

УДК: 004.81

Тезисы: Предсказание fMRI-изображений по
аудио временным рядам

А.Я. Герман Д.Д. Дорин

А.В. Грабовой

Московский физико-технический институт

Обзор

Понимание того, как мозг преобразует звуки в нейронную активность, имеет ключевое значение как для фундаментальной науки, так и для медицины. Метод функциональной магнитно-резонансной томографии (фМРТ) позволяет отслеживать нейронную активность, измеряя изменения уровня кислорода в крови — так называемый BOLD-сигнал [1, 2].

В этом исследовании мы предлагаем линейную регрессионную модель, способную предсказывать изменения BOLD-сигнала на основе акустических признаков. В качестве признаков используются мел-частотные кепстральные коэффициенты (MFCC), которые имитируют восприятие звука человеком — с помощью преобразования Фурье, мел-шкалы, логарифмического сжатия и косинусного преобразования [3, 4]. Для учёта задержки между звуком и нейронной реакцией мы применяем свёртку аудиофичей с канонической функцией гемодинамического отклика (HRF), подбирая параметр задержки Δt методом перебора [5].

Предложенный подход объединяет прозрачность, интерпретируемость и высокую вычислительную эффективность, предлагая альтернативу сложным нейросетевым моделям [6, 7].

Подготовка эксперимента

Метод протестирован на открытом мультимодальном датасете, включающем данные 30 участников в возрасте от 7 до 47 лет. Участникам показывали 390-секундный фильм с детально размеченными аудио- и видеособытиями [8].

BOLD-объёмы размером $40 \times 64 \times 64$ вокселя регистрировались с частотой 1,64 Гц и проходили предобработку: удаление артефактов движения и физиологических шумов при сохранении стимульных реакций [9]. Аудио имело частоту дискретизации 44,1 кГц, разрезалось на перекрывающиеся

фреймы, и из каждого фрейма извлекались MFCC-признаки размерности 15.

Синхронизация между аудио и фМРТ осуществлялась по формуле $k_\ell = (\ell/\mu), \nu$, а в качестве входных признаков в регрессию использовались MFCC-вектора с учётом задержки Δt . Все временные ряды нормировались методом «мин-макс» в диапазон $[0, 1]$ для обеспечения сопоставимости между участниками.

Эксперименты

Для каждого voxelя (i, j, k) решалась задача регрессии

$$\mathbf{Y}_{ijk} = \mathbf{X}\theta + \varepsilon,$$

где \mathbf{Y}_{ijk} — изменения BOLD-сигнала, а \mathbf{X} — матрица признаков, составленная из MFCC с учётом временной задержки. Подбор параметра Δt проводился в диапазоне от 2 до 8 секунд: сначала для всего мозга, затем — только в слуховой коре для исключения шумов от так называемой «сети режима покоя» [10].

На примере трёх участников (№7, №22, №31) модель достигла очень низких ошибок: $6,39 \times 10^{-5}$, $7,46 \times 10^{-5}$ и $9,00 \times 10^{-5}$. Это говорит о высокой точности воссоздания пространственных паттернов активности. Визуализация показывает соответствие между реальными и предсказанными фМРТ-срезами.

Анализ коэффициентов модели θ выявил широкий спектр значимых признаков, подтверждая, что модель учитывает разнообразие компонент MFCC. Зависимость ошибки от задержки Δt имела чёткий минимум в слуховой области, что подтверждает корректность выбора этого параметра.

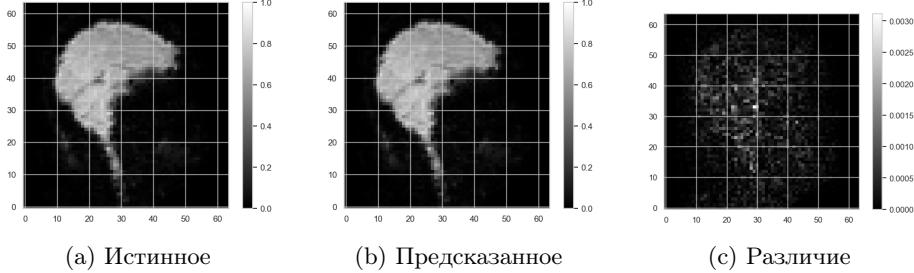


Figure 1: Пример работы алгоритма

Обсуждение

Предложенный метод линейной регрессии обеспечивает прозрачную интерпретацию вклада каждого MFCC-признака в BOLD-ответы voxelей, в отличие от

глубоких сетей. Однако наличие фоновой активности сети режима покоя и остаточные физиологические артефакты подчёркивают важность сбалансированных методов денойзинга [9]. Ограниченный объём выборки ($n=30$) может снижать статистическую мощность и обобщаемость результатов, что требует привлечения более крупных и разнообразных когорт. Линейная модель успешно описывает первые приближения зависимостей, но может не уловить сложных нелинейных взаимодействий, присущих нейронным процессам. Перспективными направлениями являются байесовские модели, учитывающие гетероскедастичность и авторегрессию вокселей [11], а также методы топологического анализа с кубической персистентностью для выделения устойчивых шаблонов [12].

Заключение

Мы показали, что простая интерпретируемая линейная модель способна с высокой точностью предсказывать ответ BOLD по MFCC-признакам аудиосигнала, демонстрируя низкие ошибки и чёткую оценку гемодинамической задержки. Данный подход задаёт вычислительно эффективный эталон для декодирования звуковых стимулов и подчёркивает баланс между простотой модели и её представительной мощностью. В дальнейшем планируется расширить исследование на более крупные наборы данных и внедрить нелинейные методы для полного захвата сложности взаимодействия звука и мозга.

References

- [1] H. White. A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. doi: 10.2307/1912934.
- [2] Catherine A Cooper, Marc R Wilkins, Keith L Williams, and Nicolle H Packer. Bold—a biological o-linked glycan database. *ELECTROPHORESIS: An International Journal*, 20(18):3589–3598, 1999.
- [3] Zrar Kh Abdul and Abdulbasit K Al-Talabani. Mel frequency cepstral coefficient and its applications: A review. *IEEE Access*, 10:122136–122158, 2022.
- [4] Steven Davis and Paul Mermelstein. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE transactions on acoustics, speech, and signal processing*, 28(4):357–366, 1980.
- [5] Peter A Bandettini, Eric C Wong, R Scott Hinks, Ronald S Tikofsky, and James S Hyde. Time course epi of human brain function during task activation. *Magnetic resonance in medicine*, 25(2):390–397, 1992.

- [6] Mingqian Zhao and Baolin Liu. An fmri-based auditory decoding framework combined with convolutional neural network for predicting the semantics of real-life sounds from brain activity. *Applied Intelligence*, 55(2):1–12, 2025.
- [7] Alexander JE Kell, Daniel LK Yamins, Erica N Shook, Sam V Norman-Haignere, and Josh H McDermott. A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. *Neuron*, 98(3):630–644, 2018.
- [8] Julia Bereznitskaya, Mariska J Vansteensel, Erik J Aarnoutse, Zachary V Freudenburg, Giovanni Piantoni, Mariana P Branco, and Nick F Ramsey. Open multimodal ieeg-fmri dataset from naturalistic stimulation with a short audiovisual film. *Scientific Data*, 9(1):91, 2022.
- [9] Molly G Bright and Kevin Murphy. Is fmri “noise” really noise? resting state nuisance regressors remove variance with network structure. *Neuroimage*, 114:158–169, 2015.
- [10] Marcus E Raichle. The brain’s dark energy. *Science*, 314(5803):1249–1250, 2006.
- [11] Anders Eklund, Martin A Lindquist, and Mattias Villani. A bayesian heteroscedastic glm with application to fmri data with motion spikes. *NeuroImage*, 155:354–369, 2017.
- [12] Bastian Rieck, Tristan Yates, Christian Bock, Karsten Borgwardt, Guy Wolf, Nicholas Turk-Browne, and Smita Krishnaswamy. Uncovering the topology of time-varying fmri data using cubical persistence. *Advances in neural information processing systems*, 33:6900–6912, 2020.