Semantic decoding across participants and stimulus modalities

Jerry Tang (jerrytang@utexas.edu)
Department of Computer Science, The University of Texas at Austin
Austin, TX 78712, USA

Alexander G. Huth (huth@cs.utexas.edu)
Departments of Computer Science and Neuroscience, The University of Texas at Austin
Austin, TX 78712, USA

Abstract

Brain decoders that reconstruct language from semantic representations have the potential to improve communication for people with language disorders. However, training a semantic decoder for one participant currently requires many hours of brain data from that participant. In this study, we tested whether semantic decoders can be trained on data from separate reference participants, and then transferred to the goal participant. We functionally aligned the brains of the participants using responses to story and movie stimuli, and then applied the reference decoders to new responses from the goal participant. We found that cross-participant decoders outperformed within-participant decoders trained on the same amount of data. Notably, cross-participant decoding performance was high regardless of whether functional alignment was performed using story or movie stimuli. Our results demonstrate that cross-participant decoding can reduce the amount of training data required from the goal participant, and potentially enable decoding from participants who struggle to comprehend language stimuli.

Keywords: fMRI; language decoding; functional alignment

Introduction

Language production requires mapping between semantic, lexical, and motor representations, and brain decoders can target different stages of this process (Pereira et al., 2018; Tang et al., 2023; Willett et al., 2023; Metzger et al., 2023). Decoders that target semantic representations have the potential to help people with language disorders such as aphasia and apraxia of speech, who struggle to transform semantic representations into lexical or motor representations. However, in order to train a semantic decoder for a participant, many hours of training data are required to map how semantic representations are encoded in that participant's brain (LeBel et al., 2023). Another limitation is that many people with language production impairments also have language comprehension impairments (Wilson et al., 2018), and may struggle to comprehend the language stimuli that are typically used to train semantic decoders.

An alternative to training a new decoder for each goal participant is to train decoders on separate reference participants, and then transfer those decoders to the goal participant. In cross-participant decoding, responses to a shared set of functional alignment stimuli are used to align the brains of the goal and reference participants (Yamada et al., 2015; Ho et al., 2023). Separately, responses to a larger set of decoder training stimuli are used to train the reference decoders. While previous studies performed functional alignment and decoder training using stimuli of the same modality, semantic decoders operate on representations that are shared across modalities (Tang et al., 2023). Since vision and language rely on shared semantic representations (Fairhall & Caramazza, 2013; Devereux et al., 2013; Martin, 2016; Popham et al., 2021; Tang et al., 2024), an intriguing possibility is that semantic decoders trained to reconstruct language stimuli can be transferred across participants solely based on responses to visual stimuli. A method for decoding language from a goal participant without requiring any language training data from that participant could be particularly helpful for people with language comprehension impairments.

Methods

We used functional magnetic resonance imaging (fMRI) to record brain responses from three participants. We decoded each participant using the other two as references. For functional alignment, we recorded brain responses to 320 minutes of stories (LeBel et al., 2023) and 70 minutes of silent movies (Figure 1A). We adapted a cross-participant converter approach previously used to transfer vision decoders across participants (Yamada et al., 2015; Ho et al., 2023). In this approach, linear models are trained to predict the activity in each reference participant voxel using the activity in a population of goal participant voxels. Following our hypothesis that the functional alignment stimuli and decoder training stimuli can be of different modalities, we compared converters trained on responses to stories and movies.

For decoder training, we recorded brain responses to 10 hours of stories (Figure 1B). We used a GPT language model to extract quantitative features of the stimulus words, and we used linear regression to train voxel-wise encoding models that predict brain responses from the stimulus features. Given new brain responses, the we used the decoding approach from (Tang et al., 2023) to generate word sequences that make the encoding model predictions match the responses.

We tested the cross-participant decoders on single-trial brain responses that were recorded while the goal participant listened to a test story that was not used for functional alignment or decoder training (Figure 1C). We used the converters to align responses from the goal participant to the reference

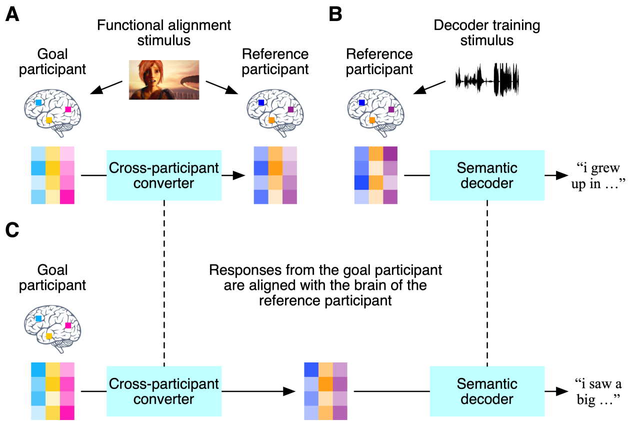

Figure 1: (A) Cross-participant converters predict brain responses in the reference participants using brain responses in the goal participant. (B) Semantic language decoders generate word sequences using brain responses in the reference participants. (C) Given new brain responses in the goal participant, the converters are used to align the responses with the reference brains, and the reference decoders are used to decode the aligned responses.

brains, and then used the reference decoders to decode the aligned responses. We combined predictions across the two reference decoders to make them less sensitive to individual differences in the reference participants. We quantified decoding performance by comparing the predicted and actual story words using BERTScore, which is a metric designed to measure similarity of semantic meaning (Zhang et al., 2019).

Results

We first compared cross-participant decoders to within-participant decoders trained on the same amount of data. For story training data, we sampled subsets of approximately 8, 16, 32, 64, and 128 minutes from 320 minutes of story stimuli (Figure 2B). For movie training data, we sampled subsets of approximately 8, 16, 32, and 64 minutes from 70 minutes of movie stimuli (Figure 2C). To train within-participant language decoders on the movie responses, we transcribed audio descriptions of the movies, and trained voxel-wise language encoding models using the transcripts. For both story and movie training data, we found that cross-participant decoders outperformed within-participant decoders.

We next compared cross-participant decoders aligned using story and movie converters trained on the same amount of data (70 minutes). We obtained a ceiling for decoding performance by training a within-participant decoder on 10 hours of data from the goal participant (Tang et al., 2023). Averaged across all of the time-points in the test story, cross-participant decoding performance for both the story (9.2 standard deviations above chance) and movie (8.3 standard deviations above chance) converters was over half of the within-participant ceiling (15.3 standard deviations above chance). Qualitatively, the cross-participant decoders recovered the meaning of the stimulus, albeit with less precision and consistency than the within-participant decoder (Figure 2C). Decoding performance using a movie converter was comparable to decoding performance using a story converter, despite the fact that the decoders were evaluated on story responses.

Conclusion

Our study demonstrates that either story or movie stimuli can be used to transfer semantic decoders across participants. As language production and comprehension impairments often co-occur, a cross-participant decoding approach may be important for adapting decoders for people with language disorders such as aphasia.

Our findings also have important implications for mental privacy (Goering et al., 2021). It was previously demonstrated that cooperation is required both to train and apply semantic decoders (Tang et al., 2023). Cross-participant decoding still requires functional alignment data collected with cooperation from the goal participant. However, using stimuli of different modalities for functional alignment and decoder training can obscure the relationship between the type of data collected from a participant and the type of data that can be decoded. Consequently, we believe that it is important to raise awareness of what brain data can be used for and establish new rights and regulations for protecting mental privacy.

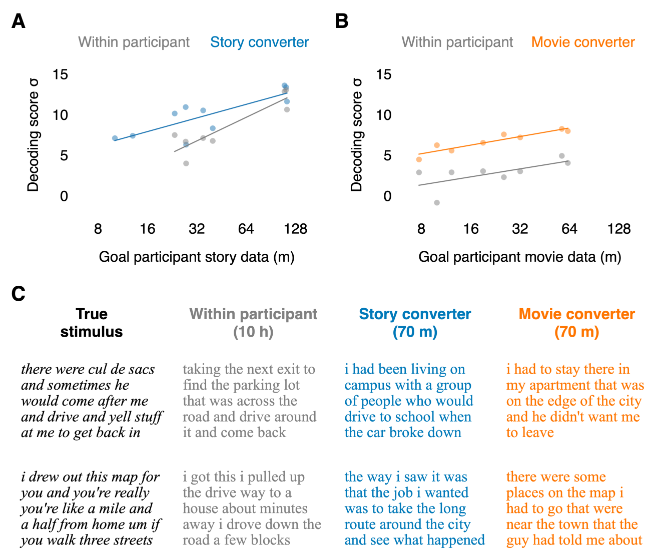

Figure 2: (A) Cross-participant decoders outperformed within-participant decoders trained on the same amount of story data. (B) Cross-participant decoders outperformed within-participant decoders trained on the same amount of movie data. (C) Two segments from the test story are shown alongside decoder predictions from the goal participant. Cross-participant decoders aligned using story and movie data recovered the meaning of the stimulus, albeit less accurately than a within-participant decoder.

Acknowledgments

This work was supported by the National Institute on Deafness and Other Communication Disorders under award number 1R01DC020088-001 (A.G.H.), the Whitehall Foundation (A.G.H.), and the Burroughs Wellcome Fund (A.G.H.).

References

Devereux, B. J., Clarke, A., Marouchos, A., & Tyler, L. K. (2013). Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. *Journal of Neuroscience*, *33*, 18906–18916.

Fairhall, S. L., & Caramazza, A. (2013). Brain regions that represent amodal conceptual knowledge. *Journal of Neuroscience*, *33*, 10552–10558.

Goering, S., Klein, E., Specker Sullivan, L., Wexler, A., Agüera y Arcas, B., Bi, G., ... others (2021). Recommendations for responsible development and application of neurotechnologies. *Neuroethics*, *14*(3), 365–386.

Ho, J. K., Horikawa, T., Majima, K., Cheng, F., & Kamitani, Y. (2023). Inter-individual deep image reconstruction via hierarchical neural code conversion. *NeuroImage*, *271*, 120007.

LeBel, A., Wagner, L., Jain, S., Adhikari-Desai, A., Gupta, B., Morgenthal, A., ... Huth, A. G. (2023). A natural language fmri dataset for voxelwise encoding models. *Scientific Data*, *10*(1), 555.

Martin, A. (2016). GRAPES—grounding representations in action, perception, and emotion systems: How object properties and categories are represented in the human brain. *Psychonomic Bulletin & Review*, *23*, 979–990.

Metzger, S. L., Littlejohn, K. T., Silva, A. B., Moses, D. A., Seaton, M. P., Wang, R., ... others (2023). A high-performance neuroprosthesis for speech decoding and avatar control. *Nature*, *620*(7976), 1037–1046.

Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S. J., Kanwisher, N., ... Fedorenko, E. (2018). Toward a universal decoder of linguistic meaning from brain activation. *Nature communications*, *9*(1), 963.

Popham, S. F., Huth, A. G., Bilenko, N. Y., Deniz, F., Gao, J. S., Nunez-Elizalde, A. O., & Gallant, J. L. (2021). Visual and linguistic semantic representations are aligned at the border of human visual cortex. *Nature Neuroscience*, *24*, 1628–1636.

Tang, J., Du, M., Vo, V., Lal, V., & Huth, A. (2024). Brain encoding models based on multimodal transformers can transfer across language and vision. *Advances in Neural Information Processing Systems*, *36*.

Tang, J., LeBel, A., Jain, S., & Huth, A. G. (2023). Semantic reconstruction of continuous language from non-invasive brain recordings. *Nature Neuroscience*, *24*, 1–9.

Willett, F. R., Kunz, E. M., Fan, C., Avansino, D. T., Wilson, G. H., Choi, E. Y., ... others (2023). A high-performance speech neuroprosthesis. *Nature*, *620*(7976), 1031–1036.

Wilson, S. M., Eriksson, D. K., Schneck, S. M., & Lucanie, J. M. (2018). A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. *PloS one*, *13*(2), e0192773.

Yamada, K., Miyawaki, Y., & Kamitani, Y. (2015). Inter-subject neural code converter for visual image representation. *NeuroImage*, *113*, 289–297.

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2019). Bertscore: Evaluating text generation with bert. In *International conference on learning representations.*