Title

High-Level features from Deep Language Models predict Subthalamic theta power during sentence processing.

Authors and Affiliations

Linyang He[1,3], Alan Bush[1,2], Latane Bullock[1,2], Yanming Zhu[1,2], Yuanning Li[4], Robert M. Richardson[1,2 *]
1 Department of Neurosurgery, Massachusetts General Hospital, Boston, MA, USA
2 Harvard Medical School, Boston, MA, USA
3 Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA
4 School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
* Correspondence to: Robert M. Richardson (mark.richardson@mgh.harvard.edu)

Introduction

The basal ganglia (BG), long associated with motor control, has been less emphasized in language processing compared to cortical regions. However, research has established links between the BG and language-related cortical areas such as the inferior frontal gyrus and other prefrontal regions (Ullman, 2006), indicating its potential involvement in language functions. This has spurred interest in the role of the BG, particularly the subthalamic nucleus (STN), in language processing. While recent studies have identified the STN's participation in motor aspects of speech production and lexical semantics (Chrabaszcz et al., 2021; Lipski et al., 2018; Tankus and Fried, 2019), its extent in high-level language aspects like syntax and contextual semantics remains an area of exploration.

On another front, deep language models (DLMs), emerging as a new tool in computational neuroscience, have become a powerful lens for exploring brain functions related to language processing (Goldstein et al., 2022; Li et al., 2023). The current project aims to leverage DLMs to investigate the role of the STN in language processing.

Methods

We used local field potential (LFP) recordings during deep brain stimulation (DBS) surgery targeting the left STN in two Parkinson's Disease patients. During the recordings, patients engaged in a sentence repetition task: repeating 10 sentences from the Harvard Psychoacoustic Sentences set with a total of 240 trials. We analyzed theta (4-8 Hz), beta (12-30 Hz), and high gamma (75-150 Hz) frequency bands of the STN-LFP data when patients were articulating sentences, correlating LFP with linguistic features derived from GPT-2 large(Brown et al., 2020).

Four types of embeddings from GPT-2 were used: full, lexical (outputs directly extracted from the GPT-2's first embedding layer, without any further contextual processing through the Transformer layers), syntactic (obtained by averaging surrogate sentences with identical syntactic structure but different lexical content, Caucheteux et al., 2021), and residual contextual (containing context-level features and removing lexical information, Toneva et al., 2020). An L2-regularized linear regression model reconstructed the LFP signals from linguistic features, with the correlation coefficient (R score) quantifying the degree of STN's potential involvement in corresponding language aspects. 5-fold cross-validation was applied to obtain reliable R scores.

Results

Our analysis revealed significant correlations across all linguistic features in theta and beta bands for all patients compared to permutation baseline (all p-values < 1e-5). For patient one, the theta band showed the most robust correlation (R=0.39±0.04, mean±std) and the beta and high gamma bands showed average R scores of 0.22±0.04 and 0.28±0.02, respectively. The second patient, despite a dominant theta band in language processing, showed a lower high-gamma R score of 0.07±0.02, possibly due to speech impairment.

Linguistic feature analysis indicated that lexical embedding had lower R scores (0.272 across frequency bands), while syntactic, residual contextual, and full embeddings exhibited similar higher R scores (0.299, 0.301, and 0.299, respectively). Paired-samples t-test was conducted to compare the mean R scores across these conditions. Results indicated a statistically significant difference between the lexical embeddings and the other three types of embeddings (p = 0.0025). In addition, different from cortical studies (Li et al., 2023), we found all Transformer layers of DLM encoded STN features similarly.

Temporal dynamics analysis, extending word onset to 100~600ms pre-onset, showed that lexical features' R scores remained relatively stable, whereas the scores for other higher-level linguistic features exhibited a strong downward trend. This may suggest that the STN's role in lexicon processing is persistent throughout speech production, whereas its involvement in higher-level language processing is more immediate and transient, differing from cortical processing patterns.

Conclusions

Through the lens of DLM, we found STN theta power can be predicted from both lexical-level and high-level language features. Interestingly, these features predict theta band power better than either beta or gamma power. Our results also suggest that the STN exhibits distinct temporal dynamics and correlations with DLM features compared to the cortex. This study is the first to apply DLMs in dissecting the neural substrates of language within the BG. This research offers a novel methodological approach that could broaden our understanding of subcortical structures' role in language processing.

References

Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D., 2020. Language Models are Few-Shot Learners [WWW Document]. arXiv.org. URL https://arxiv.org/abs/2005.14165v4 (accessed 12.1.23).

Caucheteux, C., Gramfort, A., King, J.-R., 2021. Disentangling syntax and semantics in the brain with deep networks, in: Proceedings of the 38th International Conference on Machine Learning. Presented at the International Conference on Machine Learning, PMLR, pp. 1336–1348.

Chrabaszcz, A., Wang, D., Lipski, W.J., Bush, A., Crammond, D.J., Shaiman, S., Dickey, M.W., Holt, L.L., Turner, R.S., Fiez, J.A., Richardson, R.M., 2021. Simultaneously recorded subthalamic and cortical LFPs reveal different lexicality effects during reading aloud. J. Neurolinguistics 60, 101019. https://doi.org/10.1016/j.jneuroling.2021.101019

Goldstein, A., Zada, Z., Buchnik, E., Schain, M., Price, A., Aubrey, B., Nastase, S.A., Feder, A., Emanuel, D., Cohen, A., Jansen, A., Gazula, H., Choe, G., Rao, A., Kim, C., Casto, C., Fanda, L., Doyle, W., Friedman, D., Dugan, P., Melloni, L., Reichart, R., Devore, S., Flinker, A., Hasenfratz, L., Levy, O., Hassidim, A., Brenner, M., Matias, Y., Norman, K.A., Devinsky, O., Hasson, U., 2022. Shared computational principles for language processing in humans and deep language models. Nat. Neurosci. 25, 369–380. https://doi.org/10.1038/s41593-022-01026-4

Li, Y., Anumanchipalli, G.K., Mohamed, A., Chen, P., Carney, L.H., Lu, J., Wu, J., Chang, E.F., 2023. Dissecting neural computations in the human auditory pathway using deep neural networks for speech. Nat. Neurosci. 1–13. https://doi.org/10.1038/s41593-023-01468-4

Lipski, W.J., Alhourani, A., Pirnia, T., Jones, P.W., Dastolfo-Hromack, C., Helou, L.B., Crammond, D.J., Shaiman, S., Dickey, M.W., Holt, L.L., Turner, R.S., Fiez, J.A., Richardson, R.M., 2018. Subthalamic Nucleus Neurons Differentially Encode Early and Late Aspects of Speech Production. J. Neurosci. 38, 5620–5631. https://doi.org/10.1523/JNEUROSCI.3480-17.2018

Tankus, A., Fried, I., 2019. Degradation of Neuronal Encoding of Speech in the Subthalamic Nucleus in Parkinson’s Disease. Neurosurgery 84, 378–387. https://doi.org/10.1093/neuros/nyy027

Toneva, M., Mitchell, T.M., Wehbe, L., 2020. Combining computational controls with natural text reveals new aspects of meaning composition 1–26.

Ullman, M.T., 2006. Is Broca’s Area Part of a Basal Ganglia Thalamocortical Circuit? Cortex 42, 480–485. https://doi.org/10.1016/S0010-9452(08)70382-4