АННОТАЦИЯ

диссертационной работы докторанта **Оралбековой Дины Орымбаевны** на тему: **«Разработка системы автоматического распознавания речи на основе интегрального подхода»**, представленной на соискание степени доктора философии (PhD) по специальности 8D06103 – Management information systems

Актуальность темы исследования. Прогресс информационных систем и вычислительной техники привел к улучшению процессов во многих технологиях машинного обучения, в частности в распознавании речи. Системы автоматического распознавания речи (САРР) уже стали неотъемлемой частью в нашей повседневной жизни и играют огромную роль в развитии других технологий машинного обучения, как синтез речи, машинный перевод и т.д. САРР нашли широкое применение в различных сферах деятельности, как голосовое управление автомобилем, домом и бытовой техникой, а также голосовой ввод в различных приложениях, навигационных системах и др. И это лишь некоторые примеры использования САРР. Существует традиционная система распознавания речи, которая обычно состоит из трех основных независимых элементов, и представляют из себя следующие модели, как акустические модели для прогнозирования контекстно-зависимых состояний субфонем из аудио, языковые модели и лексикон для сопоставления фонем к словам. Модели традиционных систем распознавания речи обучаются независимо друг от друга, так классическая акустическая модель может быть обучена на основе смесей гауссовых распределений и скрытых марковских моделей, а языковые модели на основе n-gram. Долгое время в задаче распознавания речи широко применялась, и была основной технологией, модель на основе скрытых марковских моделей (HMM). HMM в основном используется для динамической деформации времени на уровне кадра и смеси гауссовских распределений плотностей вероятностей (GMM) применяется для представления распределений сигналов в промежутке фиксированного небольшого периода времени, который обычно соответствует единице произношения. Продолжительное время модель HMM-GMM являлась общей структурой для распознавания речи. В последнее время глубокое обучение приносит значительные улучшения во многих исследованиях, и в развитие распознавания речи. Активное использование искусственных нейронных сетей на каждом элементе сценария классической системы распознавания речи увеличивает эффективность ее работы, что отразилась во многих исследовательских работах. С развитием технологий глубокого обучения глубоких нейронных сетей (DNN) начали применять в распознавании речи для акустического моделирования. Роль DNN заключается в расчете апостериорной вероятности состояния HMM, которое может быть преобразовано в вероятности, заменяя обычную вероятность наблюдения GMM. Таким образом, модель HMM-GMM превращается в HMM-DNN, которая достигает лучших результатов, и становится популярной

моделью автоматического распознавания речи. В научных работах было показано, что для получения эффективной акустической модели были применены глубокие нейронные сети, а в других исследовательских работах с помощью рекуррентных нейронных сетей и сетей с долгой и кратковременной памятью (LSTM) были построены языковые модели и словарь соответственно. Также сверточные нейронные сети (CNN) были применены для извлечения признаков из сигнала речи. Таким образом, многие опубликованные результаты показывают, что предложенный подход демонстрирует лучшую производительность среди всех современных систем распознавания речи, который является основой для применения различных архитектур искусственных нейронных сетей на всех модулях систем распознавания речи.

С недавних пор распространение получил интегральный метод распознавания речи с использованием методов машинного обучения. В таких системах модель реализована с использованием только одной нейронной сети. Интегральная реализация модели часто представляет лучшую производительность с позиции скорости и точности распознавания речи. Зарубежные исследовательские работы доказывают, что прогресс полученных результатов интегральных систем зависит от увеличения объема тренировочных данных для обучения сети. В настоящее время популярные приложения, как Voice to Text Messenger, Google Listen, Attend, Spell, Baidu Deep Speech и другие работают на основе интегрального подхода. Основной принцип работы заключается в том, что современные интегральные модели обучаются на основе больших данных. Из вышеуказанного можно обнаружить основную проблему, это касается распознавания языков с ограниченными обучающими данными, такие как казахский, киргизский, турецкий и т.д. Для таких малоресурсных языков не существуют большие корпусы обучающих данных. Также необходимо отметить, что не разработаны и не исследованы модели и методы интегральной архитектуры для малоресурсных языков. В настоящее время не существуют эффективных алгоритмов и программных средств интегрального распознавания для казахского языка.

Зарубежные ученые, как Dario Amodei, Chao Weng, William Chan (США), Xinhui Hu, Chiori Hori (Япония), Jinchuan Tian, Eric Chang, Jianlai Zhou, Jianwei Sun (Китай), Jan Chorowski (Польша) и другие исследователи добились конкурентноспособных результатов в совершенстовании интегральных систем для распознавания популярных языков, как английский и китайский, и ученые из ближнего СНГ, а именно из Санкт-Петербургского института информатики и автоматизации Российской академии наук Карпов А.А., Кипяткова И.С. и их коллеги исследуют область распознавания речи уже многие годы и достигли высоких показателей в разработке и улучшении интегральных систем распознавания русской речи.

Стоит отметить, что существуют разработки отечественных ученых по системам распознавания казахской речи на основе глубоких нейронных сетей. В работах ученых Евразийского национального университета им. Л.Н.Гумилева – Шарипбай А.А., Есенбаева Ж.А., Казахского национального

университета им. Аль-Фараби – Тукеева У.А., Рахимовой Д.Р., Назарбаев университета – Хасанова Е., а также в научных работах исследователей из Института информационных и вычислительных технологий – Амиргалиева Е.Н., Мусабаева Р.Р. и др. были отражены разработки системы автоматического распознавания казахской речи на основе традиционных, гибридных моделей HMM-DNN и моделей CTC с Transformer. Результаты этих работ достигли хорошего уровня распознавания речи, но до сих пор данные системы требуют улучшений в сокращении ошибок распознавания слов до человеческого уровня и в увеличении объема данных для тренировки сети. Кроме того, не разработана интегральная система для распознавания казахской речи на основе модели кодер-декодер (Encoder-Decoder) с использованием механизма внимания, которая показывает перспективные результаты. Модель с механизмом внимания может хорошо работать как с языковыми моделями, так и без них, при этом демонстрируя низкую погрешность распознавания речи. Этот подход является перспективным направлением, который может быть использован для разработки систем распознавания речи при ограниченной обучающей выборки.

Исходя из вышеизложенного, можно прийти к выводу, что в настоящий момент необходимость в эффективных методах, алгоритмах и программном обеспечении для улучшения точности распознавания казахской речи с использованием интегральных моделей особенно *актуальна*.

Цель диссертационной работы. Исследование и разработка модели, архитектуры и алгоритма для повышения точности распознавания слитной казахской речи на основе интегрального подхода.

Задачи исследования. Для реализации поставленных целей исследования решаются следующие вопросы:

1) Анализ методов и моделей распознавания речи на основе интегрального подхода.

2) Разработка речевого корпуса для интегрального распознавания казахской речи.

3) Разработка эффективной интегральной модели и алгоритма на основе кодер-декодер (Encoder-Decoder) с использованием механизма внимания для создания системы распознавания казахской речи.

4) Разработка интегральной архитектуры и программного обеспечения для распознавания казахской речи с помощью полученных в ходе исследований моделей и методов на основе интегрального подхода.

Объект исследования. Современные технологии и системы автоматического распознавания речи.

Предмет исследования. Технологии, алгоритмы, модели и методы, программное обеспечение для распознавания казахской речи на основе интегрального подхода.

Методы исследования. Методы машинного обучения, технологии и методы автоматического распознавания речи, теории вероятностей и математической статистики, методы разработки программного обеспечения, а также прикладная и компьютерная лингвистика.

Научная новизна

1) Разработаны речевой и текстовый корпусы для казахского языка.

2) Разработана интегральная модель с применением механизма внимания для распознавания казахской речи.

3) Разработан эффективный алгоритм для распознавания казахской речи на основе интегрального модуля.

4) Разработано программное обеспечение, которое автоматически преобразует речь в текст.

Теоретическое и практическое значение работы. Теоретическая значимость исследовательской работы заключается в разработке и реализации эффективных алгоритмов и интегральных моделей для распознавания казахской речи, а также в разработке обучающего корпуса речи для казахского языка.

Практическая значимость исследовательской работы заключается в применении разработанных алгоритмов и программного обеспечения для дальнейшего использования в развитии других технологий, как синтез речи, машинный перевод, голосовая аутентификация и идентификация и т.д. Разработанная система автоматического распознавания казахской речи может быть внедрено в государственных структурах, ответственных за расширение области применения национальных языков на базе информационных технологий; в мобильных устройствах (увеличение числа потенциальных покупателей за счёт внедрения речевых технологий на государственном языке); в банках (call-центры с поддержкой голосовых функций, голосовая аутентификация); и в сектор производства различных устройств с поддержкой голосовых функций.

Основное положение, выносимое на защиту

1) Для обучения модели был разработан корпус в объеме 2000 часов речи с транскрипциями. При создании корпуса учтены различные виды речи: подготовленная (чтение), спонтанная.

2) Разработана система на основе кодер-декодер (Encoder-Decoder) с вниманием. Модель на основе механизма внимания была модифицирована и расширена для распознавания схожих окончаний в словах. Была построена архитектура данной модели с помощью нейронных сетей, как LSTM и BLSTM. Результаты экспериментов показали, что построенная модель хорошо выполняет работу без применения языковых моделей для казахского языка и превзошла не только гибридные модели на основе DNN-HMM, но и другие интегральные модели и показала лучшие результаты по точности распознавания слов и символов.

Степень достоверности и апробация результатов. Исследования и результаты, относящиеся к теме диссертации, были представлены и обсуждены на различных конференциях и семинарах на основе следующих публикаций:

1) Мамырбаев О.Ж., **Оралбекова Д.О.** Онлайновые модели на основе внимания для интегральных (end-to-end) систем распознавания речи. V

Международная научно-практическая конференция "Информатика и прикладная математика (29 сентября - 1 октября **2020 г., Алматы**).

2) O. Mamyrbayev, **D. Oralbekova**, A. Kydyrbekova, T. Turdalykyzy and A. Bekarystankyzy, "End-to-End Model Based on RNN-T for Kazakh Speech Recognition," 3rd International Conference on Computer Communication and the Internet (ICCCI) (25-27 июня **2021 г., Токио, Япония**).

3) Мамырбаев О.Ж., **Оралбекова Д.О.**, Othman M., Тулендиев Д.М., Жумажанов Б., Турдалыкызы Т. Распознавание казахской речи на основе интегральной модели RNN-T. VI Международная научно-практическая конференция "Информатика и прикладная математика (29 сентября - 1 октября **2021 г., Алматы**).

4) Мамырбаев О.Ж., **Оралбекова Д.О.**, Othman M., Тулендиев Д.М., Жумажанов Б., Турдалыкызы Т. Исследование интегральной модели на основе внимания для автоматического распознавания казахской речи. Международная научная конференция в области информационных технологий, посвященной 75-летию профессора У.А. Тукеева. (8 октября **2021 г., Алматы**).

5) Мамырбаев О.Ж., **Оралбекова Д.О.**, Кыдырбекова А.С., Жумажанов Б.Ж. , Тұрдалықызы Т. Интегральная гибридная модель на основе CTC и механизма внимания для распознавания казахской слитной речи. Международная научная конференция «Сатпаевские чтения 2021» (12 апреля **2021 г., Алматы**)

6) **D. Oralbekova**, O. Mamyrbayev, M. Othman, B. Zhumazhanov, K. Mukhsina. Development of the insertion-based speech recognition method. 7th International conference on Computer Science and Engineering (14-16 сентября **2022г., Стамбул, Турция**)

Личный вклад исследователя. Докторант самостоятельно выполнил и решил задачи диссертационной работы. Разработал и реализовал модель и алгоритм для распознавания казахской речи на основе интегральной архитектуры. Разработал и расширил речевой и текстовый корпусы для казахского языка. Выполнил экспериментальную оценку разработанных моделей и алгоритмов.

Связь темы диссертации с планами научно-исследовательской работы. Проведенные научно-исследовательские работы по диссертации были выполнены в рамках двух проектов грантового финансирования: 1) «Разработка технологии мультиязычного автоматического распознавания речи с использованием глубоких нейронных сетей» (2018-2020, государственный регистрационный номер: 0118РК00139) 2) «Разработка интегральной (end-to-end) системы автоматического распознавания речи для агглютинативных языков» (2020-2022, государственный регистрационный номер: 0120РК00344) в Институте информационных и вычислительных технологий КН МОН РК.

Публикация основных результатов диссертационного исследования.

По теме диссертационной работы было получено 3 авторских свидетельства, 1 патент на изобретение и опубликовано 7 работ, из которых 3

статьи опубликованы в журналах, рекомендованных Комитетом по контролю в сфере образования и науки МОН РК, 4 статьи опубликованы в изданиях имеющие ненулевой импакт-фактор, индексирумых базой Scopus и Web of Science:

1) Mamyrbayev, O., **Oralbekova, D.,** Keylan, A. et al. A study of transformer-based end-to-end speech recognition system for Kazakh language. Sci Rep 12, 8337 (2022). https://doi.org/10.1038/s41598-022-12260-y (**Web of Science, квартиль Q1, IF=4,3**)

2) Mamyrbayev, O.Z., Oralbekova, D.O., Alimhan, K. et al. Hybrid end-to-end model for Kazakh speech recognition. International Journal of Speech Technology (2022). https://doi.org/10.1007/s10772-022-09983-8 (**Scopus, IF=1.803, процентиль 93**).

3) Mamyrbayev, O., Kydyrbekova, A., Alimhan, K., **Oralbekova, D.,** Zhumazhanov, B., Nuranbayeva, B. (2021). Development of security systems using DNN and i & x-vector classifiers. Eastern-European Journal of Enterprise Technologies, 4 (9 (112)), 32–45. doi: https://doi.org/10.15587/1729-4061.2021.239186 (**Scopus, процентиль 43**);

4) Mamyrbayev, O., Alimhan, K., **Oralbekova, D.,** Bekarystankyzy A., Zhumazhanov, B. (2022). Identifying the influence of transfer learning method in developing an end-to-end automatic speech recognition system with a low data level. Eastern-European Journal of Enterprise Technologies, 1(9(115), 84–92. https://doi.org/10.15587/1729-4061.2022.252801(**Scopus, процентиль 43**);

5) Mamyrbayev O., **Oralbekova D.** Modern trends in the development of speech recognition systems // News of the National academy of sciences of the republic of Kazakhstan. – 2020. – Vol. 4, № 332. - P. 42 – 51 // doi.org/10.32014/2020.2518-1726.64

6) Mamyrbayev O., **Oralbekova D.**, Alimhan K., Othman M., Zhumazhanov B. Realization of online systems for automatic speech recognition// News of the National academy of sciences of the republic of Kazakhstan. – 2021. – Vol. 6, № 340. - P. 66 – 72 // doi.org/10.32014/2021.2518-1726.103

7) Мамырбаев О.Ж., **Оралбекова Д.О.**, Алимхан К., Othman M., Жумажанов Б. Применение гибридной интегральной модели для распознавания казахской речи// News of the National academy of sciences of the republic of Kazakhstan. – 2022. – Vol. 1, № 341. - P. 58 – 68 // doi.org/10.32014/2022.2518-1726.117.

8) Авторское свидетельство "Система автоматического распознавания казахской речи на основе интегральной архитектуры" № 15501 от 25.02.2021, Авторы: О.Ж. Мамырбаев, **Д.О. Оралбекова**, А.С. Кыдырбекова, Б.Ж. Жумажанов, Т.Тұрдалықызы.

9) Авторское свидетельство "Система идентификации и аутентификации через речевые технологии" № 23323 от 4 февраля 2022. Авторы: **Оралбекова Д.О.**, Мамырбаев О.Ж., Алимхан К., Кыдырбекова А.С., Жумажанов Б.Ж., Турдалыкызы Т.

10) Авторское свидетельство "Система автоматического распознавания казахской слитной речи на основе модели с механизмом внимания" №24178

от 5.03.2022. Авторы: Мамырбаев О.Ж., **Оралбекова Д.О.**, Әлімхан Қ., Кыдырбекова А.С., Жұмажанов Б.Ж., Тұрдалықызы Т.

11) Патент на изобретение «Система и способ распознавания агглютинативной слитной речи на основе интегрального (end-to-end) подхода». № 35886 от 07.10.2022. Авторы: Мамырбаев О.Ж., **Оралбекова Д.О.**, Кыдырбекова А.С., Жұмажанов Б.Ж., Тұрдалықызы Т.

Структура и объем диссертационной работы. Диссертационная работа состоит из введения, 4 разделов, заключения, списка литературы из 111 наименований и 5 приложений. Работа изложена на 108 страницах и содержит 25 рисунков, 6 таблиц.

Краткое описание диссертационного исследования

Во **введении** приводятся актуальность исследуемой работы, цель и задачи диссертационной работы, научная новизна, теоретическое и практическое значение работы и методы исследования.

В **первой главе** описывается общая модель системы автоматического распознавания речи и приводится обзор интегральных систем распознавания речи, как коннекционная временная классификация и рекуррентный преобразователь, модель кодер-декодер, архитектура Transformer с механизмом внимания и модель на основе условных случайных полей. Приводится обзор сопустствующих работ по рассмотренным моделям, а также описывается сравнительный анализ данных моделей с традиционными моделями.

Во **второй главе** рассматривается работа по сбору корпуса для казахского языка, который состоит из аудиоданных с их транскрипциями (текстовое представление аудио) и полное содержание корпуса. Приводится методика составления текстовок для транскрибирования телефонных разговоров.

Третья глава описывает архитектуру модели кодер-декодер с механизмом внимания. Описывается предварительная настройка интегральной модели для кодера, декодера и механизма внимания, а также приводятся метрики оценки для распознавания речи. Приводятся описание наборов данных, используемых в экспериментах и алгоритм работы кодер-декодер с механизмом внимания. Приведены программное и аппаратное обеспечение для реализации модели с вниманием. Кроме того, приводится описание микрофона SmartMike Duo, специально разработанный компанией Philips для системы распознавания речи, который может разделить наложение двух голосовых данных на два отдельных аудиоканала. Описывается экспериментальная проверка предложенной интегральной модели. Для сравнения полученных результатов были отобраны исследовательские работы, связанные с распознаванием казахской речи.

Четвертая глава посвящена описанию системы автоматического распознавания казахской речи. Рассматривается подробная структура интегральной системы, описываются работы модулей для обучения нейронных сетей и для валидации и вывода данных системы. Приводится интерфейс программы и ее составляющие компоненты. Описывается процесс

интеграции интегральной системы с микрофоном SmartMike Duo USB PSM1010, который предоставляет уникальную возможность разработанной системе использовать два отдельных аудиоканала, что обеспечивает превосходную точность распознавания и расширенные возможности анализа речи.

В **заключительной** части приводятся полученные результаты и выводы диссертационного исследования и указываются планы на дальнейшие работы по выбранному направлению.