İTÜ

□ Zip Code : 34469
□ Tel. : +90-212-285-3601
□ Web : www.itu.edu.tr

REVIEW OF THE FOREIGN SCIENTIFIC CONSULTANT

on the dissertation work of Duisenbekkyzy Zhansaya on the topic: «The research and development of the Kazakh speech's recognition of children with speech disorders» submitted for the degree of Doctor of Philosophy (PhD) in the educational program:

8D06101 – Information Systems

The dissertation work of Duisenbekkyzy Zhansaya is devoted to a highly relevant and practically significant problem at the intersection of artificial intelligence, speech technology, and clinical linguistics: the development of an automatic speech recognition (ASR) system specifically designed for Kazakh-speaking children with speech sound disorders. This topic addresses a critical scientific and social gap in the field of low-resource language technologies, and the author has approached it with both scientific rigor and a clear awareness of its real-world clinical urgency.

The relevance of the research is thoroughly substantiated. Children's speech differs substantially from adult speech in terms of fundamental frequency, formant distribution, and articulatory maturity, resulting in a Word Error Rate increase of 40–45% when adult-trained models are applied to children's voice data. The Kazakh language compounds this challenge due to its agglutinative morphology with 150–200 grammatical forms per root word, vowel harmony constraints, and unique phonemes that are systematically misrecognized by existing systems. The documented shortage of speech-language pathologists in Kazakhstan — one specialist per 700–900 children in rural areas — makes automated screening tools not merely convenient but critically necessary.

Chapter 1 presents a comprehensive systematic review of ASR architectures spanning classical HMM/GMM approaches through DNN-HMM, CTC, Transformer, Whisper, and self-supervised learning models, as well as a thorough analysis of existing children's acoustic corpora worldwide. The author establishes convincingly that, prior to this work, no annotated acoustic dataset for Kazakh children's speech existed, providing strong and well-grounded motivation for the research.

Chapter 2 describes the creation of the first annotated Kazakh children's speech dataset, which constitutes a major standalone scientific contribution. The corpus comprises 8,594 audio recordings from children aged 2–11 years, collected through three complementary channels: a semi-automated Telegram-based collection bot (Copyright Certificate No. 56775), naturalistic home and kindergarten recordings, and controlled dictaphone sessions. The preprocessing pipeline — including noise reduction, harmonic-percussive source separation, silence trimming, amplitude normalization, and acoustic feature extraction — is methodologically thorough. The dataset has been formally registered as an intellectual property object (Copyright Certificate No. 70955), confirming its originality and long-term scientific value.

Chapter 3 presents a rigorous mathematical formalization of the proposed method. The five-stage pipeline is grounded in a clear theoretical framework: from Bayesian ASR formulation and CTC loss derivation through hybrid CTC+attention objectives and domain adaptation via Kullback–Leibler divergence minimization. A particularly strong contribution is the pathology-specific modeling of six speech disorders — dyslalia, stuttering, bradylalia, tachylalia, rhinolalia, and dysarthria — each with dedicated mathematical representations. Age-stratified Vocal Tract Length Normalization coefficients further strengthen the methodological foundation.

• Address : Maslak Mahallesi, Resat Nuri Sokak No:13, 34467 Sarıyer/Istanbul, Turkiye

İTÜ

□ Zip Code : 34469
□ Tel. : +90-212-285-3601
□ Web : www.itu.edu.tr

The three-stage post-editing pipeline (PhonNorm → MorphNorm → LLM contextual re-scoring) is a novel and technically well-justified contribution. The phonetic normalization stage addresses articulatory substitutions characteristic of children's developmental speech and Kazakh-specific phonemes; the morphological normalization stage enforces vowel harmony and grammatical correctness using the Apertium-Kaz analyzer; and the LLM re-scoring stage resolves semantically incorrect but phonetically plausible ASR outputs. This design reflects a deep and technically mature understanding of the linguistic challenges specific to Kazakh.

Chapter 4 reports rigorous three-stage experimental evaluation on a held-out test set of 1,000 audio files. The progression from zero-shot baselines (WER = 0.70–0.90 across all systems) through fine-tuning (Whisper WER reduced from 0.70 to 0.30) to the full proposed method (WER = 0.18, Accuracy = 70%) demonstrates a 74.3% overall error reduction relative to the baseline. The clinical field test at Kindergarten No. 154 in Almaty, involving 48 children aged 3–8 and expert evaluation by certified speech-language pathologists, provides convincing real-world validation of the system.

The doctoral candidate's publication record is impressive and consistent with the requirements for a PhD degree. The work includes articles in Web of Science and Scopus-indexed journals, including a Q2 WoS SCIE publication in Applied Sciences and a Q1 ESCI publication in Computers, as well as conference papers at international venues including UBMK'2025 in Istanbul. Two registered copyright certificates confirm the originality and practical significance of the developed tools and datasets.

Overall, the dissertation is completed at a high scientific and methodological level, with well-supported conclusions and a clear practical orientation. The work demonstrates scientific novelty through the development of the first annotated Kazakh children's speech corpus, a dedicated neural ASR pipeline, and a linguistically informed post-editing method, all validated in a real clinical setting.

Based on the analysis of the obtained results, it can be concluded that the dissertation work of Duisenbekkyzy Zhansaya is a completed scientific research project and can be recommended for defense for the degree of Doctor of Philosophy (PhD).

Sincerely,

Eşref Adalı, Professor
Faculty of Computer and Informatics
Engineering Istanbul Technical
University Website:
http://web.itu.edu.tr/adali/ Email:
adali@itu.edu.tr
Phone: +90-532-314-2349

Логотип: /Ыстамбұл техникалық университеті/

Пошталық индекс: 34469
Тел.: +90-212-285-3601
Веб-сайт: www.itu.edu.tr

«Сөйлеуінде бұзылыстары бар балалардың қазақ тіліндегі сөйлеуін тануды зерттеу және әзірлеу» тақырыбындағы 8D06101 - Ақпараттық жүйелер
білім беру бағдарламасы бойынша философия докторы (PhD) дәрежесін алу үшін
ұсынылған Дүйсенбекқызы Жансаяның диссертациялық жұмысына
ШЕТЕЛДІК ҒЫЛЫМИ КОНСУЛЬТАНТТЫҢ ПІКІРІ

Дүйсенбекқызы Жансаяның диссертациялық жұмысы жасанды интеллект, сөйлеу технологиялары және клиникалық лингвистика тоғысындағы аса өзекті әрі практикалық тұрғыдан маңызды мәселеге: сөйлеу дыбыстары бұзылыстары бар қазақтілді балаларға арнайы арналған автоматты сөйлеуді тану (ASR) жүйесін әзірлеуге арналған. Бұл тақырып ресурсы аз тілдер технологиялары саласындағы маңызды ғылыми және әлеуметтік олқылықты қамтиды, ал автор оған ғылыми қатаңдықпен әрі оның нақты клиникалық өзектілігін анық түсіне отырып қараған.

Зерттеудің өзектілігі жан-жақты негізделген. Балалардың сөйлеуі негізгі жиілік, форманттардың таралуы және артикуляциялық жетілу тұрғысынан ересектердің сөйлеуінен едәуір ерекшеленеді, соның нәтижесінде ересектердің дауыс деректері бойынша оқытылған модельдер балалардың дауыс деректеріне қолданылған кезде сөз қателігі коэффициенті 40-45%-ға артады. Қазақ тілі бұл міндетті түбір сөзге шаққанда 150-200 грамматикалық формаға ие агглютинативті морфологиясы, дауыстылар үндестігі шектеулері және қолданыстағы жүйелер тарапынан жүйелі түрде қате танылатын бірегей фонемалары арқылы күрделендіреді. Қазақстанда сөйлеу-тіл патологиясы мамандарының құжатталған тапшылығы - ауылдық жерлерде 700-900 балаға бір маманнан келуі - автоматтандырылған скрининг құралдарын жай ғана ыңғайлы емес, аса қажет етеді.

1-тарауда классикалық HMM/GMM тәсілдерінен бастап DNN-HMM, CTC, Transformer, Whisper және өзін-өзі бақылап оқыту модельдеріне дейінгі ASR архитектураларына жан-жақты жүйелі шолу, сондай-ақ әлемдегі қолданыстағы балалардың акустикалық корпустарына мұқият талдау ұсынылған. Автор осы жұмысқа дейін қазақ тіліндегі балалар сөйлеуінің аннотацияланған акустикалық деректер жиынтығы болмағанын сенімді түрде негіздейді, бұл зерттеу үшін күшті әрі жақсы негізделген уәждеме береді.

2-тарауда алғашқы аннотацияланған қазақ тіліндегі балалар сөйлеуі деректер жиынтығын жасау сипатталады, бұл өз алдына ірі ғылыми үлес болып табылады. Корпус 2-11 жас аралығындағы балалардан алынған 8 594 аудиожазбаны қамтиды, олар үш өзара толықтырушы арна арқылы жиналған: Telegram негізіндегі жартылай автоматтандырылған жинау боты (Авторлық құқық туралы куәлік № 56775), үй және балабақша жағдайындағы табиғи жазбалар және бақыланатын диктофондық сессиялар. Алдын ала өңдеу конвейері - шуды азайтуды, гармониялық-перкуссиялық көздерді ажыратуды, тыныштықты қиып тастауды, амплитуданы қалыпқа келтіруді және акустикалық белгілерді алуды қоса алғанда - әдіснамалық тұрғыдан жан-жақты. Деректер жиынтығы зияткерлік меншік объектісі ретінде ресми түрде тіркелген (Авторлық құқық туралы куәлік № 70955), бұл оның түпнұсқалығын және ұзақ мерзімді ғылыми құндылығын растайды.

3-тарауда ұсынылған әдістің қатаң математикалық формализациясы ұсынылған. Бес кезеңнен тұратын конвейер айқын теориялық негізге сүйенеді: Bayesian ASR формулировкасынан және CTC loss шығарылуынан бастап гибридті CTC+attention мақсаттары мен Kullback Leibler дивергенциясын минимизациялау арқылы домендік

бейімдеуге дейін. Ерекше күшті үлес - алты сөйлеу бұзылысын - дислалия, тұтықпа, брадилалия, тахилалия, ринолалия және дизартрияны - патологияға тән модельдеу, олардың әрқайсысы арнайы математикалық көріністермен берілген. Жас бойынша стратификацияланған Vocal Tract Length Normalization коэффициенттері әдіснамалық негізді одан әрі күшейтеді.

Үш кезеңді постредакциялау конвейері (PhonNorm → MorphNorm → LLM контекстік қайта бағалауы) жаңа әрі техникалық тұрғыдан жақсы негізделген үлес болып табылады. Фонетикалық қалыпқа келтіру кезеңі балалардың даму кезеңіндегі сөйлеуіне және қазақ тіліне тән фонемаларға сипаттас артикуляциялық алмастыруларды қамтиды; морфологиялық қалыпқа келтіру кезеңі Apertium-Kaz талдағышын пайдалана отырып, дауыстылар үндестігі мен грамматикалық дұрыстықты қамтамасыз етеді; ал LLM қайта бағалау кезеңі фонетикалық тұрғыдан ықтимал, бірақ семантикалық тұрғыдан қате ASR нәтижелерін шешеді. Бұл құрылым қазақ тіліне тән лингвистикалық қиындықтарды терең әрі техникалық тұрғыдан жетілген түсінуді көрсетеді.

4-тарауда 1 000 аудиофайлдан тұратын кейінге қалдырылған тест жиынтығында жүргізілген қатаң үш кезеңді эксперименттік бағалау баяндалады. Zero-shot базалық модельдерінен (барлық жүйелер бойынша WER = 0.70-0.90) fine-tuning арқылы (Whisper WER 0.70-тен 0.30-ға дейін төмендеді) толық ұсынылған әдіске дейінгі ілгерілеу (WER = 0.18, Accuracy = 70%) базалық көрсеткішпен салыстырғанда жалпы қатенің 74.3%-ға төмендегенін көрсетеді. Алматы қаласындағы № 154 балабақшада 3-8 жас аралығындағы 48 баланың қатысуымен жүргізілген және сертификатталған сөйлеу-тіл патологиясы мамандарының сараптамалық бағалауын қамтыған клиникалық далалық сынақ жүйенің нақты жағдайдағы сенімді валидациясын қамтамасыз етеді.

Докторанттың жарияланымдар тізімі әсерлі және PhD дәрежесіне қойылатын талаптарға сәйкес келеді. Жұмыс Web of Science және Scopus индекстелетін журналдардағы мақалаларды, соның ішінде Applied Sciences журналындағы Q2 WoS SCIE жарияланымын және Computers журналындағы Q1 ESCI жарияланымын, сондай-ақ Ыстамбұлдағы UBMK'2025 сияқты халықаралық алаңдардағы конференция материалдарын қамтиды. Тіркелген екі авторлық құқық туралы куәлік әзірленген құралдар мен деректер жиынтықтарының түпнұсқалығын және практикалық маңыздылығын растайды.

Жалпы алғанда, диссертация жоғары ғылыми және әдіснамалық деңгейде орындалған, қорытындылары жақсы негізделген және практикалық бағыты айқын. Жұмыс алғашқы аннотацияланған қазақ тіліндегі балалар сөйлеуі корпусын, арнайы нейрондық ASR конвейерін және лингвистикалық тұрғыдан негізделген постредакциялау әдісін әзірлеу арқылы ғылыми жаңалығын көрсетеді, олардың барлығы нақты клиникалық жағдайда валидациядан өткен.

Алынған нәтижелерді талдау негізінде Дүйсенбекқызы Жансаяның диссертациялық жұмысы аяқталған ғылыми-зерттеу жобасы болып табылады және философия докторы (PhD) дәрежесін алу үшін қорғауға ұсынылуы мүмкін деген қорытынды жасауға болады.

Құрметпен,
/Қолтаңба/
Эшреф Адалы, профессор
Компьютер және информатика инженериясы факультеті
Ыстамбұл техникалық университеті
Веб-сайт: http://web.itu.edu.tr/adali/
Электрондық пошта: adali@itu.edu.tr
Телефон: +90-532-314-2349

Мекенжайы: Маслак махаллесі, Решат Нури көшесі №13, 34467 Сарыер/Ыстамбұл, Түркия