Duisenbekkyzy Zhansaya, dissertation titled "Research and Development of a Method for Recognizing Kazakh Speech of Children with Speech Disorders", submitted for the degree of Doctor of Philosophy (PhD) in educational programme 8D06101 – Information Systems

ABSTRACT

Relevance of the research topic.

At the current stage of scientific and technological development, Automatic Speech Recognition (ASR) systems have become widely integrated into educational platforms, speech therapy diagnostic tools, and information systems for early childhood development. However, the automatic recognition of speech produced by children with speech disorders remains a scientifically underexplored and practically unsolved complex research problem.

Acoustic differences between children's and adults' speech have been confirmed by empirical studies. Specifically, the fundamental frequency (F0) of children's speech ranges from 250 to 400 Hz, compared to 100–150 Hz for adults; formant frequencies are also considerably higher due to the shorter vocal tract length in children (9–13 cm vs. 17 cm in adults). Full articulatory maturity is not achieved until the age of 12–14 years. Research demonstrates that ASR models trained on adult speech exhibit a Word Error Rate (WER) increase of 40–45% when applied to children's speech.

The Kazakh language is an agglutinative, morphologically complex low-resource language. Existing Kazakhstani ASR systems (KATEK, KazASR, KazakhTTS, KazakhBERT, Soyle) are trained exclusively on adult speech; no specialised annotated acoustic corpus of children's Kazakh speech existed prior to 2025. An independent evaluation of Kaldi, Mozilla DeepSpeech, and Google Speech-to-Text revealed that the best result was achieved by Google STT with WER = 52.97%, a level that is insufficient for practical deployment.

The scientific evidence and limitations of existing systems outlined above underscore the urgency of developing a scientific and technical solution for the automatic recognition and early diagnosis of Kazakh speech in children with speech disorders. The present dissertation introduces, for the first time, an annotated acoustic dataset of Kazakh children's speech and proposes a complex post-editing method based on the integration of ASR and LLM. The experimental results confirm that the proposed method achieves high accuracy, computational efficiency, and practical applicability.

Aim of the dissertation

To create, for the first time, an annotated acoustic database of Kazakh children's speech and to develop a complex ASR–LLM integrated post-editing

method aimed at improving the recognition accuracy of speech produced by children with speech disorders.

Research objectives:

–to conduct a comparative analytical review of global acoustic corpora of children's speech and to systematise their structural, volumetric, and acoustic characteristics; to perform a comparative analysis of state-of-the-art methods and acoustic corpora in the field of children's speech recognition;

–to collect and process an annotated acoustic dataset of Kazakh children's speech with consideration of children's age and gender characteristics;

–to investigate state-of-the-art neural ASR models (Whisper, MMS, TurkicASR, ESPnet, DeepSpeech, etc.) and to adapt them to children's speech via fine-tuning;

–to develop a complex LLM-based post-editing method for correcting ASR output after fine-tuning;

–to experimentally evaluate the effectiveness of the proposed method and conduct a comparative analysis.

Research methods: machine learning and deep learning methods, automatic speech recognition (ASR) technologies, acoustic signal processing, probability theory and mathematical statistics, software development methods.

Object of research

The speech process of children with speech disorders in the Kazakh language and its acoustic characteristics.

Subject of research

Technologies, algorithms, models, and methods for the automatic recognition of Kazakh speech of children with speech disorders, as well as the corresponding software.

Scientific novelty of the obtained results:

1. A specialised annotated acoustic dataset of Kazakh children's speech (8,594 audio recordings, ages 3–8) has been developed for the first time, taking into account the age, gender, and phonetic characteristics of the children. The created database makes a significant contribution to addressing the lack of open-access resources in the field of children's speech research.

2. A complex post-editing method combining ASR model fine-tuning and LLM-based post-processing has been proposed for the first time for the recognition of Kazakh speech produced by children with speech disorders. The proposed method improves the accuracy of Kazakh children's speech recognition.

Statements submitted for defence:

1. The developed annotated acoustic dataset of Kazakh children's speech, compiled with consideration of age, gender, and phonetic characteristics. This dataset constitutes the first specialised resource for researching Kazakh children's speech.

2. The proposed complex post-editing method, integrating ASR model fine-tuning for children's speech adaptation and LLM-based post-processing. Experimental results demonstrate a reduction in WER from 0.70 to 0.18 and CER from 0.467 to 0.065.

Volume and structure of the dissertation

The dissertation comprises an introduction, four main chapters, a conclusion, a list of references, and appendices. The total volume is 123 pages of typewritten text, including 30 figures, 43 tables, 117 references, and 4 appendices.

Chapter one provides an analytical review of research in the fields of speech recognition and speech disorders, systematically examining the evolution of ASR systems, the acoustic and phonetic characteristics of children's speech, the classification of speech disorders, and international experience with low-resource languages.

Chapter two presents the methodology for selecting textual material, methods for collecting and annotating acoustic data, and ethical and legal aspects; the dataset structure and data pre-processing methods are described. Audio data were collected exclusively on the basis of written informed consent from parents or legal representatives in accordance with the Law of the Republic of Kazakhstan «On Personal Data and Their Protection» (No. 94-V); all recordings are de-identified through the assignment of unique identifiers and stored on secured servers located within the Republic of Kazakhstan with restricted access. Participants' personal data are not published or disclosed to third parties.

Chapter three describes the architectural design of the proposed complex method, the formalisation of the speech recognition problem, the mathematical formulation of the acoustic model adapted to children's speech, and the mathematical model of the complex method.

Chapter four presents the experimental evaluation of the proposed method: a comparative analysis of ASR models, the results of fine-tuning and LLM-based post-editing, and the outcomes of the method's practical validation.

Approbation of the work

The research results were discussed at scientific seminars of Al-Farabi Kazakh National University, presented at international conferences, and underwent practical approbation at kindergarten No. 154.

Publications.

A total of 9 scientific works have been published on the topic of the dissertation, including 2 articles in journals indexed by Web of Science (SCIE, Q2) and (ESCI, Q1), 5 works in international conference proceedings, 2 works in journals recommended by the Committee for Quality Assurance in Education and Science of the Ministry of Education and Science of the Republic of Kazakhstan; 2 copyright certificates have been obtained.

- 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, percentile 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, percentile 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, percentile 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).
- Rakhimova D.R., Duisenbekkyzy Zh., Shadiyar T. Data collection and processing for the Kazakh children's corpus // International Scientific Conference «FARABI ALEMI». – Almaty, April 4–6, 2024.
- Rakhimova D.R., Duisenbekkyzy Zh., Turganbayeva M.Zh., Arabov B.B. Collection and processing of the children's acoustic corpus in Kazakh //

International Scientific Conference «ǴYLYM JÁNE BILIM – 2024». – April 2024.

– Duisenbekkyzy Zh., Rakhimova D.R., Yessirkepova A.P. Adaptation of neural ASR models for low-resource Kazakh children's speech // X International Scientific-Practical Conference on Informatics and Applied Mathematics. – Almaty, October 8–11, 2025. – P. 303–309.
– Rakhimova D.R., Duisenbekkyzy Zh., Sagyntay O.A., Turarbek A., Toleugali Sh. Analysis of noise reduction algorithms in speech recognition systems for children // Bulletin of KazNPU. – 2025. – No. 91(3). – P. 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 // Bulletin of NAS RK. – 2024. – No. 3(93).
– Copyright certificate «Dataset Loader Bot». Copyright object No. 56775. Registered: April 15, 2025. Authors: Rakhimova D., Duisenbekkyzy Zh., Aliyev R.
– Copyright certificate «Annotated acoustic database of Kazakh speech recordings of children aged 2–9 years». Copyright object No. 70955. Registered: April 27, 2026. Authors: Duisenbekkyzy Zh., Rakhimova D.