Дүйсенбекқызы Жансая на тему « Исследование и разработка метода распознавания казахской речи у детей с нарушениями речи », представленной на соискание степени доктора философии (PhD) по образовательной программе 8D06101 – Информационные системы

АННОТАЦИЯ

Актуальность темы исследования.

На современном этапе научно-технологического развития системы автоматического распознавания речи (Automatic Speech Recognition, ASR) получили широкое распространение в образовательных платформах, инструментах логопедической диагностики и информационных системах раннего развития детей. Вместе с тем задача автоматического распознавания речи детей с нарушениями речевой функции остаётся научно недостаточно изученной и не имеющей полноценного практического решения сложной научно-технической проблемой.

Акустические различия между речью детей и взрослых подтверждены эмпирическими исследованиями. В частности, основная частота (F0) у детей составляет 250–400 Гц, тогда как у взрослых этот показатель находится на уровне 100–150 Гц; форматные частоты также существенно выше — длина голосового тракта детей составляет 9–13 см, тогда как у взрослых она достигает 17 см. Полноценная артикуляция формируется лишь к 12–14 годам. Исследования показывают, что применение ASR-моделей, обученных на речи взрослых, к детской речи приводит к росту коэффициента ошибок по словам (Word Error Rate, WER) до 40–45%.

Казахский язык — агглютинативный язык с ограниченными ресурсами (low-resource). Существующие казахстанские ASR-системы (KATEK, KazASR, KazakhTTS, KazakhBERT, Soyle) ориентированы исключительно на речь взрослых; специализированный аннотированный акустический корпус детской речи на казахском языке до 2025 года отсутствовал. Независимая оценка систем Kaldi, Mozilla DeepSpeech и Google Speech-to-Text показала, что наилучший результат продемонстрировала Google STT с WER = 52,97%, что является неприемлемым для практического применения.

Изложенные научные факты и ограничения существующих систем обусловливают актуальность разработки научно-технического решения для автоматического распознавания и ранней диагностики казахскоязычной речи детей с нарушениями речевой функции. В представляемой диссертационной работе впервые создан аннотированный акустический датасет казахскоязычной детской речи и предложен комплексный метод post-editing на основе интеграции ASR и LLM. Проведённые эксперименты

подтверждают высокую точность, вычислительную эффективность и практическую применимость предложенного метода.

Цель диссертационной работы

Впервые создать аннотированную акустическую базу данных казахскоязычной детской речи и разработать комплексный метод post-editing на основе интеграции ASR и LLM, направленный на повышение точности распознавания речи детей с нарушениями речевой функции.

Задачи исследования:

–провести сравнительно-аналитический обзор мировых акустических корпусов детской речи и систематизировать их структурные, объёмные и акустические характеристики; выполнить сравнительный анализ современных методов и акустических корпусов в области распознавания детской речи;

–собрать и обработать аннотированный акустический датасет казахскоязычной детской речи с учётом возрастных и гендерных характеристик детей;

–исследовать современные нейросетевые ASR-модели (Whisper, MMS, TurkicASR, ESPnet, DeepSpeech и др.) и адаптировать их к детской речи посредством дообучения (fine-tuning);

–разработать комплексный метод постредактирования (post-editing) на основе LLM для коррекции результатов ASR после дообучения;

–экспериментально оценить эффективность предложенного метода и провести сравнительный анализ.

Методы исследования: методы машинного и глубокого обучения, технологии автоматического распознавания речи (ASR), обработка акустических сигналов, теория вероятностей и математическая статистика, методы разработки программного обеспечения.

Объект исследования

Процесс речи детей с нарушениями речевой функции на казахском языке и его акустические особенности.

Предмет исследования

Технологии, алгоритмы, модели и методы автоматического распознавания казахскоязычной речи детей с нарушениями речевой функции, а также соответствующее программное обеспечение.

Научная новизна полученных результатов:

1. Впервые разработан специализированный аннотированный акустический датасет казахскоязычной детской речи с учётом возрастных, гендерных и фонетических особенностей детей. Созданная база данных вносит существенный вклад в устранение дефицита открытых ресурсов в области исследования детской речи.

2. Впервые предложен комплексный метод post-editing, объединяющий адаптацию ASR-модели к детской речи посредством fine-tuning и постредактирование на основе LLM, для распознавания казахскоязычной речи детей с нарушениями речевой функции. Предложенный метод обеспечивает повышение точности распознавания казахскоязычной детской речи.

Положения, выносимые на защиту:

1. Разработанный аннотированный акустический датасет казахскоязычной детской речи, сформированный с учётом возрастных, гендерных и фонетических особенностей. Данный датасет является первым специализированным ресурсом для исследования казахскоязычной детской речи.

2. Предложенный комплексный метод post-editing, интегрирующий адаптацию ASR-модели к детской речи (fine-tuning) и постредактирование на основе LLM. По результатам проведённых экспериментов WER снижен с 0,70 до 0,18, CER — с 0,467 до 0,065.

Объём и структура диссертационной работы

Диссертация состоит из введения, четырёх основных разделов, заключения, списка использованных источников и приложений. Общий объём работы — 123 страницы машинописного текста, включающего 30 рисунков, 43 таблицы, 117 использованных источников и 4 приложения.

В первом разделе выполнен аналитический обзор исследований в области распознавания речи и речевых нарушений: последовательно рассматриваются эволюция ASR-систем, акустико-фонетические особенности детской речи, классификация речевых нарушений и международный опыт в области низкоресурсных языков.

Во втором разделе изложены методология отбора текстового материала, методы сбора и аннотирования акустических данных, этические и правовые аспекты; представлены структура датасета и методы предварительной обработки данных. Сбор аудиоданных осуществлялся исключительно на основании письменного информированного согласия родителей или законных представителей детей в соответствии с Законом РК

«О персональных данных и их защите» (№ 94-V); все записи деперсонализированы посредством присвоения уникальных идентификаторов и хранятся на защищённых серверах на территории Республики Казахстан с разграниченным доступом. Персональные данные участников не публикуются и не передаются третьим лицам

В третьем разделе описаны архитектура предложенного комплексного метода, формализация задачи распознавания речи, математическая постановка акустической модели, адаптированной к детской речи, а также математическая модель комплексного метода.

В четвёртом разделе проведена экспериментальная оценка предложенного метода: выполнен сравнительный анализ ASR-моделей, представлены результаты fine-tuning и постредактирования на основе LLM, а также результаты апробации метода в практической среде.

Апробация работы

Результаты исследования обсуждались на научных семинарах Казахского национального университета имени аль-Фараби, докладывались на международных конференциях, а также прошли опытную апробацию в детском саду №154.

Публикации.

По теме диссертации опубликованы 9 научных работ, в том числе 2 статьи в журналах, индексируемых в базах Web of Science (SCIE, Q2) и (ESCI, Q1), 5 работ — в сборниках международных конференций, 2 работы — в журналах, рекомендованных Комитетом по обеспечению качества в сфере образования и науки МОН РК; получены 2 авторских свидетельства.

- Rakhimova D., Duisenbekkyzy Zh., Adali E. Investigation of ASR Models for Low-Resource Kazakh Child Speech: Corpus Development, Model Adaptation, and Evaluation // Applied Sciences. – 2025. – Vol. 15, Issue 16, Article 8989. DOI: 10.3390/15168989 (WoS SCIE, Q2, IF: 2,7, процентиль 79).
- Rakhimova D., Karibayeva A., Turarbek A., Karyukin V.I., Duisenbekkyzy Zh. and Aliyev R. Development of a Children's Educational Dictionary for a Low-Resource Language Using AI Tools // Computers. – 2024. – Vol. 13, No. 10, Article 253. DOI: 10.3390/computers13100253 (WoS ESCI, Q1, CiteScore: 7,5, процентиль 84).
- Rakhimova D., Duisenbekkyzy Zh. Research on the Impact of Social Networks on the Development of Bilingualism in Kazakhstan // Advances

in Intelligent Systems and Computing. – 2024. – P. 123–134 (Springer, процентиль 39).

– Rakhimova D., Duisenbekkyzy Zh., Yessirkepova A., Adali E. Fine-Tuning Neural ASR Models for Low-Resource Kazakh Children's Speech with Preprocessing Enhancements // UBMK'2025. – Istanbul, 2025. – P. 428–432 (Scopus).

– Рахимова Д.Р., Дүйсенбекқызы Ж., Шадияр Т. Қазақ тіліндегі балалар корпусы үшін деректерді жинау және өңдеу // «ФАРАБИ ӘЛЕМІ». – Алматы, 4–6 сәуір 2024 ж.

– Рахимова Д.Р., Дүйсенбекқызы Ж., Турганбаева М.Ж., Арабов Б.Б. Балалар акустикалық корпусын қазақ тілінде жинау және өңдеу // «ǴYLYM JÁNE BILIM – 2024». – 2024.

– Дүйсенбекқызы Ж., Рахимова Д.Р., Есиркепова А.П. Нейрондық ASR үлгілерін төмен ресурсты қазақ балалар сөйлеуіне бейімдеу // «Информатика және қолданбалы математика» X конференциясы. – Алматы, 8–11 қазан 2025 ж. – Б. 303–309.

– Д.Р. Рахимова, Ж. Дүйсенбекқызы, О.А. Сағынтай, Ә. Турарбек, Ш. Төлеуғали. Балалардың сөйлеуін тану жүйелеріндегі шуды азайту алгоритмдерін талдау // ҚазҰПУ хабаршысы. – 2025. – №91(3). – Б. 230–242. DOI: 10.51889/2959-5894.2025.91.3.021

– Rakhimova D., Duisenbekkyzy Zh., Adali E., Toleugali Sh. Research and Processing of the Kazakh Children's Acoustic Corpus // ҚР ҰИА Хабаршысы. – 2024. – №3(93).

– Авторское свидетельство «Dataset Loader Bot». Объект авторского права №56775. Зарегистрировано: 15 апреля 2025 г. Авторы: Рахимова Д., Дүйсенбекқызы Ж., Алиев Р.

– Авторское свидетельство «Аннотированная акустическая база данных аудиозаписей казахскоязычной речи детей 2–9 лет». Объект авторского права №70955. Зарегистрировано: 27 апреля 2026 г. Авторы: Дүйсенбекқызы Ж., Рахимова Д.