

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

УДК 004.934.8'1

**БОВКУНОВИЧ**  
Марина Владимировна

Автоматический поиск диктора по голосу

**АВТОРЕФЕРАТ**

на соискание степени магистра информатики и вычислительной техники

по специальности 1-40 81 01 «Информатика и технологии разработки  
программного обеспечения»

Научный руководитель  
Вашкевич Максим Иосифович  
кандидат технических наук, доцент

Минск 2017

Работа выполнена на кафедре информатики учреждения образования «Белорусский государственный университет информатики и радиоэлектроники»

Научный руководитель: **Вашкевич Максим Иосифович,**  
кандидат технических наук, доцент кафедры  
электронных вычислительных средств  
учреждения образования «Белорусский  
государственный университет информатики  
и радиоэлектроники»

Рецензент:

Защита диссертации состоится «    » января 2017 г. года в    часов на заседании Государственной комиссии по защите магистерских диссертаций в учреждении образования «Белорусский государственный университет информатики и радиоэлектроники» по адресу: 220013, г.Минск, ул. П.Бровки, 6, 5 уч. корп., ауд.    , тел.: 293-89-92, e-mail: [kafei@bsuir.by](mailto:kafei@bsuir.by).

С диссертацией можно ознакомиться в библиотеке учреждения образования «Белорусский государственный университет информатики и радиоэлектроники».

## **ВВЕДЕНИЕ**

Задача распознавания дикторов является актуальной задачей области речевых технологий. За последние годы наблюдается значительное повышение качества распознавания речевой информации, однако основная проблема автоматического распознавания диктора в любой среде все еще далека от решения. Именно поэтому актуальны как исследования уже существующих алгоритмов, так и поиск новых решений в данной области.

Обработки речи может быть разбита на задачи:

- анализа/синтеза;
- распознавания;
- кодирования.

Распознавание охватывает подзадачи

- распознавания речи;
- распознавания дикторов;
- идентификации языка.

Задача распознавания личности по голосу сводится к тому, чтобы выделить, классифицировать и соответствующим образом отреагировать на человеческую речь из входного звукового потока.

## **ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ**

### **Актуальность темы исследования**

В современном мире все большее значение уделяется интерфейсам, использующим речевой ввод и вывод для взаимодействия между пользователем и компьютером. Поэтому всё большее многообразие в голосовых сообщениях приходится принимать во внимание разработчику систем распознавания речи, реализующих акустический интерфейс.

Задача распознавания речи (во многих своих проявлениях: от транскрибирования слитной речи до верификации и идентификации

диктора) в настоящее время является актуальной. Свидетельством этому служит растущее число публикаций и конференций по данной тематике, а также то, что в крупнейших транснациональных корпорациях открываются департаменты, ориентированные на исследования в данной тематике.

Исследовательские усилия в сфере речевых технологий привели к появлению большого числа коммерческих систем распознавания речи. Улучшение существующих систем распознавания языка и диктора позволит существенно упростить взаимодействие человека с компьютером в том случае, когда использование классических интерфейсов невозможно или затруднено, а также сделать работу с компьютером или иной техникой более комфортной, например, для аутентификации пользователя. Также следует отметить, что применение систем распознавания диктора играет большую роль в работе правоохранительных органов.

Существующие системы распознавания речи в основном построены на Скрытых Марковских Моделях (НММ), которые задают динамику перехода от одной фонемы в речи к другой, а моделирование вероятностного распределения признаков происходит посредством Гауссовой Смеси (GMM). Для описания речевого сигнала в системах автоматического распознавания используются так называемые мел-частотные кепстральные коэффициенты (MFCC).

Также следует отметить, что в последнее время альтернативой используемым сейчас MFCC становятся признаки, устойчивые к вариабельности речевого тракта у диктора (например, bottleneck features).

Хорошим примером может служить работа сотовой станции или call-центра, где на обработку одновременно может приходиться огромное количество заявок, требующих обработки в реальном времени.

### **Цель и задачи исследования**

Целью диссертации является изучение существующих методов идентификации и разработка алгоритма поиска диктора по голосу.

Для выполнения поставленной цели в работе были сформулированы **следующие задачи:**

- исследовать существующие методы решения задачи идентификации личности по голосу, способы оценки качества, а также существующие проблемы и ограничения;
- разработать алгоритм поиска диктора по голосу;
- провести экспериментальную проверку разработанного алгоритма.

**Объектом** исследования является система поиска диктора по голосу.

**Предметом** работы выступают методы и алгоритмы поиска диктора по голосу.

**Методы исследования.** При решении поставленных задач использовались методы и понятия цифровой обработки сигналов, распознавания образов, алгоритмы и методы обработки данных.

**Информационная база** исследования для алгоритмов и методов обработки данных и распознавания образов.

**Научная новизна** диссертационной работы заключается исследовании методов и алгоритмов идентификации по голосу, разработке алгоритма распознавания диктора.

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

1. Анализ существующего состояния в сфере распознавания диктора.
2. Сравнительный анализ методов идентификации диктора по голосу.
3. Разработанный алгоритм распознавания диктора.
4. Созданная программная реализация разработанной системы идентификации диктора.
5. Проведенные экспериментальные исследования, по оценке точности распознавания.

**Теоретическая значимость** диссертации заключается в том, что в ней

предложен алгоритм разработки систем идентификации личности. Представлена система, реализованная на основе алгоритма.

**Практическая значимость** диссертации состоит в том, что разработанную систему можно использовать в качестве вторичных средств защиты.

Основные положения работы и результаты диссертации изложены в двух опубликованных работах общим объемом 2,0 п.л.

**Структура и объем работы.** Структура диссертационной работы обусловлена целью, задачами и логикой исследования. Работа состоит из введения, пяти глав и заключения, библиографического списка и приложения. Общий объем диссертации – 57 страниц. Работа содержит 3 таблицы, 15 рисунков. Библиографический список включает 22 наименования.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

**Во введении** рассмотрено современное состояние проблемы идентификации диктора, определены основные направления исследований, а также дается обоснование актуальности темы диссертационной работы.

**В общей характеристике работы** сформулированы ее цель и задачи, показана связь с научными программами и проектами, даны сведения об объекте исследования и обоснован его выбор, приведены сведения о личном вкладе соискателя, апробации результатов диссертации и их опубликованность, а также, структура и объем диссертации.

**В первой главе** проводится обзор и анализ современных систем идентификации личности, указываются основные проблемы, возникающие при распознавании диктора и рассматриваются перспективы использования данных систем.

**Во второй главе** описываются этапы обработки речевого сигнала и формирование характеристических векторов для систем идентификации диктора.

**В третьей главе** рассматриваются методы идентификации диктора по голосу.

**В четвертой главе** приведен алгоритм и программная реализация реализации системы распознавания по голосу.

**В третьей главе** представлены исследования необходимо провести для дальнейшего усовершенствования системы.

**В приложении** приведен исходный код разработанной системы.

## ЗАКЛЮЧЕНИЕ

В данной диссертационной работе исследовались существующие методы решения задачи автоматической идентификации диктора по голосу.

Задача идентификации диктора по голосу до сих пор является до конца не решенной. Современные методы машинного обучения и цифровой обработки сигналов используют все более эффективные способы решения данной задачи.

В качестве входных данных для системы использовались мел-частотные кепстральные коэффициенты MFCC, которые предварительно проходили процедуру нормализации для очистки от шума и минимизации влияния канала передачи звуковой записи на результаты работы системы.

Анализ существующих подходов к построению систем идентификации дикторов по голосу показал, что наиболее часто применяются следующие методы для их построения:

- метод ближайшего соседа;
- метод опорных векторов (SVM);
- модели гауссовых смесей (GMM);
- метод векторного квантования (VQ);
- методы, основанные на комбинации предыдущих.

В результате был проведен полный обзор предметной области и реализован один из алгоритмов, используемых для решения поставленной задачи.

В рамках дальнейшей работы можно рассмотреть задачу идентификации на открытом множестве и задачу разделения дикторов (speaker diarization). Задача идентификации на открытом множестве заключается в том, что система выдает отказ при получении в качестве входного речевого сигнала голоса пользователя, не зарегистрированного в системе. Задача speaker diarization заключается в том, что необходимо выделить и идентифицировать голос диктора из входного речевого сигнала, содержащего запись голосов нескольких людей.

### **Список опубликованных работ**

1. Бовкунович М.В. Сравнение методов идентификации дикторов по голосу. / М.В. Бовкунович // Материалы XII Международной научно-практической конференции молодых исследователей: «Содружество наук. Барановичи-2016». – Барановичи, 2016. – с. 255.

2. Бовкунович М.В. Методы начальной обработки звукового сигнала. / М.В. Бовкунович // Материалы Международной научно-практической конференции «Экономика, технологии и право в современном мире». – Барановичи, 2016. – с. 273.