Декодирование сигналов головного мозга в аудиоданные

Набиев Мухаммадшариф Фуркатович

Московский физико-технический институт

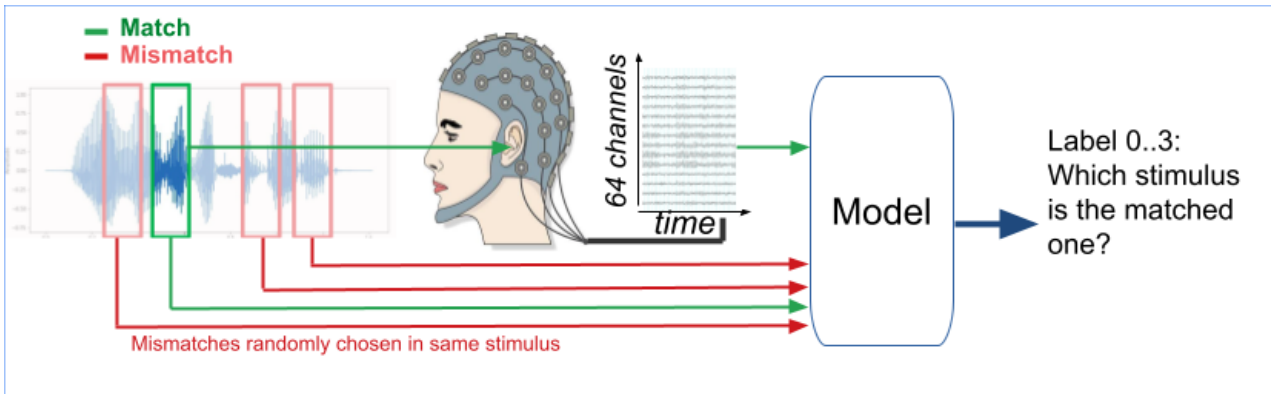
Курс: Моя первая научная статья
(практика, В. В. Стрижов)

Руководитель: аспирант П. А. Северилов

2024

Цель исследования

Цель: Исследовать влияние физико-информированных энкодеров на качество декодирвование мозговых сигналов в аудиоданные.
Задача: Решить задачу декодирования в постановке классификации, а именно определить, какой сегмент аудио вызвал конкретную мозговую активность.



Постановка задачи

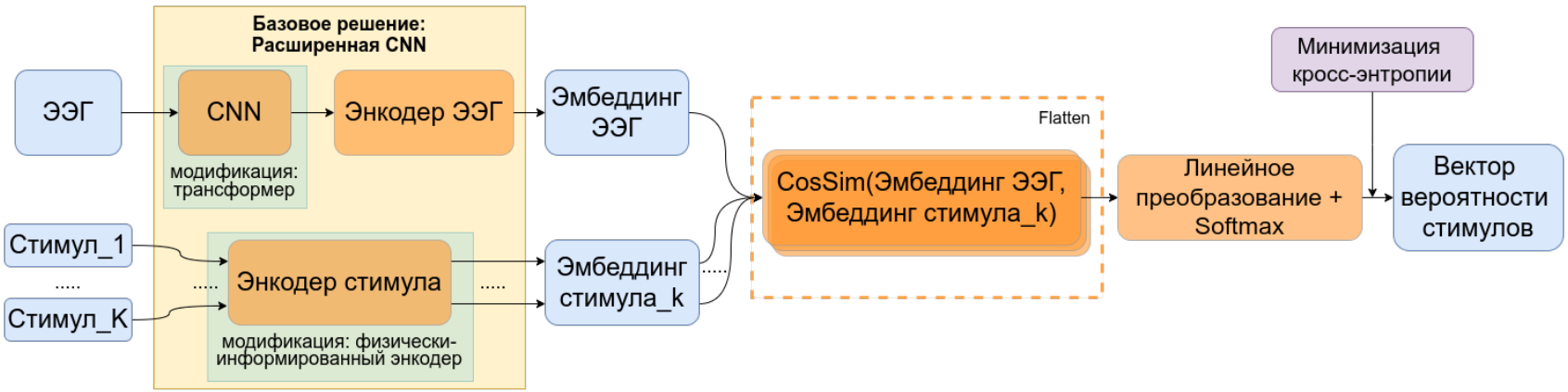
Данные: Кортеж $(\mathbf{X}^i, \mathbf{s}_1^i, \ldots, \mathbf{s}_K^i)$, где $\mathbf{X}^i \in \mathbb{R}^{64 \times T}$ — ЭЭГ-сигнал с 64 каналами, $\mathbf{s}_1^i, \ldots, \mathbf{s}_K^i \in \mathbb{R}^T$ — стимулы, а K — количество стимулов. Меткой данного объекта будет являться вектор $\mathbf{y}^i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

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

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

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

Архитектура решения



Базовое решение:
Расширенная CNN — энкодер, который переводит ЭЭГ и стимулы в латентные пространства, где считается их близость (см. [1]).

Предлагаемые улучшения:
Для ЭЭГ заменить CNN на трансформер-кодировщик и использовать физико-информированный энкодер для стимула (см. [3]).

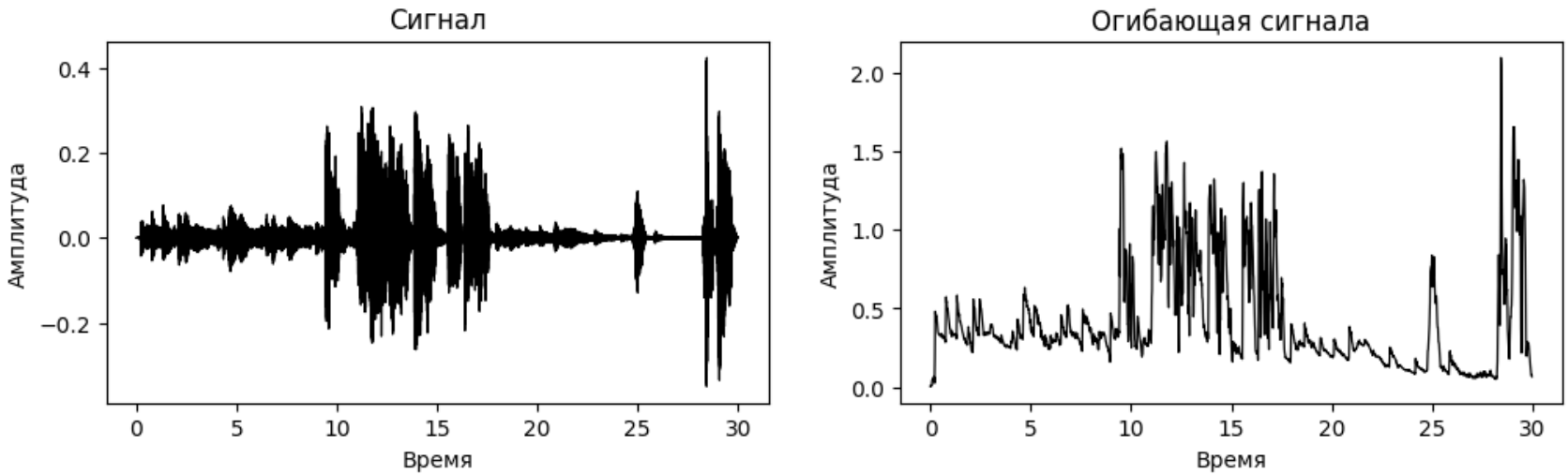
Данные для эксперимента

Эксперимент будет проверяться на данных SparrKULee (см. [2]).

- **Участники:** 85 участников.
- **Стимулы:** 6-10 аудиофрагментов разной категории, такие как аудиокниги и подкасты, каждый длительностью ≈ 15 минут.

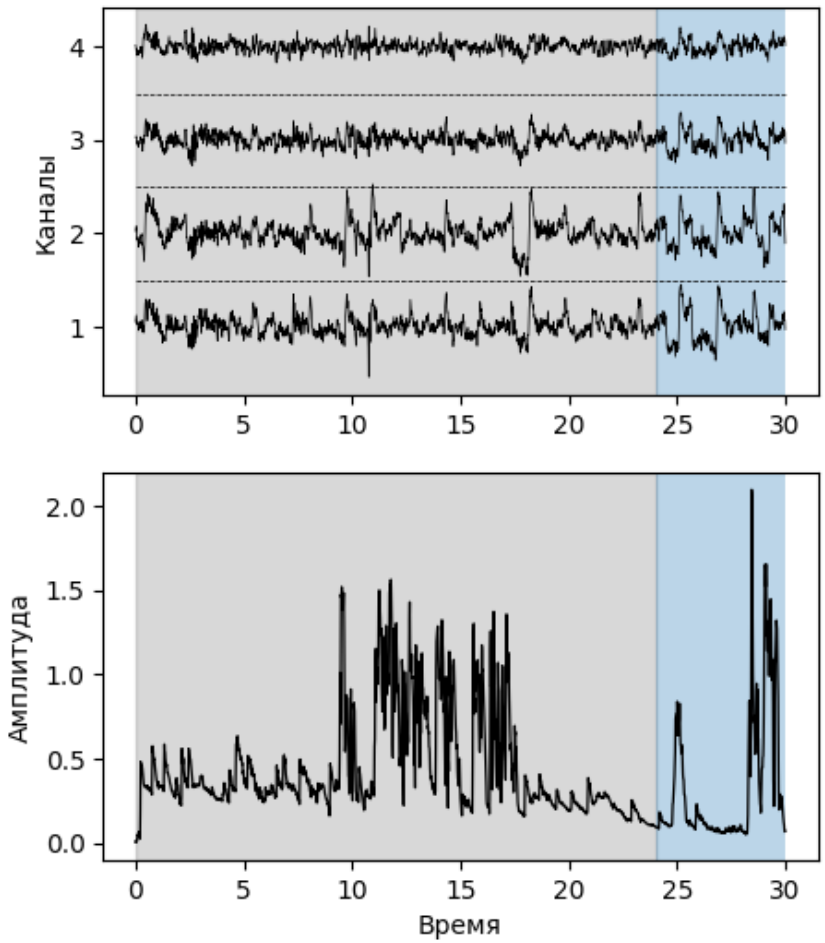
После обработки, частота дискретизации данных была понижена до 64 Гц. Для проведения эксперимента были случайно отобраны 22 участника с одинаковым количеством мужчин и женщин.

Данные для эксперимента



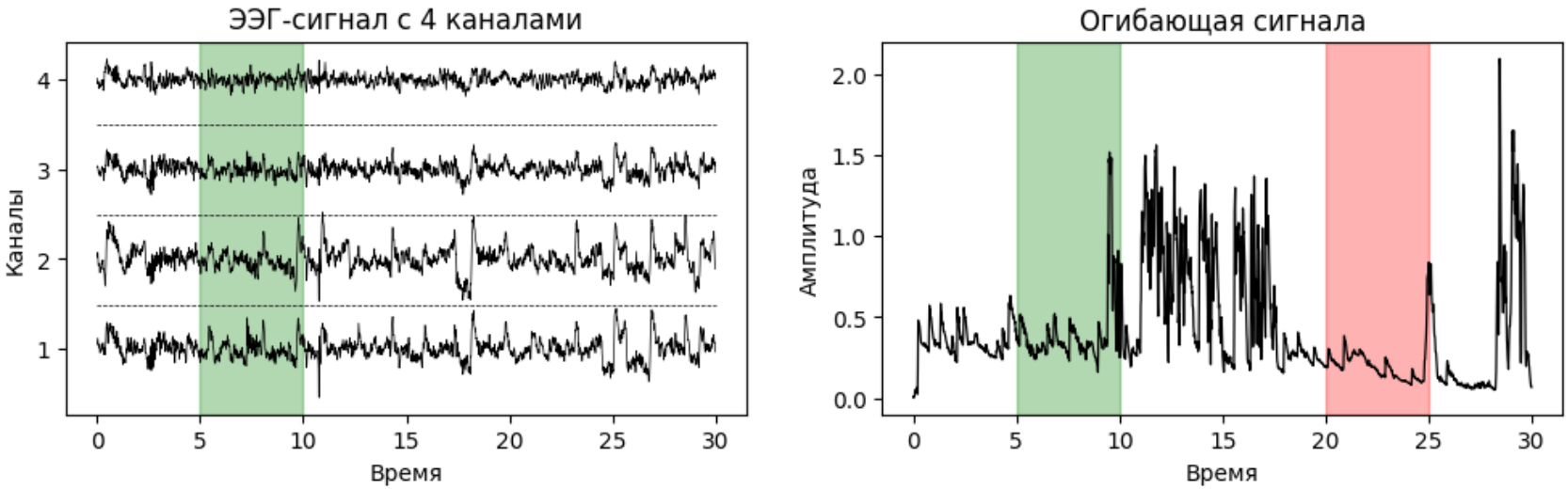
Для предложенной модели был взят аудиосигнал, а для базовой модели ее огибающая. В дальнейшем сегмент сигнала или её огибающей и будет называться стимулом.

Данные для эксперимента



Все данные были разделены в соотношении 80:20. Объединение частей с начала сигнала было использовано в качестве обучающей выборки, а объединение частей с конца было использовано в качестве тестовой выборки (см. рисунок и [1]).

Подготовка данных



Стимул вызвавший активность в мозге в соответствующий промежуток времени называется истинным, а остальные — ложные. Для генерации ложных стимулов были взяты стимулы из других пар (ЭЭГ,стимул).

Вычислительный эксперимент

Параметры эксперимента:

- Размер окна - 5 секунд
- Шаг окна - 1 секунда
- Количество ложных стимулов - 4
- Модель Wav2Vec2.0 - wav2vec2-base-960h-phoneme-reco-dutch
- Модель Whisper - whisper-small

Количество кортежей в обучающей выборке составило 612500, а в тестовой выборке 150075.

Результаты эксперимента

Обозначим множество классов, как $\{0, \ldots, K-1\}$. Учитывая это, метрика качества вычисляется по формуле

$$Score = \frac{1}{22} \sum_{i=1}^{22} \frac{1}{l_i} \sum_{j=1}^{l_i} \left[y_j^i = pred_j^i \right],$$

где $y_j^i \in \{0, \ldots, K-1\}$ — метка объекта, l_i — количество кортежей для i-го участника, а $pred_j^i$ — предсказание модели на объекте j.

Model	Score (%)
Baseline	47.68 ± 11.75
Transformer Encoder	48.15 ± 10.33
Wav2Vec2	47.92 ± 11.54
Whisper-small	48.04 ± 9.85
Transformer Encoder + Wav2Vec2	**48.70** ± 9.44
Transformer Encoder + Whisper-small	48.36 ± 9.24

Результаты эксперимента

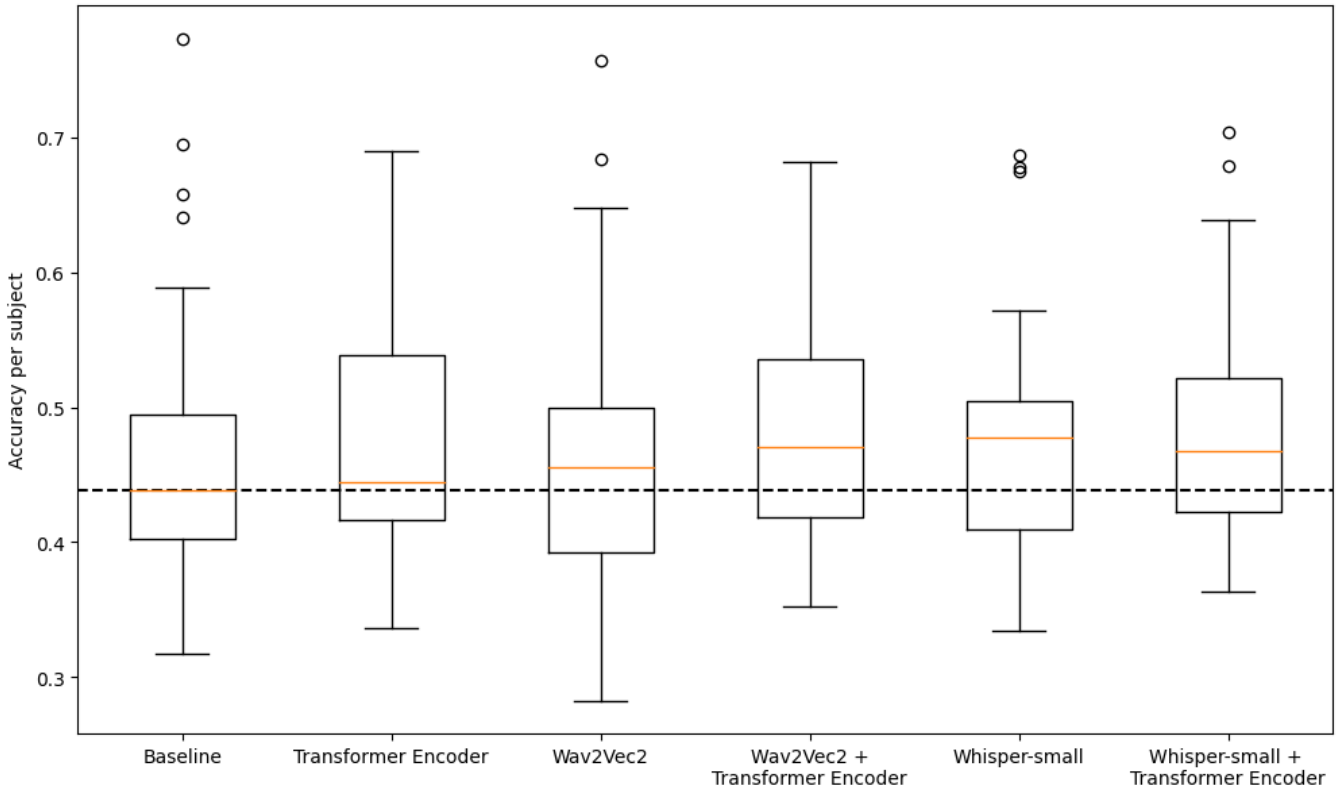


Рис.: Диаграмма размаха для тестовых данных

Наилучший результат был получен за счет комбинирования Wav2Vec2.0 и трансформера-кодировщика.

Источники

[1] Bernd Accou, Mohammad Jalilpour-Monesi, Jair Montoya-Martínez, Hugo Van hamme, and Tom Francart. Modeling the relationship between acoustic stimulus and eeg with a dilated convolutional neural network. *2020 28th European Signal Processing Conference (EUSIPCO)*, pages 1175–1179, 2021.

[2] 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.

[3] 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.