Федеральное государственное автономное образовательное учреждение
высшего образования «Национальный исследовательский университет
«Высшая школа экономики»

На правах рукописи

Петросян Артур Тигранович

Современные методы машинного обучения в задачах интерпретации электрической активности головного мозга

РЕЗЮМЕ

диссертации на соискание ученой степени
кандидата компьютерных наук

Научный руководитель:
профессор, PhD
Осадчий Алексей Евгеньевич

Москва - 2022

Диссертационная работа выполнена в федеральном государственном автономном образовательном учреждении высшего образования «Национальный исследовательский университет «Высшая школа экономики».

Научный руководитель: Осадчий Алексей Евгеньевич, PhD, директор Центра биоэлектрических интерфейсов Института когнитивных нейронаук федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет «Высшая школа экономики».

Содержание

1 Введение

1.1 Объект исследования

Интерфейсы мозг-компьютер (ИМК) напрямую связывают нервную систему с внешними устройствами [51] или другим мозгом [41]. Несмотря на то, что существует множество применений [34], клинически значимые интерфейсы мозг-компьютер представляют первостепенный интерес, поскольку они обещают реабилитировать пациентов с сенсорными, моторными и когнитивными нарушениями [53, 31].

Интерфейсы мозг-компьютер могут работать с различными сигналами, отражающими электрическую активность нейронов головного мозга [44, 27], такими как, например, электроэнцефалографические (ЭЭГ) потенциалы, измеряемые с помощью электродов, размещенных на поверхности головы [49], или сигналы, регистрируемые инвазивно с помощью внутрикорковых электродов, проникающих в кору головного мозга [40] или размещенных на кортикальной поверхности [48]. В целом, методы регистрации мозговой активности можно разделить на инвазивные и неинвазивные. В первом случае предполагается медицинская процедура имплантации электродов на поверхность коры головного мозга (субдуральная или эпидуральная) и последующего подсчета и интерпретации сигналов активности нейронной популяции. На данный момент, интерфейсы, которые регистрируют мозговую активность неинвазивным способом, не обеспечивают необходимую ширину информационного канала. Объем информации, содержащейся в инвазивно записанных сигналах, значительно перевешивает сложности и технические проблемы, связанные с этой технологией.

Перспективным и минимально инвазивным способом прямого доступа к активности коры головного мозга является использование стерео-ЭЭГ электродов (sEEG), вставляемых стереотаксическим методом через спиральное сверло или отверстие, выполненное в черепе. Последние достижения в методах имплантации, включая использование 3D-ангиографии головного мозга, МРТ и роботизированной хирургии, помогают свести риски такой имплантации практически к нулю, что делает sEEG идеальным компромиссом для приложений ИМК [10]. Полоски ЭКоГ электродов — еще один метод достижения прямого электрического контакта с кортикальной тканью с минимальным дискомфортом для пациента [22].

Одной из важных составляющих технологий нейроинтерфейсов является использование продвинутых и усовершенствованных методов машинного обучения. Из множества имеющихся подходов можно использовать как классические модели, так и методы глубокого обучения. Применение глубинных нейронных сетей в ряде математических и медицинских задач показывает хорошие результаты по сравнению с другими методами, что обуславливает попытку опробовать эти технологии в предсказательных задачах по декодированию сигналов активности мозга [50, 46]. Тем не менее, одна из проблем при декодировании сигналов мозга с помощью алгоритмов глубинного обучения связана с низкой интерпретируемостью получаемых решающих правил, что зачастую приводит к невозможности провести цензурирование получаемых решений. Это необходимо для того, чтобы, например, исключить участие электрических коррелятов нервномышечной активности в автоматическом формировании информативных признаков.

Алгоритмы, используемые для извлечения соответствующих нейронных модуляций, являются ключевым компонентом ИМК систем. Чаще всего они реализуют этапы формирования сигнала, извлечения признаков и декодирования. Современное машинное обучение предписывает выполнение двух последних шагов одновременно с использованием глубоких нейронных сетей (ГНС). [46]. ГНС автоматически находят значимые признаки в контексте задач регрессии или классификации. Корректная интерпретация вычислений, выполняемых ГНС, позволит удостовериться, что декодирование происходит непосредственно из активности мозга. Для обеспечения физиологически значимой интерпретации структура ГНС должна отвечать определенным требованиям и иметь в основе своей доменные знания, которые в данном случае предписывают использование пространственно-распределенной ритмической активности мозга [55] в качестве информационного субстрата. Интерпретируемость моделей в соответствии с традициями Explainable AI может принести пользу автоматизированному процессу обнаружения знаний [3].

Исходя из описанного выше, методы машинного обучения и, в частности, методы глубинного обучения, применяемые в задачах декодирования сигналов головного мозга, а также их интерпретация и построение интерпретируемых архитектур являются основным **объектом исследования**.

1.2 Цели и задачи исследования

Из всего вышесказанного становится очевидно, что применение методов машинного машинного обучения к задачам декодирования информации из сигналов активности головного мозга и развитие этих подходов с целью обеспечения интерпретируемости получаемых решающих правил являются **актуальными задачами**, которые напрямую влияют на практическую применимость ИМК. Искомая интерпретируемость дает гарантии надежности получаемых результатов и открывает новые возможности для изучения принципов работы головного мозга. Соответственно, основной **целью исследования** является разработка доменно-информированных архитектур нейронных сетей в сочетании с построением алгоритмов интерпретации соответствующих весовых коэффициентов и применение данного аппарата к задачам декодирования нейрональной активности в в идеомоторных и речевых нейроинтерфейсах. Диссертационное исследование выполнялось на базе Центра биоэлектрических интерфейсов НИУ ВШЭ, в котором ведутся работы по созданию инвазивных нейроинтерфейсов для замещения моторной и речевой функции. Были сформулированы следующие **задачи исследования**:

1. Разработать архитектуру компактной нейронной сети, согласованную с современными научными данными о происхождении электрофизиологической активности, механизме ее распространения в тканях и физических принципах ее регистрации.

2. Провести сравнительный анализ качества декодирования из ЭКоГ и стерео-ЭЭГ данных кинематики пальца и параметров артикуляционного тракта, достижимых при помощи предложенной компактной нейронной сети и других конкурирующих решений.

3. Разработать методы интерпретации весовых коэффициентов в предложенной архитектуре нейронной сети с целью выявления геометрических характеристик ключевых популяций нейронов и динамических свойств их активности.

4. Реализовать декодирование кинематики движения рук в реальном времени.

5. Реализовать декодирование речи на основе минимального числа пространственно-сегрегированных электродов.

1.3 Основные идеи, результаты и выводы диссертации

Мы предложили и всесторонне исследовали компактную архитектуру на основе сверточной сети для адаптивного декодирования моторных и речевых явлений из ЭКоГ и стерео-ЭЭГ данных. Также, мы предложили новый теоретически обоснованный подход к интерпретации пространственных и временных весов в нашей и подобных архитектурах, сочетающих адаптацию как в пространстве, так и во времени. Полученные пространственные и частотные паттерны, характеризующие популяции нейронов, имеющие решающее значение для конкретной задачи декодирования, подлежат дальнейшему анализу при помощи электромагнитных и динамических моделей с целью охарактеризовать локализацию и параметры активности ключевых нейронных популяций.

Сначала мы протестировали наше решение с помощью реалистичного моделирования методом Монте-Карло. Затем, применительно к данным ЭКоГ из набора данных Berlin BCI Competition IV, наша архитектура работала сравнимо с победителями конкурса, не требуя при этом никакой ручной предобработки данных. Используя предложенный подход к интерпретации весов, мы смогли раскрыть пространственные и спектральные паттерны нейронных процессов, лежащие в основе успешного декодирования кинематики пальцев из набора данных ЭКоГ, записанных в Центре биоэлектрических интерфейсов. Наконец, мы применили метод к анализу 32-канального набора данных воображаемых движений ЭЭГ и увидели физиологически правдоподобные пространственные и частотные паттерны ключевых популяций, характерные для задачи моторного воображения. Также, мы применили нашу архитектуры в реальном времени на реальном пациенте и добились высокого качества декодирования кинематики пальцев пациента исключительно из данных активности мозга. Соответствующие детали описаны в работах [12, 13].

Далее, мы дополнили нашу архитектуру LSTM слоем, применили её к задаче декодирования речи из инвазивных ЭКоГ и стерео-ЭЭГ данных. Для этого мы собрали 60 минут данных (из двух сеансов) для каждого из двух пациентов, которым

были имплантированы инвазивные электроды. Затем мы использовали только контакты, относящиеся к одному стержню стерео-EEG или одной ЭКоГ-полоске, чтобы декодировать нейронную активность в 26 слов и один класс тишины. Интерпретация весов сети дала физиологически правдоподобный результат, который совпал с результатами стимуляционного картирования.

Мы достигли в среднем 55% точности, используя только 6 каналов данных, записанных с помощью одного минимально инвазивного электрода sEEG у первого пациента, и точность 70%, используя только 8 каналов данных, записанных для одной полоски ЭКоГ, у второго пациента в классификации 26+1 произносимых слов. Наша компактная архитектура не требовала использования предварительно отобранных признаков, быстро обучалась и приводила к стабильному, интерпретируемому и физиологически значимому решающему правилу. Пространственные характеристики основных популяций нейронов подтверждают результаты картирования активной и пассивной речи и демонстрируют обратную пространственно-частотную зависимость, характерную для нейронной активности. При сравнении с другими архитектурами наше компактное решение обеспечивало более высокую точность классификации, чем алгоритмы, которые недавно упоминались в литературе по нейронному декодированию речи, и использовало во много раз меньшее число минимально-инвазивных электродов и обучалось на компактном объеме данных.

Наше исследование представляет собой первый шаг к экологичным инвазивным речевым протезам и демонстрирует принципиальную возможность их создания на основе минимально-инвазивной технологии регистрации активности головного мозга. Подробности данного исследования описаны в работе [2].

1.4 Теоретическая и практическая значимость приведенных результатов исследований

С точки зрения теории, мы:

- Впервые обосновали архитектуру нейронной сети, исходя из общепринятой в электрофизиологии модели наблюдения электрической активности мозга при помощи распределенного набора электродов.

- Впервые предложили теоретически-обоснованную методику интерпретации весов компактной нейронной сети с факторизованной пространственно-временной обработкой и провели необходимое моделирование для демонстрации работоспособности предложенной методики.

- Продемонстрировали физиологичность получаемых пространственных и частотных паттернов, характеризующих ключевые нейронные популяции. Полученная информация полностью совпала с результатами активного исследования коры головного мозга пациентов в целях поиска речевой коры. В моторной задаче, соматотопия, наблюдаемая в пространственных паттернах, полностью соответствует устоявшемуся представлению об организации моторной коры.

С практической точки зрения, мы:

- Реализовали прототип инвазивного моторного нейроинтерфейса в реальном времени.

- Предложили архитектуру и методику интерпретации весовых коэффициентов, которые могут быть использованы для построения классификаторов в нейрофизиологических исследованиях. Интерпретация весовых коэффициентов таких классификаторов позволяет добывать новые знания об изучаемых нейрофизиологических процессах.

- Реализовали и апробировали систему декодирования речи из ЭКоГ данных. Наш алгоритм работал в каузальном режиме, то есть использовал данные из прошлого по отношению к моменту времени декодирования. Это позволяет нам надеяться на успешное перенесение достигнутого качества работы нашего декодера у реального пациента с нарушениями речевой функции.

- Мы исследовали возможность работы нашего речевого интерфейса в асинхронном режиме, что имеет большое практическое значение при трансляции нашего решения в клиническую практику.

1.5 Вклад автора в проведенное исследование

Автор этого исследования является разработчиком предлагаемой методики и архитектуры нейронной сети в применении к анализу модельных и реальных данных. Разработанный подход к интерпретации весов широкого семейства архитектур был детально исследован автором в режиме Монте-Карло моделирования. Автором были получены все результаты, касающиеся точности работы предлагаемых алгоритмов в применении к реальным данным. Результаты этой работы описаны в двух статьях, опубликованных в международных журналах повышенного уровня, и в трех статьях по результатам конференций. Во всех этих работах автор является первым и основным автором.

1.6 Публикации и апробация работы

1.6.1 Публикации повышенного уровня

- **Petrosyan A. et al.** Decoding and interpreting cortical signals with a compact convolutional neural network //**Journal of Neural Engineering (Q1)**. – 2021. – Т. 18. – №. 2. – С. 026019 [7].

- **Petrosyan A. et al.** Speech Decoding From A Small Set Of Spatially Segregated Minimally Invasive Intracranial EEG Electrodes With A Compact And Interpretable Neural Network //**Journal of Neural Engineering (Q1)**. . – 2022. – Т.. – №. . – С. [2].

1.6.2 Публикации обычного уровня

- **Petrosyan A.**, Lebedev M., Ossadtchi A. Linear Systems Theoretic Approach to Interpretation of Spatial and Temporal Weights in Compact CNNs: Monte-Carlo Study //Biologically Inspired Cognitive Architectures Meeting (Q4). – Springer, Cham, 2020. – С. 365-370 [13].

- **Petrosyan A.**, Lebedev M., Ossadtchi A. Decoding neural signals with a compact and interpretable convolutional neural network //International Conference on Neuroinformatics (Q4). – Springer, Cham, 2020. – С. 420-428 [12].

- **Arthur Petrosyan**, Alexey Voskoboinikov, Alexei Ossadtchi, Compact and interpretable architecture for speech decoding from stereotactic EEG // 2021 Third International Conference Neurotechnologies and Neurointerfaces – IEEE, 2021. – С. 79-82 [5].

1.6.3 Остальные публикации

- **Petrosyan A. et al.** Compact and Interpretable Architecture for Speech Decoding From iEEG //International Journal of Psychophysiology. – 2021. – Т. 168. – С. S195 [6].
- Volkova Ksenia, **Arthur Petrosyan**, Dubyshkin Ignatii, Ossadtchi Alexei, «decoding movement time-course from ecog using deep learning and implications for bidirectional brain-computer interfacing» [30].

1.6.4 Доклады на конференциях и семинарах

- 2020 Annual International Conference on Brain-Inspired Cognitive Architectures for Artificial Intelligence (BICA*AI 2020) “Linear systems theoretic approach to interpretation of spatial and temporal weights incompact CNNs: Monte-Carlo study” (2020).
- XXII International Conference "Neuroinformatics-2020 «Decoding neural signals with a compact and interpretable convolutional neural network» (2020).
- BCI Samara - «Decoding neural signals with a compact and interpretable convolutional neural network» (2020).
- Report at the forum «Center for Bioelectric Interfaces» (2020).
- BCI Samara - «Compact and interpretable architecture for speech decoding from sEEG» (2021).
- 20th World Congress of Psychophysiology - “Compact and interpretable architecture for speech decoding from sEEG” (2021).
- The Third International Conference «Neurotechnologies and Neurointerfaces» - “Compact and interpretable architecture for speech decoding from sEEG” (2021).

2 Содержание работы

2.1 Архитектура компактной нейронной сети, отражающая современные научные представления о происхождении нейроэлектрофизиологической активности

В данном разделе приведены основные идеи статьи [13].

Вклад автора: разработана архитектура нейронной сети, разработан метод ее интерпретации, реализованы компьютерные симуляции (включая симуляции Монте-Карло).

2.1.1 Феноменологическая модель

Рисунок 1 иллюстрирует возможную взаимосвязь между моторным поведением (движениями рук), активностью мозга и записями ЭКоГ. Активность, $\mathbf{s}[n] = [s_1[n],\ldots,s_I[n]]^T \in \mathbb{R}^I$, из набора I нейронных популяции, $G_1 - G_I$, участвующих в управлении движением, преобразуются в траекторию движения, $z[n]$, посредством нелинейного преобразования H: $z[n] = H(\mathbf{e}[n])$ где $\mathbf{e}[n] = [e_1[n],\ldots,e_I[n]]^T$ - вектор огибающих $\mathbf{s}[n]$. Активность другого набора J популяций $A_1 - A_J$ не связана с движением. Записи этого действия с набором датчиков L в момент времени n представлены вектором сигналов датчиков $L \times 1$, $\mathbf{x}[n] \in \mathbb{R}^L$. В каждый момент времени n этот вектор может быть смоделирован как линейная смесь сигналов, полученных в результате применения матриц прямой модели $\mathbf{G} = [\mathbf{g}_1[n],\ldots,\mathbf{g}_I[n]] \in \mathbb{R}^{L\times I}$ и $\mathbf{A} = [\mathbf{a}_1[n],\ldots,\mathbf{a}_J[n]] \in \mathbb{R}^{L\times J}$ к столбцу вектора активности источников, связанных с задачей, в момент времени n, $\mathbf{s}[n] = [s_1[n],\ldots,s_I[n]]^T$, и несвязанные с задачей источники, $\mathbf{f}[n] = [f_1[n],\ldots,f_J[n]]^T$, соответственно:

$$\mathbf{x}[n] = \mathbf{G}\mathbf{s}[n] + \mathbf{A}\mathbf{f}[n] = \sum_{i=1}^{I} \mathbf{g}_i s_i[n] + \sum_{j=1}^{J} \mathbf{a}_j f_j[n] = \sum_{i=1}^{I} \mathbf{g}_i s_i[n] + \eta[n]. \qquad (1)$$

Векторы столбцов $\mathbf{g}_i,\ i = 1,\ldots,I$ и $\mathbf{a}_j,\ j = 1,\ldots,J$ - это топографии источников, связанных с задачей и не связанных с ней соответсвенно. Мы ссылаемся на зашумленный, не связанный с задачей компонент записи как $\eta[n] = \sum_{j=1}^{J} \mathbf{a}_j f_j[n] \in \mathbb{R}^L$. Аналогичная генеративная модель была недавно описана в [14].

Учитывая линейную генеративную модель электрофизиологических данных, обратное отображение, используемое для получения активности источников из сигналов датчиков, так же обычно ищется в линейной форме: $\hat{\mathbf{s}}[n] = \mathbf{W}^T\mathbf{X}[n]$, где столбцы $\mathbf{W}$ образуют пространственный фильтр, который противодействует эффекту объемной проводимости и уменьшает вклад зашумленных, не связанных с задачей источников.

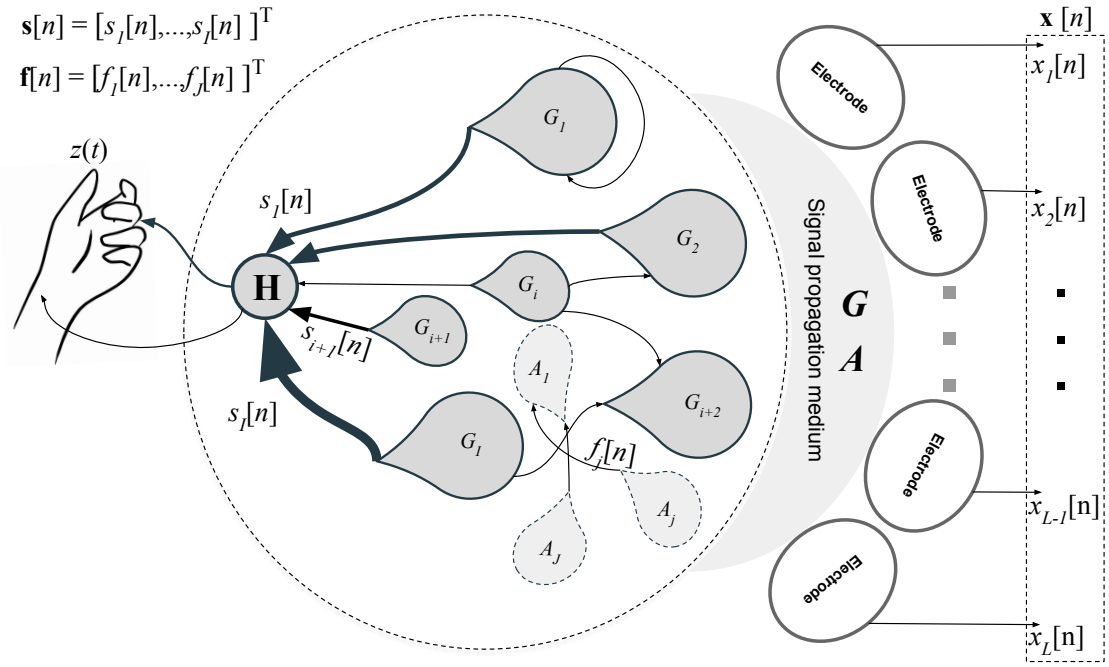

Рис. 1: Феноменологическая модель

Нейронные корреляты моторного планирования и исполнения были тщательно изучены [60]. В области кортикального ритма альфа- и бета-компоненты сенсомоторного ритма десинхронизируются непосредственно перед выполнением движения и восстанавливаются со значительным превышением после завершения двигательного акта [35]. Величина этих модуляций коррелирует со способностью человека контролировать ИМК двигательных образов [36]. Кроме того, частота бета-всплесков в первичной соматосенсорной коре обратно коррелирует со способностью обнаруживать тактильные стимулы, а также влияет на другие двигательные функции. Внутричерепные записи, такие как ЭКоГ, позволяют надежно измерять активность более быстрого гамма-диапазона, которая во времени и пространстве специфична

для моделей движения [20] и, как полагают, сопровождает контроль движений и их выполнение. В целом, основываясь на очень солидном объеме исследований, ритмические компоненты источников мозга, $\mathbf{s}[n]$, по-видимому, полезны для реализаций BCI. Учитывая линейность генеративной модели (1), эти ритмические сигналы, отражающие активность определенных популяций нейронов, могут быть вычислены как линейные комбинации узкополосных отфильтрованных сенсорных данных $\mathbf{x}[n]$.

Самый простой подход к извлечению кинематики, $z[n]$, из записей мозга, $\mathbf{x}[n]$, заключается в использовании одновременно записанных данных и непосредственном изучении отображения $z[n] = \mathcal{H}(\mathbf{x}[n])$. Чтобы практически реализовать его, необходимо параметрически описать это отображение. Длля этой цели мы использовали специфическую нейросетевую архитектуру. Архитектура была построена в тесном соответствии с уравнением наблюдения (1) и нейрофизиологическим описанием наблюдаемых явлений, проиллюстрированным на рисунке 1, что улучшило способность интерпретировать результаты.

2.1.2 Архитектура нейронной сети

Компактная и адаптируемая архитектура (ED-net), которую мы использовали здесь, показана на рисунке 2. Данная архитектура состоит из M ветвей. Каждая ветвь представляет собой адаптивный детектор огибающей со своей собственной парой временных фильтров, которым предшествует специфичный для ветви пространственный фильтр. Наш детектор огибающей аппроксимирует извлеченную огибающую как абсолютное значение аналитического сигнала, вычисленное с использованием преобразования Гильберта для входного сигнала. Используемый нами процесс обработки имитирует процесс аналогового детекторного приемника. Он использовался в других подобных компактных архитектурах сверточных нейронных сетей, которые используют раздельную обработку пространственных и временных измерений [28, 21]. Каждая ветвь нашей сети способна извлекать мгновенную мощность входного сигнала и адаптироваться к конкретной популяции нейронов и полосе частот путем соответствующей настройки пространственных и временных весов фильтров.

Как показано на схеме 2, детектор огибающей может быть реализован с использованием современных примитивов ГНС, а именно пары сверточных операций,

которые выполняют полосовую фильтрацию и фильтрацию нижних частот с одним коэффициентом нелинейности ReLU(-1) между ними, что соответствует вычислению абсолютного значения выходного сигнала первого 1-D сверточного слоя. Этот шаг выделяет сигнал, за которым следует фильтр нижних частот, который сглаживает выходной сигнал $r_m[n]$ для получения аппроксимации огибающей $e_m[n]$. Обратите внимание, что ReLU(a) теперь является стандартной нелинейностью, используемой в современных нейронных сетях и определяемой как ReLU(x,a) = $\{x, x \geq 0; ax, x < 0\}$. Чтобы нелинейность являлась явным выделением мощности сигнала, мы использовали необучаемый слой batch norm. Таким образом, мы можем использовать возможности инструментов оптимизации, реализованных в рамках подхода глубокого обучения, для настройки параметров нашей сети, которая использует пространственные фильтры с последующей оценкой огибающей в качестве блока извлечения признаков.

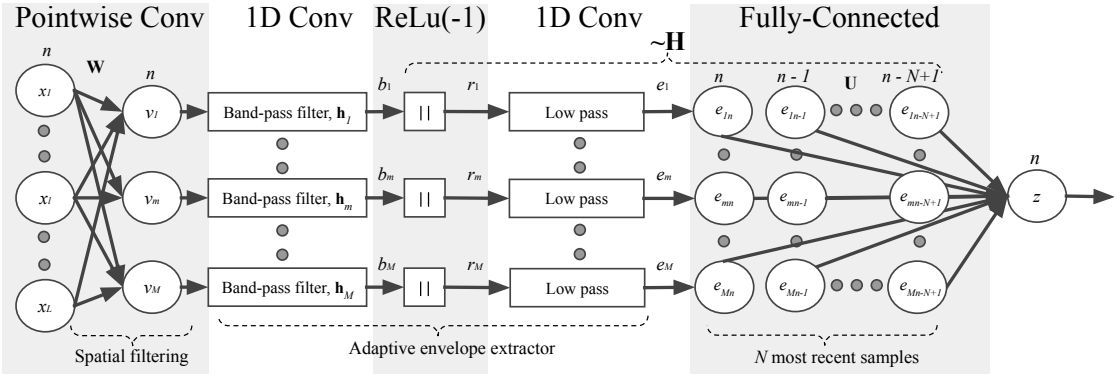

Рис. 2: Архитектура, основанная на компактной сверточной нейронной сети, включает в себя несколько ветвей - адаптивный детектор огибающей, прием пространственно несмешанных входных сигналов и вывод огибающих, чьи N самых последних значений с индексами $n - N + 1, \ldots, n$ объединяются в декодируемой переменной z полносвязанным слоем.

В нашей архитектуре детектор огибающей m-й ветви принимает в качестве входного сигнала пространственно отфильтрованный сигнал датчика $s_m[n]$, вычисленный точечным сверточным слоем. Этот слой предназначен для инвертирования процессов объемной проводимости, представленных матрицами прямой модели $\mathbf{G}$ и $\mathbf{A}$ в нашей

феноменологической модели (рисунок 1). Затем, мы аппроксимировали оператор H как линейную комбинацию запаздывающей мгновенной мощности (огибающей) временных рядов узкополосного источника blue $\mathbf{s}(t) = [s_1(t), s_2(t), \ldots, s_I(t)]$ с коэффициентами из матрицы $\mathbf{U} = \{u_{ml}\},\ m = 1, \ldots, M,\ l = 1, \ldots, N$. Это было сделано с помощью полносвязанного слоя, который смешал огибающие, $e_m[n]$, в единую оценку кинематического параметра $z[n] = \sum_{m=1}^{M} \sum_{l=1}^{N} e_m[n-l]u_{ml} + u_0$, где u_0 моделирует смещение постоянного тока, которое может присутствовать в кинематическом профиле.

2.1.3 Две задачи регрессии и интерпретация весов нейронной сети

Описанная архитектура обрабатывает данные в виде блоков заранее определенной длины из N выборок. Предположим, что длина фрагмента равна длине фильтра в 1D свертке. Рассмотрим фрагмент входных данных из L каналов, наблюдаемых за интервалом в N моментов времени, который может быть представлен с помощью матрицы Теплица $\mathbf{X}[n] = [\mathbf{x}[n], \mathbf{x}[n-1], \ldots, \mathbf{x}[n-N+1]] \in \mathbb{R}^{L \times N}$. Обработка $\mathbf{X}[n]$ первыми двумя слоями, выполняющими пространственную и временную фильтрацию, может быть описана для m-й ветви следующим образом:

$$b_m[n] = \mathbf{w}_m^T \mathbf{X}[n] \mathbf{h}_m, \tag{2}$$

где $\mathbf{w}_m \in \mathbb{R}^L$ - пространственные веса, а $\mathbf{h}_m \in \mathbb{R}^N$ - временные веса ветви m. Нелинейность, $ReLU(-1)$, в сочетании с фильтрацией нижних частот, выполняемой вторым сверточным слоем, извлекает огибающие ритмических сигналов.

Аналитический сигнал сопоставляется один к одному с его огибающей [57], а для исходных вещественных данных мнимая часть аналитического сигнала однозначно вычисляется с помощью преобразования Гильберта. Следовательно, исходный вещественный сигнал однозначно сопоставляется с его огибающей. Наш детектор огибающей вычисляет близкое приближение абсолютного значения аналитического сигнала, и поэтому мы можем утверждать, что $e_m[n]$ однозначно определяется $b_m[n]$. Таким образом, для того, чтобы получить надлежащую огибающую $e_m[n]$, достаточно получить надлежащую $b_m[n]$, что достигается путем корректировки весов пространственной и временной свертки каждой ветви сверточной нейронной сети.

Предположим, что обучение адаптивных детекторов огибающей привело к получению оптимальных весов пространственной и временной свертки, отмеченных звездочками, $\mathbf{w}_m^*$ и $\mathbf{h}_m^*$ соответственно. Дополнительно предположим, что эти оптимальные веса действительно извлекают достоверные сигналы активности населения $b_m^*[n]$, которые однозначно определяют огибающие $e_m^*[n]$, которые, в свою очередь, порождают искомую кинематику $z[n]$ при преобразовании с помощью нелинейного оператора $H(x[n]))$ аппроксимируется полносвязанным слоем нашей сети. Представим, что веса пространственного фильтра неизвестны, но веса временной свертки фиксируются на его оптимальном значении $\mathbf{h}_m^*$. Затем, мы можем найти оптимальные пространственные веса как решение выпуклой задачи оптимизации, сформулированной над пространственным подмножеством параметров:

$$\mathbf{w}_m^* = \text{argmin}_{\mathbf{w}_m}\{\parallel b_m^*[n] - \mathbf{w}_m^T\mathbf{X}[n]\mathbf{h}_m^* \parallel_2^2\} = \text{argmin}_{\mathbf{w}_m}\{\parallel b_m^*(n) - \mathbf{w}_m^T\mathbf{y}_m[n] \parallel_2^2\}, \quad (3)$$

где временные веса фиксированы на их оптимальном значении, $\mathbf{h}_m^*$, а $\mathbf{y}_m[n] = \mathbf{X}[n]\mathbf{h}_m^*$ представляет собой вектор многоканальных данных с временной фильтрацией. Аналогично, когда пространственные веса фиксируются на оптимальном значении $\mathbf{w}_m^*$, временные веса выражаются уравнением:

$$\mathbf{h}_m^* = \text{argmin}_{\mathbf{h}_m}\{\parallel b_m^*[n] - \mathbf{w}_m^{*T}\mathbf{X}[n]\mathbf{h}_m \parallel_2^2\} = \text{argmin}_{\mathbf{h}_m}\{\parallel b_m^*[n] - \mathbf{v}_m^T[n]\mathbf{h}_m \parallel_2^2\}, \quad (4)$$

где $\mathbf{v}_m[n] = [v_m[1], \ldots, v_m[N]]^T = \mathbf{X}^T[n]\mathbf{w}_m^*$ - пространственно отфильтрованный фрагмент входящих данных.

Учитывая прямую модель (1) и задачу регрессии (3) и предполагая взаимную статистическую независимость ритмических потенциалов $s_m[n]$, $m = 1, \ldots, M$, топографии основных популяций нейронов можно найти как [62, 38]:

$$\mathbf{g}_m = \mathbb{E}\{\mathbf{y}_m[n]\mathbf{y}_m^T[n]\}\mathbf{w}_m^* = \mathbf{R}_m^y\mathbf{w}_m^*, \quad (5)$$

где $\mathbf{R}_m^y = \mathbb{E}\{\mathbf{y}_m[n]\mathbf{y}_m^T[n]\}$ - пространственная ковариационная матрица $L \times L$ данных, отфильтрованных во времени, при условии, что временные ряды - это случайные процессы с нулевым средним значением, L - количество входных каналов.

Временные веса могут быть интерпретированы аналогичным образом. Временной

паттерн вычисляется как:

$$\mathbf{q}_m = \mathbb{E}\{\mathbf{v}_m[n]\mathbf{v}_m^T[n]\}\mathbf{h}_m^* = \mathbf{R}_m^v \mathbf{h}_m^*, \tag{6}$$

где $\mathbf{R}_m^v = \mathbb{E}\{\mathbf{v}_m[n]\mathbf{v}_m^T[n]\}$ - ковариационная матрица $N \times N$ пространственно отфильтрованных данных, предполагающая, что временные ряды по каналам - это случайные процессы с нулевым средним значением.

Если мы ослабим предположение о том, что длина блока данных равна длине фильтра временной свертки, мы можем прийти к представлению динамики популяции нейронов в области Фурье в виде паттерна $Q_m(f)$, полученного из спектральной плотности мощности (PSD) $P_{v_m}(f)$ пространственно отфильтрованных данных $v_m[n]$ и преобразования Фурье $H_m(f)$ временных весов вектор $\mathbf{h_m}(f)$:

$$Q_m(f) = P^{v_m}(f)H_m(f). \tag{7}$$

Важным отличием, которое отделяет наш подход к интерпретации весов от методологии, используемой в большинстве статей, использующих нейронные сети с разделяемыми операциями пространственной и временной фильтраций, является то, что наша процедура учитывает тот факт, что формирование пространственного фильтра происходит в контексте, заданном соответствующим временным фильтром, и наоборот. Кроме того, мы впервые ввели понятие паттерна частотной области $Q_m(f)$ активности нейронной популяции.

2.1.4 Реалистичные симуляции

Для интерпретации оптимальных весов временной свертки нам необходимо учитывать спектральные характеристики нейронных записей. Чтобы проиллюстрировать это, мы сначала использовали упрощенное моделирование с одним источником, связанным с задачей, занимающим диапазон частот 50-150 Гц, и одним источником, не связанным с задачей, активным в диапазоне 50-100 Гц, который является поддиапазоном полосы частот сигнала, связанного с задачей. Мы обучили одноканальный ($M = 1$) адаптивный детектор огибающей. Как видно из рисунка 3, профиль Фурье идентифицированных весов временной свертки не может быть использован для оценки спектральной плотности мощности базового сигнала, поскольку он имеет

характерное подавление в диапазоне частот, занимаемом помехой. В то же время, выражение в (7) позволяет нам получить надлежащий паттерн, который хорошо соответствует моделируемому спектральному профилю.

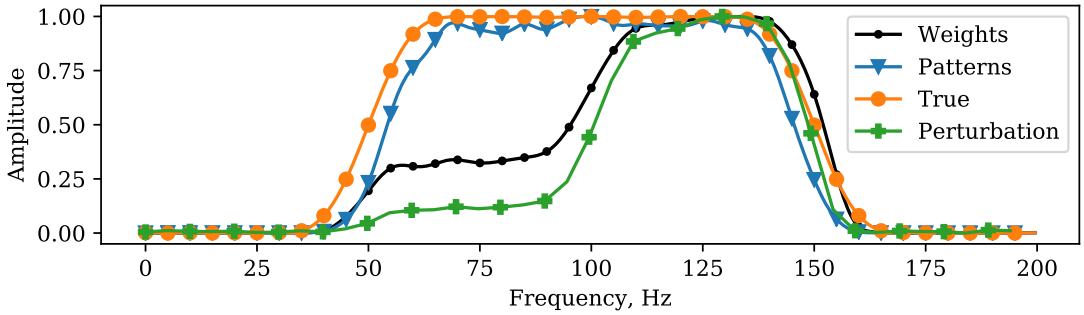

Рис. 3: Три возможных способа интерпретации весов временной свертки. Истинная картина динамической активности, т.е. спектральная плотность мощности (PSD) источника (оранжевый •). Представление в области Фурье весов временной свертки (черный •), метод Болла (зеленый +) и динамический паттерн активности источника, восстановленный с помощью предложенного подхода (синий ▼).

Чтобы изучить корректность предлагаемого подхода, мы выполнили ряд симуляций. Мы следовали процедуре, описанной на рисунке 1, для создания данных. Мы моделировали источники, относящиеся к задаче, с ритмичными LFP $s_i(t)$, занимающими разные диапазоны: 30-80 Гц, 80-120 Гц, 120-170 Гц и 170-220 Гц. Целевая кинематика $z(t)$ моделировалась как линейная комбинация огибающих ритмических LFP с вектором случайных коэффициентов. Использовались также несвязанные с задачей ритмические источники LFP в диапазонах 40-70 Гц, 90-110 Гц, 130-160 Гц и 180-210 Гц. Матрицы $\mathbf{G}$ и $\mathbf{A}$, моделирующие эффекты объемной проводимости в каждом испытании методом Монте-Карло, были сгенерированы случайным образом в соответствии с распределением $\mathcal{N}(0,1)$. Мы создали 20-минутные данные с частотой дискретизации 1000 Гц.

В результате, при отсутствии шума все методы интерпретации справлялись хорошо (4), но при наличии шума, как видно на графике 5, только *Patterns* хорошо согласуется с моделируемой топографией базовых источников. Спектральные характеристики обученных весов временной фильтрации демонстрируют характерные

глубины в полосах, соответствующих активности источников помех. После применения выражения (7) мы получаем спектральные паттерны, которые более точно соответствуют моделируемым и имеют компенсацию глубин.

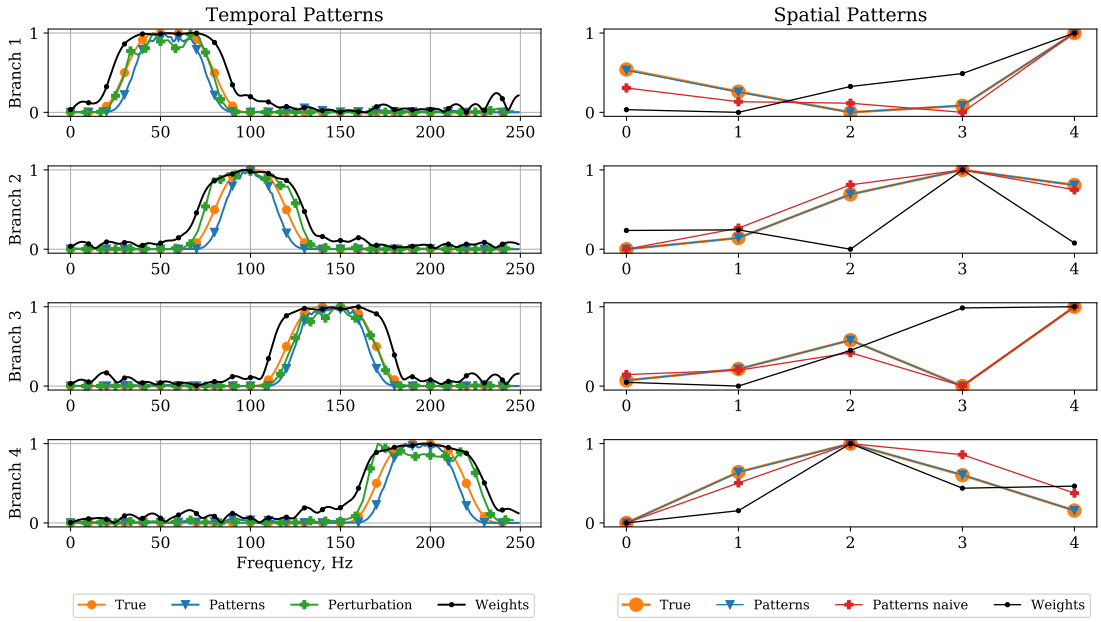

Рис. 4: Временные (слева) и пространственные (справа) паттерны, полученные для случая без шума.

Также, чтобы получить достоверные результаты, мы применили Монте-Карло симуляции с разными параметрами, на которых ясно видно, что предложенный метод дает более верную интерпретацию.

2.2 Декодирование и интерпретация кортикальных сигналов с помощью компактной сверточной нейронной сети

В данном разделе приведено краткое содержание двух статей [7, 12]. Вклад автора: разработана архитектура нейронной сети, разработан метод ее интерпретации, реализованы компьютерные симуляции (включая симуляции Монте-Карло), получены результаты качества декодировнаия и интерпретации на реальных пациентах.

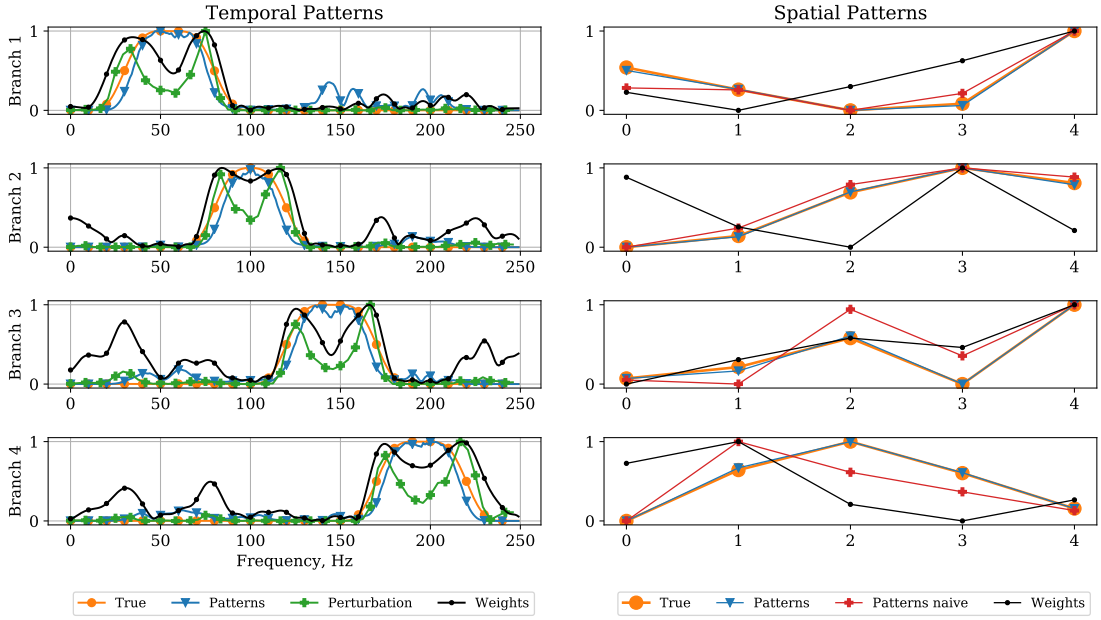

Рис. 5: Временные (слева) и пространственные (справа) паттерны, полученные для зашумленного случая.

2.2.1 Описание существующих методов в задаче декодирования моторных данных

Для обработки данных ЭЭГ и ЭКоГ было разработано несколько полезных и компактных архитектур. Работа некоторых блоков этих архитектур может быть прямолинейно интерпретирована. Таким образом, EEGNet [33] содержит явно очерченные пространственные и временные сверточные блоки. Такая архитектура обеспечивает высокую точность декодирования при минимальном количестве параметров. Однако, из-за связи между перекрестными фильтрами между любыми двумя слоями прямая интерпретация весов затруднена. Некоторое представление о правиле принятия решений можно получить, используя технику deepLIFT [29] в сочетании с анализом паттернов активации скрытых единиц. Schirrmeister и др. [28] описывают две архитектуры: DeepConvNet и его компактную версию ShallowConvNet. Последняя архитектура состоит всего из двух сверточных слоев, которые выполняют временную и пространственную фильтрацию соответственно. Авторы [24] описывают компактную архитектуру сверточной нейронной сети с разделяемыми пространственными и

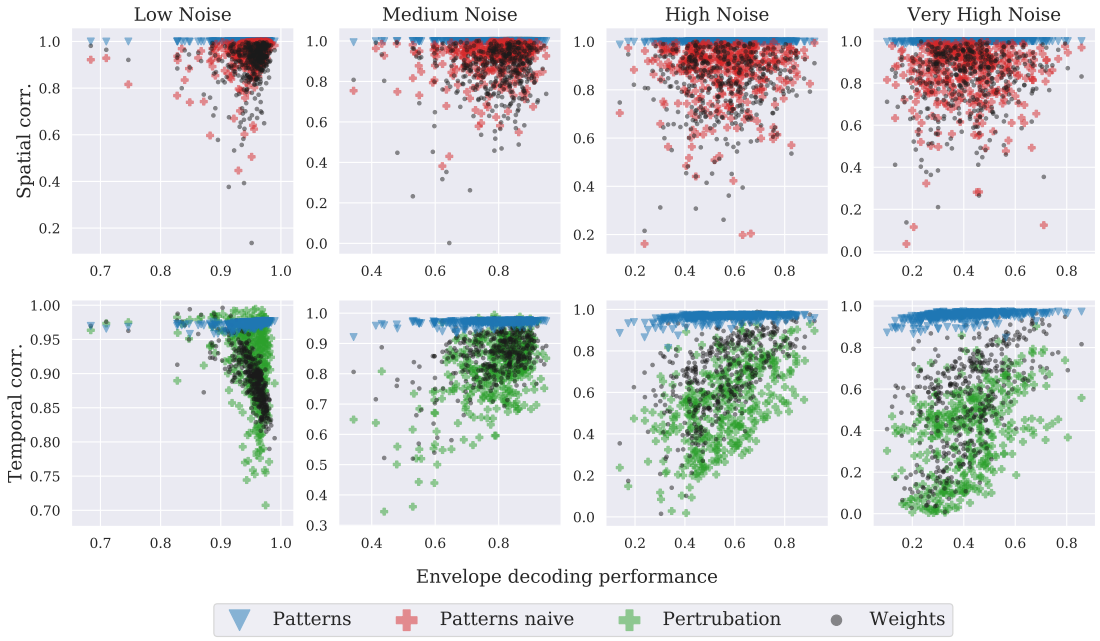

Рис. 6: Симуляции Монте Карло. Координаты точек отражают достигнутую производительность декодирования огибающей (ось x) и коэффициент корреляции с истинным паттерном (ось y) при каждом испытании методом Монте-Карло. Каждая точка определенного цвета соответствует одному испытанию методом Монте-Карло и кодирует метод, используемый для вычисления паттернов. *Weights* - прямая интерпретация весов, *Patterns naive* - интерпретация пространственных паттернов без учета временных фильтров, специфичных для отрасли, *Patterns* - предлагаемый метод.

временными свертками для выполнения классификации ЭЭГ в парадигме SSVEP. Недавнее исследование Зубарева и др. [21] сообщило о двух компактных архитектурах нейронных сетей, LF-CNN и VAR-CNN, которые превзошли другие декодеры данных MEG, включая линейные модели и более сложные нейронные сети, такие как ShallowFBCSP-CNN, EEGNet-8 и VGG19. При этом, LF-CNN и VAR-CNN содержат только одну нелинейность, которая отличает их от большинства других ГНС. Эта особенность делает веса таких архитектур легко интерпретируемыми с помощью хорошо зарекомендовавших себя подходов [62, 58, 38]. Однако, эта методология должна применяться с учетом особенностей, обусловленных разделяемостью этапов

пространственной и временной фильтрации в этих архитектурах.

Мы представляем другую компактную архитектуру, разработанную независимо, но концептуально аналогичную перечисленным выше, и используем ее в качестве тестового стенда для уточнения методов интерпретации весов в семействе архитектур, характеризующихся разделенными адаптивными этапами пространственной и временной обработки. Мы называем этот вид обработки факторизованной обработкой. Мы подчеркиваем, что при интерпретации весов в таких архитектурах мы должны иметь в виду, что эти архитектуры настраивают свои веса не только для адаптации к целевой популяции нейронов, но и для минимизации отвлечения внимания от источников помех как в пространственной, так и в частотной областях.

Решения, реализованные в [56, 52, 45, 47, 42] и элегантно обобщенные в [38], учитывают это адаптивное поведение, но непосредственно применимы только к регрессионным моделям, где к вектору данных (признаков) применяется один вектор весов. Это не относится к рассматриваемому здесь типу моделей, где за фильтрацией в одном домене следует применение фильтра в другом домене. Факторизованная обработка уменьшает количество параметров в архитектуре, но требует специального подхода к интерпретации весов, полученного здесь, чтобы точно оценить пространственные паттерны нейронных источников, лежащих в основе правила принятия решений, выученного архитектурами с факторизованной обработкой. Также, используя аргументы фильтрации Винера, мы впервые расширяем подход к интерпретации весов для анализа весов временного фильтра и показываем, как изученные веса временной свертки в сочетании с пространственно отфильтрованными данными о нейронной активности дают доступ к оценкам спектральной плотности мощности базовых популяций нейронов, имеющих решающее значение для задачи декодирования.

Чтобы проверить работу разработанной нейронной сети и методов ее интерпретации, мы применили их на трех датасетах.

2.2.2 Декодирование движений на Berlin BCI competition IV

Во-первых, чтобы сравнить нашу компактную архитектуру с имеющимися решениями, мы использовали данные, собранные Kubanek et al. и используемые в конкурсе

Berlin BCI competition IV (которые находятся в публичном доступе). В итоге, мы не наблюдали существенных различий между производительностью нашим алгоритмом и выигравшим решением Lian и Bougrain [43] (тест Манна-Уитни, $U = 103.0$, $p = 0.3543$), см. таблицу 1. Но, тем не менее, данные о расположении электродов в данном датасете не разглашаются, поэтому произвести полную интерпретацию не предоставляется возможным.

Subject 1|2|3	Thumb	Index	Middle	Ring	Little
Winner	.58|.51|.69	.71|.37|.46	.14|.24|.58	.53|.47|.58	.29|.35|.63
NET	.54|.50|.71	.70|.36|.48	.20|.22|.50	.58|.40|.52	.25|.23|.61

Таблица 1: Сравнение производительности предложенной архитектуры (NET) и решения-победителя (Winner) на конкурсе Berlin BCI competition IV.

2.2.3 Декодирования кинематики пальцев по ЭКоГ данным

Во-вторых, мы применили предложенные решения к собранным в Лаборатории Биоэлектрических Интерфейсов НИУ ВШЭ данным от двух пациентов в задачи движения пальцев рук, которым имплантировали микросетки ЭКоГ размером 8×8, размещенные поверх сенсомоторной коры головного мозга. На этих данных нам были известны расположения электродов и в результате мы получили интерпретацию, которая согласуется со знаниями с предметной области нейронаук. Пример можно видеть на рисунке 7.

2.2.4 Декодирование классификации движения по ЭЭГ данным

В третьих, в отличие от предыдущих двух наборов данных, которые требовали декодирования непрерывной траектории из инвазивного ЭКоГ, третий набор данных был записан неинвазивно в рамках парадигмы воображаемых движений ЭЭГ. Задача состояла в том, чтобы классифицировать тип выполняемых двигательных действий. Учитывая короткую продолжительность этих данных, компактная сверточная архитектура довольно хорошо решила задачу и дала в среднем 0.83 ROC AUC. На этих данных мы тоже получили интерпретацию, которая согласуется со знаниями с

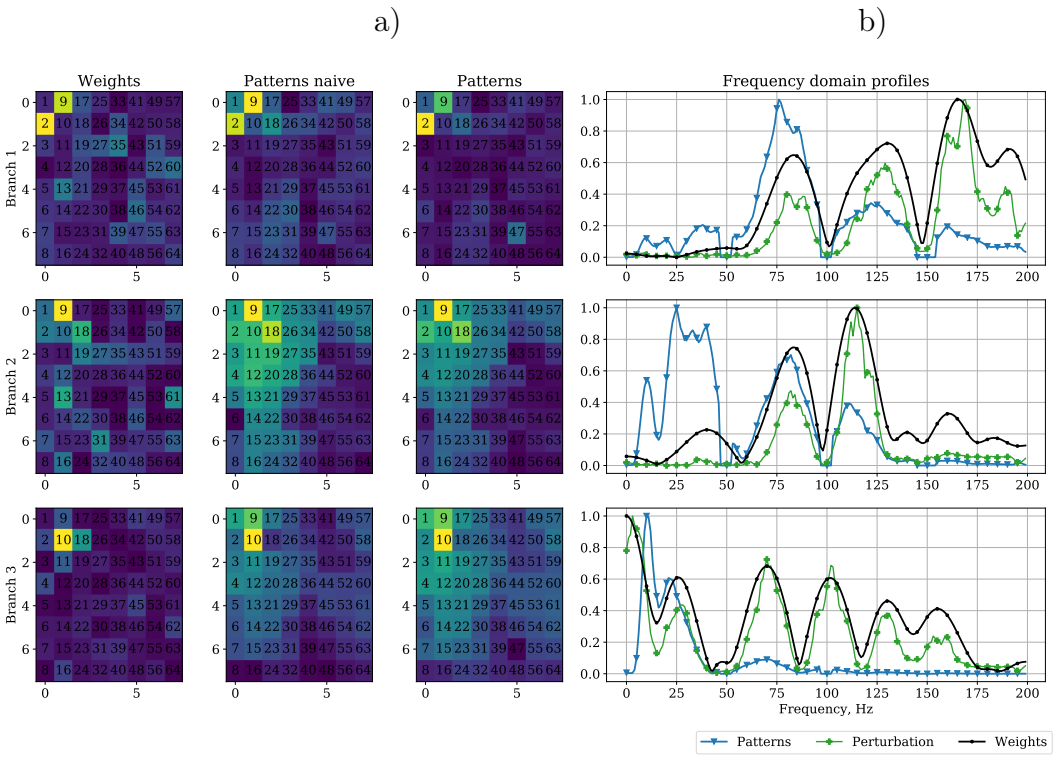

Рис. 7: Интерпретация весов сети для декодера кинематики мизинца у пациента CBI 2 (ЭКоГ). Каждая строка графиков соответствует одной из трех ветвей обученного декодера. а) В крайнем левом столбце показаны веса пространственных фильтров с цветовой кодировкой, следующие два столбца соответствуют наивно и правильно восстановленным пространственным шаблонам. Синий цвет соответствует минимальной абсолютной активации, а желтый - максимальной. б) Интерпретация весов временного фильтра в области Фурье. FFT весов фильтров - (черный •), спектральная плотность мощности (PSD) $Q_m^*[k]$ шаблон базового LFP (синий ▼), полученный в соответствии с уравнением (7). Результаты анализа чувствительности с использованием подхода возмущений показаны в (зеленый +).

предметной области нейронаук. Пример можно видеть на рисунке 8.

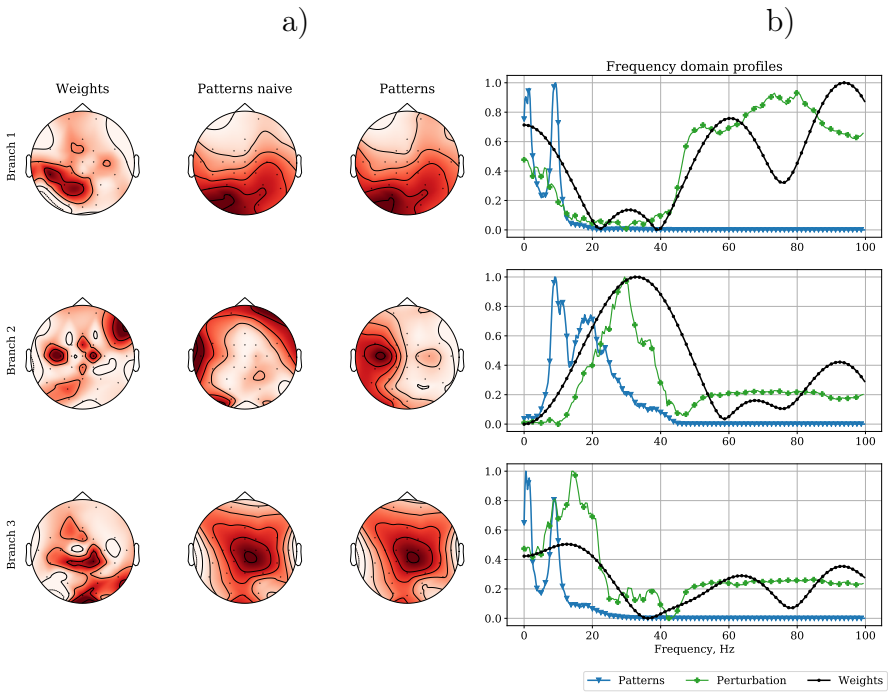

Рис. 8: Описание аналогично Рис. 7

2.3 Декодирование речи с помощью небольшого набора пространственно-разделенных минимально инвазивных внутричерепных электродов ЭЭГ с компактной и интерпретируемой нейронной сетью

В данном разделе приведено краткое содержание двух статей [2, 5]. Вклад автора: разработана архитектура нейронной сети для декодирования речи, получены результаты качества декодирования и интерпретации на реальных пациентах, произведено сравнение качества декодирования при использовании разных внутренних представлений речи, произведен анализ данных на наличие микрофонного эффекта, реализован асинхронный режим работы нейронной сети, произведено анализ взаимной информации между звуком и данными головного мозга.

2.3.1 Введение и существующие методы

Способность общаться жизненно важна для человека, и речь является наиболее естественным каналом для этого. Неспособность говорить резко влияет на качество

жизни. Ряд заболеваний может привести к утрате этой жизненно важной функции, например, детский церебральный паралич и инсульт ствола головного мозга. Кроме того, в ряде случаев после операции по радикальному удалению мозговой ткани у онкологических больных может возникнуть выраженный речевой дефицит. Хотя было предложено несколько технологий для восстановления коммуникативной функции, они по большей части основаны на контролируемом мозгом наборе текста или воображаемом почерке [8] и, по-видимому, применимы только для пациентов с тяжелыми заболеваниями. В то же время, только в Соединенных Штатах миллионы человек страдают от того, что не могут правильно использовать свой речевой аппарат. Значительная часть из них имеет патологию, не поддающуюся лечению голосовым протезом гортани [26] или устройствами "безмолвной речи"[54], и им требуется решение для восстановления речи, управляемое на основе прямого декодирования нейронной активности.

Уже было предпринято несколько успешных попыток восстановления речи на основе BCI, и достигнут значительный прогресс в декодировании фонем [16, 23, 39], отдельных слов [11, 4, 15], непрерывных предложений [11, 4, 15] и даже акустические характеристики [19, 15, 18], за которым следуют алгоритмы восстановления речи с использованием алгоритмов Griffin-Lim или Deep Neural Network (DNN), вдохновленных WaveNet[18].

Эти решения используют широкий спектр подходов к машинному обучению для декодирования речи на основе данных о мозговой активности. Начиная от линейных моделей [16], LDA [9], метрических моделей [19] до глубинны нейронных сетей (ГНС) [11, 4, 15], которые, как правило, не требуют ручной разработки функций и могут быть применены непосредственно к данным, однако иногда работает над набором функций, созданных вручную, в основном из-за высокой гамма-активности. Для задачи декодирования речи было опробовано несколько различных архитектур нейронных сетей: 1) относительно неглубокие, состоящие из нескольких сверточных слоев или слоев LSTM, 2) действительно глубокие архитектуры с начальными блоками [15] или с пропущенными соединениями, использующими остаточную технику обучения [18], а также те, которые заимствованы из приложения компьютерного зрения [25, 37], 3) ансамбли ГНС [4], что делает окончательное решение более надеж-

ным. Интересно, что линейные методы демонстрируют совместимое или, по крайней мере, близкое к ГНС качество декодирования. Более того, последние исследования показали высочайшую точность декодирования, используя всего несколько слоев поверх набора физиологически правдоподобных признаков разработанных под эту задачу [11, 4].

Большинство существующих исследований по декодированию нейронной речи основаны на сильно многоканальных измерениях мозговой активности, реализованных с помощью массивных сеток ЭКоГ [4, 11, 18, 17], охватывающих значительную площадь коры. Эти решения для считывания активности мозга не предназначены для длительного использования, связаны со значительными рисками для пациента [32] и страдают от быстрой потери качества сигнала из-за утечки спинномозговой жидкости под сеткой ЭКоГ, даже если она должным образом перфорирована. sEEG является многообещающей альтернативой, процесс имплантации которой значительно менее травматичен по сравнению с крупными сетками ЭКоГ. Использование sEEG уже изучалось для задачи декодирования речи [9], но описанный декодер снова полагался на большое количество каналов от нескольких стволов sEEG, распределенных по значительной части левой лобной и левой верхней височной долей, что снижает практичность предлагаемого решения. Решение, способное декодировать речь на основе локальной выборки мозговой активности, стало бы важным шагом на пути к созданию устройства для протезирования речи.

Точность нейронного декодирования речи повышается при использовании сжатых представлений, кодирующих кинематические или акустические характеристики речи, в качестве промежуточного представления целевой переменной [18] или для регуляризации [15]. Однако, до сих пор остается неясным, какое из сжатых речевых представлений является оптимальным для декодирования речи по электрофизиологическим данным, и как его следует использовать для достижения наилучшей точности декодирования. В дополнение к прямой практической пользе, ответ на этот вопрос вместе с соответствующей интерпретацией правила принятия решений прольет свет на нейронную основу и корковое представление процессов производства речи.

2.3.2 Архитектура нейронной сети и ее интерпретация

Здесь мы исследуем возможность декодирования отдельных слов из внутричерепной записи мозговой активности, взятой с помощью компактных зондов, имплантация которых не требовала полномасштабной трепанации черепа. В нашем исследовании участвовали два субъекта, которым имплантировали либо стержни sEEG, либо полоски ЭКоГ через компактные отверстия для сверления. Мы декодируем отдельные слова, используя либо 6 каналов данных, записанных с помощью одного вала sEEG, либо 8 каналов, отобранных с помощью одной полосы ЭКоГ. Для декодирования мы использовали нашу компактную и взаимозаменяемую архитектуру сверточной нейронной сети [7], дополненную слоем двунаправленной LSTM[61], чтобы компактно моделировать локальные временные зависимости во внутреннем речевом представлении, которое мы использовали в качестве промежуточной цели декодирования. Также, мы сравнили максимальную точность декодирования слов, достигнутую при использовании различных внутренних представлений. Наш декодер работал каузально, используя только данные из временных интервалов, предшествующих декодированному моменту времени, и поэтому полностью применим в режиме декодирования в реальном времени. В целом, наше исследование является первой попыткой достичь приемлемой точности расшифровки отдельных слов по данным кортикальной активности, отобранным с помощью компактных неинтракортикальных зондов, имплантация которых, вероятно, вызовет минимальный дискомфорт у пациента и может быть выполнена даже под местной анестезией.

Для декодирования нейронных сигналов в LMSC мы использовали компактную и интерпретируемую архитектуру сверточной сети, разработанную ранее для задачи motor BCI [7] и дополнили ее одним двунаправленным слоем LSTM с 30 скрытыми блоками для моделирования временных закономерностей. За слоем LSTM следует полносвязанный слой с выходными нейронами, каждый из которых соответствует одному mel-спектральному коэффициенту, временной профиль которого мы стремимся восстановить по данным нейронной активности, см. Рисунок 9. В поиске оптимума веса, ED не только настраиваются на такой целевой источник, но и в сторону от мешающих источников [38, 7]. Правильная интерпретация весов изученного детектора огибающей позволяет впоследствии обнаружить геометрические и динамические

свойства целевого источника.

После обучения нашей компактной архитектуры декодированию внутреннего представления речи (ВПР) в качестве нашей промежуточной цели, мы использовали сеть 2D-свертки для выполнения дискретной классификации 26 слов и класса silent, используя представления, разработанные на предпоследнем уровне компактной архитектуры, см. Рисунок 9.

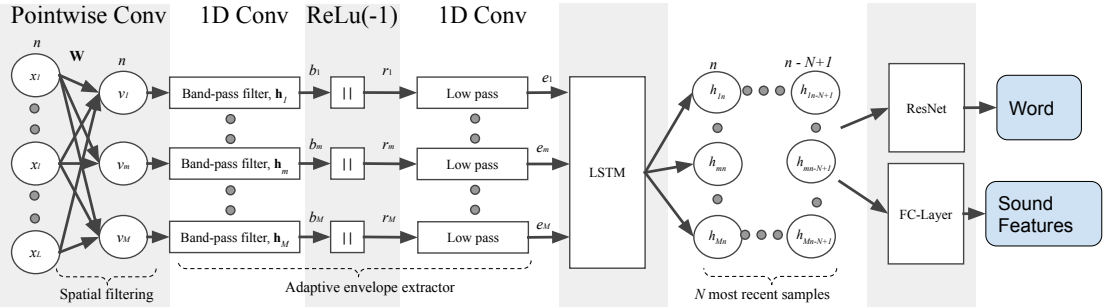

Рис. 9: Архитектура, основанная на [7] и адаптированная для задачи классификации речи. Мы использовали ту же технику детектора огибающей для извлечения надежных и значимых признаков. Затем, мы использовали слой LSTM для учета последовательной структуры mel-спектрограммы и, наконец, декодировали ее с помощью полносвязанного слоя поверх скрытого состояния LSTM (h_{ij} на рисунке). Отдельная 2D сверточная сеть была обучена и использовалась для классификации отдельных слов из активности предварительно обученного таким образом LSTM.

Мы сравнили нашу архитектуру с несколькими другими архитектурами. Мы обнаружили, что из нескольких нейронных сетей только ResNet-18 обеспечивает сопоставимую, хотя и значительно худшее качество при использовании вместо блока ED в нашей архитектуре, см. Рисунок 2. Слой LSTM также, по-видимому, очень полезен для захвата динамики объектов, извлеченных с помощью блоков ED или ResNet, см. Рисунок 10.a. Мы предполагаем, что эта ситуация может быть вызвана адекватным балансом в количестве параметров, подлежащих настройке для сети на основе ED, и объемом данных, доступных для обучения, по сравнению с несколькими другими, более сложными архитектурами.

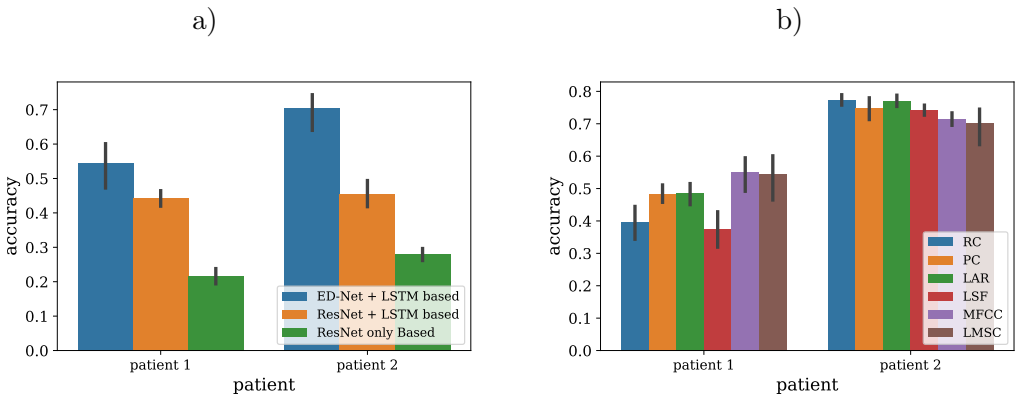

Рис. 10: Сравнительный анализ. a) Сравнение различных моделей нейронной сети b) Сравнение различных возможных промежуточных звуковых представлений, коэффициентов прогнозирования PC - авторегрессии, спектральных частот LSF-линий, коэффициентов RC-отражения, коэффициентов LAR - log-area, коэффициентов спектрограммы LMSC - log-mel, коэффициентов MFCC - mel-frequency cepstral.

2.3.3 Исследование влияния внутреннего представления речи на качество декодирования

В данной работе, мы так же задались вопросом значимости ВПР для задачи декодировния речи. Как видно из архитектуры нейронной сети, она использует дополнительный выход и обучается восстанавливать какое-то из возможных внутренних представлений речи (MELS, LPC, MFCC).

Большинство ВПР основаны на моделировании речевого сигнала, создаваемого последовательностью возбуждения, проходящей через линейный изменяющийся во времени фильтр [59]. Последовательность возбуждения - это поток воздуха в гортани, а фильтр образован элементами артикуляционного тракта (глотка, голосовые складки, язык, губы, зубы), взаимная геометрия которых меняется со временем.

Linear predictive coding (LPC) и кепстральный анализ являются двумя основными способами оценки параметров такого фильтра. Анализ LPC основан на прямой оценке коэффициентов авторегрессионного прогнозирования (PC) a_i с помощью метода Бурга [64]. Однако, сами коэффициенты прогнозирования нестабильны, поскольку их небольшие изменения могут привести к большим вариациям в спектре и, воз-

можно, к нестабильным фильтрам. Для уменьшения такой нестабильности обычно используются следующие несколько эквивалентных представлений.

Коэффициенты отражения (RC) k_i могут быть вычислены наряду с коэффициентами прогнозирования с помощью метода Бурга и представляют собой отношение амплитуд отраженной акустической волны и волны, прошедшей через разрыв.

Другой дескриптор, коэффициенты log-area ratio (LAR), g_i, равны натуральному логарифму отношения площадей смежных секций в трубном эквиваленте голосового тракта без потерь, имеющем ту же передаточную функцию, и могут быть вычислены из коэффициентов отражения как $g_i = \ln\left(\frac{1-k_i}{1+k_i}\right)$.

Линейные спектральные частоты (LSF) - еще один высокоэффективный метод сжатия речевых данных [63], поскольку ошибки в представлении одного коэффициента обычно приводят к спектральному изменению только вокруг этой частоты.

В дальнейшем мы представим результаты наших экспериментов с несколькими ВПР, но наши окончательные результаты точности декодирования основаны на использовании логарифмических спектральных коэффициентов (LMSC).

Результаты на рисунке 10.b. Мы можем видеть, что для первого пациента целевые коэффициенты спектра log-mel (LMSC) приводят к наивысшей точности декодирования слов. Интересно, что в отличие от фактической задачи декодирования ВПР, отображаемой на рисунке 12, разница в точности декодирования слов между различными ВПР, по-видимому, значительно менее выражена, чем различия в качестве декодирования каждого из таких представлений. Тем не менее, для обоих пациентов мы наблюдаем аналогичную картину: PC и LSF дают относительно худшую точность декодирования слов, чем другие ВПР. В этом анализе коэффициенты отражения LPC (RC) обеспечивают лучшую точность декодирования по сравнению с коэффициентами прогнозирования. Это наблюдение соответствует свойствам коэффициентов RC как информационно эквивалентной, но более стабильной версии исходного PC.

2.3.4 Синхронный и асинхронный режим

Традиционно BCI можно использовать в двух различных настройках: синхронной и асинхронной. В синхронной настройке команда должна быть выдана в течение определенного временного интервала. Обычно асинхронный пользователь BCI полу-

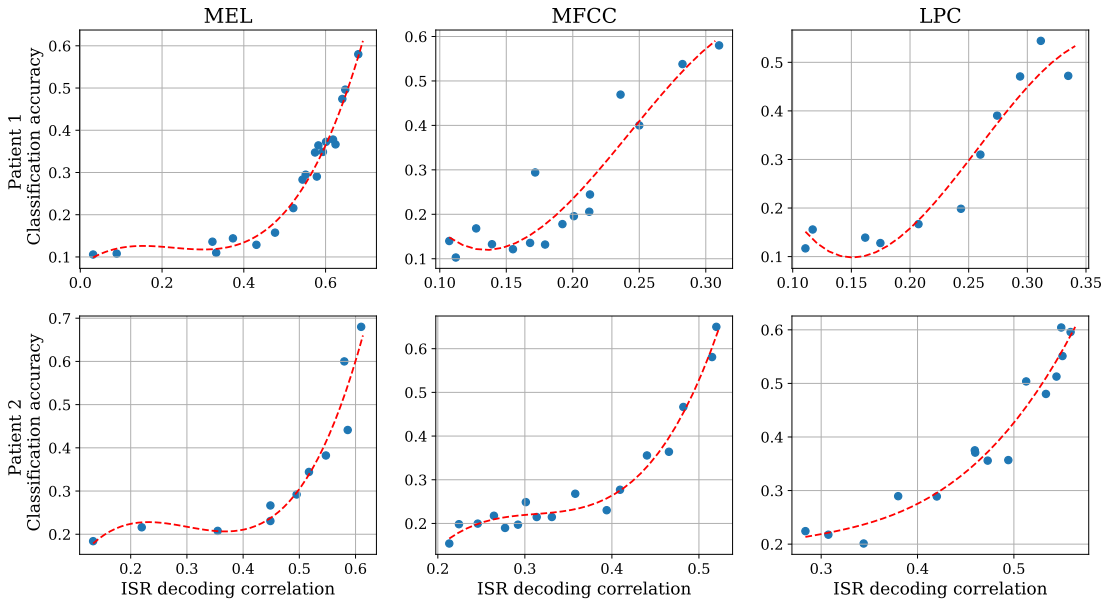

Рис. 11: Зависимость качества декодирования ВПР от точности классификации конечного слова. Красная линия - это оценка тренда третьего порядка, сделанная для визуализации.

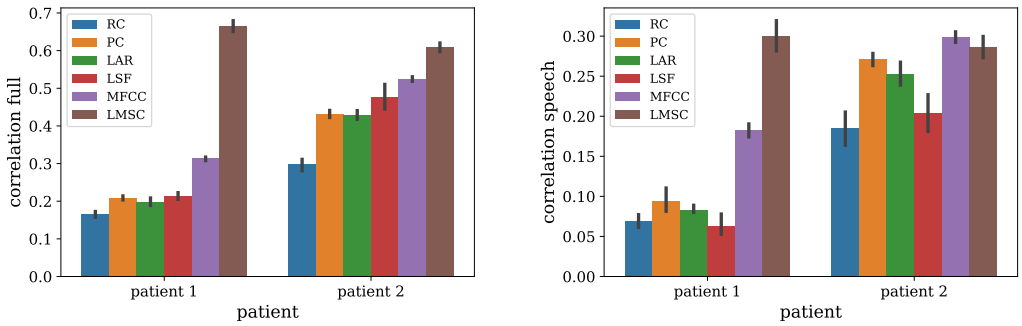

Рис. 12: Сравнение точности декодирования, достигнутой для различных ВПР: PC, LSF, RC, LAR, LMSCs, MFCC. Левая панель соответствует коэффициентам корреляции между фактическими и декодированными временными профилями, вычисленными по всему временному диапазону сегмента тестовых данных. На правой панели коэффициент корреляции вычисляется только за те временные интервалы, в которых присутствовала фактическая речь.

чает запрос в начале такого временного окна и должен выполнить команду (изменить состояние своего мозга) в течение указанного периода времени. Следовательно, алгоритм декодирования знает о конкретном сегменте данных, который необходимо обработать, чтобы извлечь информацию о команде. В асинхронном режиме BCI должен не только расшифровать команду, но и определить тот факт, что команда действительно выдается. Разграничение между синхронным и асинхронным режимами наиболее четко выражено в BCI с дискретными командами, подразумевающими использование категориального декодера.

В BCI, которые декодируют непрерывную переменную, например, кинематику рук, такое разграничение между синхронным и асинхронным режимами менее четкое. Первая часть нашего BCI реализует непрерывный декодер функций внутреннего представления речи (ВПР). Если бы это декодирование оказалось достаточно точным, его можно было бы просто использовать в качестве входных данных для механизма синтеза голоса. Такой сценарий уже был реализован в нескольких отчетах [18, 17], но в этих решениях используется большое количество электродов, что может объяснить лучшее качество декодирования ВПР. В наших условиях мы стремились создать декодер, работающий с небольшим количеством экологически имплантированных электродов, и решили сосредоточиться на декодировании отдельных слов. Сначала мы использовали непрерывно декодируемые ВПР для синхронной классификации 26 дискретных слов и одного состояния молчания. Чтобы реализовать это, мы вырезаем декодированные временные ряды ВПР вокруг произнесения каждого слова и используем их в качестве образцов данных для нашего механизма классификации.

Рисунок 13.b иллюстрирует производительность нашего BCI, работающего в полностью асинхронном режиме, когда декодер работает в течение последовательности перекрывающихся временных окон непрерывно декодируемых ВПР. Для количественной оценки производительности нашего асинхронного речевого декодера, мы использовали кривые точности и полноты, как показано на рисунке 13.a.

Хотя наблюдаемая производительность значительно превышает уровень вероятности, ее еще недостаточно для создания полноценного асинхронного речевого интерфейса, работающего с использованием небольшого количества минимально инвазивных электродов. По нашему мнению и на основе опыта работы с моторны-

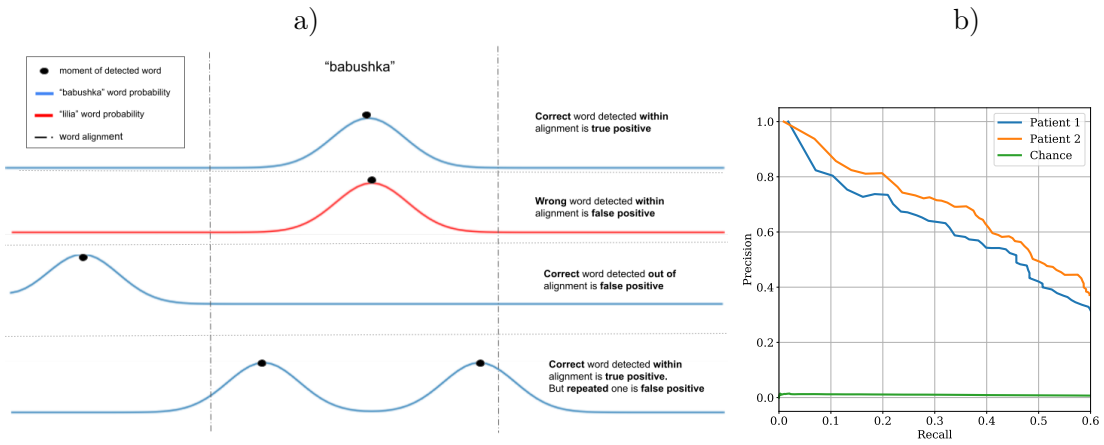

Рис. 13: а) Для каждого $i-$го слова мы вычисляем сглаженные профили вероятности $\tilde{p}_i(t)$ для каждого временного экземпляра t. Затем принимается решение о том, что слово произносится только в моменты времени, соответствующие локальным максимумам $\tilde{p}_i(t)$, которые пересекают порог θ. В случае, если выбранное слово $i-$th совпадает с тем, которое произносится в данный момент, мы отмечаем это событие как истинно положительное (TP). Если после такого обнаружения $\tilde{p}_i(t)$ остается выше порога и демонстрирует еще один локальный максимум, который превышает значения всех других сглаженных профилей вероятности, то i-е слово снова «произносится», но это событие помечается как ложноположительное (FP) даже если t принадлежит временному диапазону, соответствующему фактическому $i-$-му слову. б) Кривые PR для задачи асинхронного декодирования слов. Обратите внимание, что определение точности и полноты немного отличается от обычной бинарной классификации. Мы также показываем кривую PR полученую на уровне шанса.

ми интерфейсами, специальные протоколы для обучения пациента, в том числе с немедленной обратной связью с пользователем [1], вероятно, значительно улучшат точность декодирования в таких системах, что повысит общую осуществимость минимально инвазивных решений для протезирования речи.

3 Заключение

В данной работе выполнены два больших проекта, объединенных общей тематикой, посвященной разработке и применению современных интерпретируемых нейросетевых моделей к анализу и декодированию активности головного мозга. Работа представляет собой законченное исследование, в результате которого был разработан целый ряд программно-алгоритмических средств обработки электрофизиологических сигналов, содержащих в своей основе новый математический аппарат, так же предложенный авторами работы. Полученные решения апробированы в выступлениях на многочисленных конференциях и их научная обоснованность подтверждена рядом публикаций в ведущих международных научных журналах, в том числе в двух публикациях с первым авторством соискателя в Journal of Neural Engineering (Q1 - Scopus, Q2 - WoS). В настоящее время, все разработанные средства и алгоритмы используются в научно-исследовательской деятельности Центра биоэлектрических интерфейсов НИУ ВШЭ.

3.1 Список выносимых на защиту результатов

1. Архитектура компактной нейронной сети, отражающая современные научные представления о происхождении нейроэлектрофизиологической активности, механизме ее распространения в тканях и физических принципах ее регистрации при помощи распределенного набора электродов.

2. Результаты сравнительного анализа качества декодирования из ЭКоГ и стерео-ЭЭГ данных кинематики пальца и параметров артикуляционного тракта, демонстрирующие превосходство предлагаемой архитектуры нейронной сети по сравнению с конкурирующими решениями.

3. Теоретически обоснованная методика интерпретации весов в предложенной архитектуре нейронной сети с целью выявления геометрических характеристик ключевых популяций нейронов и динамических свойств их активности.

4. Результаты анализа зависимости итоговой точности классификации от выбора промежуточного представления речевого сигнала.

5. Реализация декодирования кинематики движения рук в реальном времени.

6. Реализация декодирования речи на основе минимального числа пространственно-сегрегированных электродов.

Список литературы

[1] Miguel Angrick и др. “Towards Closed-Loop Speech Synthesis from Stereotactic EEG: A Unit Selection Approach”. В: *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2022, с. 1296—1300.

[2] Artur Petrosyan и др. “Speech decoding from a small set of spatially segregated minimally invasive intracranial EEG electrodes with a compact and interpretable neural network”. В: *bioRxiv* (2022).

[3] Prashant Gohel, Priyanka Singh и Manoranjan Mohanty. “Explainable AI: current status and future directions”. В: *arXiv preprint arXiv:2107.07045* (2021).

[4] David A Moses и др. “Neuroprosthesis for decoding speech in a paralyzed person with anarthria”. В: *New England Journal of Medicine* 385.3 (2021), с. 217—227.

[5] Artur Petrosyan, Alexey Voskoboynikov и Alexei Ossadtchi. “Compact and interpretable architecture for speech decoding from stereotactic EEG”. В: *2021 Third International Conference Neurotechnologies and Neurointerfaces (CNN)*. IEEE. 2021, с. 79—82.

[6] Artur Petrosyan и др. “Compact and Interpretable Architecture for Speech Decoding From iEEG”. В: *International Journal of Psychophysiology* 168.S (2021), S195.

[7] Artur Petrosyan и др. “Decoding and interpreting cortical signals with a compact convolutional neural network”. В: *Journal of Neural Engineering* 18.2 (2021), с. 026019.

[8] Francis R Willett и др. “High-performance brain-to-text communication via handwriting”. В: *Nature* 593.7858 (2021), с. 249—254.

[9] Miguel Angrick и др. “Real-time Synthesis of Imagined Speech Processes from Minimally Invasive Recordings of Neural Activity”. В: *bioRxiv* (2020).

[10] Christian Herff, Dean J Krusienski и Pieter Kubben. “The potential of stereotactic-EEG for brain-computer interfaces: current progress and future directions”. В: *Frontiers in neuroscience* 14 (2020), с. 123.

[11] Joseph G Makin, David A Moses и Edward F Chang. “Machine translation of cortical activity to text with an encoder–decoder framework”. В: *Nature Neuroscience* 23.4 (2020), с. 575—582.

[12] Artur Petrosyan, Mikhail Lebedev и Alexey Ossadtchi. “Decoding neural signals with a compact and interpretable convolutional neural network”. В: *International Conference on Neuroinformatics*. Springer. 2020, с. 420—428.

[13] Artur Petrosyan, Mikhail Lebedev и Alexey Ossadtchi. “Linear Systems Theoretic Approach to Interpretation of Spatial and Temporal Weights in Compact CNNs: Monte-Carlo Study”. В: *Biologically Inspired Cognitive Architectures Meeting*. Springer. 2020, с. 365—370.

[14] David Sabbagh и др. “Predictive regression modeling with MEG/EEG: from source power to signals and cognitive states”. В: *NeuroImage* 222 (2020), с. 116893.

[15] Pengfei Sun, Gopala K Anumanchipalli и Edward F Chang. “Brain2Char: a deep architecture for decoding text from brain recordings”. В: *Journal of Neural Engineering* 17.6 (2020), с. 066015.

[16] Guy H Wilson и др. “Decoding spoken English from intracortical electrode arrays in dorsal precentral gyrus”. В: *Journal of Neural Engineering* 17.6 (2020), с. 066007.

[17] Hassan Akbari и др. “Towards reconstructing intelligible speech from the human auditory cortex”. В: *Scientific reports* 9.1 (2019), с. 1—12.

[18] Miguel Angrick и др. “Speech synthesis from ECoG using densely connected 3D convolutional neural networks”. В: *Journal of neural engineering* 16.3 (2019), с. 036019.

[19] Christian Herff и др. “Generating natural, intelligible speech from brain activity in motor, premotor, and inferior frontal cortices”. В: *Frontiers in neuroscience* 13 (2019), с. 1267.

[20] Ksenia Volkova и др. “Decoding Movement From Electrocorticographic Activity: A Review”. В: *Frontiers in neuroinformatics* 13 (2019), с. 74. ISSN: 1662-5196. DOI: 10.3389/fninf.2019.00074. URL: https://europepmc.org/articles/PMC6901702.

[21] Ivan Zubarev и др. “Adaptive neural network classifier for decoding MEG signals”. В: *NeuroImage* 197 (2019), с. 425—434.

[22] Abidemi B Ajiboye и Robert F Kirsch. “Invasive Brain–Computer Interfaces for Functional Restoration”. В: *Neuromodulation*. Elsevier, 2018, с. 379—391.

[23] Nick F Ramsey и др. “Decoding spoken phonemes from sensorimotor cortex with high-density ECoG grids”. В: *Neuroimage* 180 (2018), с. 301—311.

[24] Nicholas Waytowich и др. “Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials”. В: *Journal of neural engineering* 15.6 (2018), с. 066031.

[25] Gao Huang и др. “Densely connected convolutional networks”. В: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, с. 4700—4708.

[26] Rachel Kaye, Christopher G Tang и Catherine F Sinclair. “The electrolarynx: voice restoration after total laryngectomy”. В: *Medical Devices (Auckland, NZ)* 10 (2017), с. 133.

[27] Mikhail A Lebedev и Miguel AL Nicolelis. “Brain-machine interfaces: From basic science to neuroprostheses and neurorehabilitation”. В: *Physiological reviews* 97.2 (2017), с. 767—837.

[28] Robin Tibor Schirrmeister и др. “Deep learning with convolutional neural networks for EEG decoding and visualization”. В: *Human brain mapping* 38.11 (2017), с. 5391—5420.

[29] Avanti Shrikumar, Peyton Greenside и Anshul Kundaje. *Learning Important Features Through Propagating Activation Differences*. 2017. arXiv: 1704.02685 [cs.CV].

[30] Осадчий и др. “Интерфейс мозг-компьютер: опыт построения, использования и возможные пути повышения рабочих характеристик”. В: *Журнал высшей нервной деятельности им. ИП Павлова* 67.4 (2017), с. 504—520.

[31] Ujwal Chaudhary, Niels Birbaumer и Ander Ramos-Murguialday. “Brain–computer interfaces for communication and rehabilitation”. В: *Nature Reviews Neurology* 12.9 (2016), с. 513.

[32] Prasanna Jayakar и др. “Diagnostic utility of invasive EEG for epilepsy surgery: Indications, modalities, and techniques”. В: *Epilepsia* 57.11 (2016), с. 1735—1747. DOI: https://doi.org/10.1111/epi.13515. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/epi.13515. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/epi.13515.

[33] Vernon J Lawhern и др. “Eegnet: A compact convolutional network for eeg-based brain-computer interfaces”. В: *arXiv preprint arXiv:1611.08024* (2016).

[34] Sarah N Abdulkader, Ayman Atia и Mostafa-Sami M Mostafa. “Brain computer interfacing: Applications and challenges”. В: *Egyptian Informatics Journal* 16.2 (2015), с. 213—230.

[35] Yaron Meirovitch и др. “Alpha and Beta Band Event-Related Desynchronization Reflects Kinematic Regularities”. В: *Journal of Neuroscience* 35.4 (2015), с. 1627—1637. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.5371-13.2015. eprint: https://www.jneurosci.org/content/35/4/1627.full.pdf. URL: https://www.jneurosci.org/content/35/4/1627.

[36] Johanna Louise Reichert и др. “Resting-state sensorimotor rhythm (SMR) power predicts the ability to up-regulate SMR in an EEG-instrumental conditioning paradigm”. В: *Clinical Neurophysiology* 126.11 (февр. 2015), с. 2068—2077. DOI: 10.1016/j.clinph.2014.09.032.

[37] Christian Szegedy и др. “Going deeper with convolutions”. В: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, с. 1—9.

[38] Stefan Haufe и др. “On the interpretation of weight vectors of linear models in multivariate neuroimaging”. В: *Neuroimage* 87 (2014), с. 96—110.

[39] Emily M Mugler и др. “Direct classification of all American English phonemes using signals from functional speech motor cortex”. В: *Journal of neural engineering* 11.3 (2014), с. 035015.

[40] Mark L Homer и др. “Sensors and decoding for intracortical brain computer interfaces”. В: *Annual review of biomedical engineering* 15 (2013), с. 383—405.

[41] Miguel Pais-Vieira и др. “A Brain-to-Brain Interface for Real-Time Sharing of Sensorimotor Information”. В: *Scientific reports* 3 (февр. 2013), с. 1319. DOI: 10.1038/srep01319.

[42] Felix Bießmann и др. “Improved decoding of neural activity from fMRI signals using non-separable spatiotemporal deconvolutions”. В: *NeuroImage* 61.4 (2012), с. 1031—1042.

[43] Nanying Liang и Laurent Bougrain. “Decoding Finger Flexion from Band-Specific ECoG Signals in Humans”. В: *Frontiers in neuroscience* 6 (июнь 2012), с. 91. DOI: 10.3389/fnins.2012.00091.

[44] Luis Fernando Nicolas-Alonso и Jaime Gomez-Gil. “Brain computer interfaces, a review”. В: *Sensors* 12.2 (2012), с. 1211—1279.

[45] Benjamin Blankertz и др. “Single-trial analysis and classification of ERP components—a tutorial”. В: *NeuroImage* 56.2 (2011), с. 814—825.

[46] Steven Lemm и др. “Introduction to machine learning for brain imaging”. В: *Neuroimage* 56.2 (2011), с. 387—399.

[47] Thomas Naselaris и др. “Encoding and decoding in fMRI”. В: *Neuroimage* 56.2 (2011), с. 400—410.

[48] Gerwin Schalk и Eric C Leuthardt. “Brain-computer interfaces using electrocorticographic signals”. В: *IEEE reviews in biomedical engineering* 4 (2011), с. 140—154.

[49] Sergio Machado и др. “EEG-based brain-computer interfaces: an overview of basic concepts and clinical applications in neurorehabilitation”. В: *Reviews in the Neurosciences* 21.6 (2010), с. 451—468.

[50] Claudio Castellini и Patrick van der Smagt. “Surface EMG in advanced hand prosthetics”. В: *Biological cybernetics* 100.1 (2009), с. 35—47.

[51] Nicholas G Hatsopoulos и John P Donoghue. “The science of neural interface systems”. В: *Annual review of neuroscience* 32 (2009), с. 249—266.

[52] Aapo Hyvärinen, Jarmo Hurri и Patrick O Hoyer. *Natural image statistics: A probabilistic approach to early computational vision.* Т. 39. Springer Science & Business Media, 2009.

[53] Joseph N Mak и Jonathan R Wolpaw. "Clinical applications of brain-computer interfaces: current state and future prospects". В: *IEEE reviews in biomedical engineering* 2 (2009), с. 187—199.

[54] Michael J Fagan и др. "Development of a (silent) speech recognition system for patients following laryngectomy". В: *Medical engineering & physics* 30.4 (2008), с. 419—425.

[55] Gyorgy Buzsaki. *Rhythms of the Brain.* Oxford University Press, 2006.

[56] Lucas C Parra и др. "Recipes for the linear analysis of EEG". В: *Neuroimage* 28.2 (2005), с. 326—341.

[57] Stefan L Hahn. "On the uniqueness of the definition of the amplitude and phase of the analytic signal". В: *Signal Processing* 83.8 (2003), с. 1815—1820.

[58] Lucas Parra и др. "Single-trial detection in EEG and MEG: Keeping it linear". В: *Neurocomputing* 52 (2003), с. 177—183.

[59] Xuedong Huang и др. *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development.* 1st. USA: Prentice Hall PTR, 2001. ISBN: 0130226165.

[60] Daniel Wolpert и Zoubin Ghahramani. "Computational Principles of Movement Neuroscience". В: *Nature neuroscience* 3 Suppl (дек. 2000), с. 1212—7. DOI: 10.1038/81497.

[61] Sepp Hochreiter и Jürgen Schmidhuber. "Long short-term memory". В: *Neural computation* 9.8 (1997), с. 1735—1780.

[62] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory.* Prentice Hall, 1997.

[63] Frank Soong и B Juang. "Line spectrum pair (LSP) and speech data compression". В: *ICASSP'84. IEEE International Conference on Acoustics, Speech, and Signal Processing.* Т. 9. IEEE. 1984, с. 37—40.

[64] L. Marple. “A new autoregressive spectrum analysis algorithm”. B: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 28 (1980), c. 441—454.