Proceedings of the
9th Graz Brain-Computer Interface Conference 2024
10.3217/978-3-99161-014-4-001

WORD PREDICTION DURING NATURALISTIC SPEECH PERCEPTION

P.A. Lekhnitskaya[1]

[1] Kazan (Volga region) Federal University, Kazan, Russia

E-mail: lekhnitskaya.polina@gmail.com

ABSTRACT: The mechanisms of word prediction have not been studied in the natural speech perception paradigm, which formed the aim of the study: to explore the connection between the function of the EEG responses and the omitted words during naturalistic speech perception, confidence score of trained language model. 14 neurotypical subjects (mean age - 23,5 years; 5 males) participated in the research. EEG included 24 channels. It was proposed to listen to the story and comprehend it. The obtained results show differences in listening to omitted and non-omitted words in T3, T5, P3 electrodes. For modelling the connection between neural signals and naturalistic speech stimuli, mTRF was applied. One of the possible future directions of the research is to explore the communication processes in this paradigm.

INTRODUCTION

The human brain is a complex dynamical system that continuously processes the input information. For acoustic stimuli, as with other types of sensory information, it is important to distinguish signal from noise; and by understanding the features of signal, a person can easily percept the speech. In recent years, researchers have started to shift their attention to the use of continuous, natural speech to explore the ways the brain assesses auditory stimuli [3]. One of the possible approaches, known as system identification, is to model the obtained data based on the speech stimuli [3]. In this vein, the brain is treated as a "black box", in which there are some mappings between the features of the input speech and neurophysiological responses. Such a black box may be represented as a linear time-invariant system with obtaining a so-called temporal response function (TRF) by the connections between EEG and both acoustic and linguistic features [3].

To the best of our knowledge, the mechanisms of word prediction have not been studied in this paradigm. During speech perception, words are embedded in a broader context which facilitates meaning interpretation. Recipients can also make predictions about specific lexemes that can appear in the upcoming discourse. This task is similar to masked language modeling, where a pre-trained model predicts a masked token in a sentence (usually it is marked as [MASK]), by attending to tokens bidirectionally. In this case, a model also makes predictions about the word by its context [7]. Now the neuroscience of perception and language widely uses an integrative modeling approach in which computation and brain function are reflected in computational models [6]. Moreover, direct evidence for alpha, beta and gamma bandwidth in predictive coding is accumulated from observations of increased gamma power to stimuli with prediction errors, but differences in these rhythms with stimulus predictability are not well known [1]. However, the brain responses and the possible link between their reactions and the reactions of trained language models in word prediction have not been simulated.

Thus, the aim of the following study is to explore the connection between the function of the EEG responses during naturalistic speech perception, confidence score of trained language model. It is hypothesized that the link between EEG signals and a trained language model naturalistic speech perception exists. The expected outcome is the approximation of the mentioned link.

MATERIALS AND METHODS

14 neurotypical subjects (mean age - 23,5 years; 5 males) participated in the research. EEG recording was performed by a portable neuro-headset Mitsar-EEG-SmartBCI (Mitsar LLC, St. Petersburg) in a soundproofed room shielded from electromagnetic fields. EEG included 24 channels, in the international 10–20 system; impedance devices are maintained at a level below 10 kOhm. The experiment was implemented in the NeuroBureau program. During experiment, it was proposed to listen to the story about cosmonauts (duration = 5 min 2 sec, language - Russian), in which 48 words in word combinations were omitted (content words without functional ones). The words were chosen randomly, the main criterion was compliance with the context. The task was to understand the whole story. After listening to the recording, subjects were asked to complete a test with questions about the content of the story. Since all participants completed this task without mistakes, we can say with some probability that the missing words were recovered correctly during listening.

The time periods with omitted words were taken by the duration of the omitted word, and the time periods with non-omitted words were taken by the making shift in one second. Tools from "MNE" Python library with integrated methods were used for following EEG analysis [10]. Data preprocessing included the

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-001

filtering, interpolation, artefact removal, re-referencing, and time frequency analysis (tfr_morlet). The high-pass filter is at 1 Hz, low-pass filter is at 40 Hz. During the recording of the study, participants sat motionless with their eyes closed. Artifacts associated with minimal movement were removed using independent component analysis. For each participant was performed this sequence of actions, after that the result data was aggregated in one dataset.

Mann-Whitney U Test (w/ continuity correction) with Bonferroni correction for multiple comparisons and Machine learning classifiers ("Scikit-learn" Python library) were applied to explore differences in EEG responses to the text (TP) and omitted words (OW) comprehension. For training and testing the data was chosen randomly; the size of testing set = 0.3, the size of training one = 0,7 respectively. The preprocessing stage included only applying Standard Scaler for train and test data. The reported results were not cross-validated. The aim of applying binary classification and using so many different classifiers is to additionally prove found by Mann-Whitney U Test differences.

Spearman rank order correlations analysis was used to explore the link between separate electrodes. Next, the transformer python library was utilized (pipeline is fill-mask) [8] with a ruBert-base pre-trained model [9]. The omitted words of the text were marked by [MASK]. Separately, the score reflecting the model's confidence about the selected word was added to the new dataset. Next, the results obtained from the model with the predicted EEG responses were compared. Underlying the computational modeling framework, implemented in the language domain, is the idea that the pre-trained language model can serve as hypotheses of the computations conducted in the brain. Time domain data and other used frequency-band signals analysis is used to investigate the data in terms of complex reactions.

RESULTS

The statistically significant differences were obtained in T3 (p = 0,00, z = -13,97), T5 (p = 0,00, z = 17,47), P3 (p = 0,02, z = 10,91) electrodes in EEG responses to TP and omitted words OW comprehension. Additionally, spearman rank order correlations in OW show connections between T3 and T5 electrodes (r = 0,60, p = 0,00). This first finding shaded light on what electrodes are informative in terms of omitted word prediction.

Next, to explore possibility of the clear distinction between EEG activity while TP and OW phases, machine learning algorithms ("Scikit-learn" Python library) were applied. Data was previously preprocessed by Standard Scaler. Random Forest Classifier, K-Neighbors Classifier, Gradient Boosting Classifier, Logistic Regression, Decision Tree Classifier, MLPClassifier, and Gaussian NB showed high accuracy results among phases (table 1).

Table 1: Machine learning classification results in distinguishing omitted (OW) and non-omitted (NW) words listening

Algorithm	Words	F-score	Accuracy
Random Forest Classifier	OW	.99	.99
	NW	.99	
K-Neighbors Classifier	OW	.99	.99
	NW	.99	
Gradient Boosting Classifier	OW	.98	.98
	NW	.98	
Logistic Regression	OW	.95	.95
	NW	.95	
Decision Tree Classifier	OW	.90	.90
	NW	.89	
MLPClassifier	OW	.90	.90
	NW	.89	
Gaussian NB	OW	.87	.85
	NW	.81	

As distinct differences were found, the next aim was to model EEG responses and by this model try to predict the omitted word. For this purpose, mTRF [2] was used as a forward or encoding model to predict brain responses as the weighted sum of various acoustic and linguistic speech features. Continuous data was analyzed by dividing it into delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (14-30 Hz), and gamma (> 30 Hz) rhythms. Initially, the linguistic speech feature was the frequency of the audio. Correlation between actual and predicted response for delta rhythm r_fwd=0.51, alpha rhythm r_fwd=0.508, beta rhythm r_fwd=0.734, gamma rhythm r_fwd=0.786. The best results are beta and gamma rhythms (fig.2), for them predicted EEG responses were obtained. Although correlation results in this mTRF analysis realization does not have p-values to reveal the significance of findings, it gives possible the connection between EEG activity and the core input (audio story), approximation of the response.

Transformers Python library predicted omitted words and in the model's confidence score and EEG activity correlations were found (table 2).

Table 2: Spearman rank order correlation results between predicted activity in P3, T5 and T3 electrodes and token score given by masked language modeling model.

Rhythm	Electrode	r
Gamma	P3	-0.0005
	T5	-0.2530
	T3	0.1546
Beta	P3	0.1083
	T5	0.0639

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-001

Rhythm	Electrode	r
	T3	-0.3162*

* - statistically significant effect with p<0.05 marked

Statistically significant correlation was observed in beta-rhythm T3 electrode (r=-0.3162, p=0.00), but not in P3 (r=-0.0005, p>0.05), T5 (r=-0.2530, p>0.05), T3 (r=-0.1546, p>0.05) gamma and P3 (r=-0.1083, p>0.05), T5 (r=-0.0639, p>0.05) beta electrodes. From the table 1 we can observe the negative connection between the language model confidence and modeled EEG human language processing.

DISCUSSION

The purpose of this study was to investigate the possibility of connecting modeled EEG response and pre-trained language model to the word prediction task during naturalistic speech perception. For this purpose, the T3 electrode was the most informative, T5, P3 electrodes showed statistically significant differences.

The T3 electrode is proximate to BA44, which might be linked with the prediction of the functional elements (determiners, prepositions, morphological particles) retained within the stimuli [4]. Increased activity in the left inferior frontal gyrus (T3) has been also reported as a function of word integration in the syntactic context [4]. Furthermore, increased directed connectivity from BA44 (T3) to the posterior left middle temporal gyrus (T5 is near this zone) is observed when two-word phrases start with a function word compared to a non-predictive element, possibly reflecting the top-down transmission of a categorical expectation [4]. Machine learning results also reflect the clear difference between OW and TP trials.

For modelling the connection between neural signals and naturalistic speech stimuli, mTRF was applied. The obtained correlation between predicted and real responses denotes neural function, a generalization of the event potential obtained from averaging responses to repetitions of stimuli for continuous data. The proposed EEG model is able to accurately predict activity across neuronal populations in the human cortex during the processing of sentences with omitted words. The proposed idea is similar with the concept of predictive coding, which suggests that the brain has an internal world model. This model encodes causes of sensory inputs as parameters of a generative model. Determining which combination of the many possible causes best fits the current sensory data is achieved through a process of minimizing the error between the sensory data and the sensory inputs predicted by the expected causes [11]. Regarding the current study, the results obtained should be refined in the future to create a more accurate model.

The highest correlation between actual and predicted responses was obtained in beta and gamma rhythms. As the next step, we applied the transformers python library to a similar prediction task with marked omitted words. The model confidence was compared with predicted EEG gamma and beta responses. The predicted EEG response was used to correlate with the language model instead of the actual EEG data to explore the quality of modelling and further possibility to apply predicted EEG response to language modelling domain. We assume that correlation in this case means that model results with prediction of EEG and transformers results have common base of the language perception, and such an approach could give fruitful direction for further investigation both cortical brain organization and the large language models domains.

Statistically significant, but not strong correlation was observed in the beta-rhythm T3 electrode. This is strong evidence for the involvement of beta oscillations across grammatical and semantic processing [5]. Power decreases in beta bandwidth occurring before speech onset within a picture naming task can be provoked by the semantic context provided by a preceding sentence [5]. In our study, we got a negative correlation between beta EEG response and confidence of the language model. The possible explanation is that more predictable words by the language model may be reflected in beta oscillations in modeled human EEG responses.

Such a match between modeled EEG human language processing and the language model may be the first step to creating a semantical network for speech rehabilitation among patients with some types of aphasia. In future it may be of interest to study the communication processes in the proposed paradigm. The main limitation of this research is the sample size.

CONCLUSION

An attempt was made to explore the connection between the function of the EEG responses and the omitted words during naturalistic speech perception. The statistically significant differences were obtained in T3, T5, and P3 electrodes. Machine learning classification algorithms also show distinct differences in EEG signals during audio text comprehension. Anticipatory, likelihood-driven processes are to contribute to lexical, syntactic, and discourse processing, which were studied by mTRF method. We got the modeled brain responses for gamma and beta rhythms as the highest correlation was obtained. This model was compared with the language model. The obtained result may be regarded as the possible solution for developing a semantical network for speech rehabilitation among patients with some types of aphasia. One of the possible future directions of the research is to explore the communication processes in this paradigm and to increase the sample size.

REFERENCES

[1] Bastos, A. M., Lundqvist, M., Waite, A. S., Kopell, N., Miller, E. K. (2020). Layer and rhythm specificity for predictive routing. Proceedings of the

Published by
Verlag der Technischen Universität Graz

10.3217/978-3-99161-014-4-001

National Academy of Sciences, 117(49), 31459-31469.

[2] Bialas, Ole, Jin Dou, and Edmund C. Lalor. mTRFpy: A Python package for temporal response function analysis. Journal of Open Source Software 8.89 (2023): 5657.

[3] Lindboom, Elsa, et al (2023). Incorporating models of subcortical processing improves the ability to predict EEG responses to natural speech. Hearing Research 433: 108767.

[4] Maran, Matteo, et al (2022). Online neurostimulation of Broca's area does not interfere with syntactic predictions: A combined TMS-EEG approach to basic linguistic combination. Frontiers in psychology 13: 968836.

[5] Scaltritti, M., Suitner, C., Peressotti, F. (2020). Language and motor processing in reading and typing: Insights from beta-frequency band power modulations. Brain and Language, 204, 104758. doi:10.1016/j.bandl.2020.104758

[6] Schrimpf, M., Blank, I. A., Tuckute, G., Kauf, C., Hosseini, E. A., Kanwisher, N., Fedorenko, E., et al (2021). The neural architecture of language: Integrative modeling converges on predictive processing. Proceedings of the National Academy of Sciences, 118(45), e2105646118.

[7] Sinha, K., Jia, R., Hupkes, D., Pineau, J., Williams, A., Kiela, D. (2021). Masked language modeling and the distributional hypothesis: Order word matters pre-training for little. arXiv preprint arXiv:2104.06644.

[8] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, F., et al.. (2020). Transformers: State-of-the-Art Natural Language Processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pages 38–45, Online. Association for Computational Linguistics.

[9] Zmitrovich, D., Abramov, A., Kalmykov, A., Tikhonova, M., Taktasheva, E., Astafurov, D., et al. (2023). A family of pretrained transformer language models for Russian. arXiv preprint arXiv:2309.10931.

[10] Gramfort A., Luessi M., Larson E., Engemann D.A., Strohmeier D., et al. (2013) MEG and EEG data analysis with MNE-Python. Frontiers in Neuroscience, 7(267):1–13, doi:10.3389/fnins.2013.00267

[11] Spratling, M. W. (2017). A review of predictive coding algorithms. Brain and cognition, 112, 92-97.

Published by
Verlag der Technischen Universität Graz