

Министерство образования Республики Беларусь  
Учреждение образования  
Белорусский государственный университет  
информатики и радиоэлектроники

УДК 004.522:004.056.53

КРИЩЕНОВИЧ  
Вероника Анатольевна

**ВЕРИФИКАЦИЯ СУБЪЕКТОВ ДОСТУПА  
НА ОСНОВЕ АНАЛИЗА РЕЧЕВОГО СИГНАЛА**

Автореферат  
на соискание степени магистра  
по специальности 1-98 80 01 Информационная безопасность

---

Научный руководитель  
кандидат технических наук, доцент  
ЗАХАРЬЕВ Вадим Анатольевич

---

Минск 2023

## ВВЕДЕНИЕ

Речь – существенный элемент человеческой деятельности, позволяющий человеку познавать окружающий мир, передавать свои знания и опыт другим людям, аккумулировать их для передачи последующим поколениям. Начиная с древних времен и по сей день она является основным способом обмена информацией между людьми.

Люди в процессе общения способны подсознательно различать голоса других людей. Это связано с тем, что характеристики голоса определяются анатомическими особенностями строения речевого аппарата, которые индивидуальны для каждого человека. Уникальность голоса послужила предпосылкой создания биометрических систем, использующих голос человека в качестве входной информации – распознавания диктора, подразделяющихся в свою очередь на системы идентификации и верификации диктора.

Задача распознавания диктора по его голосу была поставлена более 40 лет назад, и исследования в этой области еще продолжаются. Ниже приведены лишь некоторые практические задачи, решение которых требует применения системы распознавания диктора.

– Биометрический контроль доступа. Системы биометрического контроля доступа предназначены для обеспечения безопасности доступа к физическим объектам, информационным и финансовым ресурсам. Верификация человека по голосу не требует применения специализированного дорогостоящего оборудования. Все, что необходимо – обычный микрофон. При этом по уровню надежности голосовая биометрия не уступает, а по некоторым характеристикам превосходит характеристики других систем биометрической идентификации.

– Сопоставление голоса подозреваемого и некоторой фонограммы. Технология автоматического распознавания диктора по голосу уже сейчас используется в современных лабораториях криминалистических исследований как средство анализа фонограмм подозреваемых.

– Автоматическое управление тангентой в радиостанциях. Полудуплексный режим работы широко используется в радиосвязи, однако в условиях занятости рук он может создавать неудобства для пользователя. В качестве решения обычно применяют детекторы речевой активности для автоматического управления тангентой. Однако при этом система может ошибочно активировать тангенту от голоса постороннего человека. Одним из перспективных способов избежать этого является добавление модуля распознавания диктора в радиостанцию.

– Голосовое управление роботом. Управление с помощью голосовых команд является одним из важнейших естественных способов взаимодействия с роботом. Однако существуют приложения, в которых необходимо ограничить круг лиц, имеющих доступ к управлению. В этом случае перед распознаванием речевой команды возможно выполнять верификацию диктора.

– Голосовое управление подсистемами автомобиля. Голосовое управление такими подсистемами, как кондиционер, навигатор, медиаплеер, уже сейчас внедряется во многие модели автомобилей. Актуальным является создание индивидуальных профилей голосового управления для разных людей, что может быть реализовано с помощью системы распознавания диктора.

Таким образом направление исследований в области верификации диктора по речевому сигналу является весьма актуальным.

**Целью работы** является анализ существующих и разработка новых алгоритмов верификации диктора для защиты технических систем от несанкционированного доступа на основе речевого сигнала.

В соответствии с указанной целью в работе поставлены и решены следующие задачи:

1. Рассмотреть существующие методы и алгоритмы распознавания идентификации и верификации диктора по речевому сигналу;
2. Проанализировать существующие модели и методы определения эмоционального состояния по речевому сигналу и выбрать один из них для дальнейшего исследования;
3. Проанализировать выбранный метод и разработать модифицированный алгоритм, на его основе.
4. Провести экспериментальные исследования предлагаемого алгоритма на основе программного прототипа.

**Объектом исследования** являются системы автоматической верификации диктора для защиты технических систем от несанкционированного доступа.

**Предметом исследования** особенности применения индивидуальных характеристик голоса диктора для организации процесса верификации субъектов доступа в системах защиты информации.

В первой главе диссертационной работы был выполнен анализ предметной области по теме диссертационной работы, произведен обзор мирового и отечественного рынка биометрических технологий. Данный анализ показал высокую степень актуальности задачи верификации на основе речевого сигнала в современных информационных системах, требующих высокого уровня безопасности при обеспечении доступа к хранимой в них информации. Рассмотрена классификация систем и основные аспекты,

связанные с разработкой систем распознавания и верификации диктора, например, такие как текстозависимый и текстонезависимый режим работы системы.

Во второй главе диссертационной работы были рассмотрены основные группы методов, используемых для решения задачи верификации диктора: методы извлечения голосовых признаков, несущих информацию об особенностях конкретного диктора, а также основные методы и модели их классификации. Представлено обоснование выбора базовых методов для выполнения анализа сигнала, выделения векторов параметров, характеризующих голос конкретного диктора, их классификации и методов построения модели диктора.

В третьей главе представлена разработка алгоритма верификации на основе анализа речевого сигнала. Приведена блок-схема алгоритма и детальное описание выполнения его шагов. Для каждого шага алгоритма приведена его реализация на языке программирования Python. Выполнена экспериментальная оценка качества работы предлагаемого базового алгоритма. Предложен вариант улучшения алгоритма за счет добавления в качестве дополнительного этапа выделения характеристического вектора глубокой искусственной нейронной сети на основе архитектуры WavLM, что позволяет повысить точность верификации диктора в том числе и в зашумленной акустической обстановке.

## **ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ**

### **Связь работы с крупными научными программами**

Тема диссертационной работы соответствует пункту 6 приоритетных направлений научной, научно–технической и инновационной деятельности Республики Беларусь на 2021 –2025 гг., утвержденных Указом Президента Республики Беларусь №156 от 7 мая 2020 г. «*Обеспечение безопасности человека, общества, государства*». Работа выполнялась в учреждении образования «Белорусский государственный университет информатики и радиоэлектроники».

### **Цель и задачи исследования**

Целью диссертационной работы является повышение эффективности методов и алгоритмов защиты от несанкционированного доступа на основе речевого сигнала.

Для достижения поставленной цели в диссертации решены следующие задачи:

1. Проанализировать существующие методы обработки и определения характеристик диктора по речевому сигналу.
2. Исследовать существующие методы верификаций.
3. Разработать программный прототип, реализующий предложенные методы защиты от несанкционированного доступа на основе речевого сигнала.

### **Личный вклад соискателя**

Соискателем выполнены все изложенные в работе разработки и исследования. Постановка задач и обсуждение результатов проводились совместно с научным руководителем и сотрудниками кафедры защиты информации Белорусского государственного университета информатики и радиоэлектроники. Соавторы опубликованных работ принимали участие в обсуждении промежуточных и конечных результатов. Обработка, интерпретация данных, а также выводы сделаны автором самостоятельно.

### **Апробация результатов диссертации**

Основные положения диссертационной работы докладывались на следующих научных конференциях:

- «Информационные технологии и системы 2021» (Минск, 2021);
- «58-я научная конференция аспирантов, магистрантов и студентов: инфокоммуникации» (Минск, 2022).

## КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

**Во введении** рассмотрены проблемы необходимости применения систем распознавания диктора по речевому сигналу, обусловленные такими факторами как: стремительный рост количества мобильных потребительских устройств, необходимостью повышения надёжности механизмов защиты информации, востребованность технологий беспарольного доступа TouchID, FaceID и т.д., ростом рынка биометрических технологий

**В общей характеристике работы** показана связь работы с приоритетными направлениями научных исследований, цель и задачи исследования, личный вклад соискателя ученой степени, апробация результатов диссертации.

**В первой** главе диссертационной работы был выполнен анализ предметной области по теме данной работы, произведен мониторинг мирового и отечественного рынков биометрических технологий, который показал высокую степень актуальности персонификации пользователя в современных информационных системах, требующих высокого уровня безопасности хранимой в них информации. Произведенное исследование показало, что механизмы биометрической верификации уже внедрены в использование, в частности, голосовая биометрия, внедрение которой в нашей стране планируется осуществить в ближайшее время в рамках Единая система идентификации физических и юридических лиц.

**Во второй** главе диссертационной работы были рассмотрены две группы методов, используемых для решения задачи идентификации диктора: методы извлечения голосовых признаков, несущих необходимую информацию, используемую при идентификации диктора, и методы их классификации.

Среди методов первой группы было уделено внимание коэффициентам линейного предсказания и MFCC-коэффициентам. Мел-частотные кепстральные коэффициенты в последнее время приобрели высокую популярность и достаточно эффективно используются в задаче автоматического распознавания диктора. Это возможно благодаря тому факту, что в спектре речи диктора отражается структура речевого тракта, которая позволяет отличаться голосам людей на физиологическом уровне. В силу своей эффективности данный метод является более предпочтительным для решения поставленной задачи.

Среди методов второй группы предпочтение в качестве базового метода реализации алгоритма верификации был выбран метод векторного квантования, поскольку данный метод может эффективно применяться в случае небольшого количества обучающих данных доступных для

формирования модели диктора. Данный случай является частым сценарием, поскольку разработчики системы не могут предлагать пользователю записывать большое количество образцов голоса, поскольку это вызывает дискомфорт может привести к отказу от использования подобной системы. Также данный метод обладает минимальной алгоритмической задержкой на платформах и устройствах с ограниченными вычислительными ресурсами, такими как мобильные телефоны или одноплатные компьютеры.

**В** **третье** главе был проведен анализ и разработка алгоритма верификации на основе речевого сигнала для определения субъектов доступа. Вначале был разработан базовый алгоритм верификации на основе векторного квантования, который показал значение F-меры равное 0,768 при соотношении сигнал/шум 18 дБ. Затем этот базовый алгоритм был доработан с использованием нейронной сети, что привело к значению F-меры равному 0,955. Что позволило увеличить точность на 18,7% при заданном.

Базовый алгоритм верификации на основе векторного квантования использовался для извлечения признаков речевого сигнала и построения кодовых книг дикторов. Однако, с использованием нейронной сети, в данном случае модели WavLM, удалось добиться более высокой точности и эффективности в процессе верификации.

Полученные результаты показывают, что использование нейронной сети в алгоритме верификации на основе речевого сигнала значительно повышает его качество. Значение F-меры, равное 0,955, свидетельствует о высокой точности и надежности алгоритма в определении схожести дикторов. Это позволяет достичь более надежной и безопасной системы идентификации субъектов доступа на основе их голосового сигнала.

Дальнейшая разработка и оптимизация алгоритма могут привести к еще более высоким показателям эффективности и точности. Например, можно использовать более сложные модели нейронных сетей или внести дополнительные этапы предварительной обработки аудио записей. Это может помочь улучшить различение между разными дикторами и добиться еще более высоких значений F-меры и общей эффективности алгоритма.

## ЗАКЛЮЧЕНИЕ

В данной диссертационной работе была выполнена разработка алгоритма верификации на основе речевого сигнала для определения субъектов доступа. В процессе исследования предметной области было обнаружено высокую актуальность персонификации пользователя в современных информационных системах, где безопасность хранимой информации является важным аспектом.

В диссертационной работе были рассмотрены основные математические методы извлечения информативных голосовых характеристик диктора, используемые в системах распознавания мел-частотные кепстральные коэффициенты и коэффициенты линейного предсказания, а также методы сопоставления данных характеристик векторное квантование, модель гауссовых смесей, нейросети. Была рассмотрена структура типичной системы распознавания диктора, которая включает этапы предварительной обработки речевых данных, обучение и тестирование.

Разработан базовый алгоритм верификации на основе векторного квантования использовался для извлечения признаков речевого сигнала и построения кодовых книг дикторов. Представлена реализация данного алгоритма на языке программирования Python, как непосредственно для процесса верификации диктора (основной сценарий использования) и так и в режиме распознавания диктора (дополнительный сценарий, для тестирования системы). Были разработаны функции реализующие основные этапы выполнения алгоритма листинг которых представлен в приложении к диссертационной работе.

Тестирование алгоритма осуществлялась на корпусе речевых аудиоматериалов, включающих голоса 8 дикторов (4 мужских и 4 женских). Аудиозаписи были подготовлены на основе фонетически сбалансированных текстов на основе материалов открытого датасета «LibriSpeech». Для каждого диктора было записано 100 аудиозаписей для тренировочной выборки и по 30 аудиозаписей для тестовой выборки. Обучение и тестирование проводилось в текстозависимом режиме. Аудиозаписи были сохранены в формате wav с частотой дискретизации 16 кГц и разрешением 16 бит.

Для оценки качества написанного алгоритма проводилось его тестирование, которое показало относительно высокую точность (0,879) и полноту (0,878) распознавания при отсутствии акустических помех. При хорошем соотношении сигнал/шум (25 дБ) точность идентификации составляет 88% ( $F = 0,878$ ), но при уменьшении соотношения сигнал/шума до 18 дБ, точность снижается на 15%.



Для повышения точности верификации диктора, как обычной, так и зашумлённой обстановке, в работе предлагается использовать глубокие нейронные сети, с учетом использования предобученных моделей для решения задач верификации диктора. Использование предобученных сетей для выделения векторов признаков дикторов после мел-кепстрального анализа и построения кодовых книг на основе данных embedding-ов является эффективным подходом, который может помочь улучшить процесс верификации

В частности, было предложено использование предобученной сети на основе архитектуры WavLM. Архитектура WavLM использует сверточные блоки для изучения локальных и глобальных зависимостей в аудиоданных. Каждый сверточный блок состоит из слоев свертки с различными фильтрами и активационными функциями, что позволяет моделировать сложные шаблоны и зависимости в речевых сигналах. В алгоритме данная сеть применяется после выделения мел-кепстральных коэффициентов для извлечения характеристических векторов (embeddings) на основе MFCC-коэффициентов и построения модели диктора (кодовой книге) уже на основе данного представления характеристик голоса диктора. Характеристические вектора (embedding) представляют сжатую и более выразительную форму представления признаков и позволяют получить более компактные векторы для каждого диктора. Применение данного подхода позволило повысить точность верификации до 0,959 и полноту 0,951, а также увеличить общую точность распознавания на 18,7% при соотношении сигнала/шум 18 дБ.

В целом, данная диссертационная работа показала, что использование комбинации методов извлечения голосовых признаков и нейронных сетей в алгоритме верификации на основе речевого сигнала приводит к значительному повышению качества и точности идентификации диктора. Открытым вопросом остается дальнейшая оптимизация и развитие алгоритма для еще более высоких показателей эффективности и надежности.

В заключение, результаты данной диссертационной работы могут быть использованы в различных информационных системах, где требуется надежная идентификация субъектов доступа на основе их голосового сигнала. Дальнейшие исследования в данной области могут привести к новым технологическим разработкам и улучшению систем безопасности, основанных на голосовой биометрии.

## СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ

1. Крищенко, В. А. Системы верификации субъектов доступа на основе речевого сигнала / В. А. Крищенко, В. А. Захарьев // Информационные технологии и системы 2021 (ИТС 2021) = Information Technologies and Systems 2021 (ITS 2021) : материалы международной научной конференции, Минск, 24 ноября 2021 г. / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: Л. Ю. Шилин [и др.]. – Минск, 2021. – С. 42–43.

2. Крищенко, В. А. Универсальная фоновая модель для задач верификации диктора / Крищенко В. А. // Инфокоммуникации : сборник тезисов докладов 58-ой научной конференции аспирантов, магистрантов и студентов, Минск, 18–22 апреля 2022 г. / Белорусский государственный университет информатики и радиоэлектроники. – Минск: БГУИР, 2022. – С. 40–42.