18.06.2022

Курсовая работа

Исследование применимости методов машинного обучения к декодированию речи из минимально инвазивных записей электрической активности головного мозга.

Выполнил: Сизов Кирилл Игоревич
Научный руководитель: Осадчий Алексей Евгеньевич

Введение в предметную область

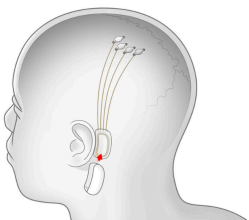
Что такое нейроинтерфейсы?

EEG & fNIRS

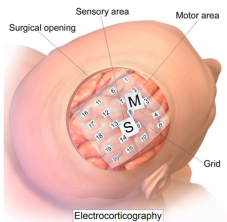
Введение в предметную область

Виды нейроинтерфейсов

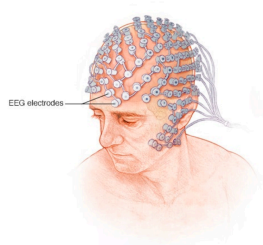
ИНВАЗИВНЫЕ



ПОЛУИНВАЗИВНЫЕ

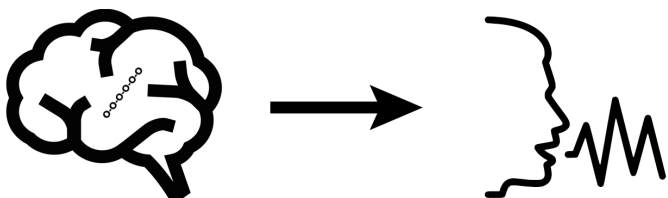


НЕИНВАЗИВНЫЕ



Введение в предметную область

Декодирование речи



"Женя широко шагает шагает в желтых штанах"

Актуальность задачи

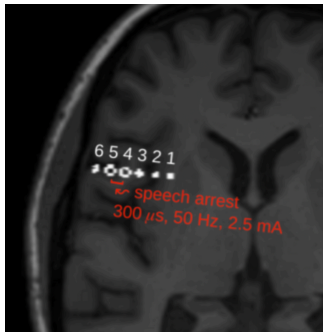
Существующие работы

- Большинство попыток строилось на неинвазивных данных, полученных при помощи ЭЭГ.
- Инвазивные решения использовали данные с устройств, имеющие большое количество каналов для считывания.
- Наиболее практичным подходом является использование данных с минимально-инвазивных устройств.

Актуальность задачи

Использование минимально-инвазивного интерфейса

В работе A. Petrosyan, M. Sinkin, M. Lebedev and A. Ossadtchi. Decoding and interpreting cortical signals with a compact convolutional neural network, J. Neural Eng. (2021). исследуют задачу декодирования, используя минимально инвазивное устройство стерео-ЭЭГ.



Актуальность задачи

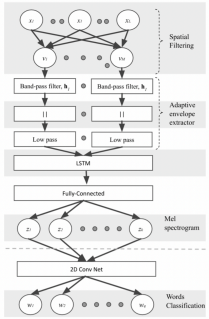
Использование минимально-инвазивного интерфейса

В результате было достигнуто в среднем 44% точности классификации 26 слов, используя только 6 каналов данных, записанных с одного стерео-ЭЭГ.

Актуальность задачи

Минусы подхода

Для распознавания слов использовалась двухступенчатая архитектура, которая сначала по данным стерео-ЭЭГ предсказывает спектрограмму, а затем по спектрограмме предсказывает слово.



Цель работы

В данной работе ставится задача восстановления речи при помощи декодирования *фонем* – коротких участков (10-100мс) речи, соответствующих определенному звуку.

Формальная постановка задачи

Работу можно разбить на несколько этапов:

1. Получить представление фрагментов звука в пространстве меньшей размерности.
2. Кластеризовать фрагменты для построения фонем.
3. Построить модель, которая по участку стерео-ЭЭГ будет распознавать фонему.

Обзор существующих методов

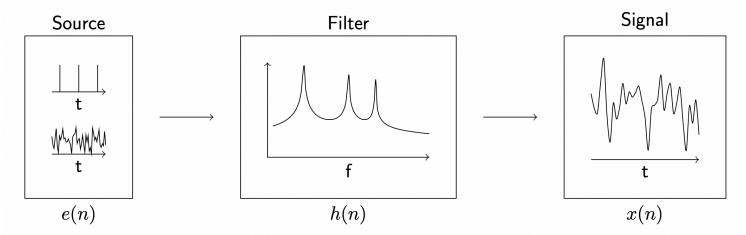
Кодирование звука

Классические способы кодирования звука:

- Спектрограмма.
- Мел-кепстральные коэффициенты (MFCC).
- Linear Predictive Coding (LPC).

Предложенные методы
Linear Predictive Coding

Source-filter модель генерации звука, которая основывается на наличие источника звука, проходящего через фильтр.



Результирующий сигнал представляет собой $x_t = (h * e)_t$

Модель предполагает, что текущий сигнал xt также зависит от p предыдущих значений: $x_t = \sum_{k=1}^{p} a_k x_{t-k} + e_t$

Коэффициенты a_k называются LPC коэффициентами. В связи с их высокой чувствительности к шуму, в работе использовались различные их представления:

- Reflection Coefficients (RC)
- Log Area Ratios (LAR)
- Line Spectral Frequencies (LSF)

Предложенные методы

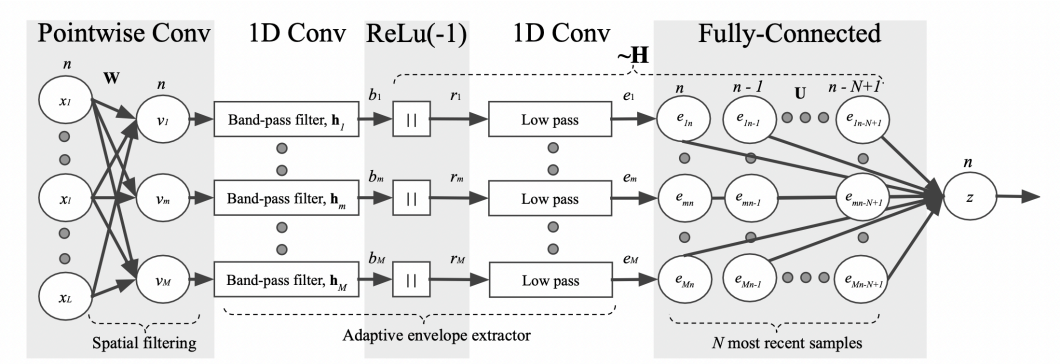
Кластеризация фонем

В данной работе использовались следующие методы кластеризации:

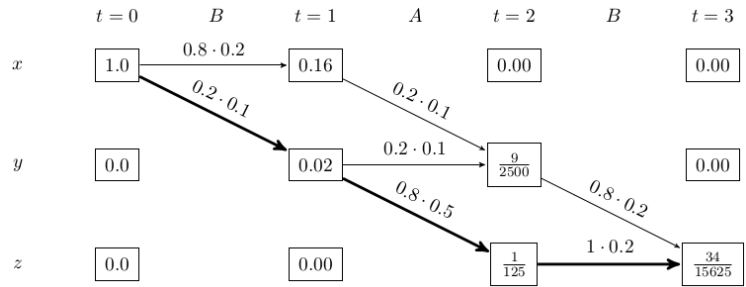
- K-Means
- Gaussian Mixture Model (GMM)
- Hidden Markov Model with Gaussian mixture emissions (GMM-HMM)

Предложенные методы

Распознавание фонем



Алгоритм Витерби использовался для улучшения качества распознования при кластеризации скрытой марковской моделью.



Экспериментальная оценка

Оценка качества кластеризации

Использовалась внутренния метрика кластеризации силуэт, которая показывает насколько объект похож на свой кластер по сравнению с другими кластерами. Также использовались внешние метрики кластеризации для так называемого stability-based validation:

- Homogenity
- Completeness
- Adjusted Rand Index (ARI)

Но самым надежным оказалось психофизиологическое тестирование.

Экспериментальная оценка

Эксперимент по кластеризации

Для отобра 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.

Экспериментальная оценка

Результаты эксперимента

- KMeans: лучше всех оказались LSF и LAR которые выдают сопоставимо одинаковое качество. От размера окна зависимость не выявилась.
- GMM: важным фактором оказался размер окна, на уровне 10мс есть небольшие вибрации при произношении слов, а при $\geq$50мс речь становится более расплывчатой. От способа кодирования зависимость незначительная.
- GMM-HMM: при LPC, RC кодировании получается неразборчивую речь, значительно лучше оказались LSF и LAR. Аналогично GMM важным фактором оказался размер окна.

Экспериментальная оценка

Результаты эксперимента

- Лучшие способы кодирования звука: LAR, LSF.
- Оптимальная длина фонемы : 30мс.
- Оптимальное количество фонем: 18.

Экспериментальная оценка

Распознавание фонем

Проводился эксперимент, в котором по 1.5с записи ЭЭГ предсказывалась фонема при разных способов кодирования звука и разных методов кластеризации.

Экспериментальная оценка

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

Экспериментальная оценка

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

Экспериментальная оценка

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

Экспериментальная оценка

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

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

Результаты

- Получили наилучшую точность предсказания 0.56 при 18 фонемах (против 0.05 при случайном угадывании)
- Текущие результаты не позволяют качественно восстановить речь.
- Дальнейшие шаги: использовать данные ЭЭГ для кластеризации, исследовать применения различных окон для фрагментов звука, попробовать другие модели.