

Ж. Дүйсенбекқызы^{1*}, А. Тұрсынхан², Г. Газиз³, Б. Илсова¹

¹Алматы технологиялық университеті, Алматы қ., Қазақстан

²Халықаралық инженерлік-технологиялық университеті, Алматы қ., Қазақстан

³Абай атындағы Қазақ ұлттық педагогикалық университеті, Алматы қ., Қазақстан

*e-mail: duisenbekkyzy.zh@kaznu.kz

ТРАНСФЕРЛІК ОҚЫТУ ӘДІСІ НЕГІЗІНДЕ ҚАЗАҚ ТІЛДІ БАЛАЛАРДЫҢ СӨЙЛЕУІН АВТОМАТТЫ ТАҢУ ЖҮЙЕСІН МОДИФИКАЦИЯЛАУ

Аңдатпа

Зерттеу жұмысы қазақ тілінің акустикалық және фонетикалық ерекшеліктерін ескере отырып, 5–12 жас аралығындағы балалардың сөйлеуін автоматты таңу тиімділігін арттыруға бағытталған. Зерттеудің мақсаты – OpenAI ұсынған Whisper-tiny архитектурасын балалар сөйлеуінің мамандандырылған аудиодеректері негізінде қайта оқыту арқылы қазақ тіліне бейімдеу. Жұмыста екі кезеңді fine-tuning стратегиясы қолданылды: бірінші кезеңде модель 235 аудиоүлгімен, ал екінші кезеңде 1936 үлгіге дейін кеңейтілген корпуспен оқытылды. Аудиожазбалардың дискреттеу жиілігі 16 кГц-ке келтіріліп, мәтіндік транскрипциялар біріздендірілді. Модель сапасы Word Error Rate (WER) және Accuracy метрикалары арқылы бағаланды. Эксперимент нәтижесінде WER көрсеткіші 0,80-нен 0,42-ге дейін төмендеп, салыстырмалы жақсару 47,5%-ды құрады, ал дәлдік 20,43%-дан 57,95%-ға дейін артты. Нәтижелер деректер көлемін ұлғайту мен доменге бейімделген қайта оқытудың балалар сөйлеуін таңу сапасына оң әсер ететінін көрсетті. Ұсынылған тәсілді аз ресурсты тілдерге арналған білім беру және интеллектуалды сөйлеу жүйелерін әзірлеуде қолдануға болады.

Түйін сөздер: автоматты сөйлеуді таңу (ASR), Whisper модельдері, қазақ тілі, балалар сөйлеуі, fine-tuning, трансферлік оқыту, WER метрикасы, нейрондық желілер.

Ж. Дүйсенбекқызы¹, А. Тұрсынхан², Г. Газиз³, Б. Илсова¹

¹Алматинский технологический университет, г. Алматы, Казахстан

²Международный инженерно-технологический университет, г. Алматы, Казахстан

³Казахский национальный педагогический университет имени Абая, г. Алматы, Казахстан

МОДИФИКАЦИЯ СИСТЕМЫ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ ДЕТСКОЙ РЕЧИ НА КАЗАХСКОМ ЯЗЫКЕ ПОСРЕДСТВОМ ТРАНСФЕРНОГО ОБУЧЕНИЯ

Аннотация

Данная работа посвящена повышению эффективности автоматического распознавания речи детей в возрасте 5–12 лет с учетом акустических и фонетических особенностей казахского языка. Цель исследования – адаптировать предложенную OpenAI архитектуру Whisper-tiny к казахскому языку посредством дообучения на специализированном наборе аудиоданных детской речи. В работе применена двухэтапная стратегия fine-tuning: на первом этапе модель обучалась на 235 аудиообразцах, а на втором – на расширенном корпусе из 1936 образцов. Аудиозаписи были приведены к частоте дискретизации 16 кГц, а текстовые транскрипции унифицированы. Качество модели оценивалось с использованием метрик Word Error Rate (WER) и Accuracy. В результате эксперимента показатель WER снизился с 0,80 до 0,42, что соответствует относительному улучшению на 47,5%, а точность увеличилась с 20,43% до 57,95%. Результаты показали, что увеличение объема данных и доменно-ориентированное дообучение положительно влияют на распознавание детской речи. Предложенный подход может применяться при разработке образовательных и интеллектуальных речевых систем для языков с ограниченными ресурсами.

Ключевые слова: автоматическое распознавание речи (ASR), модели Whisper, казахский язык, детская речь, fine-tuning, трансферное обучение, метрика WER, нейронные сети.

Zh. Duisenbekkyzy¹, A. Tursynkhan², G. Gaziz³, B. Ilesova¹

¹Almaty Technological University, Almaty, Kazakhstan

²International Engineering and Technology University, Almaty, Kazakhstan

³Abai Kazakh National Pedagogical University, Almaty, Kazakhstan

ENHANCING AUTOMATIC SPEECH RECOGNITION FOR KAZAKH PEDIATRIC SPEECH THROUGH TRANSFER LEARNING AND FINE-TUNING

Abstract

This study aims to improve automatic speech recognition for children aged 5–12 by accounting for the acoustic and phonetic characteristics of the Kazakh language. The objective is to adapt OpenAI's Whisper-tiny architecture to Kazakh through fine-tuning on a specialized children's speech dataset. A two-stage fine-tuning strategy was applied: the model was first trained on 235 audio samples and then on an expanded corpus of 1,936 samples. The audio recordings were standardized to a sampling rate of 16 kHz, and the text transcriptions were normalized. Model performance was evaluated using Word Error Rate (WER) and Accuracy. The experimental results showed that WER decreased from 0.80 to 0.42, representing a relative improvement of 47.5%, while Accuracy increased from 20.43% to 57.95%. These findings demonstrate that increasing the amount of data and applying domain-specific fine-tuning improve the recognition of children's speech. The proposed approach can support the development of educational and intelligent speech systems for low-resource languages.

Keywords: Automatic Speech Recognition (ASR), Whisper models, Kazakh language, pediatric speech, fine-tuning, transfer learning, WER metric, neural networks.

Кіріспе

Негізгі ережелер. Зерттеу жұмысы 5-12 жас аралығындағы қазақ тілді балалардың сөйлеуін автоматты тануды жетілдіруге бағытталған. OpenAI Whisper-tiny архитектурасы екі кезеңді fine-tuning стратегиясы арқылы мамандандырылған аудиодеректер жиынтығында бейімделді. Корпус бастапқыда 235 аудиоүлгіден тұрып, кейіннен 1936 сәмплге дейін кеңейтілді. Нәтижесінде WER (Word Error Rate) 0.80-ден 0.42-ге дейін төмендеп, жүйенің дәлдігі 20,43%-дан 57,95%-ға дейін артты. Алынған нәтижелер ресурсы шектеулі тілдер үшін балалар сөйлеуін тану жүйелерін әзірлеуде доменге бейімделген қосымша оқытудың тиімді екенін көрсетті. Сөйлеуді автоматты тану (ASR) технологиялары дүние жүзінде қарқынды дамып, білім беру, денсаулық сақтау, интеллектуалды интерфейстер және қолжетімділік салаларында кеңінен қолданылуда. Соңғы жылдары end-to-end архитектуралар мен transformer негізіндегі жүйелер үлкен көлемдегі деректерді пайдалана отырып, ASR дәлдігін айтарлықтай жақсартты. Алайда ғылыми әдебиеттер ASR модельдерінің өнімділігі олардың оқытылған тіліне және доменіне тікелей байланысты екенін көрсетеді [1]. Осылайша, аз ресурсты тілдер үшін сөйлеуді тану жүйелерін әзірлеу негізгі ғылыми міндеттердің бірі болып қала береді.

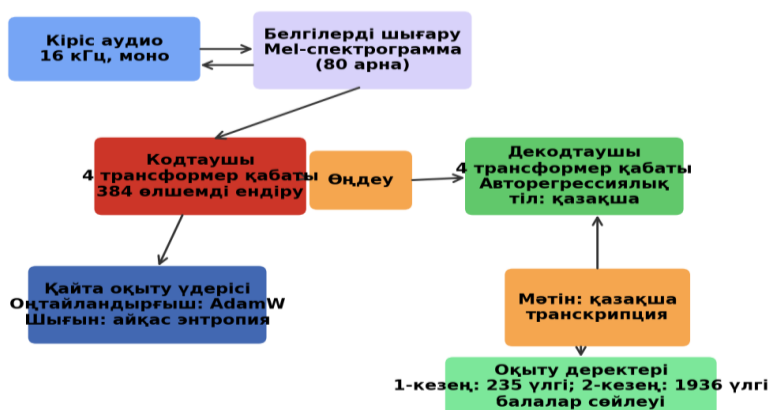
Аз ресурсты тілдер арасында қазақ тілі үшін ASR зерттеулері әлі де жеткіліксіз. Қазақ тілінің агглютинативті морфологиясы, дауысты үндестік, флексиясыз құрылым және фонетикалық тұрақсыздық сөйлеуді тану модельдері үшін айтарлықтай қиындықтар туғызады [2]. Бұл лингвистикалық ерекшеліктер алдын ала оқытылған көптілді модельдерді қазақ тіліне тікелей қолдануды шектейді.

Балалар сөйлеуін тану ASR саласындағы дербес күрделі мәселе болып саналады. Ересектер сөйлеуімен салыстырғанда балалар сөйлеуінде негізгі тон жиілігі жоғары, артикуляциясы тұрақсыз және форманттық жиіліктері өзгермелі болады [3]. Мұндай ерекшеліктер тану қателерінің көбеюіне әкеліп, балалар сөйлеуіне бейімделген модельдерді әзірлеуді қажет етеді. OpenAI әзірлеген Whisper моделі 680 000 сағаттан астам көптілді сөйлеу деректері негізінде алдын ала оқытылған. Дегенмен модель белгілі бір тілдің және қолданылу саласының ерекшеліктеріне қосымша бейімделмесе, оның тану дәлдігі төмендеуі мүмкін [4]. Трансферлік оқыту әдістері көлемі шектеулі домендік деректер негізінде модельдің нәтижелілігін арттыруға мүмкіндік береді [5]. Д.Р. Рахимова мен Ж. Дүйсенбекқызының доменге бейімделген fine-tuning жөніндегі зерттеулері [6] де осы тұжырымды растайды.

Зерттеудің мақсаты – Whisper-tiny моделін қазақ тіліндегі балалар сөйлеуін тануға бейімдеу және екі кезеңді fine-tuning стратегиясының тиімділігін эксперименттік тұрғыдан бағалау.

Зерттеу әдіснамасы

Зерттеуде OpenAI әзірлеген Whisper-tiny моделі қолданылды. Whisper — кодтаушы-декодтаушы (encoder-decoder) құрылымына негізделген трансформерлік ASR моделі; оның tiny нұсқасында 39 миллион параметр бар [7]. Архитектура үш негізгі құрамдастан тұрады: белгілерді бөліп алу модулі (Feature Extractor), кодтаушы (Encoder) және декодтаушы (Decoder). Белгілерді бөліп алу кезеңінде аудио 16 кГц жиілікке қайта дискреттеліп, 80 өлшемді мел-спектрограмма құрылады. Кодтаушы мен декодтаушының әрқайсысы төрт трансформерлік қабаттан тұрады, ал декодтау авторегрессиялық тәсілмен орындалады. Whisper-tiny моделін қосымша оқыту үдерісінің архитектурасы 1-суретте көрсетілген. Whisper-tiny моделінің негізгі конфигурациялық сипаттамалары 1-кестеде берілген.



Сурет 1. Whisper Fine-Tuning архитектурасы

Кесте 1. Whisper-tiny моделінің негізгі конфигурациялық сипаттамалары

Параметр	Сипаттама
Параметрлер саны	39 миллион
Encoder қабаттары	4
Decoder қабаттары	4
Attention Heads	6
Mel-Spectrogram өлшемі	80 bins
Pre-training көлемі	680,000 сағат аудио

Зерттеуде деректер жиыны екі кезеңде қалыптастырылды: бірінші кезеңде 235 аудиоүлгі, ал екінші кезеңде 1936 аудиоүлгі пайдаланылды (2-кесте). Барлық аудиожазба 16 кГц дискреттеу жиілігіне, моноарнаға және 16 биттік PCM пішіміне келтірілді. Мәтіндік транскрипциялар кіші әріппен жазылып, бірыңғай форматқа түсірілді. Екі кезеңдегі аудиодеректердің құрамы мен сандық сипаттамалары 2-кестеде берілген.

Кесте 2. Деректер жиынының сипаттамалары

Сипаттама	1-кезең (235)	2-кезең (1936)
Бір сөздер	183 (78%)	1479 (76,3%)
2–5 сөзді фразалар	52 (22%)	446 (23%)
6–10 сөз	0	6 (0,3%)
11+ сөз	0	7 (0,4%)
Бірегей сөздер	134	529
Жалпы токенодер	612	2858

2-кестеге сәйкес, екінші кезеңде деректер жиыны едәуір кеңейтілді. Бір сөзден тұратын аудиоүлгілер саны 183-тен 1479-ға, ал 2-5 сөзден тұратын фразалар саны 52-ден 446-ға дейін өсті. Бірегей сөздер саны 134-тен 529-ға, жалпы токендер саны 612-ден 2858-ге дейін артты. Сонымен қатар кеңейтілген корпуста 6-10 және 11-ден көп сөзден тұратын күрделі фразалар қосылды. Бірінші кезеңде модель 235 үлгі негізінде қосымша оқытылып, қазақ тілінің негізгі фонетикалық заңдылықтарына бейімделді. Екінші кезеңде модель 1936 үлгіден тұратын кеңейтілген корпуспен қайта оқытылды [8]. Шығын функциясы кросс-энтропия әдісімен есептелді: $L = -\sum \log P(y_t|y_{<t}, X)$, мұндағы X - мел-спектрограмма, ал y_t - транскрипцияның t мезетіндегі токени. Модель NVIDIA T4 графикалық процессорында AdamW оңтайландырғышын пайдалану арқылы оқытылды. ASR жүйесінің сапасын бағалау үшін екі метрика қолданылды: $WER = (S + D + I) / N$ және $Accuracy = (1 - WER) \times 100\%$. Мұндағы S - алмастырулар саны, D - түсіріп қалдырулар саны, I - артық кірістірулер саны, ал N - эталондық мәтіндегі сөздердің жалпы саны.

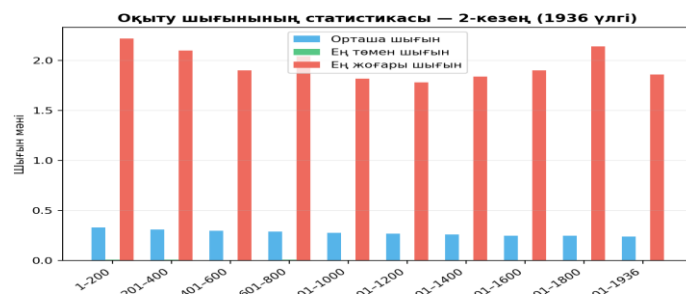
Зерттеу нәтижелері

2-суретте 1-кезең оқыту барысында (235 мысал) loss мәнінің өзгерісі көрсетілген. Бастапқыда loss өте жоғары (~8,8) болды, бірақ алғашқы бірнеше итерацияда тез төмендеді және ~0,4 деңгейінде тұрақтады. Бірінші кезеңдегі оқыту процесінің тұрақтылығын бағалау үшін loss функциясының итерациялар бойынша өзгерісі талданды. Оның динамикасы 2-суретте көрсетілген.

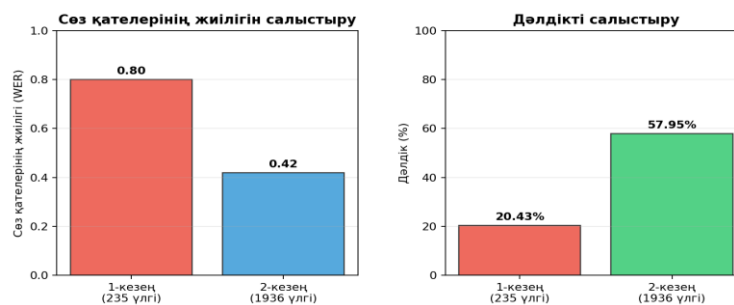


Сурет 2. Оқыту Loss динамикасы (1-кезең)

2-суреттен шығын функциясының бастапқы мәні шамамен 8,8 болғанын және алғашқы итерациялар кезінде жылдам төмендегенін көруге болады. Оқыту соңында бұл көрсеткіш шамамен 0,4 деңгейіне жетті. Мұндай динамика модельдің бастапқы 235 аудиоүлгідегі негізгі акустикалық және фонетикалық заңдылықтарды меңгергенін көрсетеді. Екінші кезеңдегі кеңейтілген fine-tuning нәтижелері 3-суретте берілген. Орташа шығын мәні 0,23-0,33 аралығында өзгерді. Ең төменгі мәнің 0,005–0,01 аралығында болуы қарапайым сөздердің жақсы танылғанын, ал ең жоғары мәндердің 1,8–2,2 аралығында болуы күрделі немесе ұзақ фразаларды тануда қателердің сақталғанын көрсетеді. Алынған статистикалық нәтижелер 3-суретте берілген. 3-суретке сәйкес, шығын функциясының орташа мәні 0,23–0,33 аралығында өзгерді. Ең төменгі мәндер 0,005–0,01 аралығында тіркеліп, қарапайым сөздердің жоғары дәлдікпен танылғанын көрсетті. Ең жоғары мәндердің 1,8–2,2 аралығында болуы ұзақ әрі күрделі фразаларды тану кезінде жекелеген қателердің сақталғанын білдіреді. Екі кезеңдегі модель нәтижелерін салыстыру үшін WER және Accuracy көрсеткіштері қолданылды. Салыстыру нәтижелері 4-суретте берілген: WER 0,80-нен 0,42-ге дейін төмендеп, 47,5% салыстырмалы жақсару байқалды; Accuracy 20,43%-дан 57,95%-ға дейін артты. Оқшауланған сөздер үшін толық сәйкестік (Exact Match) көрсеткіші 66,73%-ды құрады. Аталған метрикалардың салыстырмалы нәтижелері 4-суретте көрсетілген. 4-суретке сәйкес, екінші кезеңдегі fine-tuning нәтижесінде WER көрсеткіші 0,80-ден 0,42-ге дейін төмендеді. Accuracy көрсеткіші 20,43%-дан 57,95%-ға дейін артты.



Сурет 3. Оқыту Loss статистикасы, 2-кезең (1936 сэмпл)



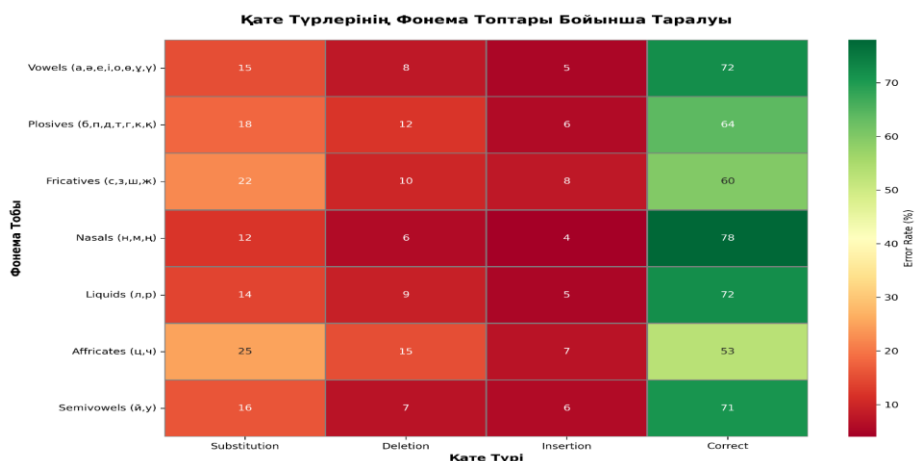
Сурет 4. WER және Accuracy салыстыруы

Зерттеуде модельді дайындау, оқыту және бағалау бірнеше өзара байланысты кезең арқылы орындалды. Ұсынылған екі кезеңді оқыту процесінің толық схемасы 5-суретте берілген.



Сурет 5. Оқыту процесінің толық схемасы

5-суретте көрсетілгендей, зерттеу процесі аудиодеректерді жинау мен алдын ала өңдеуден, модельді екі кезеңде қайта оқытудан және алынған нәтижелерді WER мен Accuracy метрикалары арқылы бағалаудан тұрады. Мұндай кезеңдік тәсіл модельді алдымен қазақ тілінің негізгі фонетикалық ерекшеліктеріне, кейін балалар сөйлеуінің күрделі үлгілеріне бейімдеуге мүмкіндік берді. 6-суреттегі график деректер жиыны өлшемі мен модель дәлдігі арасындағы тікелей байланысты суреттейді. Жалпы жақсарту метрикалары 3-кестеде берілген. Бұл байланыстың графикалық көрінісі 6-суретте ұсынылған. 6-суреттен деректер жиынының көлемі артқан сайын модель дәлдігінің де жоғарылағанын көруге болады. Бастапқы 235 сэмпл пайдаланылған кезде дәлдік 20,43%-ды құраса, 1936 сэмплмен оқытудан кейін 57,95%-ға жетті. Жалпы жақсарту көрсеткіштері 3-кестеде берілген. 3-кестеде көрсетілгендей, деректер жиынының көлемі бастапқы мөлшермен салыстырғанда шамамен 8,23 есе ұлғайды. Осы кезде Accuracy көрсеткіші 183,7%-ға өсті, WER 47,5%-ға, ал шығын функциясының мәні 84,2%-ға төмендеді. Бұл нәтижелер деректер корпусын кеңейту және екі кезеңді fine-tuning стратегиясын қолдану модель нәтижелілігін арттырғанын көрсетеді.



Сурет 6. Деректер жиыны өлшемінің модель дәлдігіне әсері

Кесте 3. Жалпы жақсарту метрикалары

Метрика	Өзгеріс (%)
Accuracy өсімі	+183,7%
WER төмендеуі	-47,5%
Loss төмендеуі	-84,2%
Деректер жиыны өлшемінің өсімі	+823,0%

Дискуссия

Алынған нәтижелер деректер жиынының көлемі мен сапасы модель нәтижелілігіне елеулі әсер ететінін көрсетті. Деректер жиынының көлемі шамамен 8,23 есе ұлғайған кезде WER 47,5%-ға төмендеп, Accuracy 183,7%-ға артты. Бұл нәтиже нейрондық желілердің тиімділігі оқыту деректерінің көлеміне тәуелді екені туралы тұжырымдармен сәйкес келеді [9]. Фонемалар деңгейіндегі қателер 7-суретте талданды.

Болжамдағы қателердің 10 мысалы



Сурет 7. Фонема тобы бойынша қателіктерді талдау

Аффрикаттар (ц, ч) тобындағы алмастыру қателерінің үлесі 25%-ды құрады. Фрикативтер мен шұғыл дауыссыздардағы қателер тиісінше 22% және 18% болды. Мұрын жолды дауыссыздар (н, м, ң) мен үнді дауыссыздар (л, р) топтарындағы қателердің үлесі төменірек болып, тиісінше 12% және 14%-ды құрады. 7-суретке сәйкес, ең жоғары алмастыру қатесі аффрикаттар (ц, ч) тобында тіркеліп, 25%-ды құрады. Фрикативтер мен шұғыл дауыссыздардағы қателердің үлесі тиісінше 22% және 18% болды. Мұрын жолды дауыссыздарда (н, м, ң) және үнді дауыссыздарда (л, р) қателер сирегірек кездесіп, тиісінше 12% және 14%-ды құрады. Нәтижелер акустикалық белгілері ұқсас кейбір фонемаларды ажырату модель үшін күрделі болғанын көрсетті. Мысалы, модель «Балқаш» сөзін «алқаш» деп, ал «бақа» сөзін «тақан» деп таныған. Wang және әріптестері [10] толассыз (end-to-end) жүйелердің дәстүрлі HMM-GMM архитектураларымен салыстырғанда тиімді екенін

көрсеткен. Бұл тұжырым трансформерлік модельді балалар сөйлеуіне бейімдеудің теориялық негіздерінің бірі болып табылады.

Д. Р. Рахимова мен Ж. Дүйсенбекқызының зерттеулерінде [11, 12] модельді қосымша оқыту оның нәтижелілігін арттыратыны көрсетілген. Кейінгі зерттеулерде деректер жиынын әртүрлі жас топтарына жататын жаңа сөйлеу үлгілерімен толықтыру, Whisper моделінің ірі нұсқаларын сынау, сондай-ақ деректерді жасанды көбейту (data augmentation) және білімді дистилляциялау (knowledge distillation) әдістерін қолдану жоспарланады.

Қорытынды

Зерттеу нәтижелері Whisper-tiny моделін қазақ тіліндегі балалар сөйлеуін тануға бейімдеуде fine-tuning әдісінің тиімді екенін көрсетті. Бірінші кезеңде 235 үлгімен оқытылған модельдің WER көрсеткіші 0,80, ал Accuracy көрсеткіші 20,43% болды. Екінші кезеңде 1936 үлгімен қосымша оқытудан кейін WER 0,42-ге дейін төмендеп, Accuracy 57,95%-ға дейін артты. Бірінші кезеңмен салыстырғанда WER 47,5%-ға төмендеп, Accuracy 183,7%-ға өсті.

Фонемалар деңгейіндегі талдау алмастыру, түсіріп қалдыру және кірістіру қателерінің фонема топтарына қарай өзгеретінін көрсетті. Ең жоғары қате үлесі аффрикаттар мен фрикативтерде байқалды. Алынған нәтижелер ресурсы шектеулі тілдерге бейімделген ASR жүйелерін әзірлеуге негіз болады және мұндай жүйелерді білім беру саласында қолдану мүмкіндігін көрсетеді.

Пайдаланылған дереккөздердің тізімі

- [1] Bai, Z.; Zhang, X.-L. Speaker recognition based on deep learning: An overview. *Neural Networks*, 2021, 140, 65–99. <https://doi.org/10.48550/arXiv.2012.00931>
- [2] Jain, R.; Barcowski, A.; Yiwere, M.; Corcoran, P.; Cucu, H. Adaptation of Whisper models to child speech recognition. *arXiv preprint*, 2023. arXiv:2307.13008.
- [3] Gulati, A.; Qin, J.; Chiu, C.C.; Parmar, N.; Zhang, Y.; Yu, J.; Han, W.; Wang, S.; Zhang, Z.; Wu, Y. Conformer: Convolution-augmented transformer for speech recognition. *Proceedings of Interspeech 2020*, 2020, pp. 5036–5040. <https://doi.org/10.21437/Interspeech.2020-3015>
- [4] Rakhimova, D.; Duisenbekkyzy, Zh.; Karibayeva, A.; Eşref, A.; Ilessova, B. Improving the Voice Recognition System for Children in Kazakh through Additional Training (Fine-Tuning). *Academic Scientific Journal of Computer Science*, 2026, No. 2, pp. 317–334. <https://doi.org/10.32014/2026.2518-1726.441>
- [5] Gretter, R.; Matassoni, M.; Falavigna, D.; Misra, A.; Leong, C.W.; Knill, K.; Wang, L. ETLT 2021: Shared task on automatic speech recognition for non-native children's speech. *Proceedings of Interspeech 2021*, 2021.
- [6] Rusu, A.A.; Rabinowitz, N.C.; Desjardins, G.; et al. Progressive Neural Networks. *arXiv preprint*, 2016. <https://arxiv.org/abs/1606.04671>
- [7] Wang, D.; Wang, X.; Lv, S. An overview of end-to-end automatic speech recognition. *Symmetry*, 2019, 11(8), 1018. <https://doi.org/10.3390/sym11081018>
- [8] Xiong, W.; Wu, L.; Allewa, F.; Droppo, J.; Huang, X.; Stolcke, A. The Microsoft 2017 conversational speech recognition system. *Proceedings of ICASSP 2018*, 2018, pp. 5934–5938. <https://doi.org/10.48550/arXiv.1708.06073>
- [9] Bengio, Y.; Goodfellow, I.; Courville, A. *Deep Learning*. MIT Press, 2016. <https://www.deeplearningbook.org/>
- [10] Koo, J.; Kim, H.; Lee, S. Fine-tuning Whisper for low-resource languages. *arXiv preprint*, 2023. <https://arxiv.org/abs/2303.00752>
- [11] 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, 15(16), 8989. <https://doi.org/10.3390/app15168989>
- [12] Duisenbekkyzy, Zh.; Rakhimova, D. Research on the Impact of Social Networks on the Development of Bilingualism in Kazakhstan. *Communications in Computer and Information Science*, 2024, 2211, pp. 106–115. https://doi.org/10.1007/978-3-031-49739-5_9