әл-Фараби атындағы Қазақ ұлттық университеті

ҚР БҒМ ҒК Ақпараттық және есептеуіш технологиялар институты

ӘОЖ 004.934:004.52:004.89

Қолжазба құқығында

ТҰРДАЛЫҰЛЫ МҰСА

Терең рекуррентті нейрожелілік модельдерді қолдану арқылы кіріккен сөйлеуді толассыз тану

6D070400 – Есептеу техникасы және бағдарламалық қамтамасыз ету

Философия докторы (PhD)
дәрежесін алу үшін дайындалған диссертация

Ғылыми кеңесшілер:
Мамырбаев О.Ж.,
PhD, қауымдастырылған профессор
Bagher BabaAli,
PhD, қауымдастырылған профессор

Қазақстан Республикасы
Алматы, 2019

МАЗМҰНЫ

НОРМАТИВТІ СІЛТЕМЕЛЕР 4

БЕЛГІЛЕУЛЕР МЕН ҚЫСҚАРТУЛАР 5

КІРІСПЕ 6

1 СӨЙЛЕУДІ ӨҢДЕУДІҢ НЕГІЗГІ МӘСЕЛЕЛЕРІ МЕН ӘДІСТЕРІ 11

1.1 Сөйлеуді автоматты тану жүйелерін құруда кездесетін қиындықтар ... 11

1.2 Сөйлеу сигналдарын автоматты өңдеу әдістері 17

1.2.1 Табиғи тіл мен сөйлеуді өңдеудің негізгі деңгейлері 18

1.2.2 Кіріккен сөйлеуді тану жүйесінің негізгі архитектурасы 21

1.2.2.1 Сигналдық параметрлік көрінісі 23

1.3 Рекуренттік нейрондық желілер 26

1.4 Рекуренттік нейрондық желілерді оқыту 27

1.4.1 Уақыт бойынша қатені кері үлестіру алгоритмі (ВРТТ) 27

1.6 Рекуренттік нейрондық желілер көмегімен тілді моделдеу 31

2 КІРІККЕН ҚАЗАҚ СӨЙЛЕУІНІҢ АКУСТИКАЛЫҚ ЖӘНЕ ТІЛДІК КОРПУСТАРЫН ТАЛДАУ ЖӘНЕ ҚҰРУ 34

2.1 Тілдік корпустарға шолу 34

2.1.1 Қазақ тілінің сөйлеу корпустары 35

2.2 Сөйлеу корпустарын талдау 35

2.3 Қазақ тілінің корпусын құру кезінде туындайтын мәселелер 36

2.4 Сөйлеу корпусын құру кезеңдері 37

2.4.1 Сөйлеу корпусының мәтіндерін дайындау 38

2.4.2 Дикторларды таңдау 38

2.5 Кіріккен қазақ сөйлеуінің акустикалық деректері 39

2.5.1 Жазу процессі 40

2.5.2 Аудиожазба туралы ақпарат 41

2.6 Қазақ тілі сөйлеуінің акустикалық корпусы 41

2.7 Қазақ тілінің мәтіндік корпусы 43

2.8 Қазақ тілі сөйлеуін тану жүйесіне арналған сөздікті құру 49

3 ҚАЗАҚ ТІЛІНІҢ АКУСТИКАЛЫҚ ЖӘНЕ ТІЛДІК МОДЕЛЬДЕРІН ҚҰРУ ЖӘНЕ ТӘЖІРИБЕЛЕР ЖҮРГІЗУ 51

3.1 Акустикалық және тілдік модельдер 51
3.1.1 Акустикалық және тілдік модельдерін құру құралы 51
3.2 Тәжірибелерді жүргізу ортасы 53
3.3 Жүргізілген тәжірибелердің сапасын бағалау 54
3.4 Сөйлеу сигналдарын өңдеу 55
3.4.1 Сөйлеу сигналдарын өңдеудің спектральді сығымдау алгоритмі 56
3.5 Модельдерді құру құралына оқытуға дайындық 58
3.6 Акустикалық модельді құру 60
3.7 Тілдік модельді құру 62
3.7.1 Рекуррентті нейрондық желілер негізіндегі тіл моделі 62
3.8 Алынған модельдердің сапасын бағалау 64
4 КІРІККЕН ҚАЗАҚ СӨЙЛЕУІН ТАНУ КЕШЕНІН ҚҰРУ 67
4.1 Kaldi Gstreamer сервері 67
4.1.1 Kaldi Gstreamer орнату және баптау 67
4.1.1 Kaldi Gstreamer серверіне қойылатын талаптар 68
4.1.2 Серверді жұмысын бастау 68
4.1.3 Жұмысшыларды іске қосу 68
4.1.4 Сервер мен клиенттік қосымшаларды байланыстыру 69
4.1.5 Сервер мен жұмысшыларды автоматтандыру 69
4.2 Сөйлеуді тану кешенінің пайдаланушы қосымшаларын құру 70
4.2.1 Сөйлеуді тану бөлімі 71
4.2.2 Қазақ сөйлеуін танудың Веб-қосымшасы 74
4.2.3 Қазақ сөйлеуін танудың мобильдік қосымшасы 78
ҚОРЫТЫНДЫ 81
ПАЙДАЛАНҒАН ӘДЕБИЕТТЕР ТІЗІМІ 82
ҚОСЫМША А. Қосымшаға авторлық куәлік 90
ҚОСЫМША Б. Мобильдік қосымшаның листингі 91

НОРМАТИВТІ СІЛТЕМЕЛЕР

Бұл диссертацияда келесі стандарттарға сәйкес сілтемелер қолданылды:

ҚР МЖМБС 5.04.034 – 2011 «Қазақстан Республикасының Мемлекеттік жалпыға міндетті білім беру стандарты. Жоғары оқу орнынан кейінгі білім. Докторантура». Негізгі ережелер ҚР білім және ғылым министрімен бекітілген. «17» маусым 2011 ж. №261. Астана, 2011.

«Диссертацияларды және авторефераттарды рәсімдеу бойынша нұсқаулық», ҚР БҒМ, Жоғары аттестаттау комитеті, Алматы, 2004.

МЕСТ 7.1-2003. Библиографиялық жазба.

БЕЛГІЛЕУЛЕР МЕН ҚЫСҚАРТУЛАР

САТ – сөйлеуді автоматты тану
DNN – Deep Neural Networks
GMM – Gaussian mixture model
ИНС – Искусственные нейронные сети
IVR – Interactive Voice Response
LVCSR – Система непрерывного распознавания речи
GMM-HMM – Hidden Markov Models and Gaussian Mixture Models
СПИИРАН - Санкт-Петербургский институт информатики и автоматизации Российской академии наук
СММ – Скрытые Марковские модели
MFCC – Мел кепстральді коэфициент
CUDA – Compute Unified Device Architecture
PDNN – Python deep learning toolkit
RNN – Рекуррентті нейрондық желі
LSTM - Long short-term memory
LDC – Linguistic Data Consortium
ҚР – Қазақтан Республикасы
БҒМ – Білім және ғылым министрлігі
ҒК – Ғылым комитеті

КІРІСПЕ

Зерттеу тақырыбының өзектілігі. Ғылым мен техниканың дамуы адам мен машинаның өзара әрекеттесу құралдарының эволюциялсымен 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 вариативтлік, сөйлеу тіліндегі акценттік және көңіл – күйлік сипатардың 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 [1]. 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 диапозонының 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 [2; 3] (300 сағаттық) ағылшын тіліндегі корпус, Фишер корпусы [4; 5] (2000 сағаттық) т.б. корпустар пайдаланылады. Ағылшын тілін зерттеушілердің көбі Лингвистикалық Мәліметтер Консорциумы (Linguistic Data Consortium, LDC) [6] дайындаған HLIB5 Evai 2000 [7] 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 уақытта 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 [8-10].

Қазіргі таңда біздің елімізде қазақ сөйлеулерін тану мәселелерімен Л.Н. Гумелев атындағы Еуразиялық ұлттық университетінің ғалымдары А.Ә. Шәріпбай, Г.Т. Бекманова [11, 12], 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 [13-17].

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 кірігу), 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 нысаны. Сөйлеуді автоматты тану жүйелері.

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

Зерттеу әдісі. 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 асыру 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.

Алынған ғылыми нәтижелердің жаңалығы:

– 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 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 қазақша айтылым кезіндегі кіріккен сөйлеуді 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 рекуррентік нейрондық желілік модельдерді қолдану сөйлеуді автоматты тану жүйелерінің тану сапасын арттырады.

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» халықаралық ғылыми конференциясы (Йоджиякарта, Индонезия, 2019).

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).

Нәтижелердің жариялынымдары. Диссертация тақырыбы бойынша алынған нәтижелері жеті баспалық жұмыста жарияланды. Оның ішінде халықаралық реферативтік базаларына енетін басылымдарында жарияланған мақалалар:

1) Persian sentences to phoneme sequences conversion based on recurrent neural networks. Open Comput. Science, № 6, 2016 (Scopus);

2) Speech Recognizer-Based Non-Uniform Spectral Compression for Robust MFCC Feature Extraction. Przegląd Elektrotechniczny, ISSN 0033-2097, №6, 2018. (Scopus және Web of Science (Clarivate Analytics));

3) Automatic Recognition of Kazakh Speech Using Deep Neural Networks. 11th Asian Conference on Intelligent Information and Database Systems, ACIIDS 2019 (Scopus).

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.

1 СӨЙЛЕУДІ ӨҢДЕУДІҢ НЕГІЗГІ МӘСЕЛЕЛЕРІ МЕН ӘДІСТЕРІ

Соңғы уақытта қазақ тілінің кіріккен сөйлеуін автоматты тану жүйесін құру бойынша зерттеулер жүргізіліп, біршама тәжірибе жинақталуда, сонымен қатар сөйлеуді автоматты тану жүйесін құруда бірқатар түткүйіл мәселелер анықталуда. Оқытатын сөйлеу деректері мен қалыпты жағдайларда өңделетін сөйлеу деректері арасында айтарлықтай айырмашылықтың болуы тану жүйелерін құру кезінде жіберілетін қателердің негізгі себебі болып табылады. Сөйлеудің айтылуы, қарқыны, стилі сынды факторлар құбылмалылығын, сонымен қатар қоршаған ортадағы шуларды, көбіне жүйені оқыту кезеңінде, алдын ала ескеру мүмкін болмайды. Бұдан басқа, өзара әрекеттестік табиғилығын қамтамасыз ету үшін пәндік саланың шектеулілігіне қарамастан, жүйе барынша орасан зор сөздікті тану үшін, оқытылуы тиіс. Өйткені диалог кезінде қолданушылар түрлі сөз жинақтарын пайдаланады. Осыған байланысты – дыбысталуы бойынша жақын сөздердің пайда болуы және өңделу жылдамдығының төмендеуі проблемасы туындайды. Ақыр соңында, қойылған міндетке байланысты танудың рұқсат етілген дәлдігі өзгеріске ұшырайды, яғни стенографиялауда барлық мәнді сөздерді тану қажет. Диалогтық жүйелерде ақпараттық ресурстарға сұрау жасау үшін кейде сөздер тізбегіндегі бірнеше негізгі сөздерді тану жеткілікті.

1.1 Сөйлеуді автоматты тану жүйелерін құруда кездесетін қиындықтар

Сөйлеуді танудың кез-келген жүйесі алдын ала оқыту кезеңін немесе белгілі бір қалыпқа келтіруді қажет етеді. Ол үшін нақты пайдалану жағдайларына неғұрлым жақын сөйлеу мен тіл материалдары болуы керек. Алайда оны әзірлеушілердің иелігінде, ең көп дегенде, оқылған сөз тіркестерінің аудио жазбасы мен пәндік саладан алынған мәтіндер ғана болуы мүмкін. Сол себепті мұндай материалдар бойынша оқытылған жүйе, әрине, елеулі ауытқулар туындатады да, сөйлеуді тану сапасына қанағаттанбаған пайдаланушы қолдануды мерзімінен бұрын тоқтатады және әдетте, мұндай жүйемен жұмыс істемейді.

Бүгінгі күні мұндай олқылықты еңсерудің бірден-бір жолы ол жүйе жұмысының алғашқы сатыларында жасырын эксперттік қолдау жүзеге асырылатын Wizard of OZ немесе OZ Paradugm (WOZ) әдістерін негізге ала отырып, бірте-бірте пайдалануға беру қажет [18, 19]. Алайда бұл тәсілдің кемшілігі – енгізу кезеңінің бастапқы сатысында өзіндік құны жоғары болып табылады, өйткені нәтижелерді жедел сараптамалық түзету мүмкіндігі бар операторлар штатын және сөйлеуді тану жүйесінің серверін қамтамасыз ету керек болады.

Сигналдарды өңдеу уақытына қойылатын қатаң талап, қойылатын міндеттер күрделі жағдай болып саналады [20]. Мысалы, ақпараттық-анықтамалық қызметтерде сөйлеуді тану жүйесін қолдануда пайдаланушы табиғи диалогты күтеді, сондықтан жүйе сөйлеу сұрауын өңдеп, оған жауап беруі үшін, ақпараттар іздеуі керек және қалыпты уақыт режимінде жауап мәтінін

құрастыруы (сөйлеуді синтездеуге сәйкес) қажет. Мұндай диалогтар бірнеше сұрақ-жауаптардан тұрады, олардың ұзақтығы бірнеше минуттан аспайды. Әдетте жүйелер берілген пәндік сала бойынша ақпарат беруге бағдарланады, ал пайдаланушы біршама ғана лексикон пайдалануы мүмкін деп күтіледі. Сондықтан тану жүйесі әдетте сөйлеу ағынындағы шектеулі негізгі кілтті сөздерді іздеуге және өңдеуге бағдарлайды [21]. Дегенмен, тану жүйесі жұмысының нәтижесі болып ендірілген сөйлеу сигналдарындағы қамтылу мүмкіндігі жоғары негізгі сөздер жиыны табылады. Мұндай сөздер жиынтығы анықтамалық ақпарат іздеуді жүргізу үшін жеткілікті. Екінші жағынан, көпшілікке қызмет көрсету жүйесі фонетикадан семантикаға дейінгі барлық деңгейде өңделінуі тиіс сөйлеу вариативтіліктерін (нұсқалылығын) ескеруі тиіс.

Кесте 1.1 – Сөйлеу сигналының түрленуіне ықпал ететін негізгі факторлар.

Ұзақмерзімді									Қысқа-мерзімді		Сигналды қабылдау және беру арнасының қасиеттері			Қоршаған ортаның қасиеттері		
Сөйлеу жолының анатомиялық ерекшеліктері				Қоршаған қоғамда сөйлесу процестерінен алынғандар		Сөйлеудің бұзылуы			Диктордың жағдаяты							
Кеуде қуысының көлемі	Дыбыс жолының ұзындығы	Ерін, тілдің көлемі мен салмағы	Сөйлеу жолының нысанын сипаттайтын басқа да параметрлер	Диалект	Сөйлеу стилі	Кекештік	Афазия	Басқа да аурулар	Физиологиялық (суық тию, сөйлеу қуысының анестезиясы, уытгану және т.б.)	Эмоционалдық (қуаныш, зерігу, жабырқау, толқу және т.б.)	Микрофон типі	Микрофоннан қашықтық	Сигнал беру арнасындағы кедергілер	Сөйлеуден тыс шу деңгейі	Диалогқа қатысушылар саны	Сөйлесу үшін қолжетімді модальдылықтар саны

Сөйлеу құруда әрбір адам бәріне бірдей артикуляциялық қозғалыстар жиынын пайдаланады деп жорамалданады. [22]. Дегенмен, әр адамның сөйлеуі – дауыстық жолдың, не болмаса сөйлеу стилінің туа біткен және жүре пайда болған ерекшеліктеріне байланысты бірегей. Сонымен қатар, сөйлеу сапасы арнаның сөйлеу сигналын қабылдау және беру қасиеттеріне айтарлықтай әсер етеді, сондай-ақ қоршаған орта жағдайлары маңызды рөл атқарады. Аталмыш параметрлер қоршаған ортаға әсер ететін шу деңгейін, диалогқа қатысушылардың саны мен мақсаттарын қамтиды [23]. Сөйлеудегі өзгеріс диктордың жағдаятының өзгерістері мен ерекшеліктерінің салдарынан [24] және сөйлеу сигналын жазудың сыртқы жағдайларынан туындайды. Сөйлеу сигналының түрленуіне ықпал ететін кейбір факторлардың жіктелімі 1.1 кесте көрсетілген.

Кестеде көрсетілген факторлар, түптеп келгенде, сөйлеуді тану жүйесіне келіп түсетін сөйлеу сигналдарының сапасында көрініс табады. Өзгеріске ұшыраған және сандық тұрғыда бағалауға болатын негізгі параметрлерді атап көрсетуге болады. Олар:

- дыбыс деңгейі;
- дыбыс қарқыны;
- негізгі дыбыс жиілігі;
- сөйлеудің спектрлік құрамы;
- спектрлік және амплитудалық аймақтағы микротербелісі деңгейі;
- дәлдік, сөйлеу сапасының интегралдық параметрі ретінде.

Сөйлеу сигналының акустикалық параметрін есептеу үшін негізгі тон жиілігін қалыптандыру әдістері [25], сигналдағы шудың компонентін ішінара жоюға мүмкіндік беретін спектрлі азайту әдістері, сондай-ақ фонетикалық контекстер мен дикторлар арасындағы айырмашылықтарды ескере отырып, фонемдер айтылуының мүмкін нұсқаларын жеткілікті деңгейде толық толтыруды қамтамасыз ететін, жай-күйі ықтималдықтар тығыздығының Гаусс үлестірілімдері қоспалары арқылы сипатталатын жасырын Марков модельдері (ЖММ) ең танымалы болып есептеледі[26].

Сонымен қатар, диалогтық жүйелерде бейретті сөйлейтінімізді және әдетте, абоненттің алғашқы екі-үш сөйлемі белгісіз және түсініксіз екенін есте ұстаған жөн. Бұл сәлемдесу немесе ауытқу үзілісі болуы мүмкін, оның барысында пайдаланушы өз сұрағын тұжырымдап, оператормен байланыс табуға тырысады [27]. Сөйлеу сигналында сөйлеу құру кезінде әртүрлі «кедергілердің» болуы (хезитациялар, аяқталмаған сөздер, сөздерді абайсыз қайталау және т.б.) негізгі сөздерді тану дәлдігін айтарлықтай төмендетеді, себебі жалған гипотезалардың пайда болу ықтималдығы елеулі түрде артады.

Сөйлеуді тану жүйесін жіктеудің бірнеше нұсқасын қарастыруға болады. Жіктеудің негізгі критерийлерінің бірі - танылатын сөздік көлемі болып табылады (1.2. кесте). Жіктеудің әлемдік тәжірибесінде [28] танудың кіші сөздігі деп ондаған сөздер қамтылған сөздікті айтады. орташа танымал сөздік жүздеген сөздерден тұрады, үлкен сөздік мыңдаған және он мыңдаған сөздерді қамтиды, аса үлкен сөздік жүздеген мыңнан бастап миллиондаған сөзді қамтиды да, ол

кейбір тілде барлық қолданыстағы және қолданыс әлеуеті мүмкін сөздерді модельдеуге мүмкіндік береді, оларды шексіз сөздік деп атайды [29, 30].

Кесте 1.2 – Сөйлеуді тану жүйелерінің сөздіктерінің көлеміне қарай жіктелуі

Тану сөздігінің көлемі	Сөздердің саны, V
Кіші	v<100
Орта	100≤v<1000
Үлкен	1000≤v<100 000
Аса үлкен	v≥100 000
Шексіз	v → ∞

Тану жүйесін жіктеудің енді бір мүмкін түрі – пайдаланушының даусына алдын ала икемдеу қажеттігі болып табылады. Мұнда жүйе дикторға тәуелді немесе дикторға тәуелсіз болуы мүмкін. Кіші және орта сөздіктермен жасалған кейбір қосымшаларда кіру сигналын бұрын жазылған сөздер мен сөздер тізбегі үлгілерімен салыстыруға негізделетін сөйлеуді танудың дикторға тәуелді жүйелері оңтайлы (мысалы, рұқсат етілмеген қолжетімділікке жол бермеу үшін) [31].

Бұл жағдайда пайдаланушы, әдетте, жүйе жадына берілген сөздіктен үлгілерінің толық жиынын енгізеді, оны акустикалық деңгейде жүйенің алдын ала оқыту сатысы деуге болады. Сөздік мөлшерінің өсуімен оқыту уақыты да (үлгі сөздер/сөз тіркестерін оқып жаздыру) тікелей өседі.

Дикторға тәуелді емес сөйлеуді тану жүйесі пайдаланушының даусына икемделмей-ақ жұмыс істей алады, яғни кез келген диктор берілген жүйені оқыту процесін жүргізбей-ақ, бірден диалог құруына болады немесе диктор сөздіктің бәрін емес, тек қандай да бір бейімделген мәтінді айтқан жағдайда (сөздік көлемі айтарлықтай үлкен болғанда), жүйе қажетті акустикалық деректер қорын икемдеуі үшін, ішінара баптап алып жұмыс істейді. Мұндай жүйелерде дикторға тәуелді болмауға тіл және сөйлеу процестерін статистикалық модельдеу жолымен жетеді. Ол үшін стохастикалық модельдер құруға мүмкіндік беретін акустикалық деректердің едәуір көлемі қажет болады. Жүйе сөйлеу корпусында оқытылған соң, ол осы корпус ұсынған орташа статистикалық диктордың сөйлеуін дәлме-дәл тануы керек.

Сөйлеуді тану модельдері, негізінен, қолданылатын сөйлеу түріне (сөйлеу стилі) байланысты: оқшауланған сөйлеу, кіріккен, оқылған және ауызекі сөйлеу (мысалы, диалогтарда қолданылатын еркін ауызекі сөйлеудің ең күрделісі болып табылады) [32].

Оқшауланған сөйлеуді тану міндеті қарапайым және шағын сөздік үшін шешілген $V=\{V_1, V_2, ..., V_n, \}$. Сөздіктің элементтері болып сөздердің немесе тұтастай сөз тіркестерінің моделі, сондай-ақ сөйлеу шекараларын анықтау үшін қажетті үзіліс моделі табылады. Мүмкін сөздер танушы болып бөлек класстар

ретінде модельденеді. Сөйлеу сигналы танылу керек кезде, танушы әрбір ықтимал класс үшін бағалауды есептейді де, ең жоғары баллға ие класс үздік шығыс гипотезасы ретінде таңдалады. Мұнда тану сапасы толықтай сөздік көлеміне байланысты. Бір жағынан, сөздердің көп болуы және сөздердің әртүрлі нұсқаларының болуы диктор тарапынан анық емес мәліметтердің орнын толтырады, ал екінші жағынан, сөздік қорын жоғарылату дыбысталуы жақын сөздердің пайда болуына әкеледі, бұл тану дәлдігін төмендетеді. Оқшауланған сөйлеуді енгізген кезде, пайдаланушы сөз арасында жасанды үзіліс жасауы тиіс немесе жүйеде белгілі берілген сөз тіркестерін бірге және нақты айтуы керек. Оқшауланып айтылған сөз тіркестерін танушылар ұқсас жұмыс істейді. Бұл жағдайда әрбір класс сөздердің ұйғарынды тізбегін сипаттайды, ал егер сөз тіркесінде сөз тәртібі сәл өзгеріске ұшыраса немесе кейбір сөз қосылса (қалдырып кетсе) жүйе бүкіл сөз тіркесін тани алмайды.

Кіріккен сөйлеуді тану – бұл күрделі міндет, онда сөздікте сақталған негізгі сөздерден құралған жеке сөз тіркестері танылады. Мұндай жүйелердің сапасы сөздік қорының толықтығына және танылған сөз тіркесіндегі сөздерді байланыстыру ережелерін белгілейтін тіл моделіне байланысты болады.

Кіріккен сөйлеуді танушылар сөйлеу сигналын бір-бірімен байланысты және мағыналы сөздердің дәйектілігі ретінде қарастырады. Әрбір ықтимал сөз жеке класс ретінде модельденеді және сөйлеу сигналына жоғары бағаланған класстардың (сөздер) дәйектілігі беріледі. Бұл рәсім сөздіктегі барлық ықтимал сөздерді модельдеуді талап етеді.

Оқылған сөйлеу оқыту үшін де, сөйлеуді тану жүйелерін сынақтан өткізу үшін де кеңінен қолданылады, себебі бейретті сөйлеуге қарағанда, оны жазу мен өңдеу әлдеқайда оңай және арзан. Мұндай сойлеу стилі тұрақты қарқынға және артикуляцияға ие (дыбыстардың анық айтылуы). Оқылған сөйлеуде, әдетте, сөздердің үзінділері, кекештену, өзін-өзі түзету сынды ешқандай мүдірістер жоқ, өйткені мұндай элементтер, егер олар пайда болса, өңдеу кезеңінде сөйлеу корпусынан алынып тасталады.

Тұтастай алғанда, фонетикалық тұрғыдан қысқа және қарапайым сөйлеулер қолданылады. Айтылған үзілістер, фальш-старттар, түзетулер немесе ұзақ үзілістер жоқ. Тиісінше, сөйлеу ішінде эмоционалды және мағыналы екпін де жоқ. Осылайша оқылған сөйлеу үшін бірсарындылық, интонацияның жоқтығы және фонемдердің айқын айтылуы (сөйлеу дыбыстары) тән.

Ауызекі сөйлеуде сөйлеудің басқа түрлерінде ескерілмейтін және бейретті сөйлеуге оқытылмаған жүйе үшін тану сапасын айтарлықтай нашарлататын көптеген ерекшеліктер бар. Сөйлеу қарқыны сөйлемнің ішінде де, жазба сессиясы кезіндегі сөйлеу арасында да, сессиялар арасында да, әртүрлі дикторлар үшін де тұрақты емес.

Артикуляция да үлкен өзгеріске ұшырайды, маңызды сөздер үшін екпінді формалар және контекстік сөздер үшін екпінсіз формалар оқылып жатқан сөйлеуде сақталмайды. Сөйлемдердің ұзақтығы артады. Бейретті сөйлеуде сөздердің айтылуы айтарлықтай түрленеді. Дыбыстарды есепке алу үшін сөздердің балама транскрипциясы жасалып, осылайша сөйлеу сөздіктерін

кеңейту үшін қажет болатын дыбыстар редукциясы мен ассимиляциясы құбылыстары туындайды.

Микрофон-гарнитураны пайдалану мүмкін емес және диктордың сөйлеуін қашықтан жазуға тура келетін сөйлеуді қашықтықтан тану ерекше жағдай болып табылады. Мысалы, ақпараттық дүңгіршектерде дауыстық функцияларды пайдаланғанда, гарнитураны пайдалану сеанстардың қысқа мерзіміне байланысты эргономикалық емес болып табылады, өйткені пайдаланушы уақытын микрофонды киюге және реттеуге жұмсағысы келмейді. Қашықтықтан жазу кезінде сөйлеу сапасы біршама нашарлайды. Біріншіден, ақпараттық дүңгіршектер, әдетте, қоғамдық орындарда, көшелерде орналасады, онда жағдай соншалықты жақсы емес. Шу компонентін жою үшін тар бағыттағы микрофондар немесе микрофондар жиыны қолданылады. Тар диаграммада бағытталатын магнитофон тек ақпараттық бөлімше жанындағы белгілі бір жұмыс саласынан тарайтын сигналдарды ғана жақсы жаза алады. Мұндай микрофонның кемшілігі – дикторлар бойының әртүрлілігіне есептелмеген.

Микрофондар жиынын қолдану әлдеқайда ұтымды болып табылады. Бірнеше арналар бойынша айналадағы акустикалық жағдайларда бір мезетте жазу – сөйлеу көздері мен шуылды оқшаулауға ғана емес, сондай-ақ пайдаланушының сөйлеуін сыртқы шудан тазарту үшін бағытталу диаграммасын бағдарламалық баптауға мүмкіндік береді. Микрофондар жиыны, сонымен бірге, жиналыстар мен семинарларды жазу кезінде де тиімді. Алайда мұндай жазуларды стенографиялау әзірше қиын, өйткені бірнеше дикторлардың бір мезеттегі үн қату жағдайларының орын алуы мүмкін, содан тану дәлдігі жоғары емес. Сондықтан қашықтықта стенография жасау әзірше қарастырылмайды.

Диалогтық жүйелерге анықтамалық қызметтерден басқа, нақты уақыт режимінде сөзді өңдеуді талап ететін міндеттер класы ретінде, техникалық және күнделікті объектілерді дауыстық бақылау жатқызылуы мүмкін, онда пайдаланушының пәрменіне жауап ретінде жүйе міндетке сәйкес келетін кейбір әрекеттерді орындауы керек. Мұндай тапсырмаларда әдетте, пайдаланушылар тобы шектеулі және алдын ала белгілі, сондықтан мұндай жүйелер көп дамыған, тіпті коммерциялық жүйелер де бар. Мысалы, кейбір ұялы телефондар мен смартфондарда түйінді атаулар бойынша телефон нөмірлерін танудың белгілі бір опциясы бар. Одан басқа, MP-3 плеерді басқарудың дикторға тәуелсіз моделін BMW компаниясы IST 2006 халықаралық көрмесінде көрсеткен.

Сөйлеуді танудың қолданбалы міндеттерінің басқа тобы болып стенографиялау табылады. Көбінесе мұндай міндеттерде белгілі бір монолог микрофон-гарнитураның көмегімен жақсы акустикалық жағдайларда жазылады. Сондықтан сөйлеудің телефон арналары арқылы және/немесе көшеде жазылу арқылы жүретін бұқаралық қызмет көрсету жүйелеріне қарағанда, автоматты стенографиялау жүйелері әлдеқайда жоғары сапада жазуға мүмкін болатын сөйлеу сигналдарын алады. Өйткені мұнда танудың жылдамдығы бойынша неғұрлым қатаң емес талаптар қойылады, сөйтіп диктор дауысына және қолданбалы міндеттерге бейімделген әдістерді қолдана отырып, жүйе сөйлеу

сигналдарын бірнеше ғана талпыныста өңдей алады [21]. Фонема модельдерін біртіндеп қайта оқыту және сөздікті жүйеге бұған дейін таныс емес сөздермен толтыру арқылы танудың максималдық дәлдігіне қол жеткізіледі. Диалогтық және стенографиялық жүйелерге қойылатын талаптардың салыстырмалы кестесі 1.3 кестеде берілген:

Кесте 1.3 – Сөйлеуді тану жүйелеріне қойылатын талаптар

Талап	Сөйлеуді танудың қолданбалы міндет топтары	
	Диалогтық жүйелер	Стенографиялау
Сөздік көлемі	Орташа (500 сөзге дейін)	Аса үлкен (100 000 астам сөз)
Дикторға тәуелсіз	Міндетті	Қалауы бойынша бейімдеу мүмкін
Нысан саласы	Тар (диалог стратегиясы арқылы шектелген)	Шектелмеген (әрбір монолог өзінің пәндік саласынан алынады)
Жазу жағдайлары	Телефон, көше шуылы	Микрофон, тыныш бөлме
Өңдеу уақыты	Нақты уақыт режимінде	Салыстырмалы түрде шектелмеген

Стенографиялау жүйелерінің тағы бір ерекшелігі – аса үлкен сөздікпен жұмыс істеуінде. Мұнда бір-біріне жақын естілетін сөздердің кездесулеріне және сөздердің орын тәртіптерінің ауысып келуіндегі ұқсастықтарға байланысты қателер пайда болады. Нәтижесінде ықтималдығы жоғары бірдей жорамалдар бір мезетте көрініс береді, бірақ олардағы сөздер әртүрлі болады. Мұндай ұқсастықтар көп жағдайларда тілдің стохастикалық моделі және пән саласының онтологиясын қолдану есебінен жойылып отырады. Алайда семантикалық білімдерді белгілі бір қалыпқа келтіру лингвисттер үшін де, жасанды интеллект саласының мамандары үшін де шешілмеген түйіткілікті мәселе болуда [33 - 35]. Аталмыш зерттеуде тілдің семантикалық деңгейі қаралмайды, өйткені бұл сала арнайы зерттеуді қажет етеді, осыған орай оны мәтінді автоматты түрде өңдеумен және машиналық аударумен айналысатын көптеген мамандар зерттеуде [36 - 39].

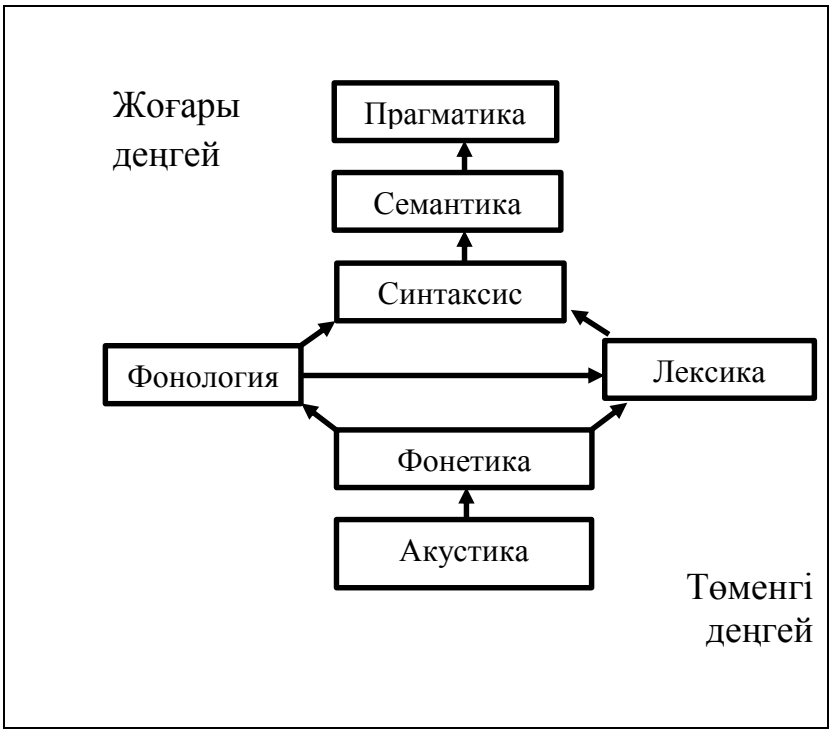
1.2 Сөйлеу сигналдарын автоматты өңдеу әдістері

Сөйлеуді қабылдаудың толық моделі тілдік сигналдың өз мағынасында жеткізілген хабарламаға айналуын қамтамасыз етуі тиіс. Толық модельдің мүмкіндікті құрылымын талдау оның ақпаратты өңдеудегі көптеген желілік/параллельдік блоктардан тұру керектігін көрсетеді. Тіл туралы ғылымның іргелі қағидаларының бірі – хабарламаның мағынасына көшуі тек сөйлеу сигналының дискреттік элементтер желісіне айналғанда ғана мүмкін деген тұжырым. Бұдан әріде сөйлеуді танудың базалық жүйелерінде

қолданылатын сөйлеу сигналдарының түрлендірілуінің негізгі деңгейлері қарастырылады.

1.2.1 Табиғи тіл мен сөйлеуді өңдеудің негізгі деңгейлері

Соңғы отыз жылда табиғи тілді өңдеу мен оны ұсыну бойынша көптеген аналитикалық шолулар жасалынған [28 – 38, 40 – 44]. 1.1-суретте сөйлеу байланысы сипаттамаларының бірнеше деңгейлері көрсетілген табиғи тілді өңдеудің жалпы сұлбасы берілген [45, 46]. Ең алдымен, сөйлеуді ұсыну мен өңдеудің екі базалық деңгейі: жоғары және төмен – бар екенін айта кеткен жөн. Деңгейі төмен болған сайын, дыбыстық сигнал физикалық сипаттауға жақын. Мұндай жоғары деңгейден төменгі деңгейге қарай сипатталатын жіктеуді өзара әрекеттестіктер мен процестердің барлық түрлеріне қолдануға болады, өйткені физикалық сигналдар мен олардың мағыналық қабылдануын әрқашан ажыратуға болады.



Сурет 1.1 – Сөйлеуді өңдеудің негізгі деңгейлері

Бұдан басқа тілдік қарым-қатынасты модельдеуде өңдеудің қандай деңгейлерін, білімнің қандай түрлерін қолдануға болатынын ескеру керектігі қаралады.

Акустикалық деңгей. Сөйлеу – дауыс жолы арқылы ауа қысымының өзгерілу есебінен пайда болатын дыбыстардың тізбегі. Аудио-сигналға ғана тікелей қатысты болғандықтан, акустикалық деңгей ең төменгі болып табылады. Акустикалық процестерді зерттеу акустикалық сигналды жазуды, сандық өңдеуді, параметрлеуді қамтиды [47 – 49]. Мұнда әрбір сатыда түрлі көріністер пайда болады: микрофоннан келетін сигнал, (аналогтық, сандық), толқындық

қалыптар, спектрлік сараптау (Фурье түрлендірулері, спектрограммалар т.б.). Өңдеудің басқа түрлерінде де пайдалы ақпараттар алуға болады (негізгі тон жиілігі, энергия, кепстр). Жалпы айтқанда, акустикалық деңгей сөйлеу сигналдарын кодтау, параметрлік ұсыну әдістерімен айналысады.

Сөйлеу сигналдарын модельдеу мен өңдеуде біршама болжамдар қолданылады. Дауыс жолы – бұл физикалық аспап, ол белгілі бір инерцияға ие және қалып күйдің шұғыл өзгерісіне жол бермейді, сондықтан сөйлеу сигналын квазистационарлық сигнал ретінде қарастыру керек, яғни біраз уақытқа дейін сигналдың сипаттамалары тұрақты тұрады. Эксперименттік мәліметтерге сүйенсек, үзілістер тәрізді қысқа сөйлеу актілерін қоспағанда, 30 мс шамасындағы қысқа уақыттық терезелерде сөйлеуді стационарлық сигнал деп тануға болады. Сөйлеу сигналдарын өңдеуде мұндай болжамды қолдана отырып, ұзақтығы бірдей (30 мс дейін) жүйелік учаскілерді анықтау үшін алдын ала сегменттеу жүргізіледі. Содан кейін параметрлік сегменттеу жүргізіледі. Содан соң параметрлеу сегменттер бойынша атқарылады, бұл шығатын деректерді едәуір қысқартады және реттейді.

Фонетикалық деңгей. Бұл деңгейде артикуляция жүйесі тудырған дыбыс деңгейіндегі сөйлеу сигналының қасиеттеріне назар аударылады [50]. Сөйлеу жасалымының физикалық процесін қысқаша сипаттауға болады. Тыныс алу - бұл энергияның негізгі көзі. Ауа дыбыстық жолдан өтіп, оның түрлі тосқауылдары ауыз бен мұрыннан шығатын ауа қысымын модельдейді. Фонетика адамдар сөйлеудің нақты дыбыстарын айтқан кезде тілдің, еріннің, тістің және басқа да органдардың көмегімен дыбыс жолының геометриялық сипаттамаларын өзгерту үшін оны бұлшық еттері басқаратынын зерттейді.

Фонологиялық деңгей. Бұл – кейбір семантикалық мәні бар бірліктердің көрініс беретін бірінші деңгей. Шын мәнінде де, фонетика әртүрлі дыбыстардың табиғи туындау мәселесін қарайтын болса, фонология белгілі бір тілде мүмкіндікті жекелеген дыбыстардың шектеулі санының айтылуында пайда болатын ырғағын, нақты сөз тіркесінде қолданылатын әуезділігін, әрбір айтылған фонеманың (просодияның) негізгі тонының жиілігін және сөз тіркесіндегі мағыналық екпінді сараптаумен айналысады. Қытай тілі сияқты кейбір тілдерде, әуезділіктің өзгеруі сөз мағынасын толық өзгертіп жібере алады. «Фонема» түсінігі айтарлықтай құрылымды: ол сөзді оған сәйкес келетін артикуляциялық ымдармен және сөйлеу сигналдарымен байланыстырып қана қоймай, сөйлеу сигналының естілу бейнесін қозғалыстармен де байланыстырады. Фонемаларға мағыналық жағынан жақын түсініктерді қолданбай, естілген мағынасыз дыбыс тіркестерін имитациялауды (сөйлеуге қайта айналдыруды) және оның әріптік жазбалары сияқты құбылыстарды сипаттау мүмкін емес. Бұл деңгей жоғары және төмен деңгейлердің арасындағы өтпелі деңгей болып табылады, өйткені бұл жерде сөйлеу сигналының параметрлік көріністерін, тілдің өзіндік арнаулы ерекшеліктерін және просодикалық сипаттамаларын бір мезетте ескеруге тура келеді. Дауыстың негізгі тонының жиілігін сараптау кейде сөйлеу сигналындағы, сондай-ақ әр

түрлі интерактивті өңделінетін бағдарламалардағы көңіл-күйді анықтау үшін қолданылады.

Лексикалық деңгей. Жоғарыда айтылғандай, әрбір тілде түрлі дыбыстардың (фонемалардың) ақырғы саны бар. Дегенмен, белгілі бір тілдің дыбыстарының ерікті тізбегін айту және осылайша мағыналы сөз тудыру мүмкін емес. Сондықтан лексикалық деңгейде сөздерді білдіретін фонемалардың барлық маңызды дәйектемелері сипатталады. Лингвистикада бұл деңгейді кейде морфологиялық деңгей деп атайды, мұнда мағыналық жүктемесі бар сөздердің қарапайым құрамдас бөліктері зерттеледі.

Мұнда сипаттаудың бірліктері болып сөз немесе сөз бөлшегі (морфемалар) танылады, ал жоғары деңгейдегі өңдеу физикалық аудио-сигналдардың қатысынсыз-ақ, тек мәтін бойынша жүргізіледі. 1-ші суретте фонологияның лексикамен де, фонетикамен де тығыз байланысы көрсетілген. Шынында да, белгілі бір тілдерде (қытай тілінде) әртүрлі просодиялық модельдерді әртүрлі сөздерді, яғни, өзгеше мағына тудыра отырып, фонемалардың ұқсас дәйектемелеріне қолданады.

Синтаксистік деңгей. Әрбір фонемдердің тізбегі сөз бола алмайтыны сияқты, сөздердің де тізбегі сөйлем бола алмайды және сөздердің тізбектерін шектейтін кейбір ережелер жиынтығы бар екенін атап өткен жөн. Сөздердің ықтимал комбинациясын сипаттайтын грамматикалық немесе стохастикалық ережелер жиынтығы синтаксистік деңгейдің мәнін құрайды. Әрбір тіл үшін сөйлемдегі әрбір сөздің қызметін анықтайтын өзінің грамматикалық жиыны болады, осылайша оның синтаксистік құрылымы сипатталынады. Жалпы алғанда, қатаң лингвистика сипаттайтын, грамматикалық ережелер жазбаша сөйлеуге тән, ол ауызекі сөйлеуді модельдеуге келмейді. Сондықтан ауызекі сөйлеуді автоматты талдау кезінде грамматикалық ережелерді сақтықпен пайдаланған жөн, өйткені олар қатаң шектеулер қойып тұрады, нәтижесінде түсінікті (бірақ қайсыбір дәлсіздіктері бар) сөз тіркестерін жүйе қабылдамай тастауы мүмкін. Есептеуіш лингивистика бойынша алғашқы жұмыстарда-ақ ауызекі және жазбаша сөйлеулердің өзара сәйкессіздігі туралы проблемалар көтеріліп, сөйлем ішіндегі байланыстар мен құрылымдарды бағалаудың неғұрлым икемді, «жұмсақ» жолдарын іздеу қажетілігі туындаған болатын [51, 52].

Семантикалық деңгей. Сөйлеу синтаксистік тұрғыдан дұрыс болғанының өзінде оның мағыналы ақпаратты қамтитынына кепілдік жоқ. Сондықтан сөйлемнің түсініктілігі мен мәнін анықтау сөйлеу коммуникациясын сипаттаудың келесі қадамы болып табылады. Семантикалық денгейде сөздердің мәнмәтіннен тәуелсіз мағыналарын және олардың сөйлем ішінде үйлесуін зерттеу басталады.

Прагматикалық деңгей. Прагматика коммуникация процесіндегі барлық мәнмәтіннен тәуелді ақпаратты тұтасымен жинақтайды және бағалайды. Мұнда тек күнделікті сөйлеу актісі процесінен алынған ағымдағы ақпарат қана емес, сонымен бірге диалогқа қатысушыларда бар барлық білім және қоршаған орта жағдайлары да қатысады. Шынында да, әңгімелесушілер әртүрлі нәрселер

туралы сөйлескенде бір-бірін бірден түсінбейді, дегенмен олар барлық сөздерді және тіпті сөз тіркестерінің дұрыс грамматикалық құрылымын қабылдайды. Мұндай жағдайлар әңгіме барысында бір тақырыптан екінші тақырыпқа жылдам көшкенде немесе әңгімелесушілердің ортақ білімдері өзара сәйкес келмеген жағдайда жиі пайда болады. Бұл білімдер ең алдымен, тұлғаның әлеуметтік-мәдени жағдайына немесе диалогтың аясына байланысты. Кейде прагматикалық деңгейді үш ішкі деңгейлерге бөледі [53]:

– Таза прагматикалық деңгей: әртүрлі бағытта айтылған жеке пікірді жеткізе алатын саналуан мәндерді зерттеу;

– Дискурс деңгейі: алдыңғы айтылған сөздің келесі айтылатын сөздің түсіндірілуіне қалай ықпал ететініне байланысты. Бұл анафора мен басқа тілдік белгісіздіктерді шешу үшін пайдалы болуы мүмкін;

– Әлем туралы білімдердің деңгейі: кейде бұл базалық білім болып табылады, сөйтіп адамның әлем туралы бар білетінін (аспан-көк, қар-суық т.б.), сондай-ақ сөздік қатынас қатысушысының өзге қатысушылардың көзқарастары мен мақсаттары туралы білетіндігін қамтиды.

Табиғи жағдайдағы сөздік қатынас әртүрлі акустикалық кедергілер мен ауытқулар жағдайларында өтетінін айта кеткен жөн. Міне, осындайда ақпараттың жоғалуы туындайды. Параметрлік түсініктен семантикаға дейінгі барлық деңгейлерде сөйлеу сигналын сипаттау кезіндегі мәнін тану дәлдігін арттыру үшін бір емес, едәуір болжамдар жиыны қалыптасады, олар келесі деңгейлерде түсіп қалып отырады. Осыған байланысты, тіпті бұрмаланған хабарлар дұрыс танылу ықтималдығын сақтайды.

Сөйлеуді автоматты танудың қазіргі заманғы жүйелерінің көпшілігі тіл туралы және қазіргі жағдай туралы әртүрлі ақпарат түрлерін ескеретін көп деңгейлі иерархиялық құрылымға ие. Ақпараттың жеткіліксіздігі немесе шулы орта жағдайында әрбір деңгейдің қорытынды шешімі жалғыз ғана емес (мысалы, фонемдердің бір тізбегі), мұндайда бірнеше ықтимал гипотезалар қалыптасады. Олардың арасындағы таңдау келесі өңдеу деңгейінде жасалады. Автоматты сөйлеуді тану туралы қарастырылған мақалалар осы салада әлі де жауабы табылмаған көптеген сұрақтар бар екенін көрсетеді.

1.2.2 Кіріккен сөйлеуді тану жүйесінің негізгі архитектурасы

Сөйлеуді автоматты тану деп сөздердің тиісті тізбегімен диктордан микрофон арқылы келіп түсетін үзіліссіз сөйлеу сигналын айтамыз. Табиғи сөйлеуді тану – бейнелерді танудың негізгі міндеттерінің бірі. Кірме бейнені жекелеген тиісті класстарға жатқызу бейнелерді тану міндеті болып табылады. Ықтимал класстар қолдануына (сарапталынатын мәліметтер типіне) байланысты модельденеді. Тану мақсаты болып қорытынды классты (қорытынды гипотезалар) бағалау танылады. Оған мейлінше ықтималдық мөлшердегі келіп енетін мәліметтер жатады.

Сөйлеу сигналы бойынша диктор айтқан сөздердің тізбегін дұрыс анықтау сөйлеуді тану жүйесінің міндеті болып табылады. Бұл мына төмендегі ұтымды критерийге сәйкес көрсетіледі [54]:

$$\hat{w} = \underset{w \in W}{\operatorname{argmax}} P(O|w) \tag{1.1}$$

мұнда $\hat{w}$- сөз тіркесінің қорытынды гипотезасы;

w - айтылған сөздер тізбегі;

W- сөздер (гипозелар) тізбегінің барлық мүмкін жиыны;

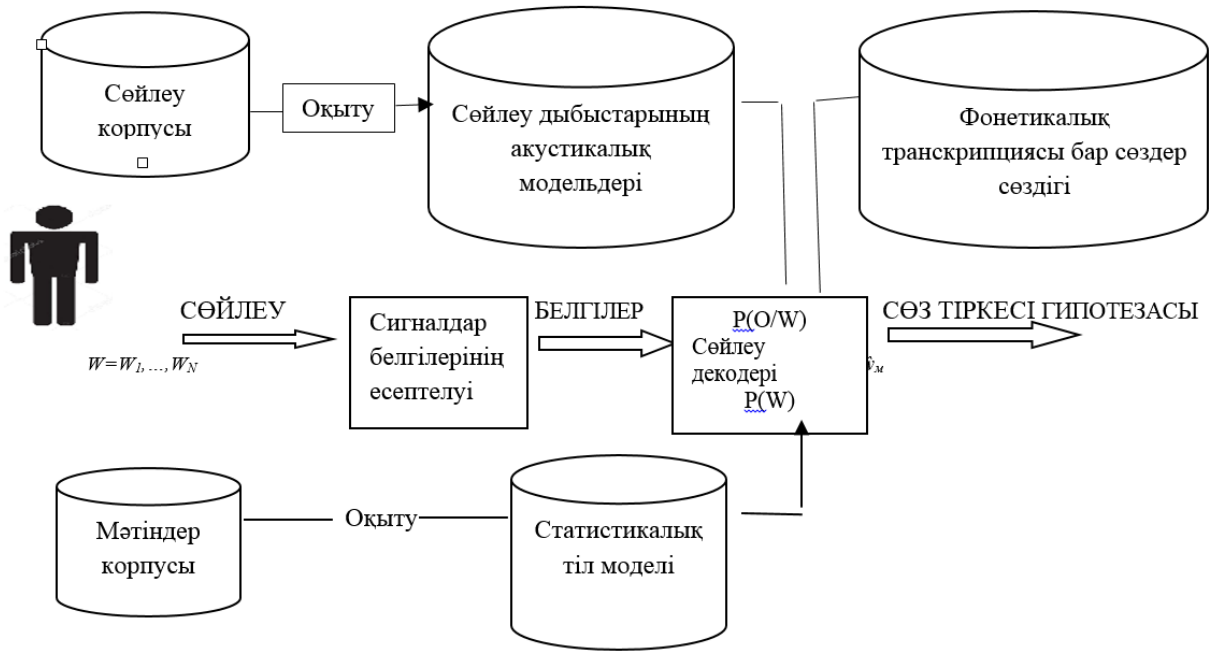
O – келіп кіретін сөйлеу сигналы бойынша есептелген векторлар белгілерінің тізбегі.

Байес формуласын қолданғанан кейін 1.1. формуласы 1.2 формуласына түрлендіріледі. $P(O)$ айтылған сөздердің w тізбегіне байланысты өзгеріске ұшырамайды, сондықтан $P(O)$-ны елемей, апостериорлық ықтималдылық максимумының критерийін ала аламыз:

$$\hat{w} = \underset{w \in W}{\operatorname{argmax}} P(O|w)P(w) \tag{1.2}$$

Мұнда $P(O|w)$ – диктор сөздер тізбегін ω білдірсе, ағымдағы вектор белгісінің O байқалу ықтималдығы. Бұл теңдеу акустикалық ықтималдылық деп аталады, ол сөйлеу сигналының декодер модельдерінің көмегімен есептелініп шығарылады. P(w) – тілдік модельдер көмегімен есептелетін сөз тіркестеріндегі кейбір сөздердің априорлық ықпалдылығы.

1.2 суретте сөйлеуді автоматты танудың базалық жүйесінің сұлбасы көрсетілген. Бұл жүйе екі режимде жұмыс істейді: оқыту және тану. Ең алдымен, акустикалық моделдер мен тіл модельдерін оқыту жүзеге асырылады. Акустикалық модельдер сөйлеу мәліметтері мен олардың белгілі транскрипциялары бойынша есептелген белгілер векторлары бойынша оқытылады. Тіл моделі пәндік сала бойынша мәтіндегі сөздердің комбинациясының кездесу жиілігін бағалау жолымен түзіледі.



Сурет 1.2 – Кіріккен сөйлеуді тану жүйесінің негізгі архитектурасы

Тану кезінде кірістік сигнал белгілері векторларының тізбегіне айналады, сөйтіп алдын ала оқытылған сөйлеу дыбыстарының (фонемалардың) модельдері мен тіл моделінің немесе грамматиканың көмегімен мейлінше ықтимал гипотеза немесе көптеген таңдаулы гипотезаларды іздеу жүзеге асырылады [54].

1.2.2.1 Сигналдық параметрлік көрінісі

Тану жүйелерінде сөйлеу сигналы қысқа сегменттерге бөлінеді және әрбір сегмент белгілер векторына айналады, нәтижесінде кіріс сигналының белгілері векторлар жүйелілігімен ұсынылады. Белгілердің векторларын есептеу процесі белгілерді алу немесе параметрлік көрініс деп аталады. Сөйлеуді тәуелсіз белгілер векторларының тізбегін жасайтын стохастикалық (кездейсоқ) процесс ретінде көрсетуге және модельдеуге болады. Стохастикалық процесс – бұл, барысы жағдайға байланысты әртүрлі болуы мүмкін және ол бір немесе басқа ықтималдығы анықталатын процесс [55]. Белгілерді алу процесінің мақсаты кіріс сигналдарын параметрлік ұсынудың кейбір ықшам түріне айналдыру болып табылады.

Микрофондар ауада тарайтын дыбыс толқындарының қысымын ұстайды, сөйтіп оларды электрлік сигналдарға айналдырады. Белгілерді алу процессінің алдында бұл сигналдар алдына ала сандық түрге келтірілген болуы тиіс. Стандартты аналогты-цифрлық түрлендіргіш (ADC) үздіксіз сигналды уақыт бойынша дискреттейді және сигналдың амплитудасын талап етілетін мәндер диапазонында өлшейді.

Бірінші қадам – уақыттық саланы дискреттеу көмегімен үздіксіз аналогтық сигналды дискреттік сигналға айналдыру болып табылады. Сигналдың шамасы t_0 белгілі бір уақытта дискреттеу кезеңімен өлшенеді. Дискреттеу жиілігі $f_n = \frac{1}{t_n}$ болып анықталады. Екінші қадамда жазылған сигналдың амплитудасаның

үздіксіз мәндері оларды машиналық кодта разрядтардың ақырғы санымен ұсыну үшін дискредтеледі. Бұл қадам кванттау деп аталады. Қазіргі тану құралдары көбінесе бір АЦТ есептеуді (бір сигнал мәнін) сақтау үшін 16 бит пайдаланады. Алдын ала өңдеудің соңғы қадамы – ауытқуды алдын ала сүзгілеу, мұнда сигнал бірінші қатардағы жоғары жиіліктің рекрусивтік сүзгі көмегімен ықшамдалады. Мұндай алдын ала бұрмалаудың мақсаты – сөйлеу сигналында жоғары жиіліктердің, әсіресе сөйлеу дыбыстарының ұяң дауыссыздар үшін төмендеуіне салдарынан төмендетілген амплитудалы сөйлеу дыбыстарының жоғары жиілікті компоненттерін көрсету болып табылады.

Белгілер векторлары әдетте сөйлеуді қысқа мерзімді кезеңде (қысқа мерзімді талдау) стационарлық деп санауға болатын болжамды пайдаланып, қысқа сигнал сегменттері үшін есептеледі. Сигналды нақтырақ сипаттау үшін, сөйлеу сегменттері жаңадан алынады.

Сөйлеу сегменттерін құру процесі терезе әдісімен жүзеге асырылады, яғни терезенің кейбір шекараларындағы бос орындардың әлсіреуі үшін сигналды бірнеше терезе функциясымен көбейту арқылы жүзеге асырылады. Әдетте осы мақсаттар үшін Хэмминг терезесі пайдаланылады, бұл жағдайда терезе функциясы мынадай түрге келтіріледі:

$$W[k]=\begin{cases} 0.54-0.46\,cos\left(\frac{k2\pi}{K-1}\right),\ k=0.1\ldots K-1 \\ 0\ ,k\neq 0.1\ldots K-1 \end{cases}, \qquad (1.3)$$

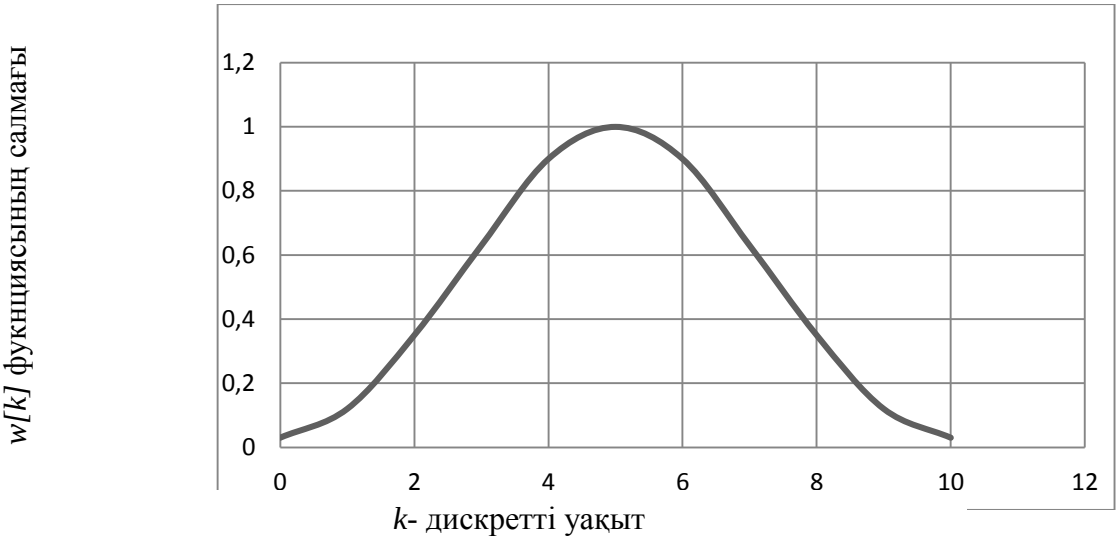
мұнда *к*– терезе ені.

Бағдарламалық іске асыруда алдына ала бұрмалау сүзгісін есептеу және Хэмминг терезесін қолдану бір мезгілде бір рекурсивті рәсімде орындалады. 1.3-суретте 10 есептен тұратын құрылымы бар Хэмминг терезесі көрсетілген. Белгілерді анықтаудың келесі қадамы – дискредті Фурье түрлендіруін қолдану арқылы әр фреймді уақытша саладан жиілік саласына (сигнал спектрін есептеу) түрлендіру болып табылады. Бұл қадам, әдетте, жылдам Фурье түрлендіруі ретінде орындалады, бұл дискретті Фурье түрлендіруін тиімді жүзеге асыру болып табылады.

Фурье түрлендіру жылдамдығын тиімді ету үшін айналдырылған ұзындық 2 дәрежелі болуы керек. 500 санақтан құралған сегменттер үшін 512 санақ болуы үшін сегменттің соңына нөлдерді қосу арқылы түрлендіру ұзындығы 512 болуы керек. Жылдам Фурье түрлендіру жылдамдығының квадраты есептеледі, өйткені жиіліктің қуат мәнін есептеу қажет.

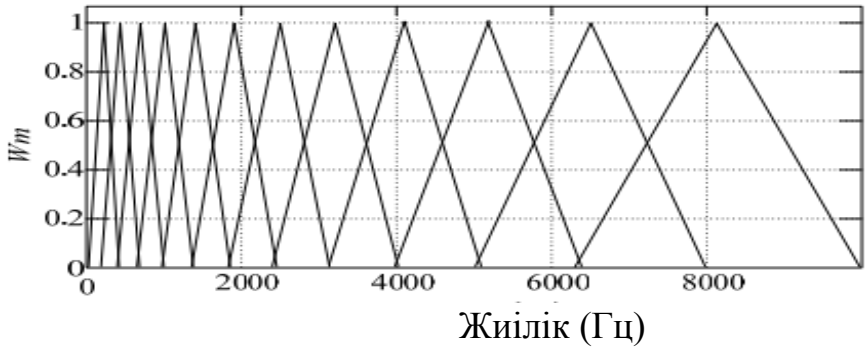
Адамның есту жүйесінің дыбыс энергиясын қабылдауын дәлме-дәл модельдеу үшін әрбір фрейм үшін мел-сүзгілер тарамы көмегімен талдау жасалынады. Мел-фильтр мел-шкалаға негізделеді, ол адамның есітуін қабылдауына ұқсас логарифмдік шкала болып табылады. Ұқсас шкала төмендегідей герцтермен өлшенетін f жиілік шкаласына салыстырмалы формуламен анықталады:

$$f_{mel}=2595*lg\left(1+\frac{f}{700}\right) \quad (1.4)$$



Сурет 1.3 – Сөйлеу сегментін өңдеуге арналған Хэмминг терезесінің үлгісі

Мел-сүзгілердің тарамы үшбұрышты қабаттасып жасалған ауырлық функциялары көмегімен орындалады.Бұл үшбұрышты ауырлық функциялары мел-шкаладағы бірдей өтпелі жолақтың 50% бөгей отырып жабады (яғни мел-сүзгі көршілес сүзгілердің орталық жиіліктері арасындағы жиіліктерді қамтиды). Сондықтан олар мел-шкалада бірдей аймақ болып көрінеді, ал жиілік диапазонындағы ені логарифмдік жиілікте өседі. Фильтрлердің ені мел-шкалада біркелкі болуы мүмкін, содан барып жиілік саласында 1.4. формуласы көмегімен көрініс табады. 1.4-суретте жиілік осіндегі 12 жолақтық мел-сүзгілердің жиынтығынан тұратын үшбұрышты ауырлық функцияларының үлгісі көрсетілген.



Сурет 1.4 – 12 жолақты мел-жиілік сүзгілерінің тарағы

Автоматтандырылған талдау кезінде спектрдің энергиясын әр мел-жолақта (сүзгі) бағалау қажет. Әрбір мел-сүзгідегі жиілік бойынша ажыратымдылықтың

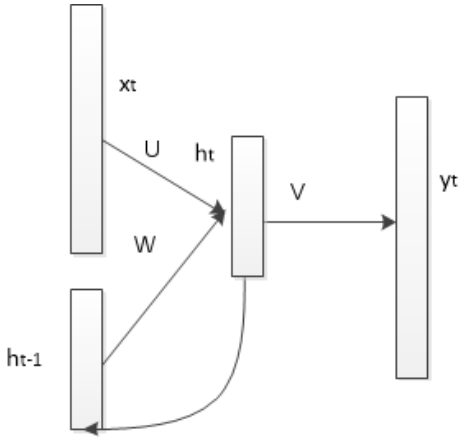
барлық элементтерінің энергиясынан тұратын вектор – мел-спектр болып табылады.

Мел-сүзгіштерінің үшбұрышты ауырлық функцияларын бөгейтініне байланысты, мел-спекторлы вектордың көршілес компоненттері бір-бірімен өзара байланысады. Логарифмдік мел-спектрде дискретті косинус түрлендіруін қолдану негізінде көршілес компоненттер арасындағы өзара корреляция айтарлықтай азаяды. Цифрлық сигналдарды өңдеу кезінде мел-жиілік кепстралды коэффициенттері пайда болады.

Мел-жиілік кепстралды коэффициенттер көбінесе басқа параметрлерге қарағанда жұмысты жақсартатындықтан, сөйлеуді тану кезінде белгілер векторларының жиынтығын пайдаланады. Көптеген сөйлеуді тану жүйелері спектралдық өзгерістерді ұстау үшін және сөйлеуді автоматты өңдеу мен тану жүйесінің жұмысын жақсарту үшін бірінші және екінші қатардағы туынды коэфициенттерді біріктіреді [56, 57].

1.3 Рекуренттік нейрондық желілер

Рекурренттік нейрондық желілер – элементтер арасындағы байланыстар бағытталған тізбекті түзейтін нейрондық желілер түрі. Осыған орай реттелген кеңістікті тізбелер немесе оқиғалар сериясын өңдеу мүмкіндігі туады.Көпқабатты перцептрондарға қарағанда, рекурренттік желілер кез келген ұзындықтың тізбегін өңдеу үшін өзінің ішкі жадысын қолдана алады. Сондықтан RNN желілерін, небір бүтіннің сегменттерге бөлінген есептерінде пайдалануға болады, мысалы: қолжазбаны тану немесе сөйлеуді тану. Рекуренттік желілер үшін, қарапайымынан бастап күрделілеріне дейінгі көптеген әр түрлі архитектуралық шешімдер ұсынылған болатын. Соңғы уақытта ұзақ мерзімді және қысқа мерзімді жадысы (LSTM) бар желілер және басқарылатын рекуренттік блок (GRU) кең тарады.



Сурет 1.5 – Элманның рекуренттік желісінің жалпы көрінісі. x_t – t қадамындағы кіріс қабат; h_t – жасырын қабат; y_t – шығыс қабат

Ең алдымен, рекуренттік нейрондық желі термині ретінде, 1990 жылы Элман ұсынған, рекуренттік архитектура түсіндіріледі (1.5-сурет) [58]. Дәлме-

дәл айтсақ, берілген архитектура жалғыз емес. Элман желісі одан бұрынғы Джордан желісі [59] мен Хопфилд желісіне [60] қарама-қайшы. Дегенмен, төменде, ыңғайлылық үшін рекуренттік нейрондық желі ретінде тек Элман архитектурасы түсіндіріледі.

Элман рекуренттік желісінің t қадамында алынған h_t жасырын қабаты келесі $t + 1$ қадамында желінің кірісіне түсетін екіқабатты нейрондық желіні білдіреді.

($x^{(t)}$, $y^{(t)}$) жұбынан тұратын, бірдей ұзындықтың уақыттық қатарларына тәуелді $\mathcal{D}$ іріктемесін қарастырайық. Мұнда ($x^{(t)}$) және ($y^{(t)}$) U және T жиындарында сәйкесінше анықталсын. Онда рекурренттік нейрондық желіде P($y^{(t)}|x^{(t)}$) шартты үлестірімін жуықтайтын функция бар, ол келесі формулалар бойынша анықталады:

$$h_t = f(W \cdot h_{t-1} + U \cdot x_t + b) \tag{1.5}$$

$$y_t = g(V \cdot h_t + d) \tag{1.6}$$

W,U,V – салмақ матрицалары, b, d – жылжытулар, $x \in$ U – t қадамындағы предикторлық реттіліктің элементі, $y \in$ T – сол t қадамындағы белгісіз реттіліктің ықтимал элементтерінің үлестіруі, $h \in$ H – желінің жасырын қабаты, f және g - активтендіру функциялары.

Мұнда f ретінде сигмоидты функция жиі алынады, немесе түзеткіш (rectifier linear unit, ReLU) [61]. g ретінде – көпкласссты логит-функция (softmax) сызықтық активтендіріледі.

Айта кету керек, алайда формалды түрде ($x_{(t)}$),($y_{(t)}$) реттіліктерінің ұзындықтары бірдей болса да, оқыту кезеңінде шығын функциясын бізді қызықтыратын нүктелерде ғана есептеу жеткілікті, демек($x^{(t)}$) және ($y^{(t)}$) шындығында әр түрлі ұзындыққа ие. Мысалы, мәтіннің эмоциялы үндестілігін бағалау есебінде, қатені бүкіл мәтінді тұтастай оқу арқылы есептеу мүмкін болғанда, бұл ыңғайлы болады.

1.4 Рекуренттік нейрондық желілерді оқыту

1.4.1 Уақыт бойынша қатені кері үлестіру алгоритмі (BPTT)

Бүгінгі күні, нейрондық желінің θ моделінің параметрлерін таңдау үшін қолданылатын және оңтайландыру әдісінің негізгісі болып градиенттік түсу әдісі саналады. Нейрондық желілік параметрлерді таңдаған жағдайда, градиенттік түсу танымал қатені кері үлестіру алгоритміне әкеледі [62].

Рекуренттік нейрондық желілер үшін ұқсас тәсіл аталған уақыт бойынша қатені кері үлестіру алгоритміне (Backpropagation through time, BPTT) алып келеді [63].

Қайсыбір t қадамында рекуренттік нейрондық желінің шығысын қарастырайық:

$$y_t = g(V \cdot h_t + d) = g(V \cdot f(U \cdot x_t + W \cdot h_{t-1} + b) + d) =$$

$$g(V \cdot f(U \cdot x_t + W \cdot f(U \cdot x_{t-1} + W \cdot h_{t-2} + b) + b) + d) =$$
$$g(V \cdot f(U \cdot x_t + W \cdot f(U \cdot x_{t-1} + W \cdot f(\ldots f\ (U \cdot x_1 + W \cdot h_0)\ldots) + b) + b) + d) \quad (1.7)$$

Реттілік жұбының әрбір қадамындағы қателер сомасы ретінде, $\mathcal{L}$ қате функциясын анықтайық:

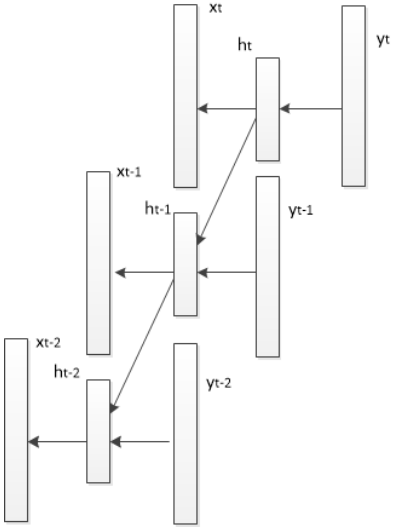
$$\mathcal{L}(\theta) = \sum_{1 \le t \le T} \mathcal{L}_t(\theta) \quad (1.8)$$

$\frac{\partial \mathcal{L}}{\partial \theta}$ градиенттері үшін келесі өрнектерді аламыз:

$$\frac{\partial \mathcal{L}}{\partial V} = \sum_{1 \le t \le T} \frac{\partial \mathcal{L}_t}{\partial y_t} \frac{\partial g(V \cdot h_t + d}{\partial d} \quad (1.9)$$

$$\frac{\partial \mathcal{L}}{\partial V} = \sum_{1 \le t \le T} \frac{\partial \mathcal{L}_t}{\partial y_t} \frac{\partial g(V \cdot h_t + d}{\partial d} \quad (1.10)$$

1.7 формуладан көруге болады, егер 1.7 формуласын 1-ден t-ға дейін толықтай жазып көрсетсек, $\frac{\partial \mathcal{L}}{\partial W}$ үшін өрнек анық түрде алынуы мүмкін. Егер 1.8 көрінісіне нейрондық желі түрінде қайта оралсақ, онда біз 1.6 суретте көрсетілгендей графты аламыз. Шындығында, біз $t-1$ қабаттары бар көпқабатты нейрондық желінің толық жұмыс циклінің нәтижесі мен W синапстарының бірдей матрицасы ретінде, y_t шығысын есептеуді ұсындық. Осылайша, $\frac{\partial \mathcal{L}}{\partial W}$ есептеу үшін біз, уақыт бойынша бастапқы рекуренттік желіні жаю жолымен алынған, нейрондық желіде уақыт бойынша қатені кері үлестіру алгоритмін пайдаланамыз.



Сурет 1.6 – Уақыт бойынша қатені кері үлестіру алгоритмін есептеу схемасы.y_k – t қадамындағы шығыс қабаты; x_t – кіріс қабаты; h_t – жасырын қабат. Нұсқаушы қатені үлестіру бағытын көрсетеді. Уақыт бойынша желіні жаю 2 қадам бойынша жасалады

Барынша формалды түсіндірулерге көшу мақсатында ыңғайлылық үшін лездік туынды ұғымын енгіземіз.

Анықтама 1 (Лездік туынды). h_t рекурсивті келісілген $h_t = f(h_{t-1},\theta)$ өрнегімен есептелсін. $\frac{\partial+\mathrm{ht}}{\partial\theta}$ лездік дербес туындысы деп, h_t айнымалысының θ бойынша туындысы аталады, егер $\frac{\partial\mathrm{ht-1}}{\partial\theta}=0$ деп қабылданса.

Анықтама 1 ескере отырып, $\frac{\partial\mathcal{L}}{\partial\mathrm{W}}$ градиентін келесі түрде жазуға болады:

$$\frac{\partial\mathcal{L}}{\partial W}=\sum_{t=1}^{T}\frac{\partial\mathcal{L}_t}{\partial h_t}\frac{\partial h_t}{\partial W}= \\ =\frac{\partial\mathcal{L}_T}{\partial h_T}\frac{\partial^+ h_T}{\partial W}+\frac{\partial\mathcal{L}_T}{\partial h_T}\frac{\partial h_T}{\partial h_{T-1}}\frac{\partial^+ h_{T-1}}{\partial W}+\sum_{t=1}^{T-2}\frac{\partial\mathcal{L}_T}{\partial h_T}\frac{\partial h_T}{\partial h_t}\frac{\partial^+ h_t}{\partial W}+ \\ +\frac{\partial\mathcal{L}_{T-1}}{\partial h_{T-1}}\frac{\partial^+ h_{T-1}}{\partial W}+\sum_{t=1}^{T-2}\frac{\partial\mathcal{L}_{T-1}}{\partial h_{T-1}}\frac{\partial h_{T-1}}{\partial h_t}\frac{\partial^+ h_t}{\partial W}+\cdots+\frac{\partial\mathcal{L}_t}{\partial h_t}\frac{\partial h_1}{\partial W}= \\ =\sum_{t=1}^{T}\frac{\partial\mathcal{L}_t}{\partial h_t}\sum_{k=1}^{t}\frac{\partial h_t}{\partial h_k}\frac{\partial^+ h_k}{\partial W} \tag{1.11}$$

Лездік туынды анықтамасы мен 1.11 формуласынан тағы да мынаны байқауға болады:

$$\frac{\partial^+ h_{T-1}}{\partial W}=\frac{\partial^+ h_T}{\partial W}\frac{\partial h_T}{\partial h_{T-1}}+\frac{\partial\mathcal{L}_{T-1}}{\partial h_{T-1}} \tag{1.12}$$

Жоғарыдағы формулаларды пайдалана отырып, рекурентті нейрондық желінің [64] келесі тиімді оқыту алгоритмін ұсынуға болады.

Алгоритм 1 Уақыт бойынша қатені кері үлестіру алгоритмі
InitRandomMatrices()- модель матрицаларының кездейсоқ бастапқы жүктеме функциясы
InitZeroMatrices()-нөлдік элементті матрицаларының бастапқы жүктеме функциясы
InitialState()- жасырын күйінің бастапқы жүктеме функциясы (0 немесе 0.1)
Кіріс: {u}, {t}// u кірісі мен **t** белгілер жүйелігі

$\mathbf{U}, \mathbf{W}, \mathbf{V} \leftarrow InitRandomMatrices()$
$\mathbf{gW}, \mathbf{gV}, \mathbf{gU} \leftarrow InitZeroMatrices()$
$\mathbf{h}_0, \leftarrow InitialState()$ // *инициализация*
қайталау
  для $t = 1, \ldots, N$
    $\mathbf{h}_t \leftarrow f(\mathbf{W} \cdot \mathbf{h}_{t-1} + \mathbf{U} \cdot \mathbf{u}_t)$ // түзу жүріс
    $\mathbf{h}_t \leftarrow g(\mathbf{V} \cdot \mathbf{h}_t)$
  $\mathbf{gV} \leftarrow \partial \mathcal{L}(\mathbf{y}_T, \mathbf{t}_T)/\partial \mathbf{V}$ // кері жүріс
  $g_\mathbf{h} \leftarrow \partial \mathcal{L}(\mathbf{y}_T, \mathbf{t}_T)/\partial \mathbf{h}_T$
  $\mathbf{gU} \leftarrow g_\mathbf{h} \cdot (\partial^+ \mathbf{h}_T/\partial \mathbf{U})$
  $\mathbf{gW} \leftarrow g_\mathbf{h} \cdot (\partial^+ \mathbf{h}_T/\partial \mathbf{W})$
  для $T-1, T-2, \ldots, 1$
    $g_\mathbf{h} \leftarrow g_\mathbf{h} \cdot (\partial \mathbf{h}_{t+1}/\partial \mathbf{h}_t) + \partial \mathcal{L}(\mathbf{y}_t, \mathbf{t}_t)/\partial \mathbf{h}_t$
    $\mathbf{gV} \leftarrow \mathbf{gV} + (\partial \mathcal{L}(\mathbf{y}_t, \mathbf{t}_t)/\partial \mathbf{V})$
    $\mathbf{gU} \leftarrow \mathbf{gU} + g_\mathbf{h} \cdot (\partial^+ \mathbf{h}_t/\partial \mathbf{U})$
    $\mathbf{gW} \leftarrow \mathbf{gW} + g_\mathbf{h} \cdot (\partial^+ \mathbf{h}_t/\partial \mathbf{W})$
  $U \leftarrow U + \alpha \mathbf{gU}$ // модельді жаңарту
  $U \leftarrow V + \alpha \mathbf{gV}$
  $U \leftarrow W + \alpha \mathbf{gW}$
әзірше Тоқтау шарты орындалған жоқ

Берілген алгоритм кіріс реттілігінің ұзындығы бойынша $\Theta(T)$ күрделілігіне ие [65].

Желіні жаюдың әрбір қадамында градиенттерді есептеу үшін төмендегідей формулалар келтірілген:

$$\mathbf{e}_y = \frac{\partial \mathcal{L}_t}{\partial y_t} g\prime(V \cdot h_t) \tag{1.13}$$

$$\frac{\partial \mathcal{L}}{\partial V} = \mathbf{e}_y \cdot \mathbf{h}_t^T \tag{1.14}$$

$$\mathbf{e}_h[t] = (V^T \cdot \mathbf{e}_y + W^T \cdot \mathbf{e}_h[t+1]) \cdot diag(f'(h_{t-1})) \tag{1.15}$$

$$\frac{\partial \mathcal{L}}{\partial W} = \mathbf{e}_h[t] \cdot \mathbf{h}_{t-1}^T \tag{1.16}$$

$$\frac{\partial \mathcal{L}}{\partial U} = \mathbf{e}_h[t] \cdot \mathbf{u}_t^T \tag{1.17}$$

1 алгоритмінде көрсетілгендей градиенттер әрбір қадамда жинақталады.

1.6 Рекуренттік нейрондық желілер көмегімен тілді моделдеу

Жоғарыда көрсетілген, Элман рекуренттік нейрондық желісімен шығарылатын есеп ($x^{(t)}$, $y^{(t)}$) ұзындығына тең уақыттық қатарлар үшін тұжырымдалған. Егер ($y(t)$) реттілігі ретінде, бір тактті кідіріспен,($x(t)$) реттілігін алсақ, онда Элман моделі көмегімен ($x(t)$) уақыттық қатарының болжамдау есебін шешуге болады. Бұл есепке шындығында тілдік моделдеу есебі сәйкес келеді. Сонымен қатар келесі жағдайлар қабылданады:

– ($x(t)$) мәндері $\mathcal{L}$ моделделетін тілінің V сөзтұлғаларының ақырғы жиынында анықталған .Осылайша, болжамдау есебі дискретті ықтимал кеңістікте есептеледі.

– $p(xt|x1 \ldots xt{-}1)$ келесі сөзін ізделіп отырған ықтималдықтарымен үлестіру Элман рекуренттік желісі арқылы есептеледі.

Рекуренттік нейрондық желідегі тілдік модель ,сонымен бірге, $\mathcal{C}$: V* → Rn кескіні операцияларының реттілігі және $\mathcal{F}$: Rn → V сызықтық жіктелуі ретінде ұсынылуы мүмкін, бұл жерде класстар түрінде V сөздігінің элементтері келтірілген. $\mathcal{F}$ түрлендіруі ретінде|V| класстарына жіктеудің кез келген алгоритмін пайдалануға болады. Одан артық, $\mathcal{C}$ түрлендіруі сондай-ақ өз еркінше таңдалуы мүмкін, бірақ нейрожелілік тәсілді қолданудың бірқатар техникалық ерекшеліктері бар: 1) күрделі сызықтық емес кескіндеуді моделдеу мүмкіндігі және 2) қарапайым және тиімді оқыту алгоритмдерінің бар болуы.

Рекуренттік нейрондық желілер негізінде тілді моделдеуші тәсіл, шындығында, Элман архитектурасымен 1.3 бөлімінде сипатталған, И.Бенджио тәсілін бейімдеуді білдіреді.

Осылайша, тәсілдің негізгі ерекшеліктері келесілер болып табылады:

1) Бенджио моделі сияқты, кіріс векторы ретінде жалғыз бірлік i координатасы бар δi векторы саналады.

2) Элман моделіндегідей, $t-1$ қадамында h_{t-1} жасырын қабаты t келесі қадамында желінің кірісіне келіп түседі.

3) y_t желісінің шығысы болып, yt шығыс қабатында softmax-белсендірудің нәтижесі ретінде есептелінетін, $p(wt+1|ht,wt)$ сөздерінің ықтималдықтарын үлестіру болып табылады.

[66] жұмысында жасырын қабатты белсендіру ретінде сигмоидты функция таңдалған, жылжытулар нөлге теңдестіріліп алынған. Осылайша, тілді моделдеу үшін рекуренттік нейрондық желінің шығысы келесі формулаларға сәйкес есептеледі:

$$h_t = \sigma(W \cdot h_{t-1} + U \cdot x_t) \tag{1.18}$$

$$y_t = softmax(V \cdot h_t) \tag{1.19}$$

бұл жерде W,U,V -салмақ матрицалары, $x \in$ U -t қадамындағы предикторлық реттіліктің элементі, $y \in$ T -сол t қадамындағы белгісіз үлестіру ықтимал элементтерінің үлестіруі, $h \in$ H -желінің жасырын қабаты, f және g - активтендіру функциялары.

σ және $softmax$ -сәйкесінше сигмоидты және софтмакс- белсендірулер.

Сонымен қатар, айқын түрде, қабаттағы қателер үшін теңдеулерді және әрбір қадамда BPTT алгоритміндегі градиенттерді мына түрде жазып алуға болады:

$$\mathbf{e}_y = \mathbf{t}_t - y_t \tag{1.20}$$

$$\frac{\partial \mathcal{L}}{\partial V} = \mathbf{e}_y \cdot \mathbf{h}_t^T \tag{1.21}$$

$$H_a = \begin{pmatrix} h[t]_1(1-h[t]_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & h[t]_H(1-h[t]_H) \end{pmatrix} \tag{1.22}$$

$$\mathbf{e}_h[t] = H_a \cdot (V^T \cdot \mathbf{e}_y + W^T \cdot \mathbf{e}_h[t+1]) \tag{1.23}$$

$$\frac{\partial \mathcal{L}}{\partial W} = \mathbf{e}_h[t] \cdot \mathbf{h}_{t-1}^T \tag{1.24}$$

$$\frac{\partial \mathcal{L}}{\partial U} = \mathbf{e}_h[t] \cdot \mathbf{u}_t^T \tag{1.25}$$

Сол сияқты әдісті жүзеге асыруда L2 регулярлау қолданылады, ол 1 алгоритмінде көрсетілмеген.

Т.Миколовты жүзеге асыру моделді жаңартудың түпнұсқалық кескінін білдірген:

1) Жасырын және шығыс қабат нейрондарының арасында V салмақ матрицаларын жаңарту әрбір уақыттық қадамда жасалады.

2) U сөздік кестесінің және W рекуренттік матрицасының жаңаруы әрбір k-ншы қадамда жасалады.

3) Уақыт бойынша кері қатені үлестіру $p < k$ қадамға кері қарай жасалады.

Осылайша, T ұзындығының оқыту реттілігімен өту кезінде және $k = T/2$, $p < T/4$ параметрлерімен, V матрицасы T рет, ал U және W матрицалары екі рет жаңартылады, сонымен бірге $\frac{\partial \mathcal{L}}{\partial \mathrm{W}}$ және $\frac{\partial \mathcal{L}}{\partial \mathrm{U}}$ нәтижелік градиенттер $T/4$ қадамдарындағы градиенттер қосындысын білдіреді.

Берілген тәсілді стохастикалық градиент әдісі ретінде түсіндіруге болады. Оның артықшылықтары болып, барынша тез және, ғылыми тәжірибелер көрсеткендей, тұрақты жинақтылық табылады. Сонымен қоса, бұл тәсілді қолдану екі жоғары параметрді енгізеді: k жиілігін және уақыт бойынша желіні жаюдың p тереңдігі. Бұл параметрлердің мәндері жоғарыда талқыланған градиенттер өшуі мәселесімен тікелей байланысты: өшу кезінде терең жаюдың мағынасы жоқ екені анық.

[67]-те сипатталынған ғылыми тәжірибелер 5-6 қадамда жаю тереңдігінде жақсы нәтижелерді көрсетеді.

2 КІРІККЕН ҚАЗАҚ СӨЙЛЕУІНІҢ АКУСТИКАЛЫҚ ЖӘНЕ ТІЛДІК КОРПУСТАРЫН ТАЛДАУ ЖӘНЕ ҚҰРУ

Сөйлеу корпусы – сөйлеу фрагменттерінің құрылымдалған жиынтығы, ал оған қолжетімділік бағдарламалық құралдар көмегімен қамтамасыз етіледі. Сөйлеу фрагменті ол корпустағы сөйлеу сигналының цифрланған негізгі бірлігі болып табылады.

2.1 Тілдік корпустарға шолу

Ең алғаш сөйлеу корпустары алдыңғы ғасырдың 80-ші жылдары АҚШ-та пайда бола бастады, оның құру жұмыстарын алдымен сол елдің Қорғаныс министірлігі қаржыландырып отырды. Осы ведомствоның қолдауының арқасында оқшауланған сандар мен сандық тізбектерді тану жүйелерін тестілеуге арналған TI-DIGITS корпусы (1984 ж.) және кілттік сөздерді (word spotting) талдау және тану үшін Road Rally корпусы мен сөйлеушіні тану (speaker recognition) жүйелеріне арналған King Corpus корпусы пайда болды. Лингвистикалық технологияларды дамыту бойынша мемлекеттік бағдарлама аясында ARPA/DARPA (the Advanced Research Projects Agency) және осы министрліктің қаржыландыруымен американдық атақты TIMIT (1980-1990 жж.) корпустары құрылды, және олар өз кезегінде басқа да көптеген сөйлеу корпустарының пайда болуына әсерін тигізді. Тағы да осы қаржыландырудың арқасында кіріккен сөйлеуді тану саласы бойынша зерттеулерге арналған Resourse Management (RM) [68] және Wall Street Journal (WSJ) мамандандырылған сөйлеу корпустары [69], сонымен қоса еріксіз сөздерді зерттеу мен диалогтық жүйелердегі табиғи тілді түсінуге арналған Air Travel Information Service (ATIS) корпустары құрылды [70].

Тәжірибеден көрсеткендей, жақсы сөйлеу корпусын құру күрделі технологиялық тапсырмалар мен айтарлықтай қаржылық және кадрлық жұмсауларды қажет етеді. Бұл процесстің ең негізгілері: әлі де қаржыландыру мәселелері мен кооперативтік күштер, сөйлеу корпустарының көпсалалылығы мен қолжетімділігін қамтамасыз ету, сөйлеу деректерін тексеру, жинау және өңдеуге арналған компьютерлік құралдарды құру және стандарттау. Бұл тапсырмаларды шешу үшін, XX ғ. 90-шы жылдары қолжетімді және стандартталған тілдік ресурстарды, соның ішінде сөйлеу ресурстарын, жинақтау, сақтау, тарату мен құру бойынша арнайы үйлестіруші орталықтар ашылды. Олардың ішінде негізгілері:

- LDC (Linguistic Data Consorcium) [6];
- CSLU (Center for Spoken Language Understanding, Oregon Graduate Institute http://www. CSLU.ogi.edu);
- ELRA (European Language Resources Association, http://www.elra.info).

Осы орталықтар ұсынатын сөйлеу корпустарының жинақтамасы жыл артқан сайын үлкейіп, оларды құру үшін көптеген мамандар жұмылдырылуда. Бұл уақытта корпустардың қуаты, түрлілігі мен компьютерлік жабдықталуы өсіп келеді [71].

2.1.1 Қазақ тілінің сөйлеу корпустары

Негізінен сөйлеу деректері көптілді болып келеді. Сөйлеу копрустары тек технологиялық қажет тілдерге ғана емес (американдық ағылшын, неміс, жапон, қытай және т.б.), сонымен қатар Еуропа Одағының көптеген ресми тілдерінде де құрылған: ағылшын тілінің британдық және шотландық түрлері, голландтық, даттық, шведтік, неміс, француз, италяндық, испандық, және де бірнеше көптілді сөйлеу корпустары бар. Copernicus ELRA бағдарламасын іске асыру арқасында шығыс Еуропа елдері тілдерінің (поляк, болгар, эстон, румын және венгр) корпустары тарала бастады. Және осылармен бірге орыс тілінің сөйлеу корпусын құру жұмыстары да жүргізілді.

Қазақ тілі сөйлеуінің корпусын құру жұмыстары жүргізілгенмен, олар туралы толық ақпарат жоқ. Жекелеген зерттеушілер көлемі кіші, деректері аз корпустарды құрды. Шәріпбай А. мен Есенбаев Ж. бастаған Л.Н. Гумилев атындағы Еуразиялық ұлттық университеттің бір топ зерттеушілері, 169 диктордың дауысы жазылған, 12500 аудио файлдан және оның мәтіндік транскрипциясы бар 20 сағаттық сөйлеу корпусын құрды [72].

2.2 Сөйлеу корпустарын талдау

Қазіргі таңда үлкен көлемді, әртүрлі және ақпаратты «бай» (көп деңгейлі) сөйлеу корпустарын құру, оларды компьютерлік бағдарламаларда қолдануға, іргелі фонетикалық зерттеулерге арналған ыңғайлы және сенімді құралдарды жасау мәселелерінің өзектілігі артып келеді. Сенімділіктің ең жоғары көрсеткіштеріне қол жеткізіп келе жатқан заманауи сөйлеуді тану жүйелері, негізінен, сөйлеу және тілдік құбылыстарды статистикалық модельдеу әдістеріне негізделіп, көптеген дикторлардан жазылған, дыбысталған сөйлеудің мазмұндамасымен берілген үлкен массивтерде оқытуды қажет етеді.

Тек дыбыс түріндегі ауызекі сөйлеуін талдау іс жүзінде мүмкін емес, дыбыс көпқырлы болғандықтан, адамның ойы мен көзқарасы бір уақытта оның маңызды бір үзіндісін бөліп ала алмайды. Сондықтан сөйлеуді дұрыс белгілеудің белгілі бір әдісі қажет. Сондай белгілеудің ең ықтимал және ыңғайлы әдісі – оны графикалық түрге аудару, яғни транскрипциясын құру. Тілдің сөйлеу корпусы кем дегенде екі компоненттен тұруы керек, олар сөйлеудің аудио жазбасы және осы жазбаларға сәйкес транскрипциясы.

Корпус – ол электронды түрдегі мәтіндердің (жазбаша және ауызша) жиыны. Сөйлеу корпусы – ауызша сөйлеудің аудиожазбасымен қоса транскрипциясы бар жинақтама. Сөйлеу корпусын бірнеше түрге жіктейді:

а) Оқытылған сөйлеу. Олар бірнеше түрге бөлінеді:

- сөздер тізімі;
- әдебиеттердің үзінділері;
- жаңалықтар үзіндісі;
- сандар тізбегі.

b) Еріксіз сөз. Диктордың еріксіз сөйлеген сөздері, оларды төмендегідей жіктеуге болады:

- диалогтар мен кездесудегі әңгімелер (2 немесе одан да көп адамдардың бір бірімен еркін сөйлесуі);
- ынталандырылған нарративтілік (бір адамның бір оқиғаны баяндауы);
- кездесуді белгілеу тапсырмасы (күні мен орны туралы);
- табиғи жағдайды қолдан қайталау.

c) Дайын корпустар:
- Datenbank Gesprochenes Deutsch (DGD) (неміс тілінің)
- CHILDES database, CMU Pittsburgh (ағылшын тілінің)
- CLAPI: Corpus des langues parlées en interaction (француз тілінің)
- CGN: Corpus Gesproken Nederlands (голланд тілінің)
- Multimedia Corpus of Spoken Bulgarian (болгар тілінің)
- Pražský mluvený korpus (Cesky národni korpus (CNC)) (чех тілінің).

d) Өзіндік корпус.

Сөйлеуді тану тапсырмаларының ерекшеліктерін жақсырақ түсіну үшін, заманауи жүйелердің барлығы дерлік бір-біріне ұқсас әдістер және алгоритмдермен жұмыс істейтінін атап өту керек. Тану жүйелеріндегі кездесетін ерекшеліктер сөйлеу типіне, сөздік көлеміне байланысты және сөйлеу бірліктерінің құрылымынан туындауы мүмкін. Олар жүйеге белгілі бір шектеулер қоятын тапсырмалардың түрлеріне қарай пайда болатындықтан, мұндай шектеулерге қажетті есептеу жылдамдығы, жадының қолдану көлемін жатқызуға болады.

2.3 Қазақ тілінің корпусын құру кезінде туындайтын мәселелер

Тіл адамдардың өзара байланысуы үшін қолданылатын, ойын, сезімін білдіре алатын, адамдардың бір-біріне әсер ететін құралы болып табылады. Аз ресурсты тіл түсінігін айтпас бұрын, тіл деген түсініктің өзі күрделі ұғым екенін атап өту қажет. Әлем тілдерінің күйі мен санын бағалау саласы бойынша зерттеу жұмыстарын «Ethnologue. Languages of the world» [73] жобасы айналысады. Бұл жоба бойынша, қолданыста бар тілдердің санын есептеу кезінде, тілді, жоқ дегенде әлем бойынша бір адам өзінің ана тілі деп санайтын болса ғана жеке тіл ретінде есепке енгізеді. Осы зерттеулер негізінде, соңғы деректер бойынша әлемде 7105 тілді бөліп көрсетеді, оның ішінде 1481 тілге «мәселелі», ал 906-сына «жойылушы» деген мәртебе берілген. 2.1-кестеде әлем тілдерінің біркелкі емес дамуы мен таралуы бойынша статистикасы көрсетілген.

Дыбысталатын сөйлеу корпусы, сонымен қоса, сөйлеу деректер базасы деп аталатын тілдік ресурстарының маңызды типін ұсынады. Корпус құрамына фонетикалық онымен бірге тілдік ресурстарды жинақтауды, құруды, ұйымдастыруды және басқаруды қамтамасыз ететін компьютерлік бағдарламалар кіреді. Сөйлеу корпустарын құруға деген қызығушылық сөйлеуді автоматты тану саласының өнімдерін әзірлеу бойынша бастамашылықтарынан туындады. Ол жерде зерттеушілер дыбыстық бірліктерінің акустикалық нұсқалылығына тап болады. Өйткені оларға нұсқалылық көзі әртүрлі – жүйелік мәнмәтінді нұсқалығынан бастап, сөйлеушінің психофизиологиялық күйі немесе

сөйлеу материалын жазу кезінде қолданатын микрофонның техникалық сипаттамаларына дейін әсер етуі мүмкін. Заманауи тану жүйелері, көп жағдайда, көптеген дикторлардың (100 адамнан кем емес) жазылған, дыбысталған үлкен көлемді сөйлеулер массивімен оқытылады.

Кесте 2.1 – Әлем тілдерінің сандық құрамының жіктелуі

Тілді иеленуші адамдардың саны	Тілдердің саны	Соңғы жиналған тілдердің пайызы	Берілген тілде сөйлеушілердің саны	Соңғы жинақталғандар бойынша сөйлеушілердің пайызы
100,000,000 – 999,999,999	8	0,1%	2,528,029,108	40,5%
10,000,000 – 99,999,999	77	1,2%	2,381,969,581	78,7%
1,000,000 – 9,999,999	308	5,5%	962,536,721	94,2%
10,000 – 999,999	2726	43,9%	355,780,848	99,87%
0 – 9,999	3986	100%	8,105,309	100%
Барлығы	7105	100%	6,236,421,567	100%

Бүгінгі таңда сөйлеуді тану жүйелерін құру кезінде кездесетін негізгі мәселелердің бірі акустикалық деректерді жинау және талдау болып табылады. Өйткені берілген тілді толықтай тануын оқыту үшін, бірнеше жылдарға созылатын күрделі жұмыс істеу керек, сол тілді нақты білетін адамдарды, сонымен қатар көптеген дауыстардың типтері мен акценттерін қамту керек. Ал қазақ тілі сөйлеу және мәтіндік деректері жеткіліксіз аз ресурсты тілдерге жатады.

Қазақ тілін автоматты тану саласы әлі де толықтай зерттелмеген. Сөйлеуді тану жүйелерін құруға арналған көлемі жеткілікті акустикалық және мәтіндік деректердің жеткіліксіздігі басты мәселелердің бірі болып табылады. Жоғарыда келтірілген статистикалық мәліметтерге сүйенсек, қазақ тілі аз ресурстарды тілдердің қатарына енеді. Сонымен қатар қазақ тілінің белгілі бір анықталып, ұсынылған фонемалық транскрипциясының жоқтығын яғни сөйлеуді тану жүйелері мен зерттеу жұмыстарын жүргізу мақсатында қолданылатын, сөйлеуді тану сөздігін құруға арналған қазақ тілі дыбыстарының фонемалық белгілерінің болмауын тағы бір мәселе ретінде анықтауға болады.

2.4 Сөйлеу корпусын құру кезеңдері

Сөйлеу корпусы – бұл құрылымдалған сөйлеу фрагменттерінің жиынтығы, оған қолжетімділік бағдарламалық құралдар көмегімен жүзеге асырылады. Сөйлеу фрагменті – корпустың негізгі бірлігі ретінде белгілі бір типтегі

қауымдасқан ақпаратымен ілесіп жүретін, сөйлеу сигналының сандық фрагменті болып табылады [71]. [74] бойынша, сөйлеу корпусын құру – белгілі бір технологиялық процесс. Оны келесі кезеңдерге бөліп көрсетуге болады:

- сөйлеу корпусын құру үшін фонетикалық қамтамасыздандыруды дайындау;
- әртүрлі деңгейдегі сөйлеу сигналының транскрипциясына арналған стандартты құру;
- мәтіндік материалдарды даярлау;
- сөйлеу корпусын құруға арналған компьютерлік бағдарлама құру;
- дикторлар құрамын анықтау;
- дикторлардың сөйлеген сөйлеу фрагменттерін жазу;
- сөйлеу фрагментін жазу кезіндегі сапасын тексеру;
- сөйлеу сигналдарын фонетикалық транскрипциялау және белгілеу бойынша нұсқаулық дайындау;
- сөйлеу фрагменттерін фонетикалық тексеру және оларды белгілеу;
- сөйлеу материалдардын автоматты белгілеуді және аннотациялауды тексеру;
- тексерілген нәтижелерді өңдеу.

2.4.1 Сөйлеу корпусының мәтіндерін дайындау

Сөйлеу корпусына арналған сөйлеу дыбыстарын жазу үшін, алдымен әр түрлі жанрдағы сөйлемдер дайындалды. Олар интернет беттерінде жарияланған жаңалықтар сайттарынан, қазақ тіліндегі википедия сайтынан, кейбір әдебиеттердің электрондық нұсқаларынан алынды. Таңдалып алынған мәтіндерден көнерген сөздер мен қысқартылған сөздер алынып тасталды. Және де арасында кездесетін сандар жазбаша түріне ауыстырылды. Дикторлардың дауыстап оқуы үшін, әрбір дикторға 100 сөйлемнен тұратын мәтіндерге бөлінді. Бір сөйлем 4-12 сөздерден тұрады. Сонымен қоса, фонетикалық «бай» мәтін құрылды [75, 76].

2.4.2 Дикторларды таңдау

Дикторлар ең алдымен Ақпараттық және есептеуіш технологиялар институтының қызметкерлерінен, кейіннен әл-Фараби атындағы Қазақ ұлттық университетінің студенттері, магистранттары мен докоранттарынан құрылды. Дикторлардың ауызша келісімімен олардың дауысының ғылыми-зерттеу жұмыстарын жүргізу мақсатында пайдаланылатыны, және де олар туралы ешқандай жеке ақпараттардың тарамайтындығы ескертілді. Дикторлар жасы 18 жастан 55 жасқа дейінгі, ер және әйел адамдардан тұрды. Жазылған дауыстардың диалектикалық және басқа да статистикаларын жүргізу мақсатында, дикторлар туралы жеке ақпарат жиналды. Оның үлгісі 2.1-суретте көрсетілген.

Диктор сауалнамасы

№	Тегі	Аты		Әкесінің аты	
1	*Тұрдалыұлы*	*Мұса*		-	
2	Жынысы*:	***Ер***		Әйел	
3	Ұлты:	*қазақ*			
4	Туған жері:	*Алматы облысы*			
5	Туған жылы:	*1990*			
6	Тұрғылықты жері:	*Алматы*			
7	Тел:	*+77788359999*			
8	e-mail:	*mkt_001@mail.ru*			
9	Білімі*:	Орта	Арнайы орта	Аяқталмаған жоғары	***Жоғары***
10	Академиялық дәрежесі*:	Бакалавр	Магистр		***PhD докторант***
11	Мамандығы:	*Ақпараттық технологиялар*			
12	Күні:	*13.07.2018*			
13	Қолы:				

Ескерту қажет болған жағдайда астын сызыңыз.*

Сурет 2.1 – Диктор сауалнамасының толтырылған мысалы

2.5 Кіріккен қазақ сөйлеуінің акустикалық деректері

Акустикалық деректерді жинау күрделі және ұзақ уақытты қажет ететін жұмыстардың бірі. Сөйлеу технологияларында сөйлеу корпустары акустикалық модельдерді құру мақсатында қолданылады (ол өз кезегінде сөйлеуді тану жұмыстарында қолданылуы мүмкін). Лингвистика саласында сөйлеу корпустары фонетика, диалектология, конверсиялық талдау және т.б. салаларда зерттеулер жүргізу кезінде қолданылады. Қазақ тілінің сөйлеуін жазу Ақпараттық және есептеуіш технологиялар институтында орналасқан Vocalbooth.com фирмасының шуылдан қорғалған жазу кабинасын (2.2-сурет) қолдану арқылы жүзеге асырылды [77].



Сурет 2.2 – Vocalbooth.com фирмасының шуылдан қорғалған аудио жазу кабинасы

Зерттеу жұмыстарын жүргізу кезінде сөйлеу сигналдарын акустикалық жағдайда сапалы жазу мақсатында Vocalbooth.com фирмасының шуылдан қорғалған кабинасын қазақ тілі сөйлеуінің корпусын құруда қолданылуда. Бұл аудио жазу кабинасы дыбыс өткізбейтін екі қабаттан, сондай герметикалық есіктен, екі қабатты мөлдір шыны терезеден тұрады, оның ішкі безендіруі қызық түсті, пирамида тектес акустикалық материалмен әрлендірілген және кабина мен сыртқы жағы шуылсыз ауа алмасу жүйесімен жабдықталған. Бұл кабина жоғары деңгейдегі дыбыстарды жазу үшін қолданылады.

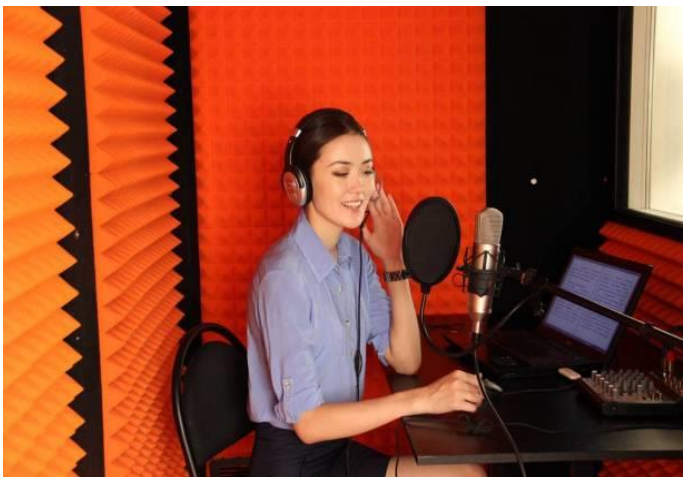
2.5.1 Жазу процессі

Аудио деректерді жазу процессі – өте баяу жүретін, көп еңбекті қажет ететін деңгей. Күніне орташа есеппен 2-4 диктордың сөйлеген сөздері компьютерге аударылып отырды. Жазу кезінде қолданылған техникалық құрылғылар келесідей:

- Windows 10 операциялық жүйесі орнатылған Lenovo G50 ноутбугы
- Микрофон
- SoundForge 11.0 аудио жазуға арналған бағдарламалық қамтамасы;
- Шуылдан қорғалған кабина

Жазу процессінің үлгісі 2.3 суретте көрсетілген. Қосалқы шулардан қорғау мақсатында және диктордың оқуына жеңіл болуы үшін, кабинаның сыртында басқарушы мәтінді сөйлем бойынша жеке жолдарға бөліп, жазу процессін басқарып отырды. Кабина ішінде диктордың алдында басқарушының компьютеріне қосылған монитор арқылы берілген мәтінді мәнерлеп оқып отырды. Еріктен тыс, жөтелу, мүдіру, есінеу және т.б жағдайларда жазу процессі тоқтатылып, кезегі келген сөйлемді жазу қайталанып отырды.

[76] жұмыста ұсынылған фонетикалық көрнекі мәтін және де қазақ тіліндегі фонемаларды барынша қамтитын 100 сөйлем оқытылды. Барлығын қоса алғанда 36 сағаттық акустикалық дерек жиналды. Олардың транскрипциясының жеке сөздері жиналып, қайталанбайтын жеке сөздерге бөлінген файл құрылды. Әрбір сөздің дыбыстарын фонемалық сипаттамасымен сипатталып, олар алфавиттік реттілікпен сұрыпталды. Қазақ тілі сөйлеуін тану жүйесін құруға арналған фонемалық транлитерациясы бар 36 сағаттық акустикалық корпустың сөздігінің фрагментінің мысалын көруге болады. Ол сөздік мағынасына қарай бір бірін қайталамайтын 29805 сөзден тұрады [78].



Сурет 2.3 – Жазу процессі

2.5.2 Аудиожазба туралы ақпарат

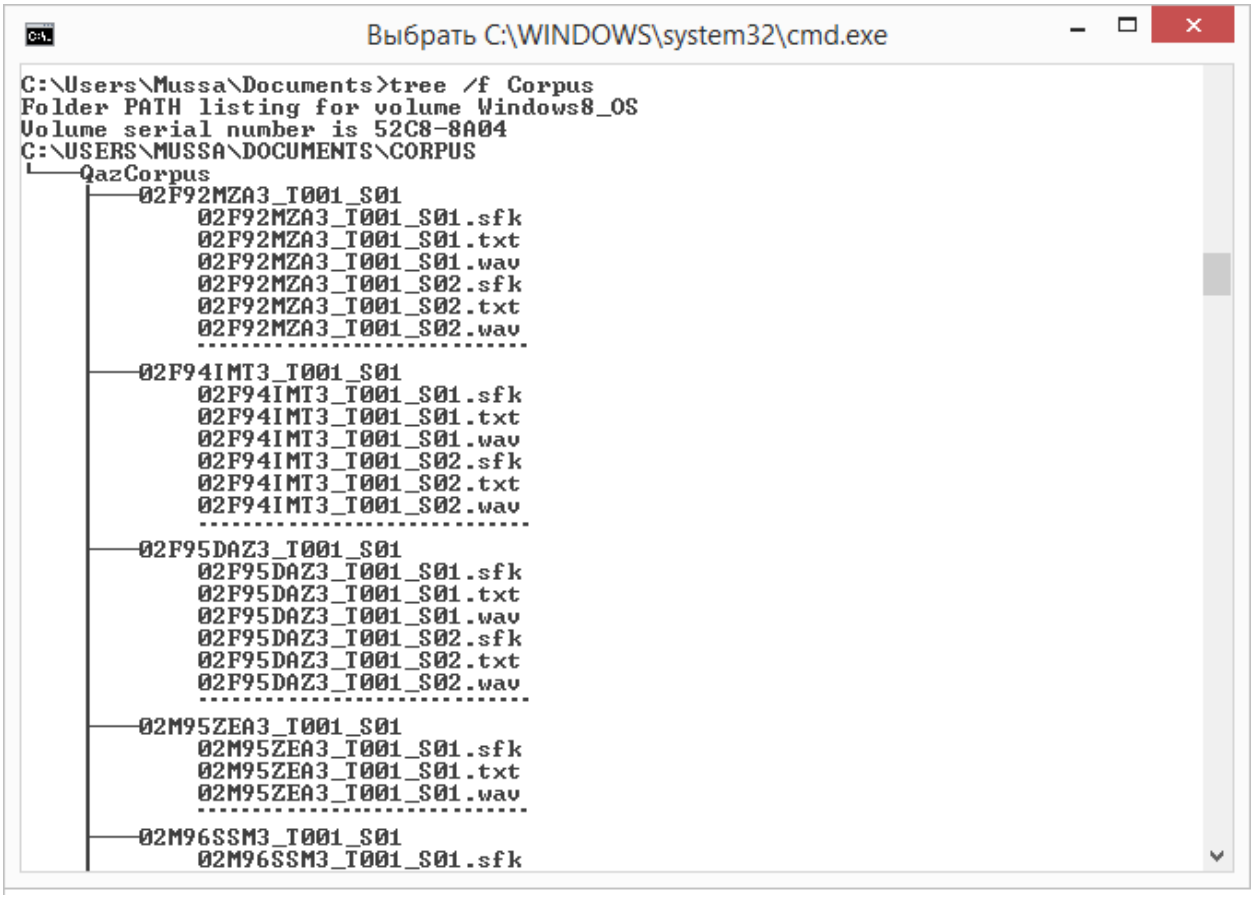
Аудиоматериалдар .wav форматында сақталып отырды. Әрбір сөйлем жеке файл ретінде сақталып, оның атын оқыған диктордың идентификациялық белгісімен бірге сөйлемнің сандық нөмірін атау келісілді. Барлық жазылған аудиоматериалдар бірдей келесі сипаттамаға ие:

- файлдың атының кеңейтілуі: *.wav;
- сигналды цифрлық түрге ауыстыру әдісі: PCM;
- дискреттеу жиілігі: 8 кГц;
- разрядтылығы: 16 бит;
- аудиоканалдардың саны: біреу (моно).

2.6 Қазақ тілі сөйлеуінің акустикалық корпусы

Құрылған қазақ тілінің сөйлеу корпусының сұлбасы 2.4 суретте көрсетілген. QazCorpus директориясында әр диктордың өзіне тән белгісімен белгіленген келесі деңгейлі директория сол диктордың жазылған аудио файлдары сәйкес транскрипциясымен бірге сақталған. Әр аудио файл бір сөйлемді қамтиды. Диктор белгісімен бірге мәтіннің нөмірі мен сөйлем нөмірі де көрсетілген. Яғни 100 сөйлемнен құралған бір мәтін.

Жоғарыда баяндалған жұмыстардың нәтижесінде, қазақ тілінің сөйлеулерін автоматты тану жүйелеріне арналған, қазақ тілінің акустикалық корпусы құрылды. Ол 36 сағаттан тұрады, және 205 диктордың (160 ер адам, 45 әйел адам) дыбысталған деректерінен тұрады (2.5 сурет).



Сурет 2.4 – Қазақ тілі корпусының сұлбасы



Сурет 2.5 – Қазақ тілі сөйелуінің акустикалық корпусын дыбыстаушы дикторлар туралы жыныстық статистикасы

2.7 Қазақ тілінің мәтіндік корпусы

Бүгінгі таңда еліміздің цифрлық жүйеге барлық салада көшуді қолға алып жатқан тұста, қазақ тілінде сөйлеу айту кезінде тануды автоматты түрде жүзеге асыру жолдары жан-жақты зерттелуде. Берілген мақалада мәтіндік тірек корпусы ретінде қазақ тілінің фонетикалық бірліктері неғұрлым көп көрініс табатын үлгісі алынған.

Алайда фонетикалық тұрғыдан толық көріністі бола алатын корпустық үлгіні дайындауда қазақ лингивистикасының фонетика саласында біршама күрделіліктер бар екенін айту керек.

Қазіргі уақытта қаза тілінің дыбыс жүйесі бірнеше теориялық бағыттың принциптері бойынша зерттеліп келеді.

Бірінші бағыт - дәстүрлі фонологиялық теория негізінде қалыптасқан. (І. Кеңесбаев, Ж. Аралбаев т.б). Дәстүрлі теорияны ұстанатын ғалымдар қазақ тіліндегі фонема – сөз мағанасын ажыраттатын шағын функционалдық бірлік деп тұжырымдайды.

Екінші бағыт – сингармониялық теория бойынша, фонема қазақ тіліне тән емес, оның қызметін сингема деген просодикалық бірлік атқарады.

Үшінші бағыт – сингармофонология деп аталады. (М.Жүсіпов т.б). бұл теория – сингармонизм фонологиясының ұғымдары мен дәстүрлі бағыттың (Мәскеулік фонология мектебі үлгісінің) ұғымдарының синтездендірілуі арқылы жасалған теория.

Сондықтан қазақ тілінің сөйлеу сигналдарының және дыбыстық сигналдарын бейімдеу мен талдаудың әдістемелік және алгоритмдік негіздерін анықтауда едәуір күрделіліктер туындайды. Ал академиялық сипатта танылған дәстүрлі теорияны басшылыққа алу фонетикалық таңдалынған мәтіндер қорын құрып дамытуға сөйлеу корпустарын қалыптастыруда қазақ тілінің өзіндік ерекшеліктері ескерілген жағдайдағы бастапқы негіз бола алады деп есептейміз.

Осы тұрғыдан қазақ фонетикасының дәстүрлі теориясын фонетикалық таңдалған мәтінді талдауда басшылыққа алғанымызды ескертеміз.

Бұл теория бойынша айтсақ, фонема дегеніміз – сөздің мағанасын ажырататын ең кіші бірлік. Қазақ тіліндегі дауысты фонемалар: а, ә, о, ө, е, ы, і, ү, ұ. Дауыссыз фонемалар: б, г, ғ, д, ж, з, й, к, қ, л, м, н, ң, п, р, с, т, х, ш, (у), (и). Фонемада вариант және вариация дейтін құбылыс байқалады [79]. Мысалы, кітап деп жазамыз, кітәп деп айтамыз. Осы жерде ә фонмасы вариант болып тұр. Хатшы деген сөзді хатчы деп айтамыз, бұл жерде таза ч айтылып тұрған жоқ, соған ұқсас дыбыс айтылып тұр. Сондықтан оны вариация дейміз.

Қазіргі кезде қазақ тілінің сөздік құрамындағы орыс сөздерінің, сондай-ақ халықаралық терминдерінің үлесі тым қомақты. Оларды орыс тіліндегідей және халықаралық терминдердің айтылу ерекшеліктеріндей сақтап айтып жазуға мәжбүрміз. Олай болса, фонологиялық талдауда дауысты дыбыстар қатарында «и» және «у» дыбыстарын, сондай-ақ «э», дыбыс фонема ретінде қолданамыз. Сондай-ақ орыс тілінен енген сөздерге байланысты «в, ф, ц, ч» фонемаларын ескермеу мүмкін емес. Фонема ретінде қаралмаса да, фонологиялық талдауда

мәтіндегі қалпына қарай қосарланып келетін «ю, я, ё» әріптерінің дыбысталуы да ескерілмей қалмайды.

Қазақ тілі екпінге емес үндесін мен үйлесімге яғни, әуезге негізделеді сондықтан Ә.Жүніспековтың «қазақ фонетикалық» атты зерттеулерінде «қазақ тілінде үнді европалық – тілдерге тән болып келетін екпін арқылы анықталатын аллофондар жоқ» деген тұжырымды қабылдаймыз [80].

Фонемалары сипаттауда негізге алатынымыз – айрым белгілері. Бір фонеманың екінші фонемаға ұқсамайтын өзгеше сипаты айрым белгілері деп аталады. Ол тілдегі дыбыстардың акустикалық жасалу орны және артикулияциялық (жасалу тәсілі) сипаттамасымен байланысты. Әр фонема инвариант (негізгі) ретінде оның айырым белгілерінің жиыны болып табылады. Айырым белгілерінің жиыны фонемалардың фонологиалық мазмұнын құрайды [81]. Қазақ тіліндегі дауысты дыбыстардың айырым белгілерін акустика – артикуляциялық ерекшеліктері ішінде талдау үшін практикада дауыстылардың тек жуан және жіңішке айырым белгілерін ғана қолданамыз. Өйткені <а> - <ә>, <о> - <ө>, <ұ> - <ү>, <ы> - <і> дауыстыларының жуан – жіңішкелілігі олардың үндесімдік қызметтеріндегі негізгі айырым белгісі болып табылады және мағына айқындау қызметін атқарады.

Сөз ағымында әр түрлі дыбыстар айтылады, олардың айтылу түрі көп, шегі жоқ, сондықтан олардың бәрін фонологиялық талдауда негізге алып қамту мүмкін емес. Біз тек солрдың өзара орайласып құрап тұрған негізгі инварианттық бір ғана типі яғни фонеманы негізге алдық.

Өз кезегінде дауыссыз дыбыстар үш типке топтастырылып қарастырылады:

1) ұяң – б, ғ, д, ж, з;

2) қатаң – қ, к, п, с, т, х, ч, ф, ш;

3) үнді – й, л, м, н, ң, р, у.

Дауысты, дауыссыз дыбыстардың жоғарыда көрсетілген жіктеулері қазақ фонетиксының дәстүрлі теориясы бойынша зерттеу жұмысының бастапқы кезеңі үшін алынып отыр. Өйткені қазақ сөйлеу тілі жүйесін лингивистикалық электорндық технологияға бейіндеп дайындау және бағалау үшін фонетикалық таңдалған мәтіндерді, сондай-ақ қазақ сөйлеу тілінің тұңғыш корпустарын қалыптастыру зәрулігі уақыт күттірмейді.

Алайда фонетикалық таңдалған мәтіндегі фонологиялық талдаудың негізіне алынған бұл теорияны пайдалану соңғы парадитматикалық жол емес екеніне көңіл аударамыз.

Зерттеу мен қолданбалы негіз жасауда одан әрі ізденістердің барысында келешекте сингармониялық теорияның және сингармофонология теориясының ұстамдарын қолану арқылы фонетикалық толықтықта қамтамасыз ету қарастырылатын болады деп сенеміз.

Фонетикалық көрінісі бар үлгіні жалпы тілдегі тірек корпустарының жиынын сараптағандағы фонетикалық бірліктер көрінісіне пайыздық сәйкестігіне қарай анықтайды. Фонетикалық көріністілікті фонетикалық тепе-теңдік деп те қарау мүмкіндігі ғылымда ескерілген. Сөйтіп фонетикалық көріністілігі бар корпусты айтылым кезіндегі сөйлеу фонетикалық сипаттау үшін

қолданады. Оларды фонетикалық ақпараттық толықтығы мен жинақтылығы пайдалануға қолайлылық тұрғызады. Сондай мәтіндік үлгінің бірі ретінде ғылыми танымдық мәтіні алынып фонетикалық сарапталынады.

Мәтінде барлығы 14599 дыбыс, дауысты 6712 дыбыс, ал дауыссыз 7887 дыбыс көрініс береді. Дауыстылардың жуаны 818, жіңішкесі 749, ал дауыссыздардың ішінде ұяңы 495, қатаңы 693, үндісі 1026. Ол төменгі 2.2-кестеде берліген.

Фонетикалық мәтіндегі фонеманың негізгі статистикалық анықтамалық корпусқа қатысты бірлік түрі 99,8% құрайды [76].

Кесте 2.2 – Дыбыстардың статистикасы

Дауысты фонема	Саны	%	Дауыссыз фонема	Саны	%	Дифтонг	Саны	%
а	409	10,40%	б	101	2,57%	и	117	2,98%
ә	54	1,37%	г	48	1,22%	у	31	0,79%
е	347	8,83%	ғ	64	1,63%			
о	101	2,57%	д	164	4,17%			
ө	64	1,63%	ж	63	1,60%			
ы	269	6,84%	з	55	1,40%			
і	239	6,08%	й	33	0,84%			
ұ	39	0,99%	к	120	3,05%			
ү	45	1,14%	қ	101	2,57%			
и	117	2,98%	л	219	5,57%			
у	31	0,79%	м	127	3,23%			
Жуан			н	283	7,20%			
а	409	10,40%	ң	43	1,09%			
о	101	2,57%	п	63	1,60%			
ы	269	6,84%	р	293	7,45%			
ұ	39	0,99%	с	101	2,57%			
Жіңішке			т	229	5,83%			
ә	54	1,37%	ф	16	0,41%			
е	347	8,83%	х	21	0,53%			
ө	64	1,63%	ч	1	0,03%			
і	239	6,08%	ш	41	1,04%			
ү	45	1,14%	у	28	0,71%			
Ашық			э	2	0,05%			
а	409	10,40%	Қатаң					
ә	54	1,37%	к	120	3,05%			
е	347	8,83%	қ	101	2,57%			
о	101	2,57%	п	63	1,60%			
ө	64	1,63%	с	101	2,57%			
Қысаң			т	229	5,83%			
ы	269	6,84%	ф	16	0,41%			
і	239	6,08%	х	21	0,53%			

2.2. - кестенің жалғас

ұ	39	0,99%	ч	1	0,03%			
ү	45	1,14%	ш	41	1,04%			
и	117	2,98%	Ұяң					
у	31	0,79%	б	101	2,57%			
Езулік			г	48	1,22%			
а	409	10,40%	г	64	1,63%			
ә	54	1,37%	д	164	4,17%			
е	347	8,83%	ж	63	1,60%			
ы	269	6,84%	з	55	1,40%			
і	239	6,08%	Үнді – мұрын жолы					
и	117	2,98%	м	127	3,23%			
Еріндік			н	283	7,20%			
о	101	2,57%	ц	43	1,09%			
ө	64	1,63%	Үнді – ауыз жолы					
ұ	39	0,99%	й	33	0,84%			
ү	45	1,14%	л	219	5,57%			
у	31	0,79%	р	293	7,45%			
			у	28	0,71%			
			Шұғыл					
			б	101	2,57%			
			г	48	1,22%			
			д	164	4,17%			
			к	120	3,05%			
			қ	101	2,57%			
			п	63	1,60%			
			т	229	5,83%			
			ч	1	0,03%			
			Ызық					
			ғ	64	1,63%			
			ж	63	1,60%			
			з	55	1,40%			
			с	101	2,57%			
			ф	16	0,41%			
			х	21	0,53%			
			ш	41	1,04%			
			Ерін					
			б	101	2,57%			
			м	127	3,23%			
			п	63	1,60%			
			у	28	0,71%			
			Тіс пен ерін					
			ф	16	0,41%			

2.2. - кестенің жалғасы

			Тіс					
			д	164	4.17%			
			ж	63	1,60%			
			з	55	1,40%			
			л	219	5,57%			
			с	101	2,57%			
			Тіл алды					
			ж	63	1,60%			
			й	33	0,84%			
			л	219	5,57%			
			н	283	7,20%			
			р	293	7,45%			
			ч	1	0,03%			
			ш	41	1,04%			
			Тіл ортасы					
			г	48	1,22%			
			к	120	3,05%			
			Тіл арты					
			ғ	64	1,63%			
			к	101	2,57%			
			н	43	1,09%			
			х	21	0,53%			
			Көмей					
			h	2	0,05%			

Фонетикалық түрде ұсынылған мәтін төменгідей түрге ие.

«Азия құрлығы Жер шарының өте үлкен бөлігін алып жатыр. Азияда тұратын халықтардың мәдениеті, мемлекеттік жүйесі, тарихы мен қоршаған ортасы біркелкі емес. Әртүрлі ұлттардың салт-дәстүрлері де бір-бірінен қатты ерекшеленеді. Жапония мемелекеті – Азия құрлығындағы экономикасы барынша дамыған ел. Ол жалпы ішкі өнім көрсеткіші бойынша дүние жүзіндегі екінші орын алады. Азияның мәдениеті – көптеген ұлттың, қоғамның, дін мен этникалық топ мұраларының көркем жиынтығы. Азия мәдениетінің маңызды бөлігі – көркемөнері, музыкасы, асханасы мен әдебиеті. Мұнда буддизм, индуизм, ислам және христиан діндері дамыған, түрлі өркениеттер қалыптасқан. Азияның дамуындағы басты факторлардың бірі – білім. Азия білімінің тарихына үш философиялық діни дәстүр әсер еткен. Олар индуизм (оның ішінде буддизм), ислам және Конфуций (оның ішінде неоконфуций ілімі) ілімі. Уақыт өте келе осы дәстүрлер бір-біріне де әсер еткен. Азияның көптеген елдерінде сауаттылық жоғары дәрежеде. Олар өзге аймақ тұрғындарының алдына шығуға дайын, инфрақұрылым, білім тұрғысынан болашағы зор. Білім тарихына көз жүгіртсек, азиялық мемлекеттерді әлемдегі алғаш білім бастаулары қалыптасқан елдер қатарынан көре аламыз. Мысалы, Қытайда, Шығыс Чжоу кезеңінде білім беру қытайлық классикалық мәтіндермен қалыптаса бастады. Жапонияда білімнің

пайда болуы дінмен тығыз байланысты. Бұрынғы кезде Жапонияда білім беру ғибадатханаларда ғана жүргізілген, дін қызметкері болғысы келген жастар балаларды белгілі бір жасқа келгенде, «теракойя» деп аталатын жерлерге оқуға шақыратын болған.

Азиядағы ең көне көркемөнердің бірі – Үнді көркемөнері. Үнді көркемөнерінің үлкен бір бөлігін мәдениет тарихы, діндері, көркемөнер туындылары мен әдеби шығармашылығы құрайды. Үнділік тау көркемөнері – өнердің ең ежелгі және ерекше түрі. Көптеген әлеуметтік және діни, аймақтық ерекшеліктерге, әртүрлі мақсаттарға сәйкес өнердің бұл түрі тоқтаусыз дамыды. Конаркадағы Күн ғибадатханасы – үнділік тау өнерінің басты үлгісі. Мыңдаған жылдар бойы Азия тарихы ірі тұлғалармен толысып отырды. Қаһарлы қолбасшы, кемеңгер философтар, саяси қолбасшылар мен бүгінгі күннің гуманистері, әлемді кезген жиһанкездер атақты тұлғалар ретінде тарихта қалды. Батыс Азия халықтары үш тілдік топқа бөлінеді, бұл бөліну мәдени тұрғыдан бөлінуге де әсер етті.

Азия құрлығы бір-бірімен тығыз байланысқан әртүрлі аймақтардан тұрады. Олар бір-бірінен климаттық өзгешеліктерімен, жер бедерлерінің әртүрлілігімен, өсімдіктер және жануарлар әлемімен ерекшеленеді. Азияда өмір сүретін халықтардың бір-бірінен үлкен айырмашылықтары бар. Әлемдегі нәсілдердің барлық түрі осы құрлықта тоғысқан.

Еуропа Еуразия құрлығының батыс бөлігін алып жатыр. Ол көлемі жағынан дүниежүзінде екінші орында тұр. Еуропаның климаты субтропиктіктен полярлыққа дейін өзгеріп отырады. Еуропада 50 мемлекет орналасқан. Жері мен халқының саны жағынан ең үлкені – Ресей, ал ең кішісі Ватикан – қала мемлекеті. Халқының тығыздығы жөнінен Азия мен Африкадан кейінгі үшінші орында. Халқының саны – 2010 жылғы мәлімет бойынша шамамен 826 миллион. Әлемдік дамуында Еуропа құрлығы маңызды рөл атқарған. Еуропалық көркемөнер, мәдениет, философия және құқықтың құрлықтағы дінге көп әсер етіп отырғанын тарих беттерінен көре аламыз. Еуропалықтардың көбі христиан дінін ұстанады. Еуропа мәдениеті әралуан. Әйгілі мәдени орталықтары – Амстердам, Лондон, Париж, Барселона, Венеция, Флоренция, Рим, Стокгольм, Прага, Берлин т.б. Еуропа Азияның батысындағы Орал таулары етегінен басталып, Атлант мұхитының жағасына дейін созылып жатыр. Оның географиясы күрделі. Бұл құрлықта өзендер, таулар, аралдар, жазықтар және үстірттер бірімен-бірі жалғасып, астасып жатады.

Мәдени тұрғыда бұл құрлық Еуропа және Скандинавия деп аталатын екі бөлікке бөлінеді. Италия, Либерия және Балқан аймағы – түбектер. Құрлықты Африкадан Жерорта теңізі бөліп тұр. Еуропаның оңтүстік аймағы таумен қоршалған. Альпі, Карпат және Пириней таулары қыраттардан, үстірт пен өзен бассейндерінен өтеді.

Еуропа әртүрлі мәдениет өкілдері тығыз қоныстанған. Еуропаның білім беру жүйесі өте жақсы дамыған және сан салалы. Қалаларында сапалы білім беретін университеттер көп. Көптеген университеттерінің мыңжылдық тарихы бар. Қазіргі таңда Еуропа үкіметі және жоғары оқу орындары Еуропалық одақпен

және басқа да қызығушылық танытқан жақтармен бірігіп, қарқынды дамуда. Олар жоғары білім мен ғылыми ынтымақтастық орнатуды көздеп отыр.

Еуропа құрлығы әлемге даңқы шыққан көрікті жерлерге бай. Шұрайлы жазықтар, табиғаты, физикалық сипаты көркем жерлердің бәрі осы аймақтан табылады. Бұл жерлер дүниежүзінде белгілі ескерткіштер мен көне тарихи жәдігерлерге тұнып тұр. Еуропада ерекше орындар көп. Көптеген туристер Еуропаға Кипр, Франция, Италия, Греция, Барселона, Дублин, Швейцария, Польша және т.б. жерлерді көру үшін келеді».

Қазақ тілінің сөйлеуін автоматты тану жүйелеріне арналған мәтіндік корпус құрылды. Ол қазақ тіліндегі көп қолданылатын сөздерді қамтиды. Мәтіндерді қазақ тіліндегі газеттердің электронды сайттарындағы мақалаларынан (мысалы, Егемен Қазақстан (https://egemen.kz/), Жас Алаш (https://zhasalash.kz/) және т.б.), әдеби жанрдағы әңгімелер, романдар және т.б. жиналды. Нәтижесінде қайталанбайтын 29805 сөзден құралған мәтіндік корпус құрылды.

2.8 Қазақ тілі сөйлеуін тану жүйесіне арналған сөздікті құру

Сөйлеуді тану жүйелерінде акустикалық модельдерді құру кезінде, тілдің ең кіші бірлігі – фонема қолданылды. Акустикалық модель шығысында тілдік бүтін сөздерін құратын фонемалар тізбегін ұсынады. Ол үшін сөйлеуді тану жүйелеріне арналған сөздікті құру қажет. Сөздікті, табиғи тілдің сөздеріне олардың фонемаларын сәйкестендіретін, яғни олардың транскрипциясын сипаттайтын сарапшы көмегімен құрылады [77]. Оны құру процесстері келесідей ерекшеліктерден тұрады:

– сөздік, оқыту жиынында қолданылатын барлық сөздерді қамтуы керек. бұл талап орындалмай акустикалық модельді оқыту процессін жүзеге асыру мүмкін емес;

– сөздік бір сөздің бірнеше транскрипциясын қамтуы мүмкін (бұл ерекшелік, кейбір тілдерде бір сөздің дыбысталуы әр түрлі болған кезде пайдалы);

– сөздік модельді оқыту үшін бір бөлек, тану үшін бір бөлек құрылуы мүмкін, өйткені тану кезіндегі сөздер саны көп мөлшерде болуы мүмкін.

Сөйлеуді тану жүйесіне арналған қазақ тілінің сөздігін жасақтайтын бағдарлама құрылды. Ол қазақ тіліндегі сөздердің фонемалық белгілеріне ауыстыруын автоматты түрде орындайды. Сөздікті құру алгоритмі төмендегідей сипатталады:

1) Алдымен қазақ тілінің дыбыстарының дыбысталуына қарай бірнеше жиындарға бөліп аламыз: бірінші дауыссыз дыбыстар жиыны, екінші дауыссыз дыбыстар жиыны және дауысты дыбыстар жиыны.

2) Кіріс мәндері бар файлды таңдау.

3) Дауыссыз дыбыстардың фонемалық белгілерін анықтау. Алдын ала анықталған жиын бойынша кіріс мәнінің бар болуын анықтау.

4) Бар болған жағдайда мәннің қай жиынға тиесілі екенін анықтау.

5) Егер таңдалған дыбыс бірінші дауыссыз дыбыстар жиынынан болып, келесі дыбыс дауысты болса, онда екеуінің арасына j латын әріпін қосамыз, егер керісінше болса дауыссыз дыбыстың латын әріпіндегі белгісін енгіземіз.

6) Егер таңдалған дыбыс екінші дауыссыз дыбыстар жиынына жататын болса, онда сол дыбыстардың латын әріптерінен құралған белгісін қоямыз.

7) Дауысты дыбыстардың фонемалық белгілерін анықтау. Егер таңдалған дыбыс дауысты дыбыстар жиынына және «яюеё» әріптер жиынына кіретін болса, онда ол дыбыстың латын әріпімен белгісін қойып, соңына латынның «j» белгісін қоямыз.

8) Егер таңдалған дыбыс дауысты дыбыстар жиынына жататын болса, онда оның латын әріпіндегі белгісін қоямыз, кері жағдайда жоғарыда анықталған дауыссыз дыбыстың фонемалық белгісін қоямыз.

9) Кіріс мәндердің барлығы аударылып болғаннан кейін оны жаңа файлға сақтаймыз.

Жоғарыда келтірілген алгоритм бойынша құрылған қосымшаның нәтижесі, қазақ тілін автоматты тану жүйелеріне арналған сөздіктің фрагментін 2.6 суреттен көруге болады [78].

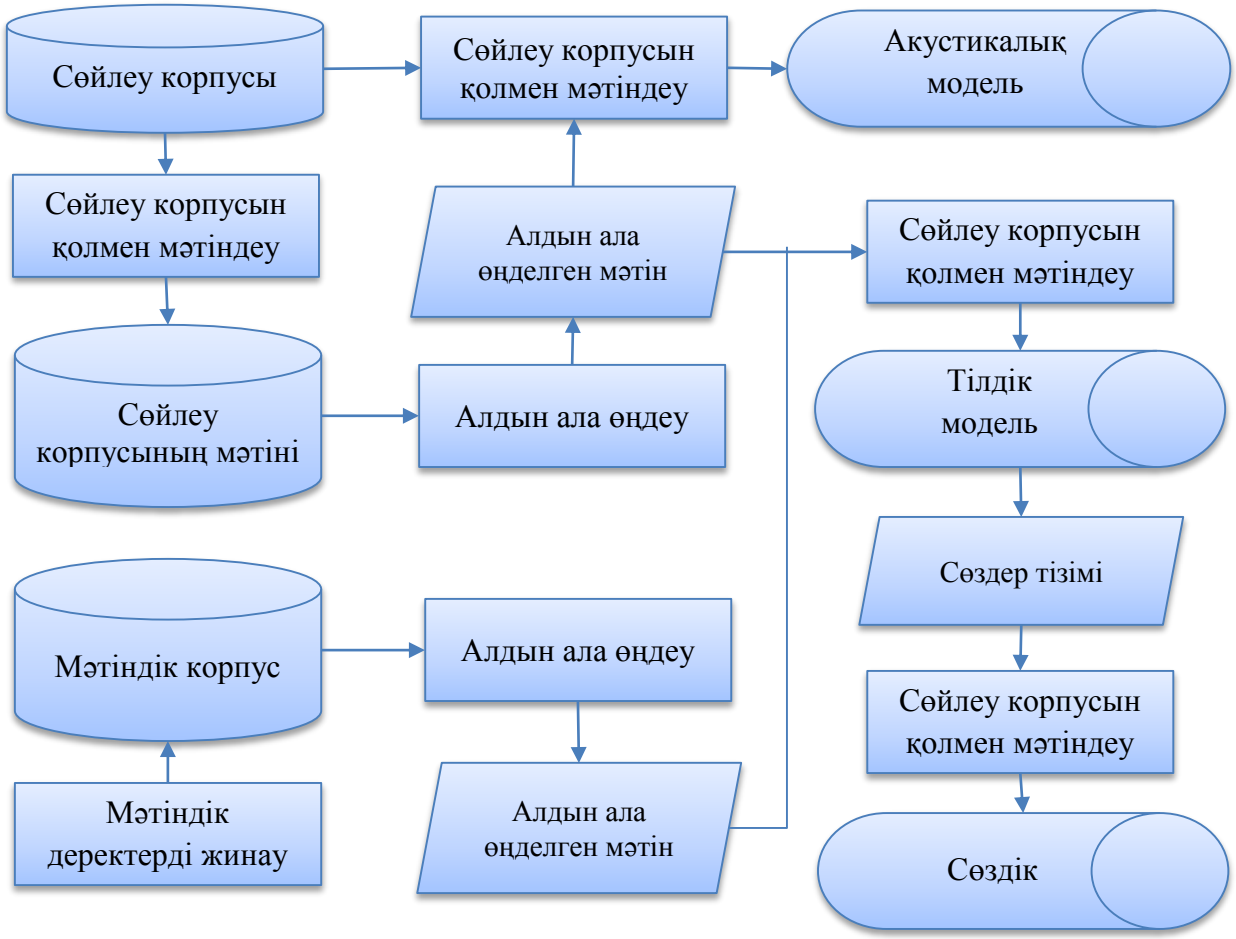
```
адасып   a d a s y p
адвокат   a d v o k a t
адвокаттар   a d v o k a t t a r
адзурраның   a d z u r r a n y ng
адиабаталық   a dj i a b a t a l y kh
админ   a d mj i n
адольф   a d o lj f
адресатқа   a d rj e s a t kh a
адресатынан   a d rj e s a t y n a n
адресін   a d rj e sj i n
адрестерін   a d rj e s tj e rj i n
адрестік   a d rj e s tj i k
адым   a d y m
адымын   a d y m y n
адыр   a d y r
адырға   a d y r gh a
адырдан-адырды   a d y r d a n a d y r d y
ажал   a zh a l
ажалдан   a zh a l d a n
ажалымды   a zh a l y m d y
ажалымнан   a zh a l y m n a n
ажалына   a zh a l y n a
ажар   a zh a r
ажарланды   a zh a r l a n d y
```

Сурет 2.6 – Қазақ тілі сөйлеуін тану жүйесіне арналған сөздік фрагменті

3 ҚАЗАҚ ТІЛІНІҢ АКУСТИКАЛЫҚ ЖӘНЕ ТІЛДІК МОДЕЛЬДЕРІН ҚҰРУ ЖӘНЕ ТӘЖІРИБЕЛЕР ЖҮРГІЗУ

3.1 Акустикалық және тілдік модельдер

Акустикалық және тілдік модельдерді құру архитектурасы 3.1 суретте көрсетілген.



Сурет 3.1 – Акустикалық және тілдік модельдерді құру архитектурасы

3.1.1 Акустикалық және тілдік модельдерін құру құралы

Акустикалық және тілдік модельдер Kaldi құралына арналып, сол құралдың модульдік қосымшалары көмегімен құрылды. Алдыңғы тарауда баяндалғандай, қазақ тілінің сөйлеу корпусын осы құралда қолдану үшін, деректер икемделді. Яғни құрылған корпус деректерінің көмегімен акустикалық және тілдік модельдерді құру жұмыстары сипатталады.

Kaldi құралы төмендегі бірнеше ерекшеліктерінің арқасында зерттеушілер арасында ең танымал құрал болып табылады [82]:

1) сөйлеуді тану бойынша заманауи әдістер мен алгоритмдердің көптеген түрлерін қолдайды;
2) зерттеушілерге өздерінің әдістері мен алгоритмдерін жүзеге асыруға мүмкіндік береді;

3) басқа құралдармен салыстырғанда, сөйлеуді тану бойынша өте жоғары көрсеткіштерге ие;

4) әр түрлі сөйлеуді тану тапсырмаларын шешуге және ұтымды жүйе құруға арналған дайын нұсқаулары бар.

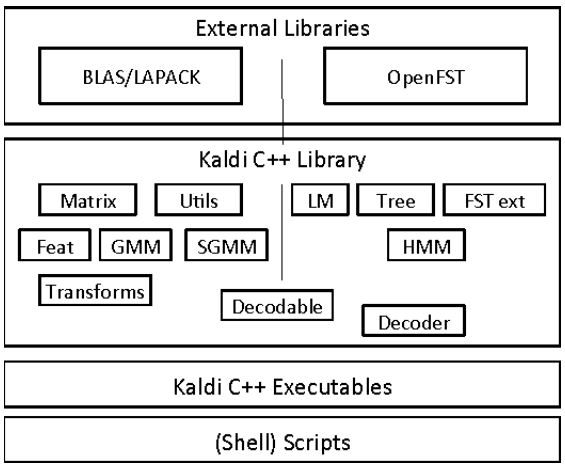
Акуситикалық және тілдік модельдерді құрудың толық нұсқасын 3.1-суреттен көруге болады.

Kaldi – C++ бағдарламалау тілінде жазылған, Apache v2.0 лицензиясы аясында лицензияланған, ашық кодты сөйлеуді тану жұмыстарын жүргізуге арналған танымал құрал. Kaldi, түсінуге, өзгертуге және өсіруге жеңіл, заманауи және икемді кодты болуды мақсат етеді. [83].

Сөйлеуді тану жүйелерін құруға арналған ашық кодты потенциалды құралдарды таңдау мүмкіндіктері көп. Дегенмен Kaldi-дың ерекшелігі оның сызықтық алгебраны қолдауы мен шексіз лицензиясы Kaldi-дің дамуына әкелді. Kaldi-дің маңызды ерекшеліктері мыналар:

- шекті түрлендіргештермен бірігуі;
- сызықтық алгебраларды кеңінен қолдауы;
- кеңейтілетін дизайн;
- ашық лицензия;
- толық анықтамасы;
- ұқыпты тестілеу.

3.2-суретте Kaldi құралының жалпы сұлбалық түрі көрсетілген. Құрал екі түрлі сыртқы кітапханаға бағынышты, біреуі шекті құрылымға арналған OpenFst екіншісі сандық алгебра кітапханасы.



Сурет 3.2 – Kaldi құралының сұлбасы

Кітапханалық қызметтерге қолжетімділік командалық жол құралы көмегімен жүзеге асырылады, олар C++ бағдарламалау тілінде орындалған және сценарийлер тілінің командалары арқылы сөйлеуді танушыны құруға және жұмысын бастауға мүмкіндік береді. Әрбір құрал өзіне тән ерекше қызметтерге ие, мысалы, жекелей орындалатын файлдар статистиканы жинауға,

аккумуляторларды қосындылауға және ақиқатқа максималды жақындауды бағалауға арналған GMM негізіндегі акустикалық модельді жаңартуға арналып өзіне тән ерекше қызметтерді атқарады.

Kaldi-дің фунцияларды алуы: оның белгілерді бөліп алуы мен сигналдар формасын оқитын кодтары MFCC және PLP стандартты функцияларын құруға арналған [84].

Алдымен қазақ тілінің деректер жиынын қолданып, Kaldi ортасына арналған деректер каталогын құру қажет. Содан кейін корпустағы сөйлеу сигналдарының белгілерін бөліп алуды орындап, кейіннен алынған белгілердің көмегімен сөйлеуді автоматты тану жүйелерін оқытамыз.

Kaldi құралының каталогының құрылымы төмендегідей сипатталады:

– steps автоматты сөйлеуді тану жүйесін құруға арналған скрипт;

– utils Kaldi-дың файлдарын өзгертуге арналған скрипттер жиынынан тұрады. Мысалы: ішкі жиын деректер каталогын одан да кіші бөлімдерге бөлу.

– local бұл каталог әдетте өзгертулер мен толықтырулардың көп орындалатын және тек корпусқа қатысты файлдарды қамтиды;

– data барлық деректерге қатысты ақпаратты сақтайды: train, test және dev каталогтары. Бұл каталог корпус деректерінен, құрылады;

– exp модельдер мен тәжірибелерді және де лог журналдарды қамтиды;

– conf белгілі бір скрипттарға арналған конфигурациялар қамтиды;

– path.sh Kaldi-дың бастапқы каталогының жолдарын қамтиды;

– run.sh жұмысты бастауға арналған негізгі файлдардың бірі, жұмыс жүргізу барысында толықтырылып, өзгертіліп отырады.

3.2 Тәжірибелерді жүргізу ортасы

Kaldi-ді орнатуға арналған ортаның минималды керекті сипаттамалары бойынша оны Unix тектес операциялық жүйесі бар кез келген есептеу машинасында қолдануға болады. Бірақ тапсырмалардың көлемі мен санына қарай, машинаның есептеу күші жетпеуі мүмкін. Сондықтан деректерді құралға ыңғайлау жұмыстары жеңіл болуы үшін алдымен Lenovo G50 ноутбугін, Ubuntu операциялық жүйесінің 16.04. нұсқасы қолданылды. Корпус деректерін өңдеу Python 2.7 бағдарламалау ортасында жүргізілді.

Сөйлеу сигналдарын өңдеп, сөйлеу белгілерін бөліп алу, акустикалық және тілдік модельдерді құру жұмыстары келесідей сипаттамалары бар серверлік машинада орындалды (сурет 3.3):

– ОЖ: Ubuntu Server 16.04;

– Процессоры: Intel(R) Xeon(R) CPU E5-2640 v4 @ 2.40GHz;

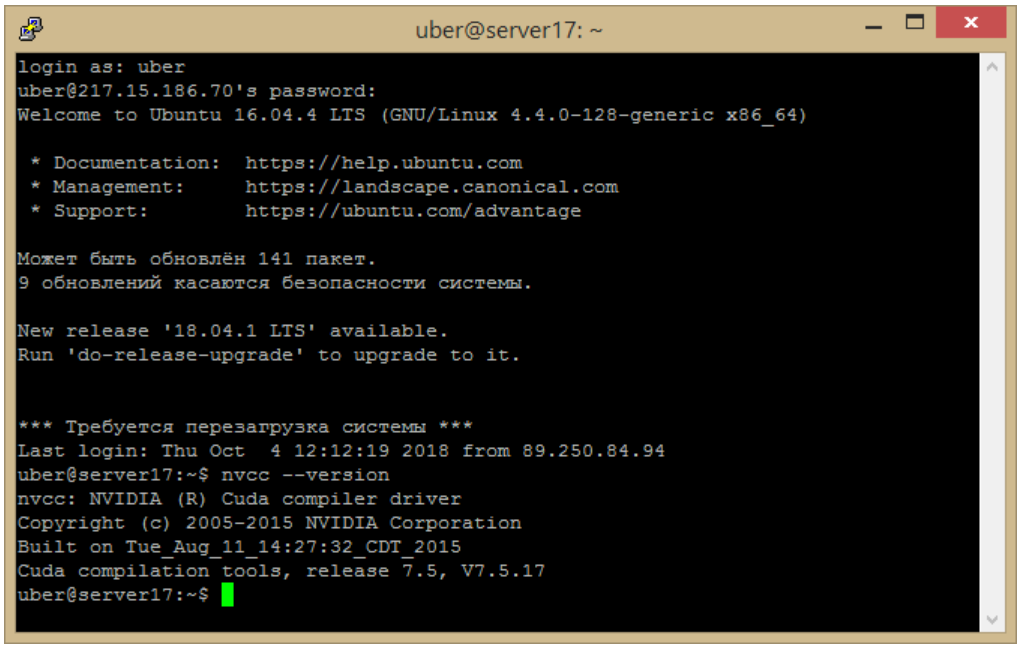
– Бейнелік процессоры: NVIDIA Corporation GM204GL [Quadro M4000] (rev a1) (сурет 3.4);

– Жадысы: 62,8 ГБ.

Акустикалық және тілдік модельдерді құру мен алынған модельдерді тесттік тәжірибелерді жүргізу кезінде парралельді есептеу технологиясы қолданылды. CUDA технологиясының 7.5 нұсқасын пайдаланып, бірқатар нәтижелер алынды.



Сурет 3.3 – CUDA технологиясын қолдайтын NVIDIA-ның Quadro M4000 моделді бейнелік процессоры



Сурет 3.4 – Тәжірибелерді жүргізу ортасы және CUDA нұсқасы

3.3 Жүргізілген тәжірибелердің сапасын бағалау

Сөйлеуді автоматты тану жүйелеріннің сапасын бағалау әдістерінің бірнеше түрі бар. Сөздердің қателік жиілігі (WER) сөйлеуді тану өнімділігінің жалпы метрикасы болып табылады [85].

Өнімділікті өлшеу бойынша басты қиындықтары, тірек сөздер тізбегімен салыстырғанда танылатын сөздер тізбегі әртүрлі ұзындықта болуы мүмкін. WER Левенштейн ара қашықтығында бойынша сөздер деңгейінде орындалады. Бұл мәселені, танылған сөздер тізбегіне, жолдарды динамикалық түзетуді қолданып, эталонды сөздер тізбегі көмегімен бірінші түзету енгізу арқылы шешуге болады.

$$WER = \frac{100 * (S + I + D)}{N} \quad (3.1)$$

мұнда:

-N: тізбектегі сөздер саны

-S: ауыстырулар саны

-I: қоюлар саны

-D: жоюлар саны.

3.4 Сөйлеу сигналдарын өңдеу

Сөйлеуді тану шудың әсерінен, оқыту деректері мен тестілеу деректерінің арасында спектрлердің сәйкес келмеуі, тиімділігі кенет төмендеп кетеді. Сондықтан шуылы бар ортада сөйлеуді тану тапсырмалары әліде күрделі болып табылады. Бұл мәселені шешу үшін зерттеушілер оның орнын толтыру мақсатында көптеген әдістер ұсынды. Жалпы ол мәселенің шешімі сигналдар, элементтер немесе модельдер кеңістігінде қолданылуы мүмкін [86]. Спектральді сығу пәндік саланың оқыту және тесттік деректерінің арасындағы сәйкессіздікті азайтуға бағытталған, белгілерді бөліп алудың сенімді тәсілі болып табылады. Әдетте кепстральді коэффициенттердің (MFCC) Mel-жиіліктердің белгілерін бөліп алуда, олардың динамикалық диапазондарын азайту үшін Mel сүзгілерінің банк қуатына логарифмдерді қолданылады [87, 88]. Түпкі кепстральді талдау лог-функцияны түрбірдің тұрақты функциясына ауыстырып, RCC коэфициенттерін береді [89]. RCC коэфициенттері шуылға үздік тұрақтылықты көрсетті. RCC әдісінде сығылған сөйлеу спектрі (3.2)-де көрсетілгендей есептеледі:

$$P_C(m) = P(m)^{\gamma},\ 0 \le \gamma \le 1 \quad (3.2)$$

мұнда, $P_C(m)$ – сығылған спектр, $P(m)$ – түпнұсқалы спектр, γ – сығу коэффициенті m – банк сүзгілерінің индексі.

Спектральді сығымдаудың әдеттегі әдістері сигнал деңгейіндегі критерийлерді қолданады, олар сигнал / шу арақатынастары үшін параметрлерін калибрлеу. Бұл мақсатқа жетудің логикалық тәсілдерінің бірі, шыншылдық гипотезасының акустикалық ықтималдығын максималдау үшін спектральді сығымдаудың коэффициентін таңдау. Ол акустикалық ықтималдығының дұрыс гипотезасы мен басқа бәсекелес гипотезаның арақашықтығын ұлғайтады, сондықтан сөйлеуді дұрыс танудың ықтималдығы жоғарылайды. Демек, алдын ала өңдеу кезеңіндегі спектральді сығымдау векторы мен акустикалық ықтималдылық арасындағы арақатынасты сақтау, декодтау кезеңінде дұрыс гипотезасы тұжырымдалады. Бұл формулалар белгілерді бөліп алу алгоритмі мен акустикалық блоктың моделіне тәуеді.

Статистикалық модельдер негізіндегі сөйлеуді тану жүйелері акустикалық бірліктердің тізбегі жақсартылған сөйлеу сигналдарынан бөліп алынған белгілер векторын түрлендіру ықтималды жоғары $Z = \{z_1, \dots, z_t\}$. Бұл байқалған

ерекшеліктер кіріс сөйлеу сигналының фунциясы болып табылатындай, спектральді сығымдаудың векторы болып та табылады. Сөйлеуді танушы байестік классификациялау оптималды формуласы негізінде ең жақын ықтималды гипотезаны алады:

$$\widehat{w} = \arg\max_{w}(P(Z(\gamma)|w)\,P(w)) \tag{3.3}$$

мұнда, байқалған белгілер векторы γ спектральді сығымдау векторының функциясы болып табылады. Бұл формулада $P(Z(\gamma)|w)$ және $P(w)$ – сәйкесінше тиісті акустикалық және тілдік бағалаулар. Біздің мақсатымыз – танудың ең үздік сапасына жететін γ векторын табу. Сөйлеушіні немесе ортаны бейімдеу әдістері секілді, γ-ны оңтайландыру үшін, белгілі транскрипциясы бар бейімдеу деректері қажет. Біз w_C сөйлемінің дұрыс транскрипциясы бар деп есептейміз. Осылайша, $P(w_C)$ мәнін елемеуге болады, өйткені бұл мән γ мәнінен тәуелсіз тұрақты болып табылады. Осыдан кейін (3.3) теңдеуді γ қатысты барынша жақсартуға мүмкіндік аламыз:

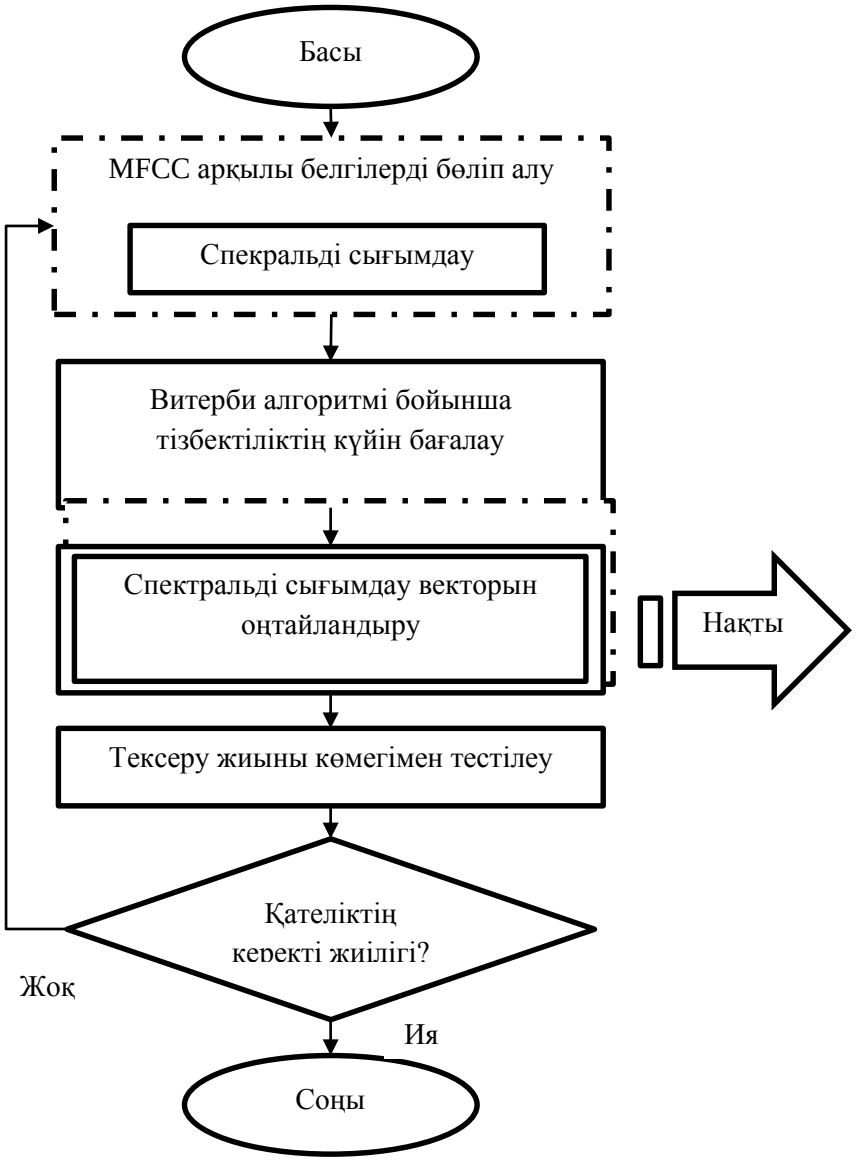
$$\hat{\gamma} = \arg\max_{\gamma}(P(Z(\gamma)|w_C)) \tag{3.4}$$

3.4.1 Сөйлеу сигналдарын өңдеудің спектральді сығымдау алгоритмі

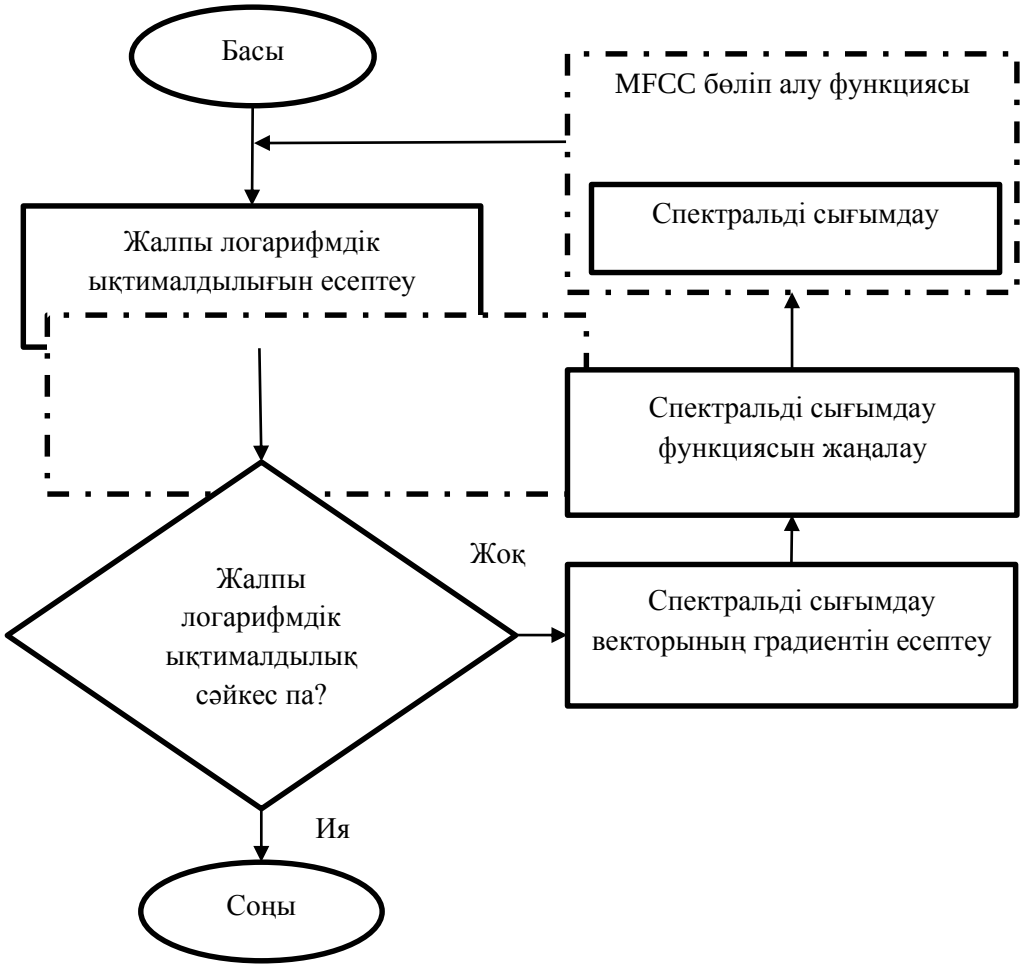
Бұл бөлімде дұрыс транскрипцияның акустикалық ықтималдығын арттыру үшін, алдыңғы бөлімде математикалық сипатталған спектральді сығымдау векторын басқарудың жаңа тәсілін ұсынамыз. Мұнда біз сұраққа жауап беруіміз керек. Егер де дұрыс транскрипция априориі белгілі болса, тану қажеттілігі болмайтын еді. Бұл сұраққа жауап беру үшін, дұрыс транскрипция бейімделу сатысында қажет, ал декодтау сатысында бұл параметрлер бекітілген екенін айтып өту керек. Алдымен, қолданушы априорлы белгілі транскрипциямен сөйлеуді дыбыстайды, содан кейін сөйлеу бастапқы параметрлермен бекітілген спектральды сығымдау блогы арқылы өтеді. Осыдан кейін, Витерби алгоритмі бойынша ең ықтимал күйлер тізбегі генерацияланады. Содан кейін күйлер тізбегін ескере отырып оңтайлы спектральды сығымдау векторы жасалады. Тану осы оңтайландырылған сүзгіні пайдаланып сынақ жинағында орындалады, егер қажетті қате жиілігі қанағаттандырылса, алгоритм аяқталады, әйтпесе жаңа күйлер тізбегі бағаланады. Итерациалық түрде спектральды сығымдау векторы табылды, ол белгілі транскрипциямен сөйлеудің толық логарифмдік ықтималдығын арттырады. Белгілер векторы алдымен жетілдірілген сөйлеу сигналынан алынады, содан кейін логарифмдік ықтималдық белгілі күйлер тізбегін ескере отырып есептеледі. Егер ықтималдық сөйкес келмесе, спектральды сығымдау векторының градиенті есептеледі және спектральды сығымдау векторы жаңартылады. Спектральды сығымдау осы жаңартылған векторды қолдана отырып орындалады, және жаңа белгілер векторлары алынады. Бұл процесс сөйкестік шарты орындалғанша қайталана береді.

Ұсынылған алгоритмде динамикаға және қоршаған ортаға бейімдеу әдістеріне ұқсас, спектральды сығымдау векторын бейімдеу бөлек автономды сеанста немесе жүйенің танудың қалыпты режиміне, ағымдағы уақыт режиміндегі қадамдық қадамын қосу арқылы іске асырылуы мүмкін. Жоғарыда түсіндірілгендей, автономды бейімдеу арқылы, әдетте арнайы бейімдеу сессиясын орындай отырып, қолданушы бейімдеу туралы хабардар, алайда онлайн-бейімдеуде қолданушы бейімдеудің 3.5 және 3.6-суретте орындалатындығын тіпті білмеуі де мүмкін.

Келіп түскен аудиодеректерді өңдеу және олардан сөйлеу сигналының белгіліерін бөліп алу [90].



Сурет 3.5 – Ұсынылған алгоритмнің блок-схемасы



Сурет 3.6 – спектарльді сығымдау векторының оңтайландыру блок-схемасы

3.5 Модельдерді құру құралына оқытуға дайындық

Train каталогының ақпараттары жүйені оқытуға арналған корпус пен құрал арасындағы ақпараттық байланыс файлдарын қамтиды. Олар келесі материалдардан тұрады (3.7 сурет):

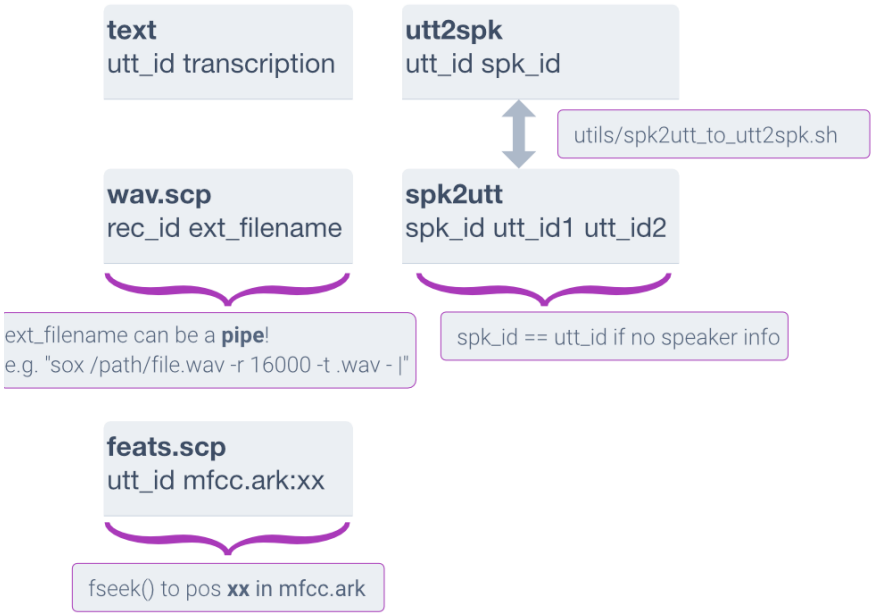
– text – сөйлеу корпусыдағы транскрипциялық файлдардың белгісі мен оған сәйкес транскрипциялық мәтінін қамтиды;

– utt2spk – сөйлеу корпусының транскрипциялық файлы белгісі мен оған сәйкес диктордың белгісін қамтитын деректер файлы;

– wav.scp – сөйлеу корпусындағы аудио файлдардың белгілері, файлдардың орналасқан жеріне дейінгі жол және sox қосымшасы арқылы файлдарды түрлендіру скрипты бар файл;

– spk2utt – бұл деректер файлы, сөйлеу корпусындағы барлық дикторлардың белгілері мен әр дикторға қатысты транскрипциялық файлдардың белгілерін қамтиды;

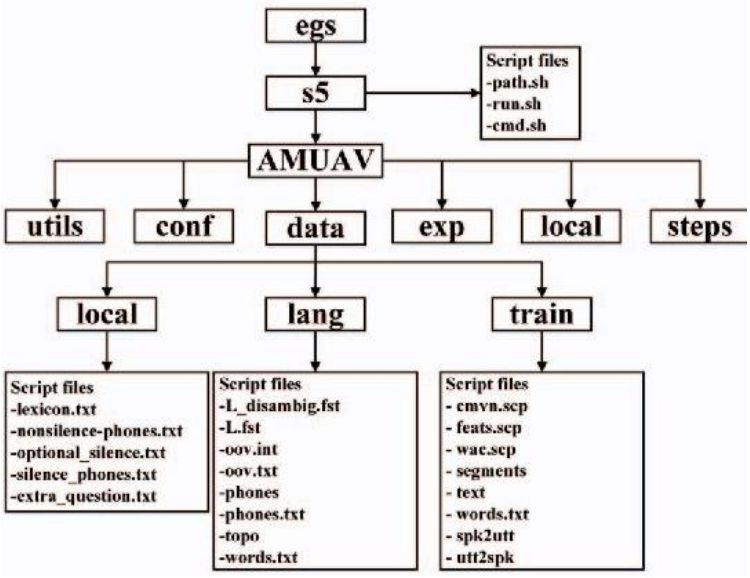
– feat.scp – сөйлеу корпусындағы транскрипциялық файлдардың белгісі мен оған қатысты сөйлеу сигналдарының бөлініп алынған белгілерін сақтайтын деректер файлы.



Сурет 3.7 – Kaldi ортасында жүйені оқыту үшін қажетті деректердің түрлері

Бұл ақпараттық файлдардың ішіндегі spk2utt және feat.scp құралдың дайын скрипттерімен тікелей автоматты түрде құрылса, қалған text, utt2spk және wav.scp ақпараттық файлдары үшін, олардың компьютер дискісінде орналасу, тілдік ерекшеліктеріне байланысты және де ұйымдастыру жұмыстарын ұтымды пайдалану арқылы, автоматты түрде орындалатын скрипттік қосымшалар құрылды. Қосымшалар Python бағдарламалау тілінде жазылған.

Тілдік және акустикалық модельдерді алу үшін, Kaldi құралының көмегімен рекуррентті желілерді оқыту жұмыстары жүргізіледі. Ол үшін ақпараттық файлдардың сұлбасы 3.8-суреттегідей орналасуы қажет. Бұл сұлба – Linux Ubuntu операциялық жүйесінде орындалған.



Сурет 3.8 – Kaldi құралындағы тілдік және акустикалық модельдерді құруға арналған сұлба

3.6 Акустикалық модельді құру

Акустикалық модель – сөйлеу сегментін дыбыстық деңгейіндегі ұқысастығын анықтап, тану сапасын бағалауға мүмкіндік береді және сөйлеуді автоматты тану үшін қолданылатын аудиосигналдағы дыбыстарды ұсынады. Әрбір дыбыс үшін алдымен, сөйлеудегі осы дыбыстың дыбысталуын анықтайтын күрделі статистикалық модель құрылады. Оның негізгі мақсаты – акустикалық толқындарды, тілдегі бір сөз бен екінші басқа сөзді ажыратуға мүмкіндік беретін, дыбыстың қарапайым тілдік бірліктері болатын, фонемалардың статистикалық құрамымен сәйкестендіру болып табылады [91].

Бұл кезеңнің тапсырмасы – ол кіріккен сөйлеуді тануды жүргізуге мүмкіндік беретін акустикалық бірліктердің моделін құру. Көп жағдайда, сөз табының дыбысталуы бойынша ажыратылатын, сонымен қатар дыбысталған сөйлемнің мағынасына әсер ететін, акустикалық бірлік ретінде тілдің фонемалары таңдалады. Акустикалық модель сөйлеу белгілерінің есептелген векторының тізбегінің дыбысталған сөйлемнің фонетикалық бірліктерінің тізбегіне ауыстыру тапсырмасын шешеді [92].

Сөйлеу сигналдарын $S = s_1, s_2, \ldots, s_T$ қысқа уақыттық кадрлар тізбегі ретінде қарастырайық. M тізбекті фонемалар келесідей болсын $F = f_1, f_2, \ldots, f_M$ және N тізбекті сөздер $W = w_1, w_2, \ldots, w_N$ түрде бейнеленсін. Сөйлеуді тану кезіндегі мақсатымыз S аудиокірісінен W сөздер жиынын алдын ала болжап табуында:

$$\widehat{W} = \underset{W}{\operatorname{argmax}} P(W|S) \tag{3.5}$$

$$\widehat{W} = \underset{W}{\operatorname{argmax}} P(S|F)\, P(F|W) P(W) \tag{3.6}$$

Мұнда $P(W)$ сөздер жолы қазақша сөйлем екенінің ықтималдығын ұсынады, яғни $P(W)$ тілдік модель болып табылады. $P(S|F)$ шамасы айтылым моделі деп аталады, ал $P(F|W)$ акустикалық модель.

Жүйені оқыту GMM-HMM бойынша келесі негізгі деңгейлерден тұрады [93]:

- 30000 сөйлемді 1000 гаусспен монофондық модельді (mono) оқыту. MFCC-дің 13-өлшемдік белгісі, бірінші және екінші туындысымен толыққан, CMN-мен бірге қолданылады;
- 100000 сөйлемнен тұратын 3200 байланысқан күймен және 30000 гаусспен оқытылған бірінші трифондық модель (tri1). MFCC-дің 13-өлшемдік белгісі, бірінші және екінші туындысымен толыққан, CMN-мен бірге қолданылады;
- 100000 сөйлемнен тұратын 4000 байланысқан күймен және 70000 гаусспен оқытылған екінші трифондық модель (tri2). MFCC-дің 13-өлшемдік белгісі, бірінші және екінші туындысымен толыққан, CMN-мен бірге қолданылады;
- барлық деректермен 6000 байланысқан күймен және 140000 гаусспен оқытылған үшінші трифондық модель (tri3). Белгі ретінде 9 көршілес кадрлар (орталық кадр және 4 оң жақтағы және 4 сол жақтағы) үшін алынған MFCC-дің

13-өлшемдік белгісі, өлшемділігі 40-қа дейін төмендетілген LDA-MLLT түрлендіруін қолданған CMN бірге қолданылады;

- барлық деректермен 6000 байланысқан күймен және 140000 гаусспен оқытылған төртінші трифондық модель (tri4). Белгі ретінде 9 көршілес кадр (орталық кадр және 4 оң жақтағы және 4 сол жақтағы) үшін алынған MFCC-дің 13-өлшемдік белгісі, өлшемділігі 40-қа дейін төмендетілген LDA-MLLT түрлендіруін және fMLLR түрлендіруі арқылы дикторға бейімдеуін қолданған CMN бірге қолданылады;

- барлық деректер бойынша bMMI критерийін қолдану арқылы төртінші трифондық модельді дискриминативті оқыту.

Сонымен бірге әрбір келесі GMM-HMM моделін оқыту үшін, алдыңғы модельдің көмегімен алынған тегістеу қолданылады (сурет 3.9).

DNN бойынша оқытуын инициализациялау оқыту сапасына үлкен әсер етеді. Бұл бөлімде акустикалық модельдерді терең нейрондық желілер алгоритмі негізінде оқыту ұсынылды.



Сурет 3.9 – Модельдердің құрылу иерархиясы

3.7 Тілдік модельді құру

Сөйлеуді тану тілдік модельдерді қажет, ол кіріс мән ретінде әрбір жаңа сөздер тізбегінің тілге қатысты екенін анықтау ықтималдығын білуге мүмкіндік береді.

Тілдік модельдерді құру кезінде Kaldi – SRILM қосымша кітапханасын қолданады, ол үшін N-граммдық моделін алуға арналған орындалатын модулі шақырылады. N-граммдық модель арнайы ARPA форматының көмегімен орындалады [94].

Униграмды модель ақпаратты іздеуде қолданылды, ол соңғы бір жағдайдағы бірнеше автоматтың комбинациясы ретінде қарастыруға болады. Контексттегі мағыналық ықтималдығын корсетуі мүмкін.

$$P(w_1 w_2 \dots w_L) = \prod_{i=1}^{L} P(w_i) \tag{3.7}$$

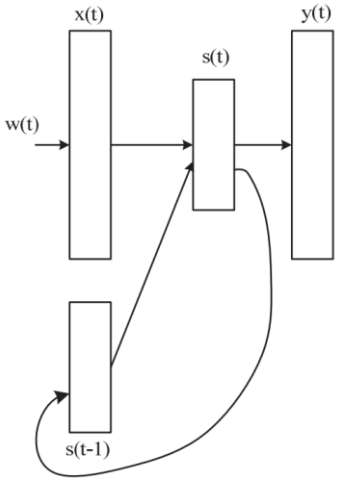
Биграмды және триграмды сөздері n-граммды тіл моделінің сәйкесінше n = 2 және n = 3 екендігін білдіреді. Шартты ықтималдық n-грамм модельдің жиілігіне қарай келесі түрде есептеуге болады:

$$P(w_1 w_2 \dots w_L) = \prod_{i=n}^{L} P(w_i | w_{i-1} w_{i-2} \dots w_{i-n}) \tag{3.8}$$

3.7.1 Рекуррентті нейрондық желілер негізіндегі тіл моделі

RNN негізіндегі тілдік модель құрылымы жалпы RNN құрылымына балама. Дегенмен, бірнеше маңызды бөлімшелері айырықшаланады.

Рекуррентті нейрондық желі алғаш рет [95] жұмыста ұсынылды. Рекуррентті нейрондық желілерде жасырын қабат алдыңғы барлық оқиғаларды сақтайды, сондықтан контекст көлемі шектеусіз болады. Рекуррентті нейрондық желілерді тілдік модельдеуің архитектурасы 8 суретте ұсынылған [96].



Сурет 3.10 – Рекуррентті нейрондық желілердің архитектурасы

Желі *x* кіріс қабатынан, *s* жасырын қабатынан және шығыс *y* қабатынан тұрады. Кіріс қабаты, ағымдағы сөзді қамтитын *w(t)* векторы мен алдыңғы қадамда алынған жасырын қабаттың шығыс мәнін ұсынатын *s(t-1)* векторының бірігуінен алынған, *x(t)* векторынан тұрады. *w(t)* векторының көлемі сөдік көлеміне тең. Әдетте жасырын қабаттың көлемі эмпирикалық таңдалады. Кіріс, жасырын және шығыс қабаттары келесідей есептеледі:

$$s(t) = w(t) + s(t-1) \tag{3.9}$$

$$s_j(t) = f\left(\sum_i x_i\,(t) u_{ji}\right) \tag{3.10}$$

$$y_k(t) = g\left(\sum_j s_j\,(t) u_{kj}\right) \tag{3.11}$$

Мұнда *f(z)* – сигмоидальді активациялық функциясы:

$$f(z) = \frac{1}{1+e^{-z}} \tag{3.12}$$

g(z) – softmax функциясы:

$$g(z_m) = \frac{e^{z_m}}{\sum_k e^{z_k}} \tag{3.13}$$

s(0)-ді инициализациялау үшін аз мәнді векторді орнатуы мүмкін, мысалы, 0,1 мәні, үлкен көлемді деректерді өңдеу кезінде инициализацияның шешуші мәні жоқ. *s(t+1)* келесі уақыттық қадамдарда *s(t)*-нің көшірмесі болып табылады. *x(t)* кіріс векторы [97]

Әдетте *s* контекстік (жасырын) қабатының көлемі 30 – 500 жасырын бірліктен тұрады. Жасырын қабаттардың көлемі оқыту деректерінің көлемін көрсетуі керек, мысалы, үлкен көлемді деректер үшін көп жасырын қабат қажет.

Желілер, оқыту корпусының барлық деректері тізбектей ұсынылған, бірнеше эпохада оқытылады. Салмақтар аз көлемді мәндермен (0,1 дисперсиясы мен нөлдік орташасымен кездейсоқ гаусстық шуыл) инициалданады. Желіні оқыту үшін біз стохастикалық градиент түсу әдісімен бірге стандартты кері үлестіру алгоритмін қолданамыз. Оқытудың бастапқы жылдамдығы 0,1-ге тең. Әрбір эпохадан кейін желі тексеру дерекерімен тексеріледі. Егер нақтылықты тексеру деректерін тіркеу ықтималдығы көбейетін болса, оқыту жаңа эпохада жалғасады. Егер айтарлықтай жақсару байқалмаса, α оқыту жылдамдығы әрбір жаңа эпоханың басында екі есе азаяды. Осыдан кейін, айтарлықтай жақсару

болмаса, оқыту аяқталады. Әдетте конвенгенция 10 – 20 эпохадан кейін орындалады [98].

Біздің тәжірибелерімізде желілер қайта оқытылмайды, өйкені көп мөлшердегі жасырын қабаттарды қолданғанның өзінде желі салмақтарының үлендігіне қарамастан айтарлықтай жақсартуларға әкелген жоқ. w(t) алдыңғы сөзге s(t-1) контекстке байланысты y(t) келесі сөздің болу ықтималдығының үлестіруін ұсынады. Ал Softmax функциясы осы ықтималдықты үлестірілуіне кепілдік береді, яғни кез келген m сөзі үшін y_m (t)> 0 және $\sum_k y_k(t) = 1$.

Оқытудың әрбір кезеңінде қателік векторлары кросс-энтропиялық критерийлерге сәйкес есептеледі, ал салмақтарының жаңаруы кері үлестіру стандартты алгоритм көмегімен орындалады. Тәжірибелерді жүргузі ортасы Kaldi ортасында орындалған (3.1 тарау).

RNN тілдік моделінің желіге кірісі ретінде сөздердің анықтауышының тізбегі беріледі. Ол жерде сөздің анықтауышы қолданыстағы сөздіктегі нақты сөзді ұсынады. Тұтастық тізбек, бір қадам солға қозғалған кіріс тізбегі болып беріледі. Мысалы, сөздік қорын ескеріп (қалай, келемін, болған, мен, сен, бүгін, ертең) және сөздерді анықтаушылардың бейнеленуі (қалай: 1, келемін: 2, болған: 3, мен: 4, сен: 5, бүгін: 6, ертең: 7]) «мен келемін» тізбегі («бүгін» сөзімен бірге тізбектің келесі сөзі болып табылады) (1, 2, 5) сөздерді анықтаушы тізбегін ұсынады. Мақсаттық тізбек (2, 5, 6) болып беріледі.

Желіге кіріс тізбегін жасырын бөліктерге берудің алдында сөздердің анықтаушысы аз мөлшерлі векторлар кеңістігіне енгізіледі, яғни әрбір кіріс сөзі оның енгізу векторымен, модельді енгізу матрицасын қолдану арқылы сәйкестендіріледі. Әдетте енгізу матрицасы кездейсоқ инициалданып, оқыту кезінде бейімделеді [99].

Оқыту корпусы алдыңғы тарауда баяндалғандай тілдік корпустың қазақша сөйлемдерін (сөздерін) қамтиды. Нейрондық желінің баптаулары RNN 30 деп бейнеледік, ол рекуррентті нейрондық желі 30 жасырын қабатта оқытылды дегенді білдіреді.

Ал LSTM-де енгізу сөздерінің тезбегі бір немесе бірнеше жасырын бірлікті қабаттарда өңделеді. LSTM-нің бөліктерінің соңғы шығыстары қосымша шығыс қабаты softmax белсендірілу функциясымен өңделеді. Әрбір кіріс сөзіне шығыс қабаты сөздікте бар сөздер мен әрбір жазба тізбектегі келесі сөздің болжамдалған ықтималдылығымен сәйкес санды векторды құрады. Желі кросс-энтропиялық, бүтіндік тізбек мен болжалды тізбек арасындағы жоғалтуын азайту жолымен оқытылады [100].

Осылай біз қазақ тілінің кіріккен сөйлеуінің рекуррентті нейрожелілік модельдері негізінде тілдік модельдерді аламыз. Оның сапасын бағалау келесі бөлімде баяндалған.

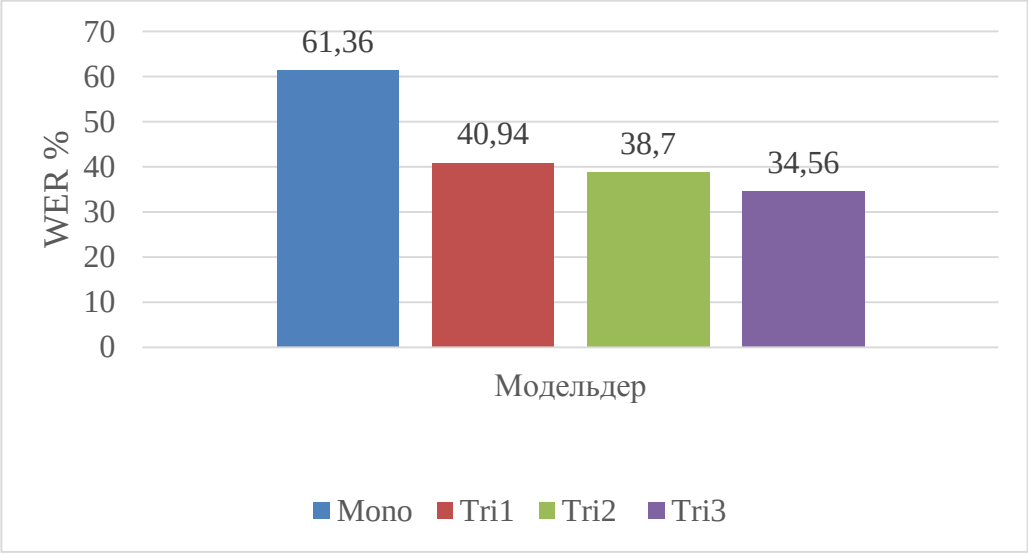
3.8 Алынған модельдердің сапасын бағалау

Мәтіндік және акустикалық деректерді сөйлеуді тану құралында оқыту арқылы алынған акустикалық және тілдік модельдердің сапасын анықтау мақсатында тәжірибелік жұмыстар жүргізілді [101]. Дәстүрлі модельдер мен

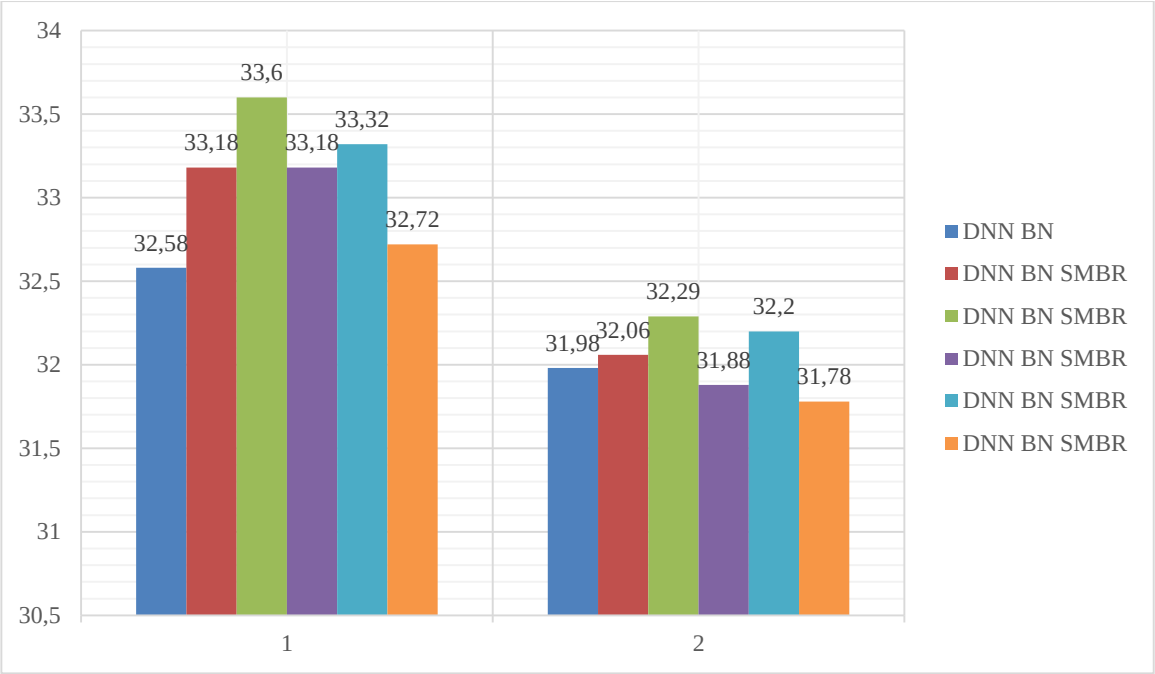
терең нейрондық желілерді бірқатар бейімделу алгоритмдерімен және рекурренттік нейрондық желі мен үйірткілі нейрондық желіні қолдану арқылы алынған модельдердің нәтижесін 3.1 кестеден көруге болады. Оң жақтағы бірінші бағанда dev жинағы бойынша WER сапасы, ал екінші бағанда test жинағының WER сапасы көрсетілген. Алынған нәтижелердің бойынша терең рекуррентті желілердің сапасы, басқа модельдермен салыстырғанда дәлдігі жоғары екенін көрсетіп отыр [102, 103]. Олардың графикалық түрін 3.11 - 3.12 - суреттерде көруге болады.

Кесте 3.1 – Алынған модельдердің сапасын бағалау

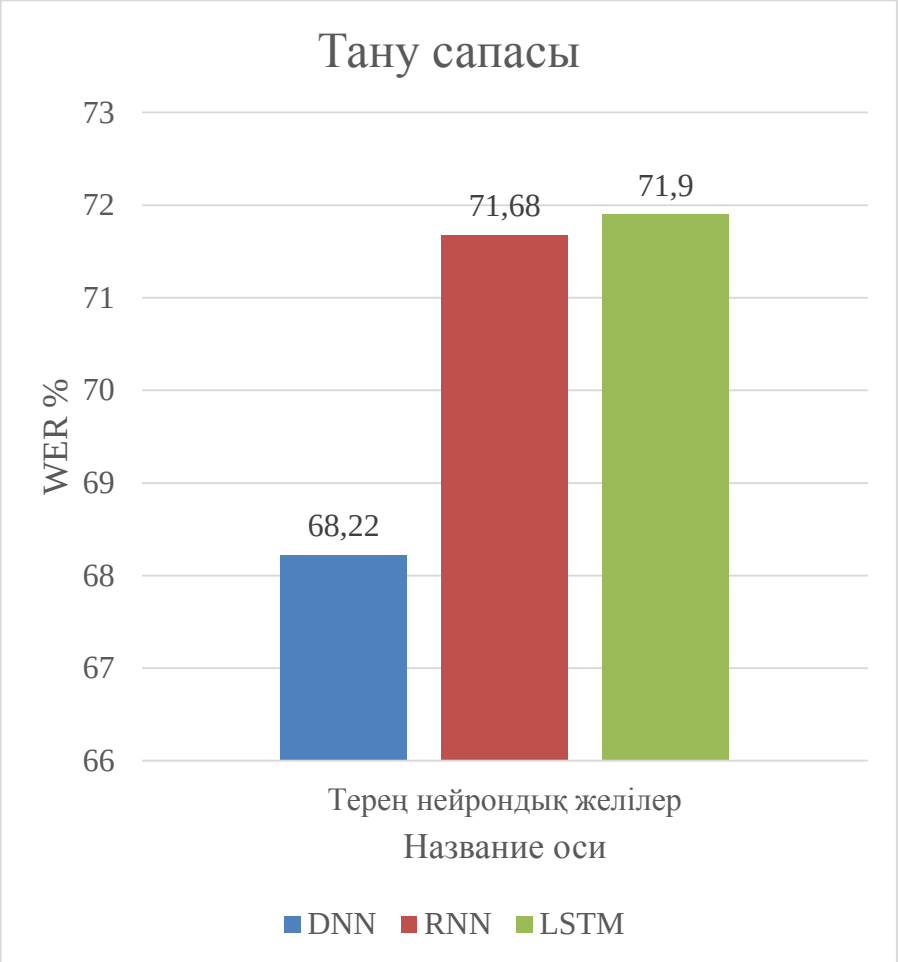
№	Модель түрлері	Сөз бойынша қате WER (dev) %	Сөз бойынша қате WER (test) %
1	Mono	62.46	61.36
2	Tri1. Deltas + Delta-Deltas	41.61	40.94
3	Tri2. LDA + MLLT	38.17	38.70
4	Tri3. LDA + MLLT + SAT	36.76	34.56
5	Терең нейрондық желі DNN алдын ала оқыту BN	32.58	31.98
6	Терең нейрондық желі DNN алдын ала оқыту BN SMBR	33.18	32.06
7	Терең нейрондық желі DNN алдын ала оқыту BN SMBR итерация1	33.60	32.29
8	Терең нейрондық желі DNN алдын ала оқыту BN SMBR итерация 2	33.18	31.88
9	Терең нейрондық желі DNN алдын ала оқыту BN SMBR итерация 3	33.32	32.20
10	Терең нейрондық желі DNN алдын ала оқыту BN SMBR итерация 4	32.72	31.78
11	RNN 30	28.25	28.32
12	LSTM	28.20	28.10



Сурет 3.11 – Алынған модельдердің сапасының графиктік көрсеткіші



Сурет 3.12 – Терең нейрондық желілердің сөйлеуді танудың қатесі



Сурет 3.13 – Терең рекуррентті нейрондық желілердің тану сапасы

4 КІРІККЕН ҚАЗАҚ СӨЙЛЕУІН ТАНУ КЕШЕНІН ҚҰРУ

Қазақ тілі сөйлеуін автоматты тану кешенін құру бойынша тапсырмаларын шешу kaldi ашық кодты құралының негізінде, Visul Studio 13 бағдарламалау ортасында, C++ бағдарламалау тілін қолдану арқылы жүзеге асырылды. Серверлік бөлімі Kaldi GStreamer құрылымдық бөлім арқылы орындалды [104].

Сервер бөлімін жүзеге асыру кезінде қолданылған технологиялар:

– Kaldi Gstreamer сервері;
– docker-kaldi-gstreamer-server;
– Recorder.js;
– dictate.js.

4.1 Kaldi Gstreamer сервері

Бұл толықфункционалды толық дуплекстік Kaldi инструментариі негізінде нақты уақыт тәртіптемесінде сөйлеуді тану сервері мен GStreamer құрылымы және Python-да жүзеге асырылған [105].

Ерекшеліктері:

– Websockets негізіндегі толықдуплексті байланыс: сөйлеу жүріп жатыр, жартылай гипотезалар пайда болады (Android дауыс жиыны туралы ойлаңыз).

– үлкен көлемде масштабталынатындар: сервер негізгі және жұмыстық компоненттерден (құрауыш) тұрады; параллельді танудың бір сеансы үшін бір жұмысшы қажет; жұмысшыларды шеберге қарамастан қашықтықтан машиналарда іске қосуға және тоқтатуға болады.

– сөйлеуді сегменттей (саралай) алады, яғни үнсіздіктерге негізделген ұзын сөйлеу сигналы қысқа сегменттерге бөлінеді.

– өз бетімен ұзақ сөйлеумен енгізуді қолдайды (мысалы, сіз оған жанды сөйлеуді бере аласыз)

– Kaldi-дің GMM үлгілерін және «онлайн-DNN»-ді қолдайды

– үлкен тілдік үлгісі бар тану торын резервті көшіруді қолдау

– сұратымдар арасындағы акустикалық үлгінің бейімделу қалпын сақтауды қолдайды

– аудиокодектің шектелмеген жиынын қолдау (іс жүзінде тек Gstreamer қолдайтындарды ғана)

– ішкі программаларды қолдану арқылы танудың бастапқы нәтижелерін көшіріп жазуды қолдайды (сөздерді сандарға айналдыру үшін қолданылуы мүмкін және т.б.)

– Python, Java, Javascript клиенттері қол жетімді

4.1.1 Kaldi Gstreamer орнату және баптау

Kaldi-ді және осы бағдарламалық қамтамасыз ету талап ететін барлық басқа пакеттерді орнату айтарлықтай қиын болуы мүмкін. Барлық алдын ала шарттарды қолдан құрудың орнына, Х.Э. Сильва жасап шығарған Docker [106] файлын қолдануға болады.

4.1.1 Kaldi Gstreamer серверіне қойылатын талаптар

Kaldi Gstreamer серверін жұмысын жүргізу үшін Python 2.7 келесі пакеттерімен бірге орнатылуы қажет:

- Tornado 4;
- ws4py (0.3.0 – 0.3.2 нұсқаларымен ғана жұмыс істейді);
- YAML;
- JSON.

4.1.2 Серверді жұмысын бастау

Бас серверді іске қосу. Негізгі сервер 217.15.186.70 адресінде 23422-ші порт бойынша қолжетімді. Оның жұмысын бастау үшін, серверде келесі команданы орындау қажет:

python kaldigstserver/master_server.py --port=23422

Сервердің жұмысын бастағаннан кейін, жұмысшылар жұмысын бастау жүзеге асырылады.

4.1.3 Жұмысшыларды іске қосу

Бас сервер сөйлеуді тануды жүзеге асырмайды, ол тек жұмысшыларға қызметтерді бөліп беру жұмыстарын атқарады. Өйткені әрбір тану сеансына бір жұмысшы қажет. Осылайша жұмысшылар саны, кем дегенде ықтимал параллельді тану сеанстарының санындай болуы керек. Бұл жағдай жұмысшылардың толықтай тәуелсіз болып, олардың бірнеше машинада жұмыс істеу мүмкіндігінің арқасында, іс жүзінде шектелмеген параллельдікті ұсынатындығы тиімді.

Жұмысшы қолдана алатын декодердің екі түрі бар: Gldreamer плагины негізіндегі онлайн-графикаға арналған Kaldi onlinegmmdecodefaster немесе kaldinnet2onlinedecoder жаңа плагины негізіндегі декодер. Біріншісі, GMM модельдерін қолдайды, соңғысына і-векторына енгізу бар DNN негізіндегі «online2» модельдерді қажет етеді [107].

Жұмысшыны іске қосу үшін, алдымен конфигурация файлын жазу қажет. GMM-HMM модельдерін қолданатын, конфигурация kazakh_worker.yaml файлылы арқылы қолжетімді. DNN негізіндегі модельдерді қолданатын жұмысшылар kazakh_dnn.yaml. файлы арқылы іске қосылады.

Жұмыстық пайдаланушы «onlinegmmdecodefaster» қолдану. Жұмысшыны іске қосу алдында, GST плагинының жолында Kaldi-дың src/gst-plugin (libgstkaldi.so файлының бар болуы қажет) каталогын қамтуы қажет. Оны келесі команда арқылы жүзеге асырамыз:

export GST_PLUGIN_PATH=~/kaldi_kazserver/src/gst-plugin

Плагин жұмысын тексеру үшін төмендегідей команданы орындаймыз:

gst-inspect-1.0 onlinegmmdecodefaster

Бұл команда Kaldi-дың GStreamer плагины жайлы ақпаратты басып шығаруы қажет.

Жоғарыдағы тапсырмаларды орындағаннан кейін, жұмысшыны қосу жүзеге асырылады. Ол келесідей команданы орындау арқылы жүзеге асырылады:

python kaldigstserver/worker.py
- u ws://217.15.186.70:23422/worker/ws/speech -c kazakh_worker.yaml

-u ws://217.15.186.70:23422/worker/ws/speech параметрі, жұмысшы қосылатын негізгі сервердің адресін көрсетеді. Ал порт серверді шақыратын портты қолданамыз, және ол ашық болуы қажет.

Жұмысшыны қосу командасын қолдану арқылы кез келген мөлшерде жұмысшыларды параллельді түрде қосуға болады.

4.1.4 Сервер мен клиенттік қосымшаларды байланыстыру

Клиенттік бөлімнің жұмысын бастау командалары client.py. файлында орналасқан. Серверлік бөлім және жұмысшылардың жұмысын бастағаннан кейін клиенттік бөлімді іске қосуға болады. Ол үшін келесі команданы шақырамыз:

python kaldigstserver/client.py -r 32000 data/audio.raw

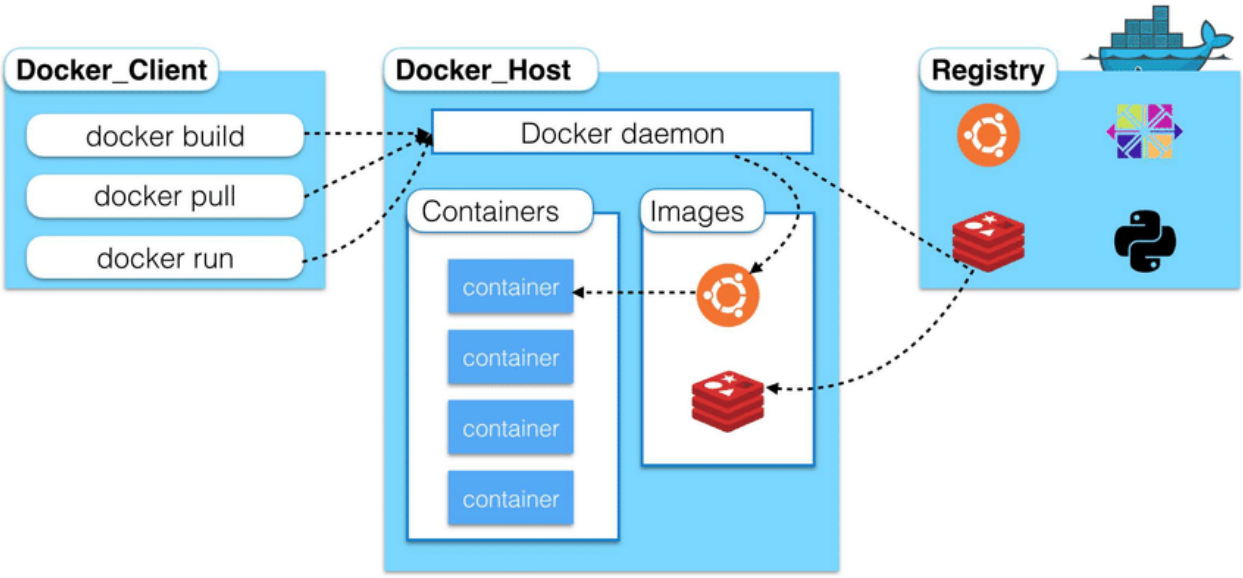
бұл командадағы «-r 32000» клиенттің серверге секундына 32000 байт жылдамдықта аудио жіберуі жайлы хабар береді. Аудио файлдың өңделмеген түрі 16000 дискреттеу жиілігін қолданады. С 16-разрядты кодтама 32000-қа ауыстырылады. Сонымен қатар *.ogg және *.wav форматындағы дыбысты файлдарды жіберуге болады, мысалы:

python kaldigstserver/client.py -r 4800 test / data / english_test.ogg

бұл мысалда файлды жіберу жылдамдығы 4800-ге тең. Өйткені *.ogg файлының биттік жылдамдығы 37,5 мыңнан тұрады, сондықтан байттық жылдамдығы 4800-ге тең.

4.1.5 Сервер мен жұмысшыларды автоматтандыру

Серверге әрбір қатынас жасаған сайын, оны қосу және тоқтату, сонымен қатар әрбір жаңа параллельді сеансты бастау үшін жұмысшыларға арналған жеке командаларды орындау жұмыстарын автоматтандыру қажет [108]. Бұл мәселелер Docker бағдарламалық қамтамасы арқылы шешіледі, яғни сервер мен жұмысшыларды қосу және тоқтату тапсырмаларын автоматы түрде жүзеге асыруға мүмкіндік береді. Оның жалпы архитектурасы 4.1 суретте көрсетілген.



Сурет 4.1 - Docker бағдарламалық қамтамасының жалпы архитектурасы

Бір докерді екі сценарийде пайдалануға болады. Сервер мен жұмысшыны бір машинада құруға немесе тек жұмысшыны құрып, оны алдын ала құрылған сервермен байланыстыруға болады..

Акустикалық және тілдік модельдер серверде орналасқандықтан бас компьютерде сервер құрамыз. Модельдер басты машинада /media/kazakh_models жолында деп көрсетілген докер контейнер құрамыз:

docker run -it -v /media/kaldi_models: /opt/kazakh_models/ mussa/docker-kaldi-gstreamer-server: /bin/bash

осыдан кейін контейнер ішінен келесі қызметті іске қосамыз:

/opt/start.sh -y /opt/models/kazakh_dnn.yaml -m 217.15.186.70 -p 23422

Ол локалды машинада жұмысшыны құрып, оны 217.15.186.70 адресіндегі бас серверге 23422 порты бойынша қосылады.

Құрылған контейнерде /opt каталогында барлық іс қимылдар жазылатын worker.log файлы құрылады.

Жұмыстық серверді тоқтату үшін контейнерде келесі команданы орындау қажет:

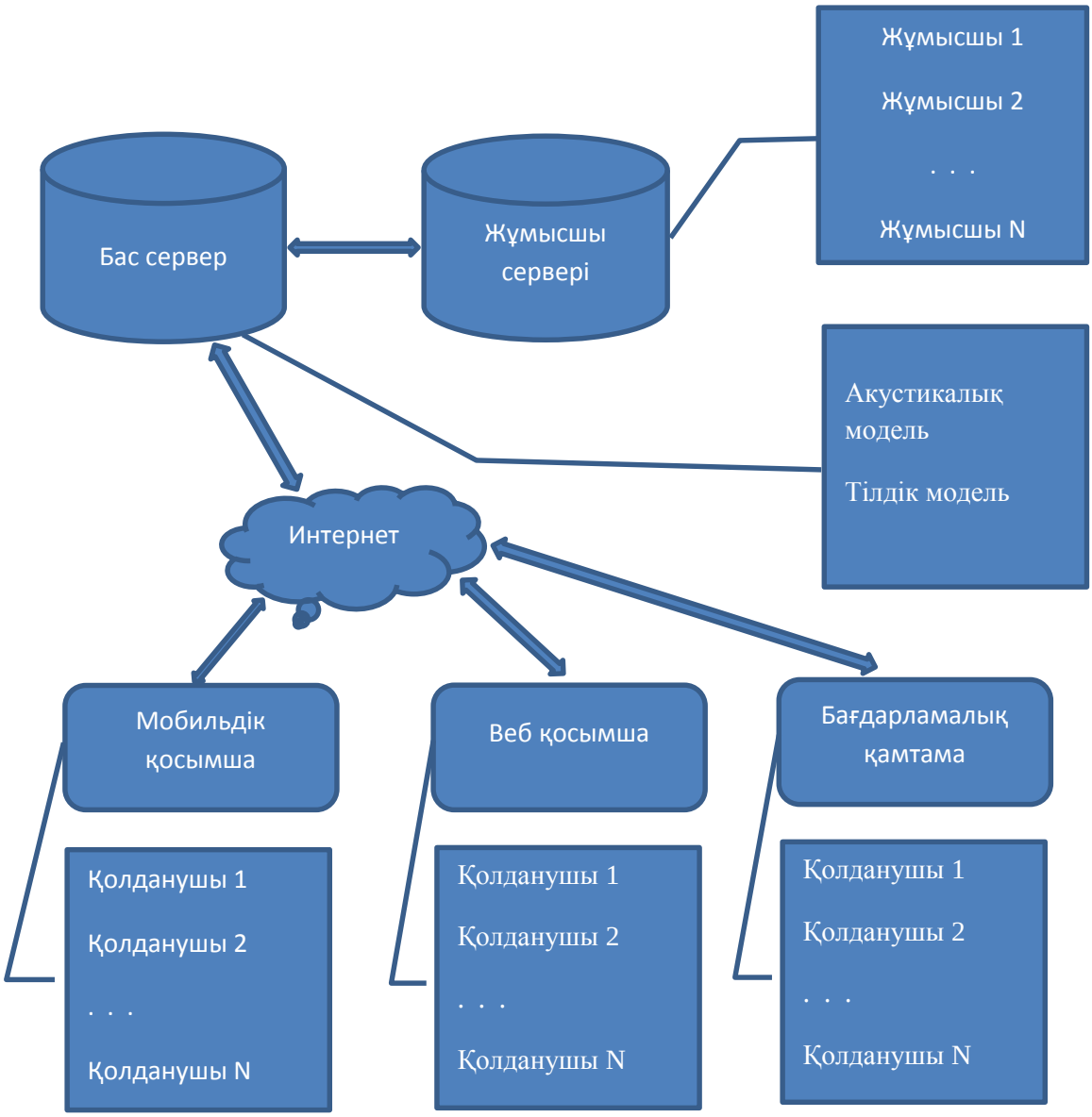
/opt/stop.sh

4.2 Сөйлеуді тану кешенінің пайдаланушы қосымшаларын құру

Кіріккен қазақ сөйлеуін тану кешені қолжетімді болуы үшін оның windows операциялық жүйесінде бағдарламалық қамтамасы, интернет браузерлердегі

қолжетімді веб-қосымшасы және смартфондарға арналған мобильдік қосымшасы құрылды.

Кіріккен қазақ сөйлеуін тану кешенінің архитектурасы 4.2 суретте бейнеленген [109].

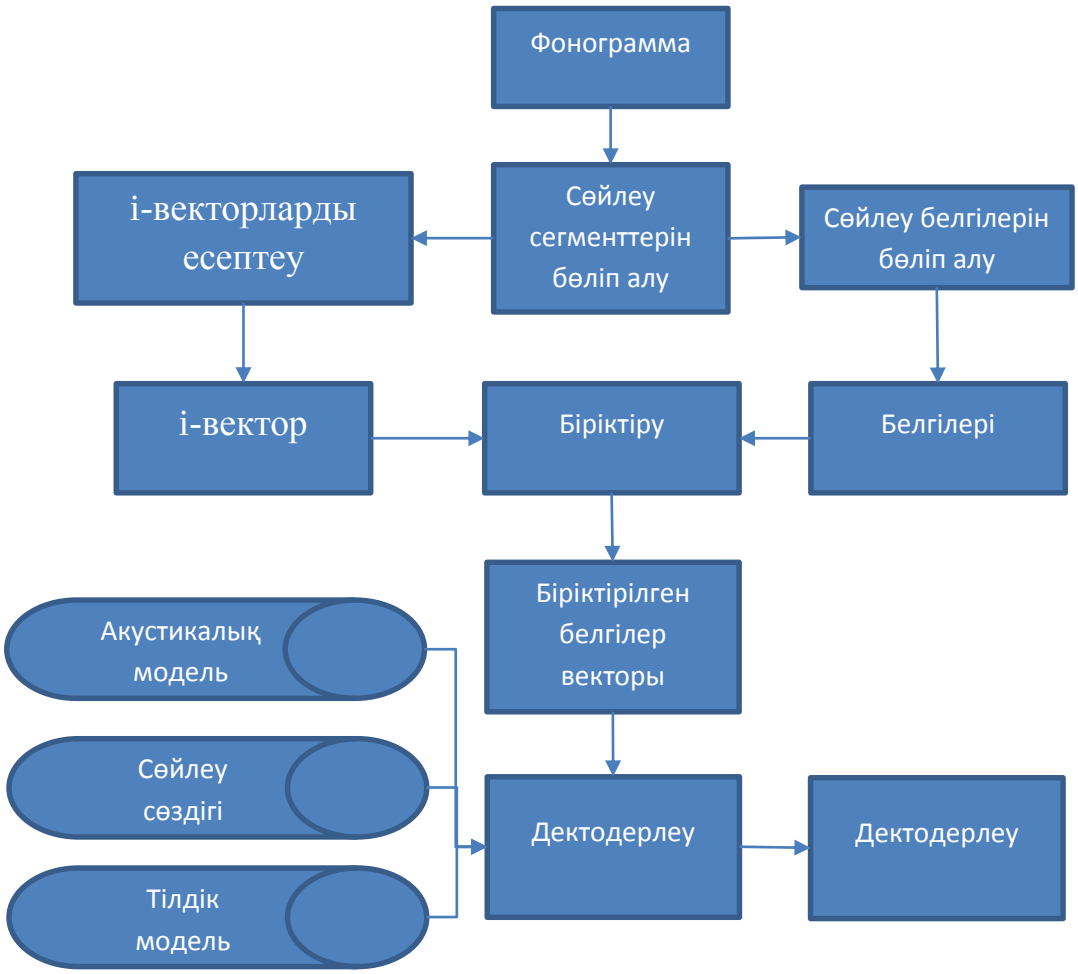


Сурет 4.2 – Кіріккен қазақ сөйлеуін тану кешенінің жалпы түрі

4.2.1 Сөйлеуді тану бөлімі

Сөйлеуді тану бөлімі 4.3 суретте ұсынылған. Бұл бөлім кіріс ретінде, 8000 Гц дискреттік жиілікте жазылған, 16 биттік wav-контейнердегі сөйлеу фонограммасын қабылдайды. Кіріс сигналдарын өңдеудің бірінші деңгейі сөйлеу сегменттерін бөліп алу болып табылады. Бөлініп алынған сөйлеу сегменттері бойынша і-векторларды есептеу фонограмманың әрбір бөліміне жүргізіледі және де 11xfbank-3 конфигурациясына сәйкес әрбір кадрге белгілерін бөліп алу орындалады. Фонограмманың сәйкес бөліміндегі әр кадрға құрылған

белгілер векторы мен і-вектор біртұтас белгілер векторына біріктіріледі. Біріктірілген белгілер векторлары бойынша терең нейрондық желілердің і-векторларын қолдану арқылы жоғарғы деңгейлі белгілер векторларын есептеліп алынады. Одан ары декодтау немесе шындыққа ұқсас сөздер тізбегін алу жүзеге асырылады. Декодтау кезінде акустикалық модель, тілдік модель және сөйлеу сөздігі қолданылады. Декодтау кезінде алынған сөздер тізбегі (жаксы гипотеза) мәтіндік жолға жазылады.



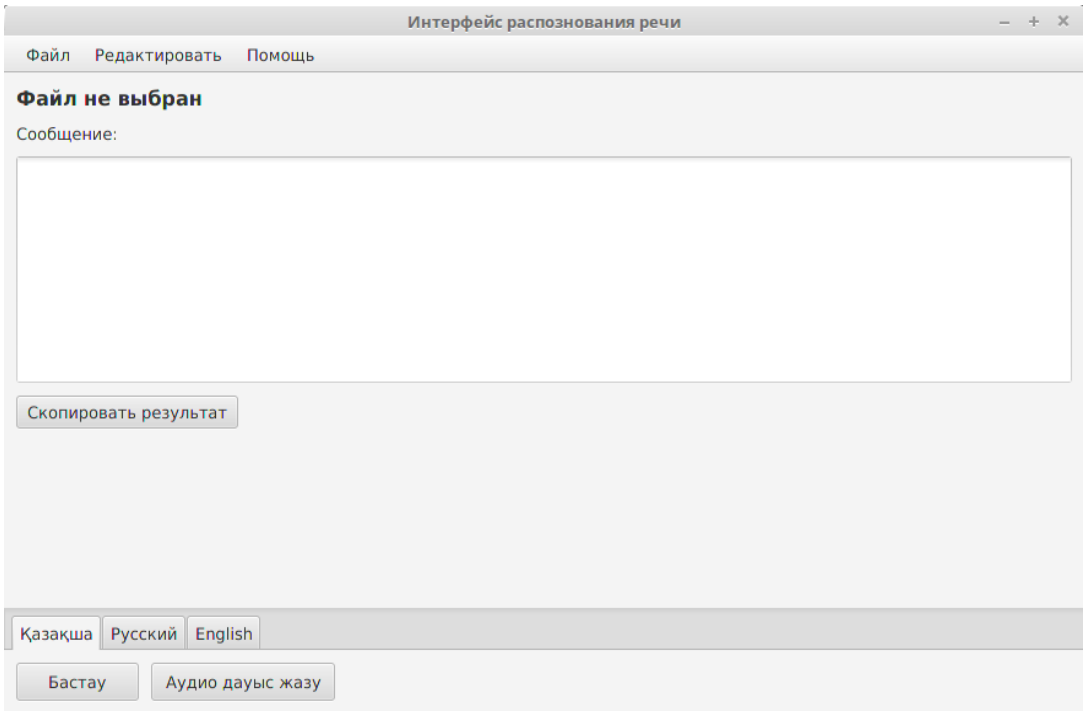
Сурет 4.3 – Кіріккен қазақ сөйлеуін тану бөлімі

Бұл бөлім Java бағдарламалау тілінде жазылған. Және келесі операциялық жүйелер мен процессорлардың архитектураларын қолдайды:

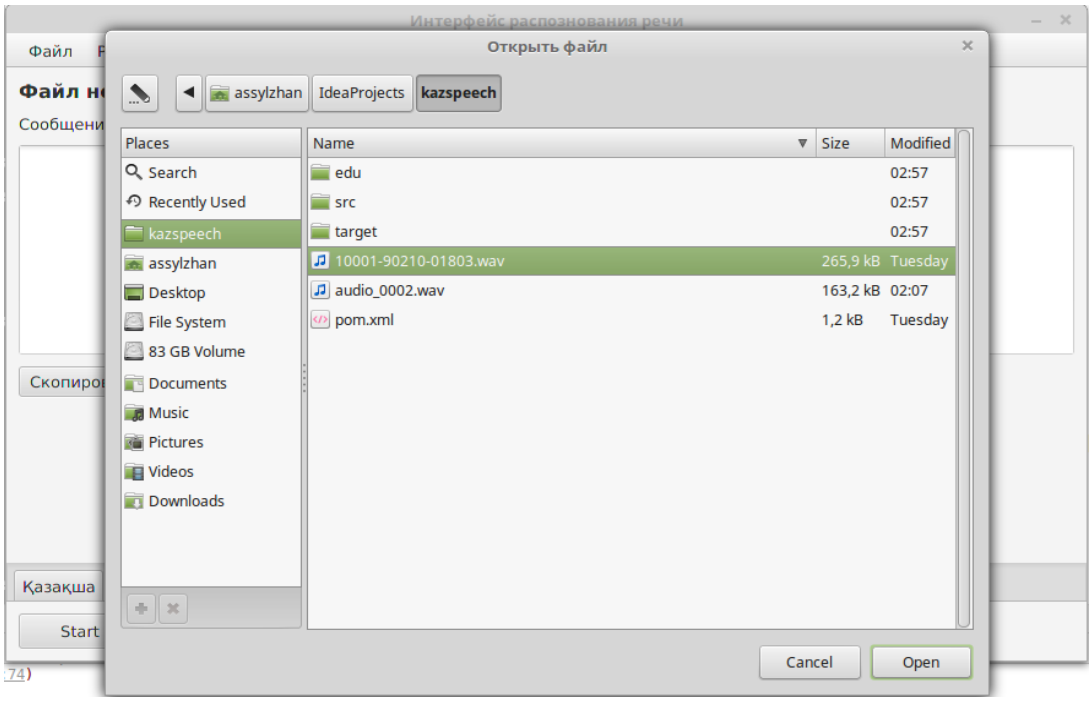
1) MS Windows XP/7/8, x86.
2) MS Windows XP/7/8, x64.
3) Linux Ubuntu 14.1, x86.
4) Linux Ubuntu 14.1, x64.

Бағдарламалық кодты компиляциялау Windows операциялық жүйесі үшін IntelliJ IDEA Community Edition 2018.2.3 құру ортасы қолданылды. Сонымен қоса Linux тектес операциялық жүйелері үшін де осы аталған құру ортасының Linux тектес операциялық жүйелеріне арналған нұсқасы қолданылды.

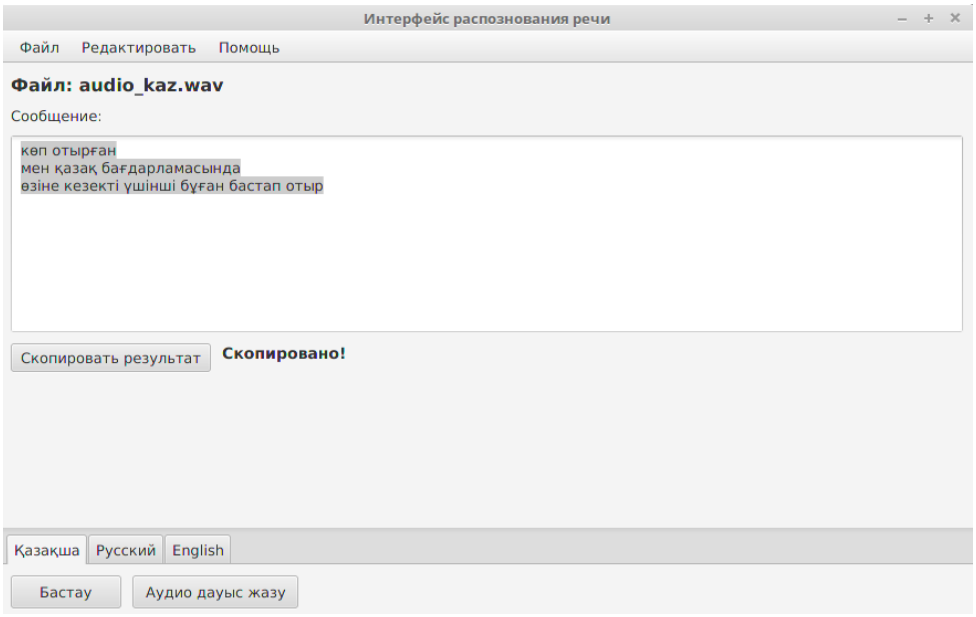
Windows операциялық жүйесінде жұмыс істеуге арналған кіріккен қазақ сөйлеуін автоматты тану жүйесі 3 түрлі режимде орындалады (4.4 – 4.6 - суреттер).



Сурет 4.4 – Кіріккен сөйлеуді тану жүйесінің негізгі беті



Сурет 4.5 – Кіріккен сөйлеуді тану жүйесінде танылатын аудио жазбаны таңдау

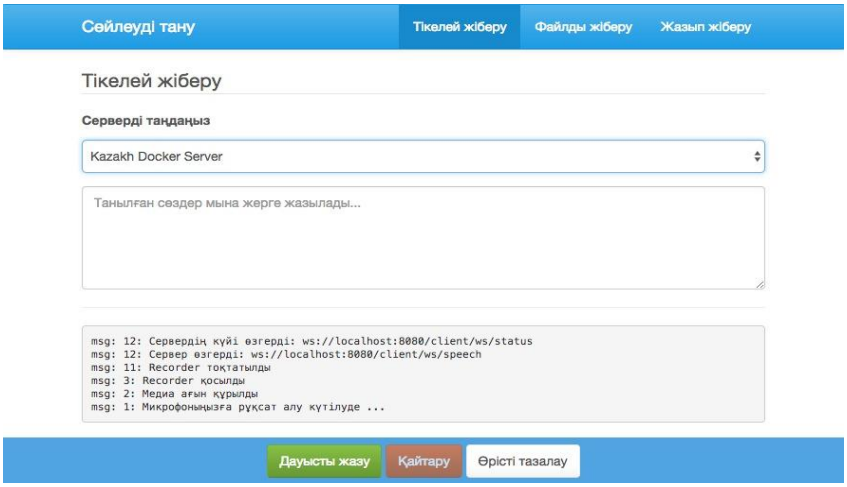


Сурет 4.6 – Кіріккен сөйлеуді тану жүйесінде тану нәтижесін алу

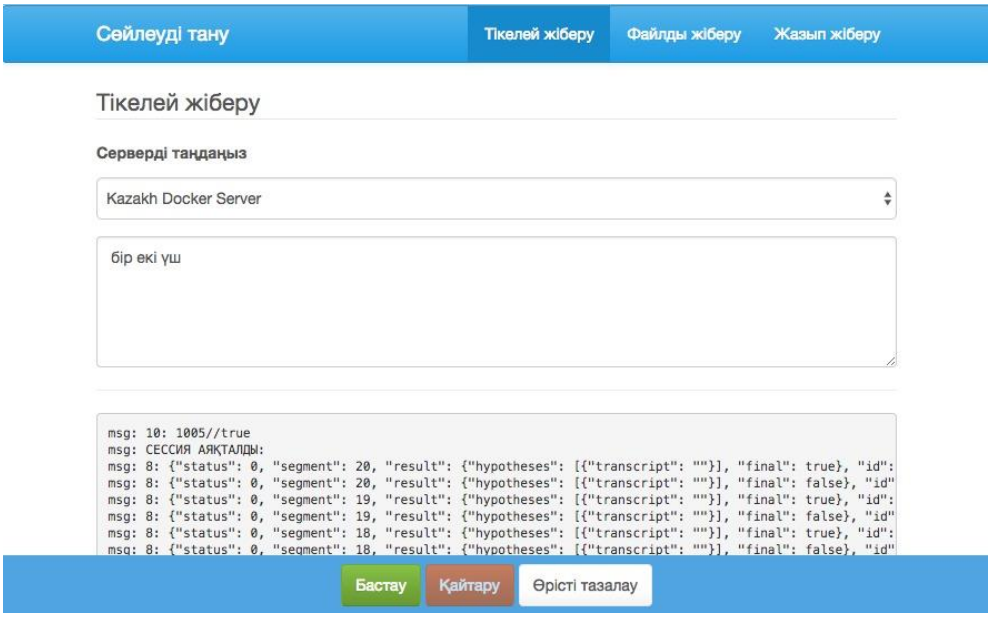
4.2.2 Қазақ сөйлеуін танудың Веб-қосымшасы

Қазақ сөйлеуін тану Веб-қосымшасында бірнеше веб технологиялардың көмегімен құрылған: HTML5, CSS3, JavaScript, Bootstrap 3, jQuery, Web Browser және т.б. Қосымшаны пайдаланар алдында браузерді ашып, веб беттің мекен жайын жазып, веб құжатты ашып алады. Браузерде ашылу барысында JavaScript арқылы қолданушы компьютерінің бірнеше құрылғыларына рұқсат алынады: микрофон, уақытша сақтау жадысы, web socket протоколы [110, 111]. Ол жерде дауысты жазып, оларды тануды жүзеге асыру үш түрлі жолмен жасауға болады:

1) Тікелей жазу барысында жіберу. Бұлай жіберу, батырманы қолданушы басқаннан бастап, оның микрофонын іске қосып дауысты жаза бастайды. Сол жазу барысында веб қосымша жазылған аудионы бөлік бөліктерін бір-бірлеп серверге жібере береді. Сервер қабылдаған ақпараттарды өңдеп, олардың нәтижелерін қолданушыға қайтадан жауап түрінде жібереді (4.7 – 4.8-суреттер).



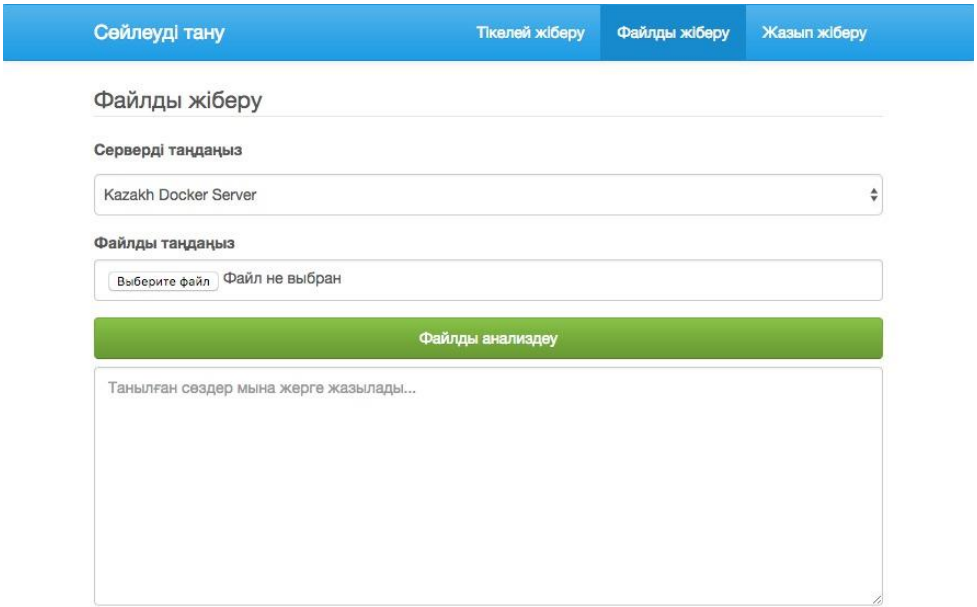
Сурет 4.7 – Қазақ сөйлеуін танудың Веб-қосымшасында сөйлеуді тікелей тану



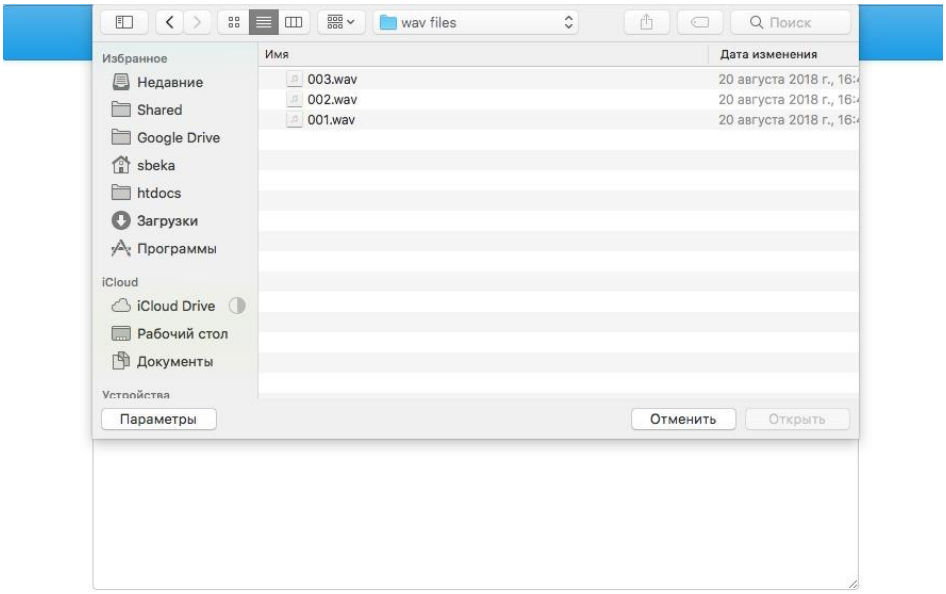
Сурет 4.8 – Қазақ сөйлеуін танудың Веб-қосымшасында сөйлеуді тікелей тану нәтижесі

2) Дауысты жазып алып жіберу. Бұл жолдың артықшылығы дауысты жазып алып, жіберу деген батырманы басқанда ғана жазылған аудионы серверге жібереді.

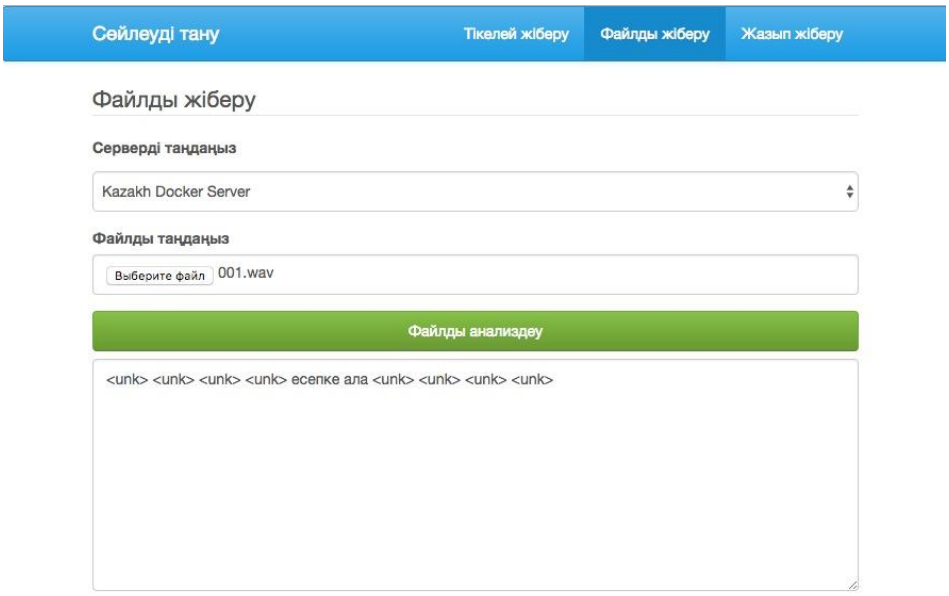
Қазақ сөйлеуін танудың Веб-қосымшасыда сөйлеуді алдын ала жазып, серверге жіберу арқылы тану (4.9 – 4.11-суреттер).



Сурет 4.9 – Қазақ сөйлеуін танудың Веб-қосымшасында сөйлеуді аудио файлдан тану



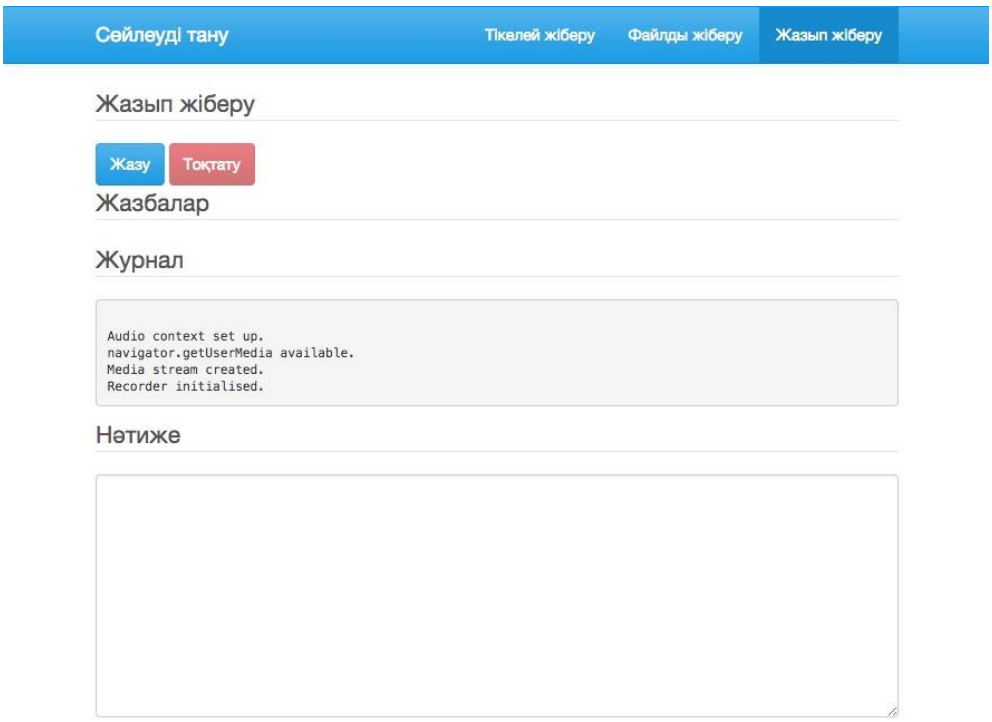
Сурет 4.10 – Қазақ сөйлеуін танудың Веб-қосымшасында аудио файлды таңдау



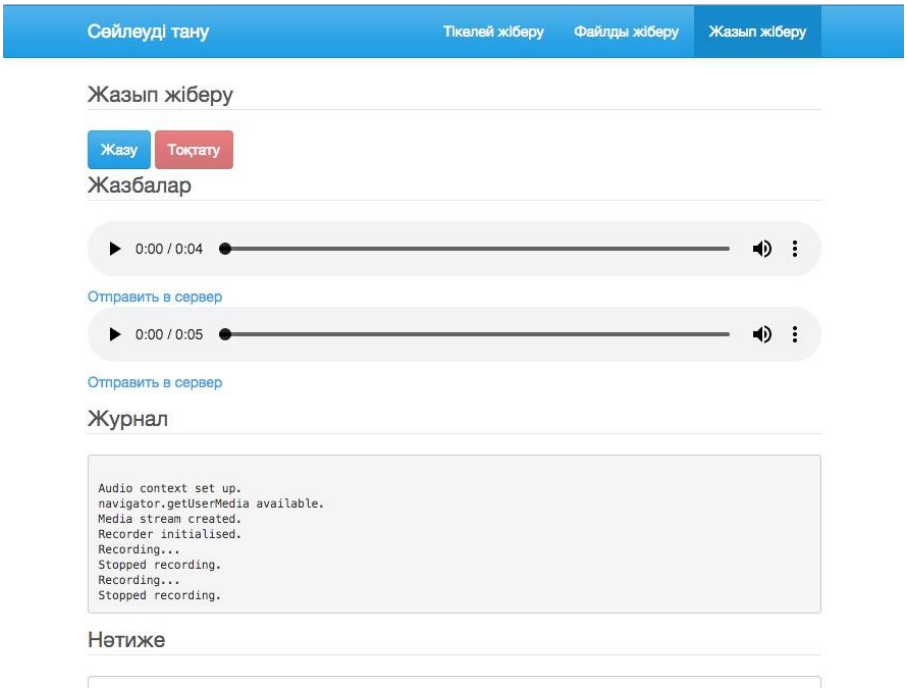
Сурет 4.11 – Қазақ сөйлеуін танудың Веб-қосымшасында аудио файлды талдау және нәтижесін алу

3. Алдын ала жазылған аудио файлды таңдап жіберу. Бұл жіберу жолы алдын ала жазылып қойылған аудиоларды жіберуге өте ыңғайлы. Яғни таңдау деген батырманы басу арқылы .mp3 немесе .wav деген аудио файлды таңдап алып, жіберу деген батырманы басқанда тану процессі жүзеге асады

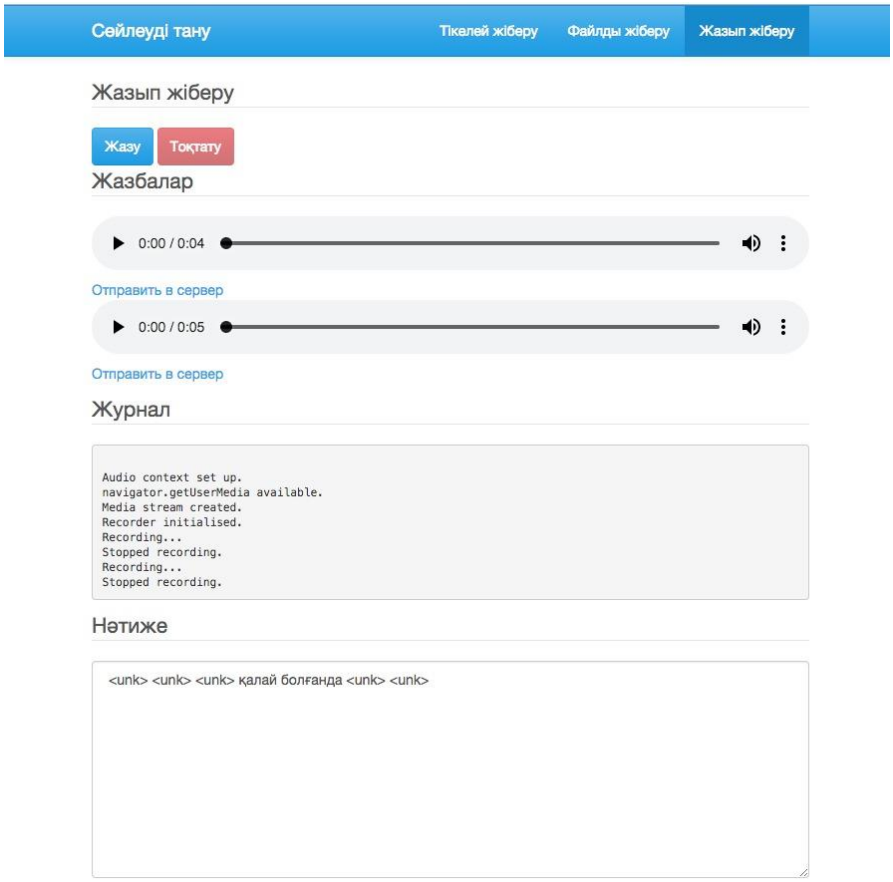
Қазақ сөйлеуін танудың Веб-қосымшасында сөйлеуді аудио файлдан тану (4.12 – 4.14-суреттер).



Сурет 4.12 – Сөйлеуді алдын ала жазу арқылы тану беті



Сурет 4.13 – Алдын ала жазылған сөйлеу файлын серверге жіберу және нәтижесін алу

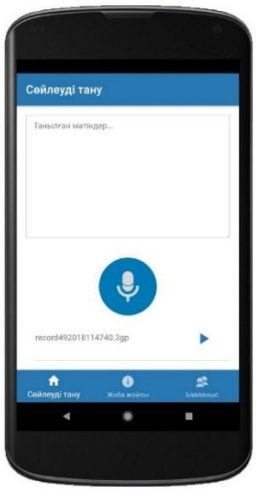


Сурет 4.14 – Сөйлеуді алдын ала жазу

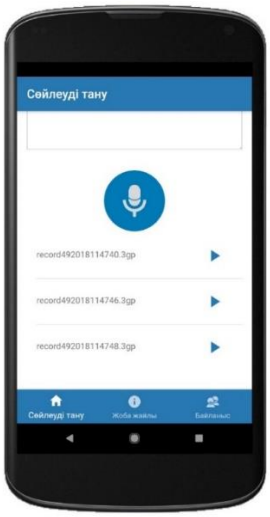
4.2.3 Қазақ сөйлеуін танудың мобильдік қосымшасы

Мобильді қосымша сайт секілді ұқсас жұмысты атқарады. Айырмашылығы бұл жерде басқа бағдарламалау тілдері мен басқа плагиндер қолданылады. Мысалға, дауысты жазу үшін мобильді телефонның микрофоны рұқсат алу, жедел жадысына рұқсат алу дегендей. Бұл рұқсат алуларды Cordova деген платформа арқылы жүзеге асырамыз. Интерфейс ionic деген фреймворктің көмегімен қолданушымен қарым-қатынас жасап, керекті әрекеттерді орындап, websocket арқылы сервермен байланыс жасап отырамыз. Cordova платформасының артықшылығы осы жазылған жобаны бірнеше мобильді платформада ашуға мүмкіндік береді: Android, iOS, Windows Phone, BlackBerry және т.б.

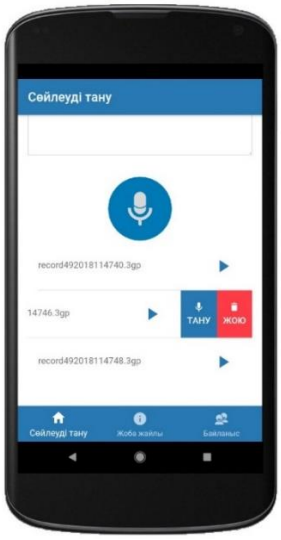
Құрылған мобильдік қосымшасының функционалдық қызметтері 4.15 – 4.18 суреттерде көрсетілген.



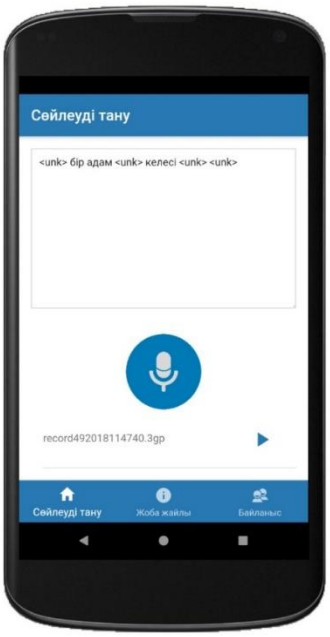
Сурет 4.15 – Сөйлеуді тану жүйесінің мобильді қосымшасының негізгі беті



Сурет 4.16 – Сөйлеуді тану жүйесінің мобильді қосымшасында сөйлеуді жазу



Сурет 4.17 – Сөйлеуді тану жүйесінің мобильді қосымшасында сөйлеуді тану



Сурет 4.18 – Сөйлеуді тану жүйесінің мобильді қосымшасында сөйлеуді тану нәтижесі

ҚОРЫТЫНДЫ

Бұл жұмыста кіріккен сөйлеуді толассыз тану жүйелерінің өзекті есептері қарастырылды. Орындалған жұмыс нәтижесінде алынған алгоритмдер мен модельдер қазақ тіліндегі кіріккен сөйлеуді толассыз тану жүйесінің сапасын арттыруға мүмкіндік берді.

Диссертациялық жұмысты жазу барысында қойылған тапсырмалар шешу кезінде келесідей нәтижелерге қол жетілді:

- Қазақ тілінің сөйлеу корпусы құрылды;
- Қазақ тілі сөйлеуін автоматты тану жүйелеріне арналған үлкен көлемді сөздігі, сөздікті автоматты түрде құратын бағдарлама құрылды. Оның алгоритмінің сипаттамасы құрылып және негізделді;
- Терең рекурренттік нейрожелілерді Nvidia CUDA-ны қолданып, қазақ тілі сөйлеу корпусымен бірнеше нәтижелер алынды;
- Кіріккен сөйлеуді толассыз танитын жүйенің архитектурасы құрылды;
- Терең рекурренттік нейрондық желілер модельдері негізінде кіріккен қазақ тілі сөйлеуін толассыз танитын бағдарламалық кешені құрылды;
- Терең рекурренттік нейрондық желілер модельдері негізінде кіріккен қазақ тілі сөйлеуін толассыз танитын Веб-қосымшасы құрылды;
- Терең рекурренттік нейрондық желілер модельдері негізінде қазақ тілі сөйлеуін толассыз танитын мобильді қосымшасы құрылды.

Құрылған акустикалық және тілдік модельдерді болашақта қазақ тіліндегі кіріккен сөйеулерді автоматты тану жүйелерін құру жұмыстарында пайдалануға болады.

ПАЙДАЛАНҒАН ӘДЕБИЕТТЕР ТІЗІМІ

1 Бондарко, Л. В. Спонтанная речь и организация системы языка // Бюллетень фонетического фонда № 8 «Фонетические свойства русской спонтанной речи». – 2001. – С. 17 – 23.

2 Godfrey, J. Switchboard: telephone speech corpus for research and development [Text] / J. Godfrey, E. Holliman, J. McDaniel // Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). - 1992. -Vol. 1. - P. 517-520.

3 Godfrey, J. Switchboard-1 Release 2 LDC97S62. Web Download [Electronic resource]. - Philadelphia : Linguistic Data Consortium, 1993. - URL: https: //catalog.ldc.upenn.edu/LDC97S62 (online; accessed: 20.11.2017).

4 Cieri, C. Fisher English Training Speech Part 1 Speech LDC2004S13. DVD [Electronic resource]. -Philadelphia : Linguistic Data Consortium, 2004. - URL: https://catalog.ldc.upenn.edu/LDC2004S13 (online; accessed: 20.11.2017).

5 Cieri, C. Fisher English Training Speech Part 2, Speech LDC2005S13. DVD [Electronic resource]. – Philadelphia : Linguistic Data Consortium, 2005. – URL: https://catalog.ldc.upenn.edu/LDC2005S13 (online; accessed: 20.11.2017).

6 Linguistic Data Consortium [Electronic resource]. -2016. -URL: https://www.ldc.upenn.edu/ (online; accessed: 20.11.2017).

7 2000 HUB5 English Evaluation Speech LDC2002S09. Web Download [Electronic resource]. - Philadelphia : Linguistic Data Consortium, 2002. - URL: https://catalog.ldc.upenn.edu/LDC2002S09 (online; accessed: 20.11.2017).

8 Санкт-Петербургский институт информатики и автоматизации Российской академии наук [Электронный ресурс]. – 2016. – URL: http://www.spiiras.nw.ru/ (дата обращения: 20.11.2017).

9 ООО «ЦРТ» [Электронный ресурс]. - 2016. - URL: http://www.speechpro.ru/ (online; accessed: 20.11.2017).

10 Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur [Electronic resource]. - 2016. - URL: https://www.limsi.fr/en/ (online; accessed: 20.11.2017).

11 Шарипбаев А.А., Бекманова Г.Т., Алтынбек Г., Адалы Е., Жеткенбай Л., Каманур У. Единый морфологический анализатор для казахского и турецкого языков // Turklang. -2017. – Р. 234 - 240.

12 А.А. Шарипбаев, Г.Т. Бекманова. Somequestions of the automatic transcription of kazakh language. The module of transcription of kazakh speech recognition system // Труды II Международной научно-практической конференции Информатизация общества. – Астана, 2010. – С. 543-551.

13 Калимолдаев М.Н., Амиргалиев Е.А., Мусабаев Р.Р., Мамырбаев О.Ж. Методы формирования словаря визем для многомодального распознавания речи // Материали за 10-а международна научна практична конференция, «Ключови въпроси в съвременната наука», Съвременни технологии на информации. – София: «Бял ГРАД-БГ» ООД, 2014. – Т. 35. – С. 60-65.

14 Калимолдаев М.Н., Мамырбаев О.Ж., Мусабаев Р.Р., Оразбеков Ж.Н. Автоматическая сегментация речи с использованием информации о средней

частоте пересечения уровней // Вестник КазНТУ им. К.И. Сатпаева. – Алматы, 2013. – №6 (100). – С. 249-257.

15 Калимолдаев М.Н., Keylan Alimhan, Мусабаев Р.Р., Мамырбаев О.Ж. Один из методов сегментации речевого сигнала на слоги // Материалы IV международной конференции «Математическое моделирование и информационные технологии в образовании и науке (ММИТОН)», посвященной 85-летию Казахского национального педагогического университета имени Абая. – Алматы, 2013. – С. 176-181.

16 Калимолдаев М.Н., Мамырбаев О.Ж., Мусабаев Р.Р., Оразбеков Ж.Н. Сегментация и обработка речевого сигнала с использованием алгоритма средней частоты пересечения уровней // Проблемы информатики. – 2014. – №1 (22). – С. 73-82.

17 Калимолдаев М.Н., Мамырбаев О.Ж., Мусабаев Р.Р. Метод модуляции речевого сигнала и его применение в системах речевой обработки // Проблемы информатики. – 2012. – №1 (13). – С. 4-10.

18 Dahlback N., Jonsson A., Ahrenberg L. Wizard of Oz Studies Why and How // Knowledge-Based Systems. – 1993. – Vol. 6, № 4. – P. 258-266.

19 Ронжин А.Л., Леонтьева А.Б. Применение техники «Гудвин» для моделирования человеко-машинного взаимодействия // Известия вузов. Приборостроение. – 2006. – Т. 49, № 11. – С. 70-75.

20 Охтилев М. Ю. Особенности технологии проектирования специализированных систем для автоматизированного анализа измерительной информации реального времени // Автоматика и вычислительная техника. – 2001, № 6. – С. 73–79.

21 Foote J. T., Jones G. J. F., Sparck J. K., Young S. J. Talker-independent keyword spotting for information retrieval // In Proc. Eurospeech 95. – 1995. – Vol. 3. – P. 2145-2148.

22 Сорокин В.Н., Цыплихин А.И. Сегментация и распознавание гласных // Информационные процессы. – 2004. – Т. 4, № 2. – С. 202-220.

23 Гойхман О.Я., Надеина Т.М. Речевая коммуникация // Инфра-М. – 2006. – 272 с.

24 Ушакова Т.Н. Проблема внутренней речи в психологии и психофизиологии // Психологические и психофизиологические исследования речи. – М.: Наука, 1985. – С. 13-26.

25 Kurematsu A., Nakano-Miyatake M., Perez-Meana H., Simancas-Acevedo E. Performance analysis of Gaussian Mixture Model speaker recognition system with different speaker features // Technical Acoustics. – 2005. – Vol. 14. – 16 p.

26 Benzeghiba M., De Mori R., Deroo O., Dupont S., Jouvet D., Fissore L., Laface P., Mertins A., Ris C., Rose R., Tyagi V., Wellekens C.. Impact of variabilities on speech recognition // In Proc. of 11-th International Conference SPECOM’2006, St. Petersburg: «Anatoliya». – 2006. – P. 493-498.

27 Александрова О.А., Иваницкий В.В. Пауза колебания – комплексный феномен современной коммуникации // Вестник Новгородского государственного университета. – Т. 25. – 2003. – С. 95–101.

28 Benesty J., Sondhi M., Huang Y. (eds.) Springer Handbook of Speech Processing. – Berlin: Springer, 2008. –1176 p.

29 Whittaker E.W.D., Woodland P.C. Efficient class-based language modelling for very large vocabularies // in Proc. ICASSP'01 Conference, Salt Lake City, USA. – 2001. – P. 545-548.

30 Vesa S., Teemu H., Mathias C., Mikko K. Unlimited vocabulary speech recognition based on morphs discovered in an unsupervised manner // Proceedings of Eurospeech. – 2003. – P. 2293–2296.

31 Косарев Ю.А. Естественная форма диалога с ЭВМ. – Л.: Машиностроение, 1989. – 143 с.

32 Cole et al. Survey of the state of the art in human language technology, Cambridge University Press, New York, NY. – 1997. –533 p.

33 Нариньяни А.С. Искусственный интеллект: стагнация или новая перспектива? // Сборник трудов VI национальной конференции по искусственному интеллекту с международным участием.- Пущино: РАИИ, 1998. - Т.1. – С.15-29.

34 Пинкер С. Язык как инстинкт / пер. с англ. ; под ред. В.Д. Мазо. – М.: УРСС, 2004. – 456 с.

35 Попов Э.В. Общение с ЭВМ на естественном языке. – М.: УРСС, 2004. – 360 с.

36 Нариньяни А.С. ТЕОН-2: от Тезауруса к Онтологии и обратно // Международный семинар по компьютерной лингвистике и ее приложениям. «Диалог 2002».- 2002. – Т. 1. – С. 149-154.

37 Скороходько Э.Ф. Семантические сети и автоматическая обработка текста. – Киев, 1983. – 112 с.

38 Смирнов А.В., Пашкин М.П., Шилов Н.Г., Левашова Т.В. Онтологии в системах искусственного интеллекта: способы построения и организации. Новости искусственного интеллекта. – 2002. – № 1. – Ч. 1. – С. 3-13.

39 Сокирко А. Семантические словари в автоматической обработке текста (По материалам системы ДИАЛИНГ): диссертация … канд.т.наук: 05.13.17. – Москва, 2001. – 120 с.: ил. РГБ ОД, 61 01-5/2234-Х.

40 Мальковский М.Г. Диалог с системой искусственного интеллекта. -М.: Изд-во МГУ, 1985. - 214 с.

41 Сулейманов Д.Ш. Аналитический обзор отечественных и зарубежных работ обработки естественного языка в аспекте прагматически-ориентированного подхода. Электронный журнал Казанского госуниверситета «Информационные технологии». www.kcn.ru/tat_en/science/fttc/vol000/st.doc (дата обращения: 22.11.2018)

42 Шаров С.А. Средства компьютерного представления лингвистической информации. Обзор. URL: http://www.ksu.ru/eng/science/ittc/vol000/002/ (дата обращения: 23.11.2018)

43 Karttunen L. Radical lexicalism. Alternative conceptions of phrase structure. – Chicago: University of Chicago Press, 1989. – P. 43-65.

44 Карпов А.А., Кипяткова И.С., Ронжин А.Л. Проектирование речевых интерфейсов для информационно-управляющих систем: учеб. пособие. – СПб: ГУАП, 2012. – 76 с.

45 Pietquin O. A framework for unsupervised learning of dialogue strategies. – Louvain-la-Neuve: Presses universitaires de Louvain, 2004. – P. 246.

46 Boite R., Bourlard H., Dutoit T., Hancq J., Leich H. Traitement de la Parole. –2nd Edition. – Lausanne: Presses Polytechnique Universitaires Romandes, 2000. – 488 p.

47 Голд Б. Рэйдер Ч. Цифровая обработка сигналов / пер. с англ. – М.: Советское радио, 1973. – 368 с.

48 Макхоул Дж., Рукос С., Гиш Г. Векторное квантование при кодировании речи. // ТИЭР. – 1985. – Т. 73. – №11. – С. 19-61.

49 Маркел Д.Д., Грей А.Х. Линейное предсказание речи: Пер. с англ. – М.: Связь, 1980. – 308 с.

50 Златоустова Л.В., Потапова Р.К., Трунин-Донской В.Н. Общая и прикладная фонетика. М.: Издательство МГУ, 1986. – 304с.

51 Пиотровский Р.Г. Текст, машина, человек. – Л.: Наука, 1975. – 327 с.

52 Chomsky N. On certain formal properties of grammars. - Inform. Control 2, 1959. – P. 137-167.

53 Allen J. Natural Language Understanding. Benjamin Cummings, 1987. – Second Edition. – 1994. – 654 p.

54 Rabiner L., Juang B. Fundamentals of speech Recognition. - Englewood Cliffs, N.J.: PTR Prentice Hall, 1993. – P. 409 – 428.

55 Прохоров А.М. (гл. ред.) Большая советская энциклопедия. Т. 23. М.: Советская энциклопедия, 1976, 638 с.

56 Brandstein M., Ward D. Microphone Arrays Signal Processing Techniques and APlications. – New York: Springer-Verlag; Berlin Heidelberg, 2001. - 398 p.

57 Нуссбаумер Г. Быстрое преобразование Фурье и алгоритмы вычисления сверток. – М.: Радио и связь, 1985. – 248 с.

58 Elman J. Finding structure in time // COGNITIVE SCIENCE. -1990. -Vol. 14, no. 2.-Pp. 179–211.

59 Cruse H. Neural Networks as Cybernetic Systems - Part II // Brains, Minds, and Media. -2006. -no. 1.

60 Hopfield J. Neural networks and physical systems with emergent collective computational abilities // Proceedings of the National Academy of Sciences of the United States of America. - 1982. - Vol. 79, no. 8. - Pp. 2554-2558.

61 R. Pascanu, ¸C. G¨ul¸cehre, K. Cho, Y. Bengio. How to Construct Deep Recurrent Neural Networks. / // CoRR. -2013. -Vol. abs/1312.6026.

62 Bishop C. Pattern Recognition and Machine Learning (Information Science and Statistics). - Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

63 Pascanu R. On Recurrent and Deep Neural Networks: Ph.D. thesis / Universit´e de Montreal.

64 Williams R., Peng J. An Efficient Gradient-Based Algorithm for On-Line Training of Recurrent Network Trajectories // Neural Computation. -1990. -Vol. 2. -Pp. 490–501.

65 . Bengio Y., Simard P., Frasconi P. Learning Long-term Dependencies with Gradient Descent is Difficult // Trans. Neur. Netw. -1994. -Vol. 5, no. 2. -Pp. 157–166.

66 Recurrent neural network based language model / T. Mikolov, M. Karafi´at, L. Burget et al. // INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September 26-30, 2010. -2010. -Pp. 1045–1048.

67 Mikolov T. Statistical Language Models Based on Neural Networks: Ph.D. thesis. -2012. -P. 129. http://www.fit.vutbr.cz/research/view_pub.php?id=10158.

68 P. Price, W. Fisher, J. Bernstein, and D. Pallett, "The DARPA 1000-Word Resource Management Database for Continuous Speech Recognition," ICASSP 88, New York, April 1988.

69 M. Liberman, "Text on Tap: the ACL/DCI," Proceedings October, 1989 DARPA Speech and Natural Language Workshop, Morgan Kanfmann Publishers, October, 1989.

70 P. Price, The ATIS Common Task: Selection and Overview," Proceedings June 1990 Speech and NaturM Language Workshop, Morgan Kaufmann Publishers, June, 1990.

71 Кривнова О.Ф., Захаров Л.М., Строкин Г.С. Речевые корпусы (опыт разработки и использование) // Труды семинара Диалог'2001 по компьютерной лингвистике и ее приложениям. М., , 2001.

72 Есенбаев Ж.А. Распознавание казахской речи по определенной словарной базе в условиях шумов. Диссертация. Астана, 2014

73 R. Gordon, Ed., Ethnologue: Languages of the World.Dallas: SIL International, 2005.

74 Кривнова О. Ф. Длительность как средство реализации словесного ударения в тексте (сопоставительный анализ разных способов оценки выраженности ударения в слове) // Язык и речь: проблемы и решения. Сборник научных трудов к юбилею проф. Л.В.Златоустовой, – М. – 2004

75 О.Ж. Мамырбаев, М. Турдалыулы, 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.

76 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 XLII Халықаралық ғылыми-практикалық конференцияның материалдары. – Алматы, 2018. – Т. 2. – Б. 81-87.

77 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.

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

79 Базарбаева З. Қазақ фонологиясының негіздері. – Алматы: Мемлекеттік тілді дамыту институты, 2012. - 120 б.

80 Мырзабеков С. Қазақ тілі дыбыс жүйесіндегі дистриубуциялық пен функционалдық. Филология ғ.д.ғылыми дәрежесін алу үшін жазылған диссертация. Алматы, 1998 ж.

81 Жүнісбеков Ә. Қазақ фонетикасы. – Алматы: «Арыс», 2009.

82 Geoffrey Hinton, Li Deng, Dong Yu, George Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara Sainath, Brian Kingsbury. Deep Neural Networks for Acoustic Modeling in Speech Recognition // IEEE Signal Processing Magazine. Volume: 29, Issue: 6, Nov. 2012.

83 Access mode: http://kaldi-asr.org/doc/ free (date of the application 20.04.2018).

84 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.

85 Wikipedia. Word error rate // Access mode: https://en.wikipedia.org/wiki/Word_error_rate free (date of the application 20.04.2018).

86 Alexandre P., Lockwood P., Root cepstral analysis: a unified view: application to speech processing in car noise environments, Speech Communication, 12 (1993), 277-288

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

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

89 Acero A., Stern R. M., Robust speech recognition by normalization of the acoustic space, in IEEE International Conference on Acoustics, Speech, and Signal Processing., (1991), 893-896

90О.Ж. Мамырбаев, М. Калимолдаев, М. Турдалыулы, 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.

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

92 Колесникова Д.С., Рудниченко А.К., Верещагина Е.А., Фоминова Е.Р. Применение современных технологий распознавания речи при создании лингвистического тренажера для повышения уровня языковой компетенции в сфере межкультурной коммуникации // Интернет-журнал «НАУКОВЕДЕНИЕ» Том 9, №6 (2017)

93 Карпов А.А., Рожин А.Л., Ли И.В. SIRIUS – система дикторнезависимого распознавания слитной русской речи // Известия ТРТУ. - 2005. –№ 10. – С. 44-53.

94О.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.

95 Elman J.L. Finding Structure in Time. Cognitive Science. 1990. vol. 14. pp. 179–211.

96 Mikolov T. et al. Recurrent neural network based language model. Proceedings of INTERSPEECH'2010. 2010. pp. 1045–1048.

97 И.С. Кипяткова, А.А. Карпов. Разновидности глубоких искусственных нейронных сетей для систем распознавания речи // Труды СПИИРАН «Методы управления и обработки информации». 2016. - № 6 (49).

98 Schwenk H., Gauvain J.-L. Training Neural Network Language Models On Very Large Corpora. Proceedings of Conference on Empirical Methods on Natural Language Processing. 2005. pp. 201–208.

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

100 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.

101 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-летнему юбилею профессора Утепбергенова И.Т., Алматы, 2019

102 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

103 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

104 Tanel Alumae. Kaldi GStreamer server, 2017.

105 D.Povey and A. Ghoshal, The Kaldi Speech Recognition Toolkit, ASRU 2011.

106 Х.Э. Сильва. Docker Kaldi GStreamer Server. https://github.com/jcsilva/docker-kaldi-gstreamer-server (сілтеме жасалған күн: 10.05.2017).

107 Chen, Berlin. “An Introduction to the Kaldi Speech Recognition Toolkit.” National Taiwan Normal University, 2014.

108 Hinton, Geoffrey, et al. "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups." IEEE Signal Processing Magazine 29.6 (2012): 82-97.

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

110 Двоеглазов Д.В., Дешко И.П., Кряженков К.Г., Тихонов А.А. Применение технологии WebSocket в системе удаленного доступа к лабораторным стендам с инфокоммуникационным оборудованием // Интернет-журнал «НАУКОВЕДЕНИЕ».- 2015. - Т. 7. - №4.

111 The WebSocket Protocol // http://tools.ietf.org/html/rfc6455 (Сілтеме жасалған уақыт: 18.02.2019).

ҚОСЫМША А. Қосымшаға авторлық куәлік

Сөздікті құруға арналған қосымшаға алынған авторлық куәлік



ҚАЗАҚСТАН РЕСПУБЛИКАСЫ

РЕСПУБЛИКА КАЗАХСТАН

АВТОРЛЫҚ ҚҰҚЫҚПЕН ҚОРҒАЛАТЫН ОБЪЕКТІЛЕРГЕ ҚҰҚЫҚТАРДЫҢ МЕМЛЕКЕТТІК ТІЗІЛІМГЕ МӘЛІМЕТТЕРДІ ЕНГІЗУ ТУРАЛЫ

КУӘЛІК

2019 жылғы « 22 » қаңтар № 1425

Автордың (лардың) жөні , аты, әкесінің аты (егер ол жеке басын куәландыратын құжатта көрсетілсе):

МАМЫРБАЕВ ОРКЕН ЖУМАЖАНОВИЧ; ТҰРДАЛЫҰЛЫ МҰСА ; МЕКЕБАЕВ НҰРБАПА ОТАНОВИЧ; ТҰРДАЛЫҚЫЗЫ ТОЛҒАНАЙ ; СЕЙТҚАЛИ БЕКЖАН НҰРЛАНҰЛЫ ДУЙСЕНБАЕВА АЙГЕРИМ ЖАНБОЛАТОВНА

Авторлық құқық объектісі: ЭЕМ-ге арналған бағдарлама

Объектінің атауы: System of automatic creation of vocabulary for ASR

Объектіні жасаған күні: 01.06.2018

СВИДЕТЕЛЬСТВО

О ВНЕСЕНИИ СВЕДЕНИЙ В ГОСУДАРСТВЕННЫЙ РЕЕСТР ПРАВ НА ОБЪЕКТЫ, ОХРАНЯЕМЫЕ АВТОРСКИМ ПРАВОМ

№ 1425 от « 22 » января 2019 года

Фамилия, имя, отчество, (если оно указано в документе, удостоверяющем личность) автора (ов):

МАМЫРБАЕВ ОРКЕН ЖУМАЖАНОВИЧ; ТҰРДАЛЫҰЛЫ МҰСА ; МЕКЕБАЕВ НҰРБАПА ОТАНОВИЧ; ТҰРДАЛЫҚЫЗЫ ТОЛҒАНАЙ ; СЕЙТҚАЛИ БЕКЖАН НҰРЛАНҰЛЫ ДУЙСЕНБАЕВА АЙГЕРИМ ЖАНБОЛАТОВНА

Вид объекта авторского права: программа для ЭВМ

Название объекта: System of automatic creation of vocabulary for ASR

Дата создания объекта: 01.06.2018

Құжат түпнұсқалығын http://www.kazpatent.kz/ru сайтының "Авторлық құқық" бөлімінде тексеруге болады https://copyright.kazpatent.kz

Подлинность документа возможно проверить на сайте kazpatent.kz в разделе «Авторское право» https://copyright.kazpatent.kz

Подписано ЭЦП

Оспанов Е. К.

ҚОСЫМША Б. Мобильдік қосымшаның листингі

Мобильдік қосымшаның листингі

```
import { Component } from '@angular/core';
import {AlertController, LoadingController, NavController, Platform} from 'ionic-angular';
import { Media, MediaObject } from '@ionic-native/media';
import {File} from '@ionic-native/file';
@Component({
  selector: 'page-home',
  templateUrl: 'home.html'
})
export class HomePage {
  loader: any;
  //For record audio
  recording: boolean = false;
  filePath: string;
  fileName: string;
  audio: MediaObject;
  statusMic: string = 'micoff.png';
  audioList: any[] = [];
  btnEnable: boolean = true;
  tab: string = 'test';
  //For response text
  resultTexts: string = '';
  constructor(
    public navCtrl: NavController,
    private media: Media,
    private file: File,
    public platform: Platform,
    public loadingCtrl: LoadingController,
    public alertCtrl: AlertController
  ) {}
  private sendToWebsocket(blob: Blob) {


    this.presentLoading(); //show loader

    let socket = new WebSocket('ws://217.15.186.70:8080/client/ws/speech');
    // let socket = new WebSocket('ws://localhost:8080/client/ws/speech');

    socket.onopen = () => {
      console.log('Соединение готово отсылать и принимать данные!');
      socket.send(blob);
    };
    socket.onmessage = (e) => {
      let data = e.data;
      let res = JSON.parse(data);
```

```
    if (res.status == 0 && res.result) {
      let resText = res.result.hypotheses[0].transcript.toLowerCase();
      if (!res.result.final) {
        this.resultTexts += resText;
        console.log(resText);
      } else {
        socket.send('EOS');
        console.log(resText);
      }
    } else {
      this.showAlert('Серверлік қате', res.status + ': ' + res.message);
    }
  };
  socket.onclose = (e) => {
    this.loader.dismiss(); //hide loader
    this.btnEnable = true;
    this.showAlert('Байланыс жабылды!',  e.code + "/" + e.reason + "/" + e.wasClean);
  };
  socket.onerror = (e) => {
    this.loader.dismiss(); //hide loader
    this.showAlert('ҚАТЕ!', 'Socket серверімен байланыс кезінде қате болды.');
  };
}
ionViewWillEnter() {
  this.getAudioList();
}
public recognition(file: string) {
  this.btnEnable = false;
  this.resultTexts = '';
  fetch(file)
    .then(res => res.blob()) // Gets the response and returns it as a blob
    .then(blob => this.sendToWebsocket(blob));
}
public recogAudio(index: number) {
  this.file.checkFile(this.file.externalDataDirectory, this.audioList[index].filename)
    .then(res => {
      this.file.readAsDataURL(this.file.externalDataDirectory, this.audioList[index].filename)
        .then((res: any) => {
          let blob = this.dataURItoBlob(res);
          this.sendToWebsocket(blob);
        });
    })
    .catch(err => console.log('ERR CheckFile', err));
}
public getAudioList() {
  if(localStorage.getItem("audiolist")) {
    this.audioList = JSON.parse(localStorage.getItem("audiolist"));
  }
```

```
  }
  public startRecord() {
    this.statusMic = 'micon.gif';
    if (this.platform.is('ios')) {
      this.fileName = 'record'+new Date().getDate()+new Date().getMonth()+new
Date().getFullYear()+new Date().getHours()+new Date().getMinutes()+new
Date().getSeconds()+'.wav';
      this.filePath = this.file.documentsDirectory.replace(/file:\/\/g, '') + this.fileName;
      this.audio = this.media.create(this.filePath);
    } else if (this.platform.is('android')) {
      this.fileName = 'record'+new Date().getDate()+new Date().getMonth()+new
Date().getFullYear()+new Date().getHours()+new Date().getMinutes()+new
Date().getSeconds()+'.wav';
      this.filePath = this.file.externalDataDirectory.replace(/file:\/\/g, '') + this.fileName;
      this.audio = this.media.create(this.filePath);
    }
    this.audio.startRecord();
    this.recording = true;
  }
  public stopRecord() {
    this.statusMic = 'micoff.png';
    this.audio.stopRecord();
    let data = { filename: this.fileName };
    this.audioList.push(data);
    localStorage.setItem("audiolist", JSON.stringify(this.audioList));
    this.recording = false;
    this.getAudioList();
  }
  public playAudio(file, idx) {
    if (this.platform.is('ios'))
    {
      this.filePath = this.file.documentsDirectory.replace(/file:\/\/g, '') + file;
      this.audio = this.media.create(this.filePath);
    }
    else if (this.platform.is('android'))
    {
      this.filePath = this.file.externalDataDirectory.replace(/file:\/\/g, '') + file;
      this.audio = this.media.create(this.filePath);
    }
    this.audio.play();
    this.audio.setVolume(0.8);
  }
  public deleteAudio(i: number) {
    //--------ANDROID
    if (this.platform.is('android')) {
      this.file.checkFile(this.file.externalDataDirectory, this.audioList[i].filename)
        .then(a => {
          if (a) {
```

```
          this.file.removeFile(this.file.externalDataDirectory, this.audioList[i].filename)
            .catch(err => console.log('ERR RemoveFile', err));
        }
      })
      .catch(err => console.log('ERR CheckFile', err));
  //--------iOS
  } else if (this.platform.is('ios')) {
    this.file.checkFile(this.file.documentsDirectory, this.audioList[i].filename)
      .then(a => {
        if (a) {
          this.file.removeFile(this.file.documentsDirectory, this.audioList[i].filename)
            .catch(err => console.log('ERR RemoveFile', err));
        }
      })
      .catch(err => console.log('ERR CheckFile', err));
  }
  this.deleteFromAudioList(i);
 }
 private deleteFromAudioList(i: number) {
  this.audioList.splice(i, 1);
  localStorage.setItem("audiolist", JSON.stringify(this.audioList));
  this.getAudioList();
 }
 private dataURItoBlob(dataURI) {
  // convert base64 to raw binary data held in a string
  // doesn't handle URLEncoded DataURIs - see SO answer #6850276 for code that does
this
  let byteString = atob(dataURI.split(',')[1]);
  // separate out the mime component
  let mimeString = dataURI.split(',')[0].split(':')[1].split(';')[0]
  // write the bytes of the string to an ArrayBuffer
  let ab = new ArrayBuffer(byteString.length);
  // create a view into the buffer
  let ia = new Uint8Array(ab);
  // set the bytes of the buffer to the correct values
  for (let i = 0; i < byteString.length; i++) {
    ia[i] = byteString.charCodeAt(i);
  }
  // write the ArrayBuffer to a blob, and you're done
  let blob = new Blob([ab], {type: mimeString});
  return blob;
 }
 presentLoading() {
  this.loader = this.loadingCtrl.create({
    content: "Күте тұрыңыз..."
  });
  this.loader.present();
 }
```

```
  showAlert(title, msg) {
    const alert = this.alertCtrl.create({
      title: title,
      subTitle: msg,
      buttons: ['OK']
    });
    alert.present();
  }
  clearTextarea() {
    this.resultTexts = '';
  }
}
```