PhD Defense

23 April 2024

Investigating Neural Mechanisms of Word Learning and Speech Perception

Akshara Soman

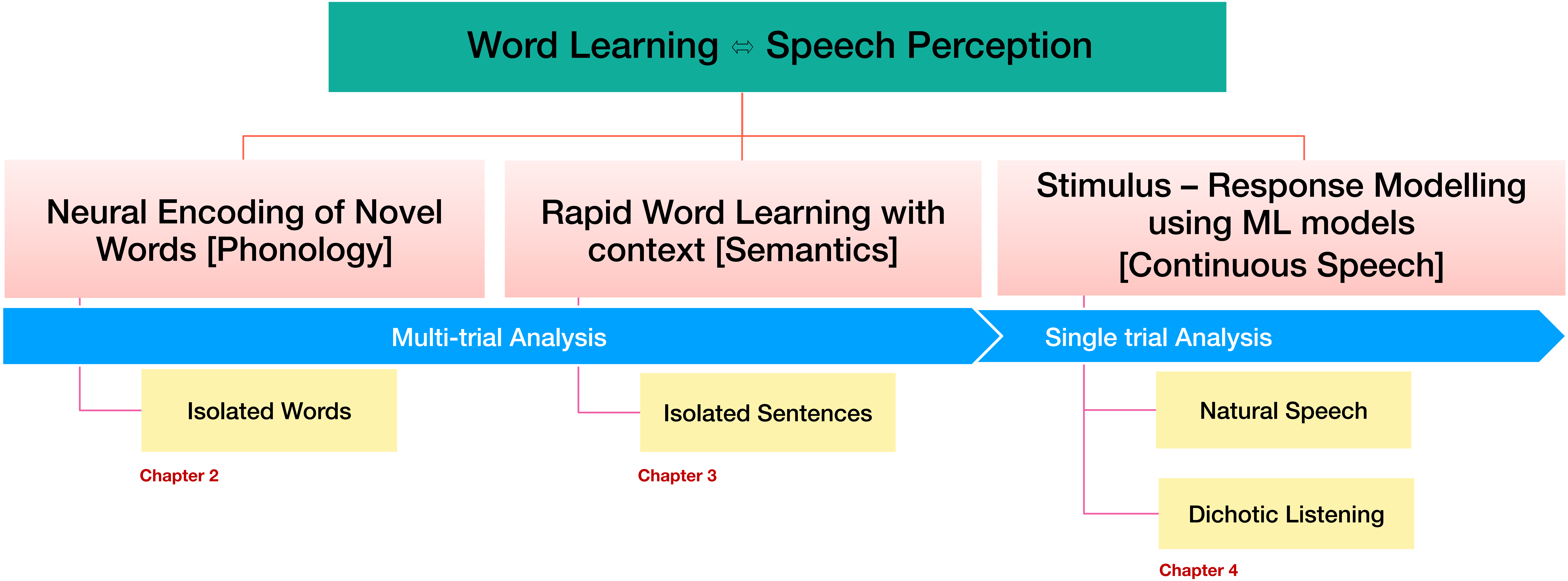
PhD,
Learning and Extraction of Acoustic Patterns (LEAP) Lab,
Electrical Engineering, Indian Institute of Science, Bangalore.

Advisor: Dr Sriram Ganapathy

LEAP

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

Outline of Presentation

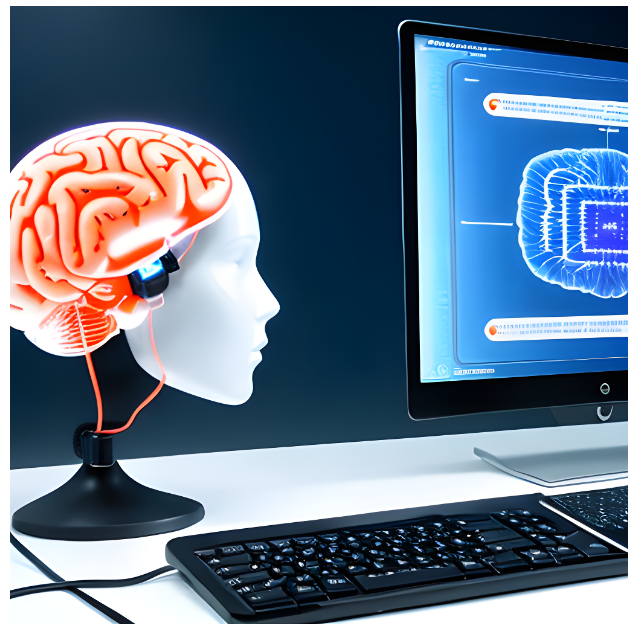


Relevance

Gain insights into how humans learn and use language

Insights on how human brain processes speech and language can be applied to improve AI systems

Develop rehabilitation strategies for individuals with speech and language disorders



Practical implications for clinical applications and brain-computer interfaces (BCIs)

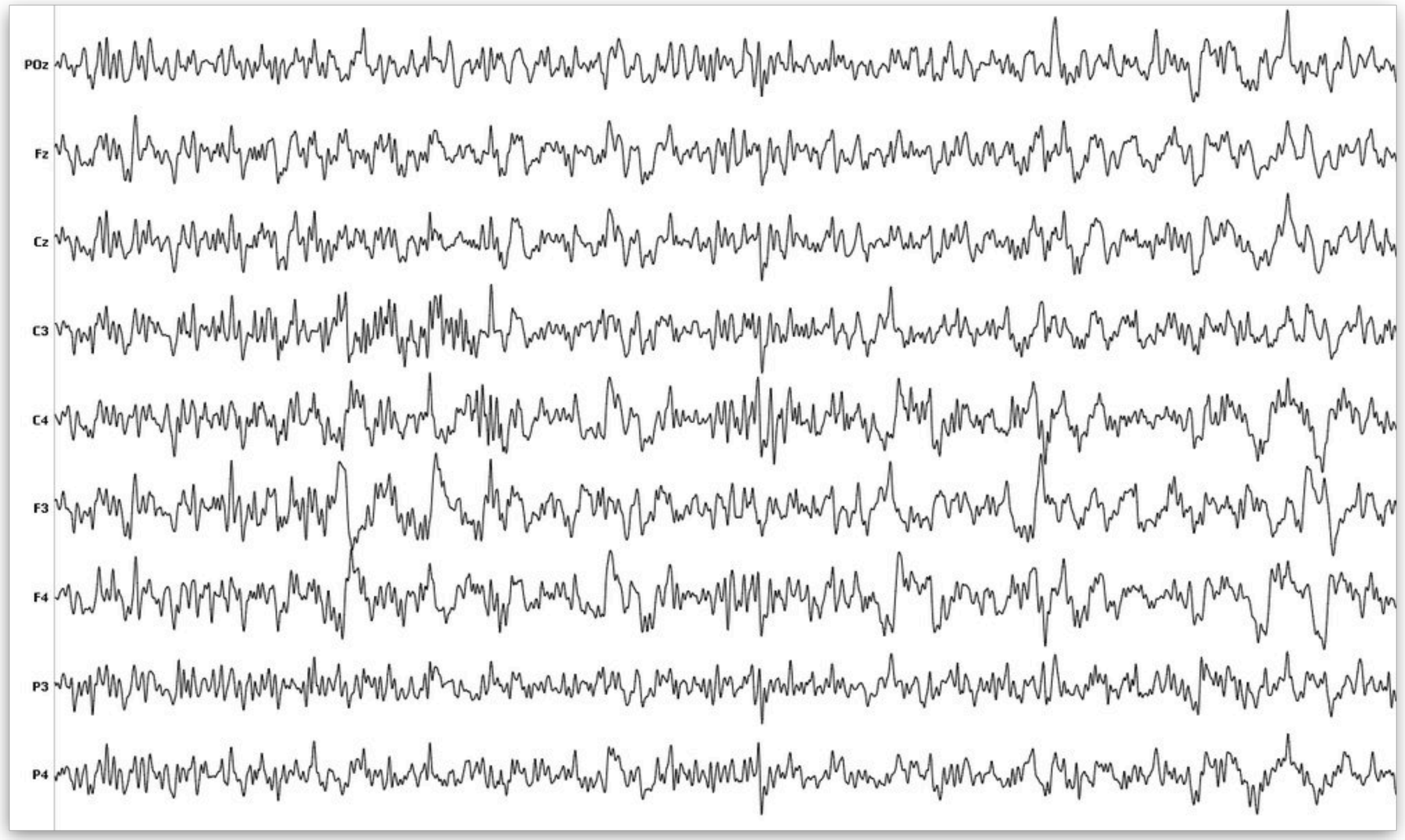
Identify effective instructional strategies and contribute to evidence-based practices in education

Main Objectives

- What are the neural correlates of language discrimination in adults learning a new language?
- What are the effects of semantic dissimilarity on EEG signals during rapid foreign language word learning?
- What is the role of word boundary information in sentence comprehension?
- In the context of a dichotic listening task, what is the relative role of semantics vs acoustics cues in speech perception?

Electroencephalogram (EEG)

- Recording of the electrical activity of the brain from the scalp.
- Signal intensity: EEG activity is quite small, measured in microvolts (μV).



- Inexpensive
- Non-invasive
- Appreciable Temporal Resolution
- Real-time capturing of cognitive processes during speech perception
- But, too noisy !!!

https://www.medicine.mcgill.ca/physio/vlab/biomed_signals/eeg_n.html

Imitate the sounds…

Word Learning ⇔ Speech Perception

- Neural Encoding of Novel Words [Phonology]
- Rapid Word Learning with context [Semantics]
- Stimulus – Response Modelling using ML models [Continuous Speech]

Motivation

- Probing neural processes in language learning
  - Learning through **repetition of sounds**.
  - Lots of open questions regarding where and how the neural representations change during a word learning process.
- In this study using EEG recordings,
  - Analyse major differences in encoding of speech from known (eg. English) and unknown (eg. Japanese) languages at the **word level**.
  - Investigate **language discriminative features** in EEG responses.

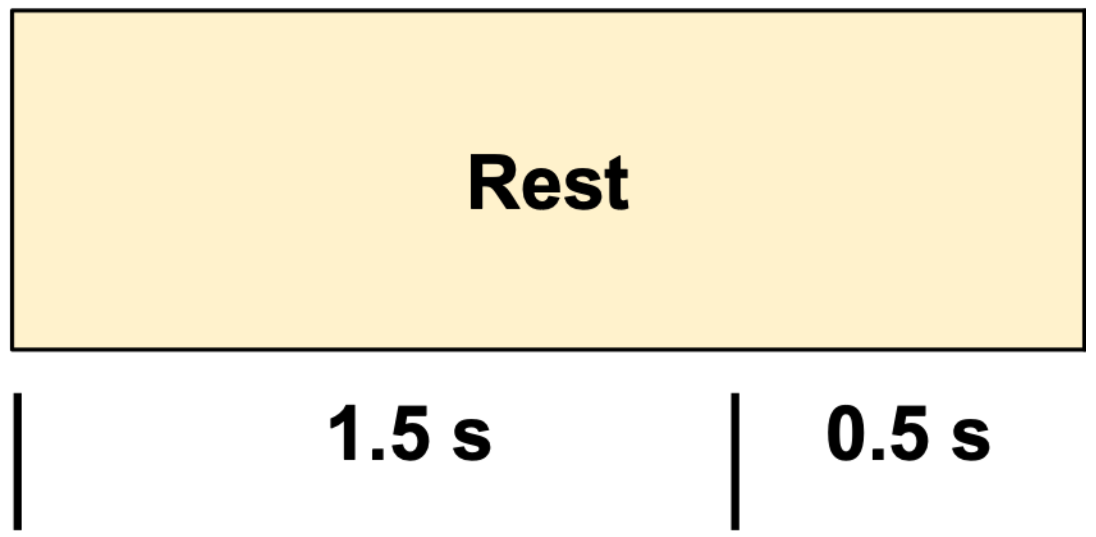
Experimental Design

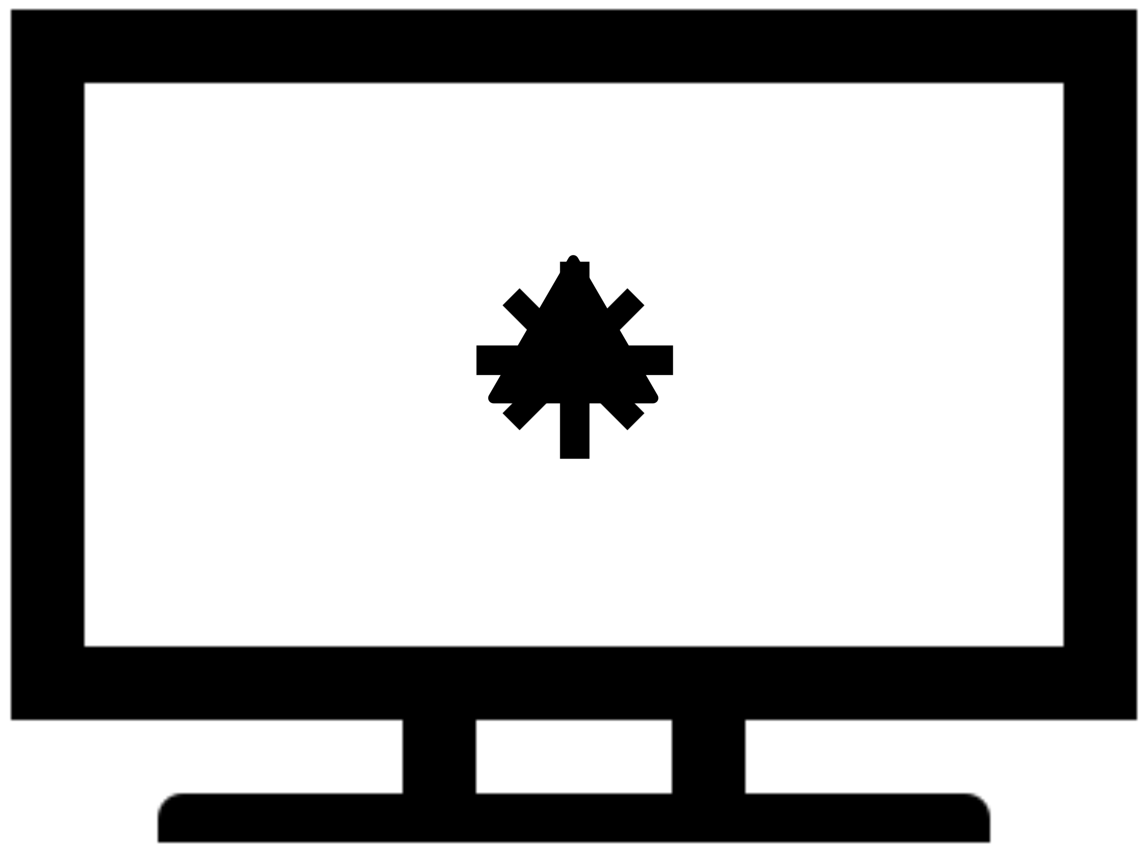
- 12 subjects: All were proficient in English and had no prior exposure to Japanese.
- 32 channel EEG cap (Axxonet System Technologies, India)
- Stimuli set consisted of 20 trials of 12 words from English / Japanese with duration of (0.5-0.8s).

English		Japanese	
Word	Length of the word in sec. (No. of speech units)	Word	Length of the word in sec. (No. of speech units)
beg	0.50 (3)	南極	0.82 (4)
cheek	0.67 (3)	抜き打ち	0.83 (4)
ditch	0.70 (3)	仏教	0.77 (3)
good	0.50 (3)	弁当	0.72 (3)
late	0.77 (3)	偶数	0.76 (2)
luck	0.64 (3)	随筆	0.83 (3)
mess	0.60 (3)	先生	0.74 (4)
mop	0.54 (3)	ポケット	0.82 (3)
road	0.59 (3)	計画	0.84 (4)
search	0.76 (3)	ミュージカル	0.83 (4)
shall	0.70 (3)	ウィークデイ	0.76 (4)
walk	0.66 (3)	行政	0.80 (3)

Experimental Design

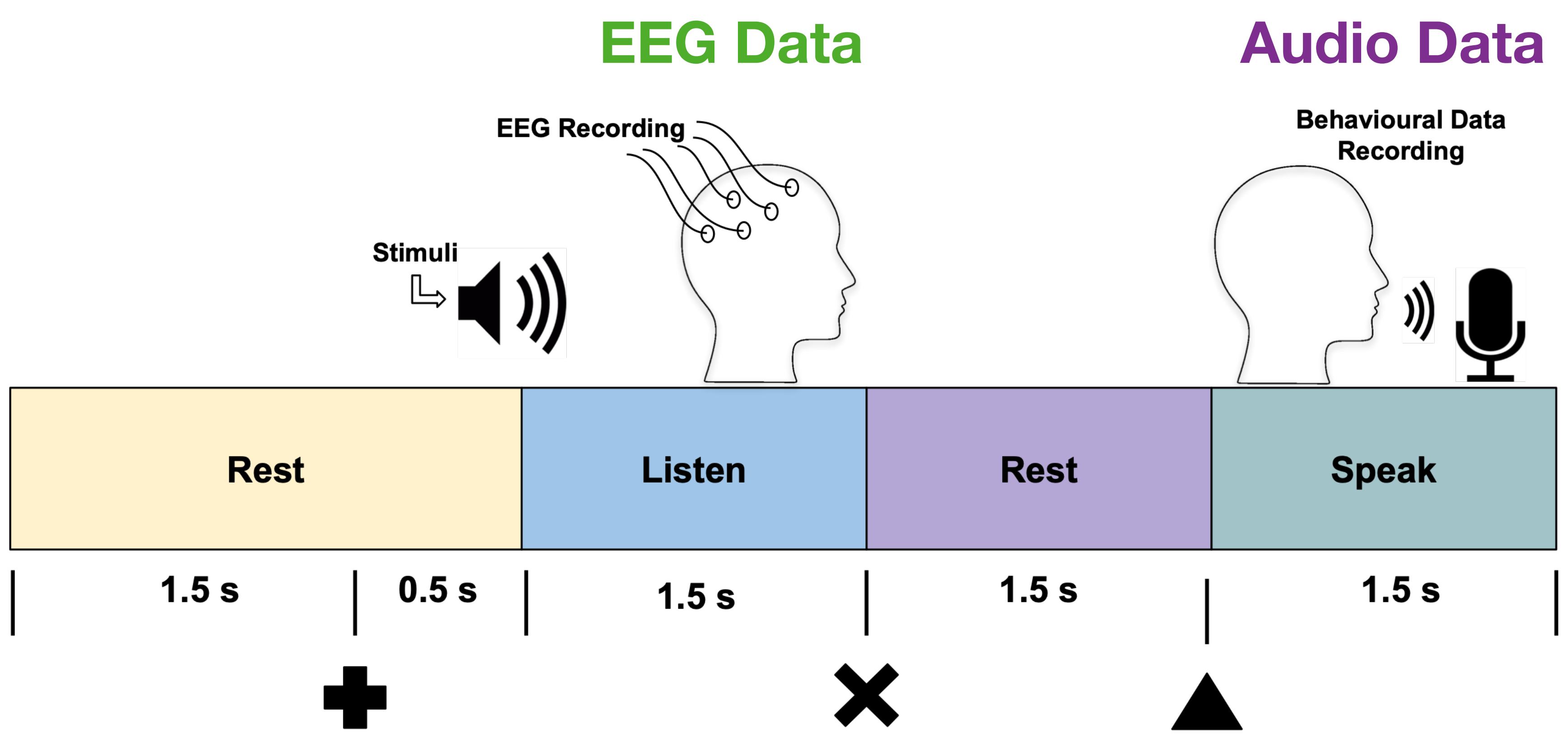
For a single word:



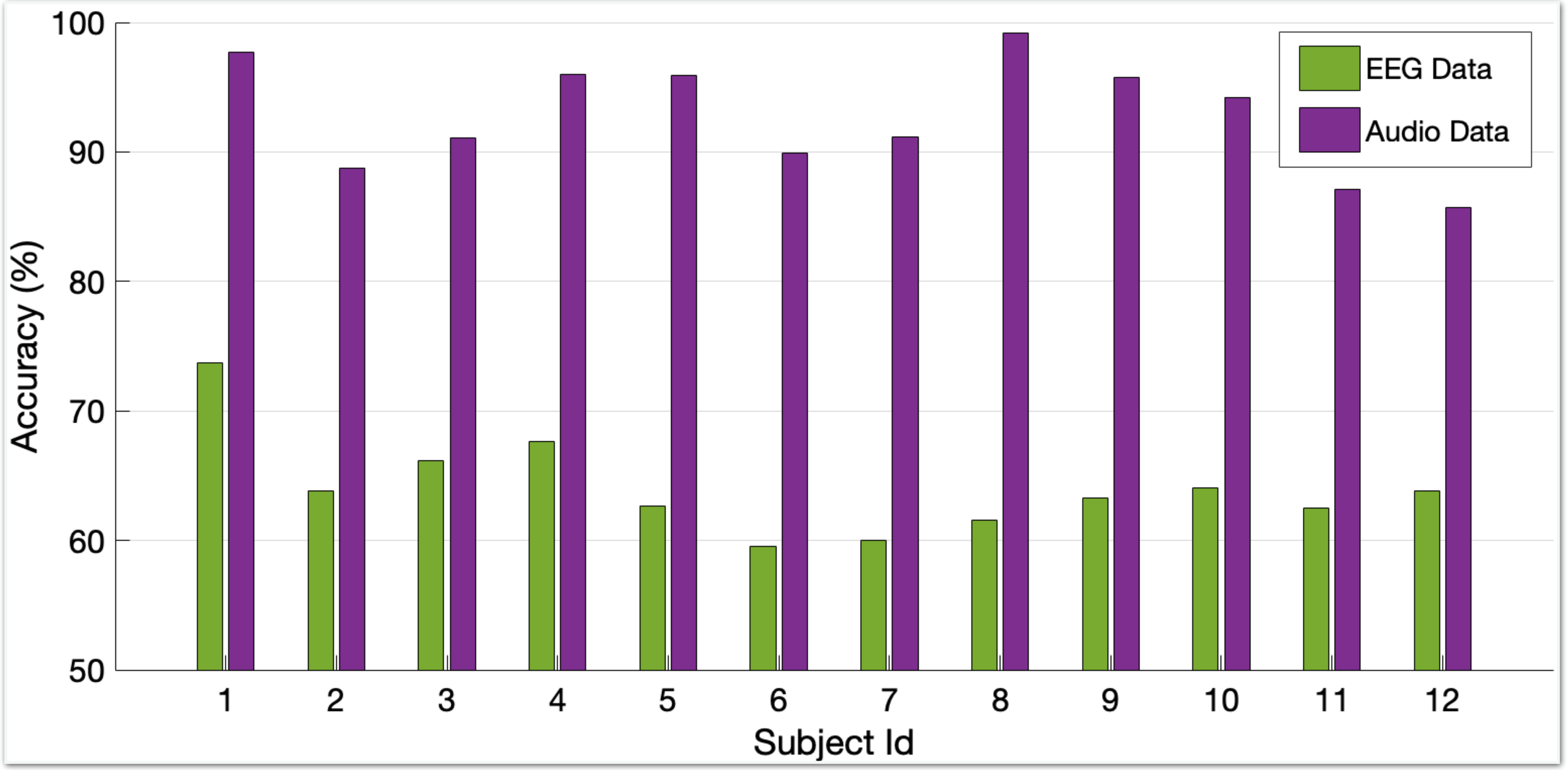


EEG analysis is performed with signals recorded during **listening**

EEG for Language Discrimination



1. Can we identify the listened language from EEG signals?

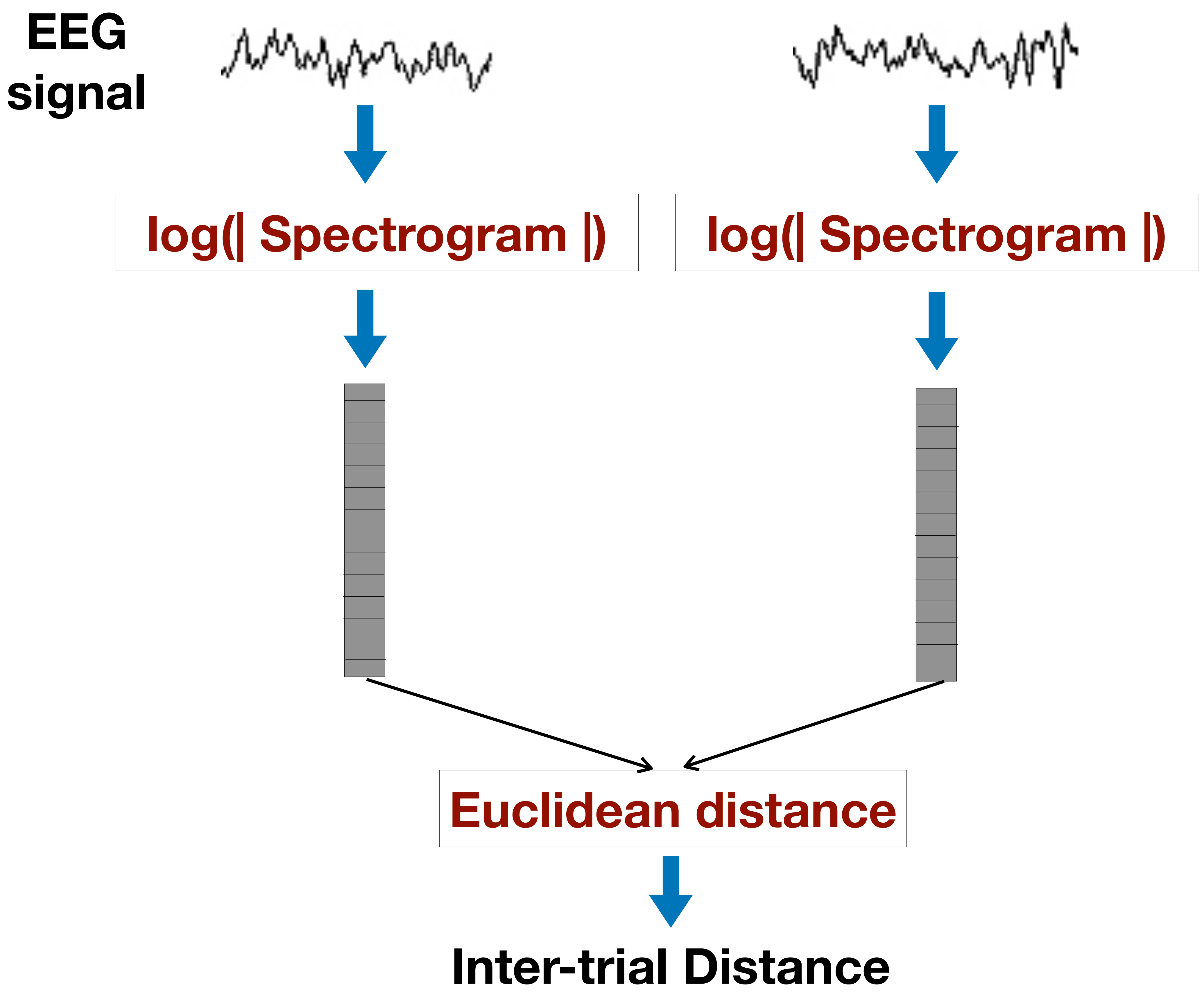


- Support Vector Machine (SVM) classifier trained on EEG spectrogram features (0-30Hz)
- Average accuracy for EEG: 64% (t-test at $\alpha = 0.01$, $p < 0.001$)
- Top-line reference: Spoken audio based classifier (using MFCC features)

✓ EEG signals encode language specific information.

Inter-trial Distance Analysis

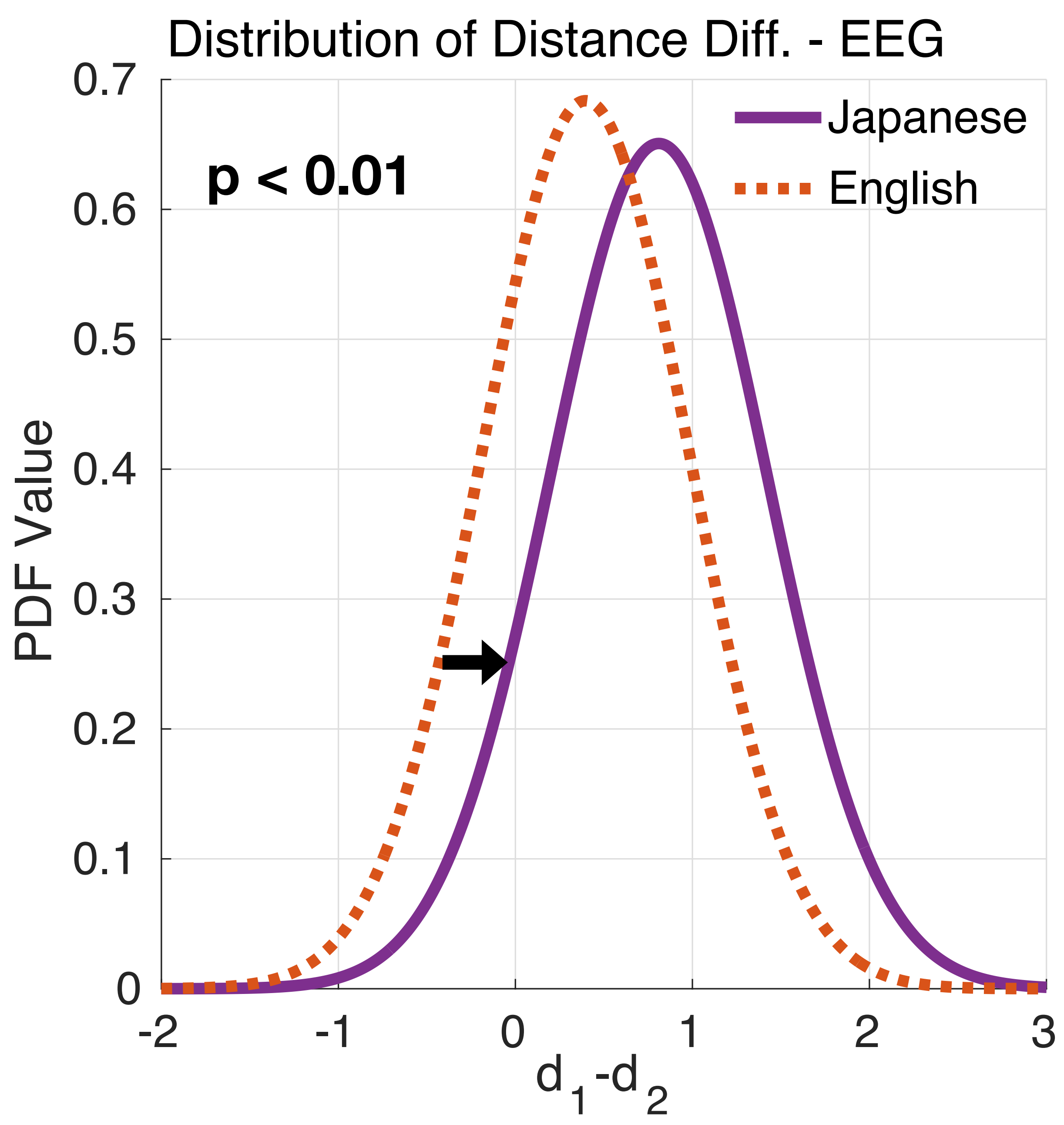
2. Are the neural encodings converging to a consistent pattern for a novel word?

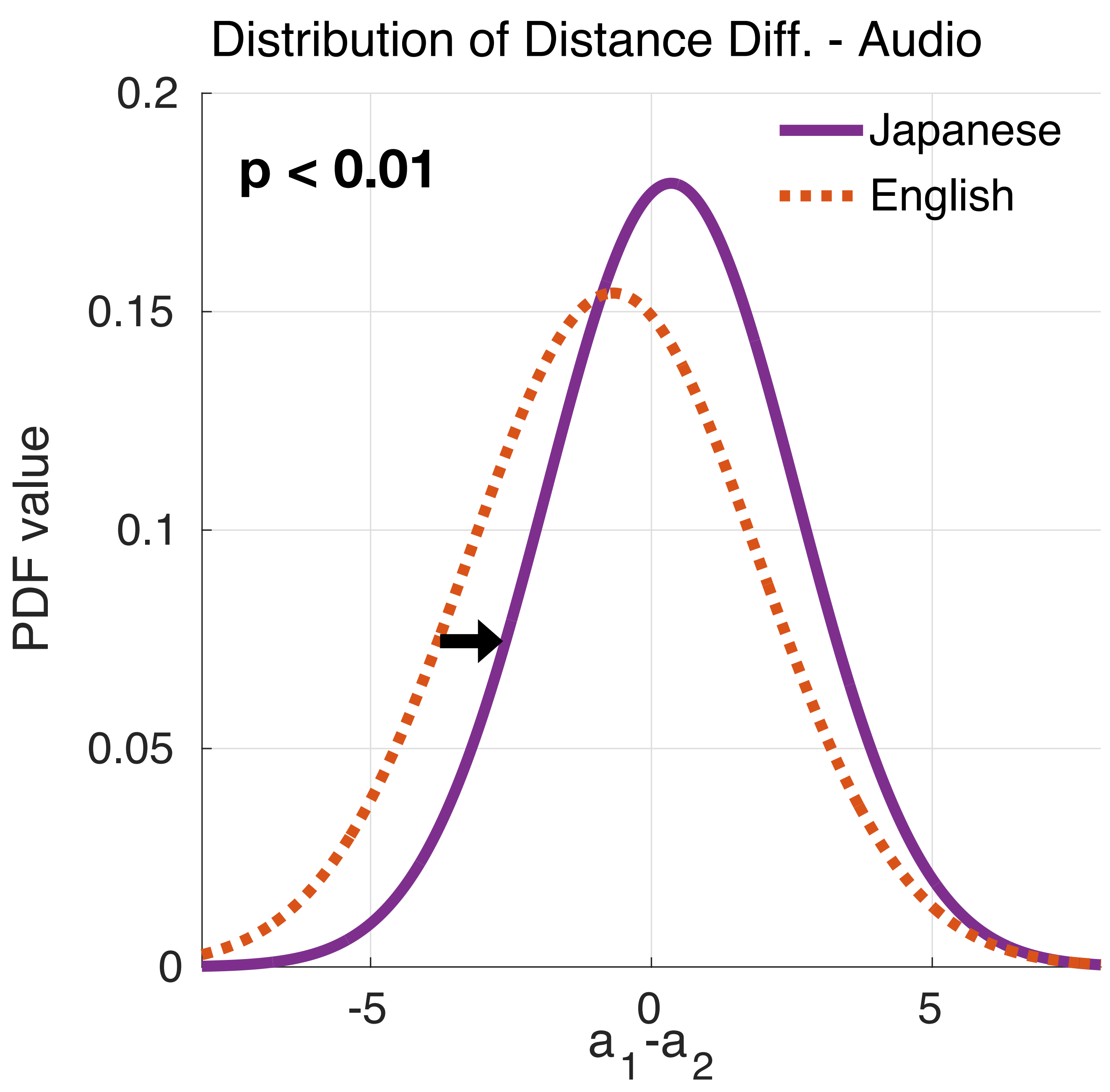


- EEG signals:
  - d_1: Average inter-trial distance of first 10 trials.
  - d_2: Average inter-trial distance of last 10 trials.
- Spoken Audio signals:
  - Dynamic time warping (DTW) based distance computation
  - a_1 & a_2

Evidence for Language Learning

2. Are the neural encodings converging to a consistent pattern for the novel words?



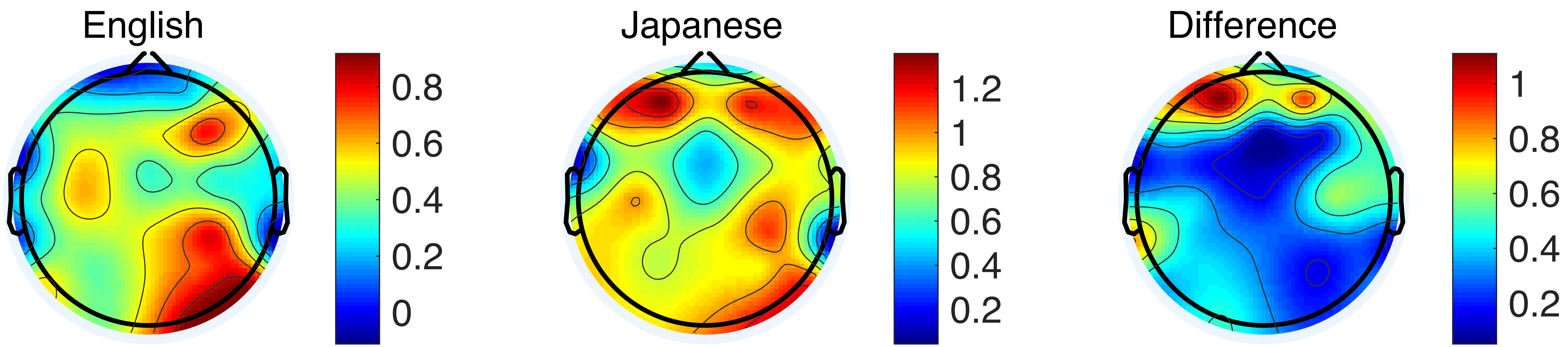


For Japanese language trials

- $d_1 - d_2 > 0$ and
- $a_1 - a_2 > 0$

✓ Listen & repeat > consistent auditory representation formation in the human brain.

Evidence for Language Learning



Scalp plots indicating the channels with higher $d_1 - d_2$ difference for English, Japanese and the difference of the two languages

✓ Language learning activity in Japanese trials are predominant in the frontal and temporal brain regions[1,2].

1. Soman, A., Madhavan, C. R., Sarkar, K., & Ganapathy, S. An EEG study on the brain representations in language learning. IOP Journal on Biomedical Physics & Engineering Express, 5(2), 25041 (2019).
2. Pallier, C. et al. Brain imaging of language plasticity in adopted adults: can a second language replace the first? Cereb. Cortex 13, 155–161 (2003).

LEAP

Summary

- Language discriminative signatures are encoded in the time-frequency representation of the EEG signals, in both magnitude and phase.
- A consistent neural representation is formed when exposed repeatedly to words from an unfamiliar language.

1. **Soman, A.**, Madhavan, C. R., Sarkar, K., & Ganapathy, S. An EEG study on the brain representations in language learning. IOP Journal on Biomedical Physics & Engineering Express, 5(2), 25041 (2019).

Learning with Semantics

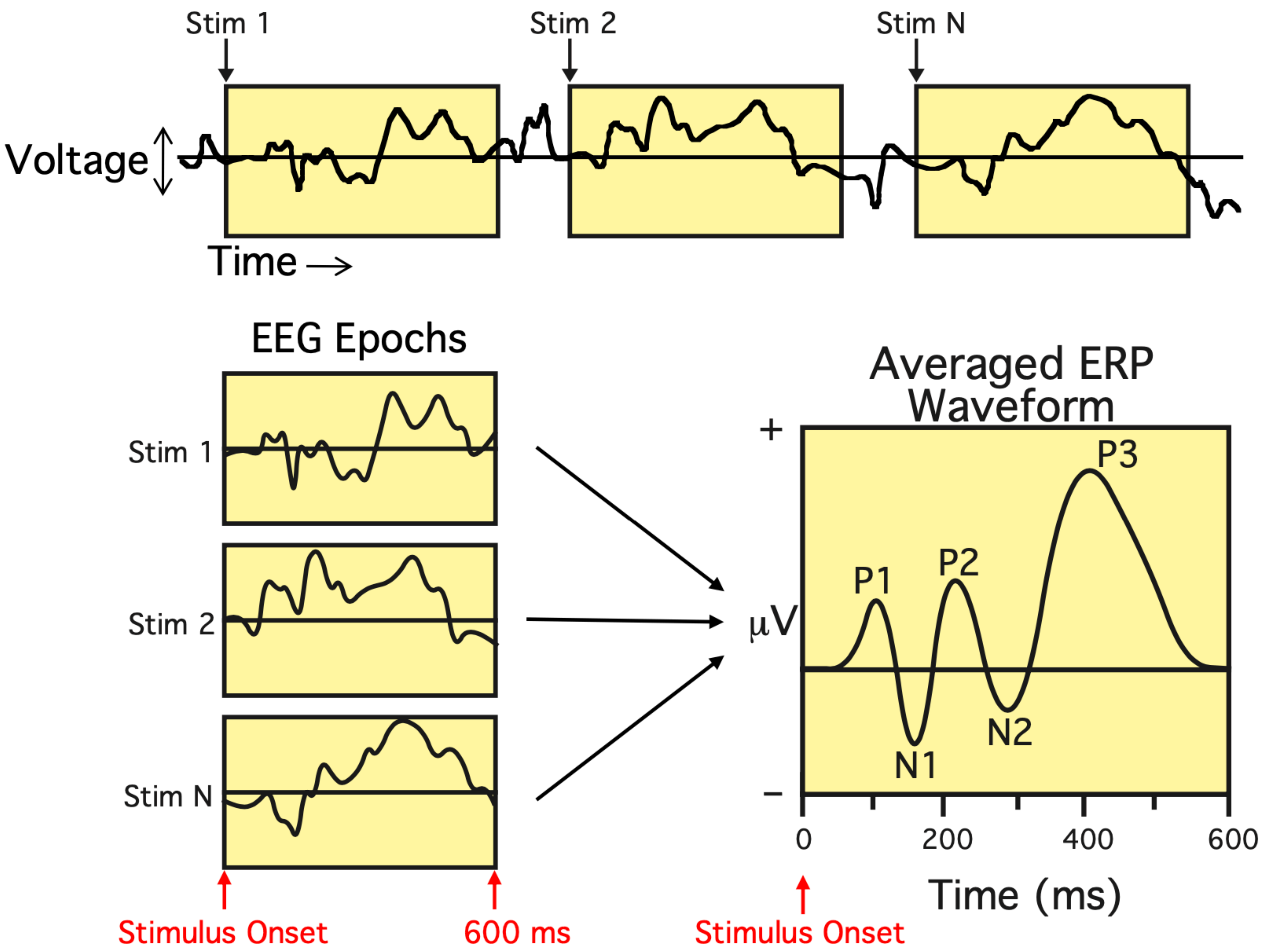
Word Learning ⇔ Speech Perception

- Neural Encoding of Novel Words [Phonology]
- Rapid Word Learning with context [Semantics]
- Stimulus – Response Modelling using ML models [Continuous Speech]

Motivation

- How does our brain adapt when we hear words from a foreign language?
- Does short-term learning incur different effect than long-term learning?
- Does similarities with a known language help in learning (transfer-learning)?

Event Related Potential (ERP)

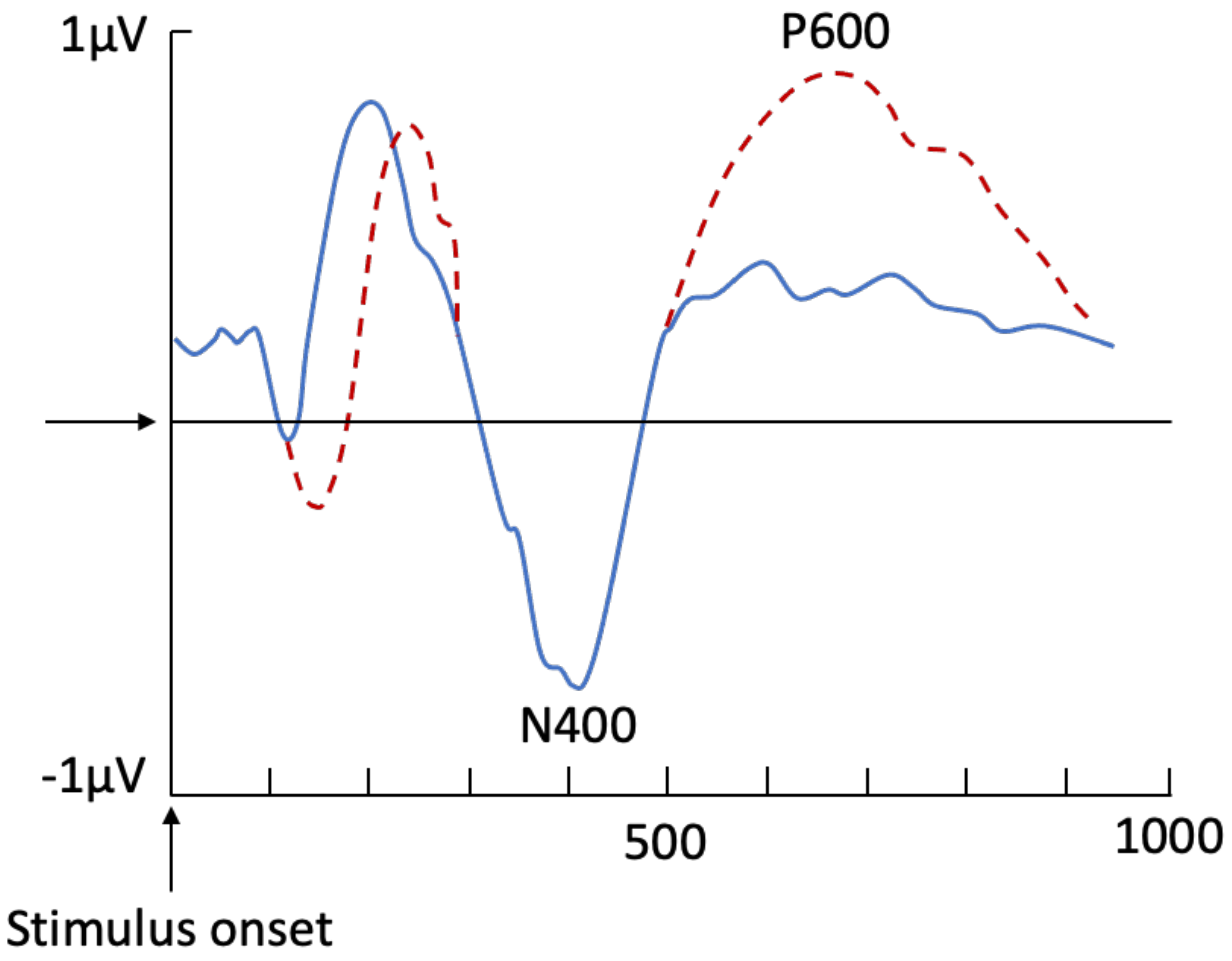


- ERP: Electrical potentials (voltages) that are related to specific events
- Average across the epochs of that event
- Random noise averages out.

https://erpinfo.org/intro-to-erps-course-materials

Using ERP for Language Research

- Allows us to investigate how language processing unfolds in time.
- Violations Paradigms
  - Expectations set up, then violated.
  - **Semantic anomaly:** I like my coffee with cream and… [sugar/socks] : N400
  - **Grammatical Anomaly:** P600



LEAP

1. Kutas, M., & Federmeier, K. D., Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology, 2011.*

Experimental Design

Japanese words- **without knowing** the meaning

The beer drinkers raised their **mugs.**
The beer drinkers raised their **hen.**
The beer drinkers raised their マッグ**maggu.**

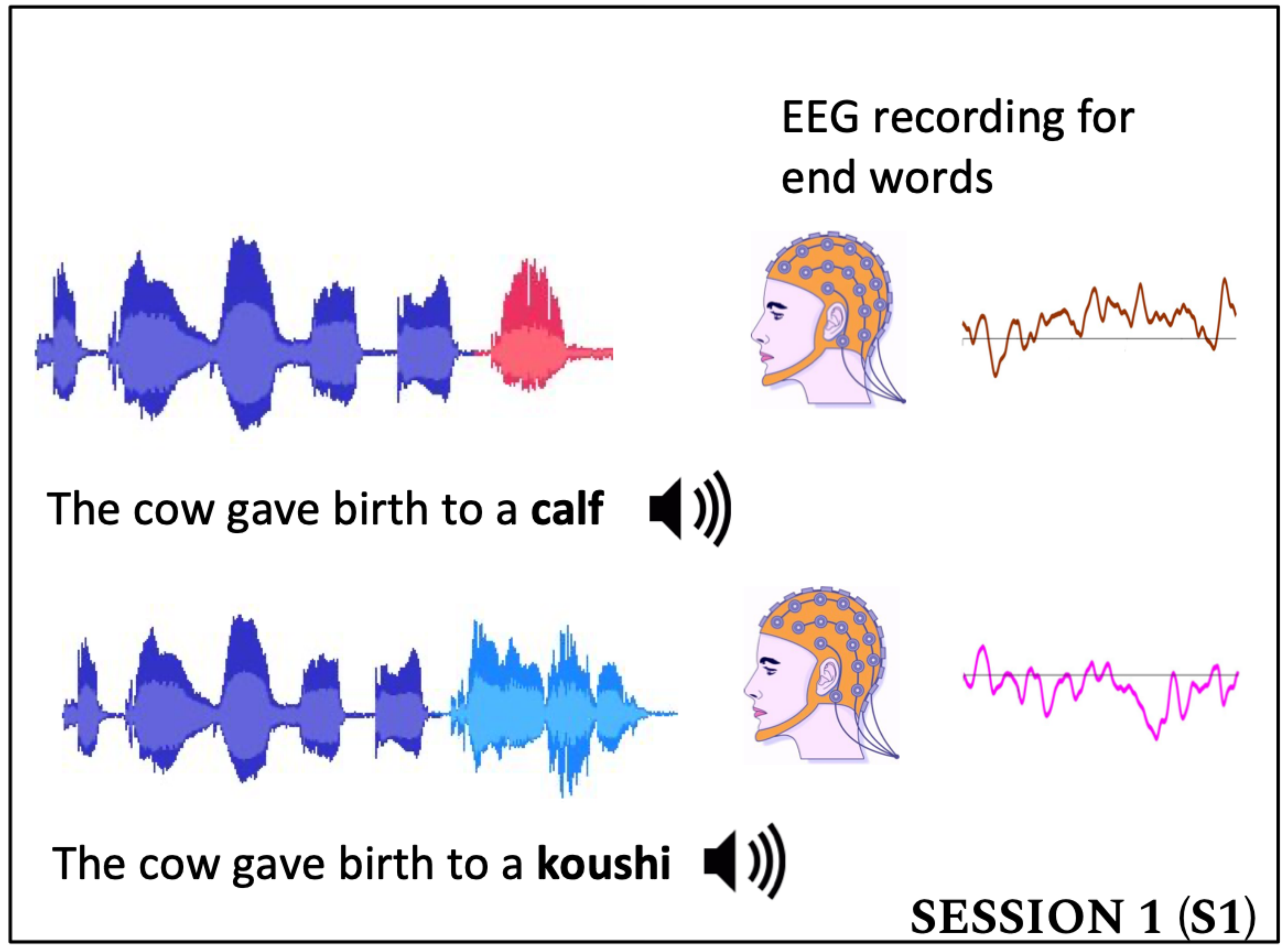
Learn meaning of **Japanese** words. (Learning & Recall)

Japanese words - **after learning** the meaning

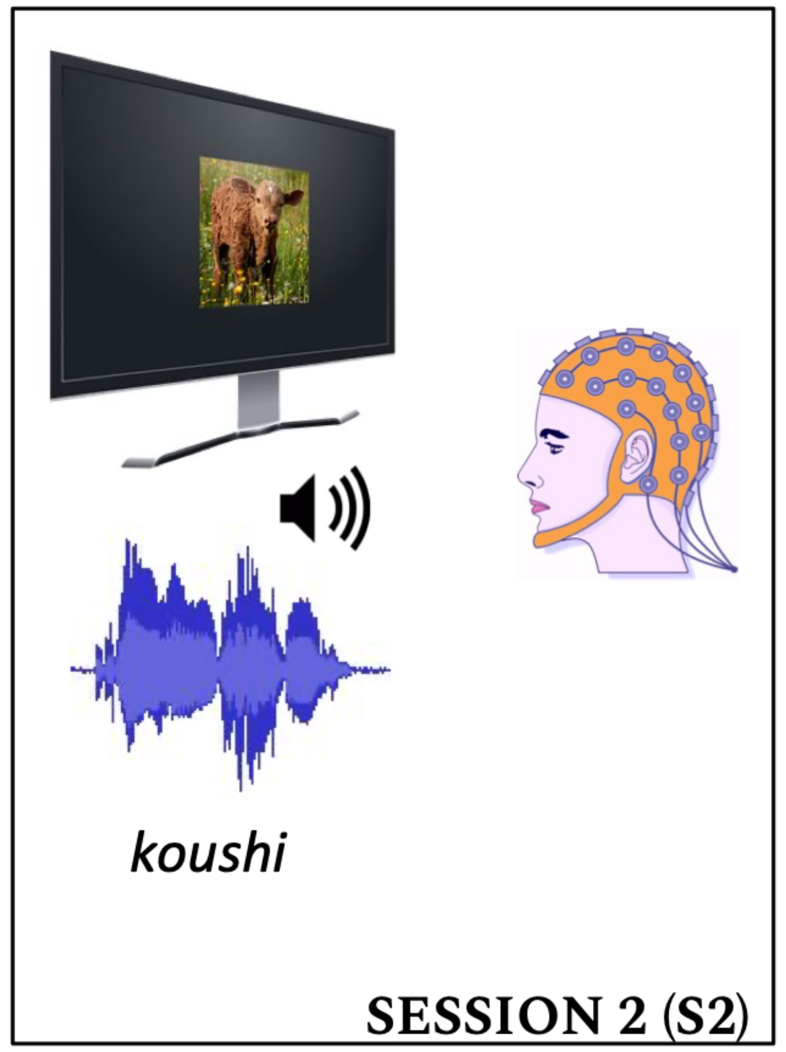
The beer drinkers raised their マッグ**maggu.**

The beer drinkers raised their もっこり **mokkori.**

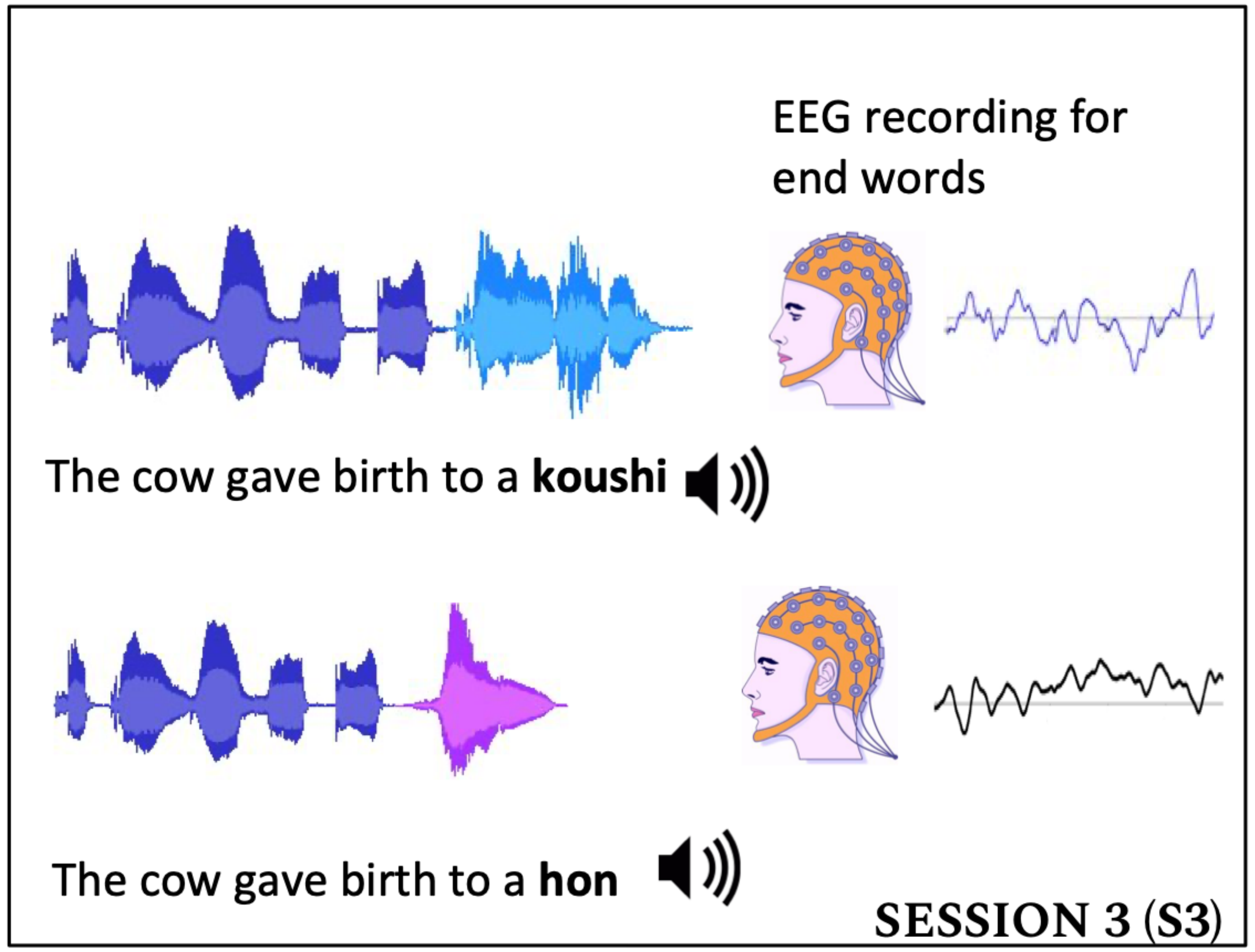
Before semantic learning



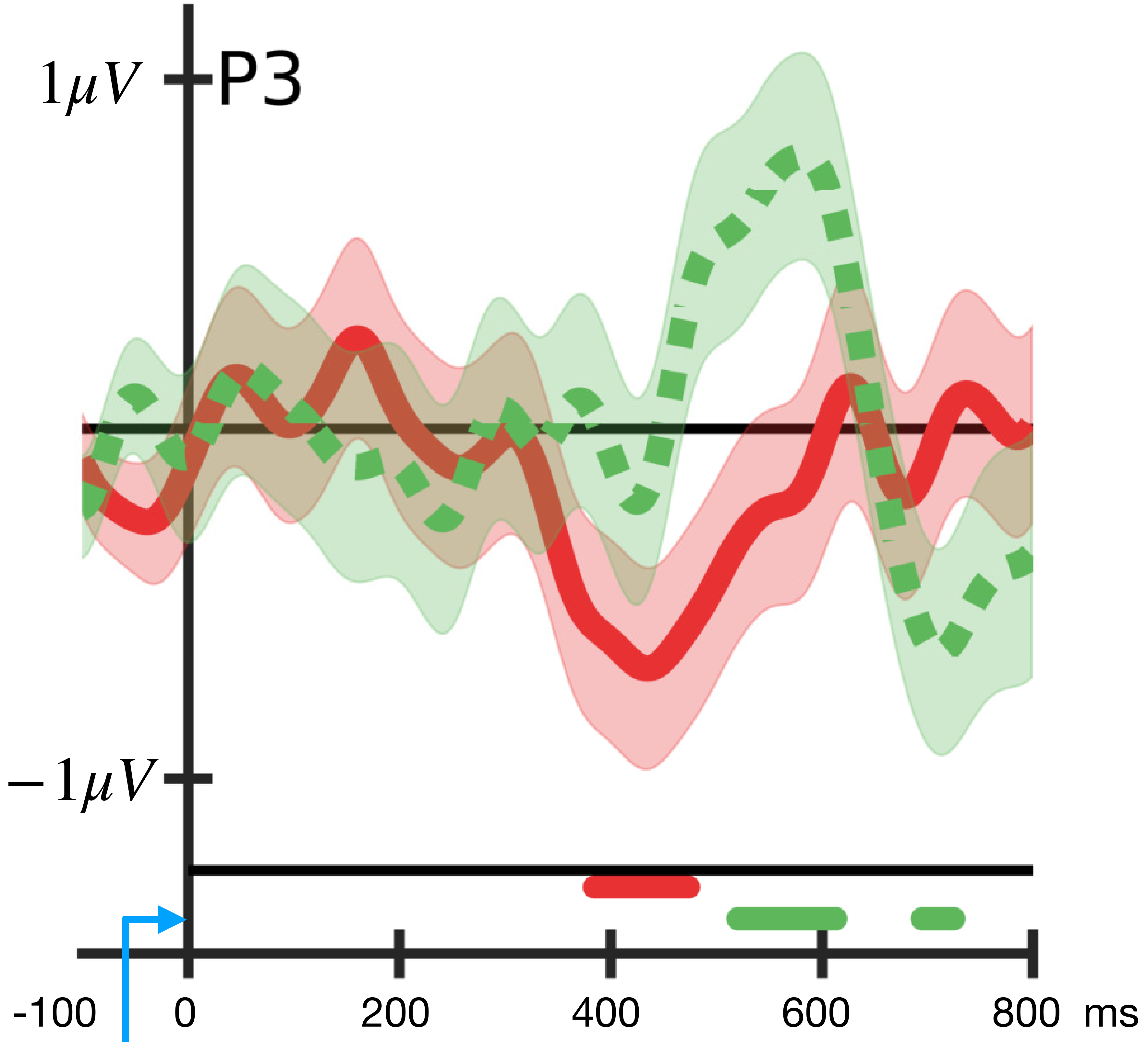
Semantic learning



After semantic learning



Short-term vs Long-term learning



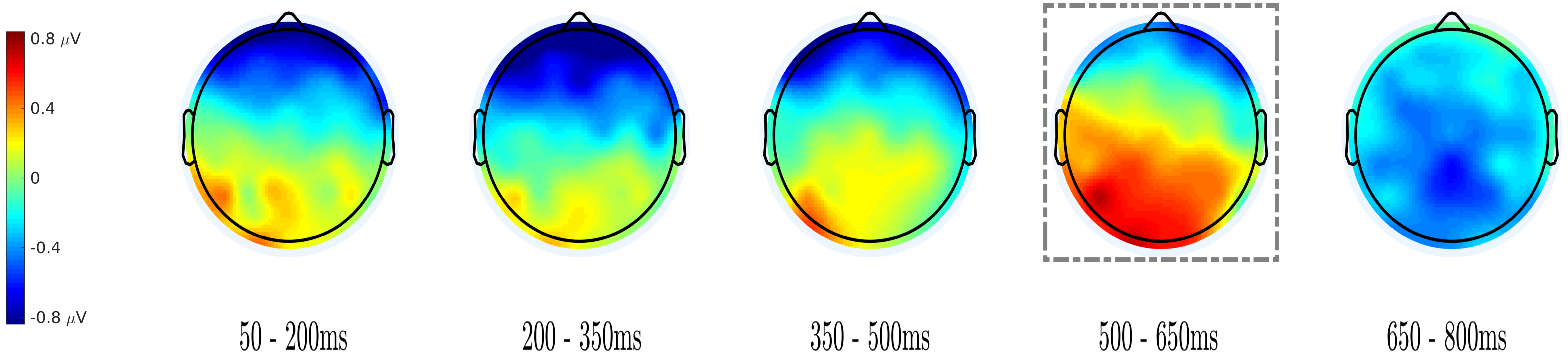
Significant regions
two-sample t-test with $p < 0.05$

Difference ERP Waveform

(English Incongruent - Congruent)

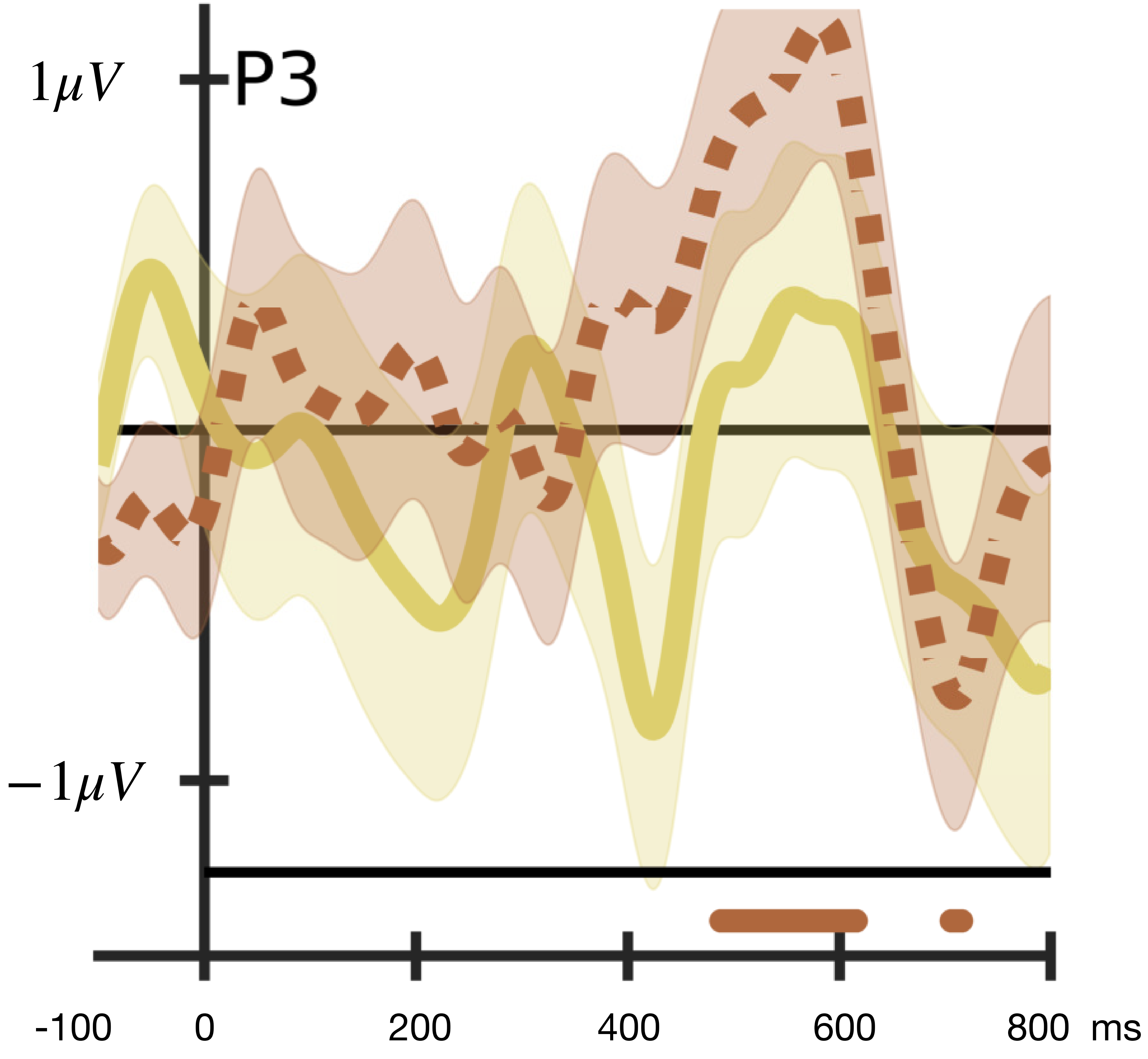
(Japanese Incongruent - Congruent)

Topographic distribution (Japanese)



Effect of similarity

Katakana vs Hiragana words



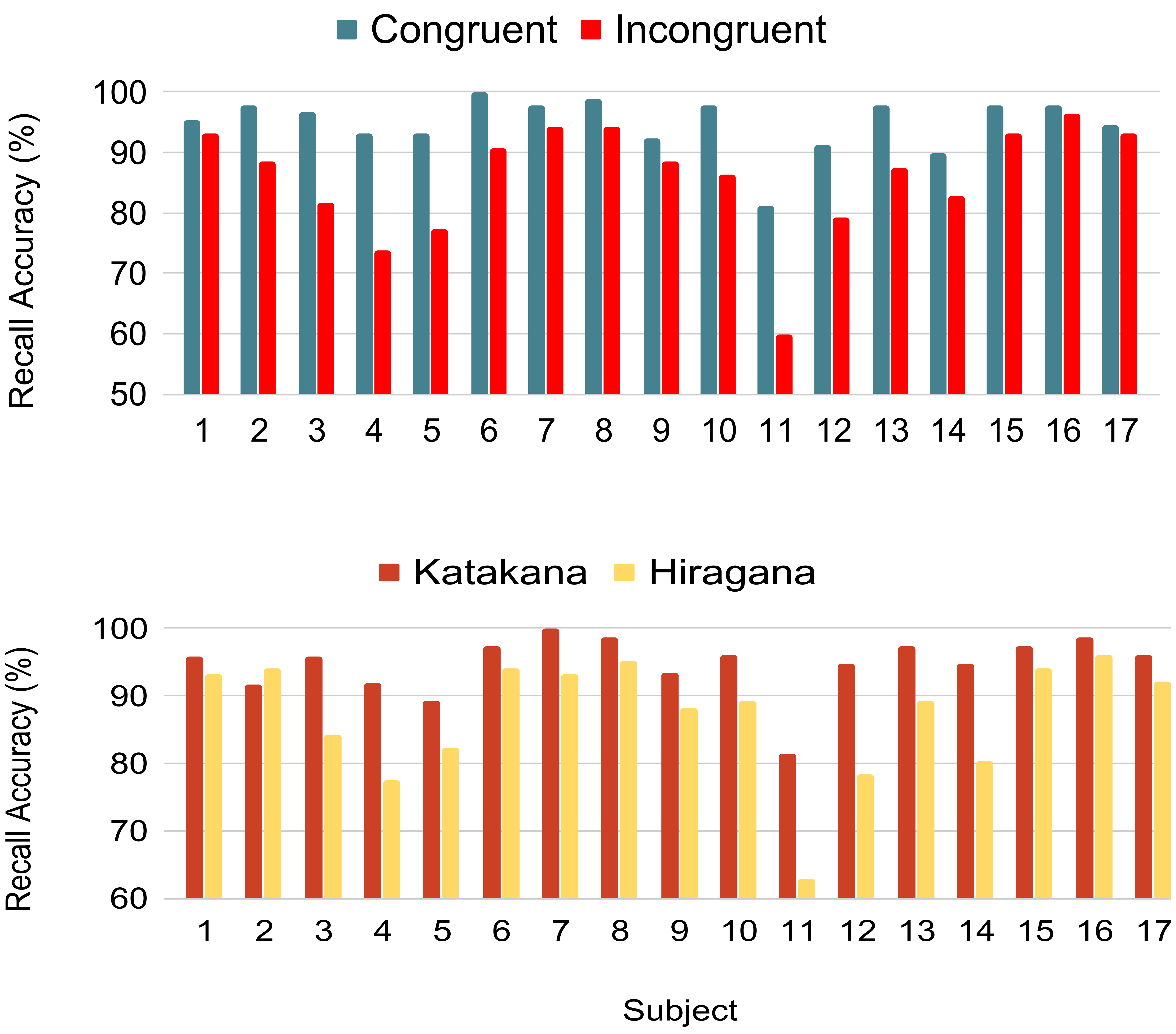
Examples:

- Katakana words: Torappu (trap), Nesuto (nest)
- Hiragana words: Sakana (fish), Hanabana (flowers)

Katakana Incongruent - Congruent

Hiragana Incongruent - Congruent

Behavioral Performance - Recall Accuracy



Summary

- Short-term learning evokes responses different from N400.
- A short-term learning task of new language words evokes a P600 component.
- Phonological similarities with known language words will aid semantic learning.

1. **A. Soman**, P. Ramachandran, and S. Ganapathy, " ERP Evidences of Rapid Semantic Learning In Foreign Language Word Comprehension," Frontiers in Neuroscience, (2022).

Continuous Speech

Word Learning ⇔ Speech Perception

- Neural Encoding of Novel Words [Phonology]
- Rapid Word Learning with context [Semantics]
- Stimulus – Response Modelling using ML models
  - Natural Speech

LEAP

INDIAN INSTITUTE OF SCIENCE

भारतीय विज्ञान संस्थान

Motivation

- Correlating continuous speech to EEG: Stimulus-response modeling by linear models [1,2]
- Deep learning models for speech-EEG decoding [3,4]

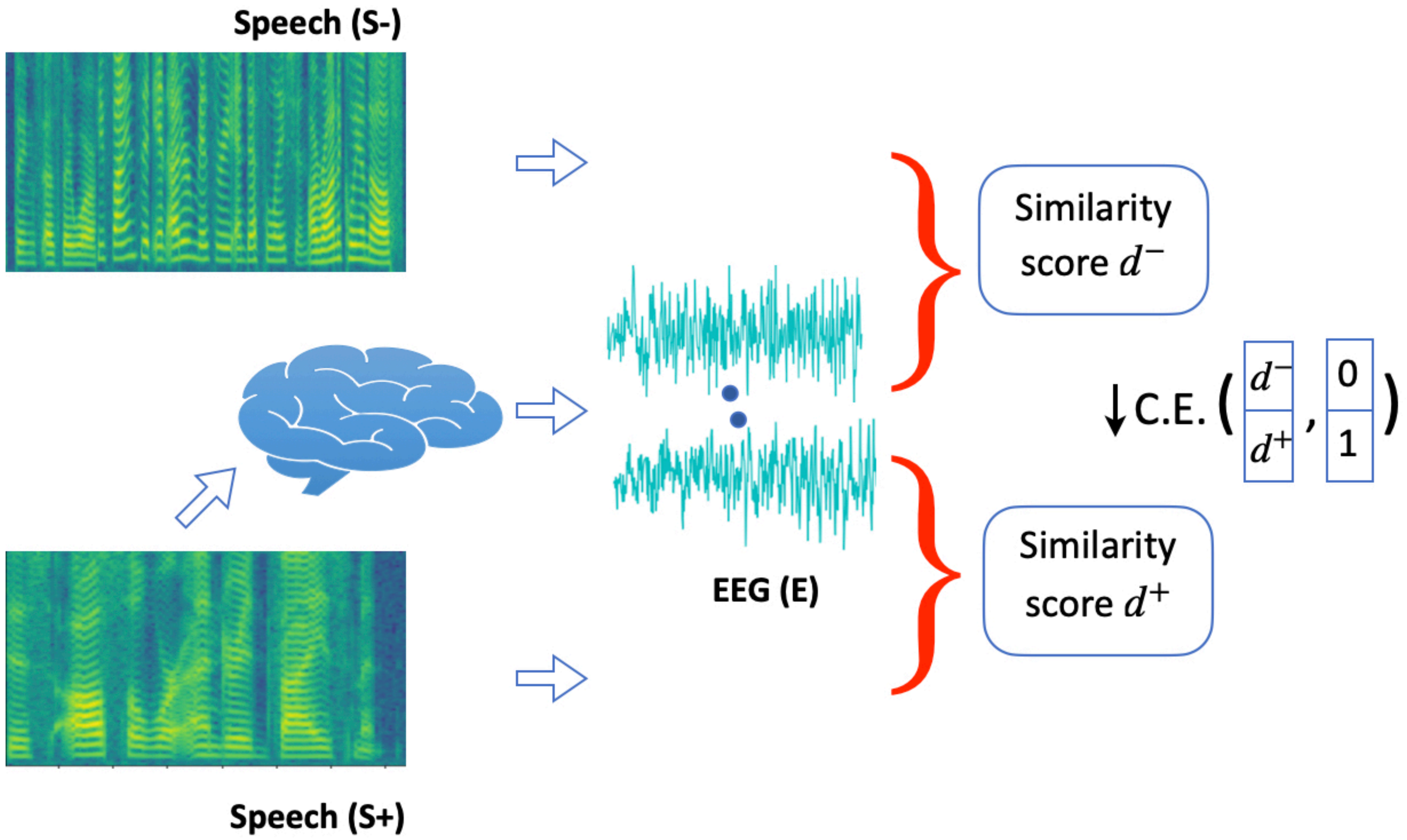
[1]. N.Ding and J.Z.Simon,“Neural coding of continuous speech in auditory cortex during monaural and dichotic listening,” Journal of Neurophysiology, 2012.
[2]. De Cheveigné, Alain, et al. "Auditory stimulus-response modeling with a match-mismatch task." Journal of Neural Engineering (2021).
[3]. Monesi, Mohammad Jalilpour, et al. "An LSTM based architecture to relate speech stimulus to EEG." ICASSP 2020.
[4]. J. R. Katthi and S. Ganapathy, “Deep correlation analysis for audio-EEG decoding,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2021.

Data

- Publicly available speech-EEG data set[1]
- Speech stimulus - Professional audio-book narration of “The old man and the Sea”.
- 19 subjects
- The data consists of 20 trials of roughly the same length
  - Each trial ≈ 180s of audio.
- Overall, speech-EEG data ≈ 19 hours.
- The sentence start and end time, and the word-level segmentation of the speech recordings are provided.
  - Word segmentation: Forced-alignment of speech with text data using Prosodylab-aligner

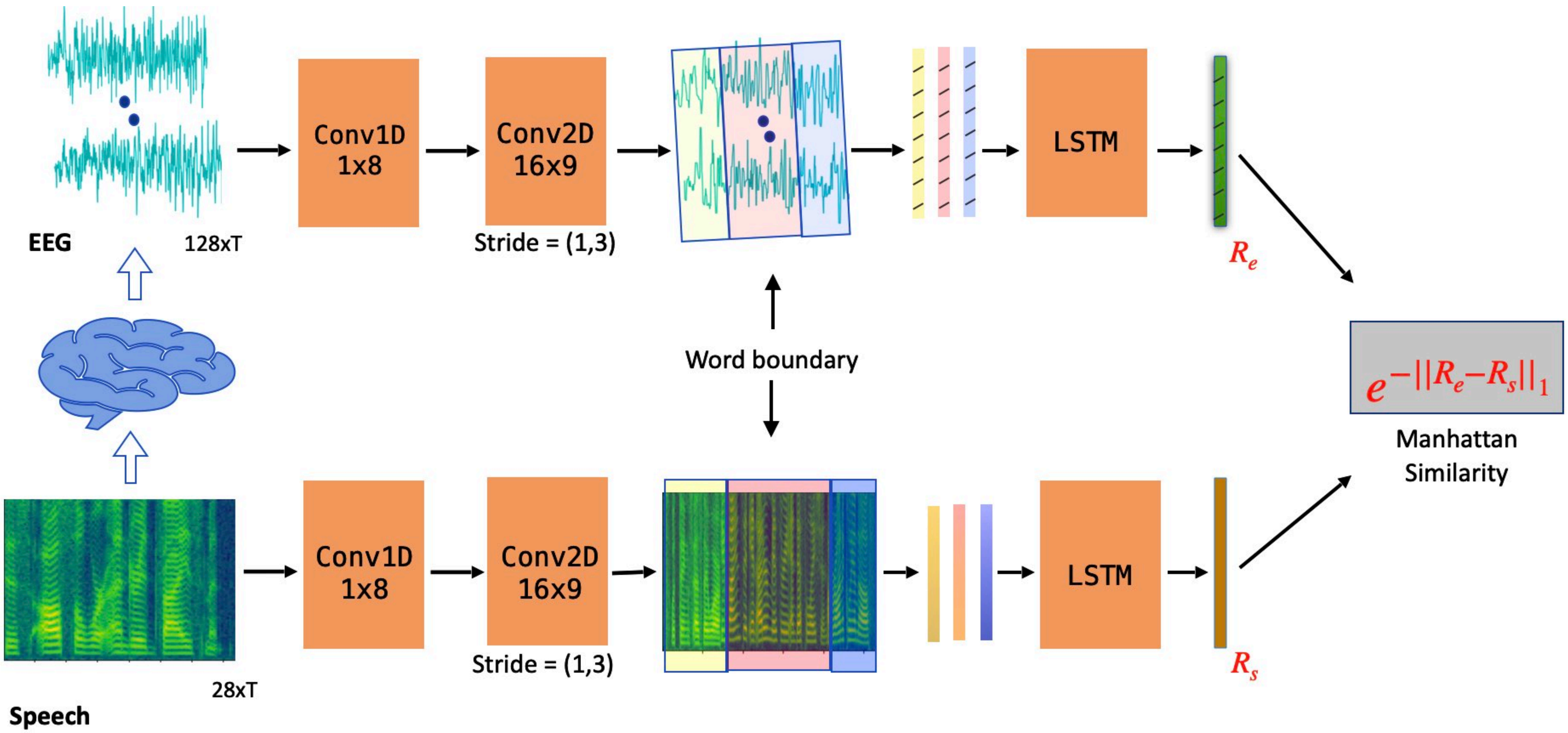
[1]. Broderick, Michael P. et al. (2019), Data from: Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech, Dryad, Dataset, https://doi.org/10.5061/dryad.070jc

Match-Mismatch Classification Task



LEAP

Proposed Model



Match-Mismatch Classification

Fixed-duration segments (Baseline Model[1])

Frame Width (sec.)	Test Accuracy (%)
1	62.21
3	72.41
5	76.12

Sentence-level

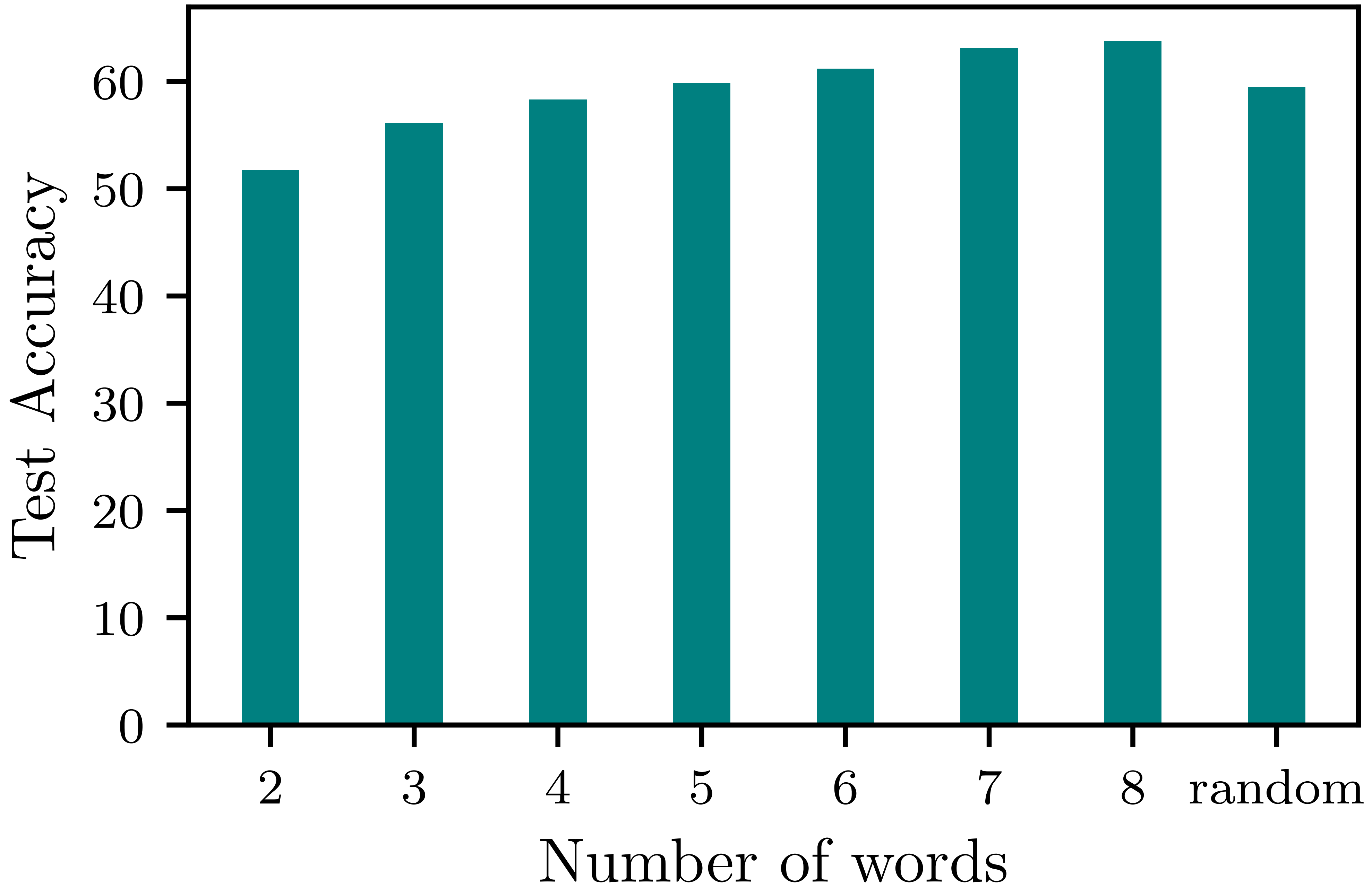
Test Set	Baseline Model	Proposed Model		
		Cos.	Euclidean	Manhattan
Fold 1	65.39	88.22	93.49	94.02
Fold 2	65.32	88.73	93.68	94.00
Fold 3	64.98	86.54	93.72	93.91
Average	65.23	87.83	93.63	**93.97**

LEAP

1. Monesi et.al , "An LSTM based architecture to relate speech stimulus to EEG," *ICASSP*, 2020.

Impact of accurate word boundaries

Random Word Boundaries



Skipping Word Boundaries

Test set	Skip-2	Skip-3	Skip-4	Skip-5
Fold 1	82.45	88.96	90.43	90.64
Fold 2	81.86	88.77	90.32	90.28
Fold 3	82.60	88.79	90.30	90.01
Average	82.30	88.84	90.35	90.31

Summary

- Proposed a match mismatch classification model that can incorporate word boundary information.
- Proposed a loss function based on Manhattan distance for the match mismatch task.
- Experimental illustration of the effectiveness of the model, where the classification performance is significantly improved over the prior works.
- A detailed set of ablation experiments to elicit the impact of word boundary information in speech EEG matching task.

LEAP

Soman, A., Sinha, V., & Ganapathy, S. (2023). Enhancing the EEG Speech Match Mismatch Tasks With Word Boundaries, Interspeech 2023.

Dichotic Listening

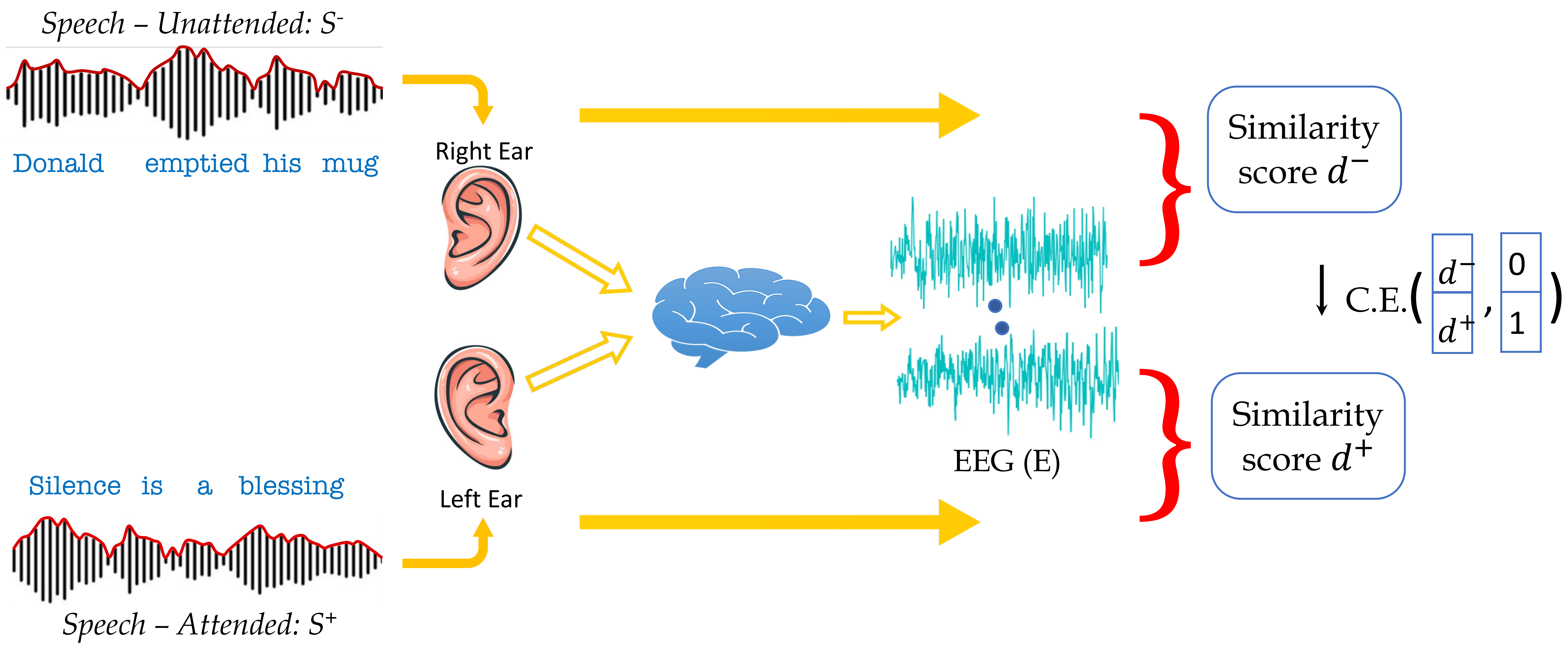
Word Learning ⇔ Speech Perception

- Neural Encoding of Novel Words [Phonology]
- Rapid Word Learning with context [Semantics]
- **Stimulus – Response Modelling using ML models**
  - Natural Speech
  - **Dichotic Listening**

LEAP

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

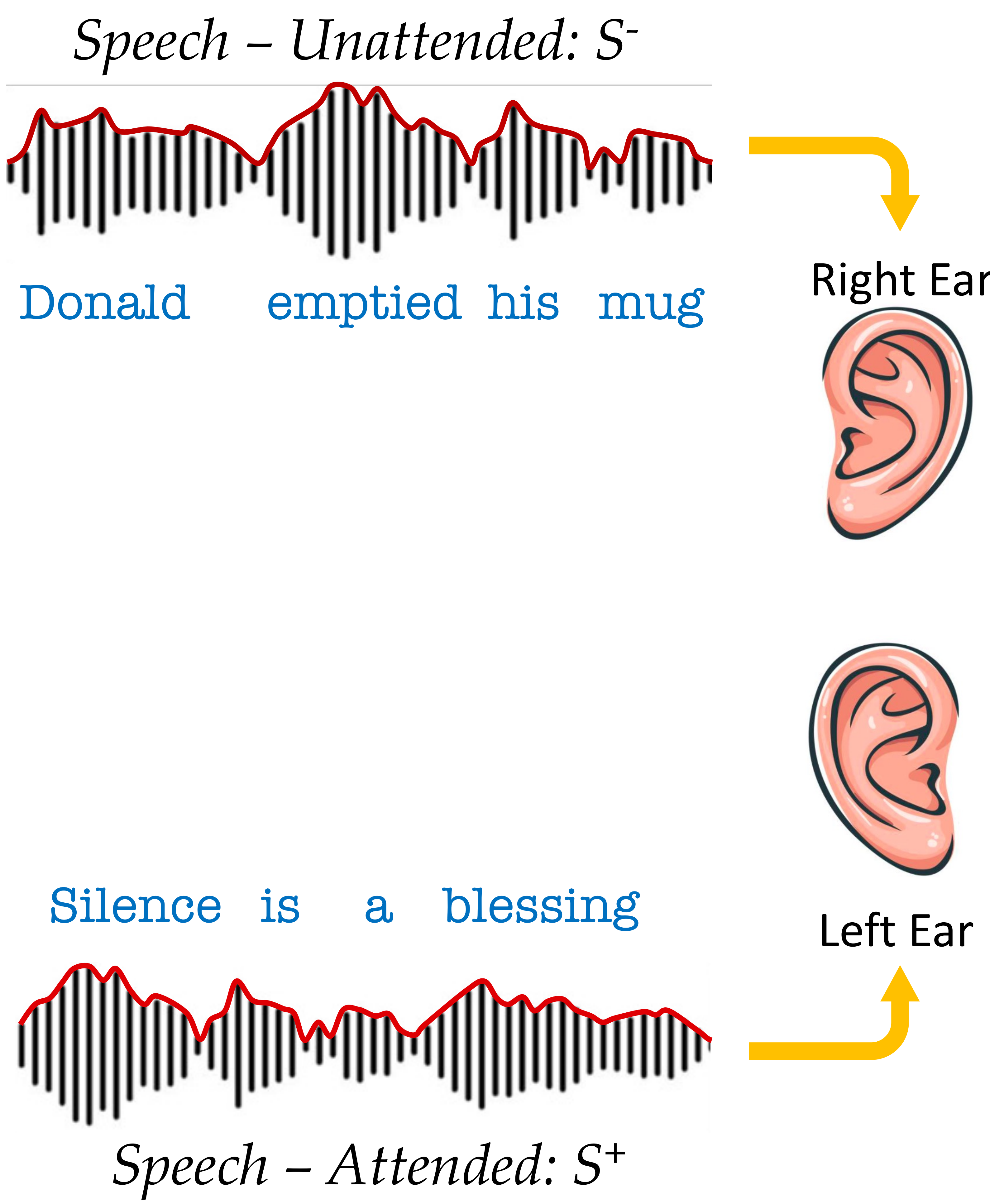
Auditory Attention Decoding as a MM task



- Dichotic listening task: Different speech sounds played to each ear simultaneously.
- Auditory Attention Detection (AAD): Identification of the speech signal to which subject paid attention.

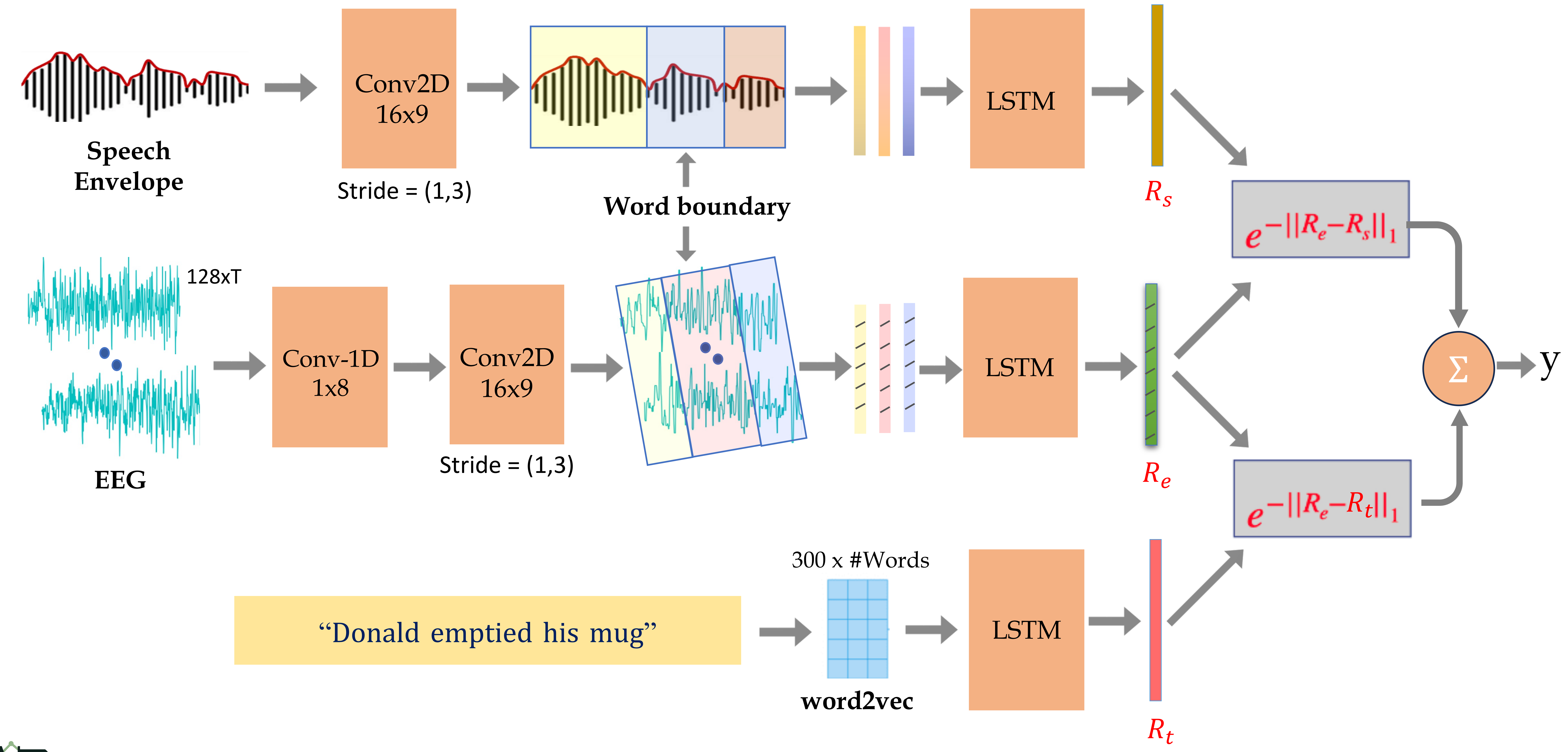
Cocktail Party Dataset

- Publicly available speech-EEG data set[1]
- Speech stimulus - Story 1 to left ear & Story 2 to right ear
- Each subject: underwent 30 trials
  - Each trial ≈ 60s of audio.
- 33 subjects
  - Divided into 2 groups (17 & 16 each).
  - Each group: Instructed to **focus on either one ear** through out all 30 trials



[1]. Broderick, Michael P. et al. (2019), Data from: Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech, Dryad, Dataset, https://doi.org/10.5061/dryad.070jc

Proposed Multi-modality Neural Network



Training and Evaluation Setup

- Subject-dependent training
- Multi-fold cross-validation performed.
  - 3 stimuli files were kept aside to the test set.
  - Natural Speech: 6-fold (only 20 stimuli files)
  - Dichotic: 10-fold (30 stimuli files)

LEAP

Natural vs Dichotic listening

Listening Condition	Speech Envelope	Text word2vec	Multi modality
Natural	93.63	93.24	93.38
Dichotic	62.12	83.06	**84.60**

- Combined training of envelope and word2vec features yields the best result for dichotic listening, 84.6%.
- For NS: acoustic features are slightly better contributor.
- For dichotic: Semantic features performs better than acoustic features in a large margin ($p < 0.001$)

✓ Humans priortize assimilating the context rather than focusing on the acoustic content of speech during difficult listening conditions.

Effect of word boundary input

Stimulus Feature	Without word-boundary	With word-boundary
Envelope	56.90	62.12
word2vec	64.06	83.06
Multi-modal	62.35	**84.60**

- Significant reduction in performance without word boundary information ($p < 0.001$).
- Word level segmentation plays an important role in auditory attention detection.

Summary

- The MM performance of text data is significantly higher than that of the audio signal.
- Human brain **priotizes semantic information than the acoustic** information during dichotic listening.
- Emphasizes the importance of **word boundary information in auditory attention decoding**.
- Proposed a **multi-modal architecture for the MM task**.
- EEG signal jointly encodes the semantic and acoustic content of the stimulus.

Concluding Remarks

Summary

Word Learning ⇔ Speech Perception

Neural Encoding of Novel Words [Phonology]

Rapid Word Learning with context [Semantics]

Stimulus – Response Modelling using ML models [Continuous Speech]

Multi-trial Analysis

Single trial Analysis

Isolated Words

Chapter 2

- Language discriminative features
- Evidence of encoding of novel sounds from repetition

Isolated Sentences

Chapter 3

- P600 as ERP marker of rapid semantic learning
- Phonological similarity aids word learning

Natural Speech

Dichotic Listening

Chapter 4

- Evidence for word level segmentation during speech perception
- Human brain priotizes semantic understanding during complex listening conditions.

Limitations

- Sample size and diversity:
  - To ensure the significance, we have used appropriate statistical tests to evaluate the results.
- EEG as a measure of neural activity : limitations in terms of spatial resolution.
- The study was conducted only with healthy adults.
  - Further studies are required for infants and patient population.

Future Perspectives

- Repetitions of sounds study on patients with language disorders
- Sensitivity of P600 magnitude to different variables like language proficiency
- Investigate generalisability to other languages
- Applying different machine learning architectures for MM classification
- Application of proposed features and ML model to BCI
- Incorporate online word segmentation module to the proposed model
- Improving the model with a larger speech-EEG dataset

Publications from this thesis

Peer-reviewed Journal Papers

1. A. Soman, P. Ramachandran, and S. Ganapathy, “ ERP Evidences of Rapid Semantic Learning In Foreign Language Word Comprehension,” Frontiers in Neuroscience, (2022): p.178.

2. A. Soman, Madhavan C. R., K. Sarkar, and S. Ganapathy, “An EEG Study On The Brain Representations in Language Learning,” IOP Journal on Biomedical Physics and Engineering Express, 5(2), (2019): p.25041.

In preparation

1. A. Soman and S. Ganapathy, “ Impact of Semantic Cues on Speech Perception During a Dichotic Listening Task,” To be submitted to Journal of Neural Engineering.

2. A. Soman, P. Ramachandran, and S. Ganapathy, “ An EEG dataset exploring semantic learning with audio-visual input,” To be submitted to Data in Brief .

Peer-reviewed Conference Papers

1. A. Soman, V. Sinha and S. Ganapathy, “Enhancing the EEG Speech Match Mismatch Tasks With Word Boundaries,” Proc. Interspeech (2023).

2. V. Krishnamohan, A. Soman, A. Gupta and S. Ganapathy, “Audiovisual Correspondence Learning in Humans And Machines,” Proc. Interspeech (2020).

3. K. Praveen, A. Gupta, A. Soman and S. Ganapathy, “Second Language Transfer Learning in Humans and Machines Using Image Supervision,” IEEE ASRU (2019).

Acknowledgements

Funding Agencies:

- MHRD
- Department of Atomic Energy
- Pratiksha Trust
- DST

Technical Discussions:

- Prof. Shihab Shamma, UMD
- Prof. Nima Mesgarani, CU, NY
- Dr. Arun Sasidharan & Dr. Aravind Kumar, NIMHANS

EEG Data Collection:

- Axxonet Technologies
- Institute Human Ethics Committee (IHEC), IISc
- Subjects
- Priya Raghavan, Japanese speaker

THANK YOU!

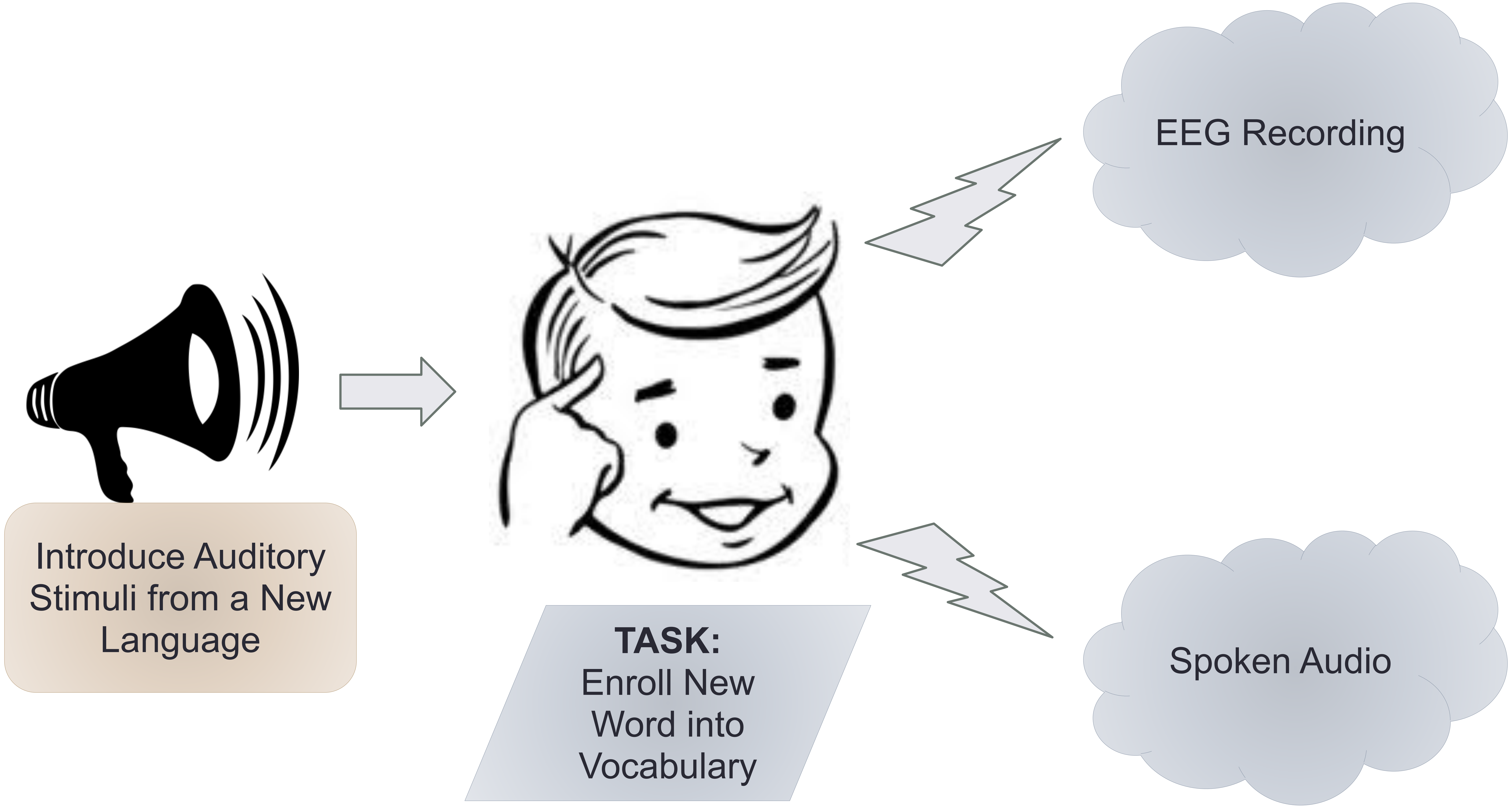
LEAP

Appendix

What makes these problems challenging?

- Not many publicly available datasets
- Cumbersome EEG data collection process
- EEG: highly noisy signal
- Many functionalities of human brain is still a mystery.
- Lot of variability in preprocessing and analysing methods based on the task performed.

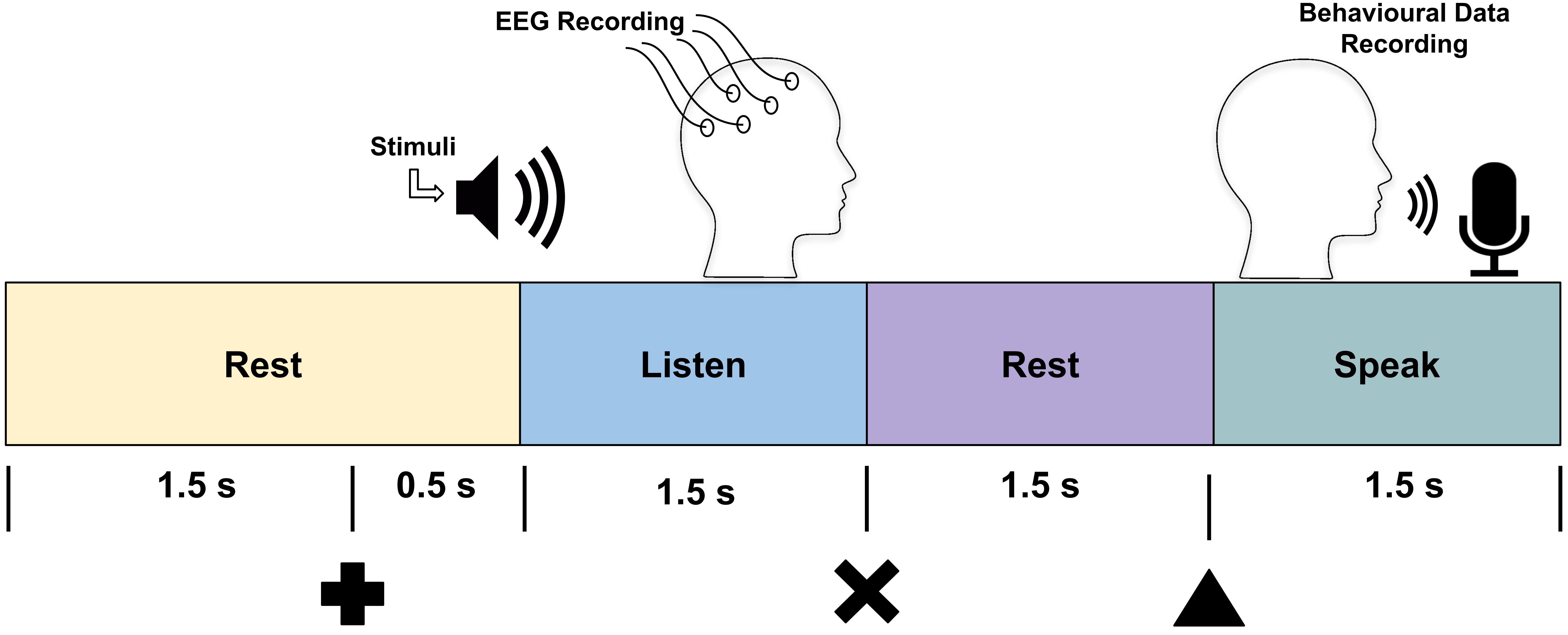
Outline of Work



Objectives

- Explore the differences in human perception while listening to a familiar and an unfamiliar language.
- Probe the language discriminative features encoded in the neural responses.
- Understand the evolution of neural representations in a language learning task.

Experimental Design

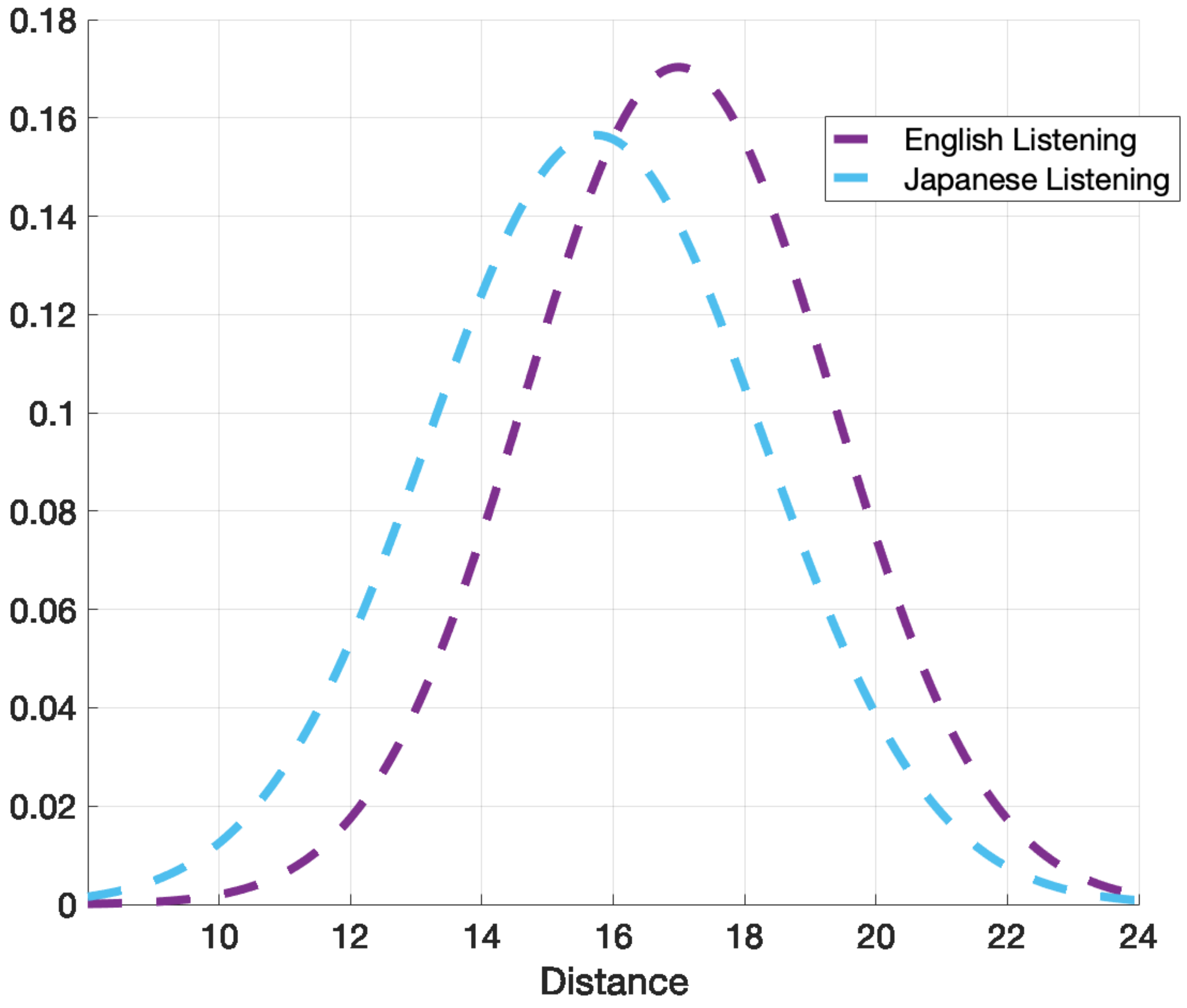


- Visual cues (at the bottom) are provided to indicate the change in state.
- All the EEG analysis performed with signals recorded during listening.

LEAP

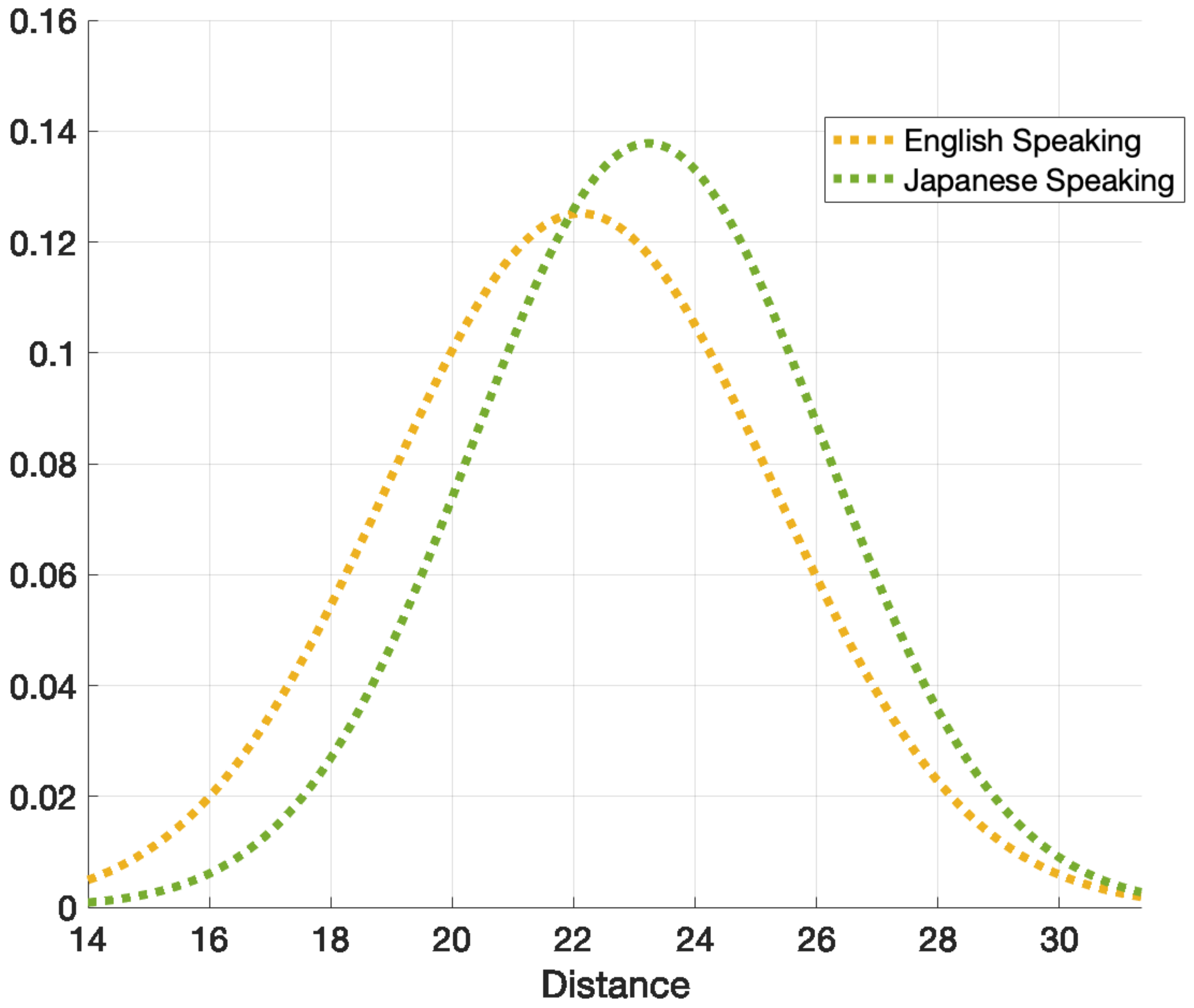
Relationship between Behavioral and Neural Activations

- Distance between envelope of EEG at the listening state and **stimuli** audio envelopes (downsampled to 64Hz).



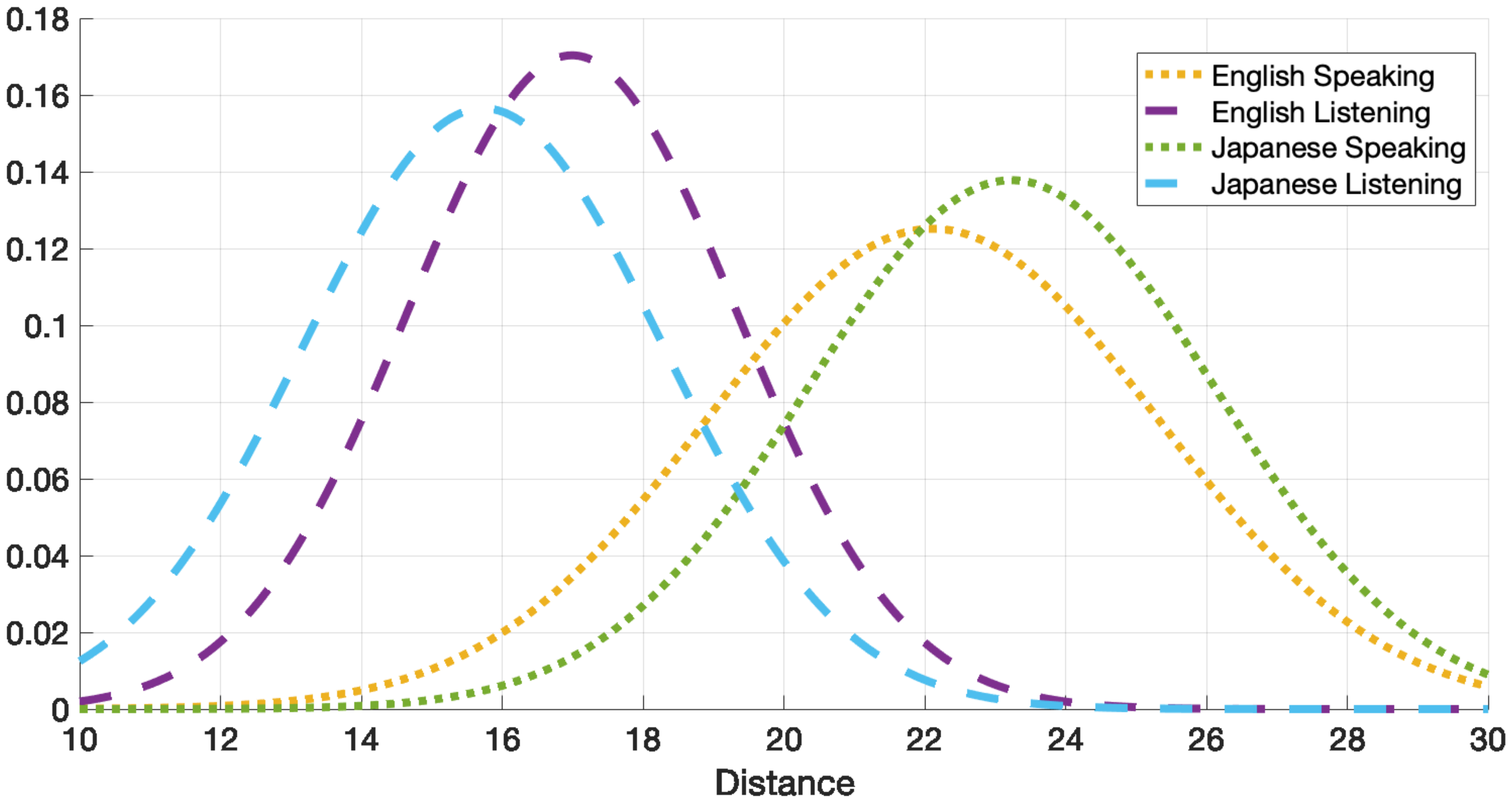
Relationship between Behavioral and Neural Activations

- Distance between envelope of EEG at the listening state and **spoken** audio envelopes (downsampled to 64Hz).



Relationship between Behavioral and Neural Activations

- Listening versus Speaking



1. Soman, A., Madhavan, C. R., Sarkar, K., & Ganapathy, S. An EEG study on the brain representations in language learning. IOP Journal on Biomedical Physics & Engineering Express, 5(2), 25041 (2019).

LEAP

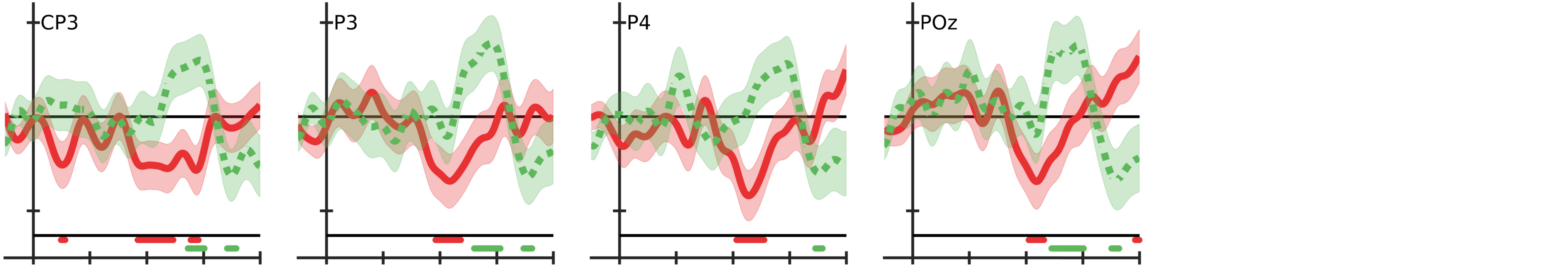
Summary

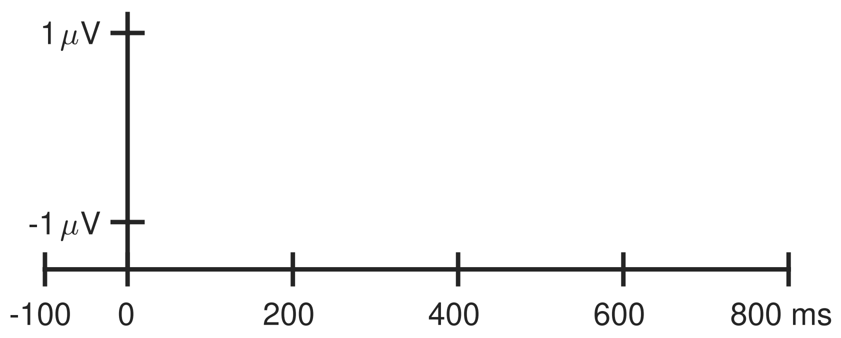
- EEG responses are different for known and unknown languages.
- Broad learning pattern in audio and EEG are correlated.
- Listening audio and EEG are more correlated (lesser distance) for unfamiliar language
  - Limited top-down processing.
- Spoken audio and EEG are less correlated (more distance) for unfamiliar language.
  - Speech production matches less with stimuli provided (and also the listening EEG).

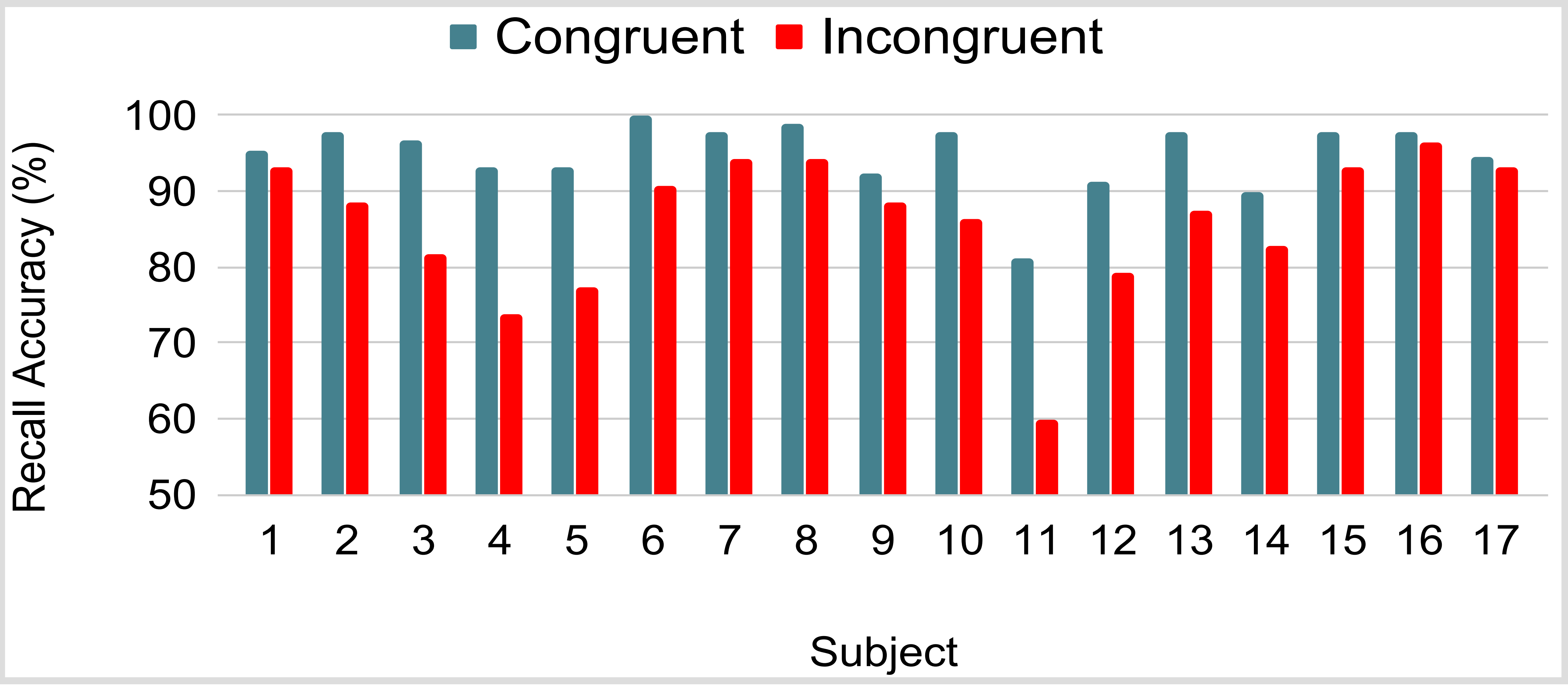
LEAP

Results

Short-term vs Long-term learning

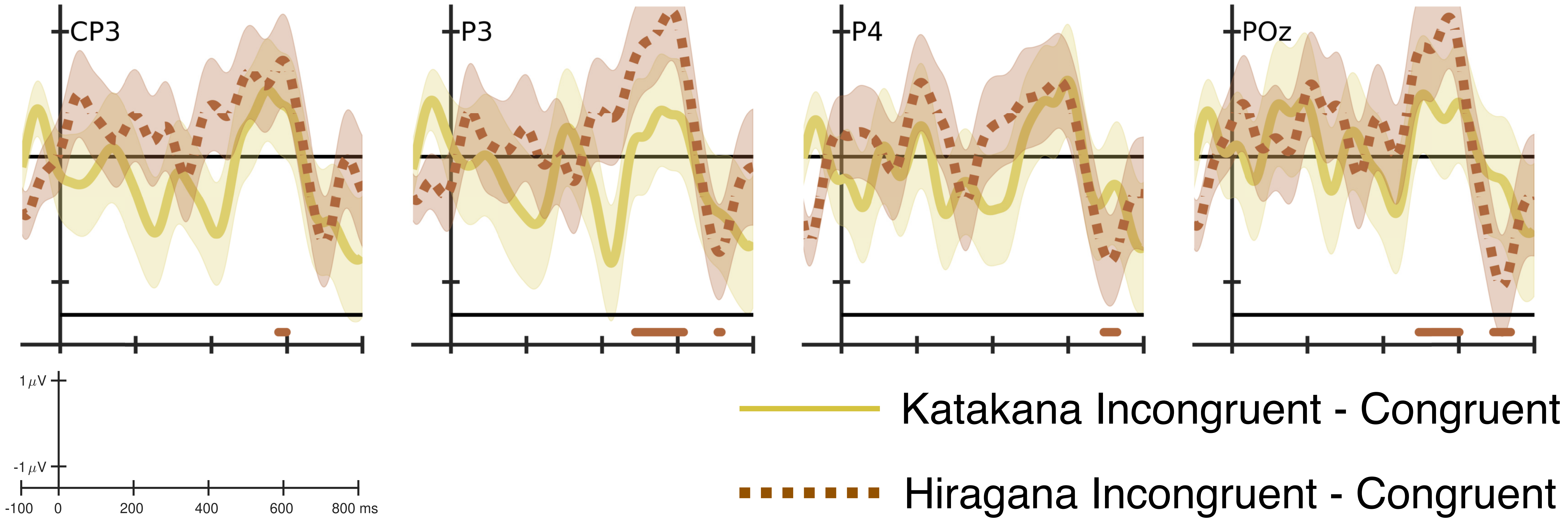


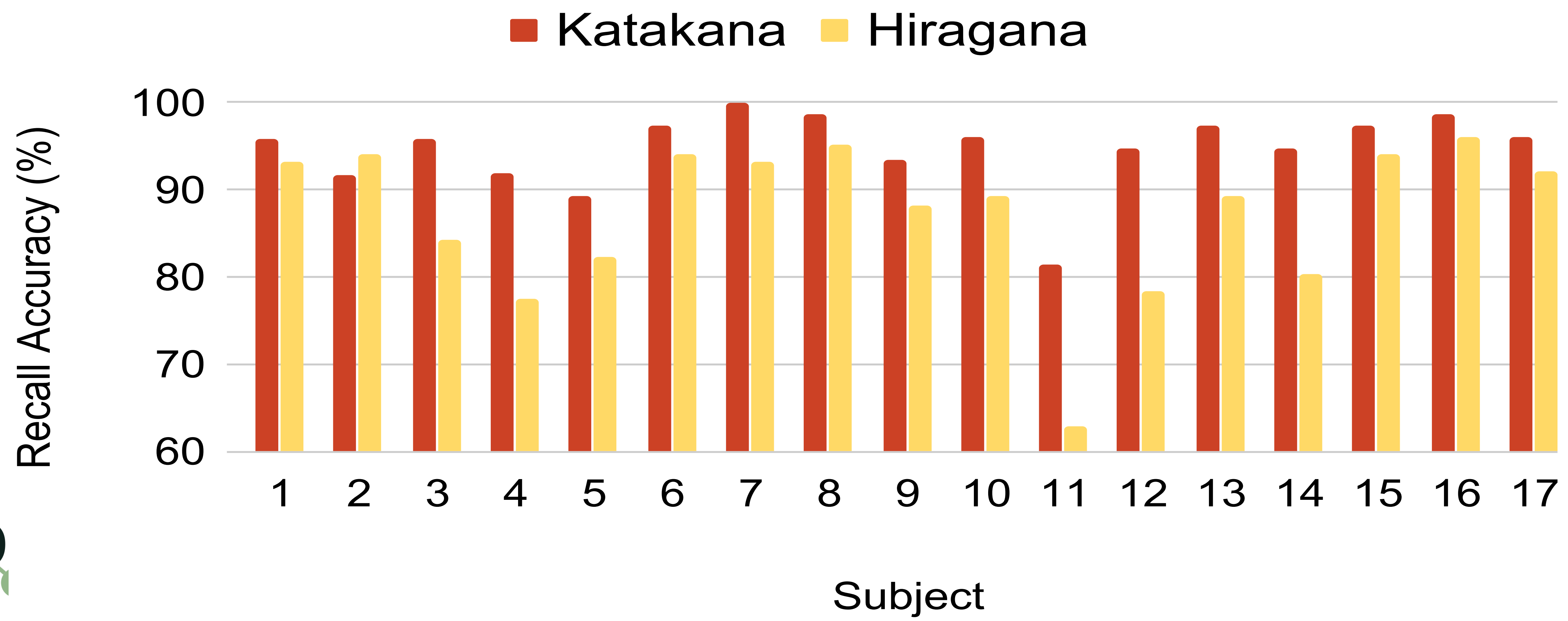




Results

Katakana vs Hiragana words



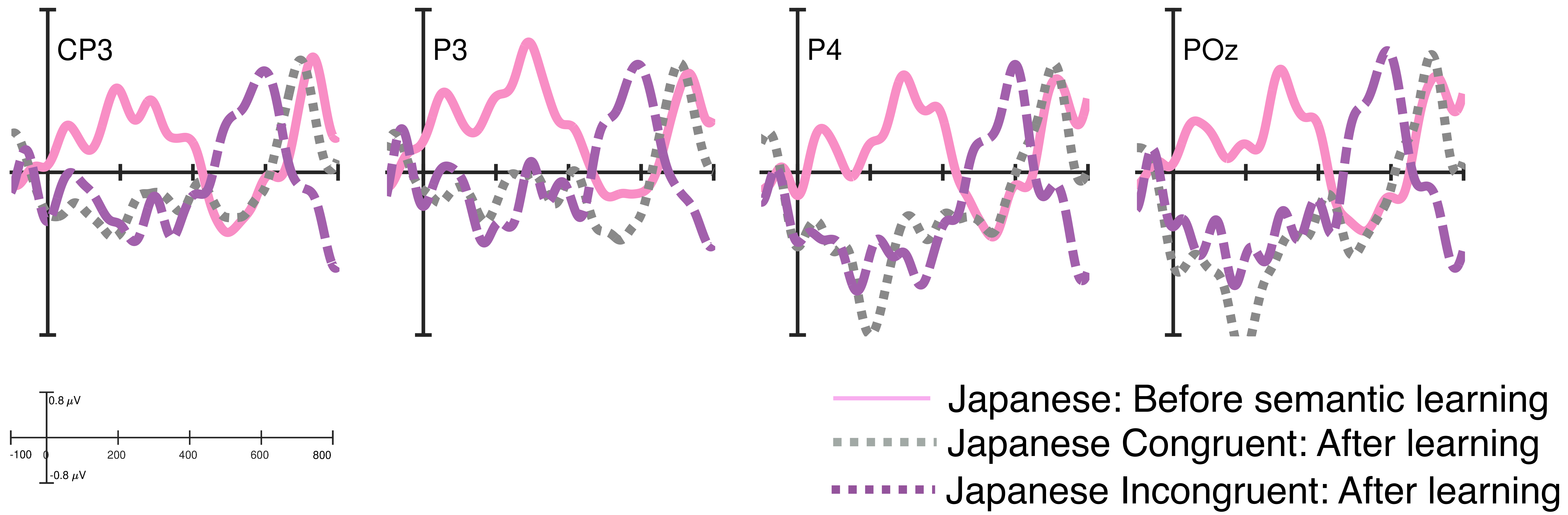


Examples:

- Katakana words: *Torappu (trap)*, *Nesuto (nest)*
- Hiragana words: *Sakana (fish)*, *Hanabana (flowers)*

Effect of Semantic Learning

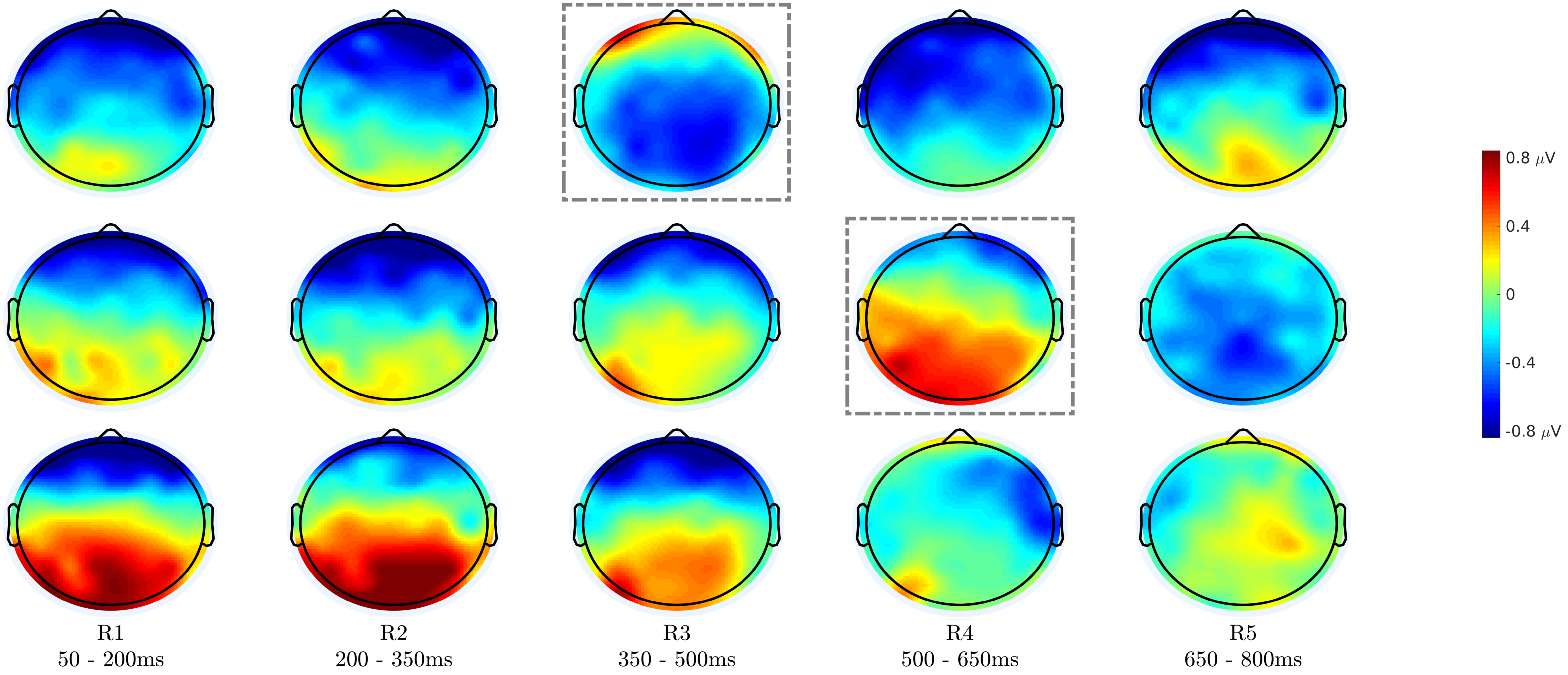
ERP responses to Japanese end words before and after learning



1. A. Soman, P. Ramachandran, and S. Ganapathy, " ERP Evidences of Rapid Semantic Learning In Foreign Language Word Comprehension," Frontiers in Neuroscience, (2022).

LEAP

Topographic distribution



- Top row: Difference of English congruent end word responses from Eng. incongruent end word responses : centro-parietal regions in R3
- Middle row: Difference of Japanese congruent end word responses from Japanese incongruent end word responses strongly positive response over centro-parietal regions (left hemisphere) in R4.
- Bottom row: Difference of Japanese end word responses before learning its meaning from the Japanese end word responses after learning

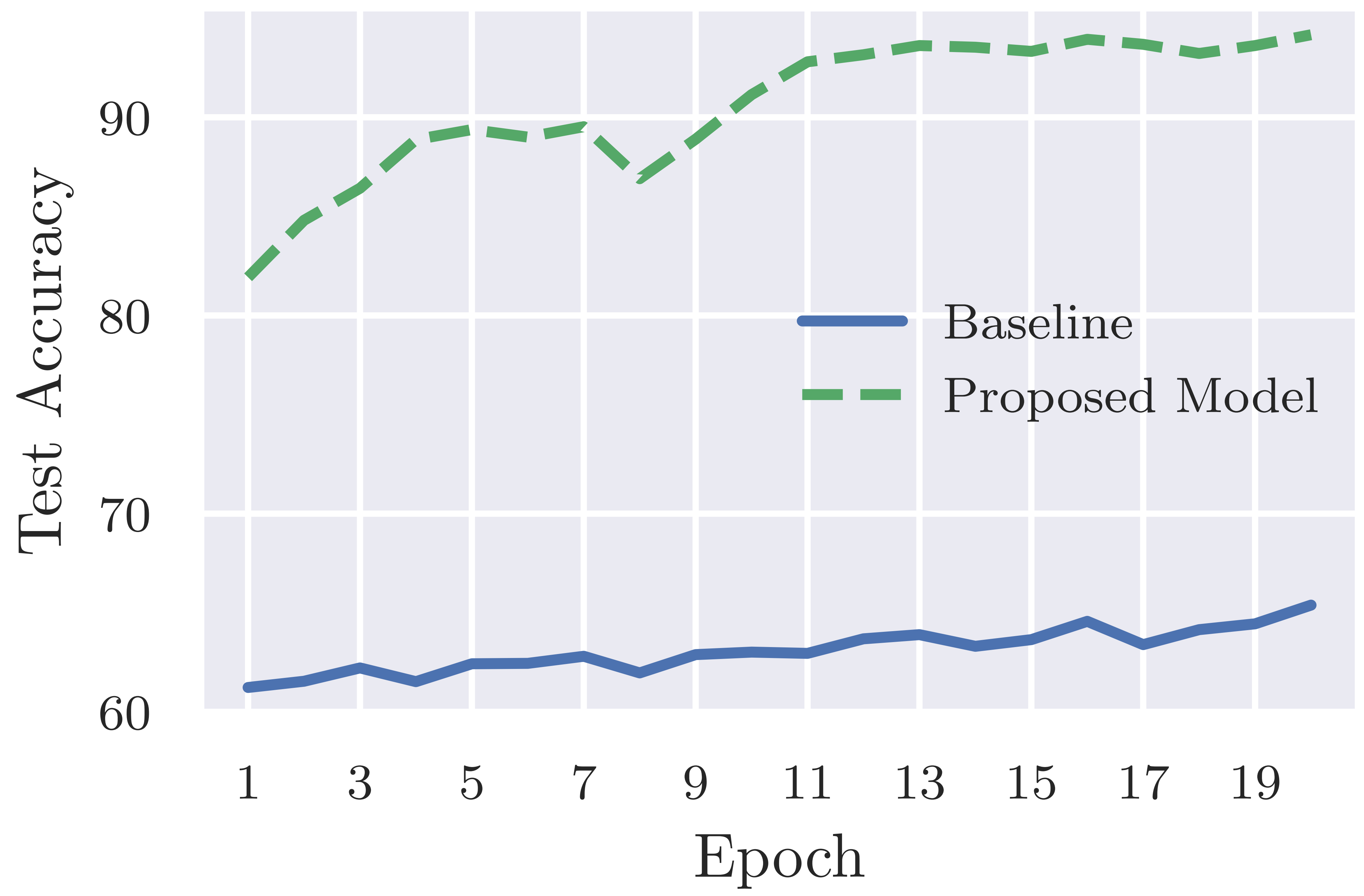
LEAP

Observations

- Variation of N200 amplitude over the parieto-occipital and occipital electrodes.
- Presence of P300 in the first exposure of the Japanese word before semantic learning.
- P600 amplitude is significantly positive for the Japanese incongruent condition.
- LPC component is elicited in the congruent condition.
- In the early time regions (0-400ms) after word onset, both congruent and incongruent conditions after semantic learning has similar differences with the ERP response for word without semantic knowledge.

Training and Evaluation Setup

- Subject-independent evaluation
- 3-fold cross-validation
- Batch-size: 32
- Adam optimizer
- Learning rate: 0.001
- Weight decay parameter: 0.0001
- Binary cross entropy loss


90
80
70
60
Test Accuracy
1
3
5
7
9
11
13
15
17
19
Epoch
Baseline
Proposed Model

Mismatch Sample Selection

Mismatch Selection Strategy	Test Accuracy (%)
Random Sentence	93.97
Next sentence	91.56

LEAP

Subjects' attentiveness and AAD accuracy

Comprehension Score based Trial-Filtering	Speech Envelope	Text word2vec	Multi modality
No Filtering	62.12	83.06	84.60
Both train and test data	66.48	83.86	84.61

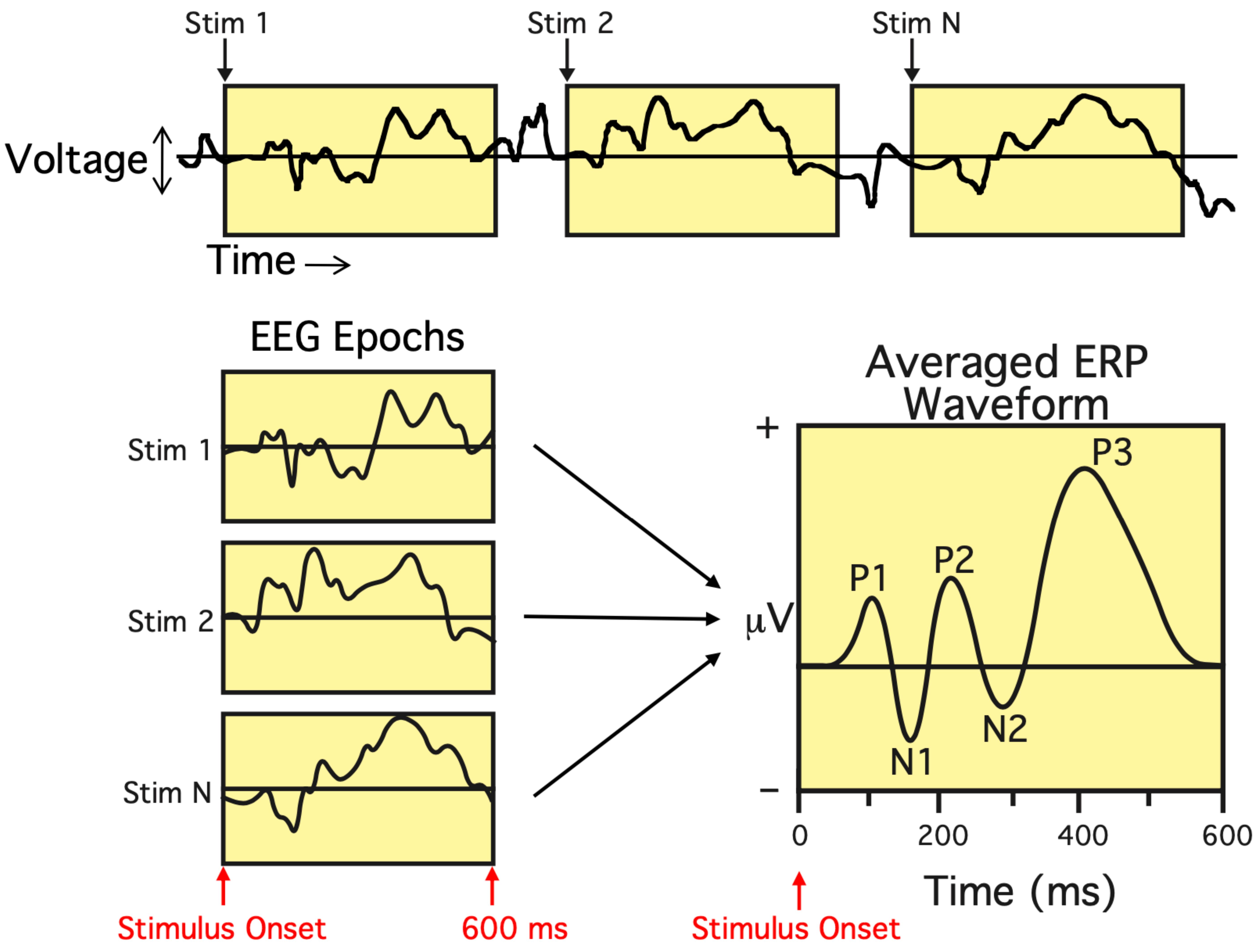
- Subjects answered multiple-choice questions about both attended and unattended stories after each trial —> Comprehension score
- Removed trials with comprehension score < 0.5
- How subjects' attentiveness affect AAD accuracy?

LEAP

Similarity measures

Similarity Measure	Multi modality
exp(-L1)	84.60
exp(-L2)	84.55
sigmoid(cos.)	83.93
-tanh(-L1)	84.03
-tanh(-L2)	**85.62**

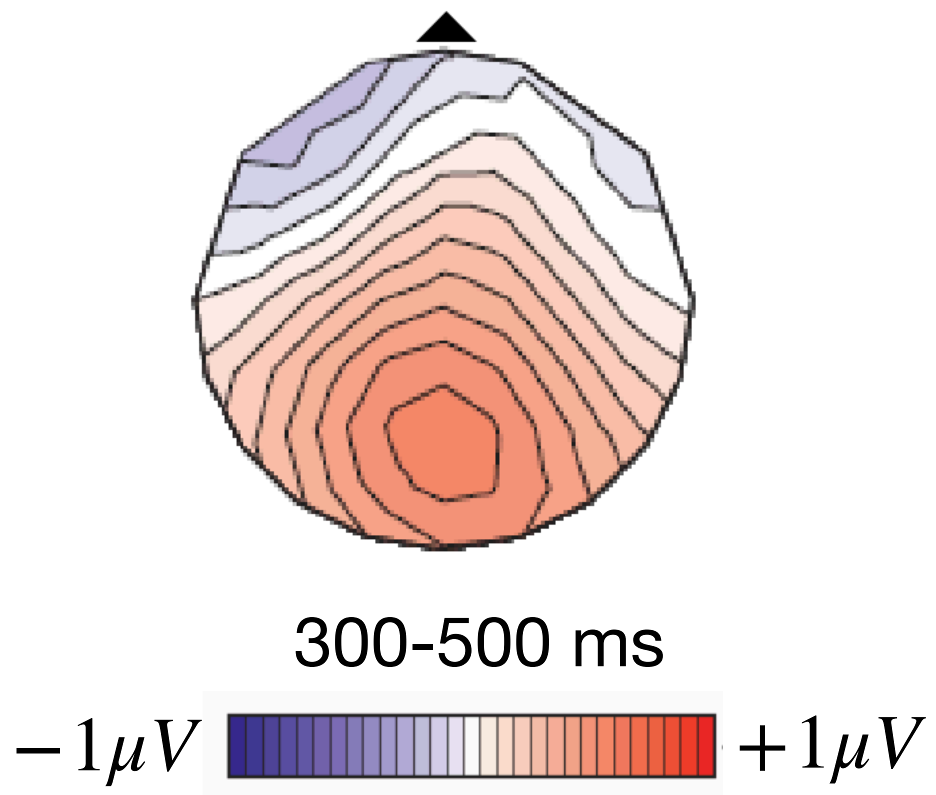
Event Related Potential (ERP)



- ERP: Electrical potentials (voltages) that are related to specific events
- Average across the epochs of that event
- Random noise averages out.

ERP metrics:

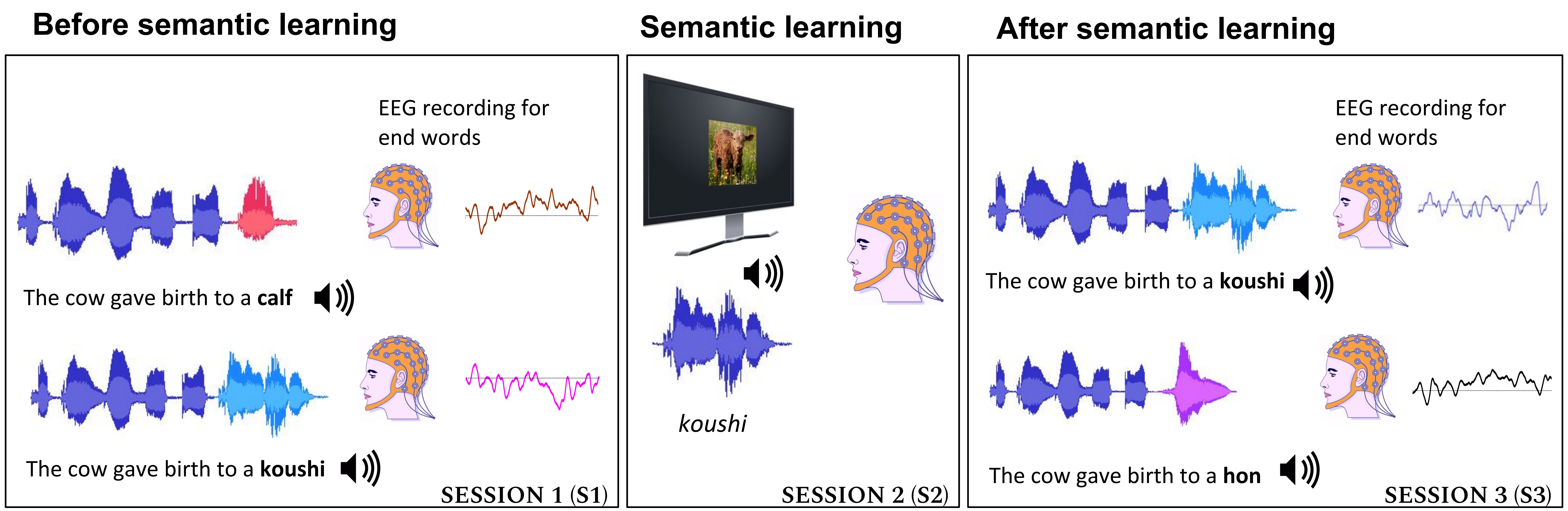
- Amplitude
- Latency
- Polarity



Scalp Topography

https://erpinfo.org/intro-to-erps-course-materials

Different Stimuli Conditions



Stimuli Condition	Example
C1 - Eng. Congruent	*The cow gave birth to a* ***calf****.*
C2 - Eng. Incongruent	*The cow gave birth to a* ***book****.*
C3 - Jap. Congruent	*The cow gave birth to a* ***koushi****.*
C4 - Jap. Incongruent	*The cow gave birth to a* ***hon****.*

- Sentences with highly predictable end word
- English end word replaced with Japanese words