

Декодирование мозговых сигналов в аудиоданные

Набиев Мухаммадшариф Фуркатович

Московский физико-технический институт

*Курс: Моя первая научная статья
(практика, В. В. Стрижов)*

Руководитель: П. А. Северилов

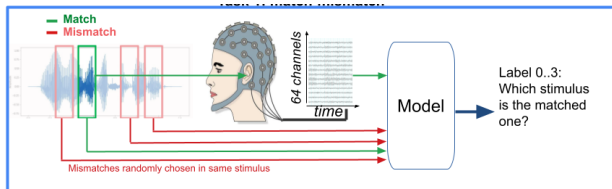
2024

Цель исследования

Задача: Декодирование мозговых сигналов, полученных с помощью электроэнцефалографии(ЭЭГ), в соответствующие аудиоданные.

Эта задача решается в двух постановках: регрессии и классификации

- ▶ Регрессия: Требуется по ЭЭГ восстановить спектрограмму аудиосегмента.
- ▶ Классификация: Определить какой сегмент аудио соответствует сегменту ЭЭГ.

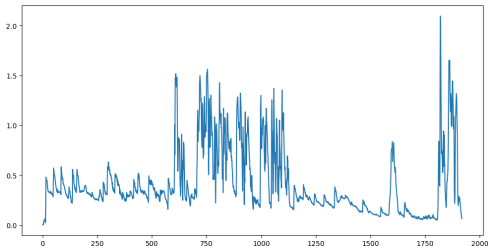


Данные для эксперимента

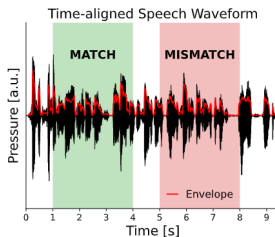
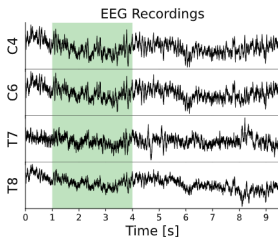
Эксперимент будет проверяться на данных SparrKULee (см. [1]).

- ▶ **Участники:** 74 женщин и 11 мужчин.
- ▶ **Стимулы:** 6-10 аудиофрагментов разной категории, такие как аудиокниги и подкасты, каждый примерно 15 минут.

После обработки частота дискретизации всех данных была понижена до 64 Гц. Аудиофрагменты представлены в виде огибающей кривой сигнала.



Подготовка данных



ЭЭГ и соответствующий аудиофрагмент делятся на сегменты фиксированной длины и для каждой пары ЭЭГ-стимул генерируются ложные стимулы.

Постановка задачи

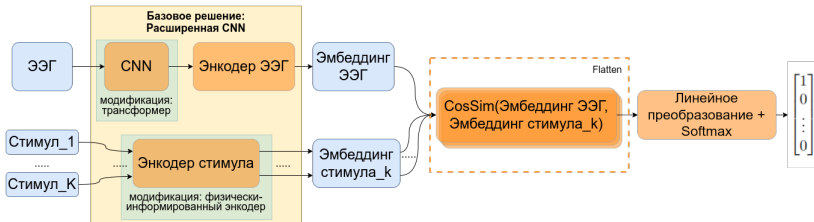
Данные: Кортеж $(\mathbf{X}^i, \mathbf{s}_1^i, \dots, \mathbf{s}_K^i)$, где $\mathbf{X}^i \in \mathbb{R}^{64 \times T}$ — ЭЭГ-сигнал, $\mathbf{s}_1^i, \dots, \mathbf{s}_K^i \in \mathbb{R}^{1 \times T}$ — стимулы, а K — количество стимулов. Меткой данного объекта будет являться вектор $\mathbf{y}^i \in \{0, 1\}^K$. Только один стимул является истинным.

Требуется по имеющимся $\mathbf{X}^i, \mathbf{s}_1^i, \dots, \mathbf{s}_K^i$ получить распределение вероятностей стимулов $\mathbf{p}^i = [p_1^i, \dots, p_K^i]^T$. Пусть модель представляет собой следующее отображение $\mathbf{F} : \mathbb{R}^{64 \times T} \times (\mathbb{R}^{1 \times T})^K \rightarrow \{0, 1\}^K$. Задача сводится к минимизации кросс-энтропии:

$$CE = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K y_k^i \log ([\mathbf{F}(\mathbf{X}^i, \mathbf{S}^i)]_k),$$

где $\mathbf{S}^i = (\mathbf{s}_1^i, \dots, \mathbf{s}_K^i)$. То есть решается задача мультиклассовой классификации.

Решение



Базовое решение:
Расширенная CNN — энкодер, который переводит ЭЭГ и стимулы в латентные пространства, где считается их близость (см. [3]).

Предлагаемые улучшения:
Для ЭЭГ заменить CNN на трансформер и использовать физико-информированный энкодер для стимула (см. [2], [4]).

Дальнейшие планы

- ▶ Добавить энкодеры для стимулов: conformer, whisper, wav2vec
- ▶ Изучить влияние полового признака на качество.

- [1] Lies Bollens, Bernd Accou, Hugo Van hamme, and Tom Francart. SparrKULee: A Speech-evoked Auditory Response Repository of the KU Leuven, containing EEG of 85 participants, 2023.
- [2] Marvin Borsdorf, Saurav Pahuja, Gabriel Ivucic, Siqi Cai, Haizhou Li, and Tanja Schultz. Multi-head attention and gru for improved match-mismatch classification of speech stimulus and eeg response. pages 1–2, 06 2023.
- [3] Bernd Accou et al. Modeling the relationship between acoustic stimulus and eeg with a dilated convolutional neural network. *2020 28th European Signal Processing Conference (EUSIPCO)*.
- [4] Bo Wang, Xiran Xu, Zechen Zhang, Haolin Zhu, Yujie Yan, Xihong Wu, and Jing Chen. Self-supervised speech representation and contextual text embedding for match-mismatch classification with eeg recording. *ArXiv*, abs/2401.04964, 2024.