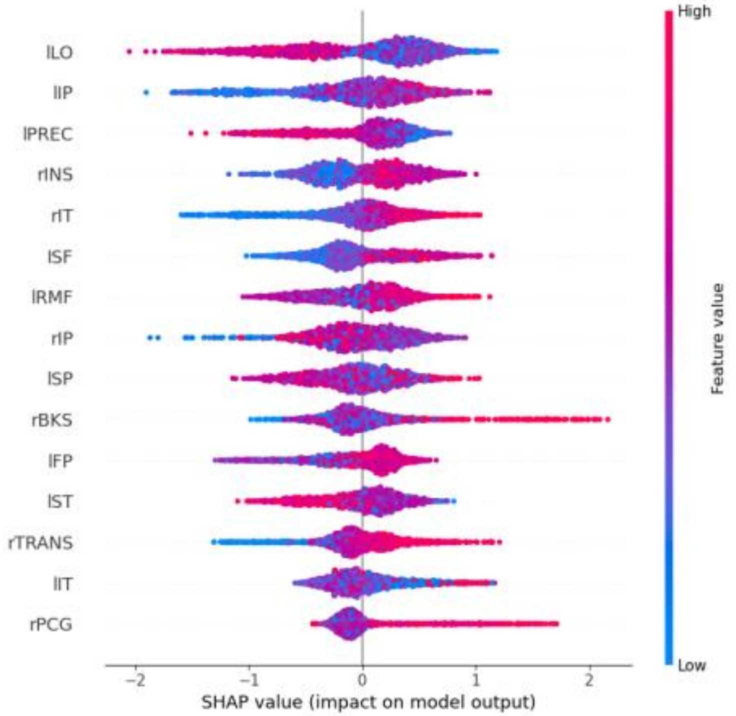

Figure 6. Density scatter plot of SHAP values obtained from XGBoost classifier. l/r: left/right; LO: lateral occipital; IP: inferior parietal; PREC: precuneus; INS: insula; IT: inferior temporal SF: superior frontal; RMF: rostral middle frontal; SP: superior parietal; BKS: bankssts ; FP: frontal pole ,ST: superior temporal; TRANS: transverse temporal; PCG: posterior cingulate.

Conclusion

- We developed a robust framework for identifying the number of microstates from brain EEG signals that exist
- Our framework showed that speech tokens can be classified most accurately 197-258 ms after speech stimulus onset
- Suggests that a surprisingly small set of brain areas actually contributes to categorizing acoustic information during speech perception
- (9/15=60%) brain regions from the LH that corroborate with LH dominant.
- Our findings could be useful to understand certain disorders that impair the perceptual mapping and learning of sound categories (e.g., dyslexia)

Thank you.

Questions?

THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER.