

Дипломная работа

Распознавание побочных акустических сигналов от нажатия клавиш клавиатуры.

Брицин Алексей Александрович

Профессия Data Scientist группа DSU-53

Москва 2025 г.

**Содержание.**

1. Введение. стр. 3

2. Постановка задачи. стр. 3

3. Признаки, описывающие акустические сигналы. стр. 5

4. Описание алгоритмов обучения и обработки данных. стр. 8

5. Результаты. стр.13

6. Выводы. стр.14

7. Список литературы. стр.15

**Введение.**

Получение данных по побочным каналам связи является одним из применяемых методов несанкционированного доступа к конфиденциальной информации. В рамках задачи защиты информации от утечек по акустическому каналу необходимо оценить практическую возможность восстановления защищаемой информации в различных условиях её перехвата. Результаты этой работы планируется использовать для совершенствования нормативно-методической базы по специальной защите технических средств, а также выделенных помещений от утечки информации.

В настоящей работе рассматривается решение задачи возможности посимвольного распознавания текста с помощью методов машинного обучения на примере побочных акустических сигналов.

В работе использовалась 2-х канальная схема приема акустических сигналов, где в качестве приемных устройств применялись микрофоны.

**Постановка задачи.**

Введем следующие обозначения. Под пространством акустических сигналов  понимается множество всех возможных акустических сигналов клавиатуры , генерируемых при нажатии клавиши, где  – количество сигналов. Классификация сигналов осуществляется с помощью алгоритма , выбранного из пространства алгоритмов . Под ответом  подразумевается класс, который присваивается сигналу алгоритмом. Под пространством ответов  подразумевается множество всех возможных классов , где  – количество классов. Алгоритм при классификации использует признаковое описание объекта. Признаковым описанием сигнала называется совокупность всех его признаков: , где  – количество признаков, а признак — это индивидуальное измеримое свойство или характеристика сигнала.

Побочный акустический сигнал от клавиши клавиатуры представляет собой случайный нестационарный процесс. В таком случае распознавание сигналов клавиатуры можно рассматривать с вероятностной точки зрения, а распознавание сводится к задаче нахождения , зная. Согласно вероятностной трактовке задачи, необходимо выдать класс клавиши, которая на данной последовательности признаков будут наиболее вероятна.

 (1)

Задачу максимизации можно переписать в виде задачи минимизации функционала ошибки , представляющей собой ошибку алгоритма  на выборке.

Таким образом, задача сводится к подбору такого алгоритма , для которого достигается минимум функционала ошибки при классификации.

 (2)

В общем случае функционалы ошибок могут отличаться друг от друга в зависимости от выбранного алгоритма, а для сравнения алгоритмов воспользуемся оценкой точности работы алгоритма, при этом необходимо обеспечить, чтобы количество элементов в каждом из классов в тестовой выборке было сбалансировано.

, (3)

где *l*– количество сигналов в тестовой выборке

Для решения задачи алгоритмами обучения с учителем, необходимо правильно выбирать пространство признаков, так как излишняя информация о сигнале затруднит процедуру распознавания в силу того, что классы будут сложно отделимы. Если данных будет недостаточно, то будет проблематично распознать объект из-за того, что одному набору признаков будет соответствовать несколько акустических сигналов.

**Признаки, описывающие акустические сигналы.**

Извлечение признаков из аудиосигналов – это важный этап анализа и обработки аудиоданных, который позволяет представить аудиосигналы в числовой форме, пригодной для обучения нейронных сетей и других методов машинного обучения. Среди таких признаков часто используют следующие [1]:

- спектрограмму,

- мел-частотные (MFCC) и линейные кепстральные коэффициенты (LFCC),

- спектральный центроид (Spectral Centroid),

- спектральный спад (Spectral Rolloff),

- скорость пересечения нуля (Zero Crossing Rate),

- вейвлет представление.

В этой работе используются мел-частотные кепстральные коэффициенты как наиболее широко используемые из-за их эффективности. MFCC представляют аудиосигнал в форме вектора, который описывает его частотные характеристики с учетом особенностей человеческого восприятия. Они позволяют снизить размерность данных, сохраняя при этом информацию, важную для анализа звука.

Процесс извлечения MFCC включает в себя следующие шаги [2]:

1. Фрейминг (Framing).

2. Дискретное преобразование Фурье (DFT).

**3. Вычисление спектра мощности (периодограммы).**

4. Применение банка мел-фильтров (Mel Filter Bank).

5. Логарифмирование (Logarithm).

6. Дискретное косинусное преобразование (Discrete Cosine Transform, DCT).

Все шаги предобработки аудио сигнала показаны на рис. 1.

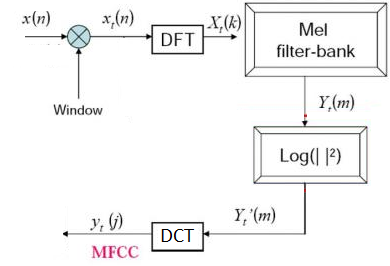


Рис. 1 Блок-схема предобработки аудио сигнала для получения мел-частотных кепстральных коэффициентов (MFCCs).

В работе [3] показано, что наилучшими для распознавания акустических сигналов являются мел-частотные кепстральные коэффициенты, при этом рекомендовалось использовать первые 12 коэффициентов из 26 подсчитанных. Отбрасывания коэффициентов связано с фильтрацией высокочастотных составляющих сигнала. В данной работе используются 128 MFCCs, так как первых 12 коэффициентов оказалось недостаточно для точности распознавания.

Для повышения точности распознавания использовались следующие параметры алгоритма. Для выделения признаков сигнал был дискретизирован с частотой sr=96 кГц, затем он обрезался до длительности duration=0.25-0.3 сек, длина каждой отдельной пачки данных – frame\_size=0.025 сек, шаг окна – frame\_stride=0.01 сек, количество отсчетов для БПФ – NFFT=2048, количество фильтров в банке фильтров – nfilt=128. Далее в зависимости от выбранного алгоритма либо производилось усреднение для сокращения набора признаков с частичным сохранением информации о временной зависимости, либо использовались нормированные мел-частотные кепстральные коэффициенты (CNN, RNN).

Для предварительной оценки возможности классификации объектов по заданному набору признаков используется алгоритм t-SNE для снижения размерности пространства признаков до 2-х. На рис. 2 показаны результаты применения алгоритма t-SNE к мел–частотным кепстральным признакам, полученным с помощью их усреднения для 1 канала (ch0). Видно, что сигналы, имеющее такое признаковое описание формируют кластеры согласно своим классам.

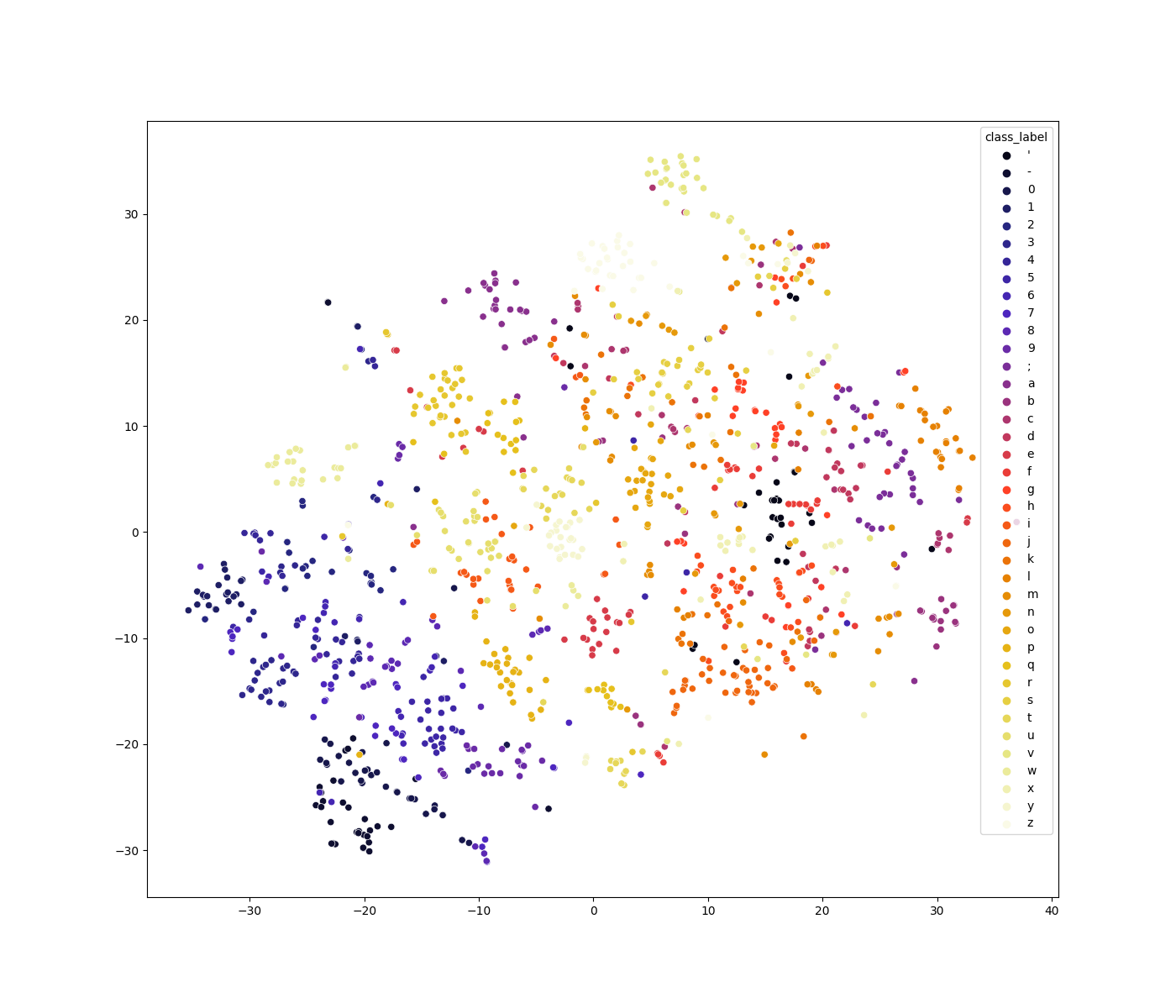


Рис. 2 Кластеры аудио сигналов, полученные с помощью алгоритма t- SNE.

**Описание алгоритмов обучения и обработки данных.**

Для задачи классификации аудио сигналов можно воспользоваться:

1. алгоритмом ближайших соседей (как базовым алгоритмом), методом опорных векторов – так как эти методы хорошо интерпретируются;
2. кусочно-линейными классификаторами (случайным лесом и градиентным бустингом на основе деревьев решений);
3. нейронными сетями - полносвязными (FCNN), свёрточными (CNN), рекуррентными (RNN) сетями;
4. композиция нейронных сетей (на основе полносвязных сетей Ensamble\_FCNN).

В ввиду малого объема выборки (порядка 30 аудио сигналов на класс) доступной для реализации классических алгоритмов машинного обучения и 39 классов (клавиши алфавита и пунктуации) для увеличения надежности в оценки точности использовался метод скользящего контроля (кросс-валидация). В этом случае производилось случайное разбиение выборки на обучающую (90 % от исходной выборки), и тестовую (10 %) с равномерным взятием элементов из каждого класса. Во время тестирования алгоритма, выбирались последовательно сигналы из тестовой выборки. В качестве ответа алгоритма использовалось имя класса сигнала из обучающей выборки, при котором наблюдалась максимальная корреляция с тестовым сигналом.

Для всех классических методов обучения, описанных выше, для поиска гиперпараметров по критерию максимизации точности распознавания использовался алгоритм поиска по сетке GridSearch и для сокращения размерности признаков использовался метод главных компонент (PCA).

Точность работы алгоритма оценивалась по формуле (3) критерия качества работы алгоритмов распознавания.

Для нейросетевых классификаторов виду малого количества объектов и большого количества признаков, обучение глубоких нейронных сетей приводит к существенному переобучению. По мере увеличения размера обучающей выборки проблема переобучения глубоких нейронных сетей становится менее серьезной. Иначе говоря, чем больше множество данных, тем сложнее (гибче) модель, которая приближает данные. Одно из грубых эвристических правил, что количество данных должно в несколько раз (в 5 или 10) превышать количество адаптивных параметров в модели [4]. Поэтому было принято решение использовать двухслойную полносвязную, небольшого размера сверточную на основе LeNet (с одноканальными 1-D свёрточными слоями) и рекуррентную (GRU из-за меньшего вычислительного ресурса) нейронные сети.

В качестве функции потерь в задаче классификации для нейронных сетей используется категориальная кросс-энтропия (categorical cross-entropy).



(4)

, где sp – значение, выданное последним слоем нейронной сети для верного класса, sj - значение, выданное последним слоем нейронной сети для j класса, C – количество классов.

Данная функция потерь широко используется в задачах мультиклассовой классификации [5]. При обучении сеть учится выдавать вероятности для всех классов. Для использования данной функции потерь метки класса кодировались с помощью прямого кодирования (one-hot).

Для ускорения сходимости в процессе обучения и для борьбы с переобучением используются слои dropout [6] с коэффициентов прореживания 0.4. В работе используется L2 регуляризатор для ограничения величины смещения и эластичный регуляризатор (Elastic L1\_L2) [7] для ограничения значений коэффициентов фильтров и значения выхода функции активации.

В качестве алгоритма оптимизации нейронной сети используется метод Adam cо значением скорости обучения равным 0.001, в качестве функции активации скрытых слоев применялась функция ReLu, для последнего слоя – функция softmax, с выходным слоем равным 39 нейронам (по количеству клавиш. Архитектуры нейросетевых классификаторов приведены на рис.3 – 5.

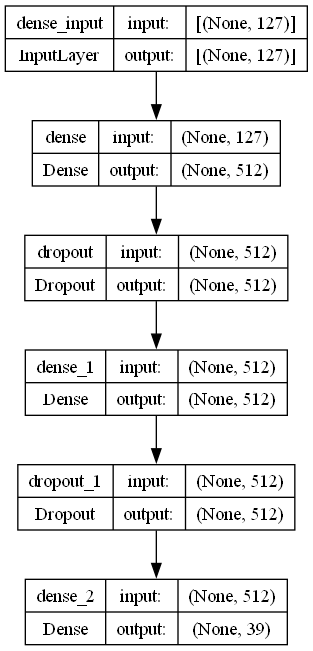


Рис. 3 Архитектура FCNN.

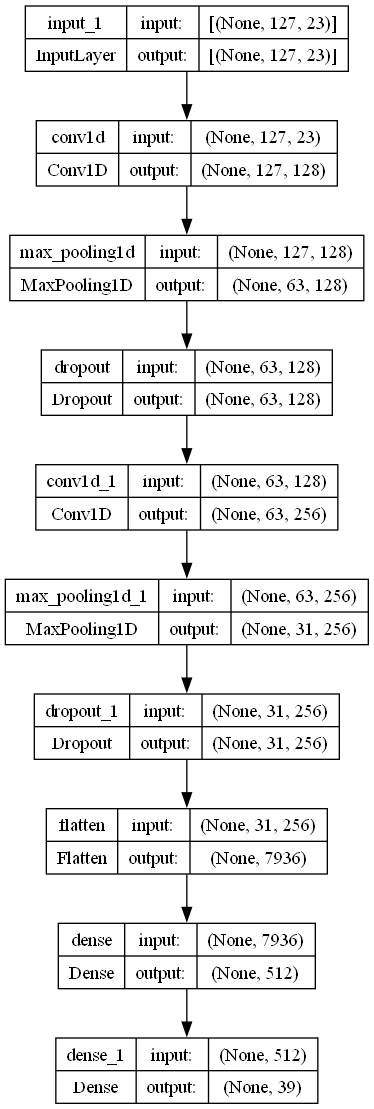


Рис. 4 Архитектура CNN.

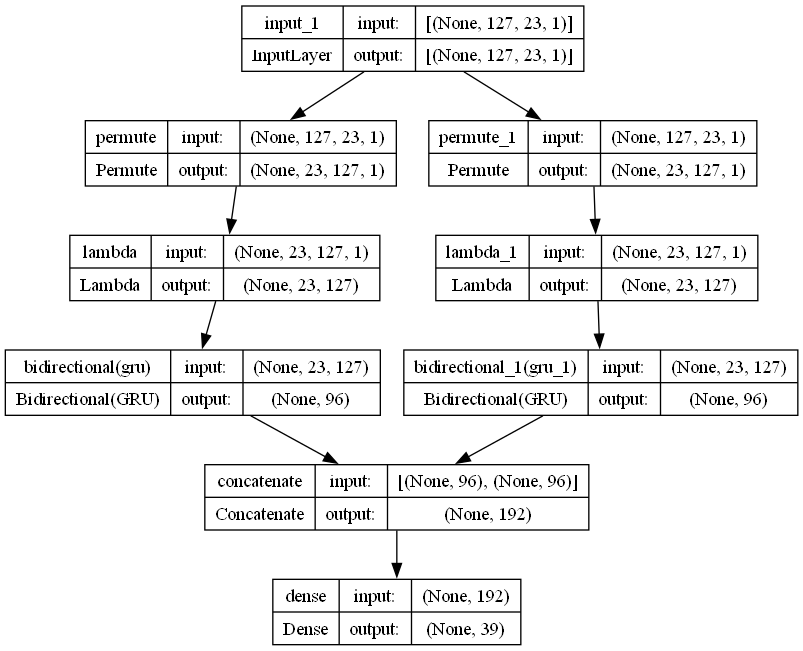


Рис. 5 Архитектура RNN (GRU) нейронной сети (2 канала из 4).

При создании архитектуры RNN (GRU) сети использовались три приема повышения качества и степени обобщения рекуррентных нейронных сетей:

1) рекуррентное прореживание (recurrent dropout) – особый способ использования прореживания для борьбы с переобучением в рекуррентных слоях

2) использование параллельно расположенных рекуррентных слоев для увеличения репрезентативности сети (за счет увеличения объёма вычислений)

3) использование двунаправленных рекуррентных слоев, которые представляют одну и ту же информацию в рекуррентной сети разными способами, повышая точность и ослабляя проблемы, связанные с забыванием.

Для дальнейшего улучшения качества распознавания можно использовать композицию нейронных сетей. Композиция алгоритмов машинного обучения — это использование нескольких (не обязательно разных) моделей вместо одной. End-to-end нейронная сеть (Ensamble FCNN) представляет собой объединение 2 полносвязных нейронных сетей в ветви (по количеству каналов приема), результирующие тензоры которых объединяются на слое конкатенации и поступают на скрытый слой, состоящий из 256 нейронов для выделения глобальных признаков (рис.6).

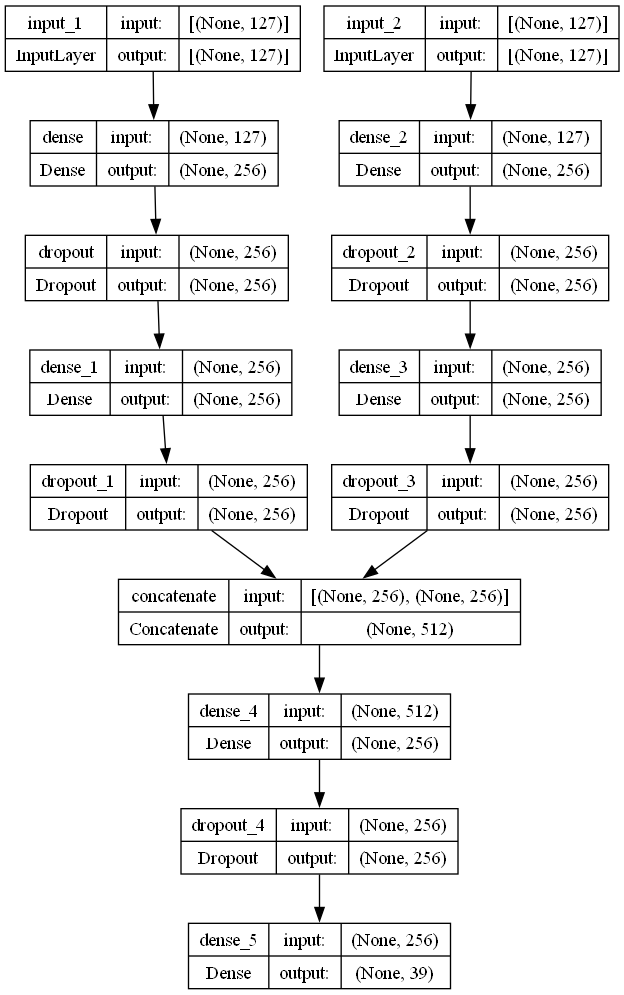


Рис.6 Архитектура Ensamble FCNN.

**Результаты.**

В таблице 1 показаны метрики точности использованных в работе алгоритмов на тестовой выборке (time – время работы алгоритма).

Из таблицы видно, что применение нейросетевых классификаторов более оправданно в задачах распознавания акустических сигналов. Из классических методов машинного обучения в данной задаче можно использовать только метод опорных векторов с использованием алгоритмов исправления ошибок.

В таблице 2 показаны вероятности попадания тестовых распознанных символов в глубину колонки 1-5 (ТОП1 - ТОП5) для нейросетевых классификаторов.

Таблица 1.

|  |  |  |  |
| --- | --- | --- | --- |
| **Алгоритм** | **CH0** | **CH1** | **Time (сек)** |
| KNN | 0.765 | 0.762 | 446 |
| SVM | 0.881 | 0.894 | 131 |
| Random Forest | 0.779 | 0.804 | 590 |
| XGBoost | 0.792 | 0.802 | 1386 |
| FCNN | 0.915 | 0.904 | 79 |
| CNN | 0.941 | 0.915 | 168 |
| RNN | 0.931 | 0.971 | 1731 |
| Ensamble\_FCNN | 0.973 | | 29 |

Таблица 2.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Алгоритм** | **TOP1** | **TOP2** | **TOP3** | **TOP4** | **TOP5** |
| FCNN: CH0  CH1 | 0.915 | 0.957 | 0.963 | 0.963 | 0.979 |
| 0.904 | 0.952 | 0.963 | 0.968 | 0.984 |
| CNN: CH0  CH1 | 0.941 | 0.968 | 0.973 | 0.979 | 0.989 |
| 0.915 | 0.963 | 0.968 | 0.984 | 0.984 |
| RNN: CH0  CH1 | 0.931 | 0.979 | 0.984 | 0.989 | 0.989 |
| 0.971 | 0.989 | 0.994 | 1.000 | 1.000 |

Исходя из результатов обучения моделей можно сделать вывод, что вероятность распознавания символов клавиатуры выше 95% можно получить, применяя RNN (GRU или LSTM) сети или композиции (ансамбли) нейронных сетей, либо усложняя архитектуры нейронных сетей.

**Выводы.**

1. Как видно из таблицы 1 среди методов классического машинного обучения лучшим оказался простой и быстродействующий метод опорных векторов (SVM), остальные классические методы ввиду малого количества объектов и большого количества признаков, возможно, приводят к переобучению даже при использовании алгоритмов регуляризации.

2. Классические методы машинного обучения значительно проигрывают нейросетевым как по точности, так и, как правило, по быстродействию.

3. Применение мел - частотные кепстральных коэффициентов приводит к лучшему результату в данной задаче из-за их малой зависимости от отношения сигнал/шум.

4. Применение нейросетевых классификаторов позволяет уменьшить глубину колонки до 2 при распознавании выше 95%, что значительно выше корреляционных методов. Использование рекуррентных сетей при наличии контекстной информации в распознаваемых символах однозначно более выгодно.

5. Применение многоканальной обработки существенно увеличивает вероятность распознавания акустических сигналов, что особенно видно на end-to-end нейросетевых моделях. Качество распознавания аудио сигналов в этом случае составляет более 95 %, что позволяет читать набираемый текст с минимальным количеством ошибок.

6. Для дальнейшего улучшения качества распознавания акустических сигналов можно использовать метод Transfer Learning для ранее предобученных моделей (Deep Speech), либо используя гибридные модели CNN-RNN, модели seq2seq с механизмом внимания или трансформеры [1]. Недостатком этих методов является требование к большим вычислительным мощностям.

**Список литературы.**

1. А.Г. Шишкин. Методы цифровой обработки и распознавания речи. - М.: ИНФРА-М, 2023 г., 345с.

2. Лайонс Р. Цифровая Обработка сигналов, - М.; ООО Бином – Пресс, 2013 г., 656 с.

3. Feng Z. Keyboard acoustic emanations revisited. – In Proceeding of the ACM CCS, 2005 г.

4. Бишоп К.М. Распознавание образов и машинное обучение. – М.: Диалектика, 2020 г., 960 с.

5. Электронный ресурс. – URL: <https://gombru.github.io/2018/05/23/cross_entropy_loss/>

6. Электронный ресурс. – G.E.Hinton, N.Srivastava. Improving neural networks by preventing co-adaptation of feature detectors, URL: [www.arxiv.org/pdf/1207.0580.pdf](http://www.arxiv.org/pdf/1207.0580.pdf) , 2012.

7. H. Zou and T. Hastie. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser.B, 67(2):301–320, 2005.