

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

УДК \_\_\_\_\_

Акунович  
Антон Андреевич

Алгоритм идентификации человека по голосу

**АВТОРЕФЕРАТ**

на соискание степени магистра технических наук

по специальности 1-39 80 02 «Радиотехника, в том числе системы и  
устройства радионавигации, радиолокации и телевидения»

---

Научный руководитель:  
Карпушкин Эдуард Михайлович  
Кандидат технических наук, доцент

---

Минск 2019

## ВВЕДЕНИЕ

Задача идентификации человека по голосу была поставлена несколько десятилетий тому назад, с тех пор она не потеряла своей актуальности. Задача распознавания по голосу может найти применение в таких областях как криминалистика, разведка, защита несанкционированного доступа, в банковской сфере и финансах и т.д. Использование систем распознавания диктора является наиболее естественным и экономичным способом решения проблем неавторизованного доступа к компьютеру или системам передачи информации, а также проблем многоуровневого контроля доступа к сетевым или информационным ресурсам. Использование речи человека для задач распознавания имеет большой потенциал, так как единственным требованием, предъявляемым к оборудованию, является наличие микрофона, а уровень сигнал/шум записи может быть довольно низок.

Распознавание человека делится на два направления: верификацию и идентификацию. При верификации человек предъявляет свой идентификатор в том или ином виде, и система должна подтвердить или опровергнуть этот идентификатор. При идентификации человек не указывает какого-либо идентификатора, и система должна установить, принадлежит ли голос данного человека одному из дикторов, прошедших обучение. Важный момент системы идентификации заключается в определении того, принадлежат ли два голоса одному и тому же человеку или нет. Эти два вида идентификации существенно отличаются по условиям принятия решений. Поэтому данный вид идентификации можно разделить на систему с обучением и идентификацию без учителя.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

В работе рассматривались различные способы и подходы, а так же был предложен алгоритм идентификации диктора по голосу. Основным предметом исследования являлся поиск способа текстонезависимой идентификации по голосу, хотя другие типы идентификации и верификации также были рассмотрены.

В ходе работы был проведён обзор литературных источников по теме исследования. Основное внимание было уделено таким моментам как классификация задач распознавания диктора, показатели эффективности систем идентификации, анализ речи и информационные признаки голоса, а также решающие правила.

Разработка алгоритма идентификации состояла из трёх этапов: разработке алгоритма цифровой фильтрации, поиск алгоритма для выделения информационных признаков голоса диктора, а также разработке алгоритма классификации дикторов.

Исследование эффективности алгоритмов идентификации происходило на основе сравнительного анализа. Анализ проводился на основе собранной базы данных голосов различных дикторов. Исследование проводилась путём варьирования различных параметров, таких как отношение сигнал/шум, размер входного вектора, частотный диапазон и т.д. Все результаты моделирования были представлены в виде таблиц и графиков функций. На основе этих данных были сделаны выводы о применимости алгоритма для тех или иных задач, а также было проведено обсуждение преимуществ и недостатков предложенных алгоритмов.

## КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Для построения качественной системы распознавания диктора по голосу, необходимо провести анализ речи и выделение признаков и характеристик, характерных для речи данного диктора.

Среди множества признаков речевого сигнала, далеко не все одинаково важны и полезны. В идеале, речевой признак должен обладать рядом важных свойств:

- Иметь большую вариативность между дикторами, и при этом иметь малую вариативность внутри голоса одного диктора;
- Быть устойчивым к воздействию шумов и искажений;
- Быть легко извлекаемым из речевого сигнала;
- Быть трудным для имитации и подражания другим диктором;
- Не быть затронутым изменениями, связанными со здоровьем или иными долговременными вариациями в голосе.

С физической точки зрения, речевые признаки можно разделить на спектральные, просодические и высокоуровневые.

В работах по распознаванию диктора по голосу наибольшую популярность приобрёл метод кепстрального преобразования спектра. Схема этого метода следующая: на интервале времени в 10 – 20 мс вычисляется текущий спектр мощности, а затем применяется обратное преобразование Фурье от логарифма этого спектра (кепстр), и находятся коэффициенты этого кепстра.

Так как, любой микрофон или записывающее устройство вносит нежелательные шумы в полезный сигнал, то записываемую речь необходимо предварительно отфильтровать. Для данной задачи применяются фильтры.

Для нашей задачи было принято решение использовать цифровые фильтры, и тому есть ряд причин. Первая причина заключается в более высокой точности воспроизведения частотной характеристики по сравнению с аналоговыми фильтрами. Следующей причиной является возможность программной реализации цифровых фильтров на микропроцессоре или других вычислительных средствах. Цифровые фильтры можно разделить на два типа: рекурсивные и нерекурсивные. В нерекурсивных фильтрах каждый отсчёт выходного сигнала получается в результате умножения задержанных отсчётов входного сигнала на соответствующие весовые коэффициенты с последующим полученных произведений. Рекурсивный фильтр кроме отсчётов входного сигнала учитывает также выходные отсчёты, полученные на предыдущих интеграциях.

Для нашей задачи применялся однородный нерекурсивный фильтр, или фильтр скользящего среднего. Данные фильтры получили широкое распространение в цифровой обработке сигналов благодаря простоте их реализации. Кроме того, данный тип фильтров является оптимальным при решении задачи подавления аддитивного шума при одновременном сохранении скорости нарастания переходной характеристики.

Были выбраны два алгоритма идентификации: на основе статистического метода и на основе искусственных нейронных сетей. Для оценки эффективности разработанного алгоритма идентификации дикторов была собрана специальная база голосов. В качестве тестовой базы голосов была выбрана база голосов *voxforge*, которая находится в свободном доступе (<http://www.voxforge.org/>). База состоит из аудиозаписей 219 голосов длиной 5 – 7 сек., принадлежащих 11 дикторам (7 мужских голосов и 4 женских). Каждому диктору принадлежит по 20 аудиозаписей.

При отношении сигнал/шум равным 65 дБ статистический метод показал точность классификации 100% , а при отношении сигнал/шум 10 дБ – 76,4% для текстозависимой идентификации. При текстонезависимой идентификации точность классификации статистического метода упала до 53,2%. Размер окна никак не повлиял на точность классификации. Для нейронной сети точность классификации составила 96,3% при отношении сигнал/шум 65 дБ и 93,6% для отношения сигнал/шум 10 дБ.

## ЗАКЛЮЧЕНИЕ

В ходе данной работы был разработан алгоритм идентификации человека по голосу. Было предложено два алгоритма идентификации: на основе статистического метода и на основе искусственных сетей. Было проведено сравнение данных методов.

В ходе моделирования оба метода показали приемлемую точность классификации при отношении сигнал/шум 65 дБ (100% - для статистического метода и 96,3 – для нейронной сети). Однако при уменьшении отношения/сигнал шум, ситуация меняется. Так при отношении С/Ш = 10 дБ точность классификации статистического метода сильно падает (78,2%), в то время как точность классификации ИНС остаётся достаточно высокой (93,6%). Ширина окна не повлияла на точность классификации.

Таким образом мы можем сделать вывод о том, что оба метода могут быть использованы для решения задачи идентификации диктора по голосу. Статистический метод является простым и не затратным по времени алгоритмом классификации, который может применяться для дикторозависимой идентификации и при малом уровне шумов. Классификатор на основе нейронных сетей является более устойчивым к шумам и способен применяться также и при текстонезависимой идентификации. Однако его недостатком является необходимость большого количества данных для обучения, а также потребность в большей вычислительной производительности, чем для статистического метода.

## Список опубликованных работ

- [1] Акунович, А.А. Распознавание человека по голосу с использованием искусственных нейронных сетей / А.А. Акунович // Радиоинформационные технологии: Тезисы докл. к 53-й научной конференции аспирантов, магистрантов и студентов БГУИР– Минск, 2017 – С.5 – 6.
- [2] Акунович, А.А. Идентификация диктора по голосу с использованием искусственных нейронных сетей / А.А. Акунович // Радиоинформационные технологии: Тезисы докл. к 55-й научной конференции аспирантов, магистрантов и студентов БГУИР– Минск, 2019.
- [3] Акунович, А.А. Алгоритм идентификации диктора по голосу на основе статистического метода / А.А. Акунович // Радиоинформационные технологии: Тезисы докл. к 55-й научной конференции аспирантов, магистрантов и студентов БГУИР– Минск, 2019.