

Reconstructing continuous language from brain signals measured by fMRI based brain-computer interface

Shurui Li¹ | Yuanning Li¹  | Ru-Yuan Zhang^{2,3} 

¹School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai, China

²School of Psychology, Shanghai Jiao Tong University, Shanghai, China

³Shanghai Mental Health Center, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

Correspondence

Yuanning Li and Ru-Yuan Zhang.

Email: liyn2@shanghaitech.edu.cn and ruyuanzhang@sjtu.edu.cn

Funding information

National Key R&D Program of China, Grant/Award Number: 2023YFF1204200; National Natural Science Foundation of China, Grant/Award Numbers: 32100901, 32371154; Shanghai Pujiang Program, Grant/Award Number: 22PJ1410500

KEYWORDS

brain-computer interface, fMRI, language processing

Brain-computer interfaces (BCIs) are designed to bridge the gap between human neural activity and external devices. Previous studies have shown that speech and text can be decoded from signals recorded from intracranial electrodes.^[1] Such applications can be used to develop neuroprostheses to restore speech function in patients with brain and psychiatric disorders.^[2] These methods largely rely on invasive intracranial neural recordings that provide signals with high spatiotemporal resolution and high signal-to-noise ratio. Despite the advantage of being non-invasive, low temporal resolution means functional magnetic resonance imaging (fMRI) has rarely been used in this context to decode continuous speech, with its application primarily limited to coarse classification tasks.^[3]

Despite this, fMRI-based neural encoding models have seen great progress in the last decade. For example, voxel-wise neural responses to continuous natural speech can be predicted using feature embeddings extracted from language models.^[4] To reconstruct continuous speech from fMRI, three obstacles must be overcome. First, the brain's semantic representation regions are not clearly defined—previous research suggests a distributed network across various brain areas. Second, due to its temporal sluggishness, a

single fMRI time point captures information from multiple preceding words within a 6–10-s window. Third, constraining the semantic space in language construction is challenging, as existing fMRI data capture only a fraction of the real semantic richness.

In a recently published study,^[5] Tang and colleagues propose a Bayesian method to decode continuous language from brain responses measured by fMRI. Unlike previous attempts to decode semantic vectors (S) directly from brain responses (R), this study used brain responses as a control condition for language generation models. The goal was to invert the encoding model to identify the most appropriate stimulus. According to Bayesian theory, the decoder estimates the posterior distribution $P(S|R)$ and finds the stimuli S that maximizes the posterior distribution given the neural response R . Instead of directly building decoders that estimate $P(S|R)$, which is usually intractable due to the aforementioned difficulties, the authors took advantage of the Bayesian decoding framework that $P(S|R) \propto P(S)P(R|S)$ and focused instead on the encoding model $P(R|S)$.

This work successfully overcame the three main barriers to fMRI-based language decoding. First, to localize the brain voxels containing semantic information, encoding

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Brain-X* published by John Wiley & Sons Australia, Ltd on behalf of Ainoohui Medical Technology.



performance was used as a metric to select voxels for decoding. Second, to deal with the temporal sluggishness of blood oxygen level-dependent (BOLD) signals, the semantic information for 10 s preceding each repetition time was used to build the encoding model. Third, to ensure that meaningful and readable sentences could be reconstructed, the language model GPT-1 was used to parameterize the prior distribution $P(S)$ over the entire semantic space. GPT-1 uses an autoregressive model to predict words based on prior context, enabling natural language generation. Additionally, a beam search algorithm was used to maintain a relatively large and stable candidate pool.

We note several differences between non-invasive fMRI-based and invasive electrophysiology-based language decoding. The success of language decoding in this study is mainly due to the distributed nature of semantic representations in the brain, and the fact that semantic representations during speech perception can be reliably captured by BOLD signals. However, semantic space is highly multi-dimensional, continuous, and infinite. Invasive speech BCIs rely on electrophysiological signals with high temporal resolution from the sensorimotor cortex; finite, discrete sets of decoding targets, such as phonemes or letters, result in relatively low word error rates. Nevertheless, the semantic reconstruction approach proposed in this study is promising for decoding higher-level amodal concepts, for example, the decoding of text from silent videos, which cannot be easily achieved by invasive speech-motor BCIs.

Despite the many advantages mentioned above, this work still has some limitations. First, in the Bayesian decoding framework, the effectiveness of the decoder depends heavily on the performance of the encoding model. GPT-1 embeddings may represent only a subset of the semantic information in the brain. For example, in this work, only well-encoded voxels were used for decoding. The remaining voxels are probably also involved in semantic representation, but cannot be encoded by GPT-1 embeddings. Second, this work assumed that the total brain response is the sum of responses to semantics in previous time points. This assumption may not be consistent with the actual activation process in the brain.

Despite its limitations, this study sheds new light on non-invasive BCI techniques. We see several promising directions for BCIs in the future. First, safer, portable, and durable invasive BCIs could help thousands of patients with neurological disorders to express their thoughts. Second, cheaper, smaller non-invasive BCIs may have clinical and entertainment applications, such as in the metaverse. Finally, it is also crucial to improve the temporal resolution of non-invasive BCIs. For example, combination with electroencephalogram or magnetoencephalography data could compensate for the low temporal resolution of fMRI. With higher temporal resolution, the decoder could use both semantic and sensorimotor information to improve reconstruction accuracy.

AUTHOR CONTRIBUTIONS

Shurui Li: Conceptualization; formal analysis; visualization; writing—original draft. **Yuanning Li:** Conceptualization; funding acquisition; investigation; resources; supervision; validation; visualization; writing—review and editing. **Ru-Yuan Zhang:** Conceptualization; formal analysis; funding acquisition; project administration; resources; supervision; validation; visualization; writing—original draft; writing—review and editing.

ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (2023YFF1204200), the National Natural Science Foundation of China (32100901) to Ru-Yuan Zhang, and the Shanghai Pujiang Program (22PJ1410500) and the National Natural Science Foundation of China General Program (32371154) to Yuanning Li.

CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

This is a commentary paper with no empirical data.

ETHICS STATEMENT

This is a commentary paper with no empirical experiment.

ORCID

Yuanning Li  <https://orcid.org/0000-0003-3277-0600>

Ru-Yuan Zhang  <https://orcid.org/0000-0002-0654-715X>

REFERENCES

1. Willett FR, Avansino DT, Hochberg LR, Henderson JM, Shenoy KV. High-performance brain-to-text communication via handwriting. *Nature*. 2021;593(7858):249-254. <https://doi.org/10.1038/s41586-021-03506-2>
2. Moses DA, Metzger Sean L, Liu JR, et al. Neuroprosthesis for decoding speech in a paralyzed person with anarthria. *N Engl J Med*. 2021;385(3):217-227. <https://doi.org/10.1056/NEJMoa2027540>
3. Mitchell TM, Shinkareva SV, Carlson A, et al. Predicting human brain activity associated with the meanings of nouns. *Science*. 2008;320(5880):1191-1195. <https://doi.org/10.1126/science.1152876>
4. Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*. 2016;532(7600):453-458. <https://doi.org/10.1038/nature17637>
5. Tang J, LeBel A, Jain S, Huth AG. Semantic reconstruction of continuous language from non-invasive brain recordings. *Nat Neurosci*. 2023;26(5):858-866. <https://doi.org/10.1038/s41593-023-01304-9>

How to cite this article: Li S, Li Y, Zhang R-Y. Reconstructing continuous language from brain signals measured by fMRI based brain-computer interface. *Brain-X*. 2024;2:e70001. <https://doi.org/10.1002/brx2.70001>