



DOI 10.51885/1561-4212_2025_4_186

FTAXP 20.23.15

МАШИНАЛЫҚ ОҚЫТУДЫ ҚОЛДАНЫП ФОНЕМА (ДАУЫСТЫ ДЫБЫСТЫ) ТАҢУ

РАСПОЗНАВАНИЕ ФОНЕМ (ГЛАСНЫХ ЗВУКОВ) С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ

PHONEME (VOWEL SOUND) RECOGNITION USING MACHINE LEARNING

Н.О. Мекебаев ^{id}¹, Д.К. Даркенбаев ^{id}^{2*}, Г.З. Зиятбекова ^{id}³, Ж.А. Орынтаева ^{id}¹,
Ғ.М. Сулейменов ^{id}¹

¹Қазақ ұлттық қыздар педагогикалық университеті, Алматы қ., Қазақстан

²Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан

³Алматы технологиялық университеті, Алматы қ., Қазақстан

*Жауапты автор: Даркенбаев Даурен Кадырович, e-mail: dauren.kadyrovich@gmail.com

Түйінді сөздер:

фонема, машиналық
оқыту, дауысты
дыбыстарды таңу,
MFCC, ANN, CNN, RNN,
сөйлеуді таңу.

ТҮЙІНДЕМЕ

Қазіргі таңда сөйлеуді автоматты түрде таңу (ASR) жүйелері жасанды интеллект пен машиналық оқытудың дамуына байланысты өзекті бағытқа айналды. Қазақ тіліндегі фонемаларды, әсіресе дауысты дыбыстарды таңу саласының жеткілікті дамымауы және цифрлық ресурстардың аздығы зерттеу өзектілігін айқындайды. Зерттеу мақсаты – қазақ тіліндегі дауысты дыбыстарды жоғары дәлдікпен танитын машиналық оқыту үлгісін жасау. Бұл үшін сөйлеу сигналдарын алдын ала өңдеу, MFCC арқылы ерекшелік белгілерін алу және Random Forest, SVM, ANN алгоритмдерін салыстыру жүргізілді. Нәтижесінде ANN моделі ең жоғары дәлдік көрсетті. Алынған нәтижелер қазақ тіліндегі сөйлеуді таңу жүйелерінің сапасын арттыруға және болашақта дауыс биометриясы мен дыбыстық интерфейстерде қолдануға мүмкіндік береді.

Ключевые слова:

фонема, машинное
обучение, распознавание
гласных звуков, MFCC,
ANN, CNN, RNN,
распознавание речи

АННОТАЦИЯ

В настоящее время системы автоматического распознавания речи (ASR) приобретают особую актуальность благодаря развитию искусственного интеллекта и машинного обучения. Актуальность исследования определяется недостаточным развитием систем распознавания фонем казахского языка, особенно гласных звуков, и нехваткой цифровых языковых ресурсов. Цель исследования – разработка модели машинного обучения для высокоточного распознавания гласных звуков казахского языка. Для этого проведены предварительная обработка речевых сигналов, извлечение признаков с помощью MFCC и сравнение алгоритмов Random Forest, SVM и ANN. В результате модель ANN показала наибольшую точность. Полученные результаты позволяют повысить качество систем распознавания речи на казахском языке и могут быть применены в области голосовой биометрии и звуковых интерфейсов.



Keywords:

phoneme, machine learning, vowel recognition, MFCC, ANN, CNN, RNN, speech recognition.

ABSTRACT

Currently, Automatic Speech Recognition (ASR) systems have become increasingly relevant due to advances in artificial intelligence and machine learning. The relevance of this study lies in the insufficient development of phoneme recognition systems for the Kazakh language, particularly for vowel sounds, and the limited availability of digital linguistic resources. The aim of the research is to develop a machine learning model capable of accurately recognizing Kazakh vowel sounds. The study involved preprocessing of speech signals, feature extraction using MFCC, and comparison of the Random Forest, SVM, and ANN algorithms. As a result, the ANN model achieved the highest accuracy. The findings can enhance the quality of Kazakh speech recognition systems and be applied in voice biometrics and sound interface technologies.

КІРІСПЕ

Адамдар бір-бірімен қарым-қатынас жасауда ең алдымен тіл арқылы сөйлесу қабілетін пайдаланады. Сөйлеуді тану — сөйлейтін адамның айтқан сөздерін түсіну және оларды мағыналық тұрғыдан интерпретациялау процесі болып табылады. Автоматты сөйлеуді тану (Automatic Speech Recognition — ASR) жүйелері адамның сөйлеуін автоматты түрде тану және оны мәтінге түрлендіру міндетін орындайды. Бұл бағыт соңғы онжылдықтарда жасанды интеллект, нейрондық желілер және сигналдарды өңдеу салаларымен тығыз байланысты өзекті зерттеу нысанына айналды.

ASR саласындағы алғашқы зерттеулерде Гаусс қоспасының үлгілері (GMM), динамикалық уақытты бұзу (DTW) алгоритмі және жасырын Марков үлгілері (HMM) сияқты дәстүрлі әдістер кеңінен қолданылды. Бұл тәсілдер сөйлеу сигналынан ерекшелік белгілерін қолмен шығарып, статистикалық сәйкестендіру арқылы тану нәтижесін анықтауға негізделген болатын. Алайда мұндай модельдер күрделі тілдік құрылымдар мен фонетикалық ерекшеліктерді толық қамти алмады.

Соңғы жылдары нейрондық желілер мен терең машиналық оқыту әдістерінің дамуы автоматты сөйлеуді тану жүйелерінің сапасын айтарлықтай жақсартуға мүмкіндік берді. ASR технологиялары бүгінде мобильді құрылғылардағы дауыс арқылы басқару, виртуалды көмекшілер, автоматты субтитрлеу және аударма жүйелері сияқты көптеген қолданбалы салаларда кеңінен енгізілуде (Yu, D., & Deng, L. (2016), Yuhua, B. P., et al. (1989), Li, J., Deng, L., Naeb-Umbach, R., & Gong, Y. (2016)). Дегенмен, аглютинативті құрылымға ие қазақ тілі үшін дыбыстық тану мәселесі ерекше күрделі болып табылады, себебі тілдің морфологиялық және фонетикалық заңдылықтары өзгеше, ал дыбыстардың артикуляциялық және форманттық сипаттамалары бір-бірінен айтарлықтай ерекшеленеді.

Сөйлеуді тануға әсер ететін негізгі факторларға айтылу екпіні, интонация, фондық шу, жаңғырық, сөйлеушінің жынысы мен жас ерекшеліктері жатады. Осы факторлардың көптүрлілігі сөйлеу сигналдарын алдын ала өңдеу мен ерекшеліктерді дәл анықтау қажеттілігін тудырады.

Қазақ тілінің дыбыстық жүйесін автоматты түрде талдау мәселесі — фонетикалық зерттеулер мен сөйлеуді тану саласындағы өзекті бағыттардың бірі. Қазақ тілі аглютинативті тіл болғандықтан, сөз құрамындағы дауысты дыбыстардың жуан-жіңішкелігі тіл үндестігі заңына бағынады. Осы заңдылықтарды компьютерлік бағдарлама арқылы талдау тілдің цифрлық моделін құруға негіз болады. Бұл зерттеу жұмысының мақсаты — қазақ тіліндегі сөйлемдерде дауысты дыбыстарды анықтайтын және олардың статистикалық сипаттамасын шығаратын қарапайым бағдарламалық құрал әзірлеу.

Қазіргі таңда сөйлеуді автоматты тану саласында машиналық оқыту мен жасанды нейрондық желілерді қолдану ең тиімді бағыттардың бірі болып саналады. Фонема — тілдің ең кіші дыбыстық бірлігі, сондықтан фонемаларды дәл және сенімді тану жалпы сөйлеуді тану жүйесінің табыстылығын айқындайды. Қазақ тіліндегі фонемаларды автоматты түрде тану өзекті ғылыми міндеттердің бірі болып табылады, себебі бұл тілдің дыбыстық жүйесі бай және үндестік заңы күрделі.

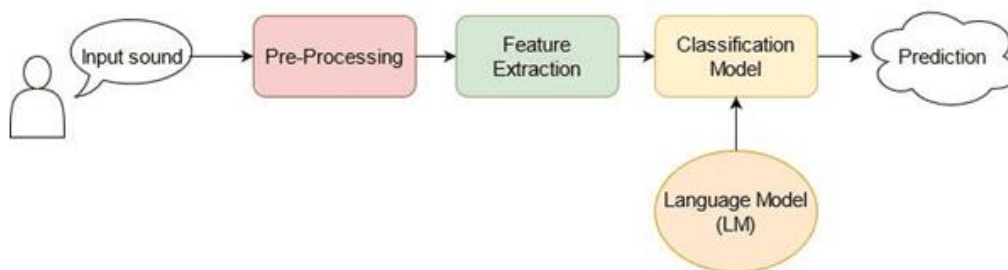
Машиналық оқыту әдістерін, әсіресе жасанды нейрондық желілердің (ANN, CNN, RNN) архитектураларын пайдалану арқылы фонемаларды танудың дәлдігін арттыруға және қазақ тіліне бейімделген интеллектуалды жүйелерді құруға мүмкіндік туады. Мұндай зерттеулердің нәтижелері қазақ тіліндегі сөйлеуді тануға арналған ақпараттық жүйелердің, виртуалды көмекшілердің, білім беру және көру қабілеті шектеулі адамдарға арналған технологиялардың дамуына елеулі үлес қосады.

Осыған орай, «Машиналық оқытуды қолданып фонема (дауысты дыбысты) тану» тақырыбы қазіргі заманғы сөйлеуді өңдеу және жасанды интеллект саласындағы ғылыми-техникалық прогрестің маңызды бағыттарының бірі болып табылады. Бұл зерттеу қазақ тілінің дыбыстық жүйесін цифрлық өңдеу саласында жаңа мүмкіндіктер ашып, ұлттық тілдің ақпараттық кеңістіктегі орнын кеңейтуге бағытталған өзекті ғылыми жұмыс болып саналады (Juang, B.-H., & Rabiner, L. R. (2005)).

Осы зерттеуде қазақ тіліндегі фонемаларды танудың ғылыми негіздері мен машиналық оқыту әдістері жүйелі түрде қарастырылады. Бірінші бөлімде сөйлеуді танудың теориялық және фонетикалық аспектілері талданады, екінші бөлімде MFCC арқылы ерекшелік белгілерін алу әдістемесі баяндалады, үшінші бөлімде нейрондық желілердің архитектурасы мен модельдері сипатталады, ал соңғы бөлімде зерттеу нәтижелері мен олардың талдауы ұсынылады.

ЗЕРТТЕУ МАТЕРИАЛДАРЫ МЕН ӘДІСТЕРІ

Сөйлеуді танудың күрделілігіне қарамастан, ол әрдайым көптеген салаларда сөзсіз рөл атқарады. Бұл көптеген адамдарға қалаған мазмұнына оңай қол жеткізуге көмектеседі. Сөйлеуді тану – көптеген маңызды қолданбалары бар дамып келе жатқан сала. Сөйлеуді тану бойынша зерттеулер жалғасады, сонымен қатар маңызды практикалық қосымшалар жасалады деп болжау оңай. Сөйлеуді тану негізгі дамып келе жатқан сала болса да, дәлдікке қол жеткізу осы саладағы басты мәселе болып табылады. Автоматты сөйлеуді тану (ASR) жалпы ағынын 1-ші суретте көрсетілген.



1-сурет. Автоматты сөйлеуді тану

Ескерту – авторлармен құрастырылған

Табиғи тілді өңдеудегі фонемаларды тану (NLP) сөйлеуді тану жүйелерін дамытудағы негізгі компонент болып табылады. Фонемалар – бір сөзді екінші сөзден ажырата алатын тілдегі ең кіші дыбыс бірліктері. Бұл фонемаларды дәл тану ауызекі тілді мәтінге айналдыру үшін маңызды. Фонеманы тану дауыспен белсендірілген жүйелер, тіл



аудармасы және т.б. сияқты қолданбаларда пайдалы. Бұл мақалада біз фонема (дауысты дыбысты тану) туралы зерттеп, машиналық оқыту арқылы фонемаларды (дауысты дыбысты) тануды жүзеге асырамыз.

Фонемалар – сөздерді бір-бірінен ажырата алатын тілдегі ең кіші дыбыс бірліктері. Фонетикада олар екі негізгі түрге бөлінеді: *дауыссыз дыбыстар* және *дауысты дыбыстар*. Әрбір түр тілде өзіндік рөл атқарады және өзіндік сипаттамалары мен вариацияларына ие.

Дауысты дыбыстар – сөйлеу кезінде ауыз қуысында кедергісіз, ашық дауыс жолы арқылы жасалатын дыбыстар. Олар дауыс желбезектерінің тербелуі нәтижесінде пайда болады және дауыс басымдығымен ерекшеленеді.

Дауысты дыбыстардың негізгі сипаттары:

1. Тілдің қатысына қарай:

- Жуан дауыстылар – *a, o, y, ы*
- Жіңішке дауыстылар – *ә, ө, ү, і, е*

2. Еріннің қатысына қарай:

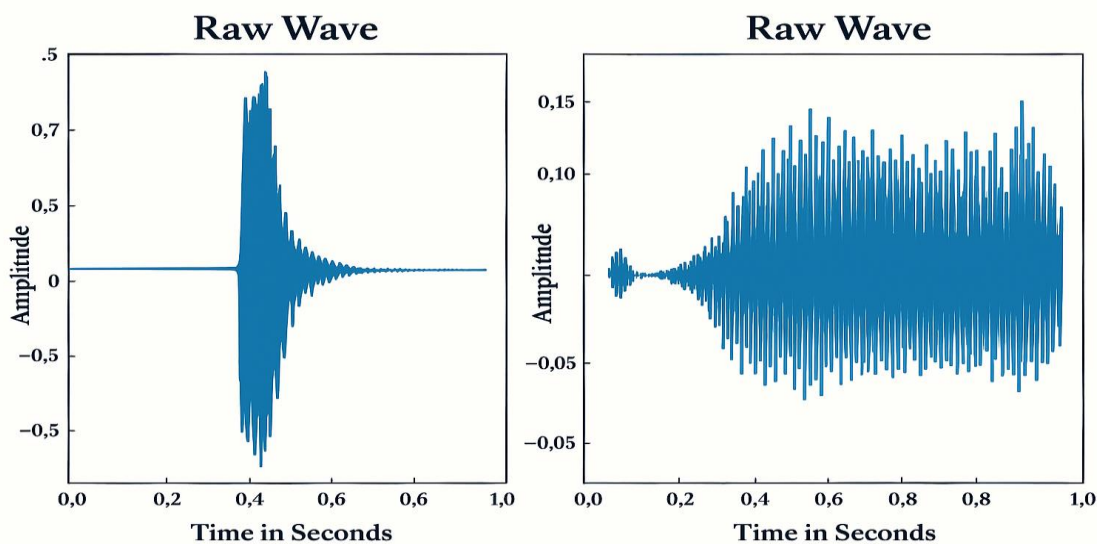
- Еріндік дауыстылар – *o, ө, y, ү* (ерін дөңгеленеді)
- Ерінсіз дауыстылар – *a, ә, e, ы, і* (ерін тартылып айтылады)

3. Тілдің көтеріңкілігіне қарай:

- Жоғары көтеріңкі – *ы, і, ү, ү*
- Орта көтеріңкі – *e, o, ө*
- Төмен көтеріңкі – *a, ә*

Дауысты дыбыстардың басты қызметі – *буын құрау*. Яғни, қазақ тіліндегі әр буында міндетті түрде бір дауысты дыбыс болады.

Дауысты дыбыстар – сөйлеу тілінің ең әуезді, үндік сипаты басым элементтері болып табылады, олар сөздің үндестігін және дыбыстық жүйесін айқындайды. 2-суретте дыбыстың аудио айтылған сигналы көрсетілген.



2-сурет. Дауысты дыбыстың сигналы

Ескерту – авторлармен құрастырылған

Фонемаларды тану әдістері

Жасырын Марков үлгілері (Hidden Markov Models, HMM)

Жасырын Марков үлгілері сөйлеу сигналының уақыт бойынша өзгерісін ықтималдық тұрғыда сипаттайды. Бұл үлгіде әрбір фонема – белгілі бір күй (state) ретінде

қарастырылып, олардың арасындағы ауысулар ықтималдық матрицасы арқылы сипатталады. НММ әдістері ерте кезеңдегі сөйлеуді тану жүйелерінде негізгі құрал ретінде пайдаланылды. Олар фонемалардың ұзындығы мен айтылу динамикасын есепке алып, уақыттық тізбектерді тиімді модельдейді.

Машиналық оқыту әдістері

Машиналық оқыту (ML) термині компьютерлерге деректерге негізделген модельдік бағдарламалауды автоматтандыруға және статистикалық маңызды деректердегі үлгілерді жүйелі түрде анықтауды қолданатын үлгілерді әзірлеуге мүмкіндік беретін әдістемелер немесе алгоритмдер жинағын білдіреді (Povey, D., et al. (2016)). Машиналық оқытудың м түрлері бақыланатын, бақыланбайтын оқыту. Қазақ фонемаларын тану жүйесін оқыту және енгізу үшін осы мақалада қолданылатын қолданбалы бақыланатын ML алгоритмдерінің негізгі түсініктері қысқаша сипатталған.

Кездейсоқ орман (Random Forest)

Кездейсоқ орман бірнеше шешім ағаштарын біріктіретін ансамбльдік оқыту моделі. Әр ағаш мәліметтердің кездейсоқ таңдамасында үйретіліп, соңғы нәтиже көпшілік дауыс немесе орташа мән арқылы анықталады. Фонемаларды тануда кездейсоқ орман әдісі сенімді болжамдар береді, себебі ол жеке ағаштардың артық үйренуін (overfitting) азайтады және деректердегі шудың әсерін төмендетеді.

Терең оқыту әдістері

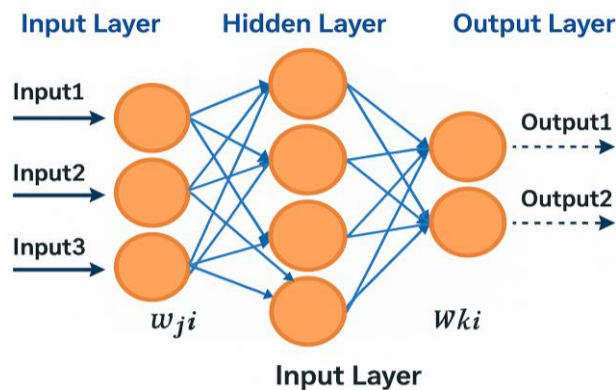
Көп қабатты нейрондық желілер (Artificial Neural Networks, ANN)

Нейрондық желілер кіріс пен шығыс арасындағы сызықтық емес байланыстарды үйрену арқылы күрделі үлгілерді бейнелейді. Фонемалық деректерде сызықтық емес заңдылықтарды үйрену арқылы ANN классификация дәлдігін арттырады және сөйлеу сигналындағы күрделі құрылымдарды тиімді модельдейді.

Көп қабатты перцептрон (MLP) - үлгілерді жіктеу үшін ең көп қолданылатын ANN архитектурасының бірі. Ол дауысты синтездеудің, фонемаларды танудың әртүрлі схемаларында қолданылған (Pineda-Jaramillo, J. D. (2019)). 3-суретте ANN құрылымы көрсетілген, ол екі қабаттан тұрады: жасырын қабат және шығыс қабаты. Шығыс қабатының мәні (1) арқылы көрсетіледі.

$$A_{c_k} = f^*(\sum_{j=0}^M W2_{kj} \times f(\sum_{i=0}^N W1_{ji} \times V_{ci})) \quad (1)$$

мұндағы A_{c_k} – жасанды нейрондық желінің (ANN) шығысы, M – шығыс элементтерінің саны, N – кіріс атрибуттарының саны, $W1_{ji}$ – бірінші қабаттың салмақ коэффициенті, $W2_{kj}$ – екінші қабаттың салмақ коэффициенті, ал f және f^* – әр қабатқа қолданылатын активация функцияларын білдіреді.



3-сурет. Жасанды нейрондық желі құрылымы

Ескерту – авторлармен құрастырылған



Қайталанатын нейрондық желі (Recurrent Neural Network, RNN)

Қайталанатын нейрондық желі (RNN) — циклдік байланыстары бар жасанды нейрондық желі (ANN) түрі, ол қарапайым ANN-мен салыстырғанда деректер тізбегін модельдеуде неғұрлым қуатты құрал болып табылады.

Кәдімгі көпқабатты перцептронда (MLP) әр қабаттың өз салмақтары мен ығысу мәндері (bias) болады. Бұл нейронның ағымдағы күйін есте сақтап, сол күйге сүйене отырып келесі шығуды (output) қалыптастыруға мүмкіндік береді (Shetty, S.K., & Siddiqua, A. (2019)).

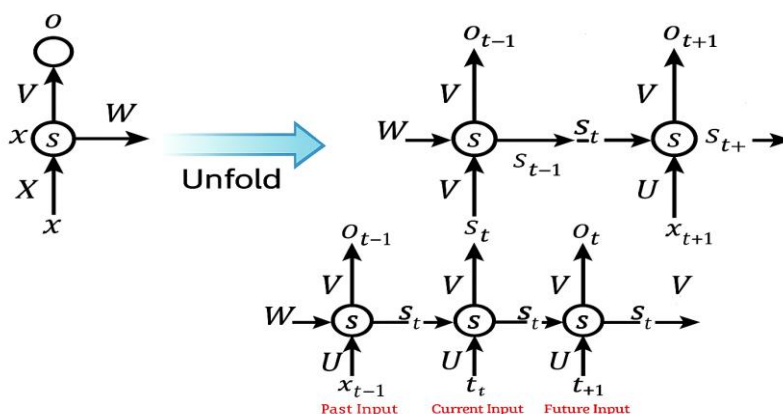
RNN-де есте сақтау қабілеті (memory) бар, ол болашақтағы болжамдарды анықтайды. Сондықтан сөйлемдегі кез келген келесі сөзді болжау үшін бұрын өңделген сөздер туралы білім қажет (Rao, M. S., Lakshmi, G. B., Gowri, P., & Chowdary, K. B. (2020)).

RNN алгоритмі 4-суретте көрсетілген. (2)-формула бойынша RNN жасырын күйді (C_t) есептейді, ол алдыңғы қадамдардағы деректерді сақтайды.

$$C_t = f(Y \times x_t + W \times x_{t-1}) \quad (2)$$

мұндағы x_{t-1} – алдыңғы кіріс дерек, x_t – уақыттың t сәтіндегі ағымдағы кіріс, Y және W – жасырын қабаттың салмақ коэффициенттері, ал V – шығыс қабатының салмағы. Уақыттың t мезетіндегі жадқа (memory) негізделе отырып, шығыс қадамы (output step) (3)-формула бойынша есептеледі.

$$A_t = \text{softmax}(V \times C_t) \quad (3)$$



4-сурет. RNN алгоритмі

Ескерту – авторлармен құрастырылған

Конволюциялық нейрондық желілер (Convolutional Neural Networks, CNN)

CNN архитектурасы спектрограммалар мен аудио сигналдардан кеңістіктік белгілерді алу үшін конволюциялық қабаттарды қолданады. Бұл тәсіл дыбыс сигналындағы жергілікті жиілік және уақытша үлгілерді тиімді анықтап, фонемаларды танудың дәлдігін арттырады.

Қайталанатын нейрондық желілер (Recurrent Neural Networks, RNN)

RNN архитектуралары тізбекті деректермен жұмыс істеуге арналған. Фонемаларды тануда RNN сөйлеу сигналындағы контексті және фонемалар арасындағы уақытша байланыстарды тиімді модельдейді.

НӘТИЖЕЛЕР ЖӘНЕ ОЛАРДЫ ТАЛҚЫЛАУ

Бұл бөлімде қазақ фонемаларының ұсынылған моделі егжей-тегжейлі талқыланады. Бұл модель төрт бөлімнен тұрады. Бірінші бөлімде мәліметтер жиынтығы дайындалады,

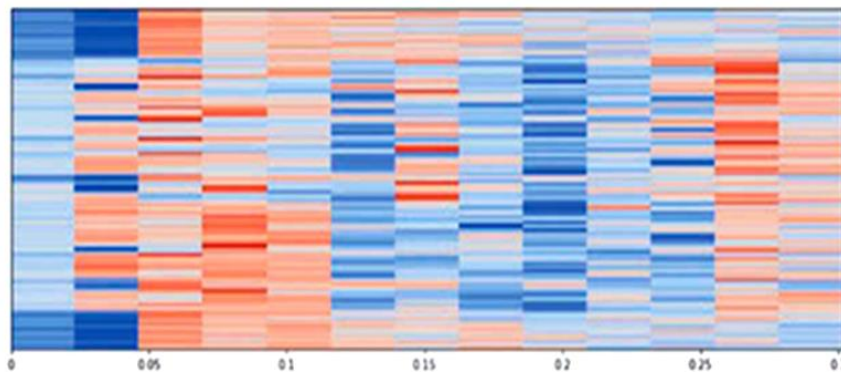
содан кейін MFCC мүмкіндіктері шығарылады. Одан кейін ұсынылған өнімділік көрсеткішімен бағаланатын тиісті ML алгоритмдері бойынша оқыту және тестілеу кезеңдері жүргізіледі. Ұсынылған қазақ фонемасының құрылымы 5-суретте көрсетілген. Бұл құрылым келесі өңдеу қадамдарын қамтиды.



5-сурет. Фонемалардың моделі

Ескерту – авторлармен құрастырылған

Бұл зерттеу жұмысында эксперимент жүргізу мақсатында қажетті деректер жиыны Ақпараттық және есептеуіш технологиялар институтының Интеллектуалдық жүйелердің компьютерлік инженериясы зертханасынан алынды. Зерттеу барысында фонемаларды (дауысты дыбыстарды) тану процесінде аудио сигналдарды алдын ала өңдеу үшін Mel Frequency Cepstral Coefficients (MFCC) әдісі пайдаланылды. Алынған сигналдар негізінде талдау жүргізу үшін машиналық оқыту модельдері қолданылды. Зерттеу нәтижесінде алынған нысандар 6-суретте көрсетілген.



6-сурет MFCC фонеманың (дауысты дыбыстың) дыбыстық сигнал функциясы

Ескерту – авторлармен құрастырылған



Сөйлеу сигналдары 16 кГц жиілікте жазылды, әрбір үлгі ұзақтығы орта есеппен 1,5 секундты құрады. MFCC ерекшеліктері 13 коэффициент арқылы есептелді:

$$c_n = \sum_{k=1}^K \log(S_k) \cos \left[n(k - 0.5) \frac{\pi}{K} \right]$$

мұнда S_k – мел-диапазондағы спектр энергиясы, n – коэффициент индексі. Алынған белгілер librosa және numpy кітапханалары арқылы есептеліп, 80 % деректер – оқытуға, 20 % – тестілеуге бөлінді.

1-кесте. Қалыпты сөйлеу деректері үшін алынған MFCC мүмкіндіктерінің үлгісі

Қатар	Мүмкіндік 1	Мүмкіндік 2	Мүмкіндік 3	Мүмкіндік 4	Мүмкіндік 5
26	-15,82301	-10,35246	14,11029	-8,54133	-15,82301
27	6,89115	3,55182	-2,15549	1,87566	6,89115
28	2,50877	-3,98701	-1,90223	-2,48391	2,50877
29	-3,10945	1,05432	-5,20401	0,65432	-3,10945
30	-7,15678	-5,22899	-6,81230	-8,99112	-7,15678
31	-0,98765	-2,11223	1,15678	-3,55401	-0,98765
32	4,00123	-2,10098	4,56789	2,01991	4,00123
33	-1,50392	5,01876	3,51234	6,00987	-1,50392
34	-4,50781	-6,99123	0,55432	0,10098	-4,50781
35	1,05678	-1,25678	1,50345	-2,50012	1,05678
36	-6,80123	-8,00567	-3,50111	-1,90567	-6,80123
37	-1,05678	0,50123	-3,90876	-1,55432	-1,05678
38	-2,55123	-2,89012	-5,50123	-7,88901	-2,55123
39	5,50123	4,56789	2,89012	4,50123	5,50123

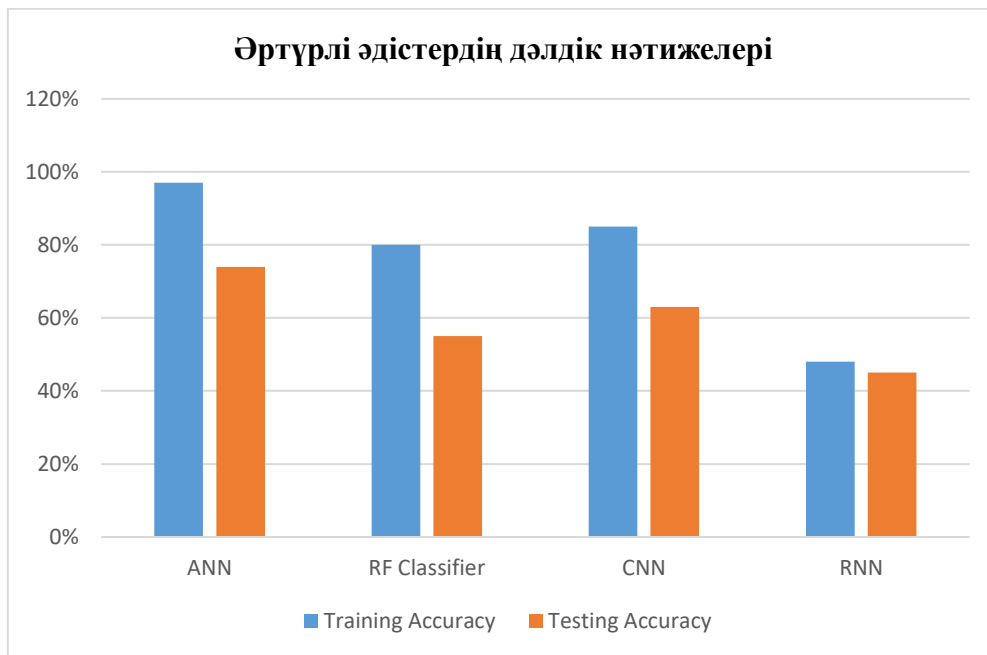
Ескерту – авторлармен құрастырылған

Кез келген басқа үлгіні тану жүйелері сияқты, фонемаларды (дауысты дыбыстарды) тануды орындау процесі екі кезеңнен тұрады, атап айтқанда: оқыту және тестілеу. Оқыту кезеңінде бүкіл деректер жиынтығынан алынған мүмкіндіктердің мәліметтер базасы құрылады. Бұл мүмкіндіктер осы жұмыста қолданылатын машиналық оқытудың ұсынылған алгоритмдерін үйрету үшін қолданылады. Қолданбалы машиналық оқыту жүйесінің модельдері ANN, CNN, RNN, RF олар дербес немесе жіктеу тапсырмасы үшін гибридті модельде. Қолданылатын гибридті модельдер CNN, RF және болып табылады, ал тестілеу кезеңінде мүмкіндіктер әрбір кіріс сигналынан (дауысты дыбыстар кездесетін мәтіндерден) алынады және бұл мүмкіндіктердің бұрын жасалған дерекқорға жататынын немесе жатпайтынын анықтау үшін мүмкіндіктерді сәйкестендіру процесі орындалады. Әрбір машиналық оқыту әдісінің әрекеті оның өнімділігін өлшеудің нақты параметрлері негізінде бағаланады. Әрбір модель үшін қолданылатын параметрлер дұрыс салыстыруды қамтамасыз ету және берілген әдістің ең жақсы нәтижелерін алу үшін бірнеше рет өзгертілді. ANN моделі кіріс қабатында 40 нейронмен, жасырын қабатта 256 нейронмен және шығыс қабатында 27 нейронмен оқытылады. Экспоненциалды сызықтық бірлік (ELU) және SoftMax белсендіру функциялары 0,001 оқу жылдамдығымен қолданылады. Пакет өлшемінің мәні 32, ал түсіру мәні кіріс үшін 0,5 және шығыс қабаты үшін 0,75 құрайды. Шамадан тыс сәйкестік мәселесін шешу үшін реттегіш мәні 0,2 пайдаланылады.

CNN моделі үшін жүйе бір жинақтау қабатын, бір максималды біріктіру қабатын және бір тегістелген қабатты қолдану арқылы оқытылады. Кіріс қабаты – 40 нейрон, әр қабат үшін 256 нейроннан тұратын 2 жасырын қабат және 27 нейроннан тұратын шығыс қабаты. Қолданылатын белсендіру функциялары түзетілген сызықтық блок (ReLU), ELU және SoftMax, ал партия өлшемі 64. Шығару мәні-0,6, ал реттегіш-0,02.

9-суретте фонемаларын (дауысты дыбысты) танудың ұсынылған әдістерімен алынған жаттығулар мен тестілеу кезеңдерінің дәлдік нәтижелері қалыпты және ELU негізіндегі сөйлеуге байланысты көрсетілген. Суретте көрсетілгендей, ANN моделі тестілеудің дәлдігі 73,23 % болатын басқа модельдерден асып түседі. ANN – бұл деректерді жіктеу мәселелерін шешудің салыстырмалы түрде жеңіл әдісі, әсіресе деректер жиынтығының шектеулі шарттары үшін.

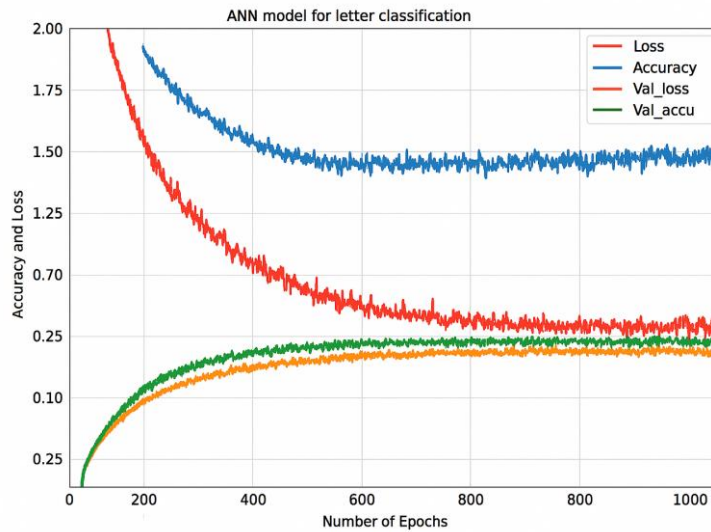
Оқытуды оқыту қисығы оқу деректер жиынтығынан есептеледі. Бұл әдісті сәтті үйрену туралы түсінік береді, ал валидация туралы мәліметтер жиынтығынан есептелген валидацияны оқыту қисығы жалпы модельдің қаншалықты жақсы екендігі туралы түсінік береді. Сондай-ақ, кросс-энтропияның жоғалуына сәйкес тапсырмаларды оңтайландыру үшін оқу қисықтарын құру жиі кездеседі және модельдің өнімділігі жіктеу дәлдігін қолдана отырып бағаланады. 7 суретте әртүрлі әдістердің дәлдік нәтижелері көрсетілген.



7-сурет. Әдістердің дәлдік нәтижелері

Ескерту – авторлармен құрастырылған

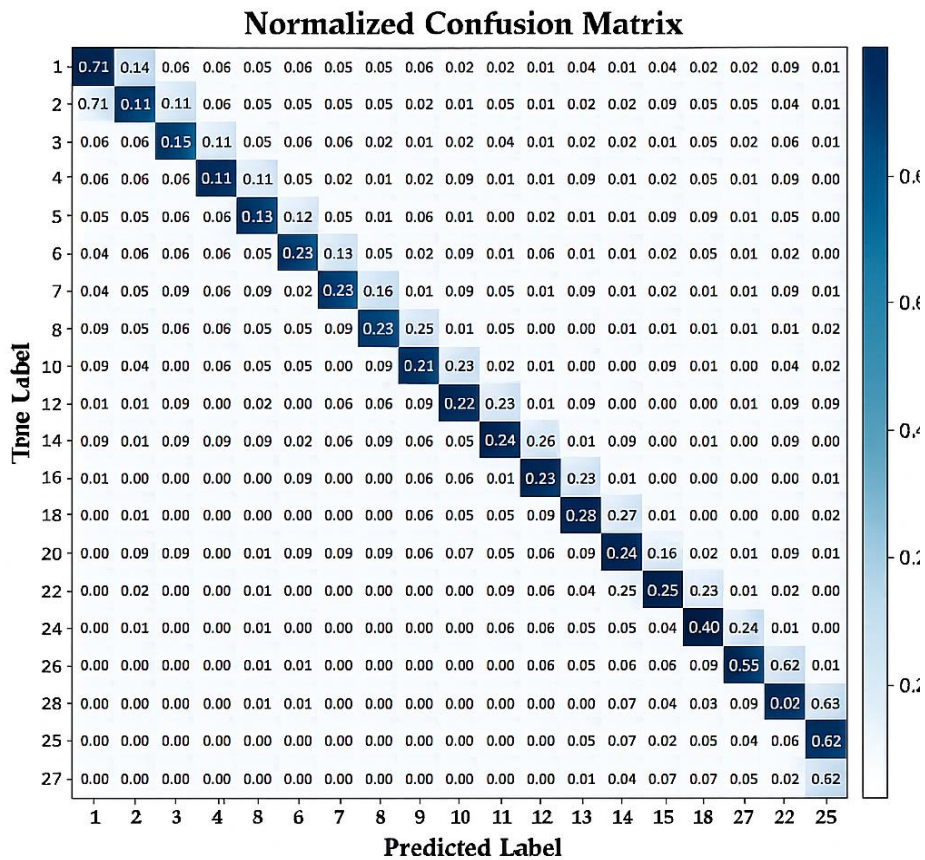
8-суретте жаттығу дәлдігі артқан сайын ANN моделінің оқу қисықтары көрсетілген. Тексеру дәлдігі алдымен артады, содан кейін шамадан тыс сәйкестік әсеріне байланысты белгілі бір аралықтар санынан (өсу немесе кему) кейін төмендей бастайды. Бұл жаттығу кезінде жүйелеуді қолдану және оқуды тастап кету қабаттарын қосу туралы болды. Шығын модельдің қателігін өлшейді, жаттығу барысында азаяды, бұл модельдің жақсы өнімділігін көрсетеді.



8-сурет. ANN моделінің оқыту қисықтары

Ескерту – авторлармен құрастырылған

9-суретте ANN-нің матрицасы көрсетілген. Фонемалардың (дауысты дыбыстардың) көпшілігі дұрыс жіктелгені сөйлемдегі дауысты дыбыстар арасында жоғары мәнге ие болғаны көрсетілген. ANN моделінің мәні үшін дәлдік мәні 77 % құрайды. 2-ші кестеде машиналық оқыту модельдерінің салыстырмалы нәтижелері көрсетілген.



9-сурет. ANN моделінің матрицасы

Ескерту – авторлармен құрастырылған



2-кесте. Машиналық оқыту модельдерінің салыстырмалы нәтижелері

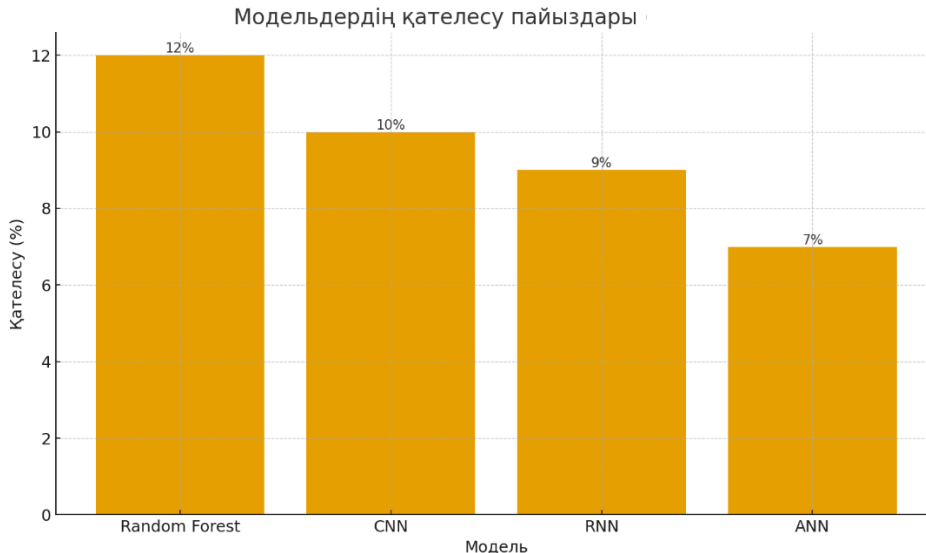
Модель	Дәлдік (%)	Анықтау уақыты (с)	Ерекшелік саны
Random Forest	68,4	0,12	40
CNN	72,1	0,25	128
RNN	75,8	0,28	256
ANN	77,0	0,19	256

Ескерту – авторлармен құрастырылған

Кестеден байқағандай, ұсынылған ANN моделі дәлдік көрсеткіші бойынша (77,0%) басқа әдістерден (Random Forest – 68,4%, CNN – 72,1%, RNN – 75,8%) жоғары нәтиже көрсетті. Сонымен қатар, ANN моделінің анықтау уақыты салыстырмалы түрде төмен (0,19 с), бұл оның тиімділігін дәлелдейді. Бұл көрсеткіштер нейрондық желілердің қазақ тіліндегі фонемаларды танудағы әлеуетін растайды.

Алынған нәтижелерге терең талдау жүргізу барысында ANN моделінің ең жоғары дәлдік көрсеткішімен, кейбір фонемаларды шатастыру байқалды. Мысалы, “ы” және “і” дыбыстары спектрлік тұрғыда ұқсас болғандықтан, модель оларды кейде қате сәйкестендірді. CNN моделінде кеңістіктік белгілерді анықтау күштірек болғанымен, уақыттық контексті есепке алмауы дәлдікке әсер етті. Ал RNN архитектурасы уақыттық тәуелділікті жақсы сақтағанымен, оқу уақыты ұзағырақ және артық үйренуге бейім болды.

Модельдердің салыстырмалы қателесу пайыздары бойынша ANN моделі жалпы 7 %, RNN – 9 %, ал Random Forest – 12 % қате жібергені байқалды. Бұл нәтиже нейрондық желілердің қазақ тіліндегі сөйлеу жүйесіне бейімделу әлеуетін дәлелдейді. 10-суретте модельдердің шатасу пайызы көрсетілген.



10-сурет. Модельдердің қателесу пайыздары

Ескерту – авторлармен құрастырылған

ҚОРЫТЫНДЫ

Бұл зерттеуде қазақ тіліндегі дауысты фонемаларды тану үшін машиналық оқытудың әртүрлі әдістері (ANN, CNN, RNN және RF) талданды. Эксперимент барысында фонемаларды тану үшін Mel Frequency Cepstral Coefficients (MFCC) арқылы ерекшелік белгілері алынды және алынған деректер машиналық оқыту модельдерімен өңделді.



Зерттеу нәтижелері бойынша жасанды нейрондық желі (ANN) моделі басқа әдістермен салыстырғанда ең жоғары дәлдік көрсетті – 77 %. Бұл модель қазақ тіліндегі фонемаларды тану жүйелерін жетілдіруде тиімді құрал бола алатынын көрсетті.

Ұсынылған әдіс қазақ тіліндегі сөйлеуді тану сапасын арттыруға, сонымен қатар дауыс биометриясы, интеллектуалды көмекшілер және дыбыстық интерфейстер секілді салаларда қолдануға мүмкіндік береді. Болашақта зерттеу жұмысы мәліметтер жиынын кеңейту, шулы ортада тану дәлдігін арттыру және гибриді нейрондық желілерді (мысалы, CNN–RNN немесе CNN–RF) одан әрі жетілдіру бағытында жалғастырылуы мүмкін.

Алынған нәтижелер негізінде қазақ тіліндегі сөйлеуді тану жүйелерінде ANN архитектурасын MFCC ерекшеліктерімен біріктіру ең тиімді шешім болып табылады. Дегенмен, зерттеу барысында деректер жиынының шектеулі болуы және фондық шудың әсері толық зерттелмеді. Болашақта үлкен деректер базасы мен гибриді (CNN–RNN) архитектураларды қолдану арқылы жүйенің жалпы өнімділігін арттыру жоспарлануда.

МҮДДЕЛЕР ҚАЙШЫЛЫҒЫ: Автор(лар) мүдделер қайшылығы жоқ екенін мәлімдейді.

ҚАРЖЫЛАНДЫРУ: Бұл зерттеу сыртқы қаржыландыруды алған жоқ.

АЛҒЫС БІЛДІРУ: Авторлар мақаланың сапасын жақсартуға ықпал еткен құнды пікірлері мен ескертулері үшін анонимді рецензенттерге алғыс білдіреді. Сондай-ақ, зерттеу барысындағы идеялар мен нәтижелердің талқылануына үлес қосқан әріптестерге және ғылыми қауымдастық өкілдеріне ризашылық білдіріледі.

ӘДЕБИЕТТЕР ТІЗІМІ

- Al-Karawi, K. A., Al-Noori, A. H., Li, F. F., & Ritchings, T. (2015). Automatic speaker recognition system in adverse conditions: Implication of noise and reverberation on system performance. *International Journal of Information and Electronics Engineering*, 5(5), 423–427. <https://doi.org/10.7763/IJIEE.2015.V5.564>
- Comert, Z., & Kocamaz, A. F. (2017). A study of artificial neural network training algorithms for classification of cardiocography signals. *Bitlis Eren University Journal of Science and Technology*, 7(2), 93–103. <https://doi.org/10.17678/beuscitech.338085>
- Kalimoldayev, M., Mamyrbayev, O., Mekebayev, N., & Kydyrbekova, A. (2020). Algorithms for detection of gender using neural networks. *International Journal of Circuits, Systems and Signal Processing*, 14, 154–159. <https://doi.org/10.46300/9106.2020.14.23>
- Li, J., Deng, L., Haeb-Umbach, R., & Gong, Y. (2016). *Robust automatic speech recognition: A bridge to practical applications*. Academic Press. <https://doi.org/10.1016/C2014-0-00305-8>
- Mamyrbayev, O., Mekebayev, N., Turdalyuly, M., Oshanova, N., Medeni, T. I., & Yessentay, A. (2019). Voice identification using classification algorithms. In *Open Access Peer-Reviewed Chapter*, IntechOpen. <https://doi.org/10.5772/intechopen.87160>
- Mekebayev, N., Tuyebaev, Ch., Sabrayev, K., & Yerkebay, A. (2022). Research of acoustic and linguistic modeling based on repetitive neural networks for speech recognition of children. *Bulletin of Physics & Mathematical Sciences*, 1(77), 119–126. <https://doi.org/10.51889/2022-1.1728-7901.16>
- Nugroho, K. (2020). Javanese gender speech recognition based on machine learning using random forest and neural network. *SISFORMA*, 6(2), 83–90. <https://doi.org/10.24167/sisforma.v6i2.2402>
- Pineda-Jaramillo, J. D. (2019). A review of machine learning (ML) algorithms used for modeling travel mode choice. *DYNA*, 86(211), 32–41. <https://doi.org/10.15446/dyna.v86n211.79743>



- Povey, D., Peddinti, V., Galvez, D., Ghahremani, P., Manohar, V., Na, X., Wang, Y., & Khudanpur, S. (2016). Purely sequence-trained neural networks for ASR based on lattice-free MMI. *Interspeech 2016* (pp. 2751–2755).
<https://doi.org/10.21437/Interspeech.2016-595>
- Rao, M. S., Lakshmi, G. B., Gowri, P., & Chowdary, K. B. (2020). Random forest-based automatic speaker recognition system. *International Journal of Analytical and Experimental Modal Analysis*, 12(4), 526–535. <https://doi.org/10.5281/zenodo.3767442>
- Sarma, M., & Sarma, K. K. (2013). An ANN-based approach to recognize initial phonemes of spoken words of Assamese language. *Applied Soft Computing*, 13(5), 2281–2291. <https://doi.org/10.1016/j.asoc.2013.01.004>
- Shetty, S. K., & Siddiq, A. (2019). Deep learning algorithms and applications in computer vision. *International Journal of Computer Sciences and Engineering*, 7(7), 195–201. <https://doi.org/10.26438/ijcse/v7i7.195201>
- Yuhas, B. P., Goldstein, M. H., Sejnowski, T. J., & Jenkins, R. E. (1989). Integration of acoustic and visual speech signals using neural networks. *IEEE Communications Magazine*, 27(11), 65–71. <https://doi.org/10.1109/35.41401>
- Yu, D., & Deng, L. (2016). *Automatic speech recognition: A deep learning approach*. Springer. <https://doi.org/10.1007/978-1-4471-5779-3>
- Zhang, Y., Wang, Z., & Zhao, J. (2023). End-to-end transformer-based model for multilingual speech recognition. *IEEE Access*, 11, 5123–5135. <https://doi.org/10.1109/ACCESS.2023.3239121>

Авторлар туралы мәліметтер
Информация об авторах
Information about authors



Мекебаев Нурбапа Отанович – PhD, Қазақ ұлттық қыздар педагогикалық университетінің қауымдастырылған профессоры, Алматы қ., Қазақстан

Мекебаев Нурбапа Отанович – PhD, ассоциированный профессор Казахского национального женского педагогического университета, г. Алматы, Казахстан

Mekebayev Nurbapa – PhD, Associate Professor, Kazakh National Women's Teacher Training University, Almaty, Kazakhstan,
e-mail: nurbapa@gmail.com,
ORCID: <https://orcid.org/0000-0002-9117-4369>,



Даркенбаев Даурен Кадырович – PhD, әл-Фараби атындағы Қазақ ұлттық университетінің қауымдастырылған профессоры, Алматы қ., Қазақстан

Даркенбаев Даурен Кадырович – PhD, ассоциированный профессор КазНУ имени аль-Фараби, г. Алматы, Казахстан

Dauren Darkenbayev – PhD, Associate professor, Al-Farabi Kazakh National University, Almaty, Kazakhstan
e-mail: dauren.kadyrovich@gmail.com,
ORCID: <https://orcid.org/0000-0002-2935-5003>



Зиятбекова Гулзат Зиятбеккызы – PhD, қауымдастырылған профессор Алматы технологиялық университеті; әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан

Зиятбекова Гулзат Зиятбеккызы – PhD, ассоциированный профессор Алматинский технологический университет; КазНУ имени аль-Фараби, г. Алматы, Казахстан

Gulzat Ziyatbekova – PhD, Associate professor, Almaty Technological University; Al-Farabi Kazakh National University, Almaty, Kazakhstan
e-mail: ziyatbekova1@gmail.com

ORCID: <https://orcid.org/0000-0002-9290-6074>



Орынтаева Жанна Абдыхамитовна – магистр, Қазақ ұлттық қыздар педагогикалық университетінің аға оқытушысы, Алматы қ., Қазақстан

Орынтаева Жанна Абдыхамитовна – магистр, старший преподаватель Казахского национального женского педагогического университета, г. Алматы, Казахстан

Oryntaeva Zhanna – master, Kazakh National Women's Teacher Training University, Almaty, Kazakhstan
e-mail: zannaoryntaeva0@gmail.com

ORCID: <https://orcid.org/0000-0002-1993-9566>



Судейменов Габиден Мухтарович – Қазақ ұлттық қыздар педагогикалық университетінің оқытушысы, Алматы қ., Қазақстан

Судейменов Габиден Мухтарович – Преподаватель Казахского национального женского педагогического университета, г. Алматы, Казахстан

Suleimenov Gabiden – Teacher of Kazakh National Women's Teacher Training University, Almaty, Kazakhstan,
e-mail: gabiden1604@gmail.com

ORCID: <https://orcid.org/0000-0002-9341-7408>