

## Нейрондық желі көмегімен дауыстық интерфейс жүйесін әзірлеу

ЖАРИЯЛАНДЫ  
07.04.2023

СІЛТЕМЕ  
<https://bilimger.kz/136123/>

Есенғалиева Жанна Сержановна, PhD, доцент,

Аскербай Дариға Бақдәулетқызы, магистрант,

Л.Н.Гумилев атындағы Еуразия ұлттық университеті, Қазақстан, Астана, Пушкин 11.

**Аңдатпа.** Мақалада бүгінгі таңда, өзекті болып табылатын сөйлеуді автоматты тану жүйесі мен алгоритмдері қарастырылған. Дауыстық интерфейс жүйесі қазіргі уақытта дербес компьютерлерде, интернет-браузерлерде, нақты уақыт жүйелерінде, барлық гаджеттерде кеңінен күнделікті қолданылады. Олар жағдайды басқаруға, шешім қабылдауға, қажетті ақпаратты іздеуге, қажетті деректерді алуға және т. б. мүмкіндік береді. Сонымен қатар, Яндекс пен Google-ден сөйлеуді тану функциясын жүзеге асыруды дауыстық сөйлеу арқылы ақылды үйді басқару сияқты әртүрлі жобаларда қолдануға болады. Олар жағдайды бақылауға, шешім қабылдауға, қажетті ақпаратты іздеуге, қажетті деректерді алуға және т.б. мүмкіндік береді. Сөйлеуді тану жүйелері мүгедектер үшін арнайы қызметтер ретінде дами бастады.

Қазіргі заманғы ақпараттық технологияларды сәтті іске асыруда тиісті бірыңғай ақпараттық кеңістіктер маңызды рөл атқарады. Киберфизикалық және мобильді робототехникалық жүйелердің пайда болуына және кеңінен енгізілуіне байланысты дыбыстық интерфейстер арқылы осы жүйелердің тиімді өзара әрекеттесуін қамтамасыз ететін бірыңғай ақпараттық кеңістікті қалыптастыру мәселелері ерекше өзекті болып табылады.

Зерттеу жұмысында сөйлеуді автоматты түрде тану арқылы дауыстық сөйлеуді мәтінге айналдыру мәселелері көрсетілген. Сөйлеуді тануда кеңінен қолданылатын тәсілдер ретінде нейрондық желілер зерттеліп, сөйлеу белгілерін, сондай-ақ, сөйлеушіні анықтайтын модель мен алгоритм құру жұмысы жүргізілді. Сонымен қатар, осы саладағы еңбектер қарастырылып, толықтай библиографиялық шолу жасалынды. Python 3.10 ортасында сөйлеуді танудың бағдарламалық жасақтамасын әзірлеуде тәжірибе жүргізіліп, қосымша SpeechRecognition және PyAudio кітапханалары пайдаланылып, сонымен қатар микрофон түріндегі қосымша жабдықтар қолданылды. Сөйлеуді тану әдістеріне талдау жасалып, сөйлеуді автоматты түрде тану

программасын жасау арқылы зерттеу жұмысы жүргізілді.

**Кілттік сөздер:** сөйлеуді тану, Марковтың жасырын модельдері, нейрондық желілер, деректерді интеллектуалды талдау, Python кітапханалары.

**Кіріспе.** Компьютерлік технологияны дамытудағы маңызды мәселенің бірі адам-машина интерфейсінің құралдарын теориялық және практикалық негіздеу. Сөйлеуді тану өзекті мәселенің бірін шешу мақсатында қолданылатын көптеген заманауи әдістер кең ауқымды есептеуіш ресурстарын талап етеді, ал көп жағдайда ол ресурстардың көлемі шектеулі. Ауызша сөйлеуді танудың белгілі әдістері мен алгоритмдерін қолдану арқылы, деректерді интеллектуалды және математикалық талдап, сөйлеуде тақырыптың контентін автоматты түрде танытын бағдарлама әзірлеу.

Л.Р. Рабинердің “A tutorial on hidden Markov models and selected applications in speech recognition” оқу құралында Марковтың жасырын модельдері мен сөйлеуді танудағы таңдалған қосымшалар туралы сипатталған. Марковтың модельдерінің негізгі теориясы және әдістері туралы практикалық мәліметтерге, сондай-ақ, сөйлеуді танудың әртүрлі мәселелерін шешуге бағытталған [1]. Осы саладағы салада зерттеу жұмысын жүргізіп, жалғастыруда қажетті ақпаратты алудың бірыңғай көзін қамтамасыз етеді. Теория монеталарды лақтыру және классикалық шарлар жүйесі арқылы сипатталған. Автор Марковтың дискретті тізбектерінің теориясын және ықтималдық функциясы болып табылатын жасырын модельдер ұғымын қалай тиімді пайдалануға болатындығын көрсетеді.

“Development of an Intelligent Mobile Application for Teaching English Pronunciation” мақаласы ағылшын тілінің айтылуын үйретуге арналған интеллектуалды мобильді қосымшаны әзірлеуге негізделген. Н. Кавустың мақсаты тілді үйренуде, мобильді құрылғылардағы сөйлеуді тану механизмін қолдана отырып интеллектуалды мобильді қосымшаның дамуын сипаттауда [2]. Зерттеуде әзірленген қосымша арқылы ұялы телефонда айтылған сөздерді тануда сөйлеуді тану механизмі қолданылған, сондықтан кез-келген қателіктерді оңай анықтап, түзетуге болады.

А. Н. Романенконың “Feature combination for the task of neural network acoustic model learning” атты жұмысында сөйлеуді тану сапасын жақсартуға бағытталған нейрондық желінің акустикалық модельдерін оқыту мәселесі үшін функцияларды біріктіру әдісі ұсынылған [3]. Нейрондық желі кірісіне басқа сипаттағы белгілердің біріктірілген векторын беруден айырмашылығы, ұсынылған әдіс жасырын қабаттар деңгейінде кешіктірілген біріктіруді пайдаланады. Ол мүмкіндіктердің әрбір түрі үшін жеке кіріс ағындарын пайдаланады. Мұндай ағындар мүмкіндіктердің әрбір түрі үшін нақты үлгілерді шығарып, содан кейін оларды нейрондық желінің акустикалық моделінің жасырын қабатында біріктіре алады. Жұмыстың нәтижелерін автоматты түрде сөйлеуді тану жүйелерін әзірлеу үшін пайдалануға болады.

Импульсті байланысы бар нейрондық желіні қолдана отырып, электрлоринкс дауысын тану мәселесін Ф. Арифин өзінің “Electrolarynx Voice Recognition Utilizing Pulse Coupled Neural Network” еңбегінде қарастырды [4]. Автор мақаласында дауысты тану технологиясын пайдаланып, қалыпты

сөйлей алмайтын науқастарға электроларингологиялық сөйлеуді қолдану және сөйлеу құралдың пайдасын анықтап көрсеткен. Электроларинкс сөйлеуді тану жүйесі сипатталып, жүйенің екі негізгі бөлігі – белгілерді алу және үлгіні тану көрсетілген. Импульстік байланысы бар нейрондық желі-PCNN электрлік ларинкстің сөйлеу ерекшеліктері мен сипаттамаларын алу үшін қолданылады. Сонымен қатар, спикерді, яғни белгілі бір адамның нақты сөйлеуін танып, жақсы жұмыс істейтін жүйені көрсеткен.

Автоматты сөйлеуді тану инвестициялар жасайтын коммерциялық қосымшаларға негізделгені жайлы Тампель И.Б. және Карпов А.А. жұмыстарында айтылған [5]. “Автоматическое распознавание речи” оқу құралында мұндай қосымшалардың ішінде, ең алдымен, call-орталықтардың немесе IVR-жүйелердің операторды айналып өтіп, ақпаратқа автоматты түрде қол жеткізе алатындығы баяндалған. Сөйлеуді автоматты тану жүйелері ақпаратты енгізуді талап ететін медициналық зерттеулерде немесе автономды басқару қажет болғанда кеңінен қолданылады. Оқу құралында сөйлеу мәселелеріне арналған қабылдау алға қойылған.

Сөйлеу ерекшеліктері визуалды сөйлеу саласында үнемі қызығушылық тудырады. Мүмкіндіктерді алу арқылы пайдалану жақсы нәтиже бере алады, бірақ адамдарға нәтижелерді түсіну және байланыстыру қиынға соғады. Адамның көзқарасына негізделген бет штрих-кодтарын психологиялық зерттеуге негізделген мүмкіндіктерді алудың жаңа, қарапайым тәсілін Сюэцзе Чжан “Visual Speech Recognition with Lightweight Psychologically Motivated Gabor Features” еңбегінде ұсынған [6]. Бұл тәсіл минималды өңдеу кезінде төмен өлшемді ерін параметрлерін сәтті алуға мүмкіндік береді. Осы Gabor негізіндегі функцияларды пайдалану мен дәстүрлі DCT немесе CNN функциялары үшін қазіргі заманғы сән сияқты басқа мүмкіндіктерді пайдалану арасындағы басты айырмашылықтардың бірі – бұл адамдарға бағытталған.

Автор Руди Адипраната “Implementasi Sistem Pengenalan Suara Menggunakan SAPI 5.1 Dan Delphi 5” деген жұмысында SAPI 5.1 Dan Delphi 5 көмегімен дауысты тану жүйесін енгізу жайлы пайымдайды [7]. Сөйлеуді тану саласындағы әйгілі зерттеулердің бірі – Windows платформасына арналған Microsoft корпорациясы. Microsoft сонымен қатар сөйлеуді тану механизмдері үшін стандартты әзірлейді, атап айтқанда SAPI (сөйлеу қосымшаларын бағдарламалау интерфейсі) COW көмегімен бағдарламашы сөйлеуді танудың кез-келген механизмін қолдана отырып, сөйлеуді тануды жүзеге асыра алады, сонымен қатар бағдарламалық кодты өзгертпестен механизмді өзгерте алады. Бұл мақалада SAPI 5.1, Microsoft Speech Engine және Delphi 5 көмегімен сөйлеуді тану жүйесін ағылшын тілінде дыбысты тануға арналған мәтіндік қосымшаны құру үшін бағдарламалау тілі қолданылған.

Сөйлеуді автоматты тану жүйелерін өнеркәсіптік қолдану жайында “Industrial Applications of Automatic Speech Recognition Systems” жұмысында баяндалып, негізделген [8]. Сөйлеуді автоматты түрде тануды, синтетикалық сөйлеуді және табиғи тілді өңдеуді қамтитын сөйлеу сигналдарын өңдеу бизнеске, өнеркәсіпке және жеке компьютерлердің жұмысына айтарлықтай әсер ете бастайды. Сонымен қатар, бұл адам миының жұмыс істеуінің күрделі механизмін тереңірек түсінуге ықпал етеді. Сондай-ақ, бұл өнеркәсіптік қосымшалардағы сөйлеуді тану жүйелерінің

уақыт өте келе дауыстық қызметтердің жаңа буынын ашу үшін қалай дамитынын көрсетеді. Бұл мақаланың мақсаты – қол жетімді әдебиеттерде сипатталған сөйлеуді тану технологиясына тиімді шолу жасау және жеке зерттеулер мен әзірлемелерді зерттеу барысында алынған идеяларды біріктіру. Сөйлеуді тану қосымшалары медицина, өнеркәсіптік робототехника, қорғаныс және авиация саласындағы қосымшаларға ерекше назар аудара отырып сипатталған.

Адамның сөйлеуін тану бұрыннан бері жасанды интеллект пен деректерді зерттеушілер арасында қызықты мәселесі бұл оқулықта қарастырылған. Сөйлеуді тану әдістеріне сыни шолу және талдау “A critical review and analysis on techniques of speech recognition: The road ahead” еңбегінде баяндалады. Бұл зерттеу жарамды сөйлеуді анықтау стратегиясына шолу жасауды қарастырады [9]. Бұл зерттеудің негізгі мотивациясы – зерттеушілер осы саладағы барлық қажетті көрсеткіштерді қосып, қолданыстағы шектеулерді жеңе алатындай етіп сөйлеуді танудың қолданыстағы стратегияларын зерттеу. Бұл шолуда сөйлеуді тану әдіснамасына енгізілген әртүрлі сұрақтар ерекшеленеді және белгілі бір жүйеге қандай қасиеттер тән және қайсысы еленбейтінін анықтау үшін сөйлеуді танудың ерекше процедуралары зерттелген.

М.Лутц “Изучаем Python” кітабінің 4-ші шығарылымында сонымен қатар, “Программирование на Python” кітабінің I томында Python тіліне толыққанды шолу жүргізілген [10-11]. Оқу құралында, Python тілін оқытудан бастап, түрлі қолданбалы бағдарламаларды әзірлеу оңай жолмен түсіндіріліп, оны қолдану саласы мен мүмкіндіктері көрсетілген. Кітап практикалық курстардың материалдарына негізделген, бақылау сұрақтары мен түрлі дәрежедегі жаттығуларды қамтиды. Бағдарламалау тілі негізгі қолданбалы салаларда – атап айтқанда, жүйелік басқаруда, графикалық интерфейстер мен веб-қосымшаларды құруда қолданылады. Нұсқаулықта, деректер базасымен жұмыс істеу, желінің өзара әрекеттесуін бағдарламалау, сценарийлер үшін интерфейстер құру, мәтінді өңдеу және т.б. зерттелген. Басылымда тілдің синтаксисі мен даму әдістерін сипаттай келе, типтік бағдарламалау идиомаларын және олардың дұрыс қолданылуын суреттейтін көптеген мысалдар қамтылған.

Сөйлеуді танудың тәсілдері мен алгоритмдері жайлы Ж.С.Есенғалиева және А.Ж.Кабатованың өздерінің “Исследование методов и алгоритмов распознавания голосовой речи” мақаласында сипаттайды [12]. Зерттеу жұмысында жасырын Марков модельдері мен нейрондық жүйелер әдістері дауыстық сөйлеуді автоматты тану мәселесінде пайдаланылып, қарастырылған. Баяндамада бағдарламалық жасақтама арқылы дауыстық сөйлеуді танудың теориялық және практикалық аспектілері зерттелген.

Осылайша, жоғарыда көрсетілген дерек көздерін қорыта келе, дауыстық сөйлемнен контентті тану тақырыбына шолу жасадық. Нақты айтсақ, Марковтың жасырын модельдерін және нейрондық желілерді зерттеу жұмысында толығымен қарастырып өттік. Ақпараттық технологиялардың қарқынды дамуымен байланысты ақпарат алмасуда дауыстық интерфейс жүйесі кеңінен қолданылады. Осыған байланысты, сөйлеуді танудың дамуы маңызды рөл атқарады.

**Зерттеу әдістері.** Зерттеу жұмысы барысында сөйлеудегі контентті автоматты тануда

нейрондық желілер әдістері қолданылады. Қазіргі таңда машналық оқыту алгоритмдерін қолдану кең етек алуда. Машиналық оқыту алгоритмдері бұл деректер үлгілерін іздеу және тиісінше бағдарлама әрекеттерін өзгерту үшін қолданады.

Сөйлеуді танудың кейбір кезеңдері үшін нейрондық желілерге негізделген әдістер жиі қолданылады. Себебі, нейрондық желілерді қолдану қарапайым және жұмыс жылдамдығының жеткілікті түрде жоғары. Сөйлеу сигналын акустикалық-фонетикалық модельдеу үшін қолданылатын модельдердің басқа классы болып жасанды нейрон желілерінің модельдері. Оның құрылымы мен жұмыс істеу принциптері нерв жүйелерінің биологиялық модельдеріне, әсіресе ми моделіне негізделеді.

Жасанды нейрондық желілер технологиясы сөйлеуді тану және сегментацияның ұқсас мәселелерін тануға қолданылады. Нейрондық желі арқылы дауысты тану – адам сөйлейтін сөзді дұрыс анықтауға қабілетті модель жасалу арқылы жүзеге асады. Соңғы үлгіні алу үшін нейрондық желі деректердің үлкен корпусында оқытылды және желі мақсатты деректерге орнатылды.

Заманауи сөйлеулерді тану жүйелерінің архитектураларын қарастырсақ:

- көп қабатты перцептрондар (Multilayer Perceptron; MLP) немесе терең нейрондық желілер (Deep Neural Networks; DNN);
- жиналмалы нейрондық желі (Convolutional Neural Networks), CNN);
- рекурренттік нейрондық желілер (recurrent Neural Networks, RNN);
- ұзақ және қысқа мерзімді жадысы бар нейрондық желілер (Long Short Term Memory: LSTM)];
- басқарылатын рекурренттік блок (gated Recurrent Unit; GRU);
- екі бағытты рекуррентті нейрондық желілер (Bidirectional Recurrent Neural Networks BRNN)];
- қалдықты желілер (Residual Networks).

Нейрон желілері өз бетінше ұйымдастырылып, реттілігіне қарай алгоритмдердің көптүрлігі ретінде қаралады, өзара байланысқан көптеген бір типті және бір қатарда қызмет ететін элементтер немесе нейрондық және арнайы ұйымдасқан байланыстар көмегімен жүзеге асырылады. Сөйлеуді танудағы нейрон желілерінің құрылымдық схемасы көпқабатты перцептрон (КП) ең белгілі және мейлінше кең таралған нейрон желінің моделі болып табылады.

Табиғи тілді өңдеу. Математика, информатика және лингвистика салаларында маңызды рөл атқаратын табиғи тілді өңдеуді тақырыбы (Natural Language Processing, NLP) қарастырылады. Бұл тәсіл арқылы компьютер мәтіндерді түсініп және анализ жасай алады. Енгізілген тексттің тақырыбын анықтауда қолданылатын модель. Қарастырылып отырған тексттердің жиынтығы әдетте корпус деп аталады. Табиғи тілді өңдеуге арналған кітапханалар пакетін Natural Language Toolkit немесе NLTK және т.б. Python кітапханалары импортталады.

Бастапқыда текстті сөйлемдерге бөлу немесе токенизациялау қолданылады. Бірақ, тыныс беліглерін, қысқартулар қолданылуда, сондай-ақ, бос орын қалып кету мәселесі болғандықтан, өңдеуде қателіктер кетуі мүмкін. Бұл мәселені шешу үшін NLTK sent\_tokenize кітапханасының стандартты әдісі пайдаланылады. Лемматизация, стемминг және сөз қаптары тәсілдері арқылы анализ жасалып, тексттегі сөздерді бастапқы формаға келтіру, сонымен қатар, мәтінде жиі кездесетін сөздер қатарын көрсету мәселесі бойынша талдау нәтижесінде текст тақырыбын анықтауға болады.

Марковтың жасырын модельдері мен жасанды нейрондық желілердің комбинациясынан тұратын бірқатар гибриді жүйелер бар, олар модельдеудің екі әдісін де қолданады.

**Нәтижесі.** Математикалық және үлкен деректерді талдау әдістерін, сигналдарды сандық өңдеу тәсілдерін пайдалана отырып, программалық қамтамасыздандыру әзірленді. Бағдарламалық жасақтама Python ортасында әзірленіп, бірнеше сөйлеуді тану API қолдау механизмін қамтитын SpeechRecognition және микрофон мен динамик арқылы дыбысты енгізу/шығару операциялары үшін PyAudio кітапханалары қолданылды.

Нәтижесінде SpeechRecognition, PyAudio кітапханаларын пайдаланып дауыстық тану программа ретінде алгоритмі зерттелді.

**Қорытынды.** Сонымен, жоғарыда көрсетілген әдебиеттерді қорыта келе, дауыстық сөйлемнен контентті тану тақырыбына шолу жасап, түрлі әдіс-тәсілдермен, алгоритмдерімен таныстық. Алдағы уақытта дауыстық сөйлемнен контентті танитын программалық қамтамасыздандыру әзірлеу барысында қарастырылған тәсілдер пайдаланылады. Сөйлеуді автоматты танудың міндеті мен жалпы құрылымдары сипатталып, нейрондық желілер әдістері қарастырылды. Табиғи тілді өңдеу саласындағы текст тақырыбын анықтау мәселесі бойынша зерттеу жұмыстары жүргізілді. Дыбысты тану алгоритмі мен деректерді талдау арқылы программалық қамтамасыздандыру әзірленіп, талдау нәтижесі ұсынылады. Даусытық интерфейстер қазіргі заманғы интеллектуалды киберфизикалық жүйелерде қолданылады.

#### ӘДЕБИЕТТЕР ТІЗІМІ

1. Л.Р. Рабинер, "A tutorial on hidden Markov models and selected applications in speech recognition", in Proc. IEEE'89, vol.77, 1989, 257 - 286 б.
2. Н. Кавус, "Development of an Intellegent Mobile Application for Teaching English Pronunciation", Elsevier: Procedia Computer Science, vol.102, 2016, 365-369 б.
3. А. Н. Романёнок, Feature combination for the task of neural network acoustic model learning. Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 2018, vol. 18, №2, 350-352 б (Ресей). doi: 10.17586/2226-1494-2018-18-2-350-352.
4. Ф. Арифин, "Electrolarynx Voice Recognition Utilizing Pulse Coupled Neural Network", June 2010 IPTEK The Journal for Technology and Science 21(3), vol. 21, №3. DOI:10.12962/j20882033.v21i3.45.

5. И.Б. Тампель, А. А. Карпов, “Автоматическое распознавание речи”, Оқу құралы. — Санкт-Петербург: Университет ИТМО, 2016, 138 б.
6. С. Чжан, Й. Сюй, А.К.Абель, Л.С.Смит, Р. Ватт, А Хуссейн, К. Гао, “Visual Speech Recognition with Lightweight Psychologically Motivated Gabor Features”, Entropy, 2020, Art. №1367. <https://doi.org/10.3390/e22121367>.
7. А. Нортджахьяна, Р. Адипраната, “Implementasi sistem pengenalan suara menggunakan sari 5.1 dan Delphi 5”, Информатика журналы, North America, 4-маусым,
8. Jayashri Vajpai et al. Int. Journal of Engineering Research and Applications, ISSN: 2248-9622, Vol. 6, Issue 3, (1-бөлім) наурыз 2016, 88-95 б.
9. Р. Адипраната, “A critical review and analysis on techniques of speech recognition: The road ahead”, Халықаралық білімге негізделген және интеллектуалды инженерлік жүйелер журналы, 22(1):39-57. DOI:10.3233/KES-180374, 2018, 39-57 б.
10. М. Лутц, “Изучаем Python”, 4-ші шығарылым. – Пер. ағыл. – СПб.: Символ-Плюс, – 1280 с.
11. М. Лутц, “Программирование на Python”, I том, 4-ші шығарылым. – Пер. ағыл. – СПб.: Символ-Плюс, 2011. – 992 б.
12. Ж. Есенғалиева, А. Кабатова, “Исследование методов и алгоритмов распознавания голосовой речи”, Алматы: Вестник КазННТУ. — №3 (66), 2020, 593–601 б.
13. Веб-ресурс: <https://www.dmitrymakarov.ru/intro/topic-identification-19/>

**ҚМ АА** Куәлік нөмірі: **KZ45VPY00102718** — ҚР Мәдениет және Ақпарат министрлігі

© 2026 **Bilimger.kz** Ақпараттық-танымдық білім порталы. Барлық мазмұн авторлық құқықпен қорғалған.