

8D06103 – Management information systems мамандығы бойынша
философия докторы (PhD) ғылыми дәрежесін алу үшін
Оралбекова Дина Орымбаевнаның
«Интегралдық әдіс негізінде автоматты сөйлеу тану жүйесін құру»
атты докторлық диссертациясына

АНДАТПА

Зерттеу тақырыбының өзектілігі. Сөйлеуді тану – бұл машинаның немесе бағдарламалық қамтаманың ауызекі тілдегі сөздер мен сөз тіркестерін тану және оларды машина танитын форматқа, яғни мәтінге түрлендіру болып табылады. Автоматты сөйлеуді тану жүйелері (АСТЖ) қазірдің өзінде күнделікті өміріміздің ажырамас бөлігіне айналды және сөйлеу синтезі, машиналық аударма және т.б. сияқты машиналық оқытудың басқа технологияларын дамытуда үлкен рөл атқарады. АСТЖ автомобильді, ақылды үй және тұрмыстық техниканы дауыспен басқару, сондай-ақ әртүрлі қолданбаларда, навигациялық жүйелерде және т.б. дауыспен енгізу сияқты қызметтің әртүрлі салаларында кең қолданыс тапты. Бұл АСТЖ қолданудың кейбір мысалдары ғана. Сөйлеуді танудың дәстүрлі жүйесі әдетте үш негізгі бір-біріне тәуелсіз элементтерден тұрады, олар келесі модельдер болып табылады: дыбыстан қосалқы фонемалардың контекске тәуелді күйлерін болжауға арналған акустикалық модельдер, тілдік модельдер және фонемаларды сөздерге сәйкестендіруге арналған сөздік (лексика). Мұндай жүйелер бір-бірінен тәуелсіз оқытылатын әртүрлі элементтерден тұратындықтан, классикалық акустикалық модельді Гаусс үлестірілімдері қоспалары мен жасырын Марков модельдері негізінде, ал тілдік модельді n-грамм моделі негізінде оқытуға болады. Ұзақ уақыт бойы сөйлеуді тану мәселесінде жасырын Марков моделі (НММ) кеңінен қолданылды және негізгі технология болды. НММ негізінен кадр деңгейінде динамикалық уақытты бұрмалау үшін қолданылады және Гаусс үлестірілімдері қоспалары (GMM) әдетте айтылу бірлігіне сәйкес келетін бекітілген қысқа уақыт аралығындағы сигнал үлестірімдерін көрсету үшін пайдаланылады. Ұзақ уақыт бойы НММ-GMM моделі сөйлеуді танудың жалпы негізгі моделі болған. Жақында терең оқыту әдістері көптеген зерттеулерде және сөйлеуді тануды дамытуда айтарлықтай жақсартулар әкелді. Сөйлеуді танудың классикалық жүйесінің сценарийінің әрбір элементінде жасанды нейрондық желілерді белсенді қолдану оның жұмысының тиімділігін арттырады, бұл көптеген ғылыми еңбектерде орын алған. Терең оқыту технологияларының дамуымен сөйлеуді танудағы акустикалық модельдер үшін әдеттегі GMM орнына терең нейрондық желілер (DNN) қолданыла бастады. DNN рөлі әдеттегі GMM бақылау ықтималдығын ауыстыра отырып, ықтималдықтарға түрлендіруге болатын НММ күйінің апостериорлық ықтималдығын есептеу болып табылады. Осылайша, НММ-GMM моделі жақсы нәтижелерге қол жеткізетін және танымал автоматты сөйлеуді тану үлгісіне айналатын НММ-DNN моделіне айналады. Ғылыми жұмыстарда тиімді акустикалық модельді алу

үшін терең нейрондық желілер пайдаланылғаны көрсетілді, ал басқа ғылыми еңбектерде рекуррентті нейрондық желілер мен ұзақ-қысқа мерзімді есте сақтау желісі (LSTM) арқылы тілдік модельдер мен сөздік құрастырылғаны көрсетілген. Сондай-ақ басқа ғылыми жұмыстарда конволюциялық нейрондық желілер (CNN) сөйлеу сигналынан дыбыс белгілерді алу үшін қолданылды. Осылайша, жарияланған зерттеулер негізінде алынған нәтижелер арқылы ұсынылып отырған әдіс сөйлеуді тану жүйелерінің барлық кезеңдерінде жасанды нейрондық желілердің әртүрлі архитектураларын қолдану үшін негіз болып табылатын барлық заманауи сөйлеуді тану жүйелері арасында ең жақсы өнімділікті алғанын көрсетеді.

Қазіргі уақытта машиналық оқыту әдістерін қолданатын сөйлеуді танудың интегралдық әдісі кең таралған. Мұндай жүйелерде модель тек бір нейрондық желіні қолдану арқылы жүзеге асырылады. Интегралдық модель көбінесе сөйлеуді тану жылдамдығы мен дәлдігі бойынша ең жақсы өнімділікті көрсетеді. Көптеген ғылыми-зерттеу жұмыстарында интегралдық жүйенің көмегімен алынған нәтижелерінің жақсаруы желіні оқытуға арналған сөйлеу корпусы көлемінің ұлғаюына байланысты екендігі дәлелденді. Қазіргі уақытта Voice to Text Messenger, Google Listen, Attend, Spell, Baidu Deep Speech және басқа да танымал программалық қосымшалар интегралдық әдіске негізделіп жасалынған. Осы уақытқа дейін белгілі интегралдық модельдердің негізгі ерекшелігі – үлкен деректер негізінде оқытылуы. Жоғарыда келтірілген мәселелерден негізгі проблеманы анықтауға болады, бұл оқыту деректері шектеулі тілдерді тануға қатысты, мұндай тілдерге қазақ, қырғыз, түрік және т.б. жатқызуға болады. Мұндай аз ресурсты тілдер үшін үлкен көлемді сөйлеу корпустары жоқ. Сондай-ақ қазақ тілі сияқты ресурсы аз тілдерге арналған интегралдық архитектураның модельдері мен әдістері әзірленбегенін және зерттелмегенін атап өткен жөн. Қазіргі таңда қазақ тілін интегралды тану үшін тиімді алгоритмдер мен бағдарламалық қамтама жетілдірілмеген.

Dario Amodei, Chao Weng сияқты шетелдік ғалымдар және басқа зерттеушілер ағылшын және қытай сияқты танымал тілдерді тану үшін интегралдық жүйелерді жетілдіруде жоғары нәтижелерге қол жеткізді, ал ТМД ғалымдары, атап айтқанда Ресей ғылым академиясының Санкт-Петербург ақпараттық және автоматизация институтының зерттеушілері Карпов А.А., Кипяткова И.С. және олардың әріптестері көптеген жылдар бойы сөйлеуді тану саласын зерттеп, орыс тілін тану үшін интегралдық жүйелерді әзірлеу мен жетілдіруде жақсы нәтижелерге қол жеткізді.

Сондай-ақ отандық ғалымдардың терең нейрондық желілерге негізделген қазақша сөйлеуді тану жүйелері бойынша зерттемелері бар екенін айта кеткен жөн. Л.Н.Гумилев атындағы Еуразия ұлттық университеті ғалымдарының – Шәріпбай А.А., Есенбаев Ж.А., Әл-Фараби атындағы Қазақ ұлттық университеті ғалымдарының – Тукеев У.А., Рахимова Д.Р., Назарбаев Университетінен – Хасанов Е., сонымен қатар Ақпараттық және есептеуіш технологиялар институтының ғылыми қызметкерлері – Амиргалиев Е.Н., Мусабаев Р.Р. және т.б. еңбектерінде дәстүрлі, гибридті HMM-DNN модельдері және CTC мен Transformer көмегімен интегралдық моделі

негізінде қазақша сөйлеуді автоматты тану жүйесінің дамуы көрініс тапты. Бұл жұмыстардың нәтижелері сөйлеуді дұрыс танудың жақсы деңгейіне жеткізген, бірақ осы уақытқа дейін бұл жүйелер сөйлеуді тану дәлділігін адам деңгейіне дейін жетілдіріп және жүйені оқыту үшін деректер көлемін арттыруды жақсартуды талап етеді. Сонымен қатар, назар аудару механизмін пайдалана отырып, кодер-декодер (Encoder-Decoder) моделі негізінде қазақша сөйлеуді танудың интегралдық жүйелері әзірленбеген. Назар аудару механизмі бар модель сөйлеуді тану дәлділігін арттыра отырып, тілдік модельдердің көмегінсіз жақсы жұмыс істей алатын перспективалы нәтижелерді көрсетеді. Бұл әдіс шектеулі оқыту сөйлеу корпусы бар тілдер үшін пайдаланылуы мүмкін перспективалы бағыт болып табылады.

Жоғарыда айтылғандарға сүйене отырып, қазіргі уақытта интегралдық модельдерді пайдалана отырып, қазақша сөйлеуді тану дәлдігін жақсарту үшін тиімді әдістер, алгоритмдер және бағдарламалық қамтама қажеттілігі *өзекті* мәселе болып табылады деп қорытынды жасауға болады.

Диссертациялық жұмыстың мақсаты. Интегралдық әдіс негізінде қазақша сөйлеуді тану дәлдігін жақсарту үшін тиімді модельді, оның архитектурасы мен алгоритмін әзірлеу.

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

1) Интегралдық модельдер негізінде заманауи сөйлеуді тану әдістерін және модельдерін талдау.

2) Қазақша сөйлеуді интегралды тану үшін сөйлеу корпусын құру.

3) Қазақша сөйлеуді тану жүйесін құру үшін назар аудару механизмін пайдалана отырып, кодер-декодер (Encoder-Decoder) негізінде интегралдық модельдің тиімді әдісі мен алгоритмін жасау.

4) Интегралдық модель негізінде зерттеу барысында алынған әдістер мен алгоритмді пайдалана отырып, қазақ сөйлеуін тануға арналған интегралды архитектура мен бағдарламалық қамтамасын әзірлеу.

Зерттеу нысаны. Сөйлеуді автоматты танудың заманауи технологиялары мен жүйелері.

Зерттеу пәні. Интегралдық әдіс негізінде қазақша сөйлеуді танудың технологиялары, алгоритмдері, модельдері мен бағдарламалық қамтама.

Зерттеу әдісі. Машиналық оқыту әдістері, сөйлеуді автоматты тану технологиялары мен әдістері, ықтималдықтар теориясы және математикалық статистика, бағдарламалық қамтаманы әзірлеу әдістері, сонымен қатар қолданбалы және есептеу лингвистикасы.

Жұмыстың ғылыми жаңалығы.

1) Қазақ тілінің жаңа сөйлеу және мәтін корпустары әзірленді.

2) Қазақша сөйлеуді танудың назар аудару механизмі бар кодер-декодер негізіндегі интегралды модель әзірленді.

3) Интегралдық модуль негізінде қазақша сөйлеуді тиімді тану алгоритмі әзірленді.

4) Сөйлеуді мәтінге автоматты түрде түрлендіретін бағдарламалық қамтама әзірленді.

Жұмыстың теориялық және практикалық маңызы.

Зерттеу жұмысының теориялық маңыздылығы қазақша сөйлеуін танудың интегралдық модельдерінің әдістері мен алгоритмдерін жасау мен енгізуде, сонымен қатар қазақ тіліне арналған сөйлеу корпусын әзірлеуде.

Зерттеу жұмысының практикалық маңыздылығы әзірленген алгоритмдер мен бағдарламалық қамтама әрі қарай сөйлеуді синтездеу, машиналық аударма, дауысты аутентификация және идентификация сияқты басқа технологияларды әзірлеуде қолдануға мүмкіндік береді. Қазақша сөйлеудің автоматты танудың әзірленген жүйесі ақпараттық технологиялар негізінде мемлекеттік тілді қолдану аясын кеңейтуге жауапты мемлекеттік органдарда іске асырылуы мүмкін; ұялы телефондарда (мемлекеттік тілде сөйлеу технологияларын енгізу есебінен әлеуетті сатып алушылар санының артуы); банктерде (дауыс функцияларын қолдайтын байланыс орталықтары, дауысты аутентификация); дауыс функцияларын қолдайтын әртүрлі құрылғыларды өндіру секторы және т.б.

Қорғауға шығарылатын ғылыми тұжырымдар.

1) Жүйені оқыту үшін транскрипциясы бар 2000 сағаттық көлемді құрайтын сөйлеу корпусы әзірленді. Корпусты құру кезінде сөйлеудің әртүрлі формалары ескерілді: дайындалған (оқылым), кездейсоқ сөйлеу.

2) Назар аудару механизміне негізделген модель сөздердегі ұқсас жалғауларды тану үшін түрлендірілді және кеңейтілді. Бұл модельдің архитектурасы LSTM және BLSTM сияқты нейрондық желілер арқылы құрастырылған. Эксперименттік нәтижелер құрастырылған модельдің қазақ тіліне арналған тілдік модельдерді қолданбай-ақ жақсы жұмыс істейтінін және DNN-HMM негізіндегі гибриді модельдерді ғана емес, сонымен қатар сөйлеу корпусының аз көлемімен оқытылған басқа да интегралды модельдерден асып түсетінін және сөз бен символ негізінде тану дәлдігі бойынша жақсы нәтиже бергенін көрсетті.

Сенімділік дәрежесі мен апробациялау нәтижелері. Диссертациялық жұмыстың тақырыбына қатысты зерттеулер мен нәтижелер келесі жарияланымдар негізінде әртүрлі конференциялар мен семинарларда ұсынылып, талқыланды:

1) V Халықаралық ғылыми-практикалық конференция «Информатика және қолданбалы математика» (29 қыркүйек - 1 қазан **2020 ж.**, Алматы).

2) 3rd International Conference on Computer Communication and the Internet (ICCCI) (25-27 маусым **2021 ж.**, Токио, Жапония).

3) VI Халықаралық ғылыми-практикалық конференция «Информатика және қолданбалы математика» (29 қыркүйек - 1 қазан 2021 ж., Алматы).

4) Профессор У.А. Тукеевтің 75 жылдығына арналған ақпараттық технологиялар саласындағы халықаралық ғылыми конференция (8 қазан **2021 ж.**, Алматы).

5) Халықаралық ғылыми конференция «Сәтбаев оқулары 2021» (12 сәуір **2021 г.**, Алматы)

6) 7th International conference on Computer Science and Engineering (14-16 қыркүйек **2022**, Стамбул, Түркия)

Зерттеушінің жеке үлесі. Докторант диссертациялық жұмыстың тапсырмаларын өзі орындап, шешті. Интегралдық модельдер негізінде қазақша сөйлеуді тану әдістері мен алгоритмдері әзірленіп, іске асырылды. Қазақша сөйлеу корпусын әзірлеп, кеңейтті. Жасалған әдістер мен алгоритмдерге эксперименттік бағалау жүргізді.

Диссертация тақырыбының ғылыми-зерттеу жұмыстарының жоспарларымен байланысы. Диссертация бойынша жүргізілген ғылыми-зерттеу жұмыстары екі гранттық қаржыландыру жобасының аясында жүзеге асырылды: 1) «Терең нейрондық желілерді пайдаланатын мультитілдік автоматты сөйлеуді тану технологиясын құру» (2018-2020 жж., мемлекеттік тіркеу нөмірі: 0118RK00139) 2) ҚР БҒМ Ақпараттық және есептеуіш технологиялар институтында «Агглютинативті тілдер үшін сөйлеуді автоматты түрде танудың интегралды (end-to-end) жүйелерін әзірлеу» (2020-2022жж., мемлекеттік тіркеу нөмірі: 0120RK00344).

Диссертациялық зерттеудің негізгі нәтижелерінің жарияланымдары. Диссертациялық жұмыстың тақырыбы бойынша 3 авторлық куәлік, 1 патент, 7 мақала жарияланды, оның ішінде 3 мақала Қазақстан Республикасы Білім және Ғылым министрлігінің білім және ғылым саласында сапаны қамтамасыз ету комитеті ұсынған мақалалар бойынша, 4 мақала Scopus және Web of Science деректер базасымен индекстелген, нөлдік емес импакт-факторы бар басылымда жарияланған:

1) Mamyrbayev, O., **Oralbekova, D.**, Keylan, A. et al. A study of transformer-based end-to-end speech recognition system for Kazakh language. Sci Rep 12, 8337 (2022). <https://doi.org/10.1038/s41598-022-12260-y> (**Web of Science, Q1, IF=4,3**)

2) Mamyrbayev, O.Z., Oralbekova, D.O., Alimhan, K. et al. Hybrid end-to-end model for Kazakh speech recognition. International Journal of Speech Technology (2022). <https://doi.org/10.1007/s10772-022-09983-8> (**Scopus, IF 1.803, процентилі 93**).

3) Mamyrbayev, O., Kydyrbekova, A., Alimhan, K., **Oralbekova, D.**, Zhumazhanov, B., Nuranbayeva, B. (2021). Development of security systems using DNN and i & x-vector classifiers. Eastern-European Journal of Enterprise Technologies, 4 (9 (112)), 32–45. doi: <https://doi.org/10.15587/1729-4061.2021.239186> (**Scopus, процентилі 43**);

4) Mamyrbayev, O., Alimhan, K., **Oralbekova, D.**, Bekarystankyzy A., Zhumazhanov, B. (2022). Identifying the influence of transfer learning method in developing an end-to-end automatic speech recognition system with a low data level. Eastern-European Journal of Enterprise Technologies, 1(9(115)), 84–92. <https://doi.org/10.15587/1729-4061.2022.252801>(**Scopus, процентилі 43**);

5) Mamyrbayev O., **Oralbekova D.** Modern trends in the development of speech recognition systems // News of the National academy of sciences of the republic of Kazakhstan. – 2020. – Vol. 4, № 332. - P. 42 – 51 // doi.org/10.32014/2020.2518-1726.64

6) Mamyrbayev O., **Oralbekova D.**, Alimhan K., Othman M., Zhumazhanov B. Realization of online systems for automatic speech recognition// News of the

National academy of sciences of the republic of Kazakhstan. – 2021. – Vol. 6, № 340. - P. 66 – 72 // doi.org/10.32014/2021.2518-1726.103

7) Мамырбаев О.Ж., **Оралбекова Д.О.**, Алимхан К., Othman M., Жумажанов Б. Қазақша сөйлеуді тану үшін гибриді интегралдық модельді қолдану // News of the National academy of sciences of the republic of Kazakhstan. – 2022. – Vol. 1, № 341. - P. 58 – 68 // doi.org/10.32014/2022.2518-1726.117.

8) Авторлық куәлік "Система автоматического распознавания казахской речи на основе интегральной архитектуры" № 15501, 25.02.2021, Авторлары: О.Ж. Мамырбаев, **Д.О. Оралбекова**, А.С. Кыдырбекова, Б.Ж. Жумажанов, Т.Тұрдалықызы.

9) Авторлық куәлік "Система идентификации и аутентификации через речевые технологии" № 23323, 4.02.2022. Авторлары: **Оралбекова Д.О.**, Мамырбаев О.Ж., Алимхан К., Кыдырбекова А.С., Жумажанов Б.Ж., Тұрдалықызы Т.

10) Авторлық куәлік "Система автоматического распознавания казахской слитной речи на основе модели с механизмом внимания" №24178, 5.03.2022. Авторлары: Мамырбаев О.Ж., **Оралбекова Д.О.**, Әлімхан Қ., Кыдырбекова А.С., Жумажанов Б.Ж., Тұрдалықызы Т.

11) Өнертабысқа патент «Агглютинативті үздіксіз сөйлеуді интегралды (end-to-end) тану тәсілі және оны жүзеге асыратын жүйе» № 35886 07.10.2022 ж. Авторлары: Мамырбаев О.Ж., Оралбекова Д.О., Кыдырбекова А.С., Жумажанов Б.Ж., Тұрдалықызы Т.

Диссертациялық жұмыстың құрылымы мен көлемі. Диссертациялық жұмыс кіріспеден, 4 тараудан, қорытындыдан, 111 атаудан тұратын әдебиеттер тізімінен және 5 қосымшадан тұрады. Жұмыс 108 бетте берілген және 25 суреттен, 6 кестеден тұрады.

Диссертациялық зерттеудің қысқаша сипаттамасы.

Кіріспеде зерттелетін жұмыстың өзектілігі, диссертациялық жұмыстың мақсаты мен міндеттері, ғылыми жаңалығы, жұмыстың теориялық және практикалық маңыздылығы және зерттеу әдістері көрсетілген.

Бірінші тарауда автоматты сөйлеуді тану жүйесінің жалпы моделі сипатталды және уақытша коннекционды классификациясы және рекуррентті трансформатор, кодер-декодер моделі және назар аудару механизмі бар Transformer архитектурасы, шартты кездейсоқ өрістерге негізделген модель сияқты сөйлеуді тану интегралды жүйелеріне шолу жасалды. Қарастырылған модельдер бойынша тиісті жұмыстарға шолу келтірілген, сонымен қатар осы модельдердің дәстүрлі модельдермен салыстырмалы талдауы сипатталды.

Екінші тарауда Екінші тарауда қазақ тіліне арналған корпусты жинау жөніндегі жұмыс қарастырылады, ол олардың транскрипциясы (аудионың мәтіндік көрінісі) бар аудио деректерден және корпустың толық мазмұнынан тұрады. Телефонмен сөйлесулерді транскрипциялау үшін мәтіндерді құрастыру әдістемесі келтірілген.

Үшінші тарауда назар аудару механизмі бар кодер-декодер моделінің архитектурасы сипатталған. Кодер, декодер және назар аудару механизмі үшін интегралды модельдің алдын ала конфигурациясы сипатталған, сондай-ақ

сөйлеуді тану үшін бағалау метрикасы берілген. Тәжірибелерде қолданылатын деректер жиынының сипаттамасы және назар аудару механизмі бар кодер-декодер алгоритмі келтірілген. Назар аудару механизмі бар модельді жүзеге асыруға арналған бағдарламалық және аппараттық құралдар ұсынылған. Бұған қоса, сөзді тану үшін Philips компаниясы арнайы әзірлеген SmartMike Duo микрофоны сипатталған. Егер 2 адам сөйлескен кезде, әсіресе қатысушылар бір уақытта сөйлескенде дауыстың қабаттасуына әкелсе, сонда микрофон көмегімен екі бөлек аудио арнаның қабаттасуын ажыратуға болады. Сонымен қатар, ұсынылған интегралдық модельді эксперименттік тексеру сипатталған. Алынған нәтижелерді салыстыру үшін қазақша сөйлеуді тануға байланысты зерттеу жұмыстары таңдалды.

Төртінші тарау қазақ тілін автоматты түрде тану жүйесін сипаттауға арналған. Біріктірілген жүйенің егжей-тегжейлі құрылымы қарастырылды, нейрондық желілерді оқытуға және жүйе деректерін тексеруге және шығаруға арналған модульдердің жұмысы сипатталған. Бағдарламаның интерфейсі және оның құрамдас бөліктері берілген. Біріктірілген жүйені SmartMike Duo USB PSM1010 микрофонымен интеграциялау процесі сипатталған, бұл әзірленген жүйеге сапалы танудың дәлдігі мен сөйлеуді талдаудың кеңейтілген мүмкіндіктерін қамтамасыз ететін екі бөлек аудио арнаны пайдаланудың бірегей мүмкіндігін береді.

Қорытынды бөлімінде диссертациялық зерттеудің нәтижелері мен қорытындылары берілген және таңдалған бағытты дамыту бойынша алдағы жұмыстардың жоспарлары көрсетілген.