



На правах рукописи

Мкртчян Грач Маратович

**Разработка методов и средств нейросетевой обработки  
акустической информации**

Специальность 2.3.8 —  
«Информатика и информационные процессы»

Автореферат  
диссертации на соискание учёной степени  
кандидата технических наук

Москва — 2025

Работа выполнена на кафедре «Математическая кибернетика и информационные технологии» ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики» (МТУСИ).

Научный руководитель: кандидат технических наук, доцент  
**Городничев Михаил Геннадьевич**

Официальные оппоненты: **Шадрин Сергей Сергеевич**,  
доктор технических наук, доцент,  
«Московский политехнический университет»,  
профессор передовой инженерной школы  
электротранспорта

**Ракитский Антон Андреевич**,  
кандидат технических наук, доцент,  
«Сибирский государственный университет телекоммуникаций и информатики»,  
доцент кафедры прикладной математики и кибернетики

Ведущая организация: Федеральное государственное бюджетное образовательное учреждение высшего образования «Поволжский государственный университет телекоммуникаций и информатики»

Защита состоится 22 мая 2025 г. в 13 часов на заседании диссертационного совета 55.2.002.02 на базе ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики» (МТУСИ) по адресу: Москва, ул. Авиамоторная 8А.

С диссертацией можно ознакомиться в библиотеке и на сайте МТУСИ:  
<https://dis.mtuci.ru/dis/02/>

Автореферат разослан \_\_\_\_\_.

Ученый секретарь  
диссертационного совета 55.2.002.02  
(приказ ректора МТУСИ от 19.02.2025  
г., №50-о)  
доктор технических наук, доцент

Юрий Владимирович Полищук

## Общая характеристика работы

**Актуальность темы.** Современные и перспективные технические системы требуют информационной поддержки, обеспечивающей обработку информации об их состоянии для принятия решений по управлению, развитию и оптимизации.

С каждым годом наблюдается значительный рост числа автотранспортных средств, увеличение загрузки дорог и возрастание интеллектуальной нагрузки на водителей при управлении транспортным средством. Эти изменения подчеркивают актуальность разработки и внедрения передовых методов и технологий обеспечения безопасности, соответствующих современным направлениям развития автотранспорта и организации дорожного движения. Одним из ключевых инструментов в этой области являются системы помощи водителю (ADAS). Однако такие системы в основном опираются на визуальные данные, поступающие с камер и лидаров. Их эффективность существенно снижается в условиях плохой видимости, неблагоприятных погодных явлений или при наличии препятствий, затрудняющих обзор.

Использование акустических данных даёт возможность анализировать текущую обстановку на дороге, идентифицируя акустические сигналы, исходящие от различных объектов и событий. Это могут быть акустические сигналы приближающихся транспортных средств, сирены экстренных служб, шумы аварийных ситуаций и другие акустические сигналы.

Современные исследования подтверждают перспективность применения акустических данных в системах безопасности. Они включают разработку методов классификации транспортных средств на основе акустических сигналов, анализ акустических сцен с использованием спектральных характеристик и технологий машинного обучения. Такие методы помогают более точно классифицировать различные дорожные ситуации и окружающую среду.

Одним из наиболее перспективных подходов в данной области является использование нейросетевых технологий. Нейронные сети демонстрируют высокую эффективность при обработке акустических данных, включая классификацию акустических сцен, распознавание транспортных средств на основе их акустических подписей и оптимизацию обработки данных за счёт снижения их размерности. Эти технологии подтверждают свою значимость и перспективность для создания интеллектуальных систем оценки дорожной обстановки и принятия решений при управлении транспортными средствами и дорожным движением.

**Степень разработанности темы исследования.** Своевременность темы подтверждается большим количеством исследований в этой

области. Задачи анализа акустических сигналов окружающей среды представлены в работах: Ю. Леженин, Н. Богач, Ю. Фурлетов, С. Шадрин, Ли, Шваб, Ашхад, Барчизеи, Шао, Море, Ибаньес-Гусман, Суноу, Перкус, Тоффа, Миньот, Нанни, Чжао, Инь, Чжан, Лю, Линь, а также Заммана и их соавторов. Эти авторы внесли значительный вклад в разработку методов и технологий анализа акустической информации, разработку алгоритмов глубокого обучения для решения задач классификации акустических сцен, транспортных средств и экологических шумов, полученные результаты могут служить основой для дальнейших исследований. Несмотря на достигнутые успехи, в области анализа акустических сигналов остаются нерешенные задачи и перспективные направления для дальнейших исследований:

- **Улучшение качества классификации в условиях акустического шума:** создание устойчивых к помехам моделей, способных эффективно работать в реальных условиях с высокой степенью фонового шума.
- **Анализ многоканальных акустических данных:** разработка методов обработки пространственных признаков, позволяющих более точно локализовать источники акустических сигналов и анализировать акустические сцены.
- **Интеграция методов мультисенсорного анализа:** комбинирование акустических данных с визуальными или вибрационными данными для повышения точности классификации.
- **Энергоэффективные алгоритмы для встроженных систем:** разработка легковесных моделей глубокого обучения, пригодных для работы на мобильных устройствах и IoT-устройствах.

Целью диссертационной работы является разработка методов и средств нейросетевой обработки акустической информации о дорожных событиях для повышения безопасности дорожного движения посредством добавления дополнительного модуля цифровой обработки сигнала в существующие системы помощи водителям.

Для достижения поставленной цели необходимо было решить следующие задачи:

1. Исследовать существующие методы и алгоритмы классификации акустических данных применительно к повышению безопасности движения транспортных средств.
2. Разработать метод сбора и аннотирования акустической информации о дорожно-транспортной обстановке.
3. Спланировать эксперимент сбора, аннотирования и исследования нейросетевых методов классификации акустических данных дорожных событий.
4. Разработать устойчивый алгоритм обучения нейронной сети в условиях выбросов и шумов в обучающем наборе данных за счёт

применения робастных функций потерь совместно с дистилляцией знаний.

5. Разработать алгоритм классификации акустических данных дорожных событий позволяющий достигать необходимой точности в рамках предметной области.
6. Разработать архитектуру программно-аппаратного комплекса сбора и цифровой обработки акустических данных дорожных событий.

**Объектом исследования** является математические и технические средства и методы анализа и классификации акустической информации дорожных событий в условиях городской среды.

**Предметом исследования** являются алгоритмическое и техническое обеспечение нейросетевой системы классификации акустической информации дорожных событий.

**Методы исследования.** Для решения указанных задач применялись методы автоматической обработки акустических данных, статистического анализа, цифровой обработки сигналов и программирования.

**Научная новизна результатов диссертации** заключается в разработке совокупности взаимосвязанных алгоритмических, программных, технических и организационных решений, направленных на повышение безопасности дорожного движения путем применения методов обработки акустической информации с использованием нейросетей.

В процессе выполнения диссертационной работы получены следующие оригинальные научные результаты:

1. Метод сбора и аннотирования акустической информации о дорожно-транспортной обстановке, *отличающийся* внедрением предобученной модели распознавания, *позволяющий* повысить скорость аннотирования данных не менее чем на 30%, а также минимизировать человеческий фактор (2.3.8, п.7).
2. Алгоритм повышения устойчивости при обучении нейронной сети, предназначенной для классификации акустических данных дорожных событий, основанный на применении робастной функции потерь совместно с дистилляцией знаний, *позволяющий* минимизировать влияние выбросов и шумов в обучающем наборе данных при добавлении до 15% зашумленных данных, без значимой потери качества (2.3.8, п.4).
3. Алгоритм классификации акустической информации дорожных событий, *отличающийся* от существующих применением слоев Колмогорова-Арнольда, *позволяющий* достигнуть точности не менее 95% в условиях городского шума (2.3.8, п.4).
4. Архитектура программно-аппаратного комплекса сбора, хранения и классификации акустической информации дорожных событий, обладающая возможностью непрерывной обработки цифрового

сигнала на борту транспортного средства, *позволяющая* интегрировать в существующие информационные системы помощи водителю дополнительный модуль цифровой обработки акустического сигнала для повышения точности определения дорожной обстановки (2.3.8, п.9).

**Теоретическая и практическая значимость** определяется возможностью повышения безопасности дорожного движения путем интеграции разработанных методов и алгоритмов классификации акустического окружения в системы помощи водителю (ADAS). Такой подход позволяет дополнить информацию от визуальных сенсоров акустическими данными, что повышает объективность оценки реальной обстановки, точность обнаружения потенциальных источников опасности, особенно в условиях плохой видимости или ограниченного поля зрения камер. Создание и испытания действующего прототипа бортовой системы обработки акустической информации позволяют сделать вывод о возможности практической реализации системы в рамках подсистемы ADAS, что может ускорить распространение и применение подобных систем на дорогах, делая вождение более безопасным и прогнозируемым. Результаты диссертационной работы могут применяться в отраслях, где требуется классификация акустических сигналов, например, для обеспечения безопасности в общественных местах, на производстве.

### **Основные положения, выносимые на защиту:**

1. Метод сбора акустической информации дорожных событий, *позволяющий* повысить эффективность подготовки набора данных и минимизировать влияние человеческого фактора, что достигается за счёт использования предобученной модели, исключающей вероятность пропуска событий из-за человеческой невнимательности или утомляемости.
2. Впервые представлен набор данных об акустической информации дорожных событий, состоящий из 5 классов общим размером 2600 образцов, собранный в реальных условиях дорожного движения.
3. Алгоритм повышения устойчивости процесса обучения нейронной сети классификации акустических данных, *позволяющий* осуществить перенос информации из крупной модели в компактную, уменьшив её размер до 0.19 млн параметров при сохранении высокой точности (не менее 92%). Это предоставляет возможность использовать модель на устройствах с ограниченными вычислительными ресурсами.
4. Алгоритм классификации акустических данных о дорожных событиях, позволяющий повысить точность компактных нейросетевых моделей не менее чем 3% в условиях зашумленной обстановки.

5. Архитектура нейросетевого программно-аппаратного комплекса сбора, хранения и обработки цифрового сигнала, позволяющего повысить безопасность передвижения транспортных средств на дорогах общего пользования за счет интеграции разработанных методов и средств обработки акустической информации в существующие информационные системы помощи водителям, тем самым при принятии решения анализируется большое количество информации.

**Степень достоверности и апробации результатов** работы обеспечиваются использованием в качестве базы современных методов и моделей, применяемых для классификации и распознавания акустических данных. Математическую основу исследования составляют адаптированные для решения поставленных задач методы теории обработки сигналов, машинного обучения, математической статистики и спектрального анализа. Результаты были представлены и обсуждались на ряде значимых международных конференций, в том числе Core A, посвящённых обработке сигналов, телекоммуникациям и применению электроники в информационных системах. Результаты работы докладывались и осуждались на Российских и международных конференциях:

- 2024 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF);
- 2024 Systems of Signals Generating and Processing in the Field of on Board Communications;
- 2023 Systems of Signals Generating and Processing in the Field of on Board Communications;
- 2023 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF);
- 2022 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO);
- 2024 INTERSPEECH: Conference of the International Speech Communication Association

Результаты также обсуждались на заседании кафедры, а также на научных семинарах в МТУСИ.

#### **Личный вклад.**

В ходе исследования автором лично были разработаны и реализованы ключевые подходы, направленные на повышение эффективности и надежности анализа акустических данных в системах помощи водителю :

- обеспечение устойчивости модели нейронной сети для классификации акустических данных, что дало возможность улучшить способность модели сохранять точность предсказаний при наличии внешних возмущений и шумов, характерных для дорожной среды;
- применение метода дистилляции для оптимизации работы модели без потери точности, что позволило уменьшить объем вычислений

- и ресурсов, необходимых для работы модели, что особенно важно для её применения в условиях ограниченных вычислительных мощностей в реальном времени;
- программно-аппаратный комплекс (прототип) для практического применения и испытания предложенной модели, включающий необходимое программное обеспечение и специализированное оборудование, что позволяет интегрировать решение в системы помощи водителю и другие приложения;
  - тестирование и оценка эффективности предлагаемых решений, как в условиях симуляции, так и в реальных условиях для оценки точности и устойчивости модели к различным внешним факторам, оценки её надежности и эффективности при различных сценариях эксплуатации.

**Реализация и внедрение.** Алгоритмы и архитектура программно-аппаратного комплекса, разработанные в настоящей работе, внедрены в следующих организациях:

- «МКАД» (ООО) (г. Гудермес) и «ЭР СИ ТЕХНОЛОДЖИС» (ООО) (г. Москва) как модуль общего комплекса оценки дорожной ситуации;
- в учебный процесс кафедры «Математическая кибернетика и информационные технологии» Московского технического университета связи и информатики (МТУСИ).

Подтверждается соответствующими актами внедрения результатов диссертационной работы.

**Соответствие специальности.** Тематика и результаты диссертации соответствуют следующим направлениям специальности: 2.3.8 — «Информатика и информационные процессы».

- п.4. «Разработка методов и технологий цифровой обработки аудиовизуальной информации с целью обнаружения закономерностей в данных, включая обработку текстовых и иных изображений, видео контента. Разработка методов и моделей распознавания, понимания и синтеза речи, принципов и методов извлечения требуемой информации из текстов» .
- п.7. «Разработка методов обработки, группировки и аннотирования информации, в том числе, извлеченной из сети интернет, для систем поддержки принятия решений, интеллектуального поиска, анализа» .
- п.9. «Разработка архитектур программно-аппаратных комплексов поддержки цифровых технологий сбора, хранения и передачи информации в инфокоммуникационных системах, в том числе, с использованием «облачных» интернет-технологий и оценка их эффективности».



**Публикации.** Основные результаты по теме диссертации изложены в 12 печатных изданиях, 3 из которых изданы в журналах, рекомендованных ВАК, 9 — в периодических научных журналах, индексируемых Web of Science и Scopus, в том числе Q2. Зарегистрированы 3 программы для ЭВМ.

## Содержание работы

Во **введении** обосновывается актуальность темы, связанной с необходимостью повышения безопасности дорожного движения и оптимизации управления автотранспортными системами. Определены объект, предмет, цель и задачи исследования, направленные на разработку методов обработки акустических данных с использованием нейросетевых технологий для интеграции в системы помощи водителю (ADAS).

Выделена научная новизна, заключающаяся в разработке алгоритмов классификации акустических данных и создании архитектуры программно-аппаратного комплекса. Практическая ценность работы заключается в применении предложенных решений для повышения безопасности дорожного движения. Кратко изложено содержание диссертации по главам: анализ существующих подходов, разработка датасета, алгоритмов и комплекса. Результаты апробированы на международных конференциях и опубликованы в научных изданиях.

В **первой главе** рассматриваются подходы к разработке программно-аппаратных решений для анализа акустических данных в дорожной среде. Обоснована необходимость создания эффективных алгоритмов классификации акустических данных, устойчивых к шумам и изменяющимся условиям. Рассматриваются современные методы анализа, включающие машинное обучение, глубокие нейронные сети и обработку аудиосигналов, а также их применение для повышения безопасности дорожного движения.

Рассмотрены существующие подходы к обработке акустических данных, включая методы предобработки сигналов, такие как фильтрация, нормализация и разбиение на временные фреймы. Также описаны способы представления акустических данных, например, спектрограммы, мел-спектрограммы и кепстральные коэффициенты (MFCC), которые позволяют извлекать ключевые характеристики сигналов для последующей классификации.

Особое внимание уделено нейросетевым методам классификации акустических данных, включая сверточные нейронные сети (CNN), трансформеры (AST, PaSST) и методы дистилляции знаний. Эти подходы позволяют обучать компактные модели-ученики, сохраняющие точность сложных моделей-учителей. Трансформеры демонстрируют свои преимущества благодаря способности учитывать глобальные взаимосвязи между

элементами аудиосигналов, что делает их особенно эффективными в задачах аудиоклассификации.

Актуальность формирования репрезентативного набора данных подчёркивается отсутствием подходящих публичных ресурсов. Выделены ключевые классы акустических событий, такие как гудки, сирены экстренных служб, акустические сигналы резкого торможения и аварийные шумы, которые имеют критическое значение для анализа дорожной обстановки. Глава завершается выводами, подтверждающими перспективность применения трансформеров и методов дистилляции знаний для задач классификации акустических данных, а также постановкой задач для дальнейшего исследования.

Во **второй главе** описывается процесс анализа существующих данных, а также эксперимент, связанный со сбором акустических данных дорожных событий и анализ нейросетевых методов классификации, применённых к этим данным. Отдельно рассматриваются вопросы организации и проведения выездов, включая выбор маршрута и условий записи, что позволило охватить широкий спектр дорожных ситуаций и акустических особенностей .

Формирование разнообразного набора акустических записей стало ключевым шагом для обучения и тестирования нейронных сетей, рассмотренных в данной главе. Был выполнен сравнительный анализ нескольких архитектур глубокого обучения, позволяющий определить наиболее эффективные подходы к автоматической классификации дорожных аудиособытий. Такой комплексный экспериментальный подход к сбору данных и их последующей обработке делает описанные методы важным инструментом дальнейших исследований.

Существует множество наборов данных, предназначенных для обучения и оценки моделей классификации акустических сигналов окружающей среды. В исследовательских целях были выбраны 3 открытых, доступных набора данных — ESC-50, UrbanSound8K, FSD50K, которые максимально охватывают те классы, которые необходимы для предметной области:

- **ESC-50** — Набор данных, состоящий из аудиозаписей продолжительностью 5 секунд, организованных в 50 классов (по 40 файлов в каждом классе), каждый класс принадлежит одной из 5 категорий.
- **UrbanSound8K** — Набор аудиоданных, содержащий 8732 помеченных акустических фрагмента длительностью до 4 секунд акустического сигнала городской среды, разделенных на 10 классов.
- **FSD50K** — Открытый набор из 51197 акустических файлов длительностью от 0.3 до 30 секунд, неравномерно распределенных по 200 классам.

Характеристики рассмотренных наборов данных представлены в таблице № 1.

Таблица 1 — Сравнение наборов данных

Наборы данных	ESC-50	UrbanSound8K	FSD50K
Классы	50	10	200
Формат данных	wav, PCM 16bit	wav, PCM 16bit	wav, PCM 16bit
Объем (шт.)	2000	8732	51197
Длительность (файла)	5 с	$\leq 4$ с	0.3–30 с
Длительность (общая)	2.7 часа	27 часов	108.3 часа
Разделение	5-секций	10-секций	train/val

Оценка алгоритмов классификации важна для определения их сильных и слабых сторон, выбора моделей для решения определенной задачи и сравнения результатов разных моделей. Ниже приведены некоторые из используемых метрик:

- **Средняя доля правильных ответов** при  $K$ -секционной кросс-валидации:

$$\text{acc}_{K\text{-fold}} = \frac{1}{K} \sum_{j=1}^K \sum_{i=1}^N \frac{y_i = f_j(x_i)}{N} \quad (1)$$

- **Средняя точность (mAP):**

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N (\text{AP}_i) \quad (2)$$

где  $N$  — количество классов объектов,  $\text{AP}_i$  — Average Precision для класса  $i$ .

в таблице № 2 представлена оценка моделей на рассмотренных наборах данных.

Таблица 2 — Оценки моделей на наборах данных UrbanSound8K, ESC-50, FSD50K

Модель	UrbanSound8K, $\text{acc}_{10\text{-fold}}$	ESC-50, $\text{acc}_{5\text{-fold}}$	FSD50K, mAP
1DCNN	0.89	0.35	-
EsResNet	0.854	0.915	-
AST	-	0.957	-
PaSST	-	0.968	0.6555
mn_40as	-	0.9745	0.656

Рассмотренные наборы данных обладают ограничениями, такими как несбалансированность и неоднородность аудиозаписей. Это требует формирования собственного набора данных для надежного обучения моделей.

Эксперимент по сбору акустических данных дорожных событий проводился в реальных условиях дорожного движения для получения достоверных данных. Планирование включало:

- Определение времени выездов: утро (8:00–10:00), день (13:00–15:00), вечер (17:00–19:00).

- Составление маршрутов в разных районах (см. таблицу № 3).
- Контроль качества оборудования перед выездами.

Таблица 3 — Пересечения улиц по маршруту сбора данных

№	Пересечение улиц
1	Авиамоторная улица
2	Солдатская улица
3	Госпитальная набережная
4	Гольяновская улица
5	Проспект Ветеранов
6	Ростокинский проезд
7	Рижская эстакада
8	Садовая-Черногрязская улица
9	Площадь Крестыанская Застава
10	Шоссе Энтузиастов
11	Улица Лапина

Процесс аннотирования акустических данных был полностью ручным и крайне трудозатратным. Разметка одного часа аудиозаписи занимала в среднем 4,7 часа, что делало работу неэффективной при больших объёмах данных. Кроме того, ручная аннотация требовала значительных ресурсов для выполнения и проверки результатов. Качество разметки снижалось из-за ошибок, вызванных монотонностью работы. Операторы теряли концентрацию, что приводило к пропуску важных акустических событий или неверному их обозначению. Такие ошибки негативно влияли на точность последующего обучения нейросетевых моделей. Эти проблемы потребовали создания автоматизированного инструмента с поддержкой предразметки и встроенными функциями контроля качества. Для разметки данных была разработана система аннотирования акустических данных – LabelTool с использованием предобученной модели BEATs, для классификации целевых акустических событий. LabelTool представляет собой распределённую систему с микросервисной архитектурой, разработанную для эффективной разметки аудиоданных. LabelTool поддерживает многоуровневую систему аутентификации и авторизации с ролевым управлением пользователями. В системе реализованы роли: Администратор, отвечающий за управление пользователями и задачами, Разметчик, выполняющий аннотацию аудиофайлов, и Верификатор, контролирующий качество разметки. Такая модель обеспечивает безопасность и чёткое распределение обязанностей. На рисунке 1 представлен пользовательский интерфейс разметки аудиоданных в Labeltool.

Благодаря внедрению системы аннотирования LabelTool, время разметки одной часовой аудиозаписи сократилось с 4,7 до 2,1 часов, что



Рисунок 1 — Пользовательский интерфейс разметки Labeltool.

привело к увеличению производительности на 55%. На рисунке 2 представлена архитектура программного обеспечения.

Система предназначена для одновременной работы нескольких пользователей, что позволяет эффективно распределять задачи и повышать производительность. Автоматизированный алгоритм назначения задач учитывает текущую загрузку пользователей и их производительность, распределяя работу равномерно. Для контроля качества применяется механизм перекрёстной проверки — один и тот же фрагмент может проверяться несколькими разметчиками, а степень перекрытия настраивается.

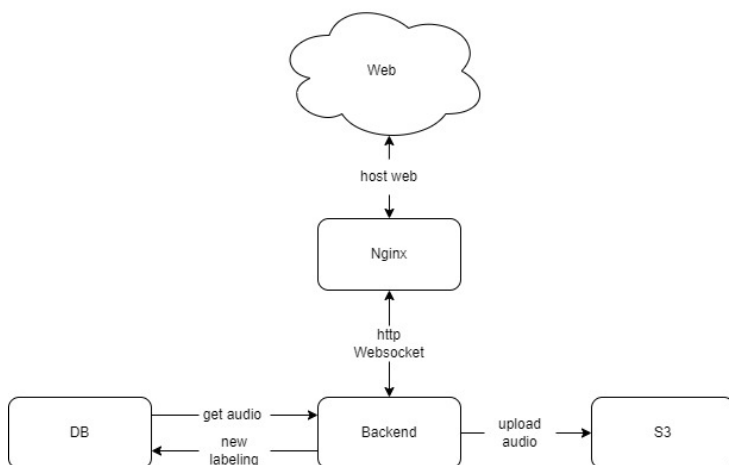


Рисунок 2 — Архитектура ПО Labeltool.

На собранных данных была проведена серия экспериментов с различными моделями для классификации акустических событий. В ходе исследования было собрано порядка **300 часов** аудиозаписей, из которых выделено **5 классов** акустических событий общей численностью **2600 семплов**. Полученные данные стали основой для обучения и сравнения эффективности различных моделей. Результаты представлены в таблице 4.

Таблица 4 — Точность моделей классификации

<b>1DCNN</b>	<b>PIPMN</b>	<b>FACE</b>	<b>EsResNet</b>	<b>BEATs</b>
47.89%	84.04%	90.23%	91.48%	97.06%

Результаты экспериментов свидетельствуют о том, что модель **BEATs** продемонстрировала наилучшую производительность на собранных данных, превосходя другие архитектуры благодаря своей способности эффективно обрабатывать сложные акустические паттерны.

В данной главе была разработана система аннотирования акустических данных дорожных событий с использованием предобученной модели, что позволило сократить время аннотирования, одновременно улучшив качество разметки. Итеративное дообучение модели обеспечило дополнительный рост точности и более тонкую адаптацию к специфике акустических дорожных событий.

В **третьей главе** рассматриваются подходы к созданию нейросетевых алгоритмов, ориентированных на классификацию акустических данных дорожной среды и удовлетворяя одновременно нескольким критериям: компактности, быстродействию и устойчивости к шумам. Решения, предложенные в рамках данной работы, нацелены на уменьшение размеров моделей и сокращение времени их работы при сохранении высокой точности.

**Проблема крупногабаритных моделей и потребности в оптимизации.** Современные высокоточные архитектуры (например, BEATs) могут содержать свыше 90 млн параметров в формате float32, что требует сотен мегабайт оперативной памяти и мощного процессора или выделенного графического ускорителя. В реальных системах (встраиваемых модулях транспортных средств, переносных устройствах) подобные требования трудно удовлетворить. В результате встаёт вопрос о создании более «лёгких» моделей, сохраняющих высокую точность распознавания акустических данных.

**Дистилляция знаний: суть и способы применения.** В работе применён метод дистилляции знаний, предполагающий перенос «знаний» от крупной, предварительно обученной модели (учителя) к более компактной архитектуре (ученику). Рассматриваются несколько подходов (основанный на ответах, на признаках, на связях).

В ходе экспериментов наиболее эффективной оказалась дистилляция, основанная на ответах: точность компактной модели превосходила 93%, при том что исходная (BEATs) демонстрировала около 97%. При этом объём памяти упрощившейся сети (MobileNetv3) составляет менее 0,2 млн параметров, что соответствует 1–2% от ресурсов, необходимых для работы BEATs.

**Учет шумов и выбросов в дорожных аудиозаписях: робастные функции потерь.** Значительная часть дорожных аудиозаписей содержит фоновый шум, накладывающиеся акустические сигналы и внедоменные примеры (речь, резонансы кузова, иные помехи), и потому стандартная перекрёстная энтропия не всегда устойчива. Включение робастных функций потерь помогает смягчать воздействие выбросов. Были протестированы:

**Биквадратная функция Тьюки (Tukey loss).** Эта функция игнорирует большие отклонения, обнуляя градиент для крупных ошибок. Она хорошо подходит для подавления выбросов, сохраняя влияние только на ошибки малой величины. Её поведение определяет пороговый параметр  $\lambda$ , после которого ошибки считаются выбросами.

Её выражение представлено в формуле (3):

$$\text{loss}(z_j, t_j) = \begin{cases} \frac{(z_j - t_j)^6}{6\lambda^2}, & |z_j - t_j| < \lambda, \\ \frac{(z_j - t_j)^4}{2\lambda^2} + \frac{(z_j - t_j)^2}{2}, & |z_j - t_j| \geq \lambda \end{cases} \quad (3)$$

где  $z$  — предсказанное значение,  $t$  — истинное значение,  $\lambda$  — порог. График функции Тьюки показан на рисунке 3.

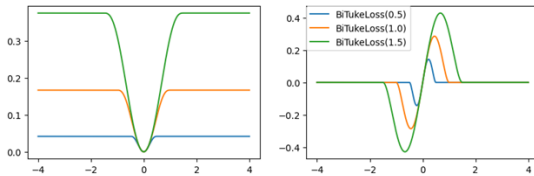


Рисунок 3 — Биквадратная функция потерь Тьюки и её производная.

**Функция Хьюбера (Huber loss).** Функция Хьюбера представляет собой гибрид квадратичной и линейной функции: квадратичная для малых ошибок и линейная для больших. Это снижает влияние выбросов, но позволяет модели учитывать их в процессе обучения. Переключение между режимами происходит при пороге  $\lambda$ .

Её выражение приведено в формуле (4):

$$\text{loss}(z_j, t_j) = \begin{cases} \frac{1}{2}(z_j - t_j)^2, & \text{если } |z_j - t_j| \leq \lambda, \\ \lambda|z_j - t_j| - \frac{1}{2}\lambda^2, & \text{если } |z_j - t_j| > \lambda. \end{cases} \quad (4)$$

где  $\lambda$  — порог между квадратичной и линейной частью. График функции Хьюбера приведён на рисунке 4.

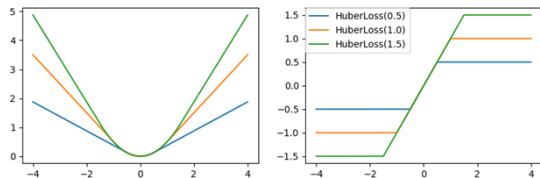


Рисунок 4 — Функция потерь Хьюбера и её производная.

**Функция Эндрюса (Andrews loss).** Эта функция эффективно подавляет влияние выбросов благодаря косинусному сглаживанию для больших отклонений. Для ошибок меньше порога  $\pi\lambda$  функция изменяется по косинусному закону, а для больших ошибок остаётся постоянной. Она хорошо подавляет шумы, сохраняя устойчивость модели.

Её выражение показано в формуле (5):

$$\text{loss}(z_j, t_j) = \begin{cases} \lambda \left(1 - \cos\left(\frac{z_j - t_j}{\lambda}\right)\right), & \text{если } |z_j - t_j| < \pi\lambda, \\ 2\lambda, & \text{если } |z_j - t_j| \geq \pi\lambda. \end{cases} \quad (5)$$

График функции Эндрюса приведён на рисунке 5.

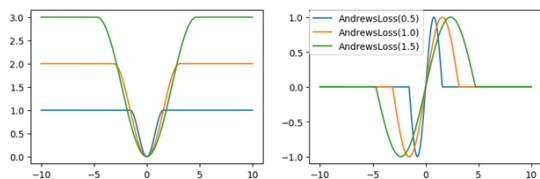


Рисунок 5 — Функция потерь Эндрюса и её производная.

### Результаты применения робастных функций потерь

Модель обучалась на данных дорожных событий при различных уровнях выбросов: 5%, 10%, 15%. Точность на валидационной выборке приведена в таблице 5.

Таблица 5 — Точность модели при разных функциях потерь и выбросов

Функция потерь	5%	10%	15%
Биквадратная Тьюки	<b>0.928</b>	<b>0.905</b>	<b>0.905</b>
Хьюбера	0.882	0.892	0.863
Эндрюса	0.882	0.892	0.892



При сравнительных тестах с разной долей искусственно добавленных «лишних» аудиозаписей вне целевых классов (5–15 %) самая высокая устойчивость продемонстрирована у модели с функцией Тьюки. Тогда как при 5% выбросов точность у всех робастных функций потерь проседала незначительно, при увеличении шума до 10–15% преимущество биквадратной функции Тьюки стало очевидным: снижение метрик не превышало нескольких десятых процента.

**Улучшение способности к обобщению: сети Колмогорова–Арнольда (KAN).** Для дополнительного повышения точности изучались сети KAN — архитектуры, основанные на теореме Колмогорова–Арнольда, гласящей, что любую многомерную непрерывную функцию можно представить в виде суперпозиции одномерных функций. Такая композиция даёт возможность эффективно аппроксимировать сложные карты признаков аудиосигнала.

KAN основаны на теореме Колмогорова — Арнольда, согласно которой любую многомерную непрерывную функцию можно представить как:

$$f(X) = \sum_{q=1}^{2n+1} \sum_{p=1}^n \psi_{q,p}(x_p), \quad (6)$$

где  $\psi_{q,p}$  — одномерные функции.

В данном исследовании модель MobileNetv3 была модифицирована путём замены финальных слоёв на блоки KAN, которые применяют В-сплайны для гладкой аппроксимации. **Модификации KAN:**

1. **Сжатие входов:** MNv3 использована для уменьшения размерности признаков.
2. **Использование В-сплайнов** для гладких активаций:

$$B_{i,0}(x) = \begin{cases} 1, & t_i \leq x < t_{i+1}, \\ 0, & \text{иначе,} \end{cases} \quad (7)$$

$$B_{i,k}(x) = \frac{x - t_i}{t_{i+k} - t_i} B_{i,k-1}(x) + \frac{t_{i+k+1} - x}{t_{i+k+1} - t_{i+1}} B_{i+1,k-1}(x), \quad (8)$$

3. **Перенос активаций на рёбра** с использованием взвешенных В-сплайнов.

Итоговый слой KAN:

$$x_{l,p} = \sum_{q=1}^{r_{l-1}} \psi_{l,q,p}(x_{l-1,q}). \quad (9)$$

**Результаты:** После замены финальных слоёв MNv3 на KAN и обучения с аугментацией и дистилляцией знаний из BEATs достигнута точность 95,11%. Результаты выбранных моделей представлены в таблице 6.

Таблица 6 — Результаты выбранных моделей

KAN мод. + дистилляция	Дистилляция	MLP
0.95	0.93	0.91

Этот подход дал прибавку в точности приблизительно до 95,1%, что на 1–2% выше, чем у аналогичного варианта без KAN. Дополнительные эксперименты показали, что оптимальное количество В-сплайнов, а также глубина слоёв KAN зависят от характера обучающего корпуса: при большом разнообразии шумов повышение гладкости активаций помогает сглаживать выбросы.

Тесты продемонстрировали, что систематическое сочетание дистилляции знаний, робастных функций потерь и KAN даёт синергетический эффект. Модель способна работать на CPU и GPU с ограниченными ресурсами (например, на встроенных модулях), сохраняя хорошую точность и иммунитет к выбросам.

Кроме того, разработанные подходы обобщаются на различные сценарии, выходящие за пределы дорожной тематики. Представленные в главе методы делают возможным надёжное и экономичное решение задачи распознавания критически важных дорожных событий (сигналы аварий, сирены, акустические сигналы столкновений), обеспечивая при этом удовлетворение требования к устойчивости в условиях городского шума и ограничениям по вычислительным ресурсам.

В **четвертой главе** изложен комплексный подход к проектированию и разработке аппаратно-программного комплекса для сбора и классификации акустических событий на дороге. Рассматривается задача выбора аппаратной основы, топологии микрофонного массива, методов предварительной и нейросетевой обработки, а также приводятся результаты теоретических и практических испытаний системы.

В систему аннотирования данных необходимо импортировать синхронизированную пару аудио и видео. Для синхронизации аудио- и видеопотоков используется метод кросс-корреляции. Описанная система позволяет эффективно собирать акустические данные для последующей классификации дорожных событий в реальном времени.

Ключевым элементом комплекса является структура из восьми микрофонов, расположенных равномерно по окружности. Это обеспечивает равномерное покрытие по азимуту, что упрощает определение направления на акустическое событие и облегчает расчёт задержек для алгоритмов формирования диаграммы направленности. В качестве управляющей платформы используется микроконтроллер STM32 NUCLEO-L476RG. Архитектура программно-аппаратного комплекса представлена на рисунке 6.

**Предобработка сигналов и формирование диаграммы направленности.** Для усиления сигнала, приходящего из заданного направления, в комплексе применяется алгоритм *Delay And Sum (DAS)*. На основе

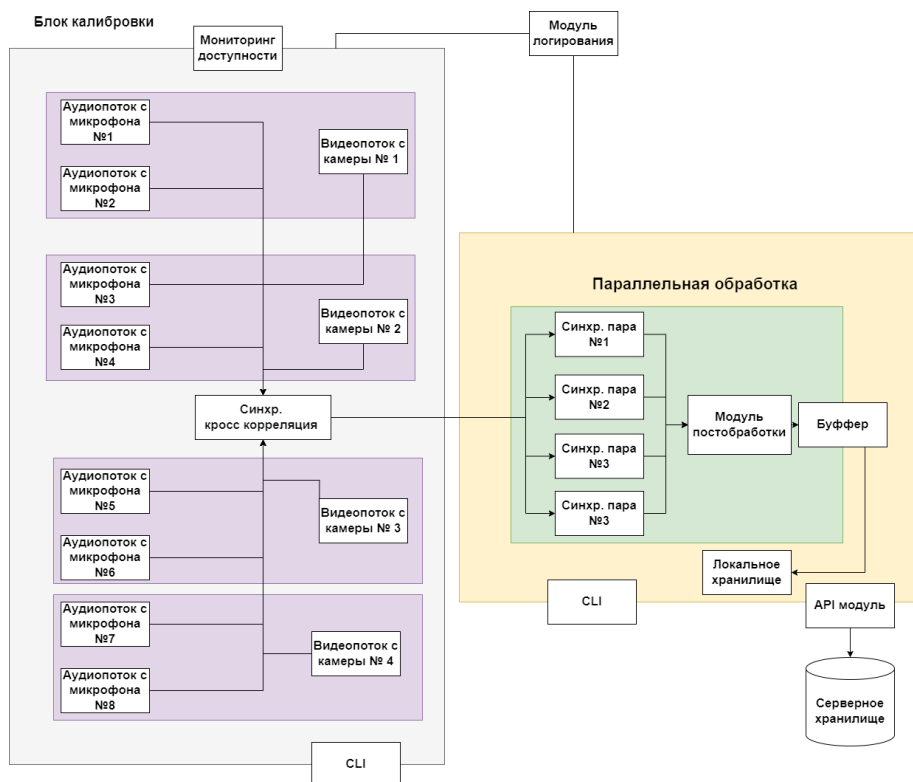


Рисунок 6 — Архитектура программно-аппаратного комплекса сбора данных.

известной геометрии массива и предположения о приблизительно параллельном фронте волны рассчитываются временные задержки для каждого микрофона. Эти задержки компенсируют различия во времени прихода сигнала, что позволяет складывать фазы акустических волн когерентно, усиливая сигналы из определённого направления и подавляя шумы и сигналы от других углов.

Рассмотрены симуляции и практические испытания системы классификации акустических событий с использованием микрофонного массива и алгоритма DAS.

**Классификация аудиозаписей после DAS:** Для непрерывной и параллельной обработки аудиоданных используется сегментация на окна длительностью  $T$  с перекрытием  $P$ . Вычислительно затратные процедуры формируют несколько лучей DAS (каждый соответствует своему углу), а получившиеся сигналы поступают в нейросетевой классификатор. На рисунке 7 представлено распределение вероятностей по классам.

**Замеры времени обработки:**

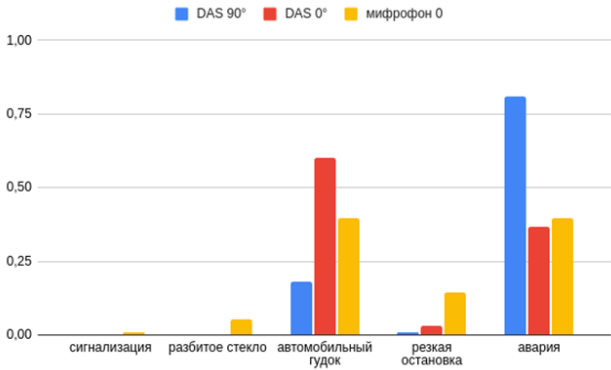


Рисунок 7 — Распределение вероятностей классов.

- Предобработка:  $\bar{T} = 56.64$  мс,  $J = 0.31$  мс
- Инференс:  $\bar{T} = 159.93$  мс,  $J = 0.25$  мс
- Постобработка:  $\bar{T} = 7.55$  мс,  $J = 0.08$  мс

Общая латентность:  $T_{\text{total}} = 224.12$  мс (ниже дедлайна 250 мс), общий джиттер:  $J_{\text{total}} = 0.64$  мс.

Система эффективно классифицирует акустические события в реальном времени, сохраняя точность и соответствуя дедлайну 250 мс.

**Испытания и результаты.** *Теоретические испытания* включали симуляцию источников звука на разном расстоянии от микрофонного массива. *Практические испытания* проводились в черте города, включая загруженные перекрёстки и тихие улицы. Зафиксировано несколько типов акустических событий: наиболее часто встречались гудки, для которых точность достигала 85 %, а для «сигнализации» 100 %. Реже возникали «резкие остановки» (66 %) из-за схожести с другими тормозными шумами и меньшего количества обучающих примеров. Результаты показывают, что система способна уверенно распознавать распространённые дорожные звуки.

В **заклучении** приведены основные результаты работы, которые заключаются в следующем: В рамках данного исследования была разработана и реализована комплексная система акустического обнаружения и классификации акустических событий для применения в городской среде. Работа охватила широкий спектр задач: от сбора и обработки акустических данных до разработки и оптимизации алгоритмов машинного обучения и их практического применения в реальных условиях.

Ключевые достижения исследования включают:

1. Изучены методы классификации акустических данных с целью повышения безопасности движения транспортных средств. Проведён глубокий анализ существующих алгоритмов, таких как методы

на основе свёрточных нейронных сетей, архитектуры на базе трансформеров и традиционные статистические подходы. Это позволило выявить преимущества и ограничения каждого метода и выбрать наиболее эффективные подходы для дальнейших исследований. **(Задача № 1)**

2. Предложен метод сбора и аннотирования акустической информации дорожных событий, позволяющий повысить эффективность подготовки набора данных и минимизировать влияние человеческого фактора, что достигается за счёт использования предобученной модели, исключающей вероятность пропуска событий из-за человеческой невнимательности или утомляемости. **(Задача № 2)**
3. Разработана система сбора и аннотирования данных о дорожных событиях в реальных условиях городской среды, что позволило сформировать уникальный набор данных. Система включает микрофонную решётку из 8 микрофонов, размещённых по периметру транспортного средства, с обеспечением 360° покрытия камеры и оборудование для синхронизации данных. Также была создана специальная система разметки данных, основанная на предобученной модели BEATs, что минимизировало вероятность пропуска событий из-за человеческого фактора. Набор данных, включающий 5 классов акустических сигналов и состоящий из 2600 сэмплов, стал основой для выявления новых закономерностей в аудиоданных и значительного повышения точности классификации. Внедрение инструмента LabelTool позволило сократить время разметки почти вдвое и повысить качество аннотаций. **(Задача № 3, Положения № 1, 2)**
4. Разработано алгоритмическое обеспечение нейросетевой обработки акустических данных, включающее алгоритм устойчивого обучения нейронной сети и алгоритм классификации акустических данных дорожных событий. Алгоритм устойчивого обучения нейронной сети учитывает выбросы и шумы в данных. Применение робастных функций потерь, таких как функции Хьюбера и биквадратная функция Тьюки, обеспечило устойчивость моделей к аномалиям. Использование метода дистилляции знаний позволило значительно уменьшить размер модели с 90,3 млн до 0,19 млн параметров, сохранив точность классификации (92%). Это открывает возможность применения модели на устройствах с ограниченными вычислительными ресурсами, таких как транспортные системы. Разработанный алгоритм классификации акустических данных дорожных событий с применением модификаций KAN, обеспечивает точность не менее 95% в условиях городской среды. Алгоритм демонстрирует устойчивость к шумам и сложным акустическим условиям благодаря использованию современных нейронных сетей

на основе архитектур трансформеров. Проведён сравнительный анализ архитектур нейронных сетей, в результате которого последняя модель показала наилучшую точность на собранном наборе данных. (**Задача № 4,5, Положение № 3,4**)

5. Разработана архитектура и реализован комплекс для сбора и цифровой обработки акустических данных, включающий специализированное оборудование и программное обеспечение. Реализация метода Delay And Sum (DAS) для формирования направленной диаграммы чувствительности микрофонной решётки позволила локализовать источники акустического сигнала. Оптимизация алгоритмов обработки в режиме низкой задержки обеспечила непрерывность классификации и возможность обработки нескольких направлений одновременно. Разработанная архитектура программно-аппаратного комплекса включает сервер приёма данных, базу данных PostgreSQL и распределённое хранилище аудиозаписей на базе Ceph. Практические испытания комплекса в городских условиях подтвердили его эффективность: точность классификации ряда акустических событий достигла не менее 95%. (**Задачи № 6, Положение № 5**)

В заключение можно сказать, что проведённое исследование вносит значительный вклад в развитие технологий акустического мониторинга и открывает новые возможности для создания интеллектуальных систем управления и безопасности. Разработанная система демонстрирует высокий потенциал для практического применения и развития, что может существенно повлиять на уровень безопасности и эффективность управления городской инфраструктурой.

Предложенные методы и средства, полученные в рамках диссертационного исследования создают основу для дальнейших инноваций в области акустического анализа и мониторинга. Объединение технологий машинного обучения, обработки сигналов и распределённых вычислений открывает перспективы для создания более умных, безопасных и экологически устойчивых городских пространств будущего.

## **Основные публикации по теме диссертации**

### **Работы, опубликованные в изданиях из перечня ВАК РФ**

1. *Мкртчян, Г. М.* Обзор методов классификации звуков городской среды [Текст] / Г. М. Мкртчян, Н. А. Кравченко // Вестник Рязанского государственного радиотехнического университета. — 2024. — (**Перечень рецензируемых научных изданий № 730 от 09.12.2024**).

2. *Мкртчян, Г. М.* О задаче применимости сетей Колмогорова-Арнольда в легковесной нейросетевой модели классификации акустических данных [Текст] / Г. М. Мкртчян // Экономика и качество систем связи. — 2024. — **(Перечень рецензируемых научных изданий № 1728 от 09.12.2024)**.
3. *Мкртчян, Г. М.* Сравнение методов акустического детектирования, классификации и локализации для решения задач обнаружения событий и сцен [Текст] / Г. М. Мкртчян, М. Г. Городничев, А. В. Тимчук // Экономика и качество систем связи. — 2024. — **(Перечень рецензируемых научных изданий № 1728 от 09.12.2024)**.

#### **Статьи, индексируемые в международных базах WoS/Scopus**

4. *Mkrtchian, G.* Capsule-based and TCN-based Approaches for Spoofing Detection in Voice Biometry [Текст] / G. Mkrtchian, K. Borodin, [ др.] // Engineering, Technology amp; Applied Science Research. — 2024. — **(Журнал Scopus Q2)**.
5. *Mkrtchian, G.* AASIST3: KAN-Enhanced AASIST Speech Deepfake Detection using SSL Features and Additional Regularization for the ASVspoof 2024 Challenge [Текст] / G. Mkrtchian, K. Borodin, [ др.] — 2024. — **(Core A Interspeech 2024)**.
6. *Mkrtchian, G.* About Choosing the Audio Signal Representation in the Task of Vehicle Pattern Recognition [Текст] / G. Mkrtchian, M. Gorodnichev, S. Simonov // . — 2023 Wave Electronics, its Application in Information, Telecommunication Systems (WECONF), 2023.
7. *Mkrtchian, G.* On the Task of Localization of Road Infrastructure Objects Based on Audio Signals [Текст] / G. Mkrtchian, S. Simonov, [ др.] // . — 2023 Wave Electronics, its Application in Information, Telecommunication Systems (WECONF), 2023.
8. *Mkrtchian, G.* Intelligent Detection and Object Localization System based on Acoustic Data [Текст] / G. Mkrtchian, N. Kravchenko, A. Timchuk // . — 2024 Systems of Signals Generating, Processing in the Field of on Board Communications, 2024.
9. *Mkrtchian, G.* On the Use of an Acoustic Sensor in the Tasks of Determining Defects in the Roadway [Текст] / G. Mkrtchian, K. Polyantseva // . — 2024 Systems of Signals Generating, Processing in the Field of on Board Communications, 2024.
10. *Mkrtchian, G.* Research and Development of a Sound Pattern Classifier in Complex Urban Acoustic Environments [Текст] / G. Mkrtchian, M. Gorodnichev, [ др.] // . — 2024 Wave Electronics, its Application in Information, Telecommunication Systems (WECONF), 2024.

11. *Mkrtchian, G.* Classification of Environmental Sounds Using Neural Networks [Текст] / G. Mkrtchian, Y. Furletov // . — 2022 Systems of Signal Synchronization, Generating, Processing in Telecommunications (SYNCHROINFO), 2022.
12. *Mkrtchian, G.* On the Task of Classifying Sound Patterns in Transport [Текст] / G. Mkrtchian, M. Gorodnichev, Y. Furletov // . — 2023 Systems of Signals Generating, Processing in the Field of on Board Communications, 2023.

### **Свидетельства о государственной регистрации программ для ЭВМ**

13. *Свидетельство о гос. регистрации программы для ЭВМ.* AudioHarvest — программный комплекс для сбора аудио данных [Текст] / Г. М. Мкртчян ; Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики». — Оpubл. 13.12.2024, 2024690295.
14. *Свидетельство о гос. регистрации программы для ЭВМ.* LabelSpeech — программный комплекс, предназначенный для аннотации различных видов данных, включая аудио, видео, текст и изображения [Текст] / Г. М. Мкртчян ; Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики». — Оpubл. 13.10.2023, 2023681411.
15. *Свидетельство о гос. регистрации программы для ЭВМ.* RID Acoustic Model — система аудиодетекции дорожных событий в реальном времени [Текст] / Г. М. Мкртчян ; Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики». — Оpubл. 16.12.2024, 2024690558.