ҒТАМР 28.23.17

# СИПАТТАМАЛЫҚ СПЕКТРОГРАММАЛАР МЕН НЕЙРОНДЫҚ ЖЕЛІ НЕГІЗІНДЕ СӨЙЛЕУШІНІ ТАНУ

**[1]Н.О. Мекебаев✉, [2]Д.К. Даркенбаев✉, [1]Ж.А. Орынтаева, [2]Н.А. Модовов**

*[1]Қазақ ұлттық қыздар педагогикалық университеті, Алматы, Қазақстан,*
*[2]Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан*

✉ *Корреспондент-автор:nurbapa@gmail.com, dauren.kadyrovich@gmail.com*

Сөйлеушінің айтылу сипаттамаларын алу үшін нейрондық желіге негізделген әдіс ұсынылады, ол қысқа мерзімді спектрограммалардың сызықтық суперпозициясынан сөйлеушінің айтылуының тұрақты көрінісін беру үшін сипаттамалық спектрограммаға қол жеткізу үшін арнайы программа статистикасын пайдаланады. Дәстүрлі SOM (AC-SOM) нейрондық желісіне негізделген ресурстармен шектелген құрылғылардағы динамиктерді тану жүйелері үшін желіні баяу оқыту және тану жылдамдығы мәселесін шешу үшін адаптивті кластерлік өзін-өзі ұйымдастыратын мүмкіндіктер картасы SOM (AC-SOM) алгоритмі ұсынылады. Бұл алгоритм кластерлер саны динамиктер санына сәйкес келгенше танылатын динамиктер санына қарай бәсекелестік деңгейіндегі нейрондар санын автоматты түрде реттейді. Ұсынылған AC-SOM моделіне спектрограммалардың сипаттамалық үлгілерінің 100 сөйлеушінің дерекқоры құрастырылды және қолданылды, бұл максималды оқу уақытын небәрі 304 секундқа, ал үлгіні танудың максималды уақытын 28 мс-ден аз уақытқа қамтамасыз етті. Басқа тәсілдермен салыстырғанда, ұсынылған әдіс тым жоғары тану дәлдігін жоғалтпай, айтарлықтай жақсартылған оқыту мен тану жылдамдығын қамтамасыз етеді. Перспективалы нәтижелер ұсынылған әдіс сөйлеушіні танудың басқа әдістеріне қарағанда edge intelligence жүйелері үшін нақты уақыттағы деректерді өңдеу және орындау талаптарын жақсырақ қанағаттандыратынын көрсетеді.

**Түйін сөздер:** сөйлеуді тану, алгоритм, MFCC, спектограмма, MLP, AC-SOM.

# РАСПОЗНАВАНИЕ ГОВОРЯЩЕГО НА ОСНОВЕ НЕЙРОННОЙ СЕТИ С ИСПОЛЬЗОВАНИЕМ ХАРАКТЕРИСТИЧЕСКИХ СПЕКТРОГРАММ

**[1]Н.О. Мекебаев✉, [2]Д.К. Даркенбаев, [1]Ж.А. Орынтаева, [2]Н.А. Модовов**

*[1]Казахский национальный женский педагогический университет, Алматы, Казахстан,*
*[2]Казахский национальный университет им. Аль-Фараби, Алматы, Казахстан,*
*e-mail: nurbapa@gmail.com, dauren.kadyrovich@gmail.com*

Предлагается метод, основанный на нейронной сети, для извлечения характеристик произношения говорящего. Он использует специальные триграммные статистики для получения характерной спектрограммы, способной обеспечить устойчивое представление произношения говорящего на основе линейной суперпозиции краткосрочных спектрограмм. Для решения проблемы медленного обучения и низкой скорости распознавания в системах идентификации динамиков на устройствах с ограниченными ресурсами предлагается алгоритм адаптивной кластерной самоорганизующейся карты признаков (AC-SOM), основанный на традиционной нейронной сети SOM. Данный алгоритм автоматически регулирует количество нейронов на уровне конкуренции в зависимости от количества распознаваемых динамиков до тех пор, пока количество кластеров не станет соответствовать количеству говорящих. Для предложенной модели AC-SOM была собрана и использована база данных спектрограммных образцов от 100 говорящих, что обеспечило максимальное время обучения всего 304 секунды и максимальное время распознавания образца менее 28 миллисекунд. По сравнению с другими подходами, предложенный метод обеспечивает значительно более высокие скорости обучения и распознавания без существенной потери точности распознавания. Полученные перспективные результаты демонстрируют, что предложенный метод лучше удовлетворя-

ет требованиям к обработке и исполнению данных в реальном времени в системах edge intelligence по сравнению с другими методами распознавания говорящего.

**Ключевые слова:** распознавание речи, алгоритм, MFCC, спектограмма, MLP, AC-SOM.

## SPEAKER RECOGNITION BASED ON NEURAL NETWORKS USING CHARACTERISTIC SPECTROGRAMS

**[1]N. Mekebayev✉, [2]D. Darkenbayev, [1]Zh. Oryntaeva, [1]N. Modovov**

*[1]Kazakh National Women' s Teacher Training University, Almaty, Kazakhstan,*
*[2]Al-Farabi Kazakh National University, Almaty, Kazakhstan,*
*e-mail: nurbapa@gmail.com, dauren.kadyrovich@gmail.com*

A neural network-based method is proposed for extracting speaker articulation characteristics. It employs specialized trigram statistics to obtain a representative spectrogram that provides a stable representation of the speaker's articulation through a linear superposition of short-term spectrograms. To address the issues of slow training and recognition speed in speaker recognition systems on resource-constrained devices, an Adaptive Clustering Self-Organizing Map (AC-SOM) algorithm is introduced, based on the traditional Self-Organizing Map (SOM) neural network. This algorithm automatically adjusts the number of neurons at the competitive layer based on the number of speakers to be recognized, until the number of clusters matches the number of speakers. A dataset comprising characteristic spectrogram patterns from 100 speakers was developed and used for the proposed AC-SOM model, achieving a maximum training time of only 304 seconds and a maximum recognition time of less than 28 milliseconds per sample. Compared to other approaches, the proposed method significantly improves training and recognition speed without sacrificing high recognition accuracy. The promising results indicate that, in contrast to existing speaker recognition techniques, the proposed method better meets the real-time data processing and execution requirements of edge intelligence systems.

**Keywords:** speech recognition, algorithm, MFCC, spectogram, MLP, AC-SOM.

**Кіріспе.** Сөйлеу адамның қарым-қатынасының негізгі көзі болып саналады. Ол биометриялық қауіпсіздік жүйелері сияқты адам мен машинаның өзара әрекеттесу әрекеттерінде кеңінен қолданылады. Бұл ақпарат алмасудың табиғи және тиімді құралдарын қамтамасыз ететін адамдар арасындағы қарым-қатынастың негізгі әдісі болып қала береді. Сөйлеу сигналдарын талдау эмоционалдық күйлерді анықтаудың, жеке тұлғаларды сипаттаудың, диалектілерді түсінудің, жасты бағалаудың және жеке басын, жынысын, тілін және денсаулық жағдайын қамтитын күшті құрал ретінде қызмет етеді. Әрбір жеке тұлғаның дауыс жолдарының тербеліс заңдылықтары бойынша қалыптасқан ерекше вокалдық сипаттамалары бар [1]. Сөйлеуді автоматты түрде тану (ASR) технологиясы соңғы жылдары айтарлықтай прогреске қол жеткізді, бұл негізінен тереңдетілген оқыту жетістіктері мен ауқымды деректер жиынының қолжетімділігіне байланысты. Қазіргі уақытта бұл жүйелер дауыстық көмекшілер мен транскрипция құралдарынан бастап нақты уақыттағы аударма және қол жетімділік қызметтеріне дейінгі қолданбаларға кеңінен біріктірілген. Алайда ASR технологияларының артықшылықтары тілдер арасында бірдей бөлінбеген. Сөйлеу сигналдары эмоциялар, сөйлеу жылдамдығы, дауыс жолдарының өлшемдері, жынысы, дауыс қатпарларының діріл жиілігі және екпін және басқа әсерлер сияқты факторларға байланысты жеке адамдар арасында өзгергіштікті көрсетеді. Зерттеушілер динамиктерді тиімді ажырату үшін осы ерекше белгілерді пайдаланады. Сөйлеу сигналдарының акустикалық ерекшеліктері толқын пішінінің бастапқы өлшеміне қатысты тиімділікке, ең аз резервтілікке және ықшамдылыққа бағытталған динамиктерді тану жүйелерін әзірлеуде өте маңызды. Сонымен қатар, сөйлеу сигналдары нақты уақыттағы аударма қосымшаларында қолданылады, мұнда жетілдірілген алгоритмдер сөйлеу тілін талдайды және оны бір-

ден дерлік басқа тілге түрлендіреді. Сонымен қатар, сөйлеу сигналдары тергеу барысында әңгімелер немесе оқиғалар туралы түсінік беретін маңызды сот-медициналық дәлелдер ретінде қызмет етеді. Олар сондай-ақ жеке тұлғаларды биологиялық және мінез-құлық ерекшеліктеріне қарай автоматты түрде анықтайтын биометриялық танудың ажырамас бөлігі болып табылады [2]. Сөйлеуді тануды қолданатын қосымшалар адамдардың күнделікті өмірінің бір бөлігіне айналады және адамдардың жүйелермен өзара әрекеттесуін жақсартуға көмектеседі. Бағдарламалық жасақтамаға негізделген икемділігінің арқасында дауысты тану технологиясы оны қолдануға болатын қосымшалардың мүмкіндіктері бойынша әмбебап болып табылады. Дауыс сапасы арқылы пайдаланушыларды анықтау және аутентификациялау контактісіз, жылдам және аудио сенсоры бар әртүрлі жағдайларда оңай жүзеге асырылады. Бірнеше мысалдар сөйлеуді тану жүйелерінің беріктігін оңай көрсете алады. Адамның сөйлеуі бірнеше мүшелердің, яғни өкпенің, дауыс жолының, дауыс байланыстарының және еріннің бірлескен әрекеті арқылы жасалады. Осы күрделі құрылымның арқасында біз сөйлеу сигналдары арқылы берілетін ақпаратты статистикалық талдау арқылы адамның айтылу сипаттамаларын білдіретін белгілерді ала аламыз [3]. Осы күрделі құрылымның арқасында біз сөйлеу сигналдары арқылы берілетін ақпаратты статистикалық талдау арқылы адамның айтылу сипаттамаларын білдіретін белгілерді ала аламыз. Бұл белгілерді жалпы бес жалпы түрге бөлуге болады: қысқа мерзімді спектрлік белгілер, дауыс көзінің белгілері, спектрлік уақыт белгілері, просодикалық белгілер және жоғары деңгей белгілері. Көптеген сөйлеушілерді тану жүйелері осы белгілердің бірнешеуін қатар қолданады, соның ішінде сөйлеудің әртүрлі аспектілері және оларды дәлірек тануға қол жеткізу үшін оларды қосымша тәсілдермен қолдану [4].

Сөйлеу адамның қарым-қатынасының негізгі көзі болып саналады. Ол биометриялық қауіпсіздік жүйелері сияқты адам мен машинаның өзара әрекеттесуінде кеңінен қолданылады. Соңғы жылдары сөйлеуді автоматты түрде тану (ASR – Automatic Speech Recognition) технологиясы айтарлықтай прогреске қол жеткізді, бұл негізінен терең оқыту жетістіктері мен ауқымды деректер жиынының қолжетімділігіне байланысты. Алайда қазақ тіліндегі ASR жүйелері жеткілікті дамымаған. Сондықтан сөйлеуді танудың тиімді әдістерін жетілдіру және қазақ тілінде бейімделген деректер қорын құру өзекті мәселелердің бірі болып табылады.

*Зерттеу мақсаты* – сөйлеушінің дауыс сигналынан ерекшелік белгілерді тиімді алу арқылы қазақ тіліндегі сөйлеушіні танудың жылдамдығы мен дәлдігін арттыратын әдіс ұсыну.

*Проблемалық ережелер* – қазақ тіліндегі сөйлеушіні тану саласында ірі көлемді деректер қорының болмауы, дәстүрлі MFCC және LPCC әдістерінің кей жағдайда жоғары дәлдік бермеуі, ресурстары шектеулі құрылғыларда сөйлеушіні танудың тиімділігінің төмендігі.

*Өзектілігі:* Қазіргі ақпараттық қоғамда дауыстық интерфейстердің рөлі артып келеді. Қазақ тілінде сөйлеушіні автоматты түрде тану жүйелерін жетілдіру – ұлттық тілдегі интеллектуалды технологияларды дамыту үшін маңызды.

*Жаңалығы:* Мақалада ұсынылған адаптивті кластерлік өзін-өзі ұйымдастыратын карта (ACSOM) алгоритмі дәстүрлі MFCC және LPCC әдістерімен салыстырғанда қазақ тіліндегі сөйлеушілерді тануда жылдамдық пен дәлдікті арттырады.

*Ғылыми маңыздылығы:* Қазақ тіліндегі сөйлеуді өңдеу және тану бойынша жаңа деректер базасының құрылуы әрі тиімді ерекшелік алу әдістерінің енгізілуі тілге бейімделген интеллектуалды жүйелерді дамытуға үлес қосады.

*ASR (Automatic Speech Recognition)* – сөйлеуді автоматты түрде тану. Бұл ағылшын тіліндегі халықаралық термин, себебі ғылыми қауымдастықта стандартты аббревиатура ретінде қолданылады.

*MFCC (Mel-Frequency Cepstral Coefficients)* – Mel жиілігі бойынша кепстральды коэффициенттер. Бұл да ағылшын тіліндегі халықаралық қысқартпа, себебі зерттеулердің басым бөлігі

ағылшын тілінде жарияланады.

*Ғылыми зерттеудің өзекті мәселелері*

- Қазақ тілінде сөйлеушіні тану үшін арнайы үлкен деректер қорының болмауы.

- Нақты уақыт режимінде жұмыс істейтін тиімді алгоритмдердің жеткіліксіздігі.

- Ресурстары шектеулі құрылғыларда (смартфон, IoT) сөйлеуді тану жүйелерінің өнімділігінің төмендігі.

- MFCC және LPCC сияқты дәстүрлі әдістердің кей жағдайда жоғары дәлдік бермеуі.

**Материалдар мен әдістер.** Спектрограммалар-бұл уақыт өте келе өзгеретін сигналдағы жиілік спектрінің визуалды көрінісі. Қарапайым тілмен айтқанда, олар әр жиіліктің белгілі бір сәтте деңгейін көрсету үшін түс немесе қарқындылықты қолдана отырып, әр түрлі жиіліктердің (мысалы, дыбыстардың) уақыт өте келе қалай өзгеретінін көрсетеді. Сөйлеуді тануда спектрограммалар әсіресе пайдалы, өйткені олар әртүрлі фонемаларды, интонацияларды және екпінның өзгеруін ажыратуға көмектесетін сөйлеу дыбыстарының маңызды ерекшеліктерін түсіреді. Спектрограммалар-бастапқыда екінші дүниежүзілік соғыс кезінде сүңгуір қайықтарды анықтау және жау кодтарын ашу үшін әзірленген, бірақ кейінірек тіл білімі саласында сөйлеу спектрінің карталары қолданылған [5].

Ауызша сөйлеу сияқты дыбыстық сигналдар спектрограммаларға айналғанда, әзірлеушілер деректердегі заңдылықтарды тиімдірек талдай алады. Мысалы, спектрограммада сөйлеу түрлі-түсті жолақтар түрінде көрсетіледі, мұнда әртүрлі түстер әртүрлі жиіліктердегі әртүрлі энергия деңгейлерін білдіреді [6]. Бұл белгілі бір жиілік диапазондарын алатын дауысты және дауыссыз дыбыстарды анықтауды жеңілдетеді. Осы спектрограммалардан тиісті белгілерді алу арқылы машиналық оқыту модельдерін берілген аудио кіріске негізделген сөздерді немесе сөз тіркестерін болжауға үйретуге болады.

Практикалық қосымшаларда бұл автоматтандырылған транскрипция қызметтері немесе виртуалды көмекшілер сияқты жүйелер сөйлеу командаларын өңдеу үшін спектрограммаларды қолданады дегенді білдіреді. Пайдаланушы сөйлеген кезде оның дауысы спектрограммаға айналады және жүйе сөздерді тану үшін оны талдайды. Әзірлеушілер сөйлеуді тану үлгілерінің дәлдігін жақсарту үшін спектрограммалардан алынған мель-жиілік кепстральды коэффициенттері (MFCC) сияқты әдістерді қолдана алады. Бұл тәсіл жылдамдық немесе екпін сияқты сөйлеу вариацияларын жақсырақ өңдеуге мүмкіндік береді, осылайша адам сөйлеуін сенімдірек түсінетін сенімдірек қолданбаларды жасауға көмектеседі.

Сөйлеуді автоматты түрде тануда және адамның дауысын тануда, Mel жиілігінің кепстральды коэффициенттерін (MFCC) сөйлеу сигналдарын сипаттау үшін жиі қолданылады, өйткені оларда адамның есту жүйесі қабылдайтын акустикалық ақпарат бар.

Міне, Mfcc сөйлеуді түсінуге қалай үлес қосады:

*Сигналды Талдау:* Сөйлеу-бұл әр түрлі жиілік пен амплитудамен сипатталатын күрделі сигнал. Mfcc бұл сигналдарды уақыт өте келе дыбыс толқындарының өзгеру жылдамдығы мен сипаттамаларын көрсететін қарапайым компоненттерге бөлуге көмектеседі.

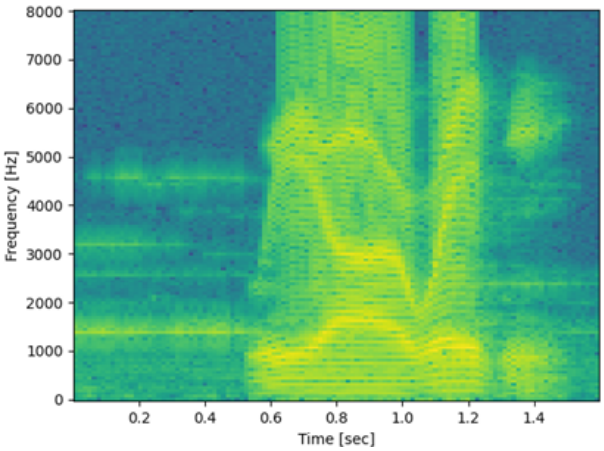
*Жиілікті Түрлендіру:* Адамдар жиіліктерді сызықтық шкала бойынша қабылдамайды. Сондықтан MFCC адамның есту жүйесінің реакциясына жақын келетін mel шкаласын пайдаланады, ол жоғары жиіліктерге қарағанда төменгі жиіліктердің өзгеруіне сезімтал.

Кепстральды көрініс: mel шкаласына ауысқаннан кейін сигнал қайтадан кепструм деп аталатын уақыт доменінің көрінісіне айналады. Кепструм сигналдың периодты вариациясын (қадамын) баяу вариациядан (тембрден) ажыратады, соңғысына назар аударады, ол сөйлеуді тануға қатысты ақпараттың көп бөлігін алып жүреді.

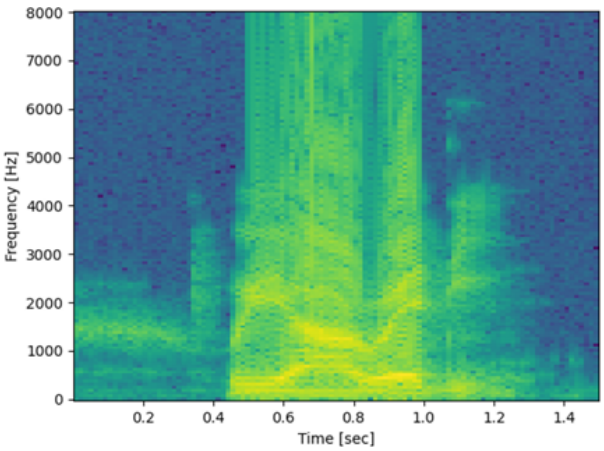
Спектрограммалардың жиілігі шағын өткізу қабілеттілігі және уақыт бойынша үлкен өткізу қабілеттілігі бар.

Сондықтан олар жоғары жиілікті ажыратымдылықты қажет етеді және сөйлеудің әртүрлі гармоникасын анық көрсете алады. Мысалы, 1-

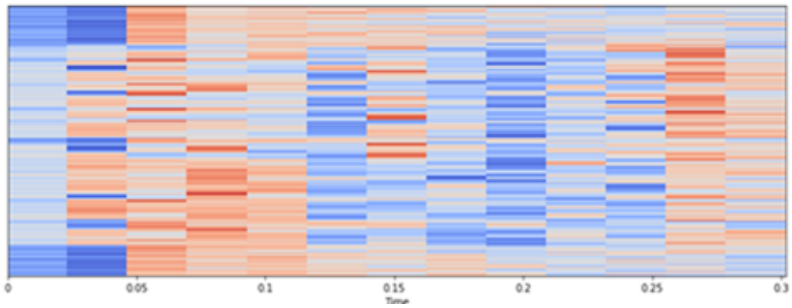
суретте көрсетілген ”біз дайынбыз” сөйлеу тіркесінің спектрограммасында дауыс жиілігі мен гармоникасын анық көруге болады.1-суретте көлденең бағыттағы төмен жолақтардың жиілік диапазоны қадам жиілігін білдіреді. Осы көлденең жолақтардың ішінде олардың кейбіреулері бір уақытта басқа көлденең жолақтарға қарағанда күңгірт түсті болады. Бұл қараңғы көлденең жолақтар сөйлеуінің резонанстық шыңын білдіреді. Атап айтқанда, жергілікті жерлерде бірнеше қараңғы көлденең жолақтар пайда болуы мүмкін, олар бірнеше резонанстық көлемді құрайды.2-суретте бір адам қалыпты атмосферада жазған ”біз дайынбыз” тіркесінің қалыпты сөйлеуінің спектрограммасы көрсетілген.




**1-сурет. ”Біз дайынбыз” сөйлеуінің спектрограммасы**




**2-сурет. ”Біз дайынбыз”қалыпты сөйлеу спектрограммасы**



**3-сурет. Кепстральды төмен жиілікті коэффициенттер**

**Нәтижелер мен талқылау.** (Mel Filter Bank) Mel сүзгі банкі кіріс қуатының спектрін Mel сүзгі банкі арқылы сүзеді. Шығыс деректері-әдетте Mel spectrum деп аталатын сүзілген мәндер жиыны, олардың әрқайсысы жеке сүзгі арқылы кіріс спектрін сүзу нәтижесіне сәйкес келеді. Бұған төмендегі формула арқылы қол жеткізуге болады:

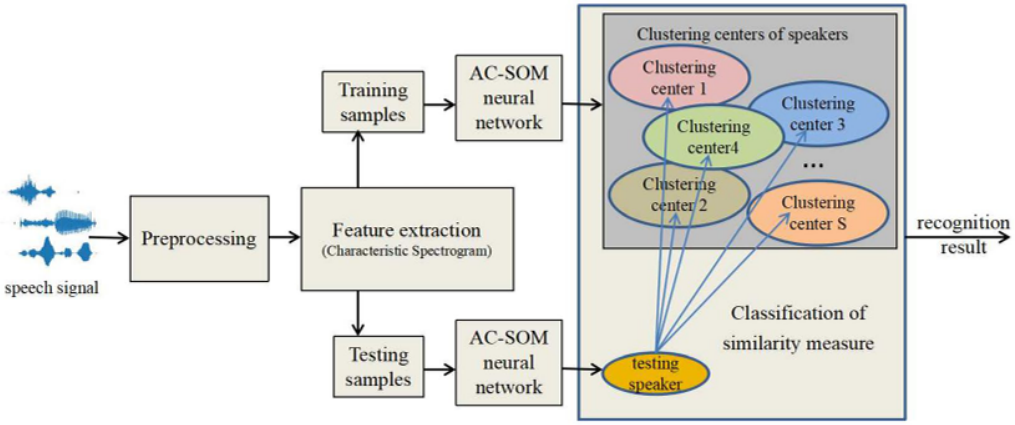
$$Y_n[n] = \sum_{i=0}^{N/2} Y_3[i] \times MelWeright[n][i], \;\; 0 < n < k$$

мұндағы k-сүзгілер саны. Жалпы модель үшін Mel мен сызықтық шкаладағы жиіліктер арасындағы қатынас келесідей: mel(f) = 2595*log10 (1+ f/700).3

Mel сүзгі банкі шығаратын мәндер диапазоны әр мәнді оның табиғи логарифміне ауыстыру арқылы төмендейді:

$$Y_n[n] = ln(Y_4[n]) \;\; 0 < n < k$$

Алдымен спектрді шағын масштабты формата көрсету үшін сүзгілердің тіркесімін есептеу керек. Mel сүзгісі-бұл жиілік диапазонындағы энергияны қосатын және mel коэффициенттерін есептейтін үшбұрышты терезе. Коэффициенттердің санын білетіндіктен, біз он сүзгі жиынтығын жасай аламыз (сурет.3).




**4-сурет. Эксперименттік динамикті сәйкестендіру жүйесіне шолу**

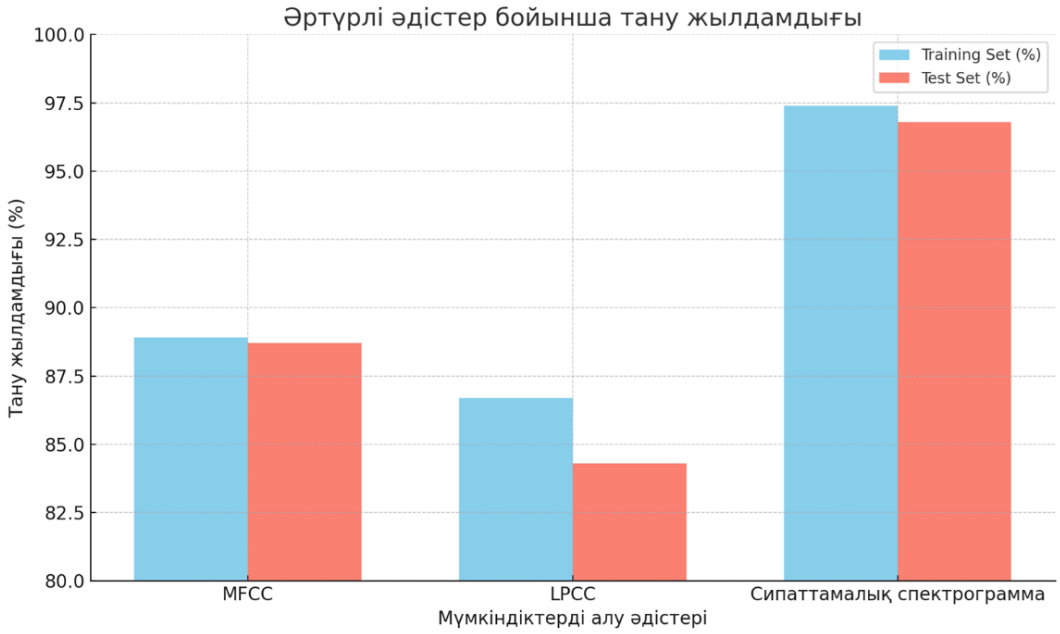
Зерттеу жұмысына қажетті деректер «Ақпараттық және есептеуіш технологиялар» институтының ғылыми базасынан алынып, тәжірибелік жұмыстарда пайдаланылды [7]. Тәжірибелер үшін біз 100 спикердің (50 ер және 50 әйел) жазбаларын қамтитын Қазақ тіліндегі мәліметтер базасын жасадық. Әрбір жазбаның ұзындығы шамамен 7 минутты құрады және зертханада 16 кгц іріктеу жиілігінде ДК аудио жазу бағдарламалық құралын пайдаланып жасалған және WAV пішімінде сақталған. Әр сөйлеуші газет, журналдардан және т.б. басқа мәтіндер алды және қалыпты қарқынмен сөйледі. Осы сөйлеу үлгілерін пайдалана отырып, біз сөйлеушілерге тән спектрограммалардың дерекқорын жасадық. Әрбір сөйлеуші жазбасы ұсталып, 4000 қысқа мерзімді спектрограммаға бөлінді. Содан кейін, 5, 40 қысқа мерзімді спектрограммалардың әр тобынан бір суперпозициялық спектрограммалар жасалды, нәтижесінде бір сөйлеушіге 100 осындай спектрограмм берілді. Кескінді өңдеу кезінде сызықтық суперимпо қимасы бірнеше кескіннің сәйкес пикселдерінде орташа өлшенген операцияны орындауды білдіреді. Бұл тәжірибелерде топ ретінде ”1” салмағымен орташа өлшенген операцияны орындау үшін 40 қысқа мерзімді спектрограмма пайдаланылады, оны біз бір суперпозиция деп атаймыз. Осыдан кейін дәл осындай орташа өлшенген операция 10-шы бір суперпозициялық спектрограммалар тобында қайтадан орындалады, оларды біз төрттік-суперпозиция деп атаймыз. Үлгілердің санын азайту және айтылымның тұрақты ерекшеліктерін алу үшін квадраттық сызықтық суперпозицияны қолдана отырып, 10-ын бір суперпозициялық спектрограммадан тұратын топтар біріктіріліп, соңында бір динамикке 10 сипаттамалық спектрограмма алынды. Осылайша, біз 100 динамиктен тұратын дерекқорымыздан 1000 сипаттамалық спектрограмма үлгілерін жасадық. 4-суретте динамикті тану жүйесіне шолу берілген. Бұл экспериментте біз әр сөйлеушінің деректерінің 80% - оқыту үшін, ал қалған 20% - тестілеу үшін пайдаландық, ал сынақ жиынтығындағы үлгілердің ешқайсысы оқыту үшін пайдаланылмады. Cubic convolution әдісіне сүйене отырып, шығыс өлшемі 420x560 болатын сипаттамалық спектрограммалар 42x56 өлшем-

ді суреттерді іріктеу үшін іріктеліп алынды, содан кейін үлгі суреттері оқуға a AC-SOM арналған нейрондық желісіне дәйекті түрде енгізілді. Кіріс қабатының өлшемі n = 42×56 = 2352 болды. S сөйлеуші бойынша тренингтен кейін біз s кластерлік орталықтарын алдық. Содан кейін біз әрбір үлгі мен барлық кластерлік орталықтар арасындағы Евклидтік қашықтық тангенсінің ұқсастығын есептей отырып, сынақ үлгілерін жіктедік, әрбір үлгі үшін тану нәтижесін ең жоғары ұқсастығы бар кластерлік орталық ретінде анықтадық.

5-ші суретте көрсетілгендей 1-кестеде бірдей тану әдісі бойынша IPCC динамигінің мүмкіндіктерін шығару әдісі ең төменгі тану жылдамдығына ие. MFCC мүмкіндіктерін шығару әдісін тану жылдамдығы IPCC-ге қарағанда жоғары. Бірақ екі тану көрсеткіші де 90% - дан аз. Ұсынылған спектрограммаға негізделген мүмкіндіктерді шығару әдісі бірдей динамиктер саны үшін жиі қолданылатын MFCC және LPCC әдістеріне қарағанда жақсырақ жұмыс істейді және динамиктердің айтылу мүмкіндіктерін тиімді түрде шығара алады.

**1-кесте. Үш айырмашылық үшін динамикті тану өнімділігін салыстыру ерекшеліктерді шығару әдістері**

| **Сөйлеушілер** | **Feature Extraction Methods Мүмкіндіктерді алу әдістері** | **Recognition Rate Тану жылдамдығы (Training Set) (%)** | **Recognition Rate Тану жылдамдығы (Test Set) (%)** |
|---|---|---|---|
| 20 | MFCC | 88.9 | 88.7 |
| 20 | LPCC | 86.7 | 84.3 |
| 20 | Сипаттамалық спектрограмма | 97.4 | 96.8 |
| 40 | MFCC | 87.5 | 85.2 |
| 40 | LPCC | 81.8 | 79.3 |
| 40 | Сипаттамалық спектрограмма | 93.3 | 91.1 |




**5-сурет. Әртүрлі әдістер бойынша тану жылдамдығы**

Бұл бөлімде ұсынылған әдіс жоғарыда сипатталған мәліметтер базасын қолдана отырып, сөйлеушіні танудың әр түрлі эксперименттерін құра отырып, сәйкестіктерімен бағаланады. Өнімділік динамиктерді танудың басқа әдістерімен салыстырылады, яғни терең сенім желісі (DBN) [8], конволюциялық нейрондық желі (CNN) [9]. Сонымен қатар, біздің көзқарасымыз басқа мүмкіндіктерді алу әдістерімен, атап айтқанда MFCC [10] және LPCC-мен салыстырылады. Оқу жылдамдығын салыстыру үшін барлық әдістер GPU қолданады. Барлық эксперименттерде біз тану жылдамдығын сыналған үлгілердің жалпы санынан негізгі rect сәйкестіктерінің саны ретінде келесідей есептедік:

$$= - \times 100\%$$

Осы формуланың көмегімен сынақ және жаттығу жиынтығының тану көрсеткіштері сәйкесінше барлық сынақ және жаттығу жиынтығының үлгілерін сынау арқылы алынды. Жаттығу уақыты динамикті тану жүйесіне қажетті уақыт (секундпен) ретінде анықталады команда оқу процесін аяқтауы керек, ал тестілеудің жалпы уақыты-бұл барлық сынақ жиынтықтарының үлгілерін жіктеуге кететін уақыт, ал бір сынақ уақыты-бір сынақ үлгісін жіктеуге кететін орташа уақыт.

Сөйлеушінің санының оқытудың орташа деңгейіне және тестілік жиынтықты тану көрсеткіштеріне әсерін зерттеу үшін біз жүйені келесі әдістермен тексердік 20, 40, 60, 80, 90 және сол эксперименттік жағдайда 100 сөйлеуші болды. Біз әр экспериментті үш рет өткіздік және орташа нәтижелер 2-кестеде келтірілген.

**2-кесте. Сөйлеушінің санының тану жылдамдығына, жаттығу жылдамдығына және тану жылдамдығына әсері**

Соңында, біз өз көзқарасымызды 20, 60 және 100 динамиктерді тану жылдамдығы мен жылдамдығы бойынша сөйлеушіні танудың басқа төрт әдісімен салыстырдық. CNN екі конволюциялық тиональды қабаттан, екі біріктіру қабатынан, екі толық қосылған қабаттан және бір softmax қабатынан тұрды, ал MLP құрылымы 2351×1000 ×500×250×100 сөйлеушіні саны. Эксперименттің үш кезеңінде танудың орташа жылдамдығы 6-суретте көрсетілген.

**6-сурет. Төрт түрлі тану әдісі үшін тану жылдамдығы**

6-шы суреттен көрініп тұрғандай бірдей эксперименттік жағдайларда CNN тану жылдамдығы ең жоғары болып табылады, бірақ ол жаттығу жылдамдығы мен тану жылдамдығын әсер етеді. Осы зерттеуде ұсынылған AC-SOM нейрондық желі әдісін тану жылдамдығы CNN әдісіне қарағанда сәл ғана төмен, бірақ ұсынылған желіні оқыту жылдамдығы мен тану жылдамдығы басқа әдістерге қарағанда айтарлықтай жылдамырақ, бұл анық, басқа әдістерден жоғары және нақты уақыттағы қолданбалардың қажеттіліктерін қанағаттандыра алады.

**Қортынды.** Деректерді оқытудың баяу жылдамдығы, тану тиімділігінің төмендігі және ресурстармен шектелген құрылғыларда қолданудың нашар өнімділігі мәселелерін шешу үшін бұл мақалада әрбір динамик үшін тұрақты айтылу мүмкіндіктерін алу үшін қысқа мерзімді спектрограмма статистикасын пайдаланатын әдіс ұсынылады, содан кейін AC-SOM нейрондық желісіне негізделген адаптивті кластерлеу әдісі бар динамиктерін алдық. Қазақ тіліндегі мәліметтер базасы құрылды, онда 100 сөйлеушінің жазбалары бар. Динамиктерге тән спектрограммаларды алу үшін ерекшеліктерді шығарудың тиімді әдісі ұсынылды. Содан кейін ұсынылған AC-SOM моделінің тану тиімділігі мен жылдамдығын тексеру үшін сипаттамалық спектрограммалар қолданылды. Эксперимент нәтижелері көрсеткендей, сөйлеушіге тән спектрограмма оның айтылу бөлшектерін ғана емес, сонымен қатар тұрақты айтылу сипаттамаларын да көрсете алады, осылайша сөйлеушінің айтылу сипаттамаларын тиімді сипаттай алады. Эксперименттік нәтижелер замануи алгоритмдермен салыстырғанда, ұсынылған AC-SOM алгоритмінің тану жылдамдығына айтарлықтай әсер етпестен

оқу және тану жылдамдығын айтарлықтай жақсарта алатынын көрсетеді. Осылайша, бұл зерттеу ресурстары шектеулі құрылғыларда динамиктерді танудың озық интеллектуалды жүйелерін енгізудің өте перспективалы нұсқасын ұсынады.

**Әдебиеттер**

1. Saritha B., Laskar R.H., Choudhury M., Anish Monsley K. Optimizing Speaker Identification through SincsquareNet and SincNet Fusion with Attention Mechanism // Procedia Computer Science. - 2024. -Vol.233. -P.215-225. DOI 10.1016/j.procs.2024.03.211.

2. El Shafai W., et al. Optical ciphering scheme for cancellable speaker identification system // Computer Systems Science and Engineering. -2023. -Vol.45(1). -P.563-578. DOI 10.32604/csse.2023.024375.

3. Daqrouq K., Tutunji T.A. Speaker identifcation using vowels features through a combined method of formants, wavelets, and neural network classifers // Elsevier Science Publishers B. V. - 2023. -Vol.27. -P. 231-239. DOI 10.1016/J.ASOC.2014.11.016.

4. Ajmera P.K., Jadhav D.V., Holambe RS. Text-independent speaker identification using radon and discrete cosine transforms based features from speech spectrogram // Pattern Recognition. -2011. - Vol.44(10-11). -P.2749-2759. DOI 10.1016/j.patcog.2011.04.009.

5. O' Shaughnessy D. Speech Communications: Human and Machine. – 2nd ed. / IEEE Press. -2008. -600 p. ISBN 978-0-780-33449-6.

6. Cheng F., et al. Visual speaker authentication with random prompt texts by a dual-task CNN framework // Pattern Recognition. -2018. -Vol.83. -P.340-352. DOI 10.1016/j.patcog.2018.06.005.

7. Hinton G., Deng L., Yu D., Dahl G.E., et al. Deep neu ral networks for acoustic modeling in speech recognition: the shared views of four research groups // IEEE Signal Process Magazine. - 2012. -Vol.29(6). -P.82-97. DOI 10.1109/MSP.2012.2205597.

8. Mamyrbayev O., Toleu A., Tolegen G., Mekebayev N., Neural architectures for gender detection and speaker identification // Cogent Engineering.- 2020. -Vol.7(1): 1727168. DOI 10.1080/23311916.2020.1727168.

9. Kalimoldayev M.N., Mamyrbayev O.Zh., Kydyrbekova A.S., Mekebayev N.O. Voice verification and identification using i-vector representation // International Journal of Mathematics and Physics.- 2019. - Vol.10(1). -P.66-74. DOI 10.26577/ijmph-2019-i1-9.

10. Mamyrbayev O., Turdalyuly M., Mekebayev N., Alimhan K., Kydyrbekova A., Turdalykyzy T. Automatic recognition of Kazakh speech using deep neural networks //Lecture Notes in Computer Science. -2019. -Vol.11432. -P.465-474. DOI 10.1007/978-3-030-14802-7_40.

***Авторлар туралы мәліметтер***

Мекебаев Н.О. - PhD, қауымдастырылған профессор, Қазақ ұлттық қыздар педагогикалық университеті, Алматы, Қазақстан, e-mail:nurbapa@gmail.com;

Даркенбаев Д.К. - PhD, доцент м.а., әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан, e-mail: dauren.kadyrovich@gmail.com;

Орынтаева Ж.А. - магистр, аға оқытушы, Қазақ ұлттық қыздар педагогикалық университеті, Алматы, Қазақстан, e-mail: zannaoryntaeva0@gmail.com;

Модовов Н.А. - докторант, әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан e-mail: modovov@gmail.com.

***Information about the authors***

Mekebayev N.- PhD, Associate Professor, Kazakh National Women' s Teacher Training University, Almaty, Kazakhstan, e-mail: nurbapa@gmail.com;

Darkenbayev D.- PhD, Acting Associate Professor, Al-Farabi Kazakh National University, Almaty, Kazakhstan, e-mail: dauren.kadyrovich@gmail.com;

Oryntaeva Zh.- master, Kazakh National Women’ s Teacher Training University, Almaty, Kazakhstan, e-mail: zannaoryntaeva0@gmail.com;

Modovov N.- master, doctoral student, Al-Farabi Kazakh National University, Almaty, Kazakhstan, e-mail: modovov@mail.ru.