АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР
ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ
INFORMATION AND COMMUNICATION TECHNOLOGIES

АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР
ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ
INFORMATION AND COMMUNICATION TECHNOLOGIES

DOI 10.51885/1561-4212_2025_2_163
МҒТАА 28.23.15

**Т. Құрметқан[1,2], Ө.Ж. Мамырбаев[2],**
[1]әл-Фараби атындағы Қазақ Ұлттық Университеті, Алматы қ., Қазақстан
*E-mail: turdybek.narat86@gmail.com**
[2]Ақпараттық және есептеуіш технологиялар институты, Алматы қ., Қазақстан
*E-mail: morkenj@mail.ru*

# ҚАЗАҚ СӨЙЛЕУЛЕРІН ТАНУДА TRANSFORMER МОДЕЛІНІҢ ЖЕТІЛДІРГЕН ТҮРЛЕРІН ҚОЛДАНУ ЕРЕКШЕЛІКТЕРІ

# ОСОБЕННОСТИ ИСПОЛЬЗОВАНИЯ РАСШИРЕННЫХ ФОРМ МОДЕЛИ TRANSFORMER В РАСПОЗНАВАНИИ КАЗАХСКОЙ РЕЧИ

# FEATURES OF USING EXTENDED FORMS OF THE TRANSFORMER MODEL IN KAZAKH SPEECH RECOGNITION

***Аңдатпа***. *Сөйлеуді тану технологиялары соңғы жылдары жасанды интеллект және машиналық оқыту әдістері негізінде үлкен жетістіктерге жетті. Бұл технологиялар адамның сөйлеуін автоматты түрде түсініп, мәтінге айналдыруға мүмкіндік береді. Машиналық оқыту, терең оқыту әдістері негізінде сөйлеуді тану саласы көптеген ғалымдардың қызығушылығын арттырды. Қазақ сөйлеулерін тану саласы бойынша отандық және шетелдік ғалымдар зерттеу жүргізіп, түрлі модельдерді сынақтан өткізді. Зерттеу нәтижелері бойынша модельдерін ұсынып, қазақ сөйлеулерін автоматты түрде танитын қолданбаларын енгізді. Соған қарамастан қазақ сөйлеулерін тануда деректер қорының аздығы, қазақ тілінің морфологиялық және синтаксистік ерекшеліктеріне байланысты сөйлеуді тануда қиындықтар орын алады. Осыған байланысты қазақ сөйлеулерін тануда сөйлеуді тану дәлдігін арттыру әлі де өзекті. Осы мақсатта мақалада қазақ сөйлеулерін тануда қолданылған автоматты сөйлеуді тану (ASR) модельдері мен олардың нәтижелеріне шолу жасалды. Қазіргі кездегі қазақ сөйлеулерін тану барысындағы кездесетін негізгі мәселер анықталды. Қазақ сөйлеулерін тану мәселелерін шешуге икемді модельдер талданды. Сөйлеуді тануда қолданылатын Hiformer моделі сипатталды. 400 сағаттан тұратын KazASR қазақ тілінің акустикалық-мәтіндік корпусында Transformer, Hiformer, Conformer модельдері оқытылып, тестілеуден өткізілді. Зерттеу нәтижесі бойынша Hiformer моделі қазақ сөйлеулерін тануда ең жоғарғы нәтиже көрсетті. Аталған модель қазақ тілінің ерекшелігіне тура келетінін дәлелдеді. Нақты айтқанда, 400 сағаттық акустикалық-мәтіндік корпус үшін Conformer моделінде сөз қатесі (wer) 12.9 %-ға дейін төмендесе, Hiformer моделінде сөз қатесі 11.6 %-ға дейін төмендеді.*

***Түйін сөздер:*** *Автоматты сөйлеуді тану, Hiformer, Conformer, Transformer, конволюциялық нейрондық желі, терең оқыту, қазақ сөйлеулерін тану*

***Аннотация.*** *В последние годы технологии распознавания речи значительно продвинулись благодаря методам искусственного интеллекта и машинного обучения. Эти технологии позволяют автоматически воспринимать человеческую речь и преобразовывать её в текст. Область распознавания речи, основанная на методах машинного и глубокого обучения, привлекла значительное внимание исследователей по всему миру. Как отечественные, так и зарубежные ученые провели исследования по распознаванию казахской речи, протестировали различные модели и предложили решения для автоматического распознавания. Несмотря на достигнутый прогресс, распознавание казахской речи по-прежнему сталкивается с рядом сложностей, обусловленных ограниченной доступностью крупномасштабных корпусов данных, а также*

*морфологическими и синтаксическими особенностями казахского языка. В связи с этим повышение точности распознавания остаётся актуальной задачей. В данной работе проведён обзор моделей автоматического распознавания речи (ASR), применяемых для казахской речи, и их результатов. Были выявлены основные проблемы, возникающие в процессе распознавания, а также проанализированы гибкие модели, предлагаемые для их решения. В исследовании рассматривается модель CHiformer в контексте распознавания речи, а также оценивается производительность моделей Transformer, Hiformer и Conformer, обученных на 400-часовом казахском акустико-текстовом корпусе KazASR. Результаты показали, что модель Hiformer достигла наивысшей точности в распознавании казахской речи, что подтверждает её соответствие языковым особенностям казахского языка. В частности, для 400-часового корпуса уровень ошибок слов (WER) модели Conformer снизился до 12,9 %, тогда как у модели Hiformer этот показатель составил 11,6 %.*

***Ключевые слова:*** *автоматическое распознавание речи, преобразователь, Conformer, Hiformer, сверточная нейронная сеть, глубокое обучение, распознавание казахской речи.*

***Annotation.*** *In recent years, speech recognition technologies have significantly advanced due to artificial intelligence and machine learning methods. These technologies enable the automatic understanding of human speech and its conversion into text. The field of speech recognition, driven by machine learning and deep learning techniques, has garnered substantial interest from researchers worldwide. Both domestic and international scientists have conducted studies on Kazakh speech recognition, testing various models and proposing solutions for automatic recognition. Despite these advancements, Kazakh speech recognition still faces challenges due to the limited availability of large-scale datasets and the morphological and syntactic complexity of the Kazakh language. Consequently, improving recognition accuracy remains a relevant research area. This study provides a comprehensive review of the automatic speech recognition (ASR) models applied to Kazakh speech and their performance outcomes. Key challenges in the current Kazakh speech recognition process were identified, and flexible models for addressing these challenges were analyzed. The study examines the CHiformer model in speech recognition and evaluates the performance of Transformer, Hiformer, and Conformer models trained on the 400-hour KazASR Kazakh acoustic-text corpus. The results indicate that the Hiformer model achieved the highest accuracy in recognizing Kazakh speech, demonstrating its suitability for the linguistic characteristics of the Kazakh language. Specifically, for the 400-hour corpus, the word error rate (WER) of the Conformer model was reduced to 12.9%, while the WER of the Hiformer model further decreased to 11.6%.*

***Keywords:*** *automatic speech recognition, converter, Conformer, Hiformer, convolutional neural network, deep learning, Kazakh speech recognition.*

*Кіріспе.* Автоматты сөйлеуді тану жүйелері жасанды интеллект саласының дамуымен бірге қалыптасты. Дәстүрлі сөйлеуді тану модельдері бастапқыда Леуренс Робимер негізін қалаған жасырын Марков желілеріне сүйенді. Одан бөлек сөйлеуді тануға Гаусс қоспалары модельдері, шешім ағаштары, К жақын көршілер әдістері қолданылды (A Vaswani,, et al., 2017). Терең нейрондық желілер шыққанда көптеген ASR архитектуралары пайда болды (Oralbekova D., et al., 2022). Сөйлеуді тану жүйелерін құруда терең оқытудың конволюциялық нейрондық желілері (CNN) (Dauphin Y., et al., 2017), қайталанатын нейрондық желілер (RNN), қарсылықты желілер (GAN) жиі қолданылады. Бұл желілерге қарағанда соңғы кезде трансформер архитектурасы сөйлеуді тануда жоғары көрсеткіштерге қол жеткізді (Gulati A., et al., 2021, Choi H., et al., 2021).

Сөйлеуді тану жүйелері көптеген салаларда, оның ішінде мобильдік қосымшаларда, дауыстық көмекшілерде, медициналық диагностикада, білім беру және құқық қорғау салаларында кеңінен қолданылуда. Алайда, әлемдік медиада ресурсы аз тілдер үшін сөйлеуді тану технологияларын қолдану белгілі бір қиындықтар туғызады.

Зерттеу объектіміз болып отырған қазақ тілі – Қазақстан Республикасының мемлекеттік тілі, Қазақстан, Ресей, Өзбекстан, Қытай, Моңғолия сияқты бірнеше мемлекетте 20 миллионға жуық қолданушысы бар түрік тілдерінің бірі. Трансформер сияқты архитектуралардың пайда болуы қазақ сөйлеулерін тануда жоғары көрсеткішке жеткенімен, қазақ тілі үшін деректер базалары әлі де аз. Қазақ тілінің агглютинативті тілдер тобына жататыны, тілдің үндестік заңы мен дауысты дыбыстардың үйлесуі сөйлеуді тану модельдеріне қосымша қиындық туғызады (Mamyrbayev O., et al., 2021). Бұл мақалада Transformer моделінің жетілдірген түрлерін қазақ тілі корпусында сынақтан өтікізіп, зерттеу нәтижемізді жариялаймыз.

*Әдеби шолу.* Қазақ тілі агглютинативті тілдер қатарына жатады, яғни сөздерге түрлі қосымшалар жалғану арқылы жаңа сөздер мен мағыналар жасалады. Бұл сөйлеуді тану жүйелеріне айтарлықтай қиындық тудырады, себебі жалғанымдар саны өте көп және сөйлемдегі сөздердің морфологиялық құрылымы күрделі. Қазақ сөйлеулерін тануда кездесетін осы мәселелерді шешу жолында көптеген ғалымдар еңбек етіп, түрлі модельдерді зерттеді. «Терең нейрондық желілер арқылы қазақ тілін автоматты тану» атты еңбекте дәстүрлі нейрондық желілерге қарағанда 6 қабаттан тұратын терең нейрондық желі жоғары көрсеткішке жетті (Mamyrbayev O., et al., 2019). Трансферлік оқытуды пайдалана отырып, қазақ сөйлеулерін автоматты тану жүйесін әзірлеуде LSTM және BiLSTM нейрондық желілерді қолданды. Сыртқы модель ретінде орыс тілінің корпусын пайдаланған. Осы арқылы қазақ тіліндегі корпустың жетіспеушілігін азайтқан (Kozhirbayev Zh., & Islamgozhayev T., 2023). Зерттеушілер Wav2vec2.0 моделін жетілдіріп қазақ тіліне қолдану нәтижесіне, wav2vec-F моделін ұсынды (Baevski A., et al., 2020). Трансформерлік және коннекционистік уақытша жіктеу модельдерін бірлесіп қолдану қазақ тілінің тану жүйесінің өнімділігін арттырды (Mamyrbaev O., et al., 2023). Hiformer моделін қазақ сөйлеулерін тануда қолдану әртүрлі деңгей арасындағы байланысты жақсартты (Ө.Ж. Мамырбаев, Т. Құрметқан, 2024). Көптілді автоматты сөйлеуді тану жүйелері де қазақ тілі үшін тиімді нәтиже көрсетті (Bekarystankyzy A., et al., 2024). Көп тілді end-to-end ASR моделін түркі тілдерінің (қазақ, башқұрт, қырғыз, саха, татар) ортақ алфавиттері мен морфологиялық ұқсастықтарын қолдана отырып, ESPnet негізінде түрік тілдерінің көптілді оқыту жүйесі әзірленді. Бұл әдіс қазақ тілі үшін сөз қатесі деңгейін (WER) екі есе, әріп қатесі деңгейін 3 есеге дейін төмендетті. Қазақ тілі үшін жоғары ресурсты тілдерде алдын ала оқытылған модельдерді қолдану нәтижесінде сөз және әріп қателері айтарлықтай азайғаны байқалған (Bekarystankyzy A., et al., 2024). Сонымен қатар, WaveNet-CTC моделін қолдану көп тапсырмалы сөйлеуді тану жүйесі үшін жаңа әдіс ретінде ұсынылды (Narejo K., et al., 2024). Нәтижесінде, қазақ тілінің диалектілеріне арналған жүйелердің дәлдігі артты. Қарастырған зерттеулер қазақ тілін автоматты тануда көптілді және көпміндетті тәсілдердің тиімділігін дәлелдеп, қазақ тілінде сөйлеуді тану жүйесінің өнімділігін арттыруға бағытталған жаңа әдістерді ұсынады. Осы сияқты зерттеулерге қарамастан қазақ сөйлеулерін танудың негізгі мәселелерін зерттеу және оны шешу өзектілігін жоғалтпайды.

*Материалдар мен зерттеу әдістері.* Соңғы жылдары Transformer архитектурасына негізделген модельдер сөйлеуді тану және өңдеу салаларында үлкен жетістіктерге жетті. Transformer модельдерінің ерекшелігі – олардың үлкен көлемді мәтіндермен және күрделі тілдік құрылымдармен тиімді жұмыс істей алуында (Wang Q., et al., 2016). Бұл модельдер әсіресе машиналық аударма, табиғи тілдерді өңдеу және сөйлеуді тану жүйелерінде кеңінен қолданылады. Қазақ тілі үшін де трансформер негізіндегі модельдер үлкен мүмкіндіктерге ие, алайда оларды қолдану барысында белгілі бір ерекшеліктер мен қиындықтар туындауы мүмкін (Mamyrbayev O., et al., 2021).

Transformer моделін алғашқы рет 2017 жылы Google зерттеушілері ұсынған және бұл архитектурада көпқабатты өзіне назар аудару механизмдері (multi-head self-attention) қолданылады. Transformer моделінің негізгі компоненттері:

– Encoder-Decoder архитектурасы: Encoder мәтінді өңдеп, оның ішкі көрінісін жасайды, ал Decoder бұл көріністерді мәтінге немесе сөйлеуге айналдырады.

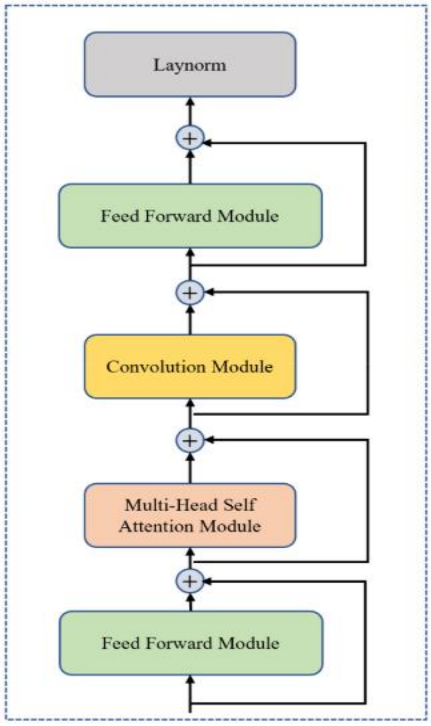
– Self-Attention механизмдері: Бұл механизм әрбір сөзді немесе сөз тіркесін бүкіл контекст бойынша салыстырады және сөйлеудің әрбір бөлігінің мағынасын анықтауға көмектеседі (Nguyen T. & Salazar J., 2019).

– Position Embeddings: Мәтін ішіндегі сөздердің орналасу тәртібін сақтау үшін трансформер модельдері позициялық кодтауды қолданады, себебі нейрондық желілер

мәтіндегі сөздердің ретін тікелей ескеріп жұмыс істемейді.

Қазақ тілінің күрделі морфологиялық құрылымын тану үшін Transformer модельдерінің өзіне назар аудару қабілеті өте тиімді. Transformer модельдері контексті түсіну үшін сөйлемдегі әрбір сөздің мағынасын басқа сөздермен салыстыра отырып, олардың өзара байланысын анықтайды.

Conformer моделінің архитектурасы. Conformer моделі сөйлеуді тану жүйелерінде кеңінен қолданылатын архитектуралардың бірі. Бұл модель 1-суреттегідей төрт негізгі модульден тұрады: алға жылжыту модулі, өзіне назар аудару модулі, конволюция модулі және тағы бір алға жылжыту модулі. Conformer архитектурасы Transformer моделіндегі назар аудару механизмдерін пайдаланады, бұл модельге ұзақ мерзімді тәуелділіктерді тиімді түрде меңгеруге мүмкіндік береді. Өзіне назар аудару модулі сөйлеудің әртүрлі бөліктері арасындағы байланысты анықтауға көмектеседі және контекстік мағынаны тереңірек түсіндіреді. Сонымен қатар, Conformer модуліндегі конволюциялық қабаттар деректердің құрылымдық ерекшеліктерін меңгеруге және қысқа мерзімді тәуелділіктерді үйренуге мүмкіндік береді. Бұл қабаттар сөйлеу сигналдарындағы фонетикалық өзгерістерді тиімді түрде өңдейді (Gulati A., et al., 2021).




**1-сурет.** Conformer моделінің архитектурасы

*Ескерту – (Gulati A., et al., 2021) негізінде құрастырылған*

***Көп басты өзіне назар аудару модулі.*** Conformer моделінде көп басты назар аудару (multi-head attention) механизмі маңызды рөл атқарады. Бұл модуль деректердің ішкі байланыстарын терең түсінуге мүмкіндік береді. Көп басты назар аудару механизмі әртүрлі ақпарат түрлеріне назар аудару арқылы сөйлеудің әр қырын зерттеуге жағдай жасайды (Gulati A., et al., 2021).

***Конволюция модулі.*** Conformer моделіндегі конволюция модулі бірнеше маңызды компоненттерден тұрады: шығыс механизмі, бір өлшемді тереңдік бойынша конволюция

қабаты және партиялық нормализация. Бұл модуль жергілікті ерекшеліктерді тануға және сөйлеу үлгілерін тиімді түрде өңдеуге көмектеседі. Тереңдік бойынша конволюция әр арнадан ерекшеліктерді жинақтап, модельдің ақпаратты дәл өңдеу қабілетін арттырады. Сонымен қатар, партиялық нормализация оқыту процесінің тұрақтылығын қамтамасыз етеді және модельдің жылдам үйренуіне ықпал етеді. Конволюция блоктары 2-суретте көрсетілген (Ө.Ж. Мармырбаев, Т. Құрметқан & R.S. Arslan. 2024).

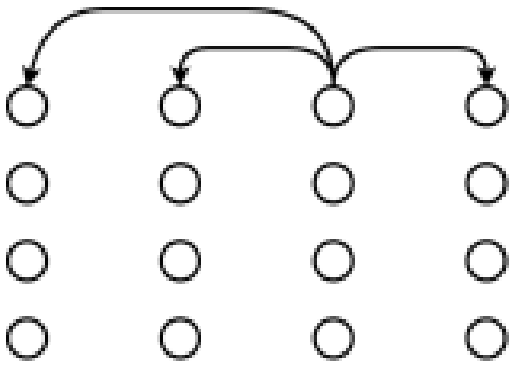

**2-сурет.** Конволюция модулі

*Ескерту – автормен құрастырылған*

Алға жылжыту модулі. Transformer архитектурасында көп басты назар аудару (MHSA) қабатынан кейін алға жылжыту модулі (Feed Forward Module) қолданылады. Бұл модуль екі сызықтық түрлендіруден және олардың арасында бейсызықты белсендіру функциясынан тұрады. Оның басты мақсаты – желінің өңдеу қабілетін күшейту және күрделі ерекшеліктерді үйрену. Қосымша қалдық байланыс қабаттары модельдің оқыту тұрақтылығын қамтамасыз етеді және ақпараттың жоғалуын азайтады. Бұл механизм градиенттердің жойылу мәселесін шешуге де көмектеседі .

Hiformer моделі – сөйлеуді тануда қолданылатын трансформер архитектурасының жетілдірілген нұсқасы. Бұл модель аудио сигналдарын өңдеп, оларды мәтінге айналдыру кезінде жоғары деңгейлі контекстік талдау мен әртүрлі ақпарат көздерін біріктіру қабілетіне ие. Модель негізінен сөйлеуді сегментациялау, дыбыстарды анықтау, фонетикалық ерекшеліктерді тану және сөйлем мағынасын дұрыс түсіну үшін қолданылады (Mamyrbayev O., et al., 2024).

*Hiformer архитектурасы.* Стандартты трансформерде қолданылғандай Hiformer моделінің өзіне назар аудару механизмі 3-суреттегідей бір қабаттағы қадамдардың әрбір жұбы арасындағы корреляцияны түсіру үшін пайдаланылады. Сұраныстар жиынтығы негізінде трансформердің l-ші қабатындағы назар аудару функциясын есептеу 1-формулада бейнеленген. S – сұраныстар, k – кіліттік мәндер, m-сәйкес мәндер *(Xixin Wu , et al.,, 2023)*.



**3-сурет.** Өзіне назар аудару механизмі

*Ескерту – (Xixin Wu., et al., 2023) негізінде құрастырылған*

$$\mathrm{FS}^{(\mathrm{l})} = \mathrm{softmax}(d_k^{-\frac{1}{2}} s^{(l)} k^{(l)} m^{(l)^T}) m^l \quad (1)$$

мұндағы $s^{(l)} \in R^{T*d_k}$, $k^{(l)} \in R^{T*d_k}$, $m^{(l)} \in R^{T*d_k}$, сұраныстар, кілттер және мәндер ретінде қабылданады. Бұл әдіс әр қабаттағы ақпаратты біріктірмей, тек бір деңгейдегі тәуелділіктерді зерттейді.

Көп деңгейлі назар аудару механизмі сұраныстарды, кілттерді және мәндерді алу үшін параллель проекцияларды қолдану арқылы өзіне назар аудару өнімділігін арттырады. Бұл процесс (1) теңдеуде көрсетілгендей назар аудару шығыстарын бөлек есептеуге мүмкіндік береді.

$$\mathrm{FS}_{x\mathrm{h}}^{(\mathrm{l})} = \mathrm{concat}(\mathrm{FS}_1^{(\mathrm{l})}, \ldots, \mathrm{FS}_1^{(\mathrm{l})})\mathrm{W}^{\mathrm{Fs}} \quad (2)$$

Бұл әдіс көп деңгейлі байланыстарды зерттеуге және әртүрлі өлшемдер бойынша өңдеуді жақсартуға мүмкіндік береді.

Қазақ тілінің ерекшеліктерін тану. Қазақ тілі агглютинативті тілдер қатарына жатады, яғни сөздерге жалғаулар мен жұрнақтар қосылып, жаңа сөздер мен мағыналар пайда болады. Сондай-ақ, қазақ тілінде дауыссыз дыбыстардың үндестік заңы бойынша өзгеруі жиі кездеседі. Hiformer моделі қазақ тіліне тән осы ерекшеліктерді тиімді тануға мүмкіндік береді:

– Сөздің өзгермелі формаларын тану: модель сөз түбірін және оған қосылған аффикстерді ажырата отырып, сөздің мағынасын дұрыс түсінеді.

– Үндестік заңы: Hiformer қазақ тіліндегі дыбыстардың бір-біріне ықпал етуін және сөздің айтылу формасын дұрыс талдай алады. Мысалы, сөздің соңғы дыбысы дауысты немесе дауыссыз болса, жалғаудың өзгеруі мүмкін және модель осы фонетикалық заңдылықтарды ескере отырып тану жасайды.

*Нәтижелер және оны талқылау.* Conformer және Hiformer моделдері арқылы қазақ сөйлеулерін тану үшін ҚР ҒжЖБМ ҒК «Ақпараттық және есептеу технологиялары институты» РМК құрастырған қазақ тілінің 400 сағаттық сөйлеу корпусы пайдаланылды. Модельдерді үйрету үшін сөйлеу корпусын 200 сағат «таза» сөйлеу және 200 сағат спонтанды телефонмен сөйлеу корпусынан тұрды. Корпуста дыбыстық файлдар оқу және сынақ бөліктеріне бөлінді. Сөйлеу корпусы әртүрлі жастағы және жыныстағы 380 қазақ тілді диктордың оқуындағы жазбалардан, көркем кітаптарды оқу деректерінен және жаңалықтар мазмұнын қамтыды.

Телефондағы сөйлесушілердің жазбалары арнайы зерттеу жұмыстарымыз үшін телекоммуникация компаниялардан сұрап алынған. Телефонмен сөйлесулерді транскрипциялау мәтіндерді құрастырудың әзірленген әдістемесі негізінде жүзеге асырылды, өйткені бұл сөйлеу стихиялы және шет тіліндегі ақпараттың араласуы мен түрлі шулардан дыбыстың анық болмауы мүмкін *(Mamyrbayev O., et al., 2024)*.

Conformer моделінде кодер үшін 12 конформер блогы және декодер үшін 6 трансформер қабаты бар. Блок өлшемі 2048, ал өзіне назар аудару модульдерінде 4 бас бар. Hiformer моделін құру үшін Conformer-ге жоғары назар аудару механизмін енгіземіз. Hiformer моделін конфигурациясы жоғары назар аудару модулін енгізуді және назар аударуды жоғалту жылдамдығын 0-ден 0,1-ге дейін арттырамыз. Жоғары назар аудару тәжірибелерде кеңейту коэффициенті 1 болатын екі тарихи қабатты қарастырады. Hiformer моделінде де кодер үшін 12, декодер үшін 6 трансформер қабаты қолданылды. Блок өлшемі 2048, ал өзіне назар аудару модульдерінде 4 бас бар. Бұл модельдер екі түрлі корпус үшін де бірдей қолданылды. Модельдерің көрсеткіштері үшін символдық қатені (CER) және сөз қатесін (CER) анықтау арқылы бағаланды.

1-кестеде көрсетілгендей Conformer моделі Transformer және CNN модельдерінің көрсеткіштеріне қарағанда жақсы көрсеткішке жеткенін байқауға болады. Сынақ таза

жинақтағы Conformer моделі Transformer моделінен WER 1,3 % төмен екенін және CER 0,9 %, CNN моделінен қателік 3 пайызға жуық төмен нәтиже көрсетті. Hiformer моделі Conformer моделінен WER 1,3 % төмен екенін және CER 0,6 % төмен нәтиже көрсеткенін көреміз.

**1-кесте.** Студияда жазылған сөйлеу корпусына модельдерді қолдану нәтижесі

| Модель | CER | WER |
|---|---|---|
| CNN | 9.8 | 17.5 |
| Transformer | 7.9 | 14.2 |
| Conformer | 7.1 | 12.9 |
| Hiformer | 6.5 | 11.6 |
| *Ескерту – автормен құрастырылған* | | |

2-кестеде көрсетілгендей телефон сөйлесулері арқылы жазылған сөйлеу корпусына да Conformer моделі Transformer және CNN модельдерінен анақұрлым жоғары нәтиже көрсетсе, жоғары назар аудару механизмі бар Hiformer моделі Transformer және Conformer модельдерімен салыстырғанда тиімді екенін көрсетеді.

**2-кесте.** Телефон сөйлесулері арқылы жазылған сөйлеу корпусына модельдерді қолданудың нәтижесі

| Модель | CER | WER |
|---|---|---|
| CNN | 17.4 | 24.3 |
| Transformer | 14.8 | 21.1 |
| Conformer | 11.6 | 17.8 |
| Hiformer | 9.5 | 15.2 |
| *Ескерту – автормен құрастырылған* | | |

Зерттеу нәтижелерімізге қарай отырып, Hiformer моделінің қазақ сөйлеулерін тануда бірқатар артықшылықтары анықталды:

Трансформердің контекстік мүмкіндіктері: модельдің трансформер архитектурасы сөйлеу барысында әрбір сөздің контекстін талдай отырып, сөздер арасындағы байланысты жақсырақ анықтайды. Бұл әсіресе көпмағыналы сөздерді дұрыс түсінуге көмектеседі.

Ұзын және күрделі сөйлемдерді тану: Hiformer моделі ұзақ сөйлемдер мен күрделі синтаксистік құрылымдарды тиімді өңдейді. Қазақ тілінде бір сөйлемде бірнеше жалғау мен жұрнақ қолданыла алады, сондықтан модель бұл құрылымдарды дұрыс талдап, сөйлемнің толық мағынасын түсінеді.

Табиғи шу мен акцентті тану: телефон сөйлесулері арқылы жазылған сөйлеу корпусына молдельдерді қолану нәтижесіне қарай отырып, модель аудио сапасының төмендеуі немесе сөйлеушінің акценті болған жағдайда да сөйлеуді тануға бейімделеді. Мысалы, түрлі аймақтардағы диалектілер мен акценттерді ескеріп, негізгі сөйлеу мазмұнын нақтылап береді.

*Қорытынды.* Автоматты сөйлеуді тану жүйелерін құруда терең оқытудың конволюциялық нейрондық желілері (CNN) мен трансформер нейрондық желілері жоғары көрсеткіштерге ие. Деседе бұл екі модельдің өзіндік артықшылықтарымен бірге

кемшіліктері де болды. Осының негізінде трансформер архитектурасының жетілдірген жаңа модельдері пайда болды. Зерттеу жұмысымызда трансформер және конволюциялық нейрондық желілердің артықшылықтарын өзара сәйкестірген Conformer моделі мен деңгейаралық және кросс-кодерлік иерархиялық ақпаратты қарастыратын иерархиялық назар аудару механизміне ие Hiformer модельдерін Қазақ тілі корпусында сынақтан өткізіп, тиімділігін көрдік. Conformer моделі қазақ сөйлеулерін тануға CNN және Transformer модельдерінен жоғары көрсеткіш көрсетсе, ал Hiformer моделі ең жоғарғы нәтиже беріп, аталған модельдер қазақ тілінің морфологиялық және синтаксистік ерекшеліктеріне байланысты мәселелерді шешуге бейім екенін көрсетті. Аталған модельдер қазақ тілінде сөйлеуді тану қосымшалары үшін қолдануға болатыны дәлелденді. Алдағы уақытта сөйлеу корпусын арттыру арқылы Hiformer моделінің нәтижесін одан да жоғарылатуды мақсат етеміз. Сондай-ақ, Conformer моделін қазақ тіліне қайта бейімдеу (Fine Tuning) әдістерін қарастыратын боламыз.

Алғыс. Бұл жұмыс Қазақстан Республикасы Ғылым және жоғары білім министрлігі тарапынан BR24993001 «Қазақ тілі мен технологиялық прогресті қолдау үшін үлкен тілдік моделін (LLM) құру» жобасы негізінде қолдау тапты.

Әдебиеттер тізімі

A. Vaswani, N., Shazeer, N., Parmar, J., Uszkoreit, L., Jones, A., Gomez, Ł., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In Proceedings of NeurIPS. Advances in Neural Information Processing Systems 30 (NIPS 2017) . DOI: https://doi.org/10.48550/arXiv.1706.03762

Oralbekova, D., Mamyrbayev, O., Othman, M., Alimhan, K., Zhumazhanov, B., & Nuranbayeva, B. (2022). Development of CRF and CTC Based End-To-End Kazakh Speech Recognition System. In Nguyen, N.T., Tran, T.K., Tukayev, U., Hong, T.P., Trawiński, B., & Szczerbicki, E. (Eds.), Intelligent Information and Database Systems. ACIIDS 2022. Lecture Notes in Computer Science (Vol. 13757). Springer, Cham.

Dauphin, Y. N., Fan, A., Auli, M., & Grangier, D. (2017). Language modeling with gated convolutional networks. In Proceedings of the 34th International Conference on Machine Learning (Vol. 70, pp. 933–941). JMLR.org.

Choi, H., Lee, J., Kim, W., Lee, J., Heo, H., & Lee, K. (2021). Neural analysis and synthesis: Reconstructing speech from self-supervised representations. In Proceedings of NeurIPS.

Mamyrbayev, O., Oralbekova, D., Alimhan, K., Turdalykyzy, T., & Othman, M. (2022). A study of transformer-based end-to-end speech recognition system for Kazakh language. In Proceedings of Springer Nature.

Gulati, A., Qin, J., Chiu, C.-C., Parmar, N., Zhang, Y., Yu, J., Han, W., Wang, S., Zhang, Z., Wu, Y., & Pang, R. (2020). Conformer: Convolution-augmented Transformer for Speech Recognition. Submitted to Interspeech 2020.

Mamyrbayev, O., Turdalyuly, M., Mekebayev, N., Alimhan, K., Kydyrbekova, A., & Turdalykyzy, T. (2019). Automatic Recognition of Kazakh Speech Using Deep Neural Networks. In ACIIDS, Lecture Notes in Computer Science (pp. 465–474).

Kozhirbayev, Zh., & Islamgozhayev, T. (2023). Cascade Speech Translation for the Kazakh Language. Applied Sciences.

Baevski, A., Zhou, Y., Mohamed, A., & Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. Advances in Neural Information Processing Systems, 33, 12449–12460.

Mamyrbayev, O., Pavlov, S., Bekarystankyzy, A., Oralbekova, D., Zhumazhanov, B., Azarova, L., Mussayeva, D., Koval, T., Gromaszek, K., Issimov, N., & Shiyapov, K. (2023). Neurorecognition visualization in multitask end-to-end speech. In Proceedings of SPIE, 12985, Optical Fibers and Their Applications 2023.

Mamyrbayev, O., Kurmetkan, T., Oralbekova, D., & Zhumazhan, N. (2024). A Study of Kazakh Speech Recognition in Hiformer Model. In Recent Challenges in Intelligent Information and Database Systems. ACIIDS 2024 (pp. 330–340).

Bekarystankyzy, A., Mamyrbayev, O., & Anarbekova, T. (2024). Integrated End-to-End Automatic Speech Recognition for Agglutinative Languages. ACM Transactions on Asian and Low-Resource Language Information Processing (TALIP).

Bekarystankyzy, A., Mamyrbayev, O., Mendes, M., Fazylzhanova, A., & Assam, M. (2024). Multilingual end-to-end

ASR for low-resource Turkic languages with common alphabets. Scientific Reports, 14, 13835.

Xixin Wu, Hui Lu, Kun Li, Zhiyong Wu, Xunying Liu, Helen Meng (2023). Hiformer: Sequence Modeling Networks with Hierarchical Attention Mechanisms. IEEE/ACM Transactions on Audio, Speech, and Language Processing. DOI: https://doi.org /10.1109/TASLP.2023.3313428

Chan, W., Jaitly, N., Le, Q., & Vinyals, O. (2016). Listen, attend and spell: A neural network for large vocabulary conversational speech recognition. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016.

Wang, Q., Li, B., Xiao, T., Zhu, J., Li, C., Wong, D. F., & Chao, L. S. (2019). Learning deep transformer models for machine translation. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 1810–1822). Association for Computational Linguistics.

Narejo, K. R., Zan, H., Oralbekova, D., Dharmani, K. P., Orken, M., & Mukhsina, K. (2024). Enhancing Emoji-Based Sentiment Classification in Urdu Tweets: Fusion Strategies With Multilingual BERT and Emoji Embeddings. IEEE Access, 12, 126587-126600. DOI: 10.1109/ACCESS.2024.3446897.

Mamyrbayev, O., Oralbekova, D., Kydyrbekova, A., Turdalykyzy, T., & Bekarystankyzy, A. (2021). End-to-End Model Based on RNN-T for Kazakh Speech Recognition. In 2021 3rd International Conference on Computer Communication and the Internet (ICCCI).

Ө.Ж. Мармырбаев, *Т. Құрметқан, R.S. Arslan. "Қазақ сөйлеулерін тануда Conformer моделін қолдану." Қазақстан Республикасы Ұлттық инженерлік академиясының хабаршысы, 2024, № 3 (93), 144-154-беттер. DOI: https://doi.org/10.47533/2024.1606-146X.57.

Ө.Ж. Мармырбаев, *Т. Құрметқан. "Қазақ сөйлеулерін тануға Hiformer моделін қолдануды талдау" Қазақстан Республикасы Ұлттық инженерлік академиясының хабаршысы, 2024, № 4 (94), 290-300-беттер. DOI: https://doi.org/10.47533/2024.1606-146X.024.

Nguyen, T. & Salazar, J. (2019). Transformers without tears: Improving the normalization of self-attention. arXiv preprint arXiv:1910.05895.

Information about authors

**Turdybek Kurmetkan** – PhD student, Institute of Information and Computational Technologies, Al-Farabi Kazakh National University, Almaty, Kazakhstan. E-mail: turdybek.narat86@gmail.com, Tel: 8771458887153

**Mamyrbayev Orken Zhumazhanovich** – PhD, Professor, Institute of Information and Computational Technologies, Almaty, Kazakhstan. E-mail: morkenj@mail.ru, Tel: +7 777 366 2727