

ФОРМИРОВАНИЕ ПРИЗНАКОВ ЗВУКОВОГО СИГНАЛА НА ОСНОВЕ МОДЕЛИ НЕРВНЫХ ИМПУЛЬСОВ ДЛЯ СИСТЕМ РАСПОЗНАВАНИЯ

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Жук И.Н.

Петровский А.А. – д.т.н., профессор

Для наиболее эффективной работы со звуком в системах классификации и автоматического распознавания речи используются модели формирования признаков. Сложность задачи заключается в том, что нам неизвестно какие признаки наиболее информативные и репрезентативные. Таким образом, ставится задача поиска и разработки наиболее эффективного способа формирования признаков звукового сигнала и предлагается способ формирования признаков, основанный на работе внутреннего уха и нейросимуляции.

Для формирования признаков звукового сигнала используются модели внутреннего уха, волосковых клеток и слухового нерва. Первым компонентом модели является банк гамма-тоновых фильтров с частотным диапазоном от 20Гц до 5.5кГц и состоящий из 200 фильтров. Затем выход каждого фильтра подвергается выпрямлению и компрессии. После этого каждый сигнал обрабатывается моделью аудиторного нерва, которая реализована с помощью метода “интегрировать и сработать” с утечками. После этого на каждой временной диаграмме фиксируются моменты времени, в которые потенциал нейрона превышает порог срабатывания и испускается импульс. Затем все импульсы для каждого сигнала объединяются в общую матрицу. Элементы матрицы могут принимать значения либо 0 - импульса нет, либо 1 - импульс есть.

Визуализация спектрограммы и матрицы нервных импульсов представлена на рисунке 1:

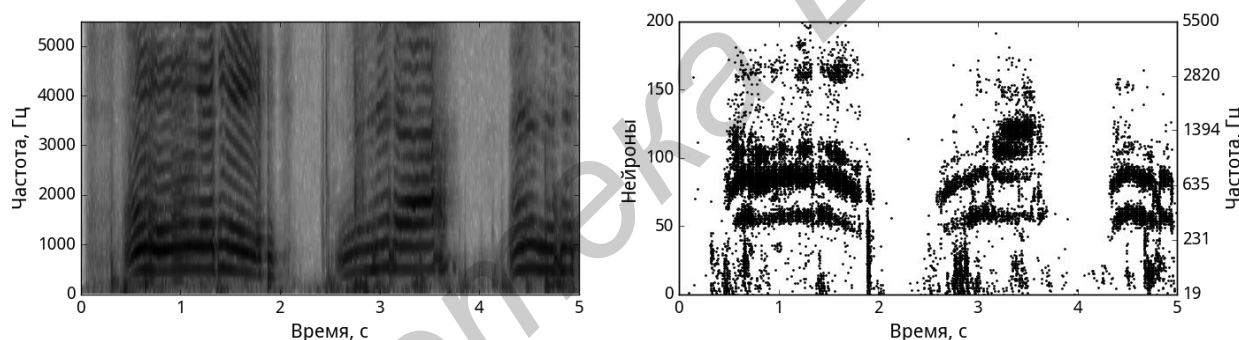


Рис. 1 - Спектрограмма и диаграмма нервных импульсов звукового сигнала

Эксперименты проводились на наборе данных ESC 10 [1], который разработан для классификации звуков окружающей среды. Он включает в себя 10 классов, в каждом из которых 40 записей. Все записи примерно одинаковой длительности 4-7 с. В итоге получилась выборка из 400 матриц размера 500x200. Архитектура сети классификатора содержит в себе 2 блока 2D сверточной сети, функции активации Rectified Linear Unit (ReLU), слои MaxPool, Dropout регуляризацию и полносвязные слои. Обучение происходит с функцией ошибки кросс-энтропия. Оптимизация весовых коэффициентов сети выполняется с помощью алгоритма adam. Используемая метрика оценки производительности - точность (accuracy).

Классификатор	Точность (accuracy)
Случайный лес с мел-кепстральными коэффициентами [1]	73.70%
Многослойный перцептрон с мел-кепстральными коэффициентами [2]	74.5%
Сверточная нейронная сеть с диаграммой нервных импульсов	85%
Производительность у человека [1]	95%

Результаты экспериментов свидетельствуют об эффективности использования признаков звукового сигнала на основе модели нервных импульсов в системах распознавания, а также необходимости дальнейших исследований в этой области.

Список использованных источников:

1. K. J. Piczak. ESC: Dataset for Environmental Sound Classification // In Proceedings of the 23rd ACM international conference on Multimedia, pp. 1015-1018, ACM, 2015.
2. Angelos Pillos, Khalid Alghamidi, Noura Alzamel, Veselin Pavlov, Swetha Machanavajhala, A real-time environmental sound recognition system for the android os // Detection and Classification of Acoustic Scenes and Events, 2016