

УДК 004.032.26

## АКУСТИЧЕСКАЯ ЛОКАЛИЗАЦИЯ ИСТОЧНИКА СИГНАЛА ПРИ ПОМОЩИ ГЛУБОКОГО ОБУЧЕНИЯ



**А. Л. Калоша**  
Магистрант БГУИР,  
инженер-программист  
JazzTeam



**М.В. Стержанов**  
Доцент БГУИР

Белорусский государственный университет информатики и радиоэлектроники,  
Республика Беларусь  
ООО «JazzTeam», Республика Беларусь  
E-mail:

### **А. Л. Калоша**

Окончил Барановичский государственный университет. Магистрант БГУИР. Сейчас работает в должности инженера-программиста. Занимается разработкой системы предиктивного анализа для классификации документов текстовых коллекций.

### **М. В. Стержанов**

Окончил БГУИР. Сейчас работает в должности доцента каф информатики. Занимается научным руководством магистрантов и аспирантов.

**Аннотация.** Цель данной работы заключается в создании системы для определения источника аудио сигнала. Для достижения данной цели использовалась нейронная сеть, которая обучена на наборе метрик, описывающих значимые характеристики звука: MFCC, спектральный центроид, скорость пересечения нуля, частоты цветности, спад спектра. Для получения данных метрики использовалась библиотека Librosa, в которой уже реализовано получение метрик из аудиофайла. Данные записывались с использованием 2 микрофонов, работающих одновременно. Обучение нейронной сети производилось на 1 000 аудиофайлов. В результате обучения нейронная сеть способна предсказать координаты источника сигнала с точностью в 81%. Улучшив качество оборудования для записи звука, размер обучающей выборки, или подобрав более точно гиперпараметры нейронной сети, можно увеличить точность системы.

**Ключевые слова:** Librosa, TensorFlow, нейронная сеть, звук.

### **Введение.**

Определение источника звука имеет важное практическое значение. Решение данной задачи играет большую роль в системах охраны, системах военного назначения или даже для отслеживания поведения животных в дикой природе. Определение источника звука может быть использовано на границах между странами для отслеживания контрабандистов, на поле боевых действий, ведь знание позиций и перемещения войск противника является критически важным. Актуальность решения данной задачи особенно повышается в условиях плохой видимости, которая может быть обусловлена ночным временем суток, туманом, бурей либо спецификой ландшафта.

Назначение разрабатываемой системы заключается в предсказании координат источника аудио сигнала. Положение источника сигнала определяется за счет разницы во времени и громкости сигнала, которой приходит на каждой из микрофонов. В этом система подобна на человека, так как мозг использует такой же механизм для определения источника сигнала.

Для обучения нейронной сети была выбрана библиотека TensorFlow как один из лучших инструментов машинного обучения. TensorFlow — это библиотека программного обеспечения с открытым исходным кодом для численного расчета с использованием графиков потока данных [1].

Нейронная сеть — это огромный распределенный параллельный процессор, состоящий из элементарных единиц обработки информации, накапливающих экспериментальные знания и предоставляющих их для последующей обработки [2].

Нейронная сеть сходна с мозгом с двух точек зрения:

Знания поступают в нейронную сеть из окружающей среды и используются в процессе обучения;

Для накопления знаний применяются связи между нейронами, называемые синаптическими весами [2].

### Архитектура нейронной сети.

На рисунке 1 проиллюстрирована архитектура нейронной сети, состоящая из 4 слоев (входной, два промежуточных и выходной слой).

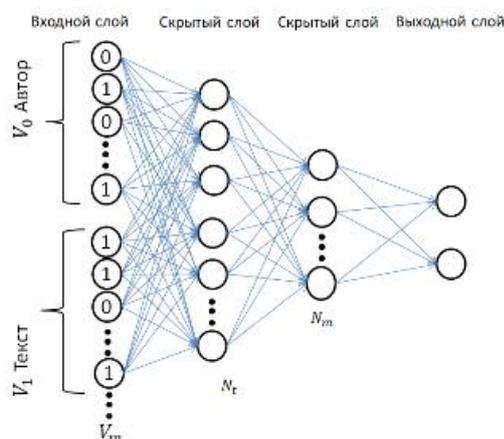


Рисунок 1. Архитектура нейронной сети

На промежуточных слоях используется функция активации ReLU, на выходном слое применяется функция sigmoid. Одной из главных проблем при обучении нейронной сети является переобучение. Когда нейронная сеть перестает обучаться, а просто запоминает ответы из обучающей выборки, для минимизации ошибки, тем самым уходя от первичной цели. Для решения этой проблемы использовалась регуляризация и dropout [3].

### Обучение нейронной сети.

В качестве локации использовалась магнитная доска с нарисованной на ней сеткой.

Квадраты на доске — это координаты, которые нейронная сеть должна предсказать. На рисунке 2 проиллюстрирована локация для сбора данных.



Рисунок 2. Локация для сбора данных

В нижней части доски находится 2 микрофона. Сверху находится камера для записи видео всего процесса сбора данных.

В ходе сбора данных магнит перемещается по доске. Моменты времени нахождения магнита в каждом из квадратов четко отмерен и взят из видео. Синхронизировав время начала на видео и на обоих микрофонах, выделены участки аудио и моменты, когда магнит находится в определенных координатах. Данные аудиозаписи разбиваются на записи в половину секунды для каждого микрофона. Это необходимо для увеличения количества данных при обучении.

Количество записей для каждого квадрата может отличаться, так как присутствует человеческий фактор. Участки с посторонним шумом по возможности удалялись.

Далее при помощи библиотеки Librosa извлекаются метрики из аудиофайлов. В векторе для обучения используются все основные метрики, которые возможно извлечь, используя данную библиотеку. К таким метрикам относятся: мел-частотные кепстральные коэффициенты (MFCC), спектральный центроид, скорость пересечения нуля, частоты цветности, спад спектра.

Таким образом, вектор для обучения состоит из 2 наборов метрик: для первого и для второго микрофона.

Обучение производилось для локации с магнитной доской. Для использования в другой локации потребуется переобучение системы. Изменив источник звукового сигнала (перемещение магнита), положение или моделей микрофонов потребует также переобучение модели нейронной сети.

Для создания универсальной системы потребуется огромное количество данных для обучения.

В данной системе использовалось 2 микрофона. Однако, увеличение количества микрофонов увеличит и точность системы в определении координат источника звука.

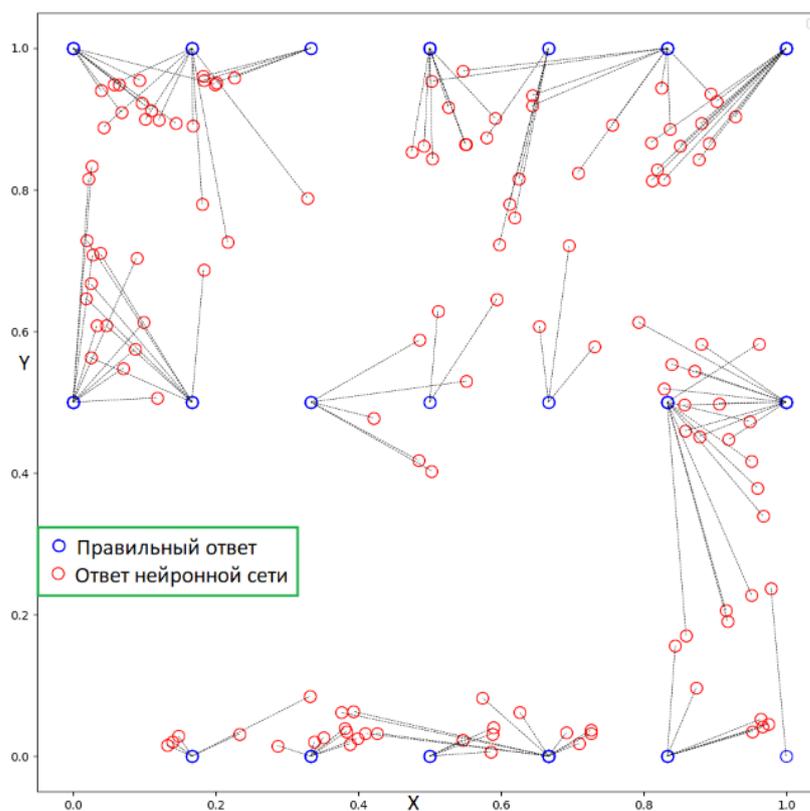


Рисунок 3. Точечная диаграмма с правильными и предсказанными координатами источника звука

### **Тестирование.**

Обучение и тестирование нейронной сети было произведено на компьютере со следующими характеристиками:

- процессор Intel Core i5-8600;
- оперативная память 48,0 ГБ;
- видеокарта Nvidia GeForce GTX 2060;
- микрофон №1: встроенный микрофон в Logitech Webcam C270
- микрофон №2: Ritmix RDM-169

Для тестирования была выбрана выборка размером 125 аудиозаписей, не участвующих в обучении.

Для каждой записи нейронная сеть делает предсказание координат источника сигнала. Согласно встроенной в TensorFlow системе расчета точности, точность обучения модели составляет 81%.

На рисунке 3 проиллюстрирована точечная диаграмма, которая отображает правильные (синие) и предсказанные нейронной сетью ответы (красные) и связи между ними.

### **Заключение.**

Обучение нейронной сети производилось на 1 000 аудиофайлов. В результате обучения нейронная сеть способна предсказать координаты источника сигнала с точностью в 81%. Улучшив оборудование для записи звука, размер обучающей выборки, или подобрав более точно гиперпараметры нейронной сети, можно увеличить точность системы. В будущем планируется провести испытания в больших масштабах, а также вне помещения.

### **Список литературы**

[1] Library for numerical computation using data flow graphs [Электронный ресурс]. – Режим доступа: <https://www.tensorflow.org/>. – Дата доступа: 15.02.2021г.

[2] Хайкин, С. Нейронные сети: полный курс, 2-е издание / С. Хайкин. – М. : Издательский дом «Вильямс», 2006. – 1104 с.

[3] Калоша, А. Л. Система анализа качества текстовых коллекций / А. Л. Калоша [и др.] // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня: сб. материалов VI Междунар. науч.-практ. конф., Минск, 20-21 мая 2020 года: в 3 ч. Ч. 2 / редкол.: В. А. Богуш [и др.]. – Минск: Бестпринт, 2020. – С. 369 – 375.

## **ACOUSTIC SOURCE LOCALIZATION USING DEEP LEARNING**

**A.L. KALOSHA**

*Master student of the BSUIR,  
software engineer JazzTeam*

**M.V. STERJANOV**

*Associate professor at BSUIR*

*Belarussian State University Informatics and Radioelectronics, Republic of Belarus*

*ООО «JazzTeam», Republic of Belarus*

*E-mail: andreikalosha@mail.ru*

**Abstract.** The purpose of this paper is to create a system for identifying the source of an audio signal. This system uses a neural network, which is trained on a set of metrics describing significant characteristics of audio: MFCC, spectral centroid, zero crossing rate, chromaticity frequencies, and spectrum decay. The Librosa library was used to retrieve metric data, which already implements retrieval of metrics from an audio file. The data was recorded using 2 microphones working simultaneously. The neural network was trained on 1,000 audio files. As a result of training, the neural network is able to predict the coordinates of the signal source with 81% accuracy. By improving the equipment for audio recording, increasing the size of the learning samples, or by selecting the hyperparameters more accurately of the neural network, you can increase accuracy of the system.

**Keywords:** Librosa, TensorFlow, neural network, sound.