Cross-Subject Semantic Decoding with Shared-Space Alignment for Generalized Neural Representation Learning

Ji-Hoon Heo, Aleksandra Wisniewska,

Seo-Hyun Lee, and Seong-Whan Lee

Korea University

IEEE SMC 2026

KOREA UNIVERSITY

Why is cross-subject decoding challenging?

Different recording configurations

Electrode locations and counts vary across participants.

Subject-specific neural responses

The same linguistic stimulus can evoke different recorded neural patterns.

Limited data from a new participant

Collecting large amounts of invasive neural data is impractical
and burdensome for the participant.

Can we identify neural representations shared across participants during language comprehension and use them to decode semantics for a new participant?

Dataset and decoding task

8
Participants

30 min
Listening to one podcast

5,136
Word occurrences

Dataset

- ECoG recordings during podcast listening [1]
- 183 selected electrodes

Preprocessing

- High-gamma power (70–200 Hz)

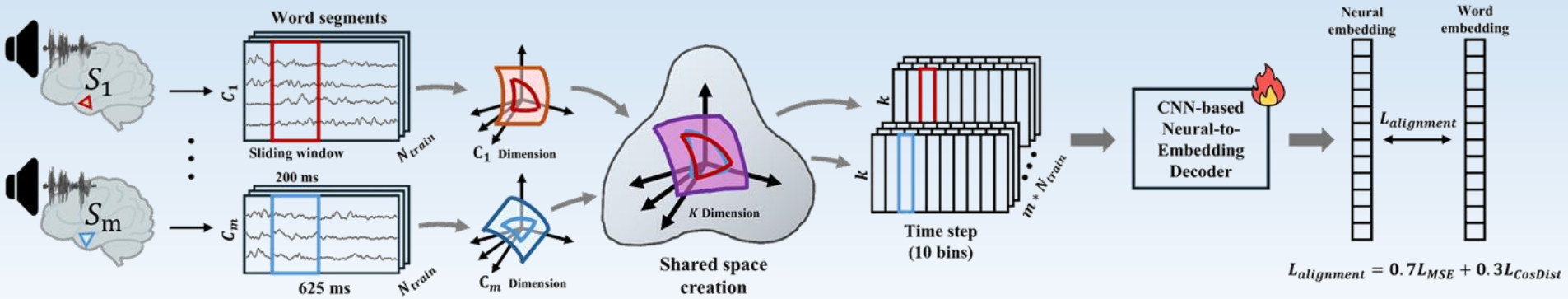
Word-level sample

- 625 ms segment centered 150 ms after word onset
- 200 ms sliding windows
- 10 overlapping temporal bins
- Each word-level neural representation is paired with its GPT-2 contextual embedding

[1] Z. Zada et al., "The 'Podcast' ECoG dataset for modeling neural activity during natural language comprehension," Sci. Data, 2025.

KOREA UNIVERSITY

Aligning participants' neural responses into a shared space



Participant-specific neural spaces over time

At each temporal bin, each participant has C_i electrodes for N word samples

$$X_i^{(t)} \in \mathbb{R}^{C_i \times N}, \quad t = 1, \ldots, T.$$

Learn a shared latent space

At each temporal bin, shared response model (SRM) [2] learns $S^{(t)}$ and a subject-specific projection $W_i^{(t)}$

$$\min_{W_i^{(t)},\, S^{(t)}} \sum_i \left\| X_{i,\text{train}}^{(t)} - W_i^{(t)} S^{(t)} \right\|_F^2$$
$$\text{s.t.} \quad {W_i^{(t)}}^\top W_i^{(t)} = I.$$

Aligned word representation

Each participant is projected into the same k-dimensional space at each temporal bin

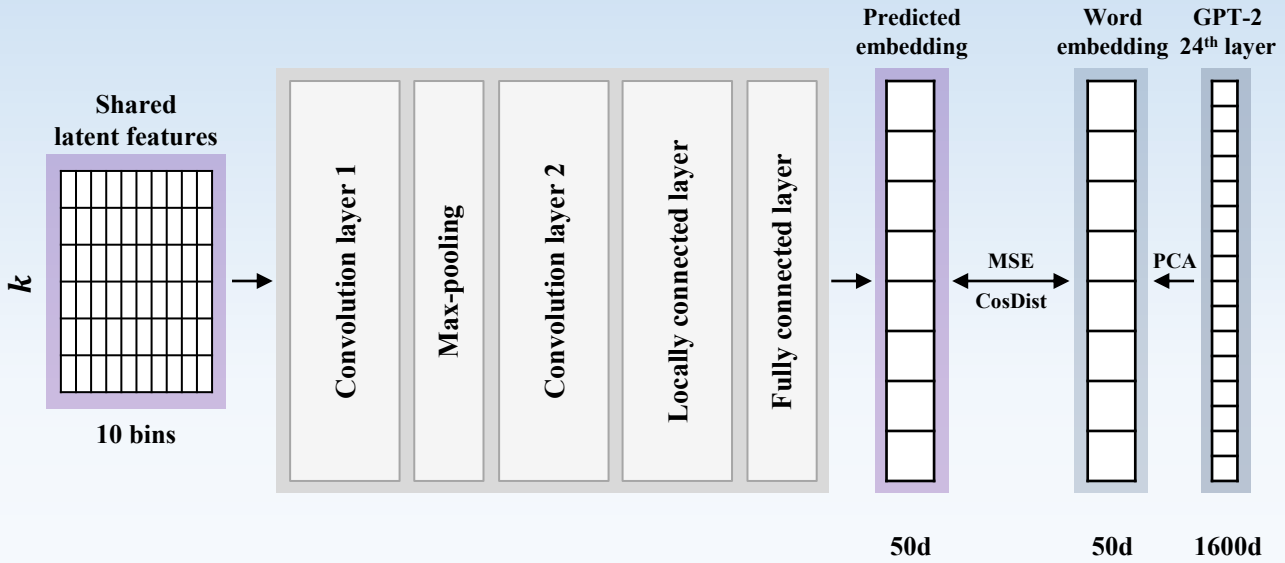
$$Z_{i,train}^{(t)} = {W_i^{(t)}}^\top X_{i,train}^{(t)} \in \mathbb{R}^{k \times N_{\text{train}}},$$

The 10 projected temporal bins are then concatenated to represent each word

[2] P.-H. C. Chen et al., "A Reduced-Dimension fMRI Shared Response Model," NeurIPS, 2015.

KOREA UNIVERSITY

Training a shared semantic decoder



Pooled source representations

Aligned representations from all source participants are pooled for training

One shared decoder

A single CNN decoder [3] maps the shared neural representation to a semantic embedding

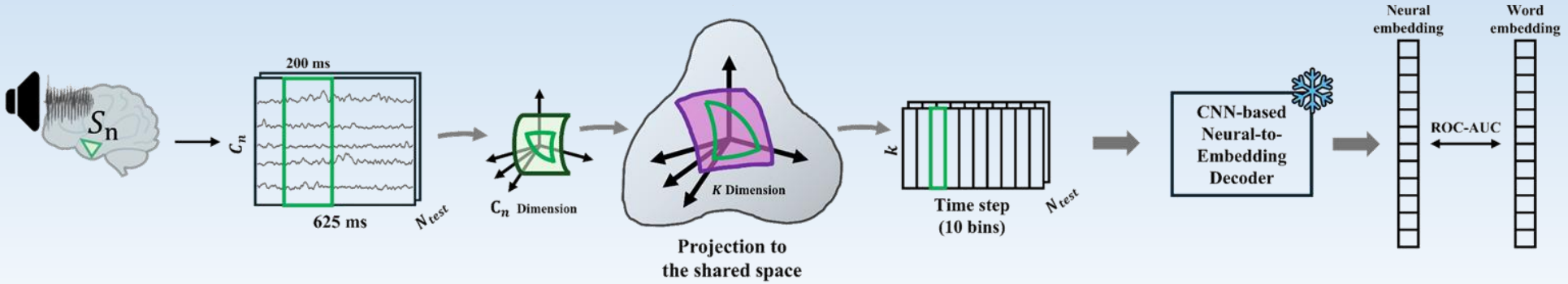
Semantic learning objective

The predicted embedding is matched to the corresponding GPT-2 contextual embedding

Loss = 0.7 MSE + 0.3 Cosine distance

[3] A. Goldstein et al., "Shared computational principles for language processing in humans and deep language models," Nat. Neurosci., 2022.

Adapting to a new participant without retraining



Keep the learned components fixed

Shared latent space and CNN decoder remain unchanged

Estimate only a participant-specific projection

Use the new participant's training neural responses to estimate $W_{target}^{(t)}$

Decode unseen test words

Project the test responses into the shared space and apply the pretrained decoder

No decoder retraining is required for the new participant

KOREA UNIVERSITY

Decoding performance on held-out subjects

Evaluation: Leave one participant out

Five splits: each used once as a test set, with the other four used for fitting the subject-specific matrix

Shared space and decoder are trained on source participants

Only the held-out participant's projection is estimated

TABLE II

CROSS-SUBJECT (HELD-OUT) DECODING PERFORMANCE.

Metric	SRM H/O	PCA+HA H/O	PCA H/O
AUC-ROC	$\mathbf{0.662 \pm 0.044}$	0.640 ± 0.040	0.553 ± 0.029
Pairwise Acc.	$\mathbf{0.656 \pm 0.035}$	0.638 ± 0.036	0.549 ± 0.032
Top-10	$\mathbf{0.051 \pm 0.016}$	0.044 ± 0.012	0.023 ± 0.006
Top-50	$\mathbf{0.209 \pm 0.039}$	0.193 ± 0.033	0.119 ± 0.020

H/O: held-out subject evaluation.
HA: hyperalignment

SRM achieves the highest held-out performance across all four metrics

(Wilcoxon signed-rank test, $p < 0.05$)

KOREA UNIVERSITY

Conclusion

Shared neural representations enable more stable cross-subject semantic decoding

Main finding

SRM achieves better held-out decoding performance and reduces
the source-to-target performance drop

Key implication

A pretrained decoder can transfer to a new participant
by estimating only a subject-specific projection

Limitation

Evaluated on perceived speech with a shared external stimulus
Attempted and imagined speech remain to be tested

KOREA UNIVERSITY

Thank you

Ji-Hoon Heo

Dept. of Artificial Intelligence | Korea University

E-mail: jihoon_heo@korea.ac.kr

KOREA UNIVERSITY