

DETECTING GAMMA-BAND RESPONSES TO THE SPEECH ENVELOPE FOR THE ICASSP 2024 AUDITORY EEG DECODING SIGNAL PROCESSING GRAND CHALLENGE

Mike Thornton¹, Jonas Auernheimer², Constantin Jehn², Danilo Mandic¹, Tobias Reichenbach²

¹Imperial College London, ²Friedrich-Alexander-Universität Erlangen-Nürnberg

ABSTRACT

The 2024 ICASSP Auditory EEG Signal Processing Grand Challenge concerns the decoding of electroencephalography (EEG) measurements taken from participants who listened to speech material. This work details our solution to the match-mismatch sub-task: given a short temporal segment of EEG recordings and several candidate speech segments, the task is to classify which of the speech segments was time-aligned with the EEG signals. We show that high-frequency gamma-band responses to the speech envelope can be detected with a high accuracy. By jointly assessing gamma-band responses and low-frequency envelope tracking, we develop a match-mismatch decoder which placed first in this task.

Index Terms— EEG decoding, deep learning, speech

1. INTRODUCTION

The ICASSP Auditory EEG Signal Processing Grand Challenge (SPGC) focuses on the decoding of electroencephalography (EEG) measurements from participants who listened to speech. The 2023 edition of the SPGC included a binary classification problem called the match-mismatch task: given a short segment of EEG signals (three seconds in duration), and two candidate speech segments (also of three seconds duration), the task is to classify which of the speech segments is time aligned, or ‘matched’, with the EEG segment [1]. For this year’s edition, the segment length was increased to five seconds. Additionally, the number of mismatched segments was increased from one to four, thereby increasing the task difficulty.

In the 2023 edition, we achieved the best score in the match-mismatch task through combining two decoders which exploited different EEG responses to speech: slow neural tracking of the speech envelope, and high-frequency speech-related frequency-following responses [2]. Both decoders shared a common deep learning architecture based on that of Accou *et al.* [3]; full details of these decoders are given in Thornton *et al.* [4]. By training multiple instances of the decoders with different weight initialisations and averaging their outputs, a further improvement in decoding accuracy was achieved. Another team showed that it was beneficial to select the mismatched speech segments randomly whilst training the decoders, rather than following the typical method of selecting mismatched segments which occurred at a fixed delay relative to the offset of the matched segments [5].

For this year’s edition of the SPGC, we adapted our previous approach by combining a low-frequency envelope-tracking decoder with a decoder which detects high-frequency (gamma-band) EEG responses to speech. Whereas we previously attempted to relate

Mike Thornton is supported by UK Research and Innovation. [UKRI Centre for Doctoral Training in AI for Healthcare grant number EP/S023283/1]. The project was supported by the German Federal Ministry of Education and Research (SEMECO, 03ZU1210XX).

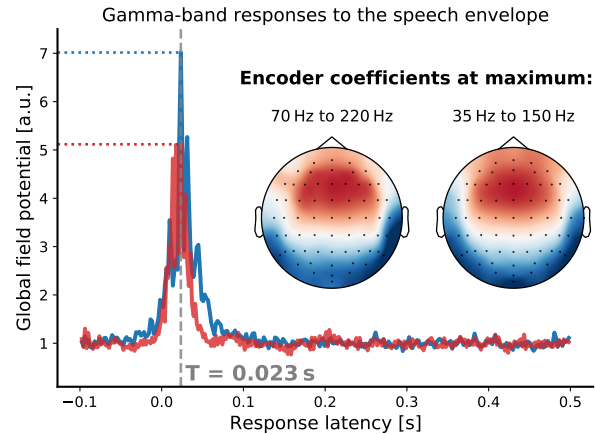


Fig. 1. High-gamma (70 Hz–220 Hz, red) and lower-gamma (35 Hz–150 Hz, blue) envelope-related EEG temporal response functions. The high-gamma range captures the typical range of the fundamental frequency of speech. The global field potentials (GFPs) shown in the figure have been normalised so that the background activity has a value of 1. Therefore, the peak GFP values represent the SNRs of the two responses.

the EEG signals in the high-gamma range (70 Hz to 220 Hz) to the high-frequency envelope-modulations feature, we show here that the speech envelope itself can be related to EEG signals in the broad gamma range. Figure 1 shows that these gamma-band responses exhibit strong signal-to-noise ratios (SNRs) when frequencies as low as 35 Hz are considered.

2. MATERIALS AND METHODS

2.1. Dataset. The SPGC organisers provided a dataset of EEG measurements taken from 85 participants who listened to speech material [6]. Data from 71 participants were used to train the decoders, while the remaining 14 participants served for evaluating the decoders prior to submitting our solutions.

The pre-processing procedure followed that of Thornton *et al.* [4]. Here, due to the short segment length, all filtering operations were performed with low-order IIR filters to reduce edge artefacts. In brief, the low-frequency envelope decoder employed broadband EEG signals and speech envelopes which were resampled to a sampling rate of 64 Hz. For the gamma-band envelope decoder, the EEG signals were filtered between 35 Hz to 150 Hz and resampled to 512 Hz; the broadband speech envelopes were also resampled to this rate.

2.2. Low-frequency (LF) envelope decoder. Given an (EEG segment, envelope segment) pair with a sampling rate of 64 Hz, this

Decoder	5 s segment length		
	Absolute decoding	One imposter	Four imposters
Gamma-band	69.52 \pm 3.32	77.43 \pm 4.35	53.51 \pm 6.78
LF Envelope	76.71 \pm 2.79	85.57 \pm 3.20	66.79 \pm 6.01
Composite	80.73 \pm 2.75	89.95 \pm 2.74	76.18 \pm 5.37

Table 1. The 95% confidence intervals of the mean decoding accuracies of our decoders for 14 heldout participants from the training dataset.

decoder produced a logit which describes how similar the two segments are. The decoder was trained in a binary classification setting using the Adam optimiser and the cross-entropy loss function. We used early stopping and a learning rate scheduler. The finding of Cui *et al.* [5] that it is beneficial to select mismatched segments at random during training was replicated.

2.3. Gamma-band envelope decoder. The gamma-band envelope decoder shared the same architecture and hyperparameters as the LF envelope decoder, and was trained in the same way. However, it used the signals which were sampled at 512 Hz.

2.4. Composite decoder. The composite decoder was formed by combining the output logits of the LF envelope decoder and the gamma-band decoder via a linear classifier (linear discriminant analysis, Figure 2). This classifier produces an estimate of the probability that a particular (EEG segment, stimulus segment) pair is matched. When one pair of segments is known to be matched, and other pairings are imposters, the matched pair is estimated as that which produces the largest predicted probability.

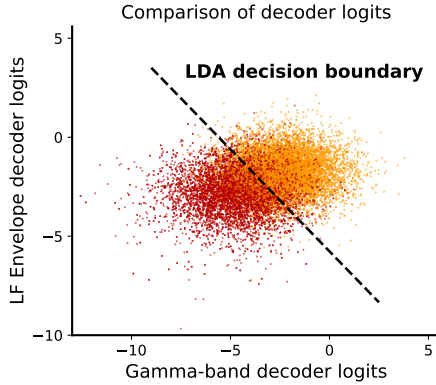


Fig. 2. Comparison between the output logits of the low-frequency (LF) and gamma-band envelope decoders, as well as the decision boundary of the composite decoder. Orange dots represent matched segments, whereas red dots represent mismatched segments.

3. RESULTS

Decoder training (including hyperparameter tuning) was performed using data from 71 of the participants in the training dataset. The remaining 14 participants were held-out for evaluation. Table 1 shows the performances of the three decoders for three related tasks. In *absolute decoding*, the decoders are required to classify whether a single (EEG segment, speech segment) pair is matched or not. In the other two tasks, a single EEG segment and several speech segments are provided, and the task is to classify which one of the speech segments is matched with the EEG.

Table 2 shows how our submitted solutions scored in the challenge. Our first submission achieved an accuracy of 59.97%, plac-

ing our team second by a narrow margin. To improve our score, we re-used the decoder averaging technique which we employed last year. By separately averaging the output logits of 16 instances of each of the envelope and gamma-band decoders, our second solution achieved a considerably higher score of 62.80%, ultimately placing our team first. The composite decoder achieved a much lower score than was expected based on our evaluation which used the 14 held-out participants from the training dataset, possibly suggesting some distributional shift between the two datasets; further investigations are required to understand the cause of this.

Decoder	Heldout accuracy	Leaderboard score
Single-instance	76.18 \pm 5.37	59.97 \pm 6.97
Multiple instances	78.43 \pm 5.21	62.80 \pm 6.88

Table 2. The average accuracies (mean \pm 95% margin of error) of our submitted solutions. Our first submission made use of single-instance decoders. For our second submission, we averaged the output logits of 16 decoder instances before applying the LDA classifier.

4. CONCLUSIONS

We have demonstrated gamma-band EEG responses to the speech envelope. Composite match-mismatch decoders, which monitor both low-frequency and gamma-band responses, have outperformed individual decoders based on either type of response. Responses to the low-gamma components of the speech envelope were particularly strong, and by incorporating these we have improved the classification accuracy of the gamma-band decoder. Finally, by averaging the output logits of multiple trained decoder instances, a considerable performance improvement has been observed.

5. REFERENCES

- [1] J. Monesi *et al.*, “ICASSP 2023: Auditory EEG challenge,” *IEEE SPS SigPort*, 2022.
- [2] M. Thornton, D. Mandic, and T. Reichenbach, “Relating EEG recordings to speech using envelope tracking and the speech-FFR,” in *Proc. ICASSP*, 2023.
- [3] B. Accou *et al.*, “Predicting speech intelligibility from EEG in a non-linear classification paradigm,” *J. Neural Eng.*, 2021.
- [4] M. Thornton, D. Mandic, and T. Reichenbach, “Decoding envelope and frequency-following EEG responses to continuous speech using deep neural networks,” *arXiv*, 2023.
- [5] F. Cui *et al.*, “Relate auditory speech to EEG by shallow-deep attention-based network,” in *Proc. ICASSP*, 2023.
- [6] B. Accou *et al.*, “SparrKULee: A speech-evoked auditory response repository of the KU Leuven, containing EEG of 85 participants,” *BioRxiv*, 2023.