Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего образования

«ЮЖНЫЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ» (ФГАОУ ВО «ЮФУ»)

Институт компьютерных технологий и информационной безопасности

АНАЛИТИЧЕСКИЙ ОБЗОР

по дисциплине «Методология научной деятельности»

Методы идентификации голоса

Выполнил

магистрант группы КТмо1-10 В. Р. Сафин

Принял

к. т. н., доцент ИКТИБ А. В. Горбунов

Таганрог 2023

**РЕФЕРАТ**

Методы идентификации голоса 60 с., 15 рис., 2 табл., 19 источн.

МАШИННОЕ ОБУЧЕНИЕ, ИДЕНТИФИКАЦИЯ ЧЕЛОВЕКА, МЕТОДЫ ИДЕНТИФИКАЦИИ, РАСПОЗНОВАНИЕ ГОЛОСА, ИСКУСТВЕННЫЙ ИНТЕЛЕКТ

Целью проведения аналитического обзора является исследование методов методов идентификации человека по голосу

Объектом исследования являются объекты и сферы, применяющие методы идентификации, предмет исследования – алгоритмы машинного обучения и методы идентификации, методы обработки голоса.

В работе применены такие методы исследования как: анализ, синтез, дедукция, индукция и сравнение.

Человеческое развитие сложного голосового аппарата привело к обретению голоса. Аналитический обзор научных работ и результатов интеллектуальной активности выявил перспективные области использования и алгоритмы идентификации голоса.

Анализ научных тенденций показал увеличение количества публикаций на эту тему в мире и в ЮФУ, что свидетельствует о повышенном интересе к данной технологии.

**СОДЕРЖАНИЕ**

[ВВЕДЕНИЕ 4](#_Toc161844219)

[1. Аналитический обзор научных публикаций и результатов интеллектуальной деятельности 6](#_Toc161844220)

[1.1 Общие данные об объекте аналитического обзора 6](#_Toc161844221)

[1.2 Анализ научных публикаций и результатов интеллектуальной деятельности 6](#_Toc161844222)

[1.3 Обзор научных публикаций в ScienceDirect 8](#_Toc161844223)

[1.4 Обзор научных публикаций РИНЦ 22](#_Toc161844224)

[1.5 Патентный поиск 32](#_Toc161844225)

[2 Анализ трендов научных исследований 38](#_Toc161844226)

[2.1 Анализ публикационного ландшафта в сервисе Dimensions 38](#_Toc161844227)

[2.2 Анализ зрелости технологий 48](#_Toc161844228)

[2.3 Выводы и рекомендации 51](#_Toc161844229)

[3 Методология исследований 52](#_Toc161844230)

[3.1 Выбор тематики выпускной квалификационной работы 52](#_Toc161844231)

[ЗАКЛЮЧЕНИЕ 56](#_Toc161844232)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 57](#_Toc161844233)

# **ВВЕДЕНИЕ**

Голос является уникальным биометрическим параметром, который может быть использован для идентификации и аутентификации личности. С ростом цифровизации и внедрением технологий в различные сферы жизни, включая финансовый сектор, медицину, безопасность и телекоммуникации, вопросы обеспечения безопасности и защиты личной информации становятся все более актуальными.

Голос, как биометрический параметр, предоставляет уникальные характеристики каждого человека, что открывает широкие перспективы для методов идентификации личности. В условиях, когда данные из различных источников, будь то разговоры, аудиозаписи или голосовые сообщения, представляют собой неструктурированные данные, важными становятся методы обработки голоса для извлечения полезной информации.

Проблема заключается в том, что существующие методы идентификации голоса подвергаются различным угрозам, таким как поддельные голоса, атаки с использованием технологий синтеза речи, а также вопросы эффективности и точности систем. Кроме того, с расширением области применения биометрических технологий появляются новые сценарии использования, что вносит дополнительные вызовы в обеспечении надежности и безопасности.

Подобно обработке текстов, где методы анализа тональности исследуют эмоциональное содержание текста, методы идентификации голоса направлены на уникальные аспекты звучания человеческого голоса. Технологии машинного обучения, спектральный анализ, и биометрические методы объединяются для создания систем, способных точно определять личность по голосовым данным.

Актуальность темы обусловлена потребностью в совершенствовании существующих методов идентификации голоса, а также поиском новых подходов, способных эффективно справляться с современными угрозами и обеспечивать высокий уровень точности и надежности в различных областях применения. Кроме того, в контексте растущих опасений по поводу защиты персональных данных и конфиденциальности, необходимо более глубоко исследовать этические аспекты использования голосовых данных для идентификации личности.

Целью данного исследования является рассмотрение современных методов идентификации голоса, их сфер применения и технологических аспектов. Для достижения этой цели предстоит провести анализ существующих методов, изучить подходы к обработке голосовых данных с использованием технологий машинного обучения, и выявить оптимальные практики в данной области.

# **1. Аналитический обзор научных публикаций и результатов интеллектуальной деятельности**

## **1.1 Общие данные об объекте аналитического обзора**

В этом разделе вы найдете результаты анализа научных публикаций и интеллектуальной активности, полученные из разнообразных информационных ресурсов, которые связаны с выбранной темой.

Был проведен патентный поиск для анализа и оценки подходов к разработке программного обеспечения, соответствующих теме исследовательской работы «Методы идентификации голоса».

Поиск научных публикаций и результатов интеллектуальной деятельности проводился в таких источниках информации, как:

– РИНЦ;

– ФИПС;

– Sciencedirect;

– Patentscope;

– Espacenet.

## **1.2 Анализ научных публикаций и результатов интеллектуальной деятельности**

В процессе исследования был проведен анализ более 50 научных статей и более пятидесяти патентных документов, которые полностью удовлетворяли критериям поиска.

По результатам поиска научных работ в системе Dimensions было выявлено 1,337,063 научные публикации, 221 публикация в системе РИНЦ, 73,547 публикаций в базе данных Sciencedirect.

На рисунке 1 видно, что материалы по поисковому запросу («Voice identification methods») присутствуют в достаточно большом объеме. Исходя из данных в базе данных Sciencedirect, это направление берет свое начало с 2001 году, в этом году количество публикаций равнялось 5,733.

На рисунке Рисунок 1 видно, что данное направление остается актуальным в плоть до нашего времени, а значит мы также сможем увидеть научные работы по этой теме и в будущем.

Исходя из данных на рисунке 1, можно более точно проследить тенденции в научной литературе по выбранному запросу. Здесь хорошо видно, что количество публикаций по данной тематике в 2015 году было около 50,000. Затем количество публикаций постепенно увеличивалось до 2022 года, после чего научные работы по данной теме немного уменьшились. В 2023 году количество публикаций составило более 91,000. То есть можно сделать вывод, что данная тематика остается популярной и в 2024 году.

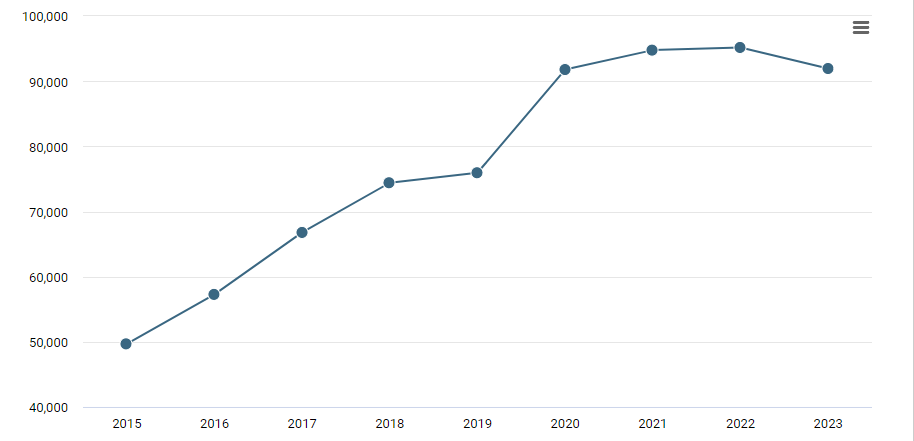


Рисунок 1 – График количества документов по годам по запросу «Voice identification methods» (Dimensions)

На рисунке 2 показано кто из авторов внес наибольший вклад в развитие данного направления. Так, например Roger J R Levesque. опубликовал по данному запросу 775 научных работ. Так же на данном рисунке можно заметить, что среди всех авторов, которые внесли большой вклад в эту область, самые цитируемые авторы из США, Великобритании, Словения, Южная Африка.

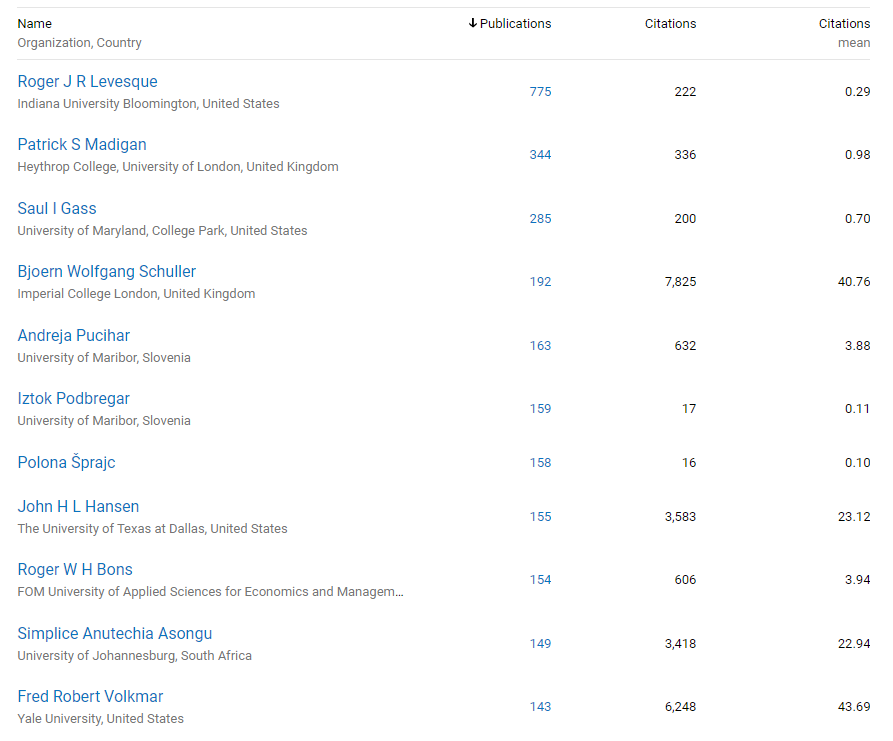


Рисунок 2 – График документов по авторам по запросу «Voice identification methods» (Dimensions)

## **1.3 Обзор научных публикаций в ScienceDirect**

Проведем обзор каждой статьи. Начнем с первой научной работы из Sciencedirect под названием «Identification of Voice Disorders Using Long-Time Features and Support Vector Machine With Different Feature Reduction Methods (1). В данной работе говорится, что идентификация голосовых расстройств играет фундаментальную роль в нашей жизни в настоящее время. Поэтому многие из этих заболеваний должны быть диагностированы на ранних стадиях до того, как они приведут к критическому состоянию. Акустический анализ может быть использован для выявления голосовых расстройств как дополнительной техники в сочетании с традиционными инвазивными методами, такими как ларингоскопия. В этой статье мы провели обширное исследование по диагностике голосовых расстройств с использованием методов статистического распознавания образов. Наконец, мы предложили комбинированную схему методов сокращения признаков, за которой следуют методы распознавания образов для классификации голосовых расстройств.

Шесть классификаторов используются для оценки векторов признаков, полученных с использованием методов анализа главных компонент или линейного дискриминантного анализа (LDA) в качестве методов сокращения признаков. Кроме того, рассматриваются методы индивидуального, прямого, обратного и метода ветвей и границ как методы выбора признаков. Эффективность каждой комбинированной схемы оценивается по точности, чувствительности, специфичности и площади под кривой ROC (AUC). Экспериментальные результаты свидетельствуют о том, что LDA в сочетании с машиной опорных векторов (SVM) показывает лучшую эффективность с уровнем распознавания 94,26% и AUC 97,94%.

Кроме того, эта структура имеет наименьшую сложность по сравнению с другими архитектурами. Среди методов выбора признаков индивидуальный отбор признаков, за которым следует классификатор SVM, показывает лучший уровень распознавания 91,55% и AUC 95,80%.

В статье исследуется способность акустических признаков выявлять нарушения голоса и предлагается новый и эффективный подход к отличению патологических голосов от нормальных.

Следующая научная работа под названием: «Employees’ voice behavior in response to corporate social irresponsibility (CSI): The role of organizational identification, issue perceptions, and power distance culture» (2) В данной работе говорится, что в то время как обширная литература исследовала роль сотрудников в корпоративной социальной ответственности (CSR), мало известно о роли сотрудников в корпоративной социальной безответственности (CSI).

Цель текущего исследования - изучить голосовое поведение сотрудников, порожденное вокруг проблемы CSI. Для этого были проведены два онлайн-опроса среди полноценных сотрудников, используя два различных сценария, связанных с ситуациями CSI: проблемами окружающей среды и разнообразием.

Результаты исследования выявили предварительную организационную идентификацию сотрудников как критически важный фактор, влияющий на их восприятие CSI в различных ситуациях. Кроме того, ситуационное восприятие CSI сотрудниками - как они воспринимают CSI как серьезную проблему, их чувство связи с ситуацией и воспринимаемые ими ограничения в решении ситуации - играли критическую роль в мотивации к голосовому поведению. Кроме того, уровень культуры дистанции власти в организации значительно модерировал отношение между восприятием CSI сотрудниками и их голосовым поведением.

Внедряя литературу по организационной идентификации и теорию публичных отношений в контексте коммуникации и управления CSI, данное исследование предоставляет теоретические и практические выводы. Понимание мотивов сотрудников высказываться еще предстоит должным образом изучить.

Исследование предлагает изучить сдерживающую роль культуры дистанцирования власти в организации во взаимосвязи между восприятием CSI сотрудниками и их речевым поведением.

Следующая научная работа под названием: «Transformational leadership and innovative work behavior: The role of identification, voice and innovation climate» (3) В данной работе говорится, что исследователи и бизнесы проявляют все больший интерес к поощрению сотрудников к инновационной деятельности в условиях высокой конкуренции. Настоящее исследование исследует основные механизмы и граничные условия, объясняющие связь между трансформационным лидерством и инновационной деятельностью сотрудников на передовой линии, основываясь на социальной теории обмена (SET) и теории сохранения ресурсов (COR).

Данные были собраны среди сотрудников на передовой линии и менеджеров, работающих в 145 отелях в четырех провинциях вдоль юго-восточного побережья Китая (Цзянсу, Чжэцзян, Фуцзянь и Гуандун). Результаты показали, что трансформационное лидерство оказывает положительное влияние на инновационное поведение, организационная идентификация/голос сотрудника усиливает отношение между трансформационным лидерством и инновационным поведением, и косвенное воздействие трансформационного лидерства на инновационное поведение (через организационную идентификацию/голос сотрудника) было сильнее, когда сотрудник работал в более инновационном климате. Это исследование предоставляет управленческие идеи в сфере гостиничного бизнеса относительно того, почему и при каких условиях сотрудники ведут себя так, как они ведут.

Исследования индустрии гостеприимства и туризма недостаточно внимания уделяют влиянию трансформационного лидерства на инновационное рабочее поведение сотрудников.

Данное исследование анализирует, как трансформационное лидерство поощряет инновационное рабочее поведение сотрудников.

Исследование предлагает теоретическую модель, объясняющую, как трансформационное лидерство влияет на инновационное рабочее поведение сотрудников отеля.

Следующая научная работа под названием: «The quality of crowdsourcing virtual community and users’ voice behavior: An analysis of stimulus-organism-response framework among Chinese users» (4) В данной работе говорится, что ачество виртуального сообщества с привлечением широких масс является важным фактором, который стимулирует восприятие пользователей принадлежности и привязанности к сообществу, влияя тем самым на их поведение.

Как предварительное условие для развития "креативного вовлечения широких масс", особенно важно изучить, как можно стимулировать поведение голоса пользователей в виртуальных сообществах. Основываясь на модели Stimulus-Organism-Response (SOR) и Теории социальной идентификации, в данном исследовании была разработана концептуальная модель, которая исследует влияние виртуальных сообществ с привлечением широких масс на качество в системе, информации, взаимодействия и обслуживания на поведение пользователей голоса. Кроме того, мы вводим идентификацию сообщества и самодисклозу для дальнейшего анализа влияющего механизма между этими двумя переменными. Данные были собраны с помощью 672 опросных листов от участников известных виртуальных сообществ с привлечением широких масс, таких как Xiaohongshu, Bilibili, Haier Hope, Test Baidu и Test China.

Используя иерархическую регрессию и анализ бутстрэпа, было выявлено положительную корреляцию между качеством виртуального сообщества с привлечением широких масс и поведением голоса пользователей, при этом идентификация сообщества действует как посредник. Более того, самодисклоза показала значительный модерирующий эффект на отношения между идентификацией сообщества и поведением голоса.

Эти результаты значительно вносят в теоретический ландшафт, продвигая модель SOR в рамках виртуального сообщества. Это не только углубляет понимание качества виртуального сообщества с привлечением широких масс, но также предоставляет теоретические и практические выводы для менеджеров и пользователей о том, как стимулировать поведение голоса.

Статья исследует взаимосвязь между идентификацией сообщества и голосовым поведением пользователей в краудсорсинговых виртуальных сообществах.

Гипотеза 1 предполагает, что идентификация сообщества влияет на голосовое поведение пользователей.

Гипотеза 2 предполагает, что самораскрытие смягчает взаимосвязь между идентификацией сообщества и голосовым поведением.

Результаты показали хорошую надежность и валидность всех переменных, а также превосходную степень извлечения показателей измерения в рамках фактора.

Следующая научная работа под названием: «Speaker identification in courtroom contexts: Individual listeners compared to forensic voice comparison based on automatic-speaker-recognition technology» (5) В данной работе говорится, что свидетельские показания экспертов допускаются только в общем праве, если они могут потенциально помочь лицу, принимающему решение, сделать выбор, который они не смогли бы сделать без помощи.

В статье рассматривается вопрос о том, будет ли идентификация диктора индивидуальным прослушиванием слушателя-непрофессионала (например, судьи) более или менее точной, чем результаты системы сравнения голоса, основанной на передовой технологии автоматического распознавания диктора (рис. 3). Слушатели слушают и делают вероятностные суждения по парам записей, отражающих условия записей спорного и известного дикторов в реальном случае.

Изображение выглядит как текст, снимок экрана, диаграмма, линия

Автоматически созданное описание

Рисунок 3 – Данные пробы в эксперименте с индивидуальным слушателем.

В соответствии с различными контекстами судебных залов, тестируются слушатели с разными языковыми фонами: некоторые знакомы с языком и акцентом, другие знают язык, но менее знакомы с акцентом, а третьи мало знакомы с языком. Также, отражая различные судебные контексты: в одном случае слушатели делают выводы исключительно на основе прослушивания, а в другом - на основе как прослушивания записей, так и учета значений отношения вероятностей, выданных системой сравнения голоса.

Статья исследует идентификацию говорящего непрофессиональными слушателями.

Эксперименты проводятся с использованием пар записей из набора данных forensic\_eval\_01.

Записи отражают условия записи допрашиваемого говорящего и записи известного говорящего из реального судебного дела.

Следующая научная работа под названием: «Speaker identification from emotional and noisy speech using learned voice segregation and speech VGG» (6) В данной работе говорится, что речевые сигналы более подвержены эмоциональным воздействиям и акустическим помехам, чем другие виды коммуникаций.

Приложения для обработки речи в реальном времени сталкиваются с трудностями при работе с шумными данными, насыщенными эмоциями. Найти надежный метод для выделения доминирующего сигнала от внешних влияний. Идеальная система должна быть способна точно идентифицировать необходимые звуковые события в сложной сцене, записанной в нежелательных условиях.

В данной работе было предложено и сделана оценка комплексной структуры для распознавания голоса в неблагоприятных условиях общения с использованием заранее обученной маски глубокой нейронной сети и голосовой модели VGG. Это исследование предлагает уникальный метод распознавания диктора в сложных условиях, включая эмоции и помехи.

Используя аудио-визуальный набор данных Райерсон, представленная модель превзошла недавнюю литературу по эмоциональным речевым данным на английском и арабском языках, сообщая среднюю ставку идентификации диктора 85,2%, 87,0% и 86,6% с использованием аудио-визуального набора данных Райерсон (RAVDESS), набора данных речи под симулированным и фактическим стрессом (SUSAS) и набора данных на арабский диалект (ESD) соответственно.

Статья обсуждает проблемы обработки сложных аудиальных сцен с помощью методов машинного обучения и интеллектуального анализа данных.

Звуковая сцена представляет собой серию звуков с различным отношением сигнала к шуму.

Цель работы - предложить способ решения этих проблем с помощью методов машинного обучения и интеллектуального анализа данных.

Глубокое обучение используется для моделирования высокоуровневых абстракций в данных.

Предложенная модель идентификации говорящего разработана на платформе глубокого обучения и оценена в шумной, стрессовой и эмоционально сложной разговорной среде.

Предлагаемая модель обеспечивает высокую производительность и устойчивость к различным акустическим искажениям и помехам.

Следующая научная работа под названием: «Influence of pitch and speaker gender on perception of creaky voice» (7) В данной работе говорится, что хриплый голос – это не модальное качество голоса, обычно описываемое как имеющее импульсную природу и низкую высоту тона. В то время как эмпирические исследования привели к противоречивым результатам относительно распространенности хриплого голоса в зависимости от пола говорящего, хриплый голос стереотипно ассоциируется с женской речью.

Прошлые исследования исследовали, улучшается ли у слушателей идентификация хриплого голоса через степень различий тональности между модальным и хриплым голосом или же через социальные предвзятости, связывающие хрип с женской речью. Однако результаты оказались относительно неоднозначными.

Настоящее исследование рассматривает этот вопрос через восприятие эксперимента, используя тщательно контролируемые стимулы. 258 слушателей было попрошено определить, слышат ли они хрип или нет, когда им предлагались стимулы, измененные по тональности и хриплому голосу из двухсловных фраз, произнесенных мужчиной и женщиной на английском языке в Австралии. Были проанализированы данные точности и время ответов.

Цель исследования - изучить, как высота тона и пол говорящего влияют на идентификацию скрипа в контексте использования.

Исследование использует высоко контролируемые стимулы для устранения влияния высоты тона и пола говорящего.

Анализ времени реакции (RT) позволит получить более детальное представление о том, как слушатели идентифицируют скрипучий голос.

Три возможных сценария могут влиять на идентификацию слушателями скрипа: контраст высоты тона, гендерная предвзятость и контраст высоты тона и гендерное смещение.

Результаты указывают на то, что говорящие на английском языке в Австралии менее полагаются на социальные предвзятости при идентификации хриплого голоса и вместо этого принимают решения на основе тональности, модулированной их ожиданиями об типичных тональных диапазонах в зависимости от пола говорящего, основанных на их опыте. Так же Результаты подчеркивают важность включения характеристик говорящего в модели восприятия хриплого голоса и качества голоса в целом.

Следующая научная работа под названием: «Can There be a Biometric Parameter for Voice?—An Investigation Into the Vocal Cry of Twins» (8) В данной работе говорится, что голос индивидуальности является результатом взаимодействия строения гортани и надгортанника человека. Поскольку строение гортани и надгортанника различается у каждого человека, голос, производимый каждым, становится уникальным и отличимым от голоса другого человека. Если эта уникальность голоса можно найти в акустических параметрах голоса, это может способствовать областям судебной идентификации дикторов и биометрии голоса. Лучшей методологией для такого исследования было бы сравнение голосов близнецов, у которых есть ряд одинаковых характеристик.

Узнать возможно ли перцептивно различить плач идентичных близнецов друг от друга. Сравнить акустические параметры плача среди пар близнецов, чтобы выяснить различающий параметр.

Это было наблюдательное сквозное исследование. В исследование были включены девятнадцать пар близнецов в возрасте менее 1 недели, завершивших 34 месяца гестационного периода без особых врожденных аномалий. Для анализа голосовых образцов использовалось программное обеспечение PRAAT.

Исследование сравнивает голоса близнецов с рядом сходных характеристик.

Цель исследования - выяснить, можно ли перцептивно отличить крик однояйцевых близнецов друг от друга.

Исследование проводилось на новорожденных близнецах и не выявило отличительных акустических параметров.

Высокое сходство акустических параметров пар близнецов предполагает, что это может быть не отдельный параметр, а набор параметров, которые могут придать индивидуальную идентичность голосу человека.

Исследование может предоставить информацию для автоматического распознавания говорящего, судебной идентификации говорящего и голосовой биометрии.

Высокая схожесть в акустических параметрах среди пар близнецов подразумевает, что скорее всего, вместо индивидуальных параметров, это может быть сложное взаимодействие различных параметров, которое может придать индивидуальную идентичность голосу человека. Исследование может предложить информацию для автоматического распознавания дикторов, судебной идентификации дикторов и биометрии голоса.

Следующая научная работа под названием: «Voice-based age, gender, and language recognition based on ResNet deep model and transfer learning in spectro-temporal domain» (9) В данной работе говорится, что в системах распознавания личности важным вопросом является определение возраста, пола и языка человека по характеристикам голосового сигнала, особенно в связи с важностью обеспечения безопасности.

Проблемы классификации возраста, пола и языка являются важными в области обработки сигналов, поскольку они используются для анализа и понимания человеческого поведения, взаимодействий и предпочтений. Это может быть особенно полезно в областях взаимодействия человека с компьютером, психологии и исследований социальных наук. В данной работе представлена новая система детектирования возраста, пола и языка диктора на основе моделей глубокого обучения.

Модели глубокого обучения показали большую эффективность в различных областях обработки сигналов. Для данной работы был проведен ряд тестов глубоких моделей, включая сверточные нейронные сети (CNN), рекуррентные нейронные сети (RNN) и архитектуру ResNet34 с тонкой настройкой. Кроме того, были применены методы, такие как перенос обучения, для улучшения эффективности предложенной системы. Входные голосовые сигналы предобрабатывались с помощью спектро-временного преобразования для получения дополнительных характеристик, которые могут быть поданы на модель ResNet34, специально разработанную для задачи обработки голосового сигнала. Набор данных, использованный в данной работе, был взят из инициативы Mozilla common voice, посвященной развитию технологий распознавания речи и идентификации языка. Производительность предложенного алгоритма была оценена в присутствии гауссовского шума для определения его устойчивости.

Экспериментальные результаты показали, что предложенный алгоритм значительно превзошел базовые алгоритмы и другие глубокие нейронные сети в области распознавания возраста и пола по голосовым сигналам.

В данной работе процедура классификации проводится с использованием глубоких моделей, которые показывают высокую производительность в различных областях обработки данных. Предложенный метод основан на архитектуре ResNet34. ResNet является сокращением от residual network, типа глубокой нейронной сети, которая продемонстрировала отличную производительность на многих задачах классификации сигналов. Эта модель использует блок остатков, где вывод каждого слоя добавляется к входу следующего слоя, тем самым избегая необходимости обрабатывать данные снова.

Система искусственного интеллекта, разработанная для данной работы, включает предварительную обработку и увеличение входных спектро-временных характеристик, которые затем подаются на тонко настроенную модель ResNet34. Входной слой модели получает входные данные, и после шага нормализации пакета, выходной класс получается из окончательного полностью связанного слоя. Обученная модель обучается классифицировать входной голосовой сигнал на девять различных языков для различных групп по полу и возрасту.

Определение возраста, пола и языка человека через анализ голосового сигнала давно является сложной задачей в системах распознавания личности, здравоохранении, безопасности, финансах и развлечениях. Эта задача классификации представляет собой сложную проблему из-за нескольких факторов, таких как сложность сигналов, изменчивость и шум в данных, ограниченные размеченные данные и изменчивость в сборе данных.

Следующая научная работа под названием: «Voice Rehabilitation After Laryngeal Transplantation: Own Experience» (10) В данной работе говорится, что ларингеальная трансплантация (LT) остаётся малопроизводимой процедурой из-за чётко определённых и ограниченных медицинских показаний. Сложности включают в себя очень сложную операцию, постоянное иммуносупрессивное лечение, а также реабилитацию голоса, речи и глотания после трансплантации. Цель данной работы - описать модель восстановления голоса и речи после LT на основе нашего собственного опыта.

Здесь представлены процессы реабилитации двух пациентов, прошедших ранее LT, а также их результаты голоса до и после реабилитации. Оба пациента прошли реабилитацию голоса, речи и глотания на пятый месяц после LT.

Один из пациентов заметно улучшил качество голоса. В то время как другой не вернул свой естественный, звучный голос, но показал очень высокий результат на тесте самооценки голоса.

Реабилитация голоса после LT - сложный и долгосрочный процесс, требующий участия специалистов. Оптимальные анатомические условия, восстановленные во время операции, мотивация пациента и поддержка профессионалов позволяют вернуться к нормальному вербальному общению.

Цель статьи - описать модель восстановления голоса и речи после ЛТ, основанную на собственном опыте авторов.

Исследование представляет собой наблюдательное исследование с описанием процессов реабилитации двух пациентов после ЛТ.

Восстановление голоса после ЛТ является сложным и длительным процессом, требующим привлечения команды специалистов.

Оптимальные анатомические условия, мотивация пациента и поддержка профессионалов играют важную роль в возвращении к нормальному вербальному общению.

## **1.4 Обзор научных публикаций РИНЦ**

В статье «Метод текстонезависимой идентификации личности по голосу» (11) предлагается иммунологический метод решения задачи текстонезависимой идентификации личности по голосу, основанный на принципах представления и обработки речевой информации, принятых в искусственных иммунных системах. Для идентификации личности по голосу используется модель Фанта, в которой, речевой сигнал образуется путем прохождения через фильтр высокого порядка. В качестве векторов признаков используются кепстральные коэффициенты, полученные на основе линейного предсказателя речи. Последующий анализ векторов признаков осуществляется на основе аппарата искусственных иммунных систем с использованием иммунологической модели отрицательного отбора. Модель реализует децентрализованное распознавание последовательно идущих фрагментов речи, путем их сопоставления со специальными, предварительно созданными распознающими элементами - детекторами, имитирующими иммунокомпетентные клетки иммунной системы.

В основе метода лежит модель Фанта, где речевой сигнал формируется через фильтр высокого порядка. Это позволяет улучшить качество обработки сигнала и повысить точность идентификации. Кепстральные коэффициенты, основанные на линейном прогнозировании речи, служат в качестве векторов признаков. Эти векторы представляют собой информацию, которую можно использовать для идентификации личности по голосу.

Анализ этих векторов признаков производится с использованием искусственных иммунных систем и модели отрицательного отбора. Эта модель реализует децентрализованное распознавание последовательных фрагментов речи, сравнивая их с заранее созданными детекторами, имитирующими клетки иммунной системы. Это позволяет системе быстро и точно определить, принадлежит ли речевой сигнал "своему" или "чужому".

Решение о принадлежности речевого сигнала к "своему" или "чужому" принимается на основе статистического подхода, основанного на частоте срабатывания детекторов. Этот подход обеспечивает высокую точность распознавания и надежность системы.

Метод прошел проверку в MATLAB и показал свою эффективность. Он предназначен для непрерывной проверки личности говорящего и обладает полной защитой от атак воспроизведения. Это делает его особенно полезным в условиях, когда требуется непрерывная и надежная проверка личности.

Однако, для повышения точности метода необходимы параллельные вычисления больших объемов данных. Это делает его перспективным для использования в многопроцессорных системах высокой производительности, где доступны значительные ресурсы для обработки данных. В свою очередь, это открывает новые возможности для развития и улучшения метода, что может привести к созданию еще более эффективных и надежных систем идентификации личности по голосу.

В статье «Методы и средства решения задачи идентификации личности по голосу» (12) рассмотрены вопросы и особенности идентификации личности по голосу, приведенапостановка задачи. Описан предложенный подход крешению задачи идентификации личности по голосу, основанный на использовании сверточных нейронных сетей для анализа «сырых» данных, без их предобработки, в комплексе с выбранным методом классификации.

Статья подробно обсуждает вопросы идентификации личности по голосу и предлагает новаторский подход, основанный на использовании сверточных нейронных сетей для анализа необработанных данных. В современном мире, где действительно ощущается растущая потребность в защите конфиденциальной информации из-за всеобщей информатизации общества, одним из наиболее обещающих способов защиты является использование биометрических технологий. В частности, статья утверждает, что голосовая идентификация занимает особое место среди этих технологий.

Авторы статьи также подробно обсуждают преимущества и недостатки различных методов биометрической идентификации, включая, но не ограничиваясь, отпечатки пальцев, сканирование сетчатки глаза и голосовую идентификацию. Они утверждают, что, несмотря на различную эффективность этих методов, каждая технология может быть наилучшим выбором для определенной задачи или обстоятельств.

В частности, голосовая идентификация может иметь значительные преимущества в определенных условиях. Например, она может быть более удобным и менее инвазивным способом идентификации для пользователей, нежели некоторые другие методы. В заключении авторы указывают, что важно продолжать исследования в этой области, чтобы улучшить точность и надежность голосовой идентификации и сделать эту технологию еще более привлекательной для широкого применения.

В статье «Основные методы идентификации диктора по голосу» (13) представлен обзор методов, используемых для решения задачи идентификации по голосу. Уделяется внимание обобщенной структуре систем идентификации. Также приводятся краткие описания распространенных методов извлечения признаков, обсуждаются методы оценки систем распознавания и представления результатов таких оценок.

Задача включает определение, кому принадлежит голосовой образец среди ранее зарегистрированных пользователей. Это означает, что голос, который был прослушан или записан, сравнивается с базой данных голосов, чтобы определить, чей голос был услышан.

В начале процесса дикторы регистрируются в системе, где они записывают свои голоса. Этот этап критически важен, поскольку он позволяет системе собрать первичные данные о голосе каждого пользователя. Эти данные затем используются для создания уникальных голосовых профилей для каждого диктора.

Затем из этих голосовых образцов извлекаются информативные признаки. Это могут быть различные аспекты голоса, такие как тембр, тон, ритм, скорость и другие уникальные характеристики, которые помогают отличить один голос от другого. Эти признаки затем используются для создания шаблонов каждого диктора.

Эти шаблоны затем используются для сравнения с образцами в процессе идентификации. Система идентификации дикторов состоит из трех основных этапов: обработка сигнала для извлечения информативных признаков, использование этих признаков для построения шаблона и использование этих шаблонов для принятия решения о том, чей голос был записан.

В целом, идентификация дикторов – это сложная задача, которая требует точности и надежности. Исследования в этой области продолжаются, поскольку они имеют огромное количество потенциальных применений в различных областях, таких как обеспечение безопасности, судебные дела, искусственный интеллект и многое другое.

В статье «Применение метода идентификации по голосу, адаптированного к тихому произнесению парольных фраз для противодействия утечки речевой информации» (14) рассматриваются особенности биометрической идентификации по голосу при условии тихого произнесения парольных фраз. Биометрическая идентификация по голосу существенно отличается от стандартных систем идентификации и систем контроля управления доступом, использующих символьные пароли и ключи. Биометрическая идентификация по голосу производится по уникальным и индивидуальным признакам личности и практически исключает вероятность несанкционированных действий, связанных с потерей, кражей или передачей пароля третьим лицам.

Широкое применение систем биометрической идентификация по голосу влечет за собой повышенный интерес со стороны злоумышленников. Наиболее частыми являются атаки с применением ранее применяемых биометрических признаков, например, аудиозапись парольной фразы. Для минимизации описанных выше атак, предложен адаптированный к тихому произнесению парольных фраз метод биометрической идентификации по голосу, основанный на применении метода обеляющего фильтра. Описана программная реализация предложенного метода – «Информационная система идентификации дикторов по голосу», позволяющая проводить биометрическую идентификацию по голосу при условии тихого произнесения парольных фраз для противодействия утечки речевой информации по акустическим каналам.

В данной статье рассматриваются уникальные особенности биометрической идентификации по голосу при использовании такого специфического подхода, как тихое произнесение парольных фраз. Как известно, биометрические системы идентификации стали значительным открытием в области защиты информации, поскольку они значительно отличаются от традиционных систем идентификации, которые используют символьные пароли и ключи.

Одним из ключевых преимуществ биометрической идентификации является то, что она практически полностью исключает возможность несанкционированных действий, связанных с потерей, кражей или передачей пароля. Однако, несмотря на это, широкое использование этих систем привлекает большой интерес со стороны злоумышленников, что делает вопрос обеспечения безопасности актуальным.

В связи с этим, для снижения риска возможных атак и повышения уровня защиты, был предложен новый метод, адаптированный для тихого произнесения парольных фраз. Этот метод основан на использовании такого инструмента, как обеляющий фильтр, который позволяет улучшить качество идентификации и устранить возможные угрозы.

В статье подробно описывается программная реализация этого метода, его преимущества и возможные сложности, которые могут возникнуть в процессе его использования. Это позволяет читателю полностью понять особенности и преимущества этого подхода, а также учесть его при выборе системы идентификации.

В статье «Методы идентификации человека по голосу» (15) голосовая аутентификация, основанная на анализе речи человека, является актуальной и перспективной задачей в современном мире. Эта задача включает распознавание голоса, который является звуком, создаваемым движением воздуха через голосовые связки в гортани человека или животного. Голос используется как основное средство коммуникации и выражения, обеспечивая общение, пение, чтение и т.д.

Голос каждого человека уникален, поскольку он зависит от множества факторов. Это включает возраст, пол, размеры голосовых связок, форму легких, гортани и носа, а также стиль речи. Голос может выдавать эмоции и намерения человека, а также содержать информацию о его возрасте, поле, социальном статусе и т.д. Благодаря этому голос играет важную роль не только в общении между людьми, но и в различных технологиях, таких как системы распознавания речи, голосовое управление и т.д.

Существуют различные характеристики голоса, которые включают тонкость, интонацию, громкость, ясность и дикцию, тембр, скорость речи, паузы и ударения. Дыхание, дикция, артикуляция и использование резонатора также влияют на характеристики голоса.

Цифровое представление голоса представляет собой уникальный набор данных для каждого человека, который может использоваться для идентификации. Это представление создается на основе различных акустических параметров, таких как частота сигнала, длительность, интонация и т.д. Цифровое представление голоса используется в различных областях, где требуется высокий уровень безопасности и идентификации, например, в системах банковской безопасности, системах контроля доступа к помещениям, системах распознавания голосового управления.

Несмотря на высокую точность идентификации по голосу, существуют и определенные проблемы. Например, простуда может повлиять на цифровое представление голоса, что понизит его точность. Также, возможно подделать звуковую атаку на систему распознавания голоса, что может привести к ошибочной идентификации. Поэтому, несмотря на высокую точность идентификации на основе цифрового представления голоса, его следует использовать с осторожностью и совмещать с другими методами идентификации личности.

Использование голоса для идентификации человека является биометрическим методом, который основывается на уникальных акустических параметрах голоса. Для идентификации человека по голосу используются специальные программы, которые сравнивают акустические параметры голоса с ранее сохраненными шаблонами голоса. Этот метод уже применяется в различных областях и продолжает развиваться.

В статье «Иммунологический метод текстонезависимой идентификации личности по голосу» (16) предлагается иммунологический подход к решению задачи текстонезависимой идентификации личности по голосу, основанный на принципах представления и обработки речевой информации, принятых в искусственных иммунных системах Речевой сигнал, подлежащий идентификации, представляется последовательностью векторов признаков, составленных из кепстральных коэффициентов. Последующий анализ такой последовательности осуществляется на основе иммунной модели отрицательного отбора.

В этой статье представлен новый метод идентификации личности по голосу, независимо от текста, который основан на принципах, применяемых в искусственных иммунных системах. Это новаторский подход, который предлагает уникальные возможности в области идентификации голоса.

Речевой сигнал, подвергающийся идентификации, представляется в виде последовательности векторов признаков, созданных из кепстральных коэффициентов. Этот процесс преобразования речи во векторные признаки является ключевым шагом в алгоритме идентификации.

Дальнейший анализ этой последовательности проводится с использованием иммунной модели отрицательного отбора. Эта модель извлекает уникальные и важные характеристики из каждого вектора признаков, что позволяет производить точную идентификацию личности по голосу.

Важно отметить, что системы идентификации личности по голосу имеют обширное применение в различных областях. Это включает телефонию, где они могут служить для проверки личности абонента; электронную коммерцию, где ускоряется процесс входа в систему и улучшается общий пользовательский опыт; криминалистику, где они могут помочь в расследовании преступлений; разведку и контрразведку, где они могут служить для идентификации особ, представляющих угрозу безопасности; антитеррористический мониторинг, где они могут помочь в быстром определении подозрительных лиц или действий; системы голосового управления, где они могут улучшить надежность и эффективность управления различными устройствами и системами. Отмечается, что потенциал применения этих систем только увеличивается с развитием технологий и увеличением требований к безопасности.

В статье «Возможности идентификации иноязычных говорящих по голосу и речи экспертными методами» (17) исследование посвящено возможности первичного распознавания говорящих на незнакомом и интерферированном русском языке с целью их дальнейшей идентификации по голосу и речи. В статье рассмотрен один из автоматизированных методов идентификации иноязычных дикторов - метод анализа мелодического контура синтагм (на основе сравнения мелодического оформления различных фрагментов речевого сигнала). Дана краткая характеристика методологической базы, а также охарактеризованы основные компоненты мелодической структуры речи.

Автором проведен сопоставительный анализ просодического оформления высказывания в речевом поведении иноязычных дикторов и определены некоторые особенности правильного распознавания говорящего, в том числе со стороны тех, для кого рассматриваемый язык является неродным. В результате исследования установлено, что опытные эксперты способны адекватно распознать и правильно идентифицировать иноязычного говорящего. Типологические различия между родным языком эксперта и языком, подлежащим идентификации, не влияют на правильность распознавания по голосу и речи.

Это исследование посвящено методам первичного распознавания и идентификации говорящих на незнакомом и интерферированном русском языке по их голосу и речи. Данный вид исследования особенно важен в контексте лингвистического анализа и создания технологий речевого распознавания.

Основной метод, который рассматривается в статье – это анализ мелодического контура синтагм. Этот метод позволяет выявлять особенности просодического оформления речи, что в свою очередь может помочь в определении личности говорящего.

В статье даны подробные описания методологической базы и основных компонентов мелодической структуры речи. Эти методы позволяют провести детальный анализ голосовых данных и обеспечивают точность в распознавании говорящих.

Также в рамках исследования проведен сопоставительный анализ просодического оформления высказывания в речевом поведении иноязычных дикторов. Это позволяет увидеть различия в акцентах и интонациях, что может быть полезно при распознавании говорящих разных языков.

В результате исследования выяснилось, что опытные эксперты могут адекватно распознавать и правильно идентифицировать иноязычного говорящего, несмотря на типологические различия между их родным языком и языком, подлежащим идентификации. Это подтверждает эффективность применяемых методов и подходов в области распознавания речи.

## **1.5 Патентный поиск**

В результате патентного поиска было найдено более 500 000 записей, которые имеют отношение к теме запроса «Voice identification methods». По результатам обзора патентных документов, было найдено 115 патентов в ФИПС, 589 500 в Espacenet и 800 в Patentscope, которые так или иначе имеют отношение к теме запроса. Некоторые из найденных патентов перечислены в таблице 1.

Таблица 1 – Результат патентного поиска

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № | Название | Автор(ы) текста | Код страны, №, свидетельства | Поисковая система |
| 1 | широковещательная система оповещения абонентов мобильной связи о возникновении экстренных ситуаций, абонентское устройство связи и способы ее функционирования | Цуриков Александр Николаевич | RU, 2598294 | ФИПС |
| 2 | Способ и система для доверенного безбумажного предъявления документов | Тамойкин Андрей Юрьевич | RU, 2701088 | ФИПС |
| 3 | Способ количественной оценки эрозионных потерь почвы с применением наземного лазерного сканера | Ермолаев Олег Петрович | RU, 2700930 | ФИПС |

Продолжение таблицы 1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № | Название | Автор(ы) текста | Код страны, №, свидетельства | Поисковая система |
| 4 | Контур обратной связи для предотвращения несанкционированной рассылки | РАУНТВЭЙТ Роберт Л. | RU, 2331913 | ФИПС |
| 5 | Способ и система оптического распознавания символов, которые сокращают время обработки изображений, потенциально не содержащих символы | Чулинин Юрий Георгиевич | RU 2571616 | ФИПС |
| 6 | Система и способ выявления и использования эталонного эмоционально-интеллектуального профиля (эи-профиля) по группам анализа | Хлевная Елена Анатольевна | RU 2720400 | ФИПС |

Продолжение таблицы 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № | Название | Автор(ы) текста | Код страны, №, свидетельства | Поисковая система |
| 7 | Метод обработки результатов идентификации по голосу, устройство и оборудование и носитель информации | Wang rui | CN 108573707 | Espacenet |
| 8 | Системы и методы идентификации голоса | LI, Lu  LU, Li  ZHANG, Xiang RAO, Feng YUE, Shuai CHEN, Bo  MA, Jianxiong  LIU, Haibo | WO 2014183373 | Espacenet |
| 9 | Метод идентификации по голосу | Chen Yi | CN 103474062 | Espacenet |

Продолжение таблицы 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № | Название | Автор(ы) текста | Код страны, №, свидетельства | Поисковая система |
| 10 | Оптимизация функций идентификации голоса и методы динамической регистрации, клиент и сервер | LIU, Gang  ZHAO, Qingen  LIU, Guangxing | WO 2018232148 | Espacenet |
| 11 | Идентификация голосовых пакетов | SAHA DEBANJAN SHAE ZON-YIN | EP 1810278 | Espacenet |
| 12 | Системы и методы для безопасной голосовой идентификации и интерфейса медицинского устройства | O'NEIL, Michael P.  ANDERSOHN, Lutz  MANNHEIMER, Paul D.  MECCA, Roger | WO 2008042121 | Espacenet |

Анализ патентного поиска показывает, что в данной сфере достигнуты значительные успехи как в России, так и за её пределами. Китай, США и Индия являются лидерами по количеству публикаций и патентов на данную тему, причём большинство авторов представляют собой китайских учёных. Отзывы научных статей также указывают на то, что многие авторы выделяют использование социальных сетей, в частности Twitter, в качестве источников данных для их сбора и анализа. Большинство авторов применяют глубокое обучение и нейронные сети в своих исследованиях для улучшения обучения их моделей.

# **2 Анализ трендов научных исследований**

## **2.1 Анализ публикационного ландшафта в сервисе Dimensions**

Для анализа трендов научных исследований был использован запрос «Voice identification methods», соответствующий тематике исследований. А Вид поискового запроса можно видеть на рисунке 4.

Изображение выглядит как текст, Шрифт, снимок экрана

Автоматически созданное описание

Рисунок 4 – Поисковой запрос.

По результатам поискового запроса было выявлено что в мире по этой тематике число публикаций равняется 1,343,194 число датасетов – 46, грантов – 451 и патентов – 2,123,750, клинические испытания – 148, программные документы – 40,335.

График динамики числа публикаций в мире представлен на рисунке 5. Из данного графика видно, что общий рост публикаций по данной тематике непрерывно возрастало до 2022 года, затем количество публикаций уменьшилось. В 2022 году количество работ составило более 90,000, а в 2023 на данный момент количество работ в данной области примерно равно 92,528

Изображение выглядит как линия, снимок экрана, текст, диаграмма

Автоматически созданное описание

Рисунок 5 – График динамики числа публикаций в мире

График динамики числа цитирований в мире (FCR) представлен на рисунке 6. Из графика видно, что на 2015 год максимальное число цитирования равнялось примерно 2,8.

Далее число снижалось постепенно до 2018 года, где оно стало равным примерно 2,39. А уже с 2022 года по настоящее время число цитирований в мире уменьшалось и стремилось к нулю. Средний коэффициент за весь период примерно равняется 2.

Изображение выглядит как линия, диаграмма, График, снимок экрана

Автоматически созданное описание

Рисунок 6 – График динамики числа цитирований в мире (FCR)

Из-за недостатка статей на эту тему в ЮФУ, было решено выбрать более общую область исследований, такую как «Машинное обучение», и дополнительно уточнить ее для мониторинга динамики в рамках ЮФУ (см. Рисунок 7).

Изображение выглядит как текст, снимок экрана, Шрифт, линия

Автоматически созданное описание

Рисунок 7 – Запрос для отслеживания динамики в пределах ЮФУ

После чего был построен график динамики числа публикаций ученых ЮФУ, приведенный на рисунке 8. Как можно увидеть из графика, с 2018 года начался рост количества статей в данной тематике, пик числа публикаций пришелся на 2020 год, количество публикаций составило примерно 737. С 2021 по 2023 год количество статей уменьшалось, количество публикаций на конец 2023 года примерно 227.

Изображение выглядит как линия, диаграмма, График, Параллельный

Автоматически созданное описание

Рисунок 8 – График динамики числа публикаций ученых ЮФУ

График динамики числа цитирований ученых ЮФУ (FCR) представлен на рисунке 9. Как можно увидеть из графика наивысшее число цитирований равно примерно 2,11, которое пришлось на 2015 год. С 2023 года число цитирований, ученных ЮФУ остается равным 0.

Изображение выглядит как линия, диаграмма, График, Параллельный

Автоматически созданное описание

Рисунок 9 – График динамики числа цитирований ученых ЮФУ (FCR)

Список ведущих авторов ЮФУ представлен на рисунке 10. В списке можно увидеть, что наивысшее количество публикаций 52 опубликовал автор: А. В. Солдатов, С. И.. Среди этих ученных, А. В. Солдатов – автор большим числом цитирований 586. А самым большим числом цитирования из списка обладает А. В. Солдатов.

Изображение выглядит как текст, снимок экрана, число

Автоматически созданное описание

Рисунок 10 – Список ведущих авторов ЮФУ

На рисунке 11 можно увидеть сеть соавторства ученых ЮФУ. Общее число исследователей в сети равно 860. Среди них есть преподаватели: Солдатов А. В., Гуда А. А., Бутакова М. А., Чернов А. В.

Изображение выглядит как текст, диаграмма, снимок экрана, линия

Автоматически созданное описание

Рисунок 11 – Сеть соавторства ученых ЮФУ

**Рассмотрим ведущих, ученных ЮФУ в данной области:**

– Солдатов А. В.: 33 соавторов, общее количество соавторств 219, 51 публикация;

– Гуда А. А.: 26 соавторов, общее количество соавторств 195, 44 публикации.

На рисунке 12 продемонстрирована сеть по анализу цитируемости ученных ЮФУ.

Изображение выглядит как текст, снимок экрана, диаграмма, карта

Автоматически созданное описание

Рисунок 12 – Анализ цитируемости ученных ЮФУ

**Рассмотрим несколько кластеров сети соавторства ученных ЮФУ:**

– Кластер 1 (красный кластер): Солдатов А. В., Бугаев А. Р., Усольцев О. А.

– Кластер 2 (синий кластер): Гуда А. А., Гуда С. А., Алгасов А. А.

– Кластер 3 (зеленый кластер): Бутаков М. А., Чернов А. В., Пашков Д. М.

Также в сервисе Dimensions можно посмотреть датасеты по интересующей тематике. По запросу «Voice identification methods» было предложено 30,606,335 датасета. Рассмотрим примеры некоторых из них:

– Датасет 1: Self-perception and voice quality of journalism students – это набор данных, содержащий заголовки из хорватских новостных изданий, помеченные настроением по отношению к названным сущностям и общим тоном заголовка (рис. 13) (18).

Целью было связать данные из анализа звучания голоса, жалоб и самоощущения голоса студентов журналистики. Методы — это наблюдательное, описательное, поперечное исследование с проспективным сбором данных, проведенное в Университете Эстадуал ду Сентро-Оэсте/УНИКЕНТРО-ПР. В исследовании приняли участие 41 студент журналистики, из них 27 женщин и 14 мужчин. Был использован протокол для сбора данных об идентификации и жалобах на голос, и проведены голосовые записи с продолжительным излучением гласного звука «а»; и связанные (подсчет чисел и месяцев года). Эти записи были проанализированы логопедом, специалистом по голосу, относительно качества голоса (адаптированного или измененного). Студенты также отвечали на протокол "Описательные термины о голосе". Результаты: несмотря на адаптированные голоса, студенты высказывали жалобы на голос. Самыми распространенными жалобами были изменение тона, фонемное производство, качество голоса и скорость речи. Студенты с адаптированным качеством голоса упоминали большее количество положительных терминов, связанных с самооценкой. Не было различий при сравнении среднего количества положительных и отрицательных терминов, использованных группой. Самые положительные термины были: приятный голос, выразительный, уверенный, женственный, сильный и кроткий.

Отрицательные термины включали: безмелодичный голос, неустойчивый, колеблющийся, неравномерный, быстрый, носовой голос, низкий и робкий. Вывод: даже имея адаптированные голоса с акустической точки зрения, студенты журналистики жалуются на голос, возможно, из-за требований, предъявляемых к ним во время учебы.

Изображение выглядит как текст, снимок экрана, число, Параллельный

Автоматически созданное описание

Рисунок 13 – Первый датасет

– Датасет 2: Voice challenge in transgender women: trans women self-perception of voice handicap as compared to gender perception of naïve listeners (Проблемы с голосом у трансгендерных женщин: самовосприятие трансгендерными женщинами недостатков голоса по сравнению с гендерным восприятием наивных слушателей) (19) – целью было проанализировать самоощущение трансгендерных женщин относительно голосового инвалидизма по сравнению с восприятием пола голоса со стороны наивных слушателей. Методы: 31 трансгендерная женщина, которым сначала была проведена голосовая акустическая оценка, а затем их голоса были записаны для измерения основной частоты и стандартных отклонений, подходили для участия в исследовании. Затем они отвечали на протокол Индекса Голосового Инвалидизма, из которого, к концу сбора данных, были выбраны наиболее подходящие вопросы для ежедневных потребностей этой группы. Затем 50 наивных помешанных глазами слушателей были подвергнуты записям и должны были определить голоса как мужские, женские или неопределенные. Для динамических характеристик использовались описательные статистики и коэффициент Спирмена для оценок протокола и определений голоса докладчика.

Результаты: средняя основная частота составляла 172,40 Гц (SD = 4,8 Гц) и треть трансгендерных женщин сообщили, что они удовлетворены своим голосом. Была найдена умеренная положительная значимая корреляция в голосах, считающихся мужскими, и сильно значимая и отрицательная в голосах, считающихся женскими. Однако для голосов, считавшихся неопределенными, значимой корреляции не было найдено. Вывод: самоощущение трансгендерных женщин относительно голосового инвалидизма непосредственно связано с восприятием их голоса по полу наивными слушателями (рис. 14).

Изображение выглядит как текст, снимок экрана, программное обеспечение, число

Автоматически созданное описание

Рисунок 14 – Второй датасет

## **2.2 Анализ зрелости технологий**

Для анализа зрелости технологии была использована методология Gartner, представленная на рисунке 15.

Изображение выглядит как текст, диаграмма, снимок экрана, линия

Автоматически созданное описание

Рисунок 15 – Методология Гартнера

На основе данных, отмеченных на кривой, были выбраны и занесены в таблицу 2 три релевантные технологии.

Machine learning (Машинное обучение) – это подраздел искусственного интеллекта, который позволяет компьютерам обучаться и делать прогнозы или принимать решения на основе данных. Машинное обучение использует алгоритмы и статистические модели для обучения компьютеров выполнению задач без явного программирования.

Code generation (Генерация кода) – это процесс, в котором компьютерная программа (называемая генератором кода) автоматически создает программный код на основе входных данных. Это может значительно упростить процесс разработки, поскольку генераторы кода могут автоматически создавать код для рутинных или повторяющихся задач.

Decentralized Identity (Децентрализованная идентификация) – это концепция, которая дает каждому пользователю возможность управлять своей цифровой идентичностью без необходимости идентификации через централизованный сервер или провайдер. Вместо этого, децентрализованная идентификация использует технологии, такие как блокчейн, чтобы создать уникальный, верифицируемый и неподдельный цифровой идентификатор.

Таблица 2 – Список релевантных технологий

|  |  |  |  |
| --- | --- | --- | --- |
| Название технологии на английском и на русском языке | Стадия (Запуск технологии; Пик завышенных ожиданий; Пропасть разочарования; Склон просвещения; Плато продуктивности) | Горизонт выхода на плато продуктивности (Менее 2-х лет; 2-5 лет; 5-10 лет; более 10 лет) | Компании-лидеры, первопроходц |
| Machine learning (Машинное обучение) | Инновационный триггер | 5-10 лет | DataToBiz,  Amazon Web Service |
| Code generation (Генерация кода) | Инновационный триггер | 5-10 лет | GitHub Copilot, OpenAI |
| Decentralized Identity (Децентрализованная идентификация) | Пик завышенных ожиданий | 5-10 ktn | Microsoft Entra Verified ID, BitMark |

## **2.3 Выводы и рекомендации**

Анализ данного раздела аналитического обзора подтверждает актуальность выбранной темы. На основе изученных в этом разделе научных исследований, за последние 9 лет наблюдался рост числа публикаций. В ЮФУ также работают специалисты в области обработки естественного языка, которые опубликовали более 700 статей. Среди них - преподаватели Солдатов А. В., Гуда А. А., Бутакова М. А., Чернов А. В. Однако, стоит отметить, что цитирование публикаций по данной теме уменьшается. Тем не менее, в ЮФУ есть эксперты, которые могут помочь с возникающими вопросами в процессе написания диссертации.

# **3 Методология исследований**

Данный раздел включает в себя выбор и обоснование актуальности темы ВКР, постановку этапов (задач) для её написания.

## **3.1 Выбор тематики выпускной квалификационной работы**

В качестве темы выпускной квалификационной работы было принято решение применить методы машинного обучения и технологий глубокого обучения для идентификации человека по голосу.

**Актуальность проблемы:** Идентификация человека по голосу является крайне важной и актуальной задачей в современном обществе. Она играет решающую роль в обеспечении безопасности, удобства и персонализации услуг во многих областях жизни.

В банковской сфере идентификация по голосу может служить дополнительным уровнем защиты от мошенничества. Это позволяет усилить безопасность транзакций и повысить уровень доверия клиентов, что важно для укрепления репутации банка и привлечения новых клиентов.

В телекоммуникациях и безопасности, голосовые данные используются для создания более надежных систем идентификации и авторизации, которые сложнее обмануть. Это позволяет усилить защиту данных, предотвратить несанкционированный доступ и обеспечить безопасность персональной информации пользователей.

С развитием технологий и увеличением количества голосовых интерфейсов, таких как голосовые помощники и автоматизированные системы ответов, актуальность этой проблемы только увеличивается. Это связано с тем, что возникает необходимость в более точной, быстрой и удобной идентификации для улучшения пользовательского опыта и повышения безопасности.

Однако идентификация по голосу не лишена сложностей. Она сталкивается с научными и техническими проблемами, связанными с изменчивостью человеческого голоса, фоновыми шумами, различиями в качестве записи и другими факторами. Эти проблемы добавляют сложности в процесс идентификации и делают ее актуальной для проведения исследований.

В дополнение к этому, голос уникален для каждого отдельного индивида и может меняться в зависимости от множества факторов, включая эмоциональное состояние, заболевания, усталость и даже время суток. Это добавляет дополнительные сложности в процесс идентификации.

Все эти факторы делают исследование в области идентификации человека по голосу крайне актуальным и важным для современного общества. Это открывает новые перспективы для научных разработок и инноваций в данной области и требует дальнейшего изучения и исследования данной проблемы.

**Основной идеей данной ВКР:** Способность человека издавать разнообразные звуки, или голос, является результатом эволюционного развития сложного голосового аппарата. Воздух, проходящий через звуковые связки, создает звуковые колебания, которые приобретают уникальные характеристики при прохождении через резонаторы, такие как трахея, носовые полости, ротоглоточная полость и гортань.

Благодаря возможности издавать звуки, у человека появилась и продолжает развиваться речь, средство сложной коммуникации между людьми. Это способствовало передаче и хранению информации, что сыграло заметную роль в формировании человека как социального существа.

Речь не только способствует обмену информацией, но и передает эмоциональное состояние, а также помогает определить личность говорящего. Часто человек интуитивно пользуется этим, а также способностью выделять речь другого человека среди фона шумов, что позволяет узнавать его голос.

Технологии распознавания речи значительно продвинулись за последние 20 лет. Современные системы не только распознают говорящего, но и понимают, что он говорит. Программное обеспечение для распознавания речи стало неотъемлемой частью электронных помощников и персональных секретарей, а системы распознавания личности по голосу используются для обеспечения безопасности доступа.

**Цель данной ВКР:** На компьютере каждого человека, как правило, находится большое количество информации. Часть этой информации не должна быть известна определенному кругу лиц. Для доступа к ней должны использоваться определенные методы идентификации, которые практически невозможно преодолеть.

**Основные решаемы задачи**:

* Изучение теоретических основ биометрической идентификации по голосу;
* Анализ существующих методов реализации программного средства идентификации личности по голосу;
* Разработка программного средства идентификации личности по голосу;
* Расчет экономической себестоимости программного средства;
* Анализ вредных факторов при разработке программного средства и расчет напряженности труда.

**Новизна и практическая значимость работы:** В данной работе подробно исследуется актуальная и сложная проблема идентификации человека по голосу. Это поле исследований было выбрано из-за его непрерывно растущей важности в эпоху цифровизации и глобализации.

Новизна этой работы заключается в использовании и комбинировании современных и продвинутых методов анализа и обработки аудиоданных, что обеспечивает создание невероятно надежной и точной системы идентификации. Этот подход, основанный на последних научных и технологических достижениях, отличает данную работу от предыдущих исследований в этой области.

Практическая значимость работы не может быть недооценена. Возможность использования разработанной системы в широком спектре областей делает ее универсальной и многофункциональной. Сферы применения включают, но не ограничиваются безопасностью, криминалистикой и сферой обслуживания клиентов. Это дает возможность для широкого применения и адаптации этой системы в различных отраслях и сферах деятельности.

В результате, данная работа может стать важным шагом на пути к созданию более безопасных и персонализированных систем идентификации. Это исследование не только обещает результаты на поле битвы с преступностью и обеспечением безопасности, но также открывает новые перспективы для улучшения качества обслуживания клиентов и создания более индивидуализированного опыта для каждого человека.

# **ЗАКЛЮЧЕНИЕ**

Исследование биометрического параметра, каковым является голос, открывает новые возможности для идентификации и аутентификации личности. В эпоху быстрого развития цифровизации и технологий в различных сферах жизни, вопросы безопасности и защиты личной информации становятся особенно актуальными.

Однако, несмотря на широкое распространение, существующие методы идентификации голоса не лишены недостатков. Одной из основных угроз являются поддельные голоса и атаки с использованием передовых технологий синтеза речи. Важность эффективных средств защиты усиливается с увеличением числа возможных угроз.

Стоит отметить, что эффективность и точность систем идентификации голоса оставляют много желать. Это подчеркивает необходимость дальнейшего исследования и улучшения текущих методов.

В свете вышеуказанных проблем, применение технологий машинного обучения, спектрального анализа и биометрических методов представляется перспективным направлением. Они могут обеспечить создание систем, способных с высокой точностью определить личность по голосовым данным.

Следует продолжать исследования в этой области для усовершенствования методов идентификации голоса и обеспечения безопасности данных. Это важно не только для отдельных людей или организаций, но и для общества в целом, учитывая растущую значимость цифровой безопасности в современном мире.

# **СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

1. **Meisam Khalil Arjmandi, Mohammad Pooyan, Mohammad Mikaili, Mansour Vali, Alireza Moqarehzadeh.** Identification of Voice Disorders Using Long-Time Features and Support Vector Machine With Different Feature Reduction Methods. *Sciencedirect.* [В Интернете] 6 Ноябрь 2011 г. https://www.sciencedirect.com/science/article/abs/pii/S0892199710001578.

2. **Katie Haejung Kim, Hyejoon Rim.** Employees’ voice behavior in response to corporate social irresponsibility (CSI): The role of organizational identification, issue perceptions, and power distance culture. *Sciencedirect.* [В Интернете] 4 11 2013 г. [Цитировано: 12 12 2023 г.] https://www.sciencedirect.com/science/article/abs/pii/S0363811123000814. CSI.

3. **Lin, Qingjin.** Transformational leadership and innovative work behavior: The role of identification, voice and innovation climate. *Sciencedirect.* [В Интернете] 8 08 2023 г. https://www.sciencedirect.com/science/article/abs/pii/S0278431923000956.

4. **Ganli Liao, Jing Wang, Qichao Zhang, Xin Ding.** The quality of crowdsourcing virtual community and users’ voice behavior: An analysis of stimulus-organism-response framework among Chinese users. *Sciencedirect.* [В Интернете] 14 09 2023 г. https://www.sciencedirect.com/science/article/pii/S2405844024029128.

5. **Nabanita Basu, Agnes S. Bali, Philip Weber, Claudia Rosas-Aguilar, Gary Edmond d 5, Kristy A. Martire b 6, Geoffrey Stewart Morrison a e 7.** Speaker identification in courtroom contexts – Part I: Individual listeners compared to forensic voice comparison based on automatic-speaker-recognition technology. *Sciencedirect.* [В Интернете] 10 12 2022 г. https://www.sciencedirect.com/science/article/pii/S0379073822003292.

6. **Shibani Hamsa, Ismail Shahin, Youssef Iraqi, Ernesto Damiani, Ali Bou Nassif d, Naoufel Werghi a.** Speaker identification from emotional and noisy speech using learned voice segregation and speech VGG. *Sciencedirect.* [В Интернете] 23 08 2023 г. https://www.sciencedirect.com/science/article/abs/pii/S095741742300372X.

7. **Hannah White, Joshua Penney, Andy Gibson, Anita Szakay, Felicity Cox.** Influence of pitch and speaker gender on perception of creaky voice. *Sciencedirect.* [В Интернете] 20 01 2023 г. https://www.sciencedirect.com/science/article/pii/S0095447023000827.

8. **Swapna Sebastian, Suma Susan Mathews, Mahasampath Gowri, Manish Kumar, John Mathew.** Can There be a Biometric Parameter for Voice?—An Investigation Into the Vocal Cry of Twins. *Sciencedirect.* [В Интернете] 20 10 2023 г. https://www.sciencedirect.com/science/article/abs/pii/S0892199723002989.

9. **Mavaddati, Samira.** Voice-based age, gender, and language recognition based on ResNet deep model and transfer learning in spectro-temporal domain. *Sciencedirect.* [В Интернете] 01 05 2023 г. https://www.sciencedirect.com/science/article/abs/pii/S0925231224002005.

10. **Hanna Mackiewicz-Nartowicz, Lidia Nawrocka, Paweł Burduk, Anna Sinkiewicz.** Voice Rehabilitation After Laryngeal Transplantation: Own Experience. *Sciencedirect.* [В Интернете] 04 03 2023 г. https://www.sciencedirect.com/science/article/abs/pii/S0892199724000213.

11. **БРЮХОМИЦКИЙ Ю.А., ФЕДОРОВ В.М.** Метод текстонезависимой идентификации личности по голосу. *РИНЦ.* [В Интернете] 30 03 2018 г. https://elibrary.ru/item.asp?id=38175897.

12. **ГАЛИАХМЕТОВ К.В., СМЕТАНИНА О.Н., ШАРИПОВА М.А.** МЕТОДЫ И СРЕДСТВА РЕШЕНИЯ ЗАДАЧИ ИДЕНТИФИКАЦИИ ЛИЧНОСТИ ПО ГОЛОСУ. *РИНЦ.* [В Интернете] 16 05 2017 г. https://elibrary.ru/item.asp?id=36345187.

13. **АЛИМУРАДОВ А.К., АЛИЕВ К.А., МУРТАЗОВ Ф.Ш.** Основные методы идентификации диктора по голосу. *РИНЦ.* [В Интернете] 08 07 2012 г. https://elibrary.ru/item.asp?id=20919363.

14. **Р.А., ВАСИЛЬЕВ.** Применение метода идентификации по голосу, адаптированного к тихому произнесению парольных фраз для противодействия утечки речевой информации. *РИНЦ.* [В Интернете] 25 06 2021 г. https://elibrary.ru/item.asp?id=45765985.

15. **П.С., ФУРСОВА.** Методы идентификации человека по голосу. *РИНЦ.* [В Интернете] 19 03 2023 г. https://elibrary.ru/item.asp?id=60047301.

16. **БРЮХОМИЦКИЙ Ю.А., ФЕДОРОВ В.М.** Иммунологический метод текстонезависимой идентификации личности по голосу. *РИНЦ.* [В Интернете] 24 09 2018 г. https://elibrary.ru/item.asp?id=35727168.

17. **И.В., КУРЬЯНОВА.** Возможности идентификации иноязычных говорящих по голосу и речи экспертными методами. *РИНЦ.* [В Интернете] 18 05 2018 г. https://elibrary.ru/item.asp?id=36485870.

18. **Ana Alice Leal dos Santos, Eliane Cristina Pereira, Juliana Marcolino, Ana Paula Dassiê-Leite.** Self-perception and voice quality of journalism students. *Dimensions.* [В Интернете] 07 06 2022 г. https://app.dimensions.ai/details/data\_set/dataset.58353050?search\_mode=content&search\_text=Voice%20identification%20methods&search\_type=kws&search\_field=full\_search.

19. **Jeanne Gabriele Schmidt, Bárbara Niegia Garcia de Goulart, Maria Elza Kazumi Yamaguti Dorfman, Gabriel Kuhl, Lauren Medeiros Paniagua.** Voice challenge in transgender women: trans women self-perception of voice handicap as compared to gender perception of naïve listeners. *Dimensions.* [В Интернете] 28 02 2018 г. https://app.dimensions.ai/details/data\_set/dataset.7879390?search\_mode=content&search\_text=Voice%20identification%20methods&search\_type=kws&search\_field=full\_search.