ЖУРНАЛ ВЫСШЕЙ НЕРВНОЙ ДЕЯТЕЛЬНОСТИ, 2023, том 73, № 6, с. 800–808

ФИЗИОЛОГИЯ ВЫСШЕЙ НЕРВНОЙ (КОГНИТИВНОЙ) ДЕЯТЕЛЬНОСТИ ЧЕЛОВЕКА

УДК 612.821.6

РАСПОЗНАВАНИЕ УСТНОЙ РЕЧИ ПО ДАННЫМ МЭГ С ИСПОЛЬЗОВАНИЕМ КОВАРИАЦИОННЫХ ФИЛЬТРОВ

В. М. Верхлютов[1,] *, Е. О. Бурлаков[2], К. Г. Гуртовой[3], В. Л. Введенский[3]

[1]Лаборатория высшей нервной деятельности человека, ФГБУН Институт Высшей Нервной Деятельности и Нейрофизиологии РАН, Москва, Россия

[2]ФГБОУ ВО Тамбовский государственный университет им. Г.Р. Державина, Тамбов, Россия

[3]Национальный Исследовательский Центр "Курчатовский Институт", Москва, Россия

**e-mail: verkhlyutov@ihna.ru*

Поступила в редакцию 02.07.2023 г.
После доработки 28.07.2023 г.
Принята к публикации 31.08.2023 г.

Распознавание устной речи по данным ЭЭГ и МЭГ является первым шагом разработки систем МКИ и ИИ для дальнейшего использования их при декодировании воображаемой речи. Большие достижения в этом направлении были сделаны с использованием ЭКоГ и стерео-ЭЭГ. В то же время существует мало работ на эту тему по анализу данных, полученных неинвазивными методами регистрации активности мозга. Наш подход основан на оценке связей в пространстве сенсоров с выделением специфического для данного отрезка речи паттерна связанности МЭГ. Мы проверили свой метод на 7 испытуемых. Во всех случаях наш конвейер обработки был достаточно надежен и работал либо без ошибок распознавания, либо с небольшим количеством ошибок. После "обучения" алгоритм способен распознавать фрагмент устной речи при единственном предъявлении. Для распознавания мы использовали отрезки записи МЭГ 50–1200 мс от начала звучания слова. Для качественного распознавания требовался отрезок не менее 600 мс. Интервалы больше 1200 мс ухудшали качество распознавания. Полосовая фильтрация МЭГ показала, что качество распознавания одинаково эффективно во всем диапазоне частот. Некоторое снижение уровня распознавания наблюдается только в диапазоне 9–14 Гц.

Ключевые слова: декодирование речи, связанность в пространстве сенсоров, МЭГ, ЭЭГ, МКИ, ИИ, тета-ритм, альфа-ритм, гамма-ритм

DOI: 10.31857/S0044467723060126, **EDN:** SWHXYB

Декодирование внутренней речи и речевых стимулов по данным мозговой активности является актуальной задачей для теоретических и прикладных целей современной нейрофизиологии. В рамках данного направления исследователи пытаются решить задачу компенсации утраченных функций при различных видах нарушений воспроизведения и восприятия речи на корковом уровне, что имеет прямое отношение к МКИ. Одновременно изучение этого вопроса помогает продвигаться по пути совершенствования систем ИИ. Существенный прогресс в этом был достигнут при использовании внутричерепных регистраций ЭКоГ (Anumanchipalli et al., 2019) и стерео-ЭЭГ (Norman-Haignere et al., 2022). Однако, инвазивные методики имеют ограниченный диапазон применений. Недавние исследования показали, что расшифровка макроскопических данных фМРТ с использованием обученной языковой модели может достаточно точно декодировать внутреннюю речь на основе семантической информации (Tang et al., 2023).

Неинвазивные методы регистрации, такие как ЭЭГ и МЭГ, доказали, что восприятие и воспроизведение речи влияет как на ритмическую (Vvedensky et al., 2023; Lizarazu et al., 2023; Neymotin et al., 2022), так и на вызванную макроскопическую электрическую активность мозга (Anurova et al., 2023). Таким образом, имеются все предпосылки для декодирования речи по данным МЭГ и ЭЭГ (Dash et al., 2020). Однако для анализа мозговой активности в этом случае используют нейросетевые технологии, результаты которых трудно интерпретируемы. Для этих целей мы предлагаем использовать более простую методику исследования связанности МЭГ в пространстве сенсоров, которая основана на наблюдениях, показавших поразительную схожесть текущей МЭГ-активности на кластерах отведений при прослушивании слов, а также динамическую перестройку этих кластеров при распознавании смысла речевого стимула (Vvedensky et al., 2023).

МЕТОДИКА

Испытуемые. В пилотном исследовании, направленном на тестирование методики, приняло участие 7 испытуемых-добровольцев (4 мужчины и 3 женщины). Один из испытуемых в возрасте 23 года был левшой. Средний возраст молодых испытуемых правшей составил 23.8 ± 0.5 лет. Возраст пожилого испытуемого-правши был 67 лет. Все испытуемые не имели неврологических и психических нарушений в анамнезе. Во всех случаях было получено письменное согласие на проведение исследования по протоколу, утвержденному этической комиссией Института высшей нервной деятельности и нейрофизиологии РАН (протокол № 5 от 15 января 2020 г.). Записи МЭГ проходили в период с 12 до 16 часов дня.

Стимулы. Испытуемому предъявляли три серии речевых стимулов в виде прилагательных русского языка. В каждой серии звучало восемь оригинальных слов, которые повторяли пять раз. Все сорок слов случайным образом перемешивали. Перед каждой серией предъявляли три слова из этой же серии слов для адаптации испытуемого, но данные регистрации при этих предъявлениях не учитывали для анализа. Серии слов различались по длительности звучания и составили соответственно 600, 800, 900 мс (1–3 наборы). Внутри набора длительность звучания не отличалась более чем на 3 мс для исключения влияния длительности звучания декодируемых слов. Громкость звучания была подобрана для каждого испытуемого индивидуально и составила от 40 до 50 дБ. Частота звуков оцифрованных слов в виде аудиофайла не превышала 22 кГц. После предъявления слова испытуемый должен был нажать на кнопку ручного манипулятора, если понял смысл предъявляемого слова. После нажатия кнопки через 500 ± 100 мс (рандоминизировано) следовал следующий стимул, но не позднее 2000 мс после предыдущего предъявления.

Процедура эксперимента. Перед началом эксперимента с помощью устройства трехмерной оцифровки “FASTRAK” (Polhemus, США) определяли координаты анатомических реперных точек (левая и правая преаурикулярные точки и переносица), а также индикаторных катушек индуктивности, прикрепленных к поверхности скальпа испытуемого в верхней части лба и за ушными раковинами. Во время эксперимента испытуемый находился в магнитноэкранированной камере из многослойного пермаллоя (AK3b, Vacuumschmelze GmbH, Германия), а его голова была помещена в стеклопластиковый шлем, являющийся частью стеклопластикового сосуда Дьюара с погруженным в жидкий гелий сенсорным массивом. Испытуемого усаживали таким образом, чтобы поверхность головы находилась максимально близко к сенсорам. Во избежание артефактов звуковые стимулы подавали через пневматическую систему, доставляющую звук от штатного аудиостимулятора. Стимулятор программировали при помощи программы Presentation (США, Neurobehavioral Systems, Inc). Испытуемого просили расслабиться и закрыть глаза. Правой рукой он касался пульта с кнопками. Он должен был нажимать указательным пальцем на одну клавишу после распознавания услышанного слова. Метка начала звучания слова отставала на 9–10 мс от реального нажатия на клавишу. По окончании одной серии предъявлений из трех, испытуемый мог отдохнуть 1–2 мин.

Регистрация. МЭГ регистрировали с помощью 306-канального аппаратно-программного комплекса “VectorView” (Elekta Neuromag Oy, Финляндия), датчики которого покрывают всю поверхность головы и состоят из 102 триплетов, содержащих один магнитометр и два планарных градиентометра, измеряющих взаимно ортогональные компоненты магнитного поля. В настоящем исследовании анализировали данные от всех 306 сенсоров. Это позволяло анализировать магнитное поле как от поверхностных, так и глубоких токовых источников в коре головного мозга испытуемого. Для регистрации глазодвигательной активности использовали два биполярных отведения электроокулограммы (ЭОГ), состоявших из четырех электродов, расположенных на внешних орбитах обоих глаз (горизонтальная составляющая), а также над и под орбитой левого глаза (вертикальная составляющая). Запись сигналов МЭГ и ЭОГ производили с частотой дискретизации 1000 Гц при полосе пропускания 0.1–330 Гц. Положение головы относительно массива сенсоров в ходе эксперимента отслеживали в реальном времени с помощью индикаторных катушек индуктивности. Удаление артефактов записи и коррекцию положения головы проводили с помощью метода пространственно-временóго разделения сигналов, реализованного в программе “MaxFilter” (Elekta Neuromag Oy, Финляндия). MaxFilter является запатентованной технологией производителя магнитометрической системы и очищает запись МЭГ от основных артефактов физического (колебания внешнего магнитного поля, радиочастотная и сетевая наводка) и физиологического происхождения (ЭКГ, окулограмма, движение головы относительно датчиков за счет дыхания, баллистического эффекта при сокращении сердца, непроизвольных движений испытуемого).

Анализ данных. Мы не использовали какихлибо дополнительных методов обработки сигнала кроме “MaxFilter” и полосовой фильтрации для выяснения вклада отдельных частотных диапазонов МЭГ от дельта до гамма в корректность распознавания слов. Актуальные отрезки МЭГ выделяли с помощью меток начала звучания слов. Эти отрезки использовали для построения ковариационных матриц следующим образом.

Для каждого слова и каждого его повтора строится вектор $\bar{M}_{nk}$ (n — номер слова, k — номер повтора), имеющий 306 компонент (по количеству датчиков магнитоэнцефалографа). Обозначим через C_{nk} соответствующую вектору $\bar{M}_{nk}$ ковариационную матрицу (1).

$$C_{nk} = \operatorname{cov}\left(\bar{M}_{nk}\right). \tag{1}$$

Определяли оператор θ_x преобразования квадратных матриц с действительными элементами, принадлежащими отрезку [0, 1], (матриц ковариации), производящий обнуление элементов главных диагоналей матриц (дисперсий $\bar{M}_{nk}$ в случае матрицы C_{nk}) и всех элементов матриц (ковариаций $\bar{M}_{nk}$ в случае матрицы C_{nk}) со значениями ниже порога x. Затем рассчитывали фильтры $\boldsymbol{F_n}$ для каждого слова (т.е. для каждого n от 1 до 8) по следующему правилу: вычтем из усредненной матрицы ковариации $\frac{1}{5}\sum_{k=1}^{5} C_{nk}$ для каждого слова усреднение $\frac{1}{40}\sum_{n=1}^{8}\sum_{k=1}^{5} C_{nk}$ матрицы ковариации по всем предъявлениям аудиальных стимулов, к результату применим оператор $\theta_{0.7}$ (здесь в качестве порога $x = 0.7$ взято значение высокой корреляции по шкале Чеддока) и итоговый фильтр обозначим через $\boldsymbol{F_n}$:

$$\boldsymbol{F_n} = \theta_{0.7}\left(\frac{1}{5}\sum_{k=1}^{5} C_{nk} - \frac{1}{40}\sum_{n=1}^{8}\sum_{k=1}^{5} C_{nk}\right). \tag{2}$$

Вес $w_n(C)$ слова с ковариационной матрицей C относительно n-го фильтра слов $\boldsymbol{F_n}$ можно оценить как

$$w_n(C) = \mathrm{Sum}\left(\mathrm{C}^\circ H\left(\boldsymbol{F_n}\right)\right), \tag{3}$$

где функционал Sum ставит в соответствие любой матрице сумму ее элементов, бинарная операция $^\circ$ представляет собой поэлементное произведение двух матриц (одной и той же размерности), а H является поэлементной функцией Хевисайда. Если вес $w_n(C)$ слова с ковариационной матрицей C по фильтру $\boldsymbol{F_{\hat{n}}}$ превышал вес при всех остальных n, то слово считалось распознанным (номер $n = \hat{n}$) Таким образом, номер распознанного слова можно найти из соотношения (4)

$$\text{номер слова } \hat{n} = \operatorname{argmax}_{n=1,\ldots,8} w_n(C). \tag{4}$$

Веса рассчитывали для всех 40 слов. Пять максимальных значений веса принадлежали распознаваемому слову. Одной ошибкой распознавания считали снижение веса одного из целевых слов ниже максимального веса выбранного из всех предъявлений 35 нецелевых слов. Таким образом, система могла допустить максимально 5 ошибок при распознавании одного слова и 40 ошибок при распознавании 8 оригинальных слов. При этом мы могли оценить успешность распознавания в процентах. При 100% распознавании идентифицировали все 5 одинаковых слов в последовательности из 40 слов. Одна ошибка (неправильно декодированное слово) снижала оценку успешности распознавания на 2.5%. Описанные математические процедуры реализовали в программном конвейере, который доступен на GitHub по адресу: https://github.com/BrainTravelingWaves/22SpeechRecognition. Опубликованы данные МЭГ и МРТ, используемые в исследовании (Verkhlyutov, 2022).

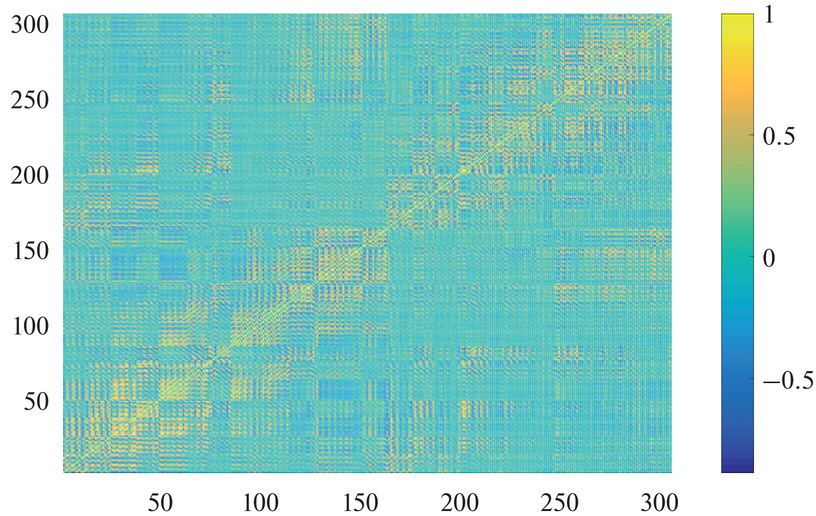

Рис. 1. Матрица корреляционных коэффициентов при сравнении каждого сенсора с каждым 1 секундного отрезка МЭГ от начала звучания первого слова первой серии у испытуемого V1. По осям абсцисс и ординат отложены номера датчиков МЭГ. Цветом кодируется уровень корреляционных коэффициентов.
Fig. 1. Matrix of correlation coefficients when comparing each sensor with each 1 second MEG segment from the beginning of sounding of the first word of the first series in subject V1. The numbers of MEG sensors are plotted along the abscissa and ordinate axes. Color codes the level of correlation coefficients.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ

Корреляционный анализ секундных отрезков МЭГ показал изменения корреляционных коэффициентов от $r < 0.9$ до $r > -0.9$ (рис. 1). Однако для анализа мы использовали только значения $r > 0.7$. Хорошо скорректированные пары датчиков могли быть не только магнитометрами, но и градиентометрами. В некоторых случаях наблюдали значения корреляции выше порога Чеддока между магнитометрическим и градиентометрическим датчиками.

Веса получали для всех 24 слов в трех сериях для каждого испытуемого. На рис. 2 показаны нормализованные значения весов декодированных слов. Звездочками отмечены целевые слова, кружочками фоновые, ковариационные матрицы сигнала МЭГ которых, также пропускали через матрицу-фильтр. Не всегда удавалось успешно распознать целевое слово с весом близким или меньшим чем у одного из фоновых слов или слов выбора. В этом случае распознавание считали ошибочным.

Была проведена попытка подбора оптимальной длинны отрезка анализа МЭГ для распознавания (рис. 3). Сокращение анализируемого отрезка МЭГ слабо сказывалось на качестве распознавания до отрезка 850–1000 мс. На рис. 3 показана успешность распознавания для испытуемого V1 в 3 сериях предъявлений. При интервалах от 200–1000 мс до 750–1000 мс в отдельных сериях наблюдается 100% уровень распознавания. При увеличении такого же анализируемый отрезка МЭГ с 0–50 мс до 0–1200 мс качество распознавания начинало увеличиваться при продолжительности анализируемого отрезка МЭГ от 0 до 600 мс (рис. 4).

Выделение для анализа 100 мс отрезков вызывало общее ухудшение распознавания слов наиболее выраженное на отрезках от 200 до 300 мс и от 1000 до 1100 мс. Наименьшее число ошибок наблюдали на отрезках 300–400 и 700–800 мс (рис. 5).

Мы оценили качество распознавания у всех испытуемых (табл. 1). При этом не выявлено какой-либо тенденции в зависимости от возраста, пола и доминантной руки (испытуемый V3 был пожилым, а испытуемый V5 был леворуким).

Не наблюдались какие-либо тенденции зависимостей между продолжительностью звучания слова-стимула и качеством их распознавания системой. Вероятнее всего качество декодирования зависело от непредсказуемых шумов и нестабильности при регистрации, которые возникали несмотря на все принятые меры по их подавлению.

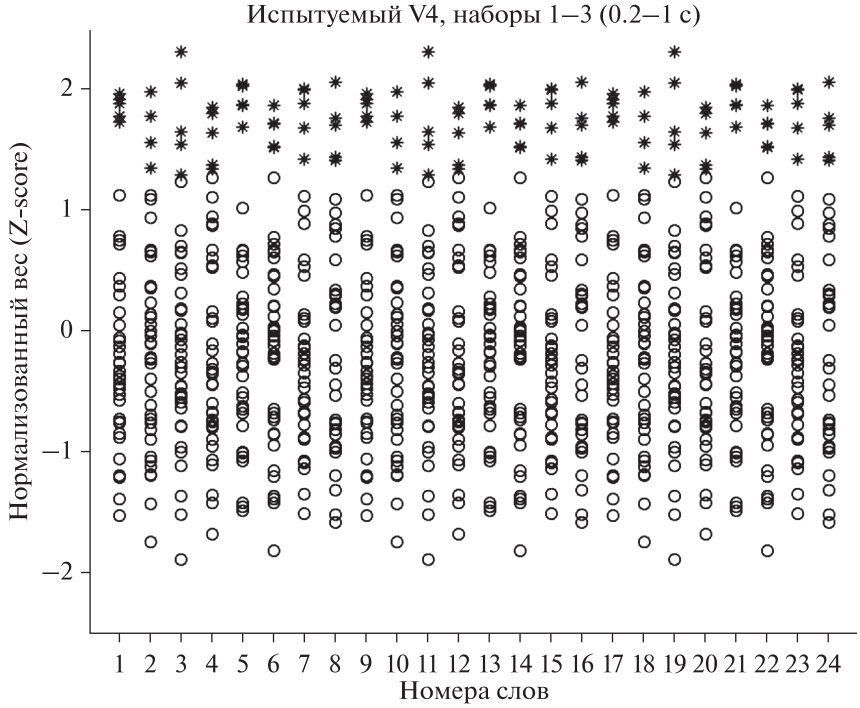

Рис. 2. Нормализованные веса для 24 оригинальных слов (5 предъявлений для каждого слова) для 3 наборов предъявлений испытуемого V4. Анализируемый интервал 0.2–1 с. В каждом столбике обозначены 40 значений весов. Звездочками показаны 5 распознаваемых слов, кружочками фоновые слова. Если звездочек меньше 5, то имеет место наслоение, т.е. веса имею очень близкие значения.
Fig. 2. Normalized weights for 24 original words (5 presentations for each word) for 3 sets of presentations of subject V4. The analyzed interval is 0.2–1 s. Each column contains 40 weights. Asterisks show 5 recognizable words, circles show background words. If there are less than 5 stars, then layering takes place, i.e., the weights have very close values.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Для поведения популяций нейронов важным фактором является их синхронизация, что позволяет многим нейронам работать параллельно и обрабатывать одновременно многие свойства входного сигнала, устанавливая многочисленные связи с другими ментальными объектами и их свойствами (Chen, 2022).

В наших экспериментах мы наблюдали, что часть сигналов от датчиков было скоррелировано при восприятии любого слова, а часть только при прослушивании определенного слова. На этом свойстве основана наша система распознавания речи. При этом мы наблюдали только амплитудную связанность, которая обусловлена удаленными связями (Rolls et al., 2022) в отличие от фазовой

Таблица 1. Процент правильного распознавания у 7 испытуемых при использовании отрезка 0–1000 мс от начала звучания слова без фильтрации для трех наборов слов
Table 1. Percentage of correct recognition in 7 subjects when using a segment of 0–1000 ms from the beginning of the sound of a word without filtering for three sets of words

Испытуемый	Пол	Возраст	Набор 1 (%)	Набор 2 (%)	Набор 3 (%)
V1	м	24	100	72.5	80
V2	ж	24	90	90	80
V3	м	67	97.5	100	82.5
V4	м	27	100	97.5	100
V5	м/л	23	100	100	85
V6	ж	20	62.5	72.5	92.5
V7	ж	24	75	100	100

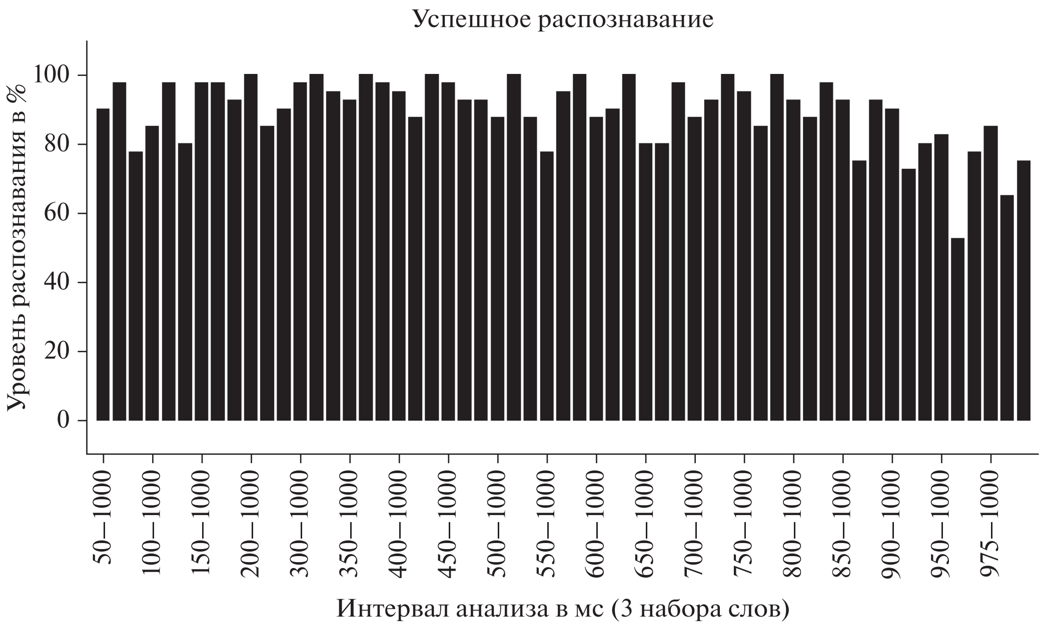

Рис. 3. Эффект снижения успешного распознавания, при сокращении анализируемый отрезка МЭГ с 50–1000 мс до 975–1000 мс от начала звучания слова у испытуемого V1. Единичный столбец обозначает процент успешного декодирования при распознавании одного набора из 8 оригинальных слов. Интервалы обозначены для трех последовательных наборов слов. Если все слова распознаны, успешность распознавания 100%.

Fig. 3. The effect of reducing successful recognition, when the analysed segment of the MEG is reduced from 50–1000 ms to 975–1000 ms from the beginning of the sound of the word in the subject V1. A single bar indicates the percentage of errors in the recognition of one set of 8 original words. The intervals are indicated for three consecutive sets of words. If all words are recognized, the recognition success rate is 100%.

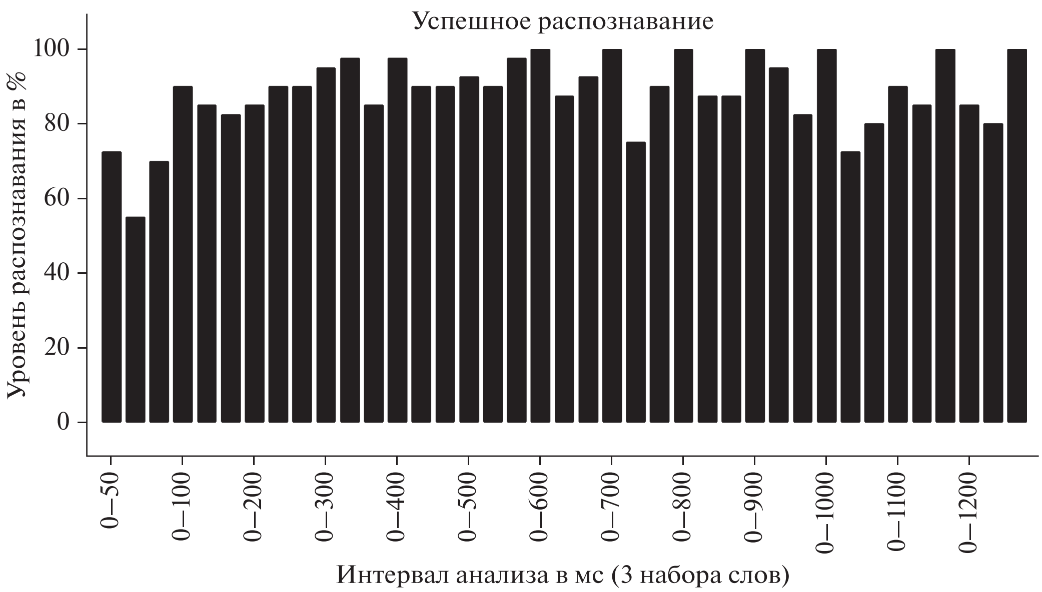

Рис. 4. Уровень успешного декодирования при увеличении анализируемый отрезка МЭГ с 0–50 мс до 0–1200 мс от начала звучания слова у испытуемого V1. Единичный столбец обозначает процент успешного декодирования при распознавании одного набора из 8 оригинальных слов. Интервалы обозначены для трех наборов слов. Уровень распознавания не достигает 100% до интервала 0–600 мс.

Fig. 4. The level of successful decoding with an increase in the analyzed segment of the MEG from 0–50 ms to 0–1200 ms from the beginning of the sound of the word in the subject V1. The single column indicates the percentage of successful decoding when recognizing one set of 8 original words. The intervals are indicated for the three sets of words. The recognition level does not reach 100% until the interval 0–600 ms.

Рис. 5. Уровень успешного декодирования при выделении отрезков 100 мс интервала 0–1000 мс от начала звучания слова у испытуемого V1. Единичный столбик обозначает % успешного распознавания при декодировании одного набора из 8 оригинальных слов повторяющихся 5 раз. Интервалы обозначены для трех наборов слов. Короткие отрезки МЭГ не позволяют декодировать 100% целевых слов.
Fig. 5. The level of successful decoding when selecting segments of 100 ms interval 0–1000 ms from the beginning of the sound of the word in the subject V1. A single bar denotes the % recognition success when decoding one set of 8 original words repeated 5 times. The intervals are indicated for the three sets of words. Short MEG segments do not allow 100% of target words to be decoded.

связанности, которая, в свою очередь, обеспечивается локальными взаимодействиями. Наличие фазовой связанности в наших экспериментах доказывает присутствие как положительно, так отрицательно скоррелированных данных. Это относительно стабильные фазовые задержки (Arnulfo et al., 2020) указывают на возможные эффекты, обусловленные вращением токовых диполей, которые возникают вследствие динамики корковых бегущих волн (Sato, 2022).

Эффективность использования отдельных отрезков МЭГ-сигнала для декодирования можно объяснить имплицитной перцепцией или предикативным кодированием, позволяющей мозгу рассматривать часть фразы как имеющую законченный смысл (Liaukovich et al., 2020). Под эффективностью мы понимали число или процент безошибочно распознанных слов. Как показано на графиках (рис. 4–6), эффективность распознавания зависела от набора предъявленных слов, от отрезка МЭГ, который мы вырезали, передвигая окно анализа в пределах от 0 до 1.2 с после начала звучания слова и от полосовой фильтрации.

Частотная полосовая фильтрация показала, что для декодирования слов можно использовать все электроэнцефалографические диапазоны мозговых волн, которые порождаются как ближним, так и дальним взаимодействием между электрическими мозговыми источниками (Proix et al., 2022).

Наряду с другими исследователями (Dash et al., 2023) мы доказали возможность декодирования речевых образов с использованием МЭГ, которая обладает сравнительно низким пространственным разрешением.

Как описано в работе Huth и соавторов (Huth et al., 2016), при использовании метода фМРТ слышимое или воображаемое слово активирует множество корковых структур, содержащих информацию, ассоциируемую с данным словом, которые выходят далеко за пределы зоны Вернике. Это можно назвать “семантическим усилением”. Зона Вернике невелика по площади, и зарегистрировать сигналы от нее без усреднения весьма проблематично. Поэтому работа с ней требует применения инвазивных методов (имплантация ЭКоГ-матриц или микроэлектродов) для декодирования слышимой или воображае-

Рис.6. Успешное распознавания, при полосовой фильтрации МЭГ интервала 200–1000 мс от начала звучания слова у испытуемого V1 в диапазонах 0.5–30, 0.5–3, 4–8, 9–14, 15–30, 30–100 Гц. Единичный столбик обозначает процент распознанных целевых слов относительно фоновых при распознавании одного набора из 8 оригинальных слов, повторно предъявляемых 5 раз. При успешном распознавании 40 (8 × 5) целевых слов при тестировании 320 (40 × 8) отрезков МЭГ уровень распознавания равнялся 100%. Полоса фильтрации обозначена для трех наборов слов.
Fig. 6. Successful recognition, with bandpass filtering of the MEG interval of 200–1000 ms from the beginning of the sound of the word in the subject V1 in the ranges of 0.5–30, 0.5–3, 4–8, 9–14, 15–30, 30–100 Hz. A single bar indicates the percentage of recognised target words relative to the background ones when recognising one set of 8 original words repeated 5 times. With successful recognition of 40 (8 × 5) target words when testing 320 (40 × 8) MEG segments, the recognition level was 100%. The filtering band is indicated for three sets of words.

мой речи. “Семантическое усиление” увеличивает площадь активированной речевым стимулом коры, которая дает более сильные сигналы и позволяет работать над декодированием внутренней речи с использованием МЭГ или ЭЭГ без усреднения данных. И как недавно было показано, это явление позволяет создать систему декодирования внутренней речи с использование нейронных сетей на основе фМРТ (Tang et al., 2023). Но если использовать наш метод в качестве предварительной обработки перед анализом нейронной сетью, возможно создание не менее эффективных систем на основе МЭГ и ЭЭГ.

ФИНАНСИРОВАНИЕ

Работа выполнена на базе Московского МЭГ-центра при поддержке Российского научного фонда, проект № 23-78-00011.

СПИСОК ЛИТЕРАТУРЫ

Anumanchipalli G.K., Chartier J., Chang E.F. Speech synthesis from neural decoding of spoken sentences. Nature. 2019. 568 (7753): 493–498. https://doi.org/10.1038/s41586-019-1119-1

Anurova I., Vetchinnikova S., Dobrego A., Williams N., Mikusova N., Suni A., Palva S. Event-related responses reflect chunk boundaries in natural speech. NeuroImage, 2022. 255 (April), 119203. https://doi.org/10.1016/j.neuroimage.2022.119203

Arnulfo G., Wang S.H., Myrov V., Toselli B., Hirvonen J., Fato M.M., Palva J.M. Long-range phase synchronization of high-frequency oscillations in human cortex. Nature Communications, 2020. 11 (1): 5363. https://doi.org/10.1038/s41467-020-18975-8

Che B., Ciria L.F., Hu C., Ivanov P.C. Ensemble of coupling forms and networks among brain rhythms as function of states and cognition. Communications Biology, 2022. 5 (1): 82. https://doi.org/10.1038/s42003-022-03017-4

Dash D., Ferrari P., Wang J. Decoding Imagined and Spoken Phrases From Non-invasive Neural (MEG)

Signals. Frontiers in Neuroscience. 2020. 14: 290. https://doi.org/10.3389/fnins.2020.00290

Défossez A., Caucheteux C., Rapin J., Kabeli O., King J.-R. Decoding speech from non-invasive brain recordings. ArXiv. 2022. 2208. 12266: 1–15. http://arxiv.org/abs/2208.12266

Huth A.G., De Heer W.A., Griffiths T.L., Theunissen F.E., Gallant J.L. Natural speech reveals the semantic maps that tile human cerebral cortex. Nature. 2016. 532 (7600): 453–458. https://doi.org/10.1038/nature17637

Liaukovich K., Ukraintseva Y., Martynova O. Implicit auditory perception of local and global irregularities in passive listening condition. Neuropsychologia, 2022. 165 (July 2020): 108129. https://doi.org/10.1016/j.neuropsychologia.2021.108129

Lizarazu M., Carreiras M., Molinaro N. Theta-gamma phase-amplitude coupling in auditory cortex is modulated by language proficiency. Human Brain Mapping, 2023. 44 (7): 2862–2872. https://doi.org/10.1002/hbm.26250

Neymotin S.A., Tal I., Barczak A., O'Connell M.N., McGinnis T., Markowitz N., Lakatos P. Detecting Spontaneous Neural Oscillation Events in Primate Auditory Cortex. Eneuro. 2022. 9 (4), ENEURO.0281-21.2022. https://doi.org/10.1523/ENEURO.0281-21.2022

Norman-Haignere S.V., Long L.K., Devinsky O., Doyle W., Irobunda I., Merricks E.M., Mesgarani N. Multiscale temporal integration organizes hierarchical computation in human auditory cortex. Nature Human Behaviour. 2022. 6 (3): 455–469. https://doi.org/10.1038/s41562-021-01261-y

Proix T., Delgado Saa J., Christen A., Martin S., Pasley B.N., Knight R.T., Giraud A.-L. Imagined speech can be decoded from low- and cross-frequency intracranial EEG features. Nature Communications, 2022. 13 (1), 48. https://doi.org/10.1038/s41467-021-27725-3

Rolls E.T., Deco G., Huang C.-C., Feng J. The human language effective connectome. NeuroImage, 2022. 258: 119352.

Sato N. Cortical traveling waves reflect state-dependent hierarchical sequencing of local regions in the human connectome network. Scientific Reports, 2022. 12 (1): 334. https://doi.org/10.1038/s41598-021-04169-9

Tang J., LeBel A., Jain S., Huth A.G. Semantic reconstruction of continuous language from non-invasive brain recordings. Nature Neuroscience. 2023. https://doi.org/10.1038/s41593-023-01304-9

Verkhlyutov V. MEG data during the presentation of Gabor patterns and word sets. Zenodo, 2022. https://zenodo.org/record/7458233

Vvedensky V., Filatov I., Gurtovoy K., Sokolov M. Alpha Rhythm Dynamics During Spoken Word Recognition. Studies in Computational Intelligence, 2023. 1064: 65–70. https://doi.org/10.1007/978-3-031-19032-2_7

RECOGNITION OF ORAL SPEECH ACCORDING TO MEG DATA BY COVARIANCE FILTERS

V. M. Verkhlyutov[a, #], E. O. Burlakov[b], K. G. Gurtovoy[c], and V. L. Vvedensky[c]

[a]Higher Nervous Activity of a Person Lab., Institute of Higher Nervous Activity and Neurophysiology of RAS, Moscow, Russia

[b]Derzhavin Tambov State University, Tambov, Russia

[c]RNC Kurchatov Institute, Moscow, Russia

[#]e-mail: verkhliutov@ihna.ru

Speech recognition based on EEG and MEG data is the first step in the development of BCI and AI systems for their further use in inner speech decoding. Great advances in this direction have been made using ECoG and stereo-EEG. At the same time, there are few works on this topic on the analysis of data obtained by non-invasive methods of recording brain activity. Our approach is based on the evaluation of connections in the space of sensors with the identification of a pattern of MEG connectivity specific for a given segment of speech. We tested our method on 7 subjects. In all cases, our processing pipeline was quite reliable and worked either without recognition errors or with a small number of errors. After "training", the algorithm is able to recognise a fragment of oral speech with a single presentation. For recognition, we used segments of the MEG recording 50–1200 ms from the beginning of the sound of the word. For high-quality recognition, a segment of at least 600 ms was required. Intervals longer than 1200 ms worsened the recognition quality. Bandpass filtering of the MEG showed that the quality of recognition is equally effective in the entire frequency range. Some decrease in the level of recognition is observed only in the range of 9–14 Hz.

Keywords: speech decoding, sensor space connectivity, MEG, EEG, BCI, AI, theta-rhythm, alpha-rhythm, gamma-rhythm