

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.ЧЕРНЫШЕВСКОГО»

Кафедра информатики и программирования

**РАЗРАБОТКА ПРИЛОЖЕНИЯ ГОЛОСОВОЙ ИДЕНТИФИКАЦИИ С
РАСПОЗНАВАНИЕМ ЭМОЦИЙ
АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ**

студента 4 курса 441 группы

направления 02.03.03 Математическое обеспечение и администрирование
информационных систем

факультета компьютерных наук и информационных технологий

Василенко Ивана Александровича

Научный руководитель:

доцент, к.э.н., доцент

Кабанова Л.В.

подпись, дата

Зав. кафедрой:

Ученая степень, звание

Огнева М.В.

подпись, дата

Саратов 2025

ВВЕДЕНИЕ

Актуальность темы. В современном мире голосовая биометрия становится все более распространенным способом аутентификации: автоматические системы идентификации говорящего используются в банковских сервисах, мобильных приложениях и системах контроля доступа. Их популярность объясняется удобством, бесконтактностью и возможностью удаленной верификации личности без необходимости использования физических носителей информации. Одновременно с этим прогресс нейросетевых методов синтеза речи позволяет создавать высококачественные аудиозаписи, имитирующие чье-либо произношение. Например, современные генеративные модели позволяют воспроизводить голос человека по коротким фрагментам записи, фактически копируя тембр говорящего и интонационные особенности. Такие модели способны генерировать реалистичные фразы, которые трудно отличить от оригинала, особенно при поверхностном аудиоанализе. Это уже приводило к реальным инцидентам мошенничества, когда голосовые deepfake-технологии использовались для обмана как частных лиц, так и корпоративных структур. Подобное создает серьезную угрозу для безопасности голосовых биометрических систем, подрывая доверие к ним и требуя разработки дополнительных уровней защиты, включая многофакторную аутентификацию и алгоритмы выявления синтетической речи.

Традиционные голосовые биометрические системы основаны на анализе акустических признаков: например, тембра, тона и ритма речи. Однако эти параметры меняются под влиянием эмоционального состояния говорящего (стресса, гнева, волнения и т. п.). В результате при проверке речи в эмоциональном отклонении сравниваются нерегулярные характеристики и системы идентификации говорящего дают сбои. Системы проверки говорящего часто страдают снижением точности при наличии речи с не нейтральной интонацией. Игнорирование этой изменчивости приводит к ухудшению качества идентификации и несправедливому отказу от

аутентификации у эмоциональных пользователей. Поэтому современные исследователи идентификации по голосу подчеркивают: учет эмоций в голосе — важный аспект, который может значительно повысить надежность биометрии. Он открывает новые возможности: анализ эмоциональных примитивов речи (например, уровня активации и валентности) позволяет компенсировать изменения говорящего и улучшить устойчивость системы к вариациям голоса.

Учитывая вышесказанное, очевидна необходимость комбинированного подхода: сочетать голосовую идентификацию с анализом эмоций. Совместная оценка вокальных характеристик и эмоционального контента сделает взлом биометрии значительно сложнее. Если поддельный голос звучит неестественно с эмоциональной точки зрения, система сможет это обнаружить. Такой подход позволит повысить надежность аутентификации за счет дополнительного уровня проверки и улучшит устойчивость к синтетическим голосовым атакам.

Таким образом разработка программы для голосовой идентификации пользователей с использованием распознавания эмоций является актуальной задачей.

Цель бакалаврской работы – разработка приложения для голосовой идентификации пользователей с использованием распознавания эмоций.

Поставленная цель определила **следующие задачи**:

1. Разобрать существующие аналоги.
2. Дать теоретическое обоснование применению голоса для идентификации.
3. Дать теоретическое обоснование влиянию эмоций на голос.
4. Дать определение нейронных сетей.
5. Описать принцип работы нейронных сетей.
6. Описать процесс обучения нейронных сетей.
7. Описать виды архитектур нейронных сетей.
8. Дать описание реализуемого функционала приложения.

9. Выбрать технологии для реализации приложения и объяснить их выбор.
10. Реализовать серверную часть приложения.
11. Реализовать клиентскую часть приложения.
12. Протестировать работу приложения.

Методологические основы голосовой биометрии представлены в работах Плешакова Е. С, Катмакова П. С, Бахметьева П. В.

Теоретическая и/или практическая значимость бакалаврской работы. Практической значимостью данной работы является демонстрация перспективной идеи защиты голоса от копирования с помощью добавления проверки речи на эмоции.

Структура и объём работы. Бакалаврская работа состоит из введения, 2 разделов, заключения, списка использованных источников и 13 приложений. Общий объем работы – 70 страниц, из них 46 страниц – основное содержание, включая 37 рисунков, список использованных источников информации – 23 наименования.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Первый раздел «Теоретические основы голосовой идентификации» посвящен теоретическому наполнению бакалаврской работы. В нем присутствуют 4 подразделы, которые дают всю необходимую для работы теорию:

Подраздел «1.1 Разбор существующих аналогов» предоставляет всю необходимую информацию о существующих аналогах реализуемого приложения.

Picovoice — компания, разрабатывающая и внедряющая различные системы и решения, связанные с машинной обработкой речи. Одним из известных продуктов является Eagle Speaker Recognition. Это программное обеспечение способно верифицировать и идентифицировать говорящего, отличая людей по уникальным характеристикам их голоса, вне зависимости от

языка. Система работает на основе технологий глубокого обучения, что обеспечивает высокую точность идентификации и устойчивость к вариациям речи.

Nuance Gatekeeper — биометрическая система безопасности, которая анализирует звуковой поток с использованием технологий глубоких нейронных сетей. Система Gatekeeper отличается многофункциональностью и высоким уровнем безопасности. Использование методов глубокого обучения и расширенной биометрии позволяет системе достигать высокой точности в обнаружении аномалий и идентификации пользователя. Применение технологии Nuance Gatekeeper актуально в финансовом секторе, где требуется усиленная защита от мошенничества, а также в государственных учреждениях. Недостатком может быть высокая стоимость лицензии и необходимость мощных вычислительных ресурсов для ее поддержки.

NICE Omilia — облачная платформа от компании NICE, используется для голосовой аутентификации пользователей в реальном времени. Она интегрируется с виртуальными агентами, что делает ее подходящей для применения в контакт-центрах. В процессе идентификации система сопоставляет голосовые отпечатки говорящих с зарегистрированными ранее образцами.

Verint Voice Biometrics — система пассивной голосовой биометрии от компании Verint представляет собой систему пассивной голосовой биометрии, автоматически идентифицирующую звонящих по их голосовым отпечаткам. Основной сферой применения являются контакт-центры, где система создает и хранит уникальные записи голосов клиентов для последующего сравнения.

Amazon Alexa Voice Profiles — голосовой помощник, имеющий функцию Voice Profiles, которая идентифицирует говорящего с использованием заранее обученных голосовых профилей.

Google Assistant Voice Match — Google Assistant обладает схожей с Amazon Alexa функцией Voice Match, которая также позволяет

идентифицировать пользователей на основе их уникальных голосовых отпечатков.

Подраздел «1.2 Голос как механизм идентификации» посвящен параметрам голоса: высоте тона, тембру, ритму речи, интонации, спектральным характеристикам и почему их можно использовать для идентификации. Приведен детальный анализ голосовых параметров и их особенностей.

Подраздел «1.3 Влияние эмоций на голос» подробно описывает влияние разных эмоциональных состояний человека на следующие характеристики голоса: акустические параметры, темп, интонационные паттерны.

Подраздел «1.4 Нейронные сети» посвящен теории нейронных сетей. Сам раздел достаточно объемный, поэтому он поделен ещё на несколько подразделов для улучшения структуры работы. В начале раздела приведена основа нейронных сетей: введено понятие искусственного нейрона и слоев из этих нейронов. Остальное подробнее описано в подразделах.

Подраздел «1.4.1 Принцип работы нейронных сетей» подробным образом описывает принцип работы нейронных сетей. Математически определен процесс обработки нейронами входных данных. Введено понятие функции активации и дано объяснение, зачем она нужна.

В Подразделе «1.4.2 Обучение нейронных сетей» описан принцип обучения нейронных сетей, состоящий из следующих последовательных этапов: подбор гиперпараметров, подготовка и предварительная обработка данных, прямой проход, оценка ошибки и контроль качества, обратное распространение ошибки, обновление параметров и оптимизация.

В Подразделе «1.4.3 Архитектуры нейронных сетей» описываются архитектуры нейронных сетей, пригодные для идентификации по голосу и распознавания эмоций. Рассмотрены такие архитектуры как: полносвязные нейронные сети, сверточные нейронные сети, рекуррентные нейронные сети, трансформеры.

Таким образом данный раздел предоставляет всю необходимую теория для реализации приложения для голосовой идентификации с распознаванием эмоций.

Второй раздел «Разработка приложения» посвящен поэтапной реализации приложения. Данный раздел поделен на подразделы для структурирования работы.

В Подразделе «2.1 Задача реализуемого приложения» подробным образом изложено, какую задачу должно выполнять приложение, из каких частей оно будет состоять и как будет происходить проверка речи с использование распознавания эмоций.

В Подразделе «2.2 Технологии реализации» описаны технологии, что будут использованы для реализации приложения с подробные пояснением, почему именно они были выбраны. Список технологий следующий: язык программирования — Python, архитектура нейронных сетей — сверточные нейронные сети, решение для реализации нейросетей — Python фреймворк PyTorch, технологии для реализации клиентской части — HTML, CSS, JavaScript.

В Подразделе «2.3 Серверная часть приложения» описан процесс реализации серверной части приложения. Реализованы следующие компоненты: загрузчик данных — загружает аудиофайлы и создает наборы для обучения нейронных сетей, преобразователь аудио в датасет признаков — выделяет аудиопризнаки из аудиофайлов и преобразовывает их в датасеты для обучения нейронных сетей, аугментатор — опциональная программа, которая работает в случае, если требуется расширить набор обучающих данных, модель-каркас — модель с общей логикой моделей. Выполняет вспомогательную роль и служит для облегчения модернизации, отладки и читаемости исходного кода проекта, программа обучения нейронных сетей, программа валидации обучения нейронных сетей, конфигурация — в конфигурационном файле задаются ключевые параметры работы серверной

части приложения, например: использовать аугментацию или нет, список эмоций, пути загрузки аудиофайлов и т. д.; обработчик запросов от клиентской части — обработчик запросов принимает запросы от клиентской части, посылает полученные данные на обработку другим компонентам серверной части и возвращает ответы обратно на клиентскую часть, модель и нейронная сеть для распознавания эмоций, модель и нейронная сеть для идентификации по голосу.

В Подразделе «2.4 Клиентская часть приложения» описан процесс реализации клиентской части приложения. Реализованы следующие компоненты: HTML страница, CSS стили для HTML страницы, JavaScript скрипты для взаимодействия клиентской части с серверной.

В Подразделе «2.5 Демонстрация работы» приведена демонстрация работы приложения с разными сценариями проверки.

Таким образом данный раздел поэтапно описывает процесс создания приложения с дальнейшей демонстрацией работы приложения.

ЗАКЛЮЧЕНИЕ

В данной работе была приведена разработка приложения для идентификации пользователей с применением распознавания эмоций.

В теоретической части были разобраны существующие аналоги, дано теоретическое обоснование применению голоса в качестве средства идентификации и обосновано влияние эмоций на голос. Дано подробное описание такой науки, как машинное обучение, обосновано ее использование и описаны подходящие модели для решения задач идентификации по голосу и распознавания эмоций. Рассказано о нейронных сетях, что это такое, каков принцип их работы, как они обучаются и какие архитектуры нейронных сетей бывают.

В практической части приведен функционал приложения. Изложены используемые технологии с объяснением их выбора. Далее поэтапно

реализованы серверная и клиентская части, а также связь между ними. После чего продемонстрирована работа готового приложения.

Основные источники информации:

1. Нургельдиева М. ГОЛОСОВАЯ БИОМЕТРИЯ: ПОТЕНЦИАЛ И РИСКИ ИСПОЛЬЗОВАНИЯ В ПОВСЕДНЕВНОЙ ЖИЗНИ / С. Бегполодов, М. Гурбангылыджов // Вестник науки. - 2024. - № 11. - С. 943-946.
2. Бахметьев П. В. ОХРАНА ИНДИВИДУАЛЬНОГО ГОЛОСА ЧЕЛОВЕКА: ОТЕЧЕСТВЕННЫЙ И ЗАРУБЕЖНЫЙ ОПЫТ // Право и практика. - 2023. - № 4. - С. 184-188.
3. Катмаков, П. С. Биометрия: учебное пособие для вузов / П. С. Катмаков, В. П. Гавриленко, А. В. Бушов. — М. : Юрайт, 2024. — 186 с.
4. Плешакова Е. С. Распознавание эмоций человека по голосу в борьбе с телефонным мошенничеством / С. Т. Гатауллин, А. В. Осипов, М. В. Коротеев // Национальная безопасность. - 2022. - № 5. - С. 11-29.
5. Xu, Y. Human voice sensing through radio frequency technologies: A comprehensive review / J. Han, Z. Jian, W. Xu // IEEE Transactions on Instrumentation and Measurement. - 2024. - vol. 73 .
6. Гудфеллоу, Я. Глубокое обучение / Я. Гудфеллоу, И. Бенджио, А. Курвилль. — М. : ДМК Пресс, 2018. — 652 с.
7. Хайкин, С. Нейронные сети / С. Хайкин. — М.: ЛитРес, 2021. — 1104 с.
8. Волхонский А. Н. АЛГОРИТМЫ НЕЧЕТКОГО И НЕЙРОННОГО УПРАВЛЕНИЯ В СИСТЕМАХ РЕАЛЬНОГО ВРЕМЕНИ // Международный студенческий научный вестник. - 2021. - № 1.
9. Джейд, К. Сверточные нейросети / К. Джейд. — М.: ЛитРес, 2024. — 200 с.
10. Шолле, Ф. Глубокое обучение на Python / Ф. Шолле. — М.: ЛитРес, 2022. — 576 с.