

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

Факультет компьютерных наук
Основная образовательная программа
Прикладная математика и информатика

КУРСОВАЯ РАБОТА
ИССЛЕДОВАТЕЛЬСКИЙ ПРОЕКТ НА ТЕМУ
ИССЛЕДОВАНИЕ ПРИМЕНИМОСТИ МЕТОДОВ
МАШИННОГО ОБУЧЕНИЯ К ДЕКОДИРОВАНИЮ РЕЧИ ИЗ
МИНИМАЛЬНО ИНВАЗИВНЫХ ЗАПИСЕЙ
ЭЛЕКТРИЧЕСКОЙ АКТИВНОСТИ ГОЛОВНОГО МОЗГА.

Выполнил студент группы 191, 3 курса,
Сизов Кирилл Игоревич

Руководитель КР:
профессор Осадчий Алексей Евгеньевич

Москва 2022

Аннотация

Нейроинтерфейс или интерфейс «мозг-компьютер» – это технология, которая позволяет считывать электрическую активность головного мозга и преобразовывать ее в команду для внешнего устройства при помощи различных методов анализа данных. Одно из наиболее интересных применений нейроинтерфейсов является декодирование речи, которое имеет большое количество потенциальных применений – от реабилитации пациентов до общения людей.

В данной работе ставится задача восстановления речи при помощи декодирования фонем – коротких участков (10-100мс) речи, соответствующих определенному звуку. Для исследования использовались данные электрической активности мозга, записанные при помощи минимально-инвазивного устройства стерео-ЭЭГ.

Построение фонем осуществлялось при помощи кластеризации звуковых фрагментов. Для преобразования звука в пространство меньшей размерности был применен подход Linear Predictive Coding. В качестве алгоритмов кластеризации были исследованы метод k-средних, модель смеси Гаусса и скрытая марковская модель с объектами гауссовой смеси. Распознавание фонем осуществлялось сверточной нейронной сетью с архитектурой схожей на DenseNet, которая хорошо себя показала в работе по распознаванию отдельных слов.

Ключевые слова: Brain Computer Interface (BCI), stereotactic EEG (sEEG), Linear Predictive Coding (LPC)

Abstract

Neurointerface or brain-computer interface is a technology that allows you to read the electrical activity of the brain and convert it into a command for an external device using various data analysis methods. One of the most interesting applications of neurointerfaces is speech decoding, which has a large number of potential applications – from the rehabilitation of patients to the communication of people.

In this paper, the task of speech restoration is set by decoding phonemes – short sections (10-100ms) of speech corresponding to a certain sound. The current study used data on the electrical activity of the brain recorded using a minimally invasive stereotactic EEG device.

Phonemes were constructed using the clustering of sound fragments. The Linear Predictive Coding approach was applied to transform sound into a smaller dimensional space. K-Means, the Gaussian Mixture Model, and Hidden Markov Model with Gaussian mixture emissions were used as clustering algorithms. Phoneme recognition was carried out by a convolutional neural network with an architecture similar to DenseNet, which proved itself well in the work of recognizing individual words.

Ключевые слова: Brain Computer Interface (BCI), stereotactic EEG (sEEG), Linear Predictive Coding (LPC)

Содержание

1 Введение	5
2 Обзор литературы	6
2.1 Данные	7
2.2 Модель	7
2.3 Результаты	8
3 Данные и предобработка	9
3.1 Предобработка	10
4 Построение фонем	10
4.1 LPC модель	11
4.2 Кластеризация	12
4.2.1 K-Means	12
4.2.2 GMM	13
4.2.3 GMM-HMM	13
4.3 Оценка качества кластеризации	14
4.3.1 Внутренние метрики	14
4.3.2 Внешние метрики	15
5 Классификация фонем	16
6 Результаты	17
6.1 Кластеризация фонем	17
6.1.1 K-Means	18
6.1.2 GMM	19
6.1.3 GMM-HMM	20
6.2 Оценка декодированного звука	20
6.2.1 K-Means	20
6.2.2 GMM	21

6.2.3	GMM-HMM	21
6.3	Классификация фонем	21
6.3.1	Алгоритм Витерби для улучшения прогнозов	22
7	Заключение	22

1 Введение

Развитие технологий машинного обучения в последнее время привело к тому, они все глубже проникают в нашу жизнь, а сфера их применения постоянно расширяется. Особенную важность представляют приложения в медицине, где при помощи анализа данных исследователи создают новые способы диагностирования и изучения заболеваний, изготовления лекарств и лечения пациентов [1].

В частности, машинное обучение используется в нейронауках – междисциплинарной области знаний, занимающейся изучением нейронных процессов. С их помощью создают функциональные протезы и помогают людям вернуть дееспособность после перенесенных болезней или травм.

Центральный подход решения этой сферы заключается в построении интерфейса мозг-компьютер (ИМК, англ. Brain Computer Interface (BCI), также его называют нейроинтерфейс), который осуществляет обмен информацией между мозгом и внешним устройством: компьютером, экзоскелетом или протезом, инвалидной коляской или искусственными органами чувств. Самый распространенный пример – прибор для электроэнцефалограммы (ЭЭГ), который используется в медицине с 1970-х годов. Обмен происходит за счет считывания электрической активности мозга и ее декодирования статистическими методами и методами машинного обучения в команду для устройства.

Устройства для регистрации активности головного мозга можно разделить по своему расположению на инвазивные и неинвазивные. В первом случае производится имплантация нейроинтерфейса на кору головного мозга, примерами таких методов могут служить электрокортикография (ЭКоГ) или стерео-ЭЭГ. Во втором случае активность считывается без непосредственного внедрения в мозг, примерами таких подходов являются ЭЭГ или магнитоэнцефалография (МЭГ). Как правило, нейроинтерфейсы реализуются на основе неинвазивных методов, однако значительное повышение пропускной способности канала между мозгом и внешним устройством возможно только

при использовании технологий, требующих хирургического вмешательства.

Текущая работа нацелена на построение нейроинтерфейса для восстановления речевой функцию, одно из самых интересных потенциальных применений технологии BCI. Вероятно в будущем такие интерфейсы смогут помочь в разработке имплантов, дающих людям с параличом или другими болезнями, связанными с нарушениями речи, возможность говорить.

Со всем кодом можно ознакомиться по [ссылке](#).

2 Обзор литературы

Попытки применения методов машинного обучения для декодирования речи по электрической активности предпринимались многими исследователями [2], [3], [4]. Однако большинство попыток строилось на неинвазивных данных, как правило полученных при помощи ЭЭГ, подробный обзор таких работ можно прочитать в статье [2]. Такие данные имеют меньшую пропускную способность и имеют на выходе более зашумленные данные, поскольку при считывании электрический сигнал проходит через черепную коробку. Также такой подход не практичен с точки зрения построения компактного импланта, которым можно будет использовать повседневно, поскольку неинвазивные устройства являются более громоздкими и требуют определенных процедур для их монтажа.

Также было предпринято несколько попыток построения инвазивных нейроинтерфейсов для декодирования речи, в частности для распознавания отдельных слов [3], так и в распознавании фонем [4]. Но эти исследования использовали данные с устройств, имеющие большое количество каналов для считывания. В работе [4] использовались данные мозговой активности, считанные внутрикорковыми датчиками, а в работе [3] данные считывались массивной сеткой ЭКоГ. Такие устройства не предназначены для длительного использования и связаны с большими рисками для пациента [7].

Поэтому для построения повседневного импланта необходимо чтобы он

был инвазивным и с небольшим количеством каналов. Решением в данном случае может послужить минимально-инвазивное устройство стерео-ЭЭГ. В этом методе глубинные электроды устанавливаются через миниинвазивные отверстия, не требующие разрезов и трепанации черепа. Стерео-ЭЭГ чаще всего используется при выявлении эпилептогенных зон и является наиболее безопасным инвазивных интерфейсом [8]

В работе [6] решается задача декодирования слов по электрической активности мозга, считанной при помощи стерео-ЭЭГ. Рассмотрим эту работу детальнее.

2.1 Данные

Данные были собраны у пациента, которому по медицинским показаниям был имплантирован стерео-ЭЭГ. Была проведена запись, в результате которой он в течение двух сессий суммарной длиной примерно час произносил одну из 6 фраз в случайном порядке, содержащие 26 слов.

2.2 Модель

Архитектуру всей модели можно разбить на две большие части. Первая часть по данным ЭЭГ учится распознавать данные звука, а именно построенную по ним спектrogramму. Вторая часть по предсказанной спектrogramме учится классифицировать слова. Схема модели отображена на иллюстрации 2.1.

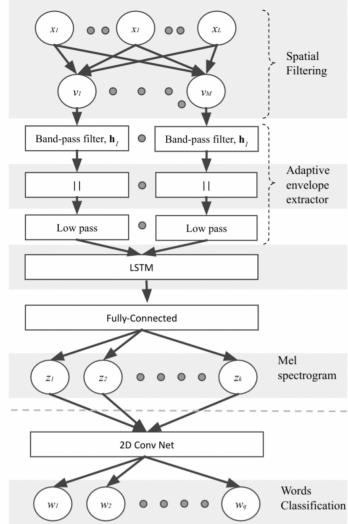


Рис. 2.1: Архитектура существующей модели, источник [6]

2.3 Результаты

В результате было достигнуто в среднем 44% точности классификации 26 слов, используя только 6 каналов данных, записанных с одного стерео-ЭЭГ. На иллюстрации 2.2 отражена матрица ошибок (confusion matrix) предсказаний, а на иллюстрации 2.3 приводится пример сравнения спектрограмм построенной по звуку и предсказанной моделью.

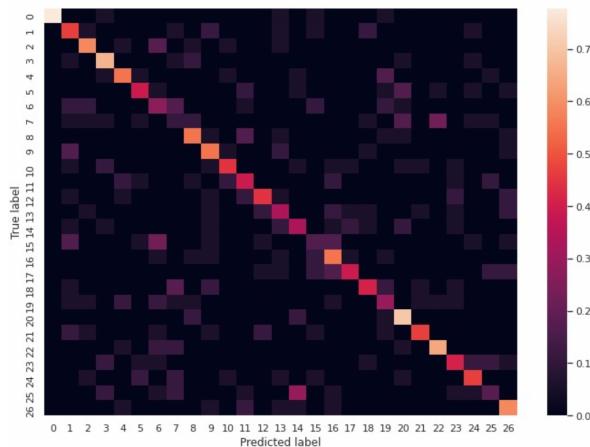


Рис. 2.2: Матрица ошибок (confusion matrix) предсказаний, источник [6]

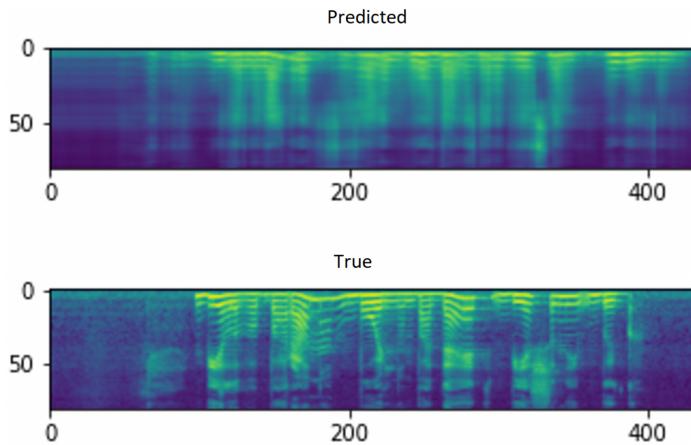


Рис. 2.3: Сравнение спектрограмм построенной по звуку и предсказанной по стерео-ЭЭГ, источник [\[6\]](#)

Минусом данного подхода является то, что модель предсказывает слово не напрямую по данным стерео-ээг, а разделена на две части: предсказание по данным стерео-ээг спектrogramму записанного звука, а затем предсказание слова по предсказанной первой моделью спектrogramмой. При этом если восстановить по этой спектrogramме звук, то получится что-то очень далекое от исходного аудио. Фактически предсказания слово происходит по какому-то представлению, близкому к мел-спектrogramме звука. Причем эти представления не поддаются интерпретации.

Идея данной работы заключается в том, чтобы обучить представления звука так, чтобы они были интерпретируемыми и по ним можно было восстановить исходный звук. Также предполагается, что эти представления будут дискретными и конечными, т.е. задача состоит в том, чтобы выделить из звука так называемые фонемы.

3 Данные и предобработка

Уникальные данные предоставлены Осадчим А.Е. – директором «Центра биоэлектрических интерфейсов». Они также использовались в статье [\[6\]](#).

Датасет включает в себя электрическую активность мозга во время произнесения речи, записанную при помощи стерео-ЭЭГ, а также записанный в процессе звук. Данные были собраны за два сеанса и в сумме составляют

примерно 1 час. В процессе записи пациент произносил одну из 6 различных случайно перемешанных фраз и делал отдых.

Запись проводилась у пациента с эпилепсией, которому по медицинским показаниям был имплантирован стерео-ЭЭГ с 6 контактами. Схематическая иллюстрация устройства с 6 каналами приведена на изображении 3.1.

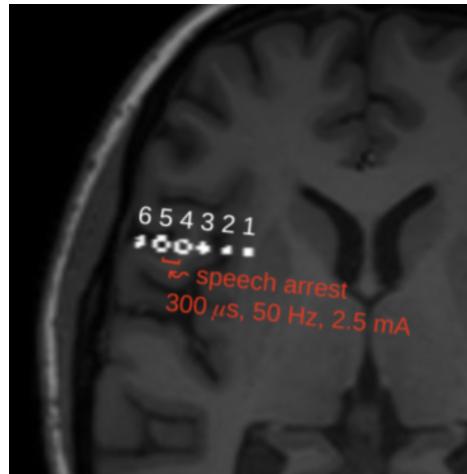


Рис. 3.1: Стерео-ЭЭГ с 6 каналами, источник [\[6\]](#)

3.1 Предобработка

В сырых данных ЭЭГ сначала понижается дискретизация. Затем применяется высокочастотный фильтр Баттервортса, чтобы очистить данные от глазных артефактов. Далее применяется IIR-Notch для того чтобы избавиться от лишних частот в данных. Также применяется низкочастотный фильтр Баттервортса, для того чтобы избавиться в данных от утечки таргета, поскольку звуковые сигналы также могут доходить до стерео-ЭЭГ.

4 Построение фонем

Основная идея заключается в следующем. Полный звуковой канал разбивается на короткие фрагменты (в данной работе исследовались длины 10-100мс), затем каждому фрагменту необходимо поставить в соответствие номер его фонемы. Как упоминалось ранее, этот процесс должен быть обратимым. Подобная задача может быть решена при помощи метода класте-

ризации, поскольку мы хотим закодировать фрагменты так, чтобы похожие участки звука относились к одной фонеме.

Однако применять алгоритмы кластеризации на исходных данных звука является мало осмысленным в виду их большой размерности. Поэтому нужно произвести какое-то преобразование звука в пространство меньшей размерности так, чтобы это преобразование было обратимым. Одним из наиболее популярных подходов решения этой задачи является метод Linear Predictive Coding (LPC), который широко используется в обработке звука, в особенностях обработки речи [9]

4.1 LPC модель

Пусть x_t это амплитуда звукового сигнала в момент времени t . Изначальная цель LPC состояла в том, чтобы смоделировать производство человеческого голоса. Она относится к так называемой source-filter модели, которая основывается на наличие источника звука, проходящего через фильтр, на схеме 4.1.

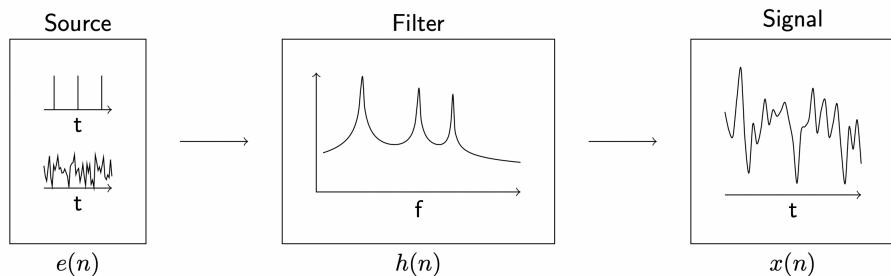


Рис. 4.1: Source-filter модель производства голоса, источник [9]

Источник e_t моделирует голосовые связки, в то время как резонансный фильтр h_t моделирует голосовой тракт. Результирующий сигнал представляет собой:

$$x_t = (h * e)_t$$

где операция $*$ означает свертку сигналов.

Далее модель предполагает, что текущий сигнал x_t также зависит от p предыдущих значений (x_{t-1}, \dots, x_{t-p}) и что источник является константой, следовательно получаем уравнение:

$$x_t = \sum_{k=1}^p a_k x_{t-k} + e_t$$

Таким образом, задача заключается в нахождении коэффициентов a_k . В приложениях временных рядов такой подход известен как авторегрессионная (AR) модель.

Одним из наиболее эффективных методов оценки коэффициентов a_k является метод Бурга, который описан в [10] и будет использоваться в этой работе.

Полученные в результате работы алгоритма оцененные коэффициенты a_k далее будем называть LPC коэффициентами. Недостаток LPC коэффициентов заключается в их высокой чувствительности к шуму. Поэтому на практике могут использоваться другие формы представления LPC коэффициентов. В данной работе помимо LPC коэффициентов будут применяться также Reflection Coefficients (RC), Line Spectral Frequencies (LSF), Log Area Ratios (LAR).

4.2 Кластеризация

В качестве методов кластеризации было выбрано три метода: k-средних (англ. K-Means), модель смеси Гаусса (англ. Gaussian Mixture Model, сокр. GMM) и скрытая марковская модель с объектами гауссовой смеси (англ. Hidden Markov Model with Gaussian mixture emissions, сокр. GMM-HMM)

4.2.1 K-Means

Один из самых используемых методов кластеризации. Алгоритм состоит из следующих шагов:

- 1 Изначально определяется количество кластеров k .
- 2 Из данных отбираются k случайных объектов как начальные центры кластеров.
- 3 Для каждого объекта определяется свой кластер как ближайший к нему центр кластера.
- 4 Для каждого кластера вычисляется центр масс (среднее по всем признакам), далее центр масс назначается центром нового кластера.
- 5 3-й и 4-й шаги итеративно повторяются до момента, когда на какой-то итерации не происходит изменения кластеров.

4.2.2 GMM

Модель смеси Гаусса — это вероятностная модель, которая предполагает, что все объекты генерируются из смеси конечного числа распределений Гаусса с неизвестными параметрами. Построение модели основывается на применении EM-алгоритма, где латентными переменными является номер гауссова распределения модели. Эти же латентные переменные будут являться номером кластера каждого объекта.

4.2.3 GMM-HMM

Скрытая марковская модель — это генеративная вероятностная модель, в которой последовательность наблюдаемых объектов X генерируется последовательностью внутренних скрытых состояний Z . Скрытые состояния не наблюдаются непосредственно. Предполагается, что переходы между состояниями образуют цепь Маркова, т.е. вероятность наступления каждого состояния зависит только от предыдущего состояния. Они могут быть заданы вероятностным распределением π начального состояния и матрицей переходных вероятностей A . Каждое состояние производит наблюдаемый объект в

соответствии с некоторым распределением, которое можно задать параметрически с параметром θ . В случае GMM-HMM модели объект в состоянии генерируется из гауссовой смеси распределений. Скрытые состояния будут являться результатами кластеризации. В данном случае, задача разбивается на две части:

- 1 По наблюдаемым данным сделать оценку параметров модели.
- 2 По параметрам модели и наблюдаемым данным оценить оптимальную последовательность скрытых состояний.

Первую задачу можно решить при помощи алгоритма Баума-Велша, который основан на ЕМ-алгоритме и описан в статье [11]. Вторую задачу можно решить при помощи алгоритма Витерби, который основан на применении динамического программирования и подробно описан в статье [12]. Данный метод представляет наибольший интерес для построения фонем, поскольку предполагается что последовательность переходов фонем можно описать цепью Маркова.

4.3 Оценка качества кластеризации

Для того чтобы отобрать лучшее разбиение на фонемы нужно выбрать оптимальный размер фрагмента фонемы, количество фонем, а также способ кодирования звука. Поэтому было предпринято несколько подходов к оценке кластеризации.

4.3.1 Внутренние метрики

Первый подход заключался в использовании внутренних метрик, которые отображают качество кластеризации только по информации в данных. Поскольку не существует априори известным разделение на фонемы, то этот подход является базовым. Наиболее информативной оказалась метрика силуэт.

Значение силуэта показывает, насколько объект похож на свой кластер по сравнению с другими кластерами. Оценка для всей кластерной структуры:

$$Sil(C) = \frac{1}{N} \sum_{c_k \in C} \sum_{x_i \in c_k} \frac{b(x_i, c_k) - a(x_i, c_k)}{\max\{a(x_i, c_k), b(x_i, c_k)\}}$$

где:

- $a(x_i, c_k) = \frac{1}{|c_k|} \sum_{x_j \in c_k} \|x_i - x_j\|$ – среднее расстояние от $x_i \in c_k$ до других объектов из кластера c_k (компактность)
- $b(x_i, c_k) = \min_{c_l \in C \setminus c_k} \left\{ \frac{1}{|c_l|} \sum_{x_j \in c_l} \|x_i - x_j\| \right\}$ – среднее расстояние от $x_i \in c_k$ до объектов из другого кластера $c_l : k \neq l$ (отделимость).

Силуэт лежит в диапазоне от -1 до 1 и качество кластеризации лучше, чем ближе он к единице.

Поскольку методы GMM и GMM-НММ являются статистическими, то для них можно посчитать логарифм правдоподобия, а также показатели AIC и BIC, но в итоге они не оказались информативными.

4.3.2 Внешние метрики

Другой подход заключается в использовании внешних метрик, которые основаны на сравнении результата кластеризации с априори известным разделением на классы. Этого распределения в данном случае нет, но используется подход stability-based validation, который описан в статье [13]. Идея основана на предположении о существовании истинного распределения кластеров, в соответствии с этим при обучении на разных выборках одним методом, должен получаться схожий результат. Поэтому оценка происходило следующим образом: выборка делилась на обучающую и тестовую, строились две модели одним методом с общими параметрами, но одна тренировалась на обучающей выборке, а другая на тестовой. Затем строились прогнозы обеих моделей на тестовой выборке и сравнивались внешними метриками кластеризации. А именно, были выбраны следующие метрики:

- Homogeneity – однородность кластеров, значение метрики должно уменьшаться при объединении в один кластер двух эталонных.
- Completeness – полнота кластеров, значение метрики качества должно уменьшаться при разделении эталонного кластера на части.
- Adjusted Rand Index (ARI), которое оценивает, насколько много из тех пар элементов, которые находились в одном классе, и тех пар элементов, которые находились в разных классах, сохранили это состояние после кластеризации алгоритмом.

5 Классификация фонем

Самый распространенный подход в работе анализа данных нейроинтерфейсов состоит из нескольких общих этапов:

- 1 Обработка сигналов.
- 2 Извлечение признаков из сигналов.
- 3 Декодирование информации из полученных признаков.

Все эти этапы могут быть оптимизированы при помощи использования глубинного обучения (DNN). Поэтому данная работа будет нацелена именно на работу с нейронными сетями.

Наиболее подходящим типом нейронных сетей для нашей задачи являются сверточные нейронные сети, поскольку они могут аппроксимироватьложения фильтров на сигнал, с чем связано большое количество подходов по работе с ЭЭГ данными.

Существует несколько архитектур по работе с ЭЭГ данными, использующих сверточные нейросети. Одной из наиболее известных является архитектура EEGNet [14], которая содержит четко разграниченный пространственные и временные сверточные блоки и при этом содержит небольшое количество параметров.

В данной работе будет использоваться архитектура концептуально похожая на EEGNet, но разработанная независимо от нее. Это архитектура, которая была предложена в статье [5] для декодирования кинематических движения пальца руки. Ее архитектура приведена на рисунке 5.1.

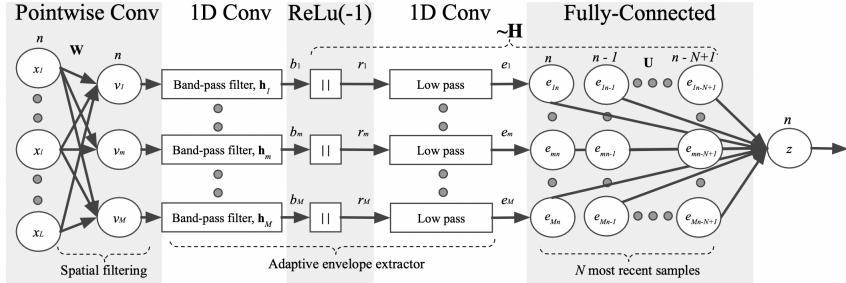


Рис. 5.1: Сверточная архитектура обработки ЭЭГ сигналов, источник [5]

Для нашей задачи дополним ее LSTM слоем как делалось в работе [6] по декодированию слов.

6 Результаты

6.1 Кластеризация фонем

Для отбора лучшего разбиения на фонемы был проведен эксперимент, в котором для каждого метода кластеризации перебирались следующие параметры:

- Способ кодирования звука: LPC, RC, LSF, LAR.
- Длина окна: 10мс, 30мс, 50мс, 100мс.
- Количество кластеров: от 16 до 30 с шагом 2.

Для каждой комбинации была построена кластеризация и оценена способами, описанными в разделе 5.3.

6.1.1 K-Means

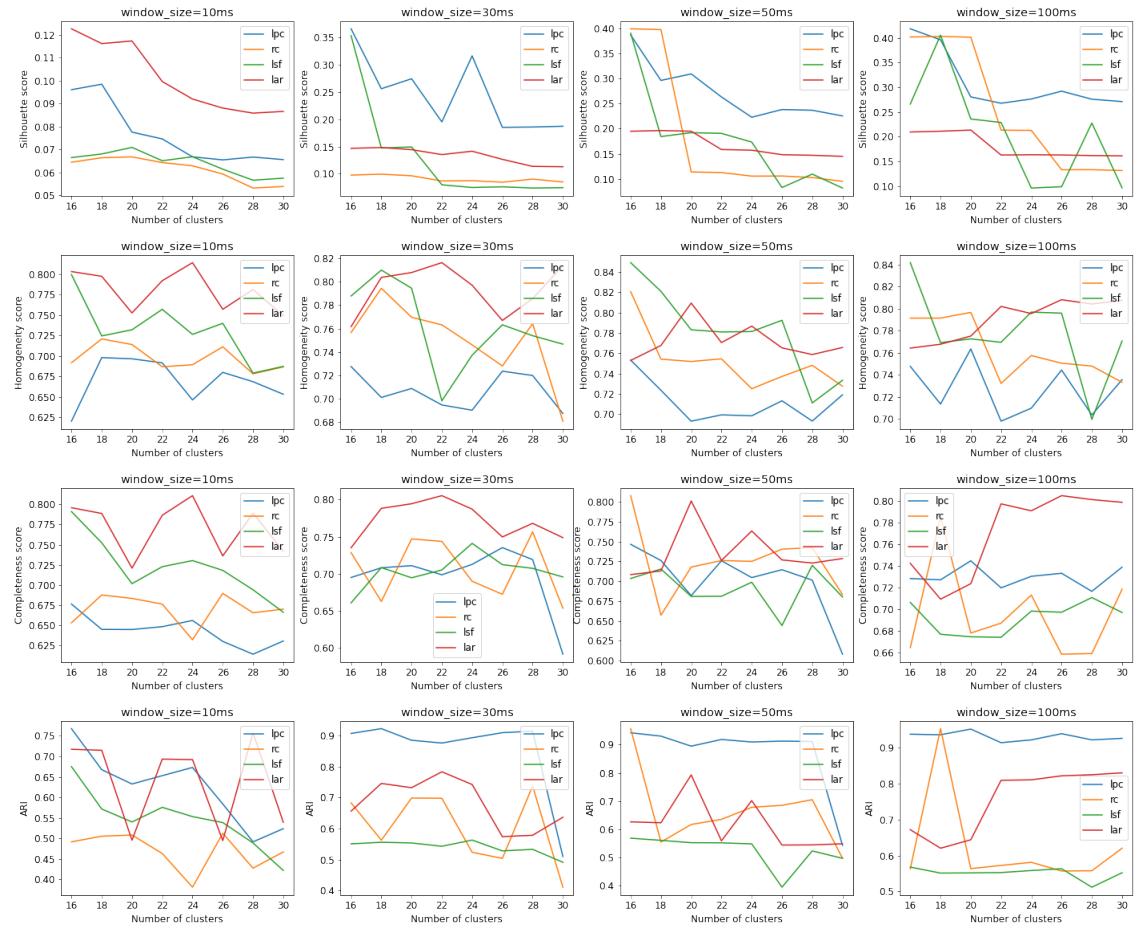


Рис. 6.1: Результаты кластеризации методом K-Means

6.1.2 GMM

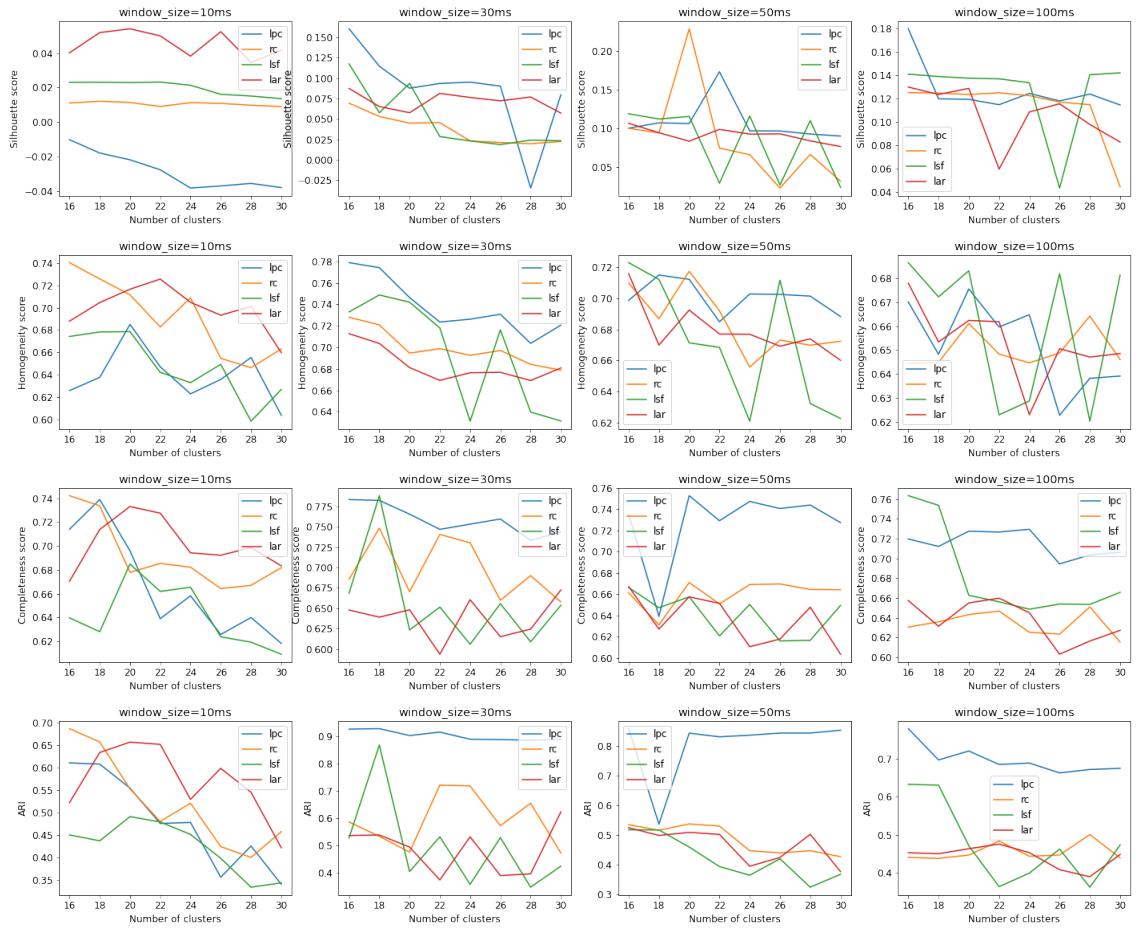


Рис. 6.2: Результаты кластеризации методом GMM

6.1.3 GMM-HMM

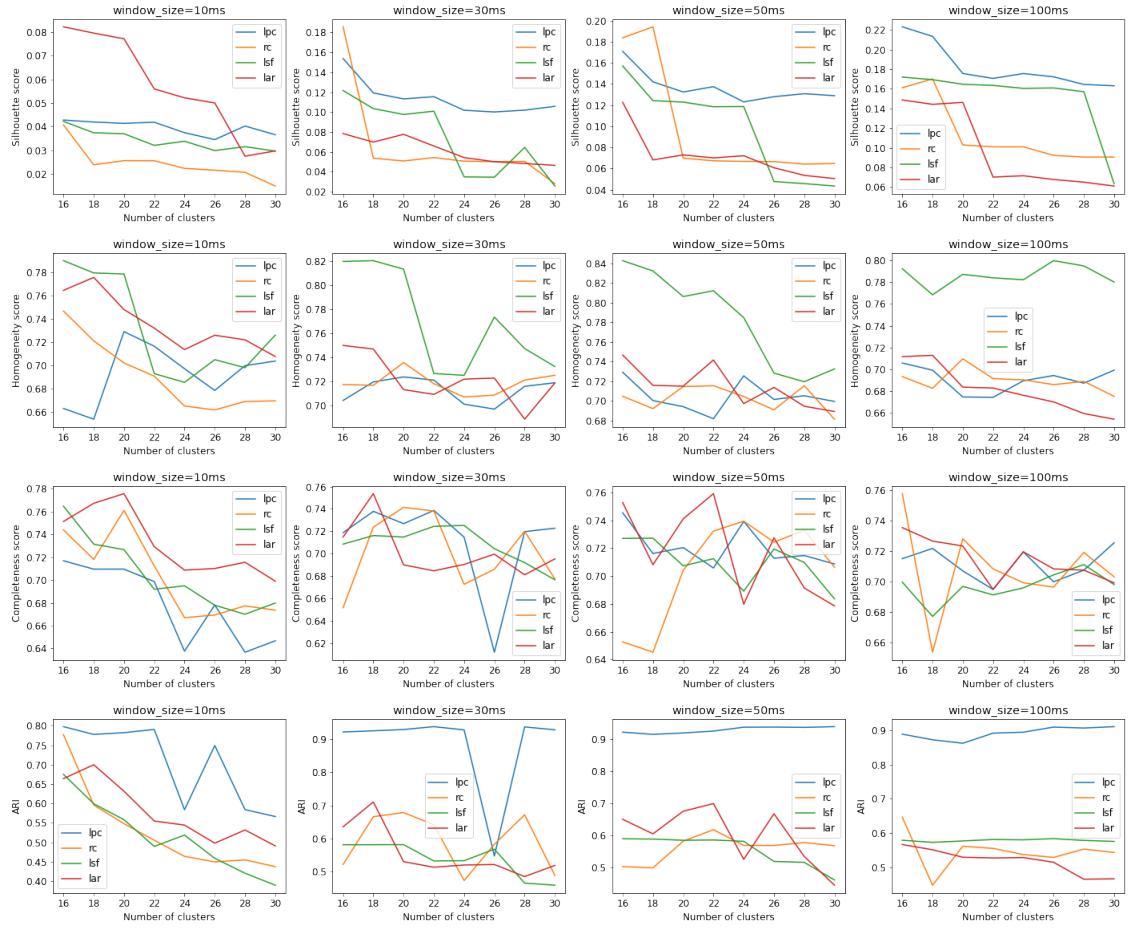


Рис. 6.3: Результаты кластеризации методом GMM-HMM

6.2 Оценка декодированного звука

Также было проведено психофизиологическое тестирование на выборке из трех человек. В результате этого тестирования закодированный LPC участок заменялся на центр кластера фонемы и затем декодировался в звук. Далее респондентам было предложено оценить получившийся звук. В эксперименте перебирались длины окон и способ кодирования (LPC, RC, LSF, LAR), оптимальное количество кластеров подбиралось из метрик кластеризации.

6.2.1 K-Means

В целом речь получилась различимая при любых параметрах. Кодирование LPC коэффициентами давало наихудший результат, RC сравнительно

лучше. Но лучше всех оказались LSF и LAR которые выдают сопоставимо одинаковое качество. От размера окна зависимость не выявила.

6.2.2 GMM

Важным фактором для этого метода оказался размер окна: на уровне 10мс есть небольшие вибрации при произношении слов, а при $\geq 50\text{мс}$ речь становится более расплывчатой, поэтому оптимальный размер окна 30мс. LPC, LSF, LAR дают сопоставимо хорошее качество, RC хуже.

6.2.3 GMM-HMM

При кодировании LPC коэффициентами получается неразборчивую речь, RC выдают качество лучше, но значительно лучше оказались LSF и LAR, которые имеют сопоставимо одинаковое качество. Аналогично GMM важным параметром является размер окна: при маленьком окне будут вибрации во время речи, при большом окне речь будет расплывчата. Оптимальный размер окна: 30мс.

6.3 Классификация фонем

Исходя из результатов кластеризации, были отобраны лучшие способы кодирования звука: LAR, LSF. Также была отобрана оптимальная длина окна фонемы : 30мс.

Затем проводились эксперименты, в котором по 1.5с записи ЭЭГ предсказывалась фонема при разных способов кодирования звука и разных методов кластеризации.

С результатами экспериментов можно ознакомиться в таблицах [6.1](#), [6.2](#), [6.3](#).

Таблица 6.1: Результаты при построении фонем методом K-Means

	train cross entropy	test cross entropy	train accuracy	test accuracy
LAR	1.78	2.84	0.38	0.14
LSF	1.13	1.68	0.66	0.56

Таблица 6.2: Результаты при построении фонем методом GMM

	train cross entropy	test cross entropy	train accuracy	test accuracy
LAR	1.59	2.46	0.50	0.27
LSF	1.38	2.02	0.58	0.47

Таблица 6.3: Результаты при построении фонем методом GMM-HMM

	train cross entropy	test cross entropy	train accuracy	test accuracy
LAR	1.74	2.83	0.39	0.15
LSF	1.33	2.06	0.59	0.44

6.3.1 Алгоритм Витерби для улучшения прогнозов

При построении фонем методом GMM-HMM, можно воспользоваться следующим приемом. Обучить модель для предсказания фонем и получить вероятности предсказания для каждой фонемы. Полученные вероятности и матрицу переходов из модели кластеризации GMM-HMM применить в алгоритме Витерби для получения лучших предсказаний с учетом матрицы переходов.

Был проведен подобный эксперимент, с его результатами можно ознакомиться в таблице 6.4

Таблица 6.4: Точность предсказаний после применения алгоритма Витерби

	source accuracy	viterbi to the whole sequence	viterbi to windows
LAR	0.15	0.14	0.14
LSF	0.44	0.51	0.52

7 Заключение

В ходе курсовой работы был проведен анализ существующих решений в области декодирования речи по данным ЭЭГ. Большинство из них строилось на неинвазивных данных, которые имеют меньшую пропускную способность

и в следствии чего гораздо хуже подходят для декодирования речи. Решения основанные на инвазивных интерфейсах, как правило использовали данные с устройств, имеющие большое количество каналов для считывания, что не предназначено для длительного использования. Поэтому наиболее подходящими для этой задачи являются минимально-инвазивные нейроинтерфесы. Примером такого интерфейса является стерео-ЭЭГ, его данные исследовались в этой работе.

Процесс восстановления речи по данным с стерео-ЭЭГ происходил при помощи декодирования фонем, которые были получены в результате кластеризации фрагментов звука. Перед применением кластеризации, данные по звуку преобразовывались в пространство меньшей размерности при помощи коэффициентов: LPC, RC, LSF, LAR. В результате экспериментов была выявлена оптимальная длина фонемы – 30мс, количество фонем – 18, а также наилучшие способы кодирования: LSF, LAR.

Распознавание фонем осуществлялось сверточной нейронной сетью с архитектурой схожей на DenseNet, которая хорошо себя показала в работе по распознаванию отдельных слов. Наилучшее качество распознавания происходило, когда звук кодировался при помощи LSF коэффициентов. Если сравнивать методы кластеризации фонем, то наилучшее качество распознавания фонем у K-Means, 0.56 точность предсказания, затем идет GMM, 0.47 точность предсказания, и хуже всех оказался GMM-HMM, 0.44 точность предсказания. Однако, если к полученным предсказанным вероятностям применить алгоритм Витерби, который будет учитывать еще вероятности переходов между состояниями, то качество значительно вырастает до точности 0.52.

Однако это все еще не качество, способное на качественное восстановление речи. Говоря о дальнейшей работе, то конечно же стоит попробовать большее количество моделей по распознаванию фонем. Также есть гипотеза о том, что качество будет выше, если обучить две модели: одна будет классификацией по распознаванию тишины, а другая будет распознавать конкретную фонему речи.

Список литературы

1. May, M. Eight ways machine learning is assisting medicine. *Nat Med* 27, 2–3 (2021).
2. Panachakel J. T., Ramakrishnan A. G., Decoding Covert Speech From EEG-A Comprehensive Review, *Frontiers in Neuroscience* (2021).
3. Makin, J.G., Moses, D.A. & Chang, E.F. Machine translation of cortical activity to text with an encoder–decoder framework. *Nat Neurosci* 23, 575–582 (2020).
4. G. H. Wilson, S. D. Stavisky, F. R. Willett, D. T. Avansino, J. N. Kelemen, L. R. Hochberg, J. M. Henderson, S. Druckmann, and K. V. Shenoy, “Decoding spoken english from intracortical electrode arrays in dorsal precentral gyrus,” *Journal of Neural Engineering*, vol. 17, no. 6, p. 066007 (2020).
5. A. Petrosyan, M. Sinkin, M. Lebedev and A. Ossadtchi. Decoding and interpreting cortical signals with a compact convolutional neural network, *J. Neural Eng.* (2021).
6. A. Petrosyan, A. Voskoboinikov and A. Ossadtchi. Compact and interpretable architecture for speech decoding from stereotactic EEG, *Third International Conference Neurotechnologies and Neurointerfaces* (2021).
7. P. Jayakar, J. Gotman, A. S. Harvey, A. Palmini, L. Tassi, D. Schomer, F. Dubeau, F. Bartolomei, A. Yu, P. Krs̄ek, D. Velis, and P. Kahane, “Diagnostic utility of invasive eeg for epilepsy surgery: Indications, modalities, and techniques,” *Epilepsia*, vol. 57, no. 11, pp. 1735–1747 (2016).
8. Bourdillon P, Ryvlin P, Isnard J, Montavont A, Catenoix H, Mauguière F, Rheims S, Ostrowsky-Coste K, Guénot M. Stereotactic Electroencephalography Is a Safe Procedure, Including for Insular Implantations. *World Neurosurg* (2017).

9. Hyung-Suk Kim, Linear Predictive Coding is All-Pole Resonance Modeling (2014).
10. L. Marple, "A new autoregressive spectrum analysis algorithm," in IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 28, no. 4, pp. 441-454 (1980).
11. Jeff Bilmes, A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models, Technical Report ICSI-TR-97-021, University of Berkeley (2000).
12. G. D. Forney, "The viterbi algorithm," in Proceedings of the IEEE, vol. 61, no. 3, pp. 268-278, (1973).
13. Lange T, Roth V, Braun ML, Buhmann JM. Stability-based validation of clustering solutions. Neural Comput. (2004).
14. Vernon J. Lawhern, Amelia J. Solon, Nicholas R. Waytowich, Stephen M. Gordon, Chou P. Hung, Brent J. Lance, EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces, J. Neural Eng (2018).