АННОТАЦИЯ

диссертационной работы Тұрдалыұлы Мұсы
на тему: «Сквозное распознавание слитной речи с использованием глубоких рекуррентных нейросетевых моделей», представленной на соискание степени доктора философии (PhD) по специальности 6D070400 – Вычислительная техника и программное обеспечение

Актуальность темы исследования. Развитие техники и науки тесно связано с эволюцией взаимодействия человека и машины. В настоящее время речевые интерфейсы набирают популярность в человеко-машинных взаимодействиях. Это основывается на то что речь является самым естественным средством взаимоотношения человека. Автоматическое распознавание речи является важным составляющим в речи человеко-машинном интерфейсе.

Одной из самых сложных проблем в сфере автоматического распознавания речи является распознавание речи в произношении. Происходящий с непосредственным участием говорящих, но осуществляемое заранее не подготовленным типом и содержанием устного сообщения– описывает речевую стиль. Сложность этих задач вытекает из следующих особенностей разговорной речи: значительная вариативность между говорящими и дикторами, наличие акцента и характеров настроения в речевом языке, изобилие разнообразных количеств используемых форм слов. Наличие хезитации увеличивает сложность задачи. Хезитация это колебания, которые появляются в связи с осуществлением при произношении в речевом языке, к ней относятся: паузы, не лексические посторонние входные звуки, «паразитные слова», нарушение предложений, обмен словами, повторения, заекания, незаконченные неопределенные пред ложения. В ситуациях отношении разговорной речи сперва надо учесть возникновение разговорной речи языка в произношении. Поэтому задача актуальности его распознавания возрастает.

Система распознавания слитной речи– один из самых востребованных и актуальных проблем. Например, информационные средства коммуникации или анализ больших архивов речи на разных встречах. Но существует ряд особенностей ухудшающее качество работы системы распознавания речи, собственную среду речи говорящего. Они: ограниченная полоса частот в диапазоне 0-4000 Гц, наличие аддитивных и неспецифических отклонении канала, также, потеря информации в процессе кодирования речевого сигнала. Эти особенности дополнительно затрудняют задачу распознавания слитной речи.

В исследованиях распознавания английской речи используют корпус на английском языке Switchboard-1 (300 часовой), корпус Фишер (2000 часовой) и т.д.. Большинство исследователей английского языка уделили большое внимание отбору результатов испытаний HLIB5 Evai 2000 изготовленную в Лингвистическом Консорциуме Данных (Linguistic Data Consortium, LDC)

.Эти базы использовали исследователи из компании IBM Brian Kingsbury, George E. Dahl, из Microsoft Li Deng, Dong Yu, Frank Seide, из компании Google Andrew Senior, Tara Sainath и т.д. В сегодняшнее время каждая страна привлекает своих ученых процессу распознавания своего языка. Сегодня, передовые системы распознавания речи для английского языка – дают возможность снизить уровень распознавания до 15%.

Распознаванию при произношении русской разговорной речи также распознаванию единых образцов посвящены работы исследователей Санкт-Петербургского института информатики и автоматизации Российской академии наук Андрея Ронжина, Алексея Карпова и других ученых из России.

Сегодня в нашей стране распознаванием речи казахского языка занимаются учёные из Евразийского Национального университета им. Л.Н. Гумилева А.Ә. Шарипбай, Г.Т. Бекманова, учёные Казахского национального университета имени аль-Фараби У.А. Тукеев, Д.Р. Рахимова, также из Института информационных и вычислительных технологий Е.Н. Амиргалиев, Р.Р. Мусабаев.

А.Ә. Шарипбай вместе с учениками предложили математическую теорию, морфологический и синтаксический порядок, синтез и анализ слов, алгоритм распознавания речи и приложение, основанную на закономерности формальной фонетики казахского языка.

Исследования по изучению казахского языка показали, что слитная речь еще не может быть достигнута. Данная ситуация свидетельствует о том, что система распознавания слитной речи на казахском языке, соответствующая уровню распознавания речи на казахском языке, еще не создана. Существует ряд причин, по которым недостаточно эффективно распознавать разговорную слитную речь в казахской речи. Во-первых, недоступность необходимого корпуса для оценки качества системы распознавания разговорной слитной речи в казахском языке; во-вторых, казахский язык, как агглютинативный язык, обладает гораздо большим количеством словоформ, чем аналитические языки. Если на казахском языке говорят в нескольких десятках тысяч слов, то казахскому языку понадобится словарь, содержащий сотни тысяч слов. В-третьих, разговорная речь на казахском языке можно назвать фонетическими особенностями, такими как артикуляционный слабый вид, ассимиляционное явление (адаптация звуков), редуцирование (сокращение длительности звука). Для казахского языка необходимо создать систему распознавания, которая сильно влияет на акустическую вариацию отличительных признаков речи, возникающих из этих фонетических особенностей.

Исходя из вышеизложенного, можно сделать вывод, что необходимость в алгоритмах, методах и программном обеспечении для обеспечения точности распознавания слитной речи в казахской речи является весьма актуальной.

Цель диссертационной работы. Разработка методов, алгоритмов и программных средств для обеспечения точности распознавания разговорной слитной речи на казахском языке и быстрого ее внедрения в системе, которая достаточно используется для практических задач.

Задачи исследования. Для достижения целей исследования решаются следующие вопросы:

1. Анализ современных методов распознавания речи при произношении.
2. Разработка акустического корпуса и текстового корпуса слитной речи казахского языка.
3. Разработка языковой модели, словаря транскрипции и акустической модели, которая является частью системы распознавания слитной речи при произношении на казахском языке.
4. Оценка качества работы разработанной системы распознавания слитной речи при произношении на казахском языке, также сравнение с зарубежными системами.
5. Подготовка методов, алгоритмов и программных средств распознавания слитной речи при произношении на казахском языке.

Объект исследования. Системы автоматического распознавания речи

Предмет исследования. Методы, алгоритмы и программные средства распознавания слитной речи при произношении на казахском языке

Методы исследования. Широко используется в прикладных научных исследованиях: подготовка целей и задач, анализ состояния исследования и существующей литературы, разработка алгоритмических и программных решений, оценка эффективности разработанных решений, апробация и анализ результатов. В экспериментальной части исследования он был осуществлен только с помощью эстественного языкового материала, здесь тестовая сортировка сходится с заданными с помощью особенностей произношения или особенностями состава дикторов.

В качестве методов исследования использовались численные методы обработки сигналов, вероятностные и математические статистические теории, машинное обучение, прикладная лингвистика, также методы разработки программного обеспечения.

Научные новизны:

– Разработан акустический корпус и текстовой корпус слитной речи казахского языка.

– Разработан алгоритм сжатия сигналов речи для увеличения качества акустических моделей при распознавании произнесенной речи.

– С использованием разработанных акустической и текстовой корпусов впервые разработана языковая модель слитной речи казахского языка на базе глубоких рекуррентных нейросетевых моделей.

– Разработаны программные средства входящие в систему распознавания слитной речи при произношении на казахском языке, которые позволяют использовать акустические и языковые модели, созданные с помощью представленных алгоритмов в диссертации.

Теоретическое и практическое значение работы. Теоретическая значимость данной работы заключается в совершенствовании существующих и новых алгоритмов акустических моделей на основе глубоких нейронных сетей для задач распознавания речи, а также описывается экспериментальным

исследованием и разработкой нового метода получения информационных символов, которые более улучшены, чем те, которые использовались ранее.

Практическое значение диссертационного исследования заключается в следующих результатах: применение алгоритмических и программных средств, разработанных при создании системы распознавания слитной речи при произношении на казахском языке; демонстрировать использование качественного распознавания и быстрого реагирования в практических задачах такие как, автоматическое преобразование в текст архивов фонограмм, поиск ключевых слов в потоке непрерывного разговорного языка, тематическая кластеризация записей.

Основное положение выносимое на защиту. Применение глубоких нейросетевых моделей в задачах сквозного распознавания слитной речи казахского языка увеличивает качества расознавания систем автоматического распознавания речи.

Личный вклад исследователя. Исследователь лично решал задачи диссертационной работы. Разработаны методы и алгоритмы распознавания речи при произношении. Проведена экспериментальная оценка разработанных методов и алгоритмов. Разработаны программные средства входящие в состав системы распознавания слитной речи при произношении на казахском языке

Связь темы диссертации с планами научно-исследовательской работы. Исследовательские работы по диссертации были выполнены в рамках проекта грантового финансирования «Разработка технологии мультиязычного автоматического распознавания речи с использованием глубоких нейронных сетей». (2018-2020, государственный регистрационный номер: 0118RK00139) в Институте информационных и вычислительных технологий КН МОН РК.

Объем и структура работы. Диссертация состоит из введения, 4 глав и заключения. Общий объем диссертации составляет 95 страниц, 41 рисунков, 6 таблиц. Список литературы состоит из 111 наименований.

Во **введении** была определена актуальность работы и показаны проблемы, связанные с темой. Показаны идея работы, цель и задачи исследования, научная новизна и практическая ценность исследования, методы исследования.

Первая глава описывает классификацию систем автоматического распознавания речи, их проблемы создания и другие использованные модели.

Вторая глава описывает работу по созданию речевого и текстового корпуса на казахском языке. Сначала были определены этапы создания речевого и текстового корпуса. Описан процесс сбор а текстовой информации на казахском языке. Были отобраны дикторы озвучивающие речь для работ по созданию акустического корпуса и были проведены их идентификационные работы. Были определены этапы создания речевого и текстового корпуса. Были определены инструменты, используемые для сбора речевой информации. Проведены работы по созданию словаря состоящий из слов созданных корпусов, и в результате создан словарь для разработки систем автоматического распознавания слитной речи на казахском языке.

В третьей главе при помощи речевого корпуса казахского языка проведены работы по построению акустической и языковой моделей этого языка. Разработан и основан алгоритм сжатия речевых сигналов. Были выбраны инструменты для создания акустического и языкового моделей, с помощью этого инструмента была проведена адаптация речевого корпуса казахского языка к устройству для обучения системы. Процесс обучения системы описан использованием Скрытой Марковской модели, Гауссовых смесей и глубоких нейронных сетей. После процесса обучения системы, для определения качества акустической и языковой моделей речи казахского языка, были проведены практические работы с монофонными, трифонными, глубокими нейронными сетями и глубокими рекуррентными сетями, и были получены результаты качества речи.

В четвертой главе с помощью полученных, по проведенным практическим работам в предыдущих главах, глубоким рекуррентным нейросетевым моделям, был разработан программный комплекс который сквозно распознает речь казахского языка с помощью акустической и языковой моделей. Программный комплекс состоит из программного приложения, работающего в операционных системах Windows и Linux, Веб-приложения, работающего на основе архитектуры клиент-сервер, и мобильного приложения, работающего в системах Android и iOS. Раскрыты функции работающие в приложении состава программного комплекса.

В заключение приводятся основные результаты и выводы диссертации и указывается их связь с будущей работой.

Уровень достоверности и результаты апробации. Обоснованность и достоверность исследования соответствуют обоснованным обязанностям задачи, анализу критериев и состоянию исследований в данной области, большому количеству проведенных экспериментов и успешной реализации их на практике. Результаты диссертации обсуждались и докладывались на следующих научно-методических конференциях:

1. II Международная научно-практическая конференция «Информационные и телекоммуникационные технологии: образование, наука, практика» (Алматы, 3-4 декабря, 2015).
2. XLII Международная научно-практическая конференция «Инновационные технологии на транспорте: образование, наука, практика» в рамках реализации Послания Президента Республики Казахстан Н.А. Назарбаева «Новые возможности развития в условиях четвертой промышленной революции». (Алматы, 18 апреля, 2018).
3. XIV международная азиатская школа-семинар "Проблемы оптимизации сложных систем" (Чолпон-Ата, Кыргызстан, 2018).
4. III Международная научно-практическая конференция «Информатика и прикладная математика», посвященной 80-летнему юбилею профессора Бияшева Р.Г. и 70-летию профессора Айдарханова М.Б. (Алматы, 26-29 сентября 2018).

5. 3rd International Conference Applied Mathematics, Computational Science and Systems Engineering (Рим, Италия, 2018).

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

По теме диссертации опубликованы 16 статей и получено 1 авторское свидетельство:

1. Yasser Mohseni Behbahani, Bagher BabaAli, and Mussa Turdalyuly. Persian sentences to phoneme sequences conversion based on recurrent neural networks // Open Computer Science. – 2016. - № 6. – P. 219-225. (Scopus)
2. Bagher BabaAli, Waldemar Wojcik, Oken Mamyrbayev, Mussa Turdalyuly, Nurbapa Mekebayev. Speech Recognizer-Based Non-Uniform Spectral Compression for Robust MFCC Feature Extraction // Przeglad Elektrotechniczny. ISSN: 0033-2097 – 2018. - № 6 (94). – P. 90-93. (Clarivate Analytics)
3. К.Н. Калимолдаев, О.Ж. Мамырбаев, М. Турдалыулы, К.Е. Нурлан, А.Е. Ибраимкулов. Нейрондық желілерді қолдану арқылы сөйлеуді автоматты түрде тану // ҚазҰТЗУ хабаршысы. – 2016. -№ 5 (117). – С. 435-438.
4. О.Ж. Мамырбаев, Н.О. Мекебаев, М. Тұрдалыұлы. Сөйлеулерді тану үрдісінде MFCC алгоритмін қолдану // ҚазҰТЗУ хабаршысы. – 2018. -№ 2 (126). – С. 389-392.
5. 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. – С. 120-129.
6. 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.
7. О.Ж. Мамырбаев, М. Тұрдалыұлы, Н.О. Мекебаев. Кіріккен қазақ сөйлеуін тану жүйесі // ҚБТУ хабаршысы. – 2018. - №3 (46). – Б. 129-133.
8. О.Ж. Мамырбаев, М. Тұрдалыұлы, Н.О. Мекебаев, И. Ахметов. MFCC негізіндегі дикторды анықтау жүйесі // ҚазҰТЗУ хабаршысы, № 2 (132), 2019
9. О.Ж. Мамырбаев, М. Турдалыулы, 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.
10. О.Ж. Мамырбаев, М. Калимолдаев, М. Турдалыулы, В. BabaAli. Методы построения многомодального распознавания речи // Труды II Международной научно-практической конференции «Информационные и телекоммуникационные технологии: образование, наука, практика». – Алматы, 2015. – Т. 1. – С. 217-221.
11. 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 XLII Халықаралық ғылыми-практикалық конференцияның материалдары. – Алматы, 2018. – Т. 2. – С. 81-87.

12. 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. – С. 344-347.

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

14. 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

15. Orken Mamyrbayev, Mussa Turdalyuly, Nurbapa Mekebayev, Alimukhan Keylan, Aizat Kydyrbekova and Turdalykyzy Tolganai. Automatic Recognition of Kazakh Speech Using Deep Neural Networks // 11th Asian Conference on Intelligent Information and Database Systems. - Yogyakarta, Indonesia, 2019

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

17. Авторское свидетельство «System of automatic creation vocabulary for ASR» от 22.01.2019 г. №1425