

МЕКЕБАЕВ НУРБАПА ОТАНОВИЧ

**«РАЗРАБОТКА И ИССЛЕДОВАНИЕ АЛГОРИТМОВ
ИДЕНТИФИКАЦИИ И ОБРАБОТКИ ПРИЗНАКОВ В ЗАДАЧАХ
РАСПОЗНАВАНИЯ РЕЧИ С ПРИМЕНЕНИЕМ МАШИННОГО
ОБУЧЕНИЯ»**

АННОТАЦИЯ

диссертации на соискание степени доктора философии (PhD) по
специальности «6D060200 – Информатика»

Актуальность темы исследования. В настоящее время одной из самых актуальных проблем информатики является проблема распознавания речи. Прежде всего, от этого зависит эффективность использования компьютерных систем, т.к. речь это наиболее распространенный и естественный для человека способ общения, который позволяет значительно ускорить процесс ввода информации и управления мобильными системами. Информационные технологии быстро развиваются и широко используются в обмене информацией и коммуникаций. В связи с этим она играет важную роль в развитии распознавания речи.

В повседневной жизни язык является естественным выражением человеческой природы. Хорошо известно, что на протяжении многих лет ученые и инженеры изучали вопрос о словесной коммуникации между пользователем и машиной по мере развития науки и техники.

Необходимо признать, что многие компании и отдельные разработчики добились определенных успехов в создании технологий распознавания речи, но необходимо признать, что в Казахстане они до сих пор не нашли широкого распространения. Это связано с особенностями дикции говорящего и наличием звуковых помех.

В связи с этим наиболее актуальной задачей решения проблемы является идентификация диктора при автоматическом распознавании речи.

Оптимизация существующих систем распознавания речи может значительно облегчить и повысить эффективность взаимодействия человека и компьютера. Использование систем распознавания речи также имеет большое значение в работе правоохранительных органов.

Актуальность проблемы распознавания речи усугубляется целым рядом факторов – различной структурой языков, особенностями произношения диктора, наложением шумов и помех, низкой скоростью работ систем.

Современные системы распознавания речи, в основном основаны на использовании скрытых моделей Маркова (СММ), которые определяют вероятность того, что одна фонема в речи может быть объединена с другой фонемой. Таким образом, дисперсия наблюдаемого сигнала достигается путем моделирования возможного деления знаков Гауссовым соединением (GF).

Этот метод был предложен Лоуренсом Робимером в 1989 году и послужил основой для моделирования длительной речи. Благодаря быстрому развитию сети Deep Belief Networks стали альтернативой СММ и смогли обеспечить высокую точность процесса распознавания.

Со времени публикации работ Л. Райнера мел кепстральный коэффициент (MFCC Mel Frequency Cepstral Coefficient) широко использовались для описания речевого сигнала в системах автоматического распознавания речи, которые были основаны Полом Мермельштейном. Кроме того, произведенная недавно замена маркера MFCC на изменения в динамике потока способствовала созданию более надежных систем.

В настоящее время многие зарубежные компании (Agnitio, Nuance, Voice Security Systems) в том числе и российские (Speech technologies, Speech Technologies Center), разрабатывают методы голосовой биометрии, имеющие небольшую ошибку 1-3%, но у этих приложений есть ряд недостатков.

Университеты США, в частности Университет Карнеги-Меллона, Университет Иллинойса Университет Орегона, Университет Восточной Финляндии, добились значительных успехов в идентификации речи. Представляют интерес и работы ученых СНГ и Казахстана в этой области (Шелепов В.Ю., (2007) А.А.Карпов, Аграновский А.Б., Леднов Д.А., Балакирев Н.Е., Малков М.А. и отечественные ученые (2008) А.Шарипбай, У.А.Тукеев, Есенбаев Ж.А., Амиргалиев Е.Н., Мусабаев Р.Р., Мамырбаев О.Ж.)

Машинное обучение стало ключевой частью процесса распознавания речи как одной из областей искусственного интеллекта. Создание речевых интерфейсов для быстрого управления информационными и коммуникационными технологиями, разработка эффективных систем распознавания речи являются одной из актуальных проблем, требующих своевременного решения.

Цель диссертационной работы: Разработка и исследование оптимальных алгоритмов и моделей распознавания речи с использованием машинного обучения в распознавании речи.

Задачи исследования:

- анализ современных систем и методов распознавания речевых сигналов и идентификация диктора в области распознавания речи;
- создание акустического корпуса для звуковых сигналов и идентификация данных диктора в процессе распознавания речи;
- сравнение современных классификационных моделей и алгоритмов идентификации диктора;
- разработка модели и алгоритма идентификации диктора и гендерных особенностей на основе нейронных сетей;

Объект исследования. Система распознавания и идентификации речи.

Область научного исследования: Акустические данные речи, методы и алгоритмы идентификации диктора в процессе автоматического распознавания речи.

Методы исследования: Теория информации, теория сигналов, теория и технологии нейронных сетей, методы и технологии распознавания речи, теория и технологии видеораспознавания, технологии проектирования и построения программного обеспечения.

Научная новизна работы:

- создан акустический корпус для речевых звуков говорящего и данных идентификации диктора;
- разработаны классификационные алгоритмы идентификации диктора с использованием машинного обучения;
- разработаны модели и алгоритмы определения гендерных особенностей и идентификация диктора на основе нейронных сетей;
- с помощью модели и алгоритма, полученных в ходе исследования, создано программное приложение для характеристики речевых признаков и идентификации диктора.

Теоретическая и практическая значимость работы. Теоретическая значимость исследования заключается в совершенствовании теории и технологий нейронных сетей, методов распознавания речи, разработки моделей и алгоритмов выявления гендерных различий в распознавании речи.

Практическая значимость диссертационного исследования заключается в использовании современных нейросетевых моделей для идентификации речевых сигналов, экспериментальном изучении новых методов выявления расширенных и специфических особенностей речи. Встроенный акустический корпус позволяет проводить дальнейшие исследования в области распознавания речи.

Основные положения, выносимые на защиту:

- рассмотрена общая структура и усовершенствованные архитектуры нейронных сетей в распознавании речи;
- разработка алгоритмов и моделей идентификации речевых особенностей диктора, зарегистрированных в базе данных;
- разработка алгоритмов и моделей, используемые для идентификации гендерных особенностей;
- разработка акустического корпуса казахского языка, состоящий из более чем 36-часовых речей, предназначенных для распознавания и идентификации признаков речи.

Личный вклад исследователя. Соискатель решил поставленные задачи самостоятельно. Разработаны алгоритм и модель гендерной идентичности и идентификации диктора, проведена экспериментальная оценка эффективности разработанных алгоритмов. Автором была создана модель идентификации и гендерных особенностей речи и идентификации диктора.

Связь темы диссертации с планами научно-исследовательской работы. Диссертационное исследование было выполнено в рамках проекта грантового финансирования «Разработка технологии многоязычного автоматического распознавания речи с использованием глубоких нейронных сетей» в Институте информационных и вычислительных технологий

Комитета науки Министерства образования и науки. (2018-2020, государственный регистрационный номер: 0118RK00139).

Объем и структура работы. Диссертация состоит из введения, трех разделов, заключения и списка литературы. Общий объем диссертации: 108 страниц письменного текста, в том числе 35 рисунка, 10 таблиц, список литературы из 110 источников, 2 приложения.

В разделе 1 анализируются особенности способов предварительной обработки речевого сигнала, характеристики речевых сигналов, методы и модели распознавания образов и распознавания речи.

Раздел 2 посвящен использованию алгоритмов и моделей машинного обучения в задачах распознавания речи и нейронных сетях в машинном обучении. Был создан акустический корпус для распознавания речи. Были проанализированы алгоритмы классификации для распознавания говорящего.

В разделе 3 представлены разработанные алгоритмы и модели для идентификации и обработки знаков с использованием машинного обучения в задачах распознавания речи. Выявлена гендерная специфичность с использованием MFCC при предварительной обработке речевого сигнала. Сравнение архитектуры нейронных сетей MLP и CNN в отношении гендерной специфики и распознавания звука выявило более высокую производительность CNN.

В заключении изложены основные результаты и выводы диссертационного исследования.

Уровень достоверности и результаты апробации. Достоверность результатов исследования комплексной оценкой критериев и методов исследования, большим количеством проведенных экспериментов и успешным внедрением их в практику. Результаты диссертации были доложены и обсуждены на следующих научно-методических конференциях:

1. «Төртінші өнеркәсіптік революция жағдайындағы дамудың жаңа мүмкіндіктері» атты ҚР Президенті Н. Назарбаевтың Жолдауын іске асыру шеңберінде «Көліктегі инновациялық технологиялар: білім, ғылым, тәжірибе» атты XLII Халықаралық ғылыми-практикалық конференциясында (Алматы, 18 сәуір, 2018);

2. Профессор Р.Г. Бияшевтың 80 жылдығына және профессор М.Б. Айдархановтың 70 жылдығына арналған «Информатика және қолданбалы математика» атты III Халықаралық ғылыми-практикалық конференциясында (Алматы, 26-29 қыркүйек 2018);

3. Мамырбаев О.Ж., Н.О. Мекебаев., М. Тұрдалыұлы. Идентификация диктора используя mfcc // IV международной научно-практической конференции "Информатика и прикладная математика", посвященной 70-летнему юбилею профессоров Биярова Т.Н., Вальдемара Вуйчика и 60-летию профессора Амиргалиева Е.Н. – Алматы, 2019. – Т. 2. – С. 384-392.

4. 3rd International Conference Applied Mathematics, Computational Science and Systems Engineering (Рим, Италия, 2018);

5. 11th Asian Conference on Intelligent Information and Database Systems (Yogyakarta, Индонезия, 6-12 сәуір 2019 ж.).

По теме диссертации опубликованы 20 статей и получено 2 авторское свидетельство:

1. О.Ж. Мамырбаев, Н.О. Мекебаев, М. Тұрдалыұлы. Сөйлеулерді тану үрдісінде MFCC алгоритмін қолдану // ҚазҰТЗУ хабаршысы. – 2018. -№ 2 (126). – Б. 389-392.

2. О.Ж. Мамырбаев, Н.О. Мекебаев, М. Тұрдалыұлы. Генетикалық алгоритм көмегімен сөйлеуді автоматты танудағы гендерлік сәйкестендіру // Алматы энергетика және байланыс университетінің хабаршысы. – 2018. – специальный выпуск. – Б. 120-129.

3. О.Ж. Мамырбаев, М. Тұрдалыұлы, Н.О. Мекебаев. Система распознавания слитной казахской речи на основе глубоких нейронных сетей // Вестник Алматинского университета энергетики и связи. – 2018. – специальный выпуск. – С. 130-135.

4. О.Ж. Мамырбаев, М. Тұрдалыұлы, Н.О. Мекебаев. Кіріккен қазақ сөйлеін тану жүйесі // ҚБТУ хабаршысы. – 2018. - № 3 (46). – Б. 129-133.

5. О.Ж. Мамырбаев, М. Тұрдалыұлы, Н.О. Мекебаев, Ахметов И. MFCC негізіндегі дикторды анықтау жүйесі // ҚазҰТЗУ хабаршысы. – 2019. – № 2 (132). – Б. 155-160.

6. Мамырбаев О.Ж., М.Тұрдалыұлы, Т.Тұрдалықызы, Шаяхметова А.С. Автоматическое распознавание казахской речи с использованием DNN // ҚБТУ хабаршысы. – 2019. – №2 (49). – Б. 134-142.

7. Мамырбаев О.Ж. Кыдырбаева А.С., Ахмедиярова А.Т., М.Тұрдалыұлы. Н.О.Мекебаев. Систематический обзор и анализ особенностей идентификации по голосу // ҚБТУ хабаршысы. – 2019. – №2 (49). – Б. 120-133.

8. Қалимолдаев М.Н., Мамырбаев О.Ж., Н.О. Мекебаев., Тұрдалыұлы М. Машиналық оқытуды қолдануда дауыстың гендерлік жіктелінуі // ҚазҰТЗУ хабаршысы – 2019. – № 6 (136). – Б.229-233.

9. Bagher BabaAli, Waldemar Wojcik, Orken Mamyrbayev, Mussa Turdalyuly, Nurbapa Mekebayev. Speech Recognizer-Based Non-Uniform Spectral Compression for Robust MFCC Feature Extraction // Przegląd Elektrotechniczny. ISSN: 0033-2097 – 2018. - № 6 (94). – P. 90-93. (Scopus) (Clarivate Analytics)

10. Orken Mamyrbayev, Mussa Turdalyuly, Nurbapa Mekebayev. Automatic Recognition of Kazakh Speech Using Deep Neural Networks // Lecture Notes in Computer Science. 11432 LNAI, с. 465-474. (Scopus)

11. Mamyrbayev O, Toleu A, Tolegen G, Mekebayev N. Neural Architectures for Gender Detection and Speaker Identification // Journal Cogent Engineering. ISSN: 2331-1916. – 2020. Volume 7, - Issue 1. (Scopus)

12. Orken Mamyrbayev, Turdalyuly, Nurzhamal Oshanova, Tolga Ihsan Medeni, Aigerim Yessentay. Voice Identification Using Classification Algorithms // We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists. June 25, 2019. London.

13. Automatic Recognition of Kazakh Speech Using Deep Neural Networks. 11th Asian Conference on Intelligent Information and Database Systems, ACIIDS 2019 (Scopus).

14. О.Ж. Мамырбаев, Н.О. Мекебаев, М. Тұрдалыұлы. Қазақ сөйлеуін тануда іргелі және қолданбалы зерттеуге арналған фонетикалық мәтін // «Төртінші өнеркәсіптік революция жағдайындағы дамудың жаңа мүмкіндіктері» атты ҚР Президенті Н. Назарбаевтың Жолдауын іске асыру шеңберінде «Көліктегі инновациялық технологиялар: білім, ғылым, тәжірибе» атты ХЛІІ Халықаралық ғылыми-практикалық конференцияның материалдары. – Алматы, 2018. – Т. 2. – Б. 81-87.

15. О.Ж. Мамырбаев, М. Тұрдалыұлы, Н.О. Мекебаев. Қазақ тілі сөйлеуінің акустикалық және тілдік модельдерін құру // Материалы XIV Международной Азиатской школы-семинара «Проблемы оптимизации сложных систем». – Алматы, 2018. – Т. 2. – Б. 344-347.

16. О.Ж. Мамырбаев, Н.О. Мекебаев, М. Турдалыұлы. Алгоритмы и архитектуры систем распознавания речи // Материалы III Международной научной конференции «Информатика и прикладная математика» посвященная 80-летию профессора Бияшева Р.Г. и 70-летию профессора Айдарханова М.Б. – Алматы, 2018. – Т. 2. – С. 108-121.

17. Мамырбаев О.Ж., Н.О. Мекебаев., М. Тұрдалыұлы. Идентификация диктора используя mfcc // IV международной научно-практической конференции "Информатика и прикладная математика", посвященной 70-летию профессоров Биярова Т.Н., Вальдемара Вуйцика и 60-летию профессора Амиргалиева Е.Н. – Алматы, 2019. – Т. 2. – С. 384-392.

18. Orken Mamyrbayev, Mussa Turdalyuly, Nurbapa Mekebayev, Kuralay Mukhsina, Alimukhan Keylan, Bagher BabaAli, Gulnaz Nabieva, Aigerim Duisenbayeva and Bekturgan Akhmetov. Continuous Speech Recognition of Kazakh Language // AMCSE 2018 - International Conference on Applied Mathematics, Computational Science and Systems Engineering. - Rome, Italy, 2018, v24 - 2019

19. О.Ж. Мамырбаев, А.С. Кыдырбекова, Н.О. Мекебаев, М. Тұрдалыұлы. Обзор методов идентификации и аутентификации пользователей по голосу // Материалы научной конференции ИИВТ КН МОН РК «Инновационные IT и Smart-технологии», посвященной 70-летию профессора Утепбергенова И.Т., Алматы, 2019. – Б. 315-321.

20. О.Ж. Мамырбаев, М. Турдалыұлы, Н.О. Мекебаев, К. Алимхан, Г.С. Набиева, Б.Ж. Мамырбаев. Фонетически представительный текст для создания систем автоматического распознавания казахской речи // Наука и Мир. – 2018. – № 6 (58). – Т. 2 – С. 49-52.

Получено 2 авторских свидетельства:

1. Компьютерное программное обеспечение `` Система автоматического создания словарного запаса для АСР " Авторское свидетельство от 22 января 2019 № 1425.

2. Программное обеспечение для компьютеров «MultiSpeech Multi-Spacing Multiprocessing», авторское свидетельство от 9 января 2020 г. № 7844.