

6D070400 – Есептеу техникасы және бағдарламалық қамтамасыз ету
мамандығының PhD докторанты

Тұрдалыұлы Мұсаның

«Терең рекуррентті нейрожелілік модельдерді қолдану арқылы кіріккен
сөйлеуді толассыз тану»

тақырыбындағы диссертациялық жұмысына

АНДАТПА

Зерттеу тақырыбының өзектілігі. Ғылым мен техниканың дамуы адам мен машинаның өзара әрекеттесу құралдарының эволюциясымен тығыз байланысты. Қазіргі әлемде адам-машина өзара әрекеттестігінде сөйлеу интерфейсі барған сайын танымалдылыққа ие болуда. Бұл сөйлеудің адам үшін ең табиғи қарым-қатынас құралы болып табылатынына негізделеді. Адам-машина интерфейсіндегі сөйлеуде аса маңызды құраушы ретінде сөйлеуді автоматты тану жүйесі танылады.

Сөйлеуді автоматты тану саласындағы аса күрделі міндеттерінің бірі – айтылым кезіндегі сөйлеуді тану. Айтушылардың тікелей қатысуымен болатын, бірақ ауызекі хабарламаның түрі мен мазмұны алдын ала дайындалмауымен жүзеге асуы – сөйлеу стилін сипаттайды. Бұл міндеттердің күрделілігі айтылым кезіндегі сөйлеудің төмендегідей ерекшеліктерінен туындайды: айтушылар яғни дикторлар арасындағы елеулі вариативтік, сөйлеу тіліндегі акценттік және көңіл – күйлік сипаттардың бар болуы, пайдаланылатын сөз тұлғаларының әралуандылық санының көптігі. Хезитацияның бар болуымен міндет күрделілігі арта түседі. Хезитация дегеніміз: сөйлеу тілінің айтылу кезінде жүзеге асуымен байланысты пайда болатын толқымалар, оған жататындар: үзілістер, лексикалық емес кірме бөгде дыбыстар, «паразит сөздер», сөйлемдердің бұзылулары, сөздердің алмасулары, қайталаулар, тұтықпалылықтар, аяқталмаған тиянақсыз сөйлемдер. Сөйлеу тілі қарым-қатынасындағы жағдаяттарда алдымен айтылым кезіндегі сөйлеу тілі пайда болатынын атап айту керек. Сондықтан оны тану міндетінің көкейкестілігі арта түседі.

Кіріккен сөйлеуді тану жүйесі – ерекше сұранысқа ие және өзекті мәселелердің бірі. Мысалы, ақпараттық байланыс құралдарындағы немесе әр түрлі отырыстардағы сөйлеудің үлкен архивтерін сараптау. Алайда сөйлеушінің сөйлегендегі ортасының өзінің, сөйлеуді тану жүйесі жұмысының сапасын төмендететін әр түрлі ерекшеліктері бар. Олар: 0-4000 Гц жиілік диапазонының өткізу жолақтарының шектеулілігі, аддитивтік және арнаға тән емес ауытқулардың болуы, сондай-ақ, сөйлеу сигналдарының кодталу үрдісінде ақпараттардың жоғалуы. Бұл ерекшеліктер кіріккен сөйлеулерді тану міндетін қосымша қиындата түседі.

Ағылшынша айтылым кезіндегі сөйлеуді тану зерттеулерінде Switchboard-1 (300 сағаттық) ағылшын тіліндегі корпус, Фишер корпусы (2000 сағаттық) т.б. корпустар пайдаланылады. Ағылшын тілін зерттеушілердің көбі Лингвистикалық Мәліметтер Консорциумы (Linguistic Data Consortium, LDC)

дайындаған HLIB5 Evai 2000 тестілік таңдау нәтижелеріне үлкен көңіл бөлген. Осы базаларды IBM компаниясының зерттеушілері Brian Kingsbury, George E. Dahl және Microsoft-тан Li Deng, Dong Yu, Frank Seide, Google компаниясынан Andrew Senior, Tara Sainath т.б. қолданды. Қазіргі таңда әрбір мемлекет өз тілін тану процессіне ғалымдарын жұмылдыруда. Бүгінгі таңда ағылшынша сөйлеуді танудың озық жүйелері – танудағы қателіктердің деңгейін 15%-ға түсіруге мүмкіндік берді.

Орысша сөйлеу тілінің айтылым кезіндегі сондай-ақ біртұтас үлгілерін тануға Ресей ғылым академиясының Санкт-Петербург ақпараттық және автоматизация институтының зерттеушілері Андрей Ронжин, Алексей Карпов және Ресей елінің басқада ғалымдарының еңбектері арналған.

Қазіргі таңда біздің елімізде қазақ сөйлеулерін тану мәселелерімен Л.Н. Гумелев атындағы Еуразиялық ұлттық университетінің ғалымдары А.Ә. Шәріпбай, Г.Т. Бекманова, әл-Фараби атындағы Қазақ ұлттық университетінің ғалымдары У.А. Тукеев, Д.Р. Рахимова, ҚР БҒМ ҒК Ақпараттық және есептеуіш технологиялар институтының ғалымдары Е.Н. Амиргалиев, Р.Р. Мусабаев т.б. ғалымдар зерттеу жүргізуде.

А.Ә. Шәріпбай өзінің шәкірттерімен қазақ тілінің формальдық фонетика заңдылығына негізделген математикалық теориясын, морфологиялық және синтаксистік тәртібін, сөзді синтездеу және талдау, сөйлеу тану алгоритімі мен қосымшасын ұсынды.

Қазақ сөйлеулерін тану саласындағы зерттеулер кіріккен сөйлеудің мағынасын әлі де жеткізе алмайтындығы анықталды. Бұл жағдай қазақша айтылым кезіндегі сөйлеуді тану жүйесінің деңгейіне сәйкес келетін қазақша айтылым кезіндегі кіріккен сөйлеуді тану жүйесі әлі де болса жасалына қоймағандығын аңғартады. Қазақша айтылым кезіндегі кіріккен сөйлеуді танудағы тиімділіктің жеткілікті болмауының бірнеше себептерін айрықша көрсетуге болады. Біріншіден, қазақша айтылым кезіндегі сөйлеуді тану жүйесінің сапасын бағалау үшін керекті корпустың қолжетімді түрде ашық болмауы; екіншіден, қазақ тілі агглютинативті тіл ретінде, аналитикалық тілдерге қарағанда, едәуір көп сөзтұлғаларға ие. Қазақша айтылым кезіндегі сөйлеу тілі бірнеше ондаған мың сөзбен тоқтаса, бұл жағдайда қазақ тілінде жүздеген мың сөзді қамтитын сөздік қажет болады. Үшіншіден, қазақша айтылым кезіндегі сөйлеуде артикуляциялық әлсіз көрініс, ассимиляциялық құбылыс (дыбыстардың кірігуі), редукциялану (дыбыс ұзақтығының қысқаруы) сияқты фонетикалық ерекшеліктерді атауға болады. Бұл фонетикалық ерекшеліктерден туындайтын сөйлеудің айырымдық белгілерінің акустикалық варианттылығына табанды тірек болатын тану жүйесін жасау қазақ тілі үшін қажет-ақ.

Жоғарыдағыларды ескере отырып, қазақша айтылым кезіндегі кіріккен сөйлеуді танудың дәлдігін арттыруды қамтамасыз ететін әдістер, алгоритмдер мен бағдарламалық құралдардың қажеттілігі **аса өзектілікті** деген қорытынды жасауға болады.

Диссертациялық жұмыстың мақсаты. Қазақша айтылым кезіндегі кіріккен сөйлеуді танудың дәлдігін арттыруды қамтамасыз ететін әдістер,

алгоритмдер мен бағдарламалық құралдар дайындау және оны жылдам әрекет жасайтын, практикалық міндеттерде жеткілікті қолданылатын жүйеде жүзеге асыру.

Зерттеудің міндеттері. Зерттеуге қойылған мақсаттарға жету үшін келесі мәселелерді шешу қарастырылады:

1. Айтылым кезіндегі сөйлеуді танудың қазіргі әдістерін сараптауды жүргізу.

2. Қазақ тілінің кіріккен сөйлеуінің акустикалық корпусын және мәтіндік корпусын құру.

3. Қазақша айтылым кезіндегі кіріккен сөйлеуді тану жүйесінің құрамына енетін акустикалық моделін, тіл моделін және транскрипциялар сөздігін құру.

4. Қазақша айтылым кезіндегі кіріккен сөйлеуді тану жүйесінің тану сапасын бағалау, сондай-ақ басқа жүйелермен салыстыру.

5. Қазақша айтылым кезіндегі кіріккен сөйлеуді танудың алгоритмдерін және бағдарламалық құралдарын құру.

Зерттеу объектісі. Сөйлеуді автоматты тану жүйелері.

Зерттеу пәні. Қазақша айтылым кезіндегі кіріккен сөйлеуді автоматты танудың әдістері, алгоритмдері және бағдарламалық құралдары.

Зерттеу әдістері. Қолданбалы ғылыми зерттеулерде кең тараған: мақсаттар мен міндеттерді рәсімдеу, зерттелу жайын және бар әдебиетті сараптау, алгоритмдік және бағдарламалық шешімдерді әзірлеу, әзірленген шешімдердің тиімділігін бағалау, нәтижелерді апробациялау және сараптау қолданылды. Зерттеудің эксперименттік бөлігін жүзеге асыруда ол тек табиғи тіл материалымен жүргізілді, мұнда тестілік сұрыптау не айтылу ерекшеліктері бойынша, не дикторлар құрамы ерекшеліктері бойынша оқытылатын берілгендермен тоғысқан жоқ.

Зерттеу әдістері ретінде сигналдарды сандық өңдеу әдістері, ықтималдық және математикалық статистикалық теориялар, машиналық оқыту, қолданбалы лингвистикалық, сондай-ақ бағдарламалық қамтамасыз етуді әзірлеу әдістері қолданыс тапты.

Ғылыми жаңашылдықтар:

– Қазақ тілінің кіріккен сөйлеуінің акустикалық корпусы және мәтіндік корпусы құрылды.

– Айтылым кезіндегі сөйлеу үшін акустикалық модельдердің сапасын арттыруға мүмкіндік беретін сөйлеу сигналдарын сығымдау алгоритмі құрылды.

– Қазақ тілінің кіріккен сөйлеуінің акустикалық және мәтіндік корпустарын қолданып алғаш рет терең рекуррентті нейрондық желі көмегімен қазақ тілінің кіріккен сөйлеуінің тілдік моделі құрылды.

– Диссертацияда ұсынылған алгоритмдер көмегімен құрылған акустикалық және тілдік модельдерді пайдалануға мүмкіндік беретін қазақша айтылым кезіндегі кіріккен сөйлеуді тану жүйесінің құрамына енетін бағдарламалық құралдар құрылды.

Жұмыстың теориялық және практикалық маңызы. Берілген жұмыстың теориялық маңызы сөйлеуді тану міндеттері үшін терең рекуррентті нейрондық желілердің негізіндегі оқытылатын акустикалық модельдердің қазіргі бар және жаңа алгоритмдерін жетілдіруде болып табылады, сондай-ақ бұрын пайдаланылып келе жатқан әдістерден де жоғары артықшылықтағы ақпараттық белгілерді туындатудың жаңа әдісін әзірлеумен және эксперименттік зерттеумен сипатталады.

Диссертациялық зерттеудің практикалық маңызы болып төмендегі нәтижелер саналады: қазақша айтылым кезіндегі кіріккен сөйлеу тану жүйесін құру кезінде әзірленген алгоритмдік және бағдарламалық құралдарды қолдану; фонограммалардың архивтерін автоматты түрде мәтіндеудегі, үзіліссіз сөйлесу тілі ағынындағы кілт сөздерді іздеп табудағы, жазылғандарды тақырыптық кластерлеулер сияқты практикалық міндеттердегі жеткілікті жоғары сапалы тану мен жылдам әрекет етуді қолдануды көрсете алады.

Қорғауға шығарылған негізгі тұжырым:

Қазақ тілінің кіріккен сөйлеуін толассыз тану тапсырмаларына терең рекурренттік нейрондық желілік модельдерді қолдану сөйлеуді автоматты тану жүйелерінің тану сапасын арттырады.

Зерттеушінің жеке үлесі. Ізденушінің жеке өзі диссертациялық жұмыстың қойылған міндеттерін шешті. Айтылым кезіндегі сөйлеу тілін танудың әдістері мен алгоритмдерін әзірледі. Әзірленген әдістер мен алгоритмдердің тиімділігін эксперименттік бағалау жүргізілді. Қазақша айтылым кезіндегі сөйлеуді тану жүйесінің құрамына енетін бағдарламалық құралдар әзірленді.

Диссертация тақырыбының ғылыми-зерттеу жұмыстарының жоспарларымен байланысы. Диссертация бойынша зерттеу жұмыстары Білім және ғылым министрлігі Ғылым комитеті Ақпараттық және есептеуіш технологиялар институтында «Терең нейрондық желілерді пайдаланатын мультитілдік автоматты сөйлеуді тану технологиясын құру» гранттық қаржыландыру жобасы аясында орындалды. (2018-2020 жж., мемлекеттік тіркеу нөмірі: 0118РК00139).

Жұмыс көлемі мен құрылымы. Диссертация кіріспеден, 4 тарау және қорытындыдан тұрады. Диссертацияның толық көлемі: 95 бет, 41 сурет, 6 кесте. Әдебиеттер тізімі 111 атаудан тұрады.

Кіріспе бөлімінде жұмыстың өзектілігі анықталды, зерттелетін тақырыпқа байланысты мәселелер көрсетілді. Жұмыстың идеясы, зерттеудің мақсаты мен міндеттері, жұмыстың ғылыми жаңалығы және практикалық құндылығы, зерттеу әдістері берілген.

Бірінші бөлімде сөйлеуді автоматты тану жүйелерінің классификациясы, олардың құру мәселелері мен қолданылған басқа модельдер көрсетілген.

Екінші бөлімде қазақ тілінің сөйлеу және мәтіндік корпустарын құру жұмыстары баяндалған. Алдымен сөйлеу және мәтіндік корпустарды құру кезеңдері анықталды. Қазақ тіліндегі мәтіндік ақпараттарды жинау процесі баяндалды. Акустикалық корпусын құру жұмыстары үшін сөйлеулерді

дыбыстаушы дикторлардың таңдалды, оларды идентификациялау бойынша жұмыстары жүргізілді. Сөйлеу және мәтіндік корпусы құру кезеңдері анықталды. Сөйлеу ақпараттарын жинақтау кезінде қолданылатын құралдары анықталды. Құрылған корпусстардың сөздерін қамтитын сөздік құру жұмыстары жүргізіліп, нәтижесінде, қазақ тілінің кіріккен сөйлеуін автоматты тану жүйелерін құруға арналған сөздік құрылды.

Үшінші бөлімде қазақ тілі сөйлеуінің корпусын қолданып, осы тілдің акустикалық және тілдік модельдерін құру жұмыстары жүргізілді. Сөйлеу сигналдарын сығымдау алгоритмі құрылып негізделді. Акустикалық және тілдік модельдерді құру құралы таңдалып, сол құрал көмегімен жүйені оқыту жұмыстарын жүргізу үшін қазақ тілінің сөйлеу корпусын құралға бейімдеу жұмыстары жүзеге асырылды. Жасырын Марков моделі мен Гаусс қоспаларын және терең нейрондық желілерді қолданып жүйені оқыту процесі сипатталды. Жүйені оқыту процессінен кейін алынған қазақ тілі сөйлеуінің акустикалық және тілдік модельдерінің сапасын анықтау мақсатында, монофондық, трифондық және терең нейрондық желілер мен терең рекуррентті желілер көмегімен тәжірибелік жұмыстар жүргізіліп, сөйлеу сапасының нәтижелері алынды.

Төртінші бөлім алдыңғы тарауларда жүргізілген тәжірибелік жұмыстар бойынша алынған терең рекуррентті нейрожелілік модельдер арқылы алынған акустикалық және тілдік модельдердің көмегімен қазақ тілі сөйлеуін толассыз танитын бағдарламалық кешен құрылды. Бағдарламалық кешен Windows және Linux типті операциялық жүйелерде орындалатын бағдарламалық қосымшадан, клиент-серверлік архитектура негізінде жұмыс істейтін Веб-қосымшасынан және Android және iOS жүйелерінде орындалатын мобильдік қосымшасынан тұрады. Осы бағдарламалық кешен құрамының қосымшаларының жұмыс істейтін функциялар ашып көрсетілген.

Қорытындыда диссертацияның негізгі нәтижелері мен тұжырымдары баяндалып, оның болашақ жұмыстармен байланысы көрсетілген.

Сенімділік дәрежесі мен апробациялау нәтижелері. Зерттеудің сенімділігі мен нәтижелерінің негізділігі міндеттердің қойылудың негізделінген жауапкершілігімен, критерийлердің және берілген саладағы зерттеулердің жай-күйінің сарапталуымен, жүргізілген эксперименттер санының көптігімен, сондай-ақ олардың практикаға табысты енгізілуімен қамтамасыз етіліп дәйектеледі. Диссертация нәтижелері мына төмендегі ғылыми-әдістемелік конференцияларда баяндалып талқыланды:

1. «Ақпараттық және телекоммуникациялық технологиялар: білім, ғылым, тәжірибе» атты II Халықаралық ғылыми-тәжірибелік конференциясында (Алматы, 3-4 желтоқсан, 2015);

2. «Төртінші өнеркәсіптік революция жағдайындағы дамудың жаңа мүмкіндіктері» атты ҚР Президенті Н. Назарбаевтың Жолдауын іске асыру шеңберінде «Көліктегі инновациялық технологиялар: білім, ғылым, тәжірибе» атты XLII Халықаралық ғылыми-практикалық конференциясында (Алматы, 18 сәуір, 2018);

3. «Кешенді жүйелерді оңтайландыру мәселелері» XIV халықаралық Азиялық семинары халықаралық ғылыми конференциясы (Чолпон-Ата, Қырғызстан, 2018);

4. Профессор Р.Г. Бияшевтың 80 жылдығына және профессор М.Б. Айдархановтың 70 жылдығына арналған «Информатика және қолданбалы математика» атты III Халықаралық ғылыми-практикалық конференциясында (Алматы, 26-29 қыркүйек 2018);

5. 3rd International Conference Applied Mathematics, Computational Science and Systems Engineering (Рим, Италия, 2018);

6. 11th Asian Conference on Intelligent Information and Database Systems (Yogyakarta, Индонезия, 6-12 сәуір 2019 ж.).

Диссертация тақырыбы бойынша 16 мақала жарияланды және 1 авторлық куәлік алынды:

1. Yasser Mohseni Behbahani, Bagher BabaAli, and Mussa Turdalyuly. Persian sentences to phoneme sequences conversion based on recurrent neural networks // Open Computer Science. – 2016. - № 6. – P. 219-225. (Scopus)

2. Bagher BabaAli, Waldemar Wojcik, Oken Mamyrbayev, Mussa Turdalyuly, Nurbara Mekebayev. Speech Recognizer-Based Non-Uniform Spectral Compression for Robust MFCC Feature Extraction // Przegląd Elektrotechniczny. ISSN: 0033-2097 – 2018. - № 6 (94). – P. 90-93. (Clarivate Analytics)

3. К.Н. Калимолдаев, О.Ж. Мамырбаев, М. Турдалыұлы, К.Е. Нурлан, А.Е. Ибраимкулов. Нейрондық желілерді қолдану арқылы сөйлеуді автоматты түрде тану // ҚазҰТЗУ хабаршысы. – 2016. -№ 5 (117). – Б. 435-438.

4. О.Ж. Мамырбаев, Н.О. Мекебаев, М. Тұрдалыұлы. Сөйлеулерді тану үрдісінде MFCC алгоритмін қолдану // ҚазҰТЗУ хабаршысы. – 2018. -№ 2 (126). – Б. 389-392.

5. О.Ж. Мамырбаев, Н.О. Мекебаев, М. Тұрдалыұлы. Генетикалық алгоритм көмегімен сөйлеуді автоматты танудағы гендерлік сәйкестендіру // Алматы энергетика және байланыс университетінің хабаршысы. – 2018. – специальный выпуск. – Б. 120-129.

6. О.Ж. Мамырбаев, М. Тұрдалыұлы, Н.О. Мекебаев. Система распознавания слитной казахской речи на основе глубоких нейронных сетей // Вестник Алматинского университета энергетики и связи. – 2018. – специальный выпуск. – Б. 130-135.

7. О.Ж. Мамырбаев, М. Тұрдалыұлы, Н.О. Мекебаев. Кіріккен қазақ сөйлеуін тану жүйесі // ҚБТУ хабаршысы. – 2018. - №3 (46). – Б. 129-133.

8. О.Ж. Мамырбаев, М. Тұрдалыұлы, Н.О. Мекебаев, И. Ахметов. MFCC негізіндегі дикторды анықтау жүйесі // ҚазҰТЗУ хабаршысы, № 2 (132), 2019

9. О.Ж. Мамырбаев, М. Турдалыұлы, Н.О. Мекебаев, К. Алимхан, Г.С. Набиева, Б.Ж. Мамырбаев. Фонетически представительный текст для создания систем автоматического распознавания казахской речи // Наука и Мир. – 2018. – № 6 (58). – Т. 2 – Б. 49-52.

10. О.Ж. Мамырбаев, М. Калимолдаев, М. Турдалыұлы, В. BabaAli. Методы построения многомодального распознавания речи // Труды II

Международной научно-практической конференции «Информационные и телекоммуникационные технологии: образование, наука, практика». – Алматы, 2015. – Т. 1. – Б. 217-221.

11. О.Ж. Мамырбаев, Н.О. Мекебаев, М. Тұрдалыұлы. Қазақ сөйлеуін тануда іргелі және қолданбалы зерттеуге арналған фонетикалық мәтін // «Төртінші өнеркәсіптік революция жағдайындағы дамудың жаңа мүмкіндіктері» атты ҚР Президенті Н. Назарбаевтың Жолдауын іске асыру шеңберінде «Көліктегі инновациялық технологиялар: білім, ғылым, тәжірибе» атты ХЛІІ Халықаралық ғылыми-практикалық конференцияның материалдары. – Алматы, 2018. – Т. 2. – Б. 81-87.

12. О.Ж. Мамырбаев, М. Тұрдалыұлы, Н.О. Мекебаев. Қазақ тілі сөйлеуінің акустикалық және тілдік модельдерін құру // Материалы XIV Международной Азиатской школы-семинара «Проблемы оптимизации сложных систем». – Алматы, 2018. – Т. 2. – Б. 344-347.

13. О.Ж. Мамырбаев, Н.О. Мекебаев, М. Турдалыұлы. Алгоритмы и архитектуры систем распознавания речи // Материалы III Международной научной конференции «Информатика и прикладная математика» посвященная 80-летию юбилею профессора Бияшева Р.Г. и 70-летию профессора Айдарханова М.Б. – Алматы, 2018. – Т. 2. – Б. 108-121.

14. Orken Mamyrbayev, Mussa Turdalyuly, Nurbapa Mekebayev, Kuralay Mukhsina, Alimukhan Keylan, Bagher BabaAli, Gulnaz Nabieva, Aigerim Duisenbayeva and Bekturgan Akhmetov. Continuous Speech Recognition of Kazakh Language // AMCSE 2018 - International Conference on Applied Mathematics, Computational Science and Systems Engineering. - Rome, Italy, 2018, v24 – 2019

15. Orken Mamyrbayev, Mussa Turdalyuly, Nurbapa Mekebayev, Alimukhan Keylan, Aizat Kydyrbekova and Turdalykyzy Tolganai. Automatic Recognition of Kazakh Speech Using Deep Neural Networks // 11th Asian Conference on Intelligent Information and Database Systems. - Yogyakarta, Indonesia, 2019

16. О.Ж. Мамырбаев, А.С. Кыдырбекова, М. Тұрдалыұлы, Н.О. Мекебаев. Обзор методов идентификации и аутентификации пользователей по голосу // Материалы научной конференции ИИВТ КН МОН РК «Инновационные IT и Smart-технологии», посвященной 70-летию юбилею профессора Утепбергенова И.Т., Алматы, 2019

17. Авторлық куәлік «System of automatic creation vocabulary for ASR» 22.01.2019 ж. №1425