

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ
Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті
Бүркітбаев ат.Өнеркәсіптік автоматтандыру және цифрландыру институты
Электроника, телекоммуникация және ғарыштық технологиялар кафедрасы

ОӘК 621.3

Қолжазба нұсқасында

Раманкулова Ботакоз Полатовна

Магистрі академиялық дәрежесін іздену үшін

МАГИСТРЛІК ДИССЕРТАЦИЯ

Диссертация тақырыбы Сөйлеу сигналының фонемалық мәндерін тану
әдістерін зерттеу

Мамандық 6М071900 – «Радиотехника, электроника,
телекоммуникациялар»

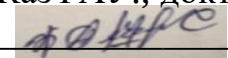
Ғылыми жетекші,
техн.ғыл.канд., ассис.профессор



Л.Б.Илипбаева

«01» шілде 2020 ж.

Сын-пікір беруші,
КазҰАУ., доктор PhD, қауым-н профессор

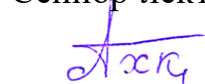


Әлібек Н.Б.

«03» шілде 2020 ж.

Нормобақылаушы:

Сениор-лектор, доктор PhD



А.Хабай

«01» шілде 2020 ж.

ҚОРҒАУҒА ЖІБЕРІЛДІ

ЭТЖҒТ каф.меңгерушісі

_____И.Сыргабаев

«____» _____2020 ж.

Алматы 2020

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ

Қ.И. Сәтбаев атындағы Қазақ ұлттық техникалық зерттеу университеті
Бүркітбаев ат.Өнеркәсіптік автоматтандыру және цифрландыру институты
Электроника, телекоммуникация және ғарыштық технологиялар кафедрасы
6М071900 – «Радиотехника, электроника, телекоммуникациялар»

БЕКІТЕМІН

ЭТжҒТ каф.меңгерушісі

_____И.Сыргабаев

«_____» _____2020 ж.

**магистерлік диссертация орындауға
ТАПСЫРМА**

Магистрант *Раманкулова Ботакөз Полатовна*

Тақырыбы: *«Сөйлеу сигналының фонемалық мәндерін тану әдістерін зерттеу»*

Университет ректорының 29.10.2018 ж. № 1194-М бұйрығымен
бекітілген.

Аяқталған жобаны тапсыру уақыты *«09» шілде 2020 ж.*

Магистерлік диссертацияға берілген алғашқы мәліметтер: *1) Алдын-ала жазылған дыбыстық файл; 2) Негізгі тонның периоды; 3) Сөйлеу сигналының параметрлері.*

Диссертацияда өңделетін сұрақтар:

а) Сөйлеудің фонемалық мәндерін танудың бар әдістеріне талдау жүргізу;

б) Қолданылатын әдістер мен алгоритмдерді жүйелеу;

в) Сөйлеуді танудың қолданыстағы жүйелерінің сипаттамаларына салыстырмалы талдау жүргізу;

г) Дауыстық басқару жүйесін бағдарламалық іске асыруда орындау.

Сызбалық материалдар тізімі (міндетті сызбалар дәл көрсетілуі тиіс):

а) Фонемалық талдау сұлбалары;

б) Дыбыс форманттары;

в) Дауысты дыбыстылар осциллограммасы;

г) Дыбыстық сигналдың спектограммасы.

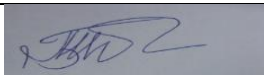
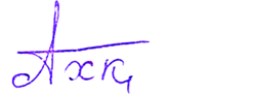
Ұсынылатын негізгі әдебиеттер:

1. Dr.Shaila, D.Apte.Speech and Audio Processing.Дели, 2012-438 б. 2. В.Н. Позднин, М.Г.Хохлов. Расчет формант по участку речевого сигнала. Технические науки, 2017, №1-6 б. 3. К.К. Томчук. Сегментация речевых сигналов для задач автоматической обработки речи. –Санкт-Петербург, 2017 –С.196.

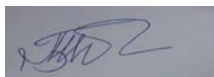
магистерлік диссертацияны дайындау
КЕСТЕСІ

Бөлімдер атауы, қарастырылатын мәселелер тізімі	Ғылыми жетекшіге және кеңесшілерге көрсету мерзімі	Ескерту
Тақырып бойынша аналитикалық шолу. Сөйлеу сигналы және оның сиппатамасы	01.09.2019 -13.12.2019	орындалды
Сөйлеудің фонема бойынша сигналдық ерекшеліктерін зерттеу	20.01.2020 -25.02.2020	орындалды
Қазақ тілі фонемдерін тануды бағдарламалық іске асыру	25.02.2020 – 24.04.2020	орындалды

Диссертациялық жұмыс бөлімдерінің кеңесшілері мен
норма бақылаушының аяқталған бөлімдерге қойған
қолтаңбалары

Бөлімдер атауы	Кеңесшілер (аты, әкесінің аты, тегі, ғылыми дәрежесі, атағы)	Қол қойылған күні	Қолы
Диссертация жұмысының тақырыбын талдау	Л.Б.Илипбаева, техн.ғыл.канд., ассис.профессор	25.02.2020 ж.	
Теориялық ақпарат	Л.Б.Илипбаева, техн.ғыл.канд., ассис.профессор	24.04.2020 ж.	
Норма бақылау	PhD докторы, ЭТЖҒТ каф.сениор-лекторы Хабай А.	01.07.2020 ж.	

Ғылыми жетекші
Л.Б.Илипбаева



Магистрант



Б.П.Раманкулова

Мерзімі

«01» шілде 2020 ж.

АҢДАТПА

Табиғи тілдік қарым-қатынас – адамзат үшін ең ыңғайлы қарым-қатынас құралы болып табылады. Сондықтан ғалымдар мен инженерлер көп жылдар бойы адам мен машинаның тілдік қарым-қатынас мәселелерін қарастыруда. Сөйлеу интерфейсінің негізгі міндеті адам тілін түсіну және оған дұрыс жауап беру. Ең негізгі мәселе – сөйлеудің дыбыстық сигналдарын тану алгоритмін ойлап табу. Өлі күнге дейін сөйлеуді тану жүйесінің дәлдігі 100 пайызға жетпеді. Бұл сөйлеуді танудың қиын мәселе екенін көрсетеді. Күрделілік айтарлықтай дәрежеде келесі факторлармен анықталады:

- сөйлеу сигналының түрлілігі мен тұрақсыздығы;
- микрофондардың амплитудалық-жиіліктік сипаттамаларының айырмашылығы;
- қоршаған орта мен аппаратурада шудың болуымен;
- сигналды өңдеуде көптеген деңгейлердің болуымен.

Сөйлеушінің сөзін тану адамның машиналармен өзара әрекеттесуінде маңызды орын алады. Бұл жұмыстың зерттеу аймағы – сөйлеудің фонемалық мәндерін тану аясында шолу және бағдарламалық жабдықтардың көмегімен сөйлеушінің сөздерін кіріс деректер ретінде қабылдауға мүмкіндік беретін әдістер мен жүйелер болып табылады. Жасалатын жұмыстың мақсаты сөйлеу сигналының фонемалық мәндерін тану жүйесінде қолданылатын әдістерін жалпылау және сөйлеуді тану алгоритмдерін зерттей отырып, оны бағдарламалық өнімде жүзеге асыру. Зерттеу нысаны – зерттеушінің сөйлеу сигналы. Зерттеу пәні – сөйлеуді ұсыну және тану модельдері мен әдістері. Зерттеу әдістері – ақпарат және кодтау әдістері, сөйлеу және сөйлеуді қабылдау теориясы, акустикалық фонетика, нысан ретіндегі сөйлеуді талдауға арналған жағдайдағы синтаксис. Зерттеу нәтижесінде сөйлеуді тануды бағдарламау арқылы зерттеушінің талабы орындалады.

АННОТАЦИЯ

Естественное речевое общение является самым удобным для человечества средством общения. Поэтому ученые и инженеры на протяжении многих лет рассматривают проблемы языкового общения человека и машины. Основной задачей речевого интерфейса является понимание языка человека и правильный ответ на него. Самая главная проблема-придумать алгоритм распознавания звуковых сигналов речи. До сих пор точность системы распознавания речи не достигла 100 процентов. Это указывает на то, что познание речи является сложной проблемой. Сложность в значительной степени определяется следующими факторами:

- разнообразие и неустойчивость речевого сигнала;
- разница амплитудно-частотных характеристик микрофонов;
- наличие шума в окружающей среде и аппаратуре;
- наличие большого количества уровней обработки сигнала.

Распознавание речи говорящего занимает важное место во взаимодействии человека с машинами. Область исследования этой работы – это методы и системы, позволяющие воспринимать речь говорящего в качестве входных данных с помощью программного обеспечения на фоне распознавания фонемных значений речи. Целью выполняемой работы является обобщение методов, применяемых в системе распознавания фонемных значений речевого сигнала и его реализация в программном продукте с изучением алгоритмов распознавания речи. Объект исследования-речевой сигнал исследователя. Предмет исследования-модели и методы представления и распознавания речи. Методы исследования-методы информации и кодирования, теория восприятия речи, акустическая фонетика, синтаксис. В результате исследования выполняются требования исследователя с программированием распознавания речи.

ANNOTATION

Natural speech communication is the most convenient means of communication for mankind. This is why scientists and engineers have been studying the problems of human-machine language communication for many years. The main task of the speech interface is to understand a person's language and answer it correctly. The most important problem is to come up with an algorithm for recognizing speech audio signals. So far, the accuracy of the speech recognition system has not reached 100 percent. This indicates that the perception of speech is a difficult problem. Complexity is largely determined by the following factors:

- diversity and instability of the speech signal;
- difference in the amplitude-frequency characteristics of microphones;
- presence of noise in the environment and equipment;
- availability of a large number of signal processing levels.

Speech recognition of the speaker occupies an important place in human interaction with machines. The research area of this work is methods and systems that allow us to perceive the speaker's speech as input data using software against the background of speech phonemic recognition. The purpose of this work is to generalize the methods used in the system for recognizing phonemic values of a speech signal and its implementation in a software product with the study of speech recognition algorithms. The object of research is the speech signal of the researcher. The subject of research is models and methods of speech representation and recognition. Research methods-methods of information and coding, theory of speech perception, acoustic phonetics, syntax. As a result of the study, the requirements of the researcher with speech recognition programming are met.

МАЗМҰНЫ

Кіріспе	8
1. Тақырып бойынша аналитикалық шолу	9
1.1 Сөйлеу сигналы және оның сипаттамасы	9
1.2 Адамда сөйлеудің қалыптасуы және қабылдануы. Қазақша сөйлеу тілі дыбыстарының (фонема) жіктелуі	12
1.3 Сөйлеудің фонемалық мәндерін тану ұғымы	15
1.4 Сөйлеуді тану жүйесінің тарихы	15
1.5 Сөйлеуді тану міндеттерін шешудің негізгі әдістері	17
1.6 Сөйлеуді тану жүйесінің және әдістерінің жіктелуі	18
1.7 Сөйлеуді автоматты танудың мақсаты, негізгі міндеті, жұмыс істеу принципі	21
1.8 Сөйлеуді тану кезеңдері	22
1.9 Matlab көмегімен сөйлеу сигналының талдауына кіріспе	25
2. Сөйлеудің фонема бойынша сигналдық ерекшеліктерін зерртеу	29
2.1 Сөйлеу сигналын өңдеуде Нейрон желілерін қолдану	29
2.2 Фонемалық дыбыстық сигнал кезінде өтпелі мәндерді талдау	33
2.3 Ағзаның жынысы мен құрылысына байланысты сөйлеу жүйесінің параметрлері	38
2.4 Қазақ дауысты дыбыстарының акустикалық мәндерін зерттеу	41
2.5 Сөйлеу сигналын құраушы фонемдердің форманттарын алу	44
2.6 Дауысты дыбыстар форманттарының шығыс қуыс, жұтқыншақ және ауыз резонатор ұзындығына тәуелділігі	47
3. Қазақ тілі фонемдерін тануды бағдарламалық іске асыру	50
3.1 Бағдарламаның жұмыс сипаттамасы	50
3.2 Visual Studio-дағы сөйлеу базасына байланысты жобаны іске асыру қадамдары	51
3.3 C# бағдарламалау тілі арқылы сөйлеу командасын орындау	53
3.4 Сөйлеу технологиясы заманауи нарықта	54
Қорытынды	57
Қысқартылған сөздер тізімі	58
Әдебиеттер тізімі	59

КІРІСПЕ

Сөйлеу - адам пайдаланатын негізгі және маңызды байланыс әдісі. Сөйлеу технологиялары жұмыс және зерттеу үдерістерін жетілдіруден бастап заманауи технологияларды дербес қолдану саласына дейін адамның автоматтандырылған ортасын дамытудағы негізгі фактор болып табылады.

Есептеу техникасы мен ақпараттық технологияларды дамытудың жоғары жылдамдығына қарамастан, сөйлеу қосымшаларының негізгі мәселелері әлі күнге дейін өзекті болып отыр. **Зерттеу тақырыбының өзектілігі** – сөйлеу сигналы құрылымының күрделілігі болып табылады: тілдің фонетикалық бірліктерінің, интонациялық бояулардың, сөйлеушінің тұлғалық ерекшеліктерінің алуан түрлілігі дауысты жазуға және беруге әсер ететін сыртқы факторлардың алуан түрлілігімен күрделене түседі. Нәтижесінде сөйлеу сигналдарын математикалық модельдердің көмегімен егжей-тегжейлі зерттеу және сипаттау өте қиын. Өте үлкен сөздікпен қазақ тілін тану жүйесінің болмауы соның көрсеткіші болып табылады.

Бұл жұмыстың мақсаты сөйлеу сигналының фонемалық мәндерін тану жүйесінде қолданылатын әдістерді және сөйлеуді тану алгоритмдерін зерттей отырып, оны бағдарламалық өнімде жүзеге асыру.

Осы мақсатқа жету үшін келесідей **тапсырмалар** орындалу керек:

- сөйлеудің фонемалық мәндерін танудың бар әдістеріне талдау жүргізу;
- қолданылатын әдістер мен алгоритмдерді жүйелеу;
- сөйлеуді танудың қолданыстағы жүйелерінің сипаттамаларына салыстырмалы талдау жүргізу;
- дауыстық басқару жүйесін бағдарламалық іске асыруда орындау қажет.

Жоғарыда баяндалғанды ескере отырып, жұмыста сөйлеу сигналдарын уақытша сегменттеу міндеттерін шешудің жолдары мен қолда бар тәсілдерді жетілдірудің жолдарын талдап және осы алгоритмдер бағдарланатын тілдің ерекшеліктерін анықтауға әрекет жасалады.

Диссертациялық жұмыстың ғылыми жаңалығы – қазақша сөйлеу фонемдерінің өтпелі аймақтарындағы өзгерістерді әртүрлі тілде айту бойынша сипаттамаларын алу арқылы зерттеу жүргізу.

Жүйені зерттеуде орындаушының алдын-ала жазылған дыбыстық файлдары қолданылады.

1 Тақырып бойынша аналитикалық шолу

Қазіргі таңда сөйлеуді тану жүйелеріне үлкен қызуғышулық танылуда. Ол бірінші кезекте ұялы құрылғылардың кең таралуымен байланысты. Осы құрылғыларға дауыстық басқаруды қолдану қолмен енгізу интерфейстерінің шектеулігіне байланысты, олардың мүмкіндіктерін кеңейтуге және пайдалануды жеңілдетуге алып келіп жатыр. Қолданыстағы тану технологияларының, әдетте, олардың бір немесе бірнеше параметрлері бойынша шектелуіне байланысты күрделенуде. Мысалы, пайдаланушылық сұраныстардың көпшілігін тануға мүмкіндік беретін жүйелердің негізгі үлесі ғаламтор желісіне қосылуды талап етеді (сұраныстарды өңдеу жүргізілетін қашықтағы серверлерге), өйткені олардың есептеуіш қуаты мен сөздікке бөлінген жады шектеулілігі жеткіліксіз. Осынша күрделі үдерістің бастауы неден алынатынын түсіну үшін тақырып бойынша ғылыми жұмыстарға талдау жүргізу міндеті қойылған. Осы жерде сөйлеуді тану (СТ) аясын шолу, сөйлеуді тану әдістері және кезеңдердің қысқаша сипаттамасы беріледі. Бұдан басқа, осы жұмыста белгілерді алудың төрт әдістерінің сипаттамасы келтіріледі: сызықтық болжаммен кодтау (СБК), мел-жиілік кепстраль (МЖК), салыстырмалы спектральды сүзгілеу және ықтималдық сызықтық ажырату талдауы (ЫСАТ).

1.1 Сөйлеу сигналы және оның сипаттамасы

Адамзат тілін сигнал түрінде ұсынуға болады. Сөйлеу сигналының көмегімен берілетін хабарлама – дискретті, яғни олардың соңғы санының символдарының кезектілігі түрінде ұсынылуы мүмкін. Сөйлеу сигналынан әр түрлі ықтималдық дәрежесімен біз ақпаратты ала аламыз: жынысы, жасы, сөйлеушінің денсаулығы туралы және тағы басқалары. Бір дыбысты әр түрлі адамдар әр түрлі айтады. Сөйлеу дыбыстарының айтылуы ықпал, көрші дыбыстар және тағы басқаларына байланысты. Бірақ әртүрлілікке қарамастан олардың айтылуында дыбыстар жалпылама сөйлеу дыбыстарының шектеулі санының физикалық реализациясы болып табылады.

Ақпаратты жеткізу үшін өткізуші орта ретінде ауа пайдаланылса, онда ауызша сөйлеу пайда болады – ол жиілігімен және амплитудасымен сипатталатын дыбыстық тербеліс. Сөйлеу физикалық табиғатта уақыт бойынша үздіксіз өзгертін акустикалық сигнал болып табылады. Осы сигналдың табиғатын ерекшелеу және басқа сигнал түрінен ажырату үшін техникалық әдебиеттерде сөйлеуді сөйлеу сигналы деп атайды.

Көптеген сигналдар (сөйлеу сигналдары да) табиғаты аналогты, сондықтан олар цифрлы компьютерлерде өңдеу кезінде аналогты-цифрлы түрлендіру (АЦТ) көмегімен дискретті сигналға түрленеді. Осы процедура көмегімен физикалық табиғатынан айырылған, ал олардың максималды және минималды мәндері АЦТ разрядтылығымен қойылатын үздіксіз сигналдың бірлік сәтінде түсірілген $s[n]$ есептеу жинағы алынады. Мысалы АЦТ

разрядтылығы 2 байтқа тең болса, онда есептеу мәндері $[2^{16-1}, 2^{16-1} - 1]$ аралығында болады. Онымен қоса түрлендірудің маңызды параметрі болып дискретизация жиілігі болып есептеледі. Ол үздіксіз сигналдың 1 секундта қанша жылдам мәндері сақталатынын анықтайды. Дискретизация жиілігі – дискретизация қадамына кері шама. Котельников теоремасы бойынша дискретті сигналдан шығынсыз тек осындай аналогты сигналды ғана қалпына келтіруге болады. Оның спектрінің жоғары f_h жиілігі f_s дискретизация жиілігінен 2 есе кіші

$$f_s > 2 \cdot f_h \quad (1.1)$$

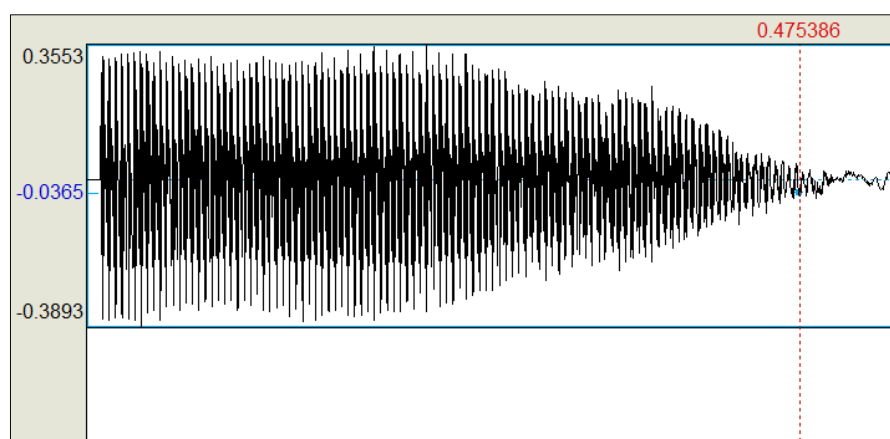
Дискретті сигналды сипаттау және түрлендіру үшін сигналдарды цифрлық өңдеу (СЦӨ) құралдары қолданылады. СЦӨ маңызды процедурасы Фурьенің дискретті түрленуі (ФДТ) болып табылады

$$S[m] = \sum_{n=1}^N s[n] \cdot e^{\frac{j2\pi nm}{N}}, m = 1, \dots, N \quad (1.2)$$

мұндағы N – ФДТ құрылатын есептеу саны;

j -жалған бірлік.

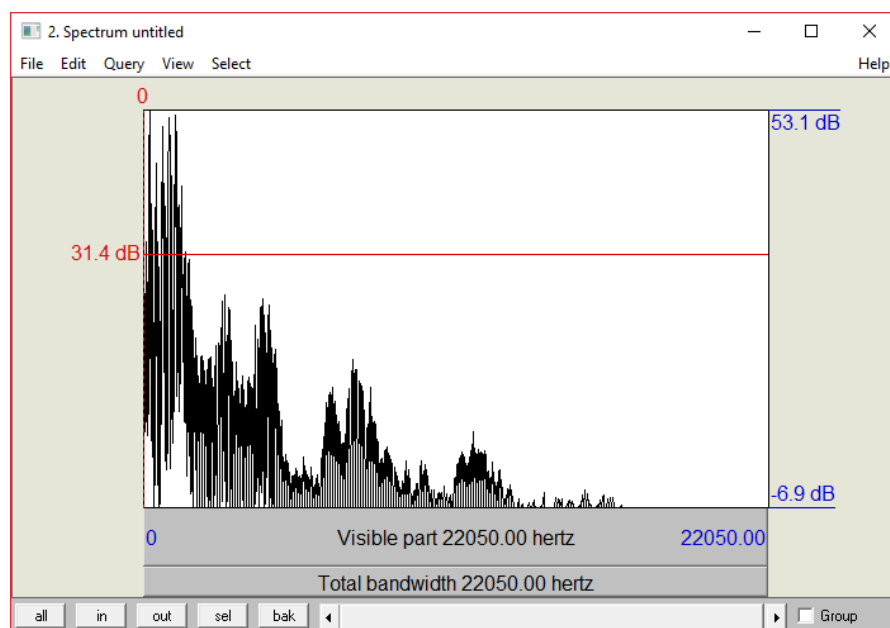
ФДТ уақыттық аймақтан жиіліктік аймаққа өту, яғни $s[n]$ – ді гармоника жинағына бөліп және гармоника амплитудасының (энергия) оның жиілігіне тәуелділігін табуға мүмкіндік береді. Сурет 1.1-де уақыттық аймақта дауысты «а» дыбысымен сөйлеу сигналының учаскесі берілген. Сонымен қатар АЦТ разрядтығынан абстрагирлеу үшін цифрланған сигналдың есептеуі салыстырмалы шамамен бейнеленеді: не максималды мәндегі үлеспен (2 байт үшін $s[n]/2^{16-1}$), не децибелмен. ФДТ табу үшін $N=1024$ есептеу шамасындағы учаске бөлінді.



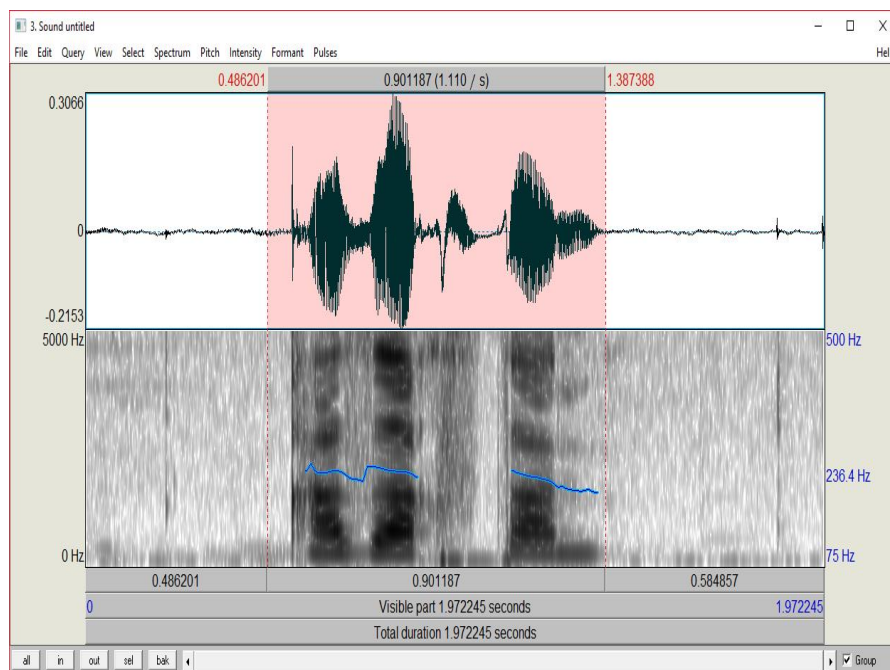
Сурет 1.1 – «а» дауысты дыбысымен сөйлеу учаскесі

Сөйлеу тұрақсыз сигнал болып табылады, яғни оның сипаттамалары уақыт бойынша өзгереді. Бұл өзгерістерді сөйлеу сигналының қатар келе

жатқан фрагменттер (фрейм) үшін ФДТ модуль графиктерін құру арқылы байқауға болады. Алынған бейне спектограмма деп аталады.



Сурет 1.2 – «а» дауысты дыбысымен сөйлеу сигналы учаскесінің ФДТ-і



Сурет 1.3 – «Қазақстан» сөзінің осцилограммасы және спектограммасы

Жазбаша сөйлеудегі сөздер символдардың соңғы жиынтығынан — тіл әліпбиінен құралатындай, сондай-ақ ауызша сөйлеу де оның барлық вариативтілігімен дыбыстық "әріптердің" шектеулі жиынтығымен құралады. Ең төменгі сөйлеу бірлігі болып фонема табылады.

1.2 Адамда сөйлеудің қалыптасуы және қабылдануы. Қазақша сөйлеу тілі дыбыстарының (фонема) жіктелуі

Дауысты құраушыларға бөлсек, негізгі параметрлерін аламыз: жиілік, күш, ұзақтығы және тембр. Оларды шамалар ретінде және жекелей талдауға болады. Шын мәнінде, алайда, мұндай талдау дауыстың нақты көрінісі болып табылмайды, өйткені бұл қасиеттер бірыңғай бөлінбейтін кешенді құрайды.

Шығарылатын дыбыстың биіктігі 1 секундта дауыс қатпарлары тербелісінің санына байланысты болады. Дауыстық қатпарлар тербелмелі қозғалыстарға толығымен ғана емес, өзінің барлық массасымен де, жеке учаскелерімен де келуге қабілетті. Тек сол бірдей дауыстық қатпарлар әр түрлі жиілікте тербелуі мүмкін: шамамен секундына 80-нен 10 000-ға дейін тербелістер және тіпті одан да көп.

Адам дауысының тондық диапазоны ең төменгі және ең жоғары дыбыстар арасындағы шекаралар шегінде дауыс аппаратымен жүргізілуі мүмкін. Адам дауысы әдетте 64-тен 1300 герц-ке дейінгі тондарды қамтиды. Адам дауысының екі түрлі көрінісінде – ән айту және сөйлесу – дауыс сапасы әр түрлі болады. Сөйлесу дауысы жалпы дауыс диапазонының 110 ғана құрайды. Әншілік дауыстың тондық қамтуы сөйлеу дауысына қарағанда кеңірек. Берілетін дыбыстың күші дауыстық қатпарлар кернеуінің қарқындығымен және байланыс кеңістігіндегі ауа қысымының шамасымен анықталады. Ол процесс те орталық жүйке жүйесімен реттеледі. Бақылау есту арқылы жүзеге асырылады. Егер осы процестер арасындағы өзара қарым-қатынас бұзылса, мысалы, қорқыныш айқайы кезінде, онда кеңірдектің ішіндегі қысымның басым болуы таза тоналдықтың болмауымен сипатталатын дыбыс тудырады.

Төменде адам дауысының ауқымы келтірілген (Герцте):

Бас-75-300;

Баритон-100-400-Тенор-120-500;

Контральто - 170-780;

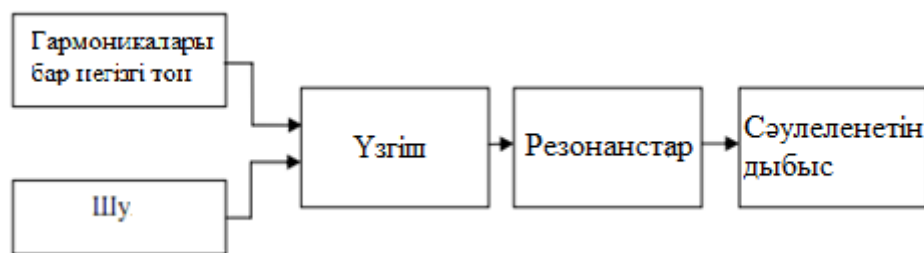
Меццо-сопрано - 200-900;

Сопрано-230-1000;

Колоратуралық сопрано-260-1400.

Осылайша, адам дауысы 75 - тен 110 Герц-ке дейін дыбыс диапазонын құрайды.

Физикалық тұрғыдан сөйлеу – топтары арасындағы кідірісі бар дыбыс тізбектілігінің жүйелігінен тұрады. Адамда сөйлеудің қалыптасуының сұлбасы сурет 1.4-те көрсетілген.



Сурет 1.4 – Адамда сөйлеудің қалыптасудың сұлбасы

Сөйлеу қарқыны қалыпты жағдайда үзіліс фразалардың арасында пайда болады. Әдетте, сөздер бірге айтылады, бірақ тыңдаушы сөздерді жеке қабылдайды. Сөйлеу жылдамдығы баяулаған кезде, мысалы, диктовка кезінде үзіліс сөздер мен тіпті олардың бөліктері арасында бөлінуі мүмкін. Сөйлеудің жиілік диапазоны 70-1400 Гц шегінде болады.

Сөйлеу дыбыстары дауысты және дауыссыз болып бөлінетіні белгілі. Дауыс байламдармен жасалатын ауа ағынының импульстары жеткілікті дәлдікпен периодты болып саналуы мүмкін. Импульстерді қайталаудың тиісті кезеңі T_0 – негізгі тонның периоды деп аталады. Кері шама – $1/T_0$ – негізгі тонның жиілігі. Барлық дауыстар үшін негізгі тонның жиілігі 70 - 450 Гц шегінде жатыр. Сөйлеу кезінде ол екпінге, дыбыстар мен сөздердің ерекшеліктерінде, сондай-ақ эмоциялар (сұраулы, лепті сөйлемді айтудағы, таңқаларлық жағдайда және т.б.) көрінісіне сәйкес үздіксіз өзгереді. Негізгі тонның жиілігін өзгерту интонация деп аталады. Әр адамның негізгі тонның өзгеру диапазоны (әдетте ол аз октавадан көп) және өзінің интонациясы бар. Соңғысы сөйлеушіні тану үшін өте маңызды.

Негізгі тоны, интонация, ауызша жазу және дауыс тембрі адамды тану үшін қызмет етеді, және нақтылық дәрежесі саусақ іздері сияқты жоғары. Негізгі тонның импульстері ара тәрізді формада болады, сондықтан олардың периодтық қайталануы кезінде жиілігі негізгі тонның жиілігіне еселенген гармоникалардың көп саны (40 дейін) бар дискреттік спектр алынады.

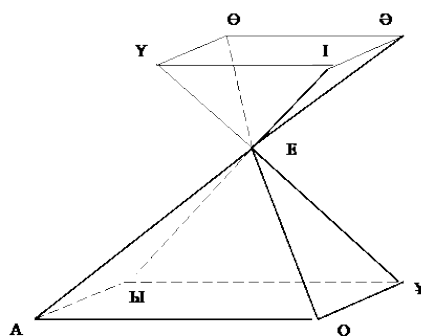
Дауыстың дыбыстары, әсіресе дауысты дыбыстарда қарқындылық деңгейі жоғары, дауыссызда төмен. Сөйлеу деңгейінің динамикалық диапазоны 35-45 дБ аралығында болады. Сөйлеудің дауысты дыбыстарының орташа ұзақтығы 0,15 с, дауыссыз - шамамен 0,08.

Хабарды телеграф және сөйлеу трактісі бойынша беру үшін әртүрлі өткізу қабілеті талап етілетіні белгілі. Телеграфтық хабарлама үшін 100 бит/с-тен артық емес өткізу қабілеті жеткілікті, ал сөйлеу хабарламасы үшін - 100000 бит/с (жолақ 7000 Гц-ке тең, динамикалық диапазон 42 дБ, яғни 7 санды код қажет, сонда $2 \cdot 7000 \cdot 7 = 98\ 000$ бит/с), яғни 100 есе үлкен.

Сөйлеу дыбыстарының пайда болуы мидың сөйлеу ортасынан сөйлеудің артикуляциялық ағзаларының бұлшық етке командаларды беру арқылы жүреді. Одан келетін жалпы хабарламалар ағыны шамамен 100 бит/с. Барлық қалған ақпарат сөйлеу сигналында ілеспе ақпарат деп аталады.

Қазақ тілі теориясының негізін қалаушы Ахмет Байтұрсынов XX ғасырдың басында қазақ тілінің 28 аса көрнекті дыбысын бөлді, оның ішінде 9 дауысты және 19 дауыссыз. 1929 жылы латын әліпбиін әзірлеудің алдында қазақ тілінің үнсіз дыбыстарының құрамына х дыбысы енгізілді. Қазіргі кезде қазақ тілінің фонетикалық және фонологиялық заңдылықтарын бұзбайтын басқа тілдегі сөздерде дауыссыз в және ф дыбыстары қолданылады.

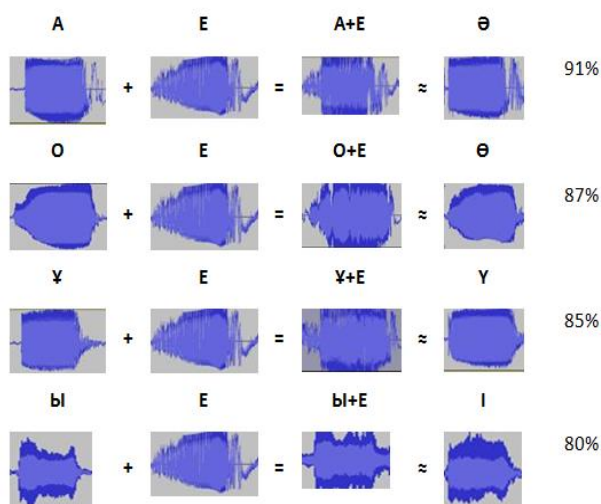
Сонымен классикалық латын әліпбиінде 26 әріп және қазақ тілінде 31 дыбыс бар (28 жасанды дыбыс, 3 енгізілген дыбыс). Дауысты дыбыстардың математикалық модельдерін құру арқылы қатты және жұмсақ дауыстар арасында мына қатынастар бар екенін дәлелденді: $a+e=\text{ә}$, $o+e=\text{ө}$, $y+e=\text{ү}$, $ы+e=\text{і}$. Бұл қарым-қатынастар жұмсақ дауысты белгілеу үшін диграфтарды (екі әріптің тіркесімі) пайдалануға болатынын көрсетеді. Мұндай әдіс басқа тілдерде қолданылады. Мысалы, ағылшын тілінде ш дыбысы s және h әріптерімен белгіленеді.



Сурет 1.5 – Қазақ тілінің дауысты дыбыстарының моделі

АОҰЫ – жуан, ӘӨҮІ – жіңішке, АОӘӨ – ашық, ЫҮҮҮ – қысаң, ОҰӨҮ – еріндік, АЫӘІ – езулік, Е – аралық.

А, О, Ұ, Ы дыбыстарының жолдарына компьютерлік тәжірибе негізінде Е дыбысының жолын қосу $A+E=\text{Ә}$, $O+E=\text{Ө}$, $Y+E=\text{Ү}$ ара қатынасының дәлелдемелері келесі оциллограмм түрінде алынады.



Сурет 1.6 – Қазақ дауыстыларының оциллограммасы

1.3 Сөйлеудің фонемалық мәндерін тану ұғымы

Сөйлеу дыбыстарын әр түрлі белгілер бойынша сипаттауға және жіктеуге болады — акустикалық (физикалық) және артикуляциялық (физиологиялық).

Сөйлеу дыбыстарының артикуляциялық белгілері бойынша жіктелуі екі негізгі сыныптарды немесе түрлерімен — дауысты және дауыссыз дыбыстарды бөлуден басталады, олар барлық тілдерде ерекшеленеді.

Тілдің фонетикалық жүйесінің ең қысқа, қарапайым бірліктері ретінде қарастырылатын сөйлеу дыбыстарымен қатар, әлемнің көптеген тілдерінде полифтонг деп аталатын неғұрлым күрделі фонетикалық бірліктер қолданылады (грек тілінен. poly - 'көп' және phthongos - 'дыбыс'). Олардың арасында дифтонгтар, трифтонгтар және тетрафтонгтар бар. Олардың барлығы бірге "әдеттегі" дыбыстарға қарсы қойылады, олар бұл жағдайда монофтонг деп аталады (грекше. monos - "бір" және phthongos - "дыбыс").

Сөйлеу сигналының басты ерекшелігі – оның көптеген параметрлер бойынша қатты түрленуі: ұзақтығы, қарқыны, дауыс биіктігі, бұрмалану, адамның дауыстық жолының үлкен өзгергіштігімен, диктордың түрлі эмоциялық күйлерімен, түрлі адамдардың дауыстарының күшті айырмашылығымен байланысты. Әр түрлі уақытта жазылған бір адам үшін де бірдей сөйлеу үзіндісінің уақыт көріністері сәйкес келмейді. Бір жағынан, оны толық сипаттайтын (яғни, бір дыбысты екінші дыбыстан ажыратуға мүмкіндік беретін) және екінші жағынан, жоғарыда көрсетілген сөйлеу нұсқаларын теңестіретін сөйлеу сигналының параметрлерін іздеу қажет. Содан кейін бұл параметрлер үлгілермен салыстырылуы тиіс, бұл сәйкестікке қарапайым салыстыру емес, ең көп сәйкестікті іздеу болуы тиіс. Бұл табылған параметрлік кеңістіктегі қашықтықтың қажетті түрін іздеуге мәжбүр етеді.

Осылайша, сөйлеуді тану рәсімі параметрлердің (белгілердің) қолайлы жүйесін пайдалануға негізделуі және ақылға қонымды алгоритмдердің көмегімен орындалуы тиіс. Төменде айтылғандардың сөйлеудің фонемдық мәндерін тану тәсілінің ерекшелігі сөйлеу сигналын жиілік көрінісінде емес, уақыттық көрінісінде қарау болып табылады.

"Фонемалық тану" термині сөздің танылатын бірлігі ретінде біз сөйлемдерді емес, морфемдерді емес, фонемаларды — яғни сөйлеу дыбыстарын, тану тілінің фонетикалық құрылымының немесе элементтерін пайдаланамыз дегенді білдіреді.

1.4 Сөйлеуді тану жүйесінің тарихы

Сөйлеуді тану жүйелерінің көпшілігі (ағылшын тілінде Automatic Speech Recognition - ASR) аналогтық сигналды талдау мен өңдеу процесінен және тану процесінен тұрады. Аналогтық сигналды талдау кезінде сөйлеуден не

айтылғанын анықтау үшін тану процесінде одан әрі қолданылатын қасиеттер ерекшеленеді. Осы екі үдерістің контекстінде ASR жүйесінің қысқаша даму тарихын қарастырайық.

ASR жүйелерін құрудың алғашқы талпыныстары 1950 жылдары жүзеге асырылды. Дикторға тәуелді жүйе салынды. Сигналдың қасиеттері ретінде сөздердегі дауыстардың спектралды резонанстары қолданылды. 1959 жылы дикторға қарамастан он дауысты тануға қабілетті модуль жасалды.

Жапонияда 60-шы жылдары арнайы спектралды талдағышты пайдалана отырып, дауысты дыбыстарды танитын бірнеше машина салынды. Сондай-ақ фонемаларды танитын құрылғы жасалды.

Сөйлеуді тану саласында 70-ші жылдары екі маңызды жаңалық жасалды: сөйлеу диалектілерін уақытша теңестіруге негізделген динамикалық бағдарламалау әдісін (ағылшын тілінде Dynamic Time Warping – DTW) пайдалану және сызықтық болжамдарды кодтау әдісі (ағылшын тілінде Linear Predictive Coding – LPC), ол төмен битрейтпен (битрейт – секундына берілетін ақпарат биттерінің саны) сигналдарды тануда сәтті пайдаланылған. AT&T Bell Laboratories-те LPC анализіне негізделген акустикалық сигналды өңдеу, ал DTW-да тану процесі құрылды.

Ары қарай 80-жылдары шаблондарды қолдануға негізделген тәсілдерден сөйлеуді тану саласындағы зерттеулер статистикалық модельдеу әдістеріне көшті. Жасырын Марковтың модельдері (ағылшын тілінде Hidden Markov Models - HMM) пайдаланылды. Бейкердің жұмысы алғашқылардың бірі болды, онда сөйлеуді тану мәселесін шешу үшін HMM қолданылған. 80-жылдардың соңында тану проблемасына жасанды нейрондық желілерге (ағылшын тілінде Artificial Neural Network – ANN) негізделген әдіс қолданылды. Біздің күндерде көптеген ASR жүйелері тану процесінде HMM пайдаланады.

Сөйлеуді тану 90-шы жылдары бірнеше жетілдірілді. Танылатын сөздер сөздігі бірнеше ондаған мыңға дейін өсті. Декодтау жылдам алгоритмдерін пайдалану нақты уақытта тануды жүргізуге мүмкіндік берді. Олардың саны жиырма мың сөзге жететін жекелеген сөздерді білетін. Қазіргі дикторға тәуелді жүйелерде қателер 0.1% - дан кем. Ал диктордан тәуелсіз жүйелерде қателер 5% - ға жуық, онда мыңдаған сөзден бірізді сөйлеу танылды.

Қазіргі заманғы әдістердің көмегімен нақты уақытта сөйлеуді тану көлемі жиі шектелген және үлкен есептеу ресурстарын талап етеді. Бүгінгі таңда көптеген алгоритмдерді кеңінен қолданудың мүмкін еместігі, мысалы, мобильді құрылғыларда зерттеушілерді неғұрлым тиімді және оңтайландырылған әдістерді іздеуге мәжбүрлейді.

Сонымен сөйлеуді тану шамамен елу онжылдықтың зерттеу саласы болды. Бұл салада көптеген өзгерістер болды, бірақ әлі күнге дейін толығымен дәл орындалатын жүйе жасалмады. Көптеген жылдар бойы әзірленген негізгі жүйелер келесі кестеде сипаттамаларымен келтіріледі.

Кесте 1.1 – Жылдар бойғы тану жүйесіндегі дамулар

Жыл	Негізгі оқиғалар
2000	CMU Sphinx – Sphinx 2, Sphinx 3 және Sphinx 4 болып табылатын танушының бірнешеуін қамтиды. Олар Жасырын Марковтың Модельдерін пайдаланады.
2001	RWTH ASR-сондай-ақ RASR ретінде белгілі, сөйлеуді тану декодері бар. Сонымен қатар, бұл жүйе акустикалық үлгілерді жасау үшін қажетті құралдарды қамтиды. Мұны тану жүйесінде қолдануға болады.
2003	Dragon NaturallySpeaking 7 Medical – ScanSoft жасаған. Бұл сөзді дәл тану есебінен медициналық қызмет көрсету құнын төмендетті.
2008	Siri Inc құрылды. iPhone-да Google Voice Search құрылды.
2010	Google Voice Search android-та іске қосылды.
2012	Siri Apple iPhone 4s іске қосылды.

1.5 Сөйлеуді тану міндеттерін шешудің негізгі әдістері

Сөйлеуді тану көрсеткіштерін жақсарту үшін қолданылатын бірінші кадам дыбысты адам құлағымен қабылдау ерекшеліктерін ескере отырып, сигналдан қасиеттердің векторларын бөлуге негізделеді. Ол тасушы жиіліктерді талдау және сигналды дыбыс қаттылығы бойынша түзеуді қамтиды. Мұндай тәсілді қолданатын ең көп таралған технологиялар тондық жиіліктің кепстральді коэффициенттерінің әдісі (ағылшын тілінде Mel Frequency Cepstral Coefficients – MFCC) және сызықтық болжам коэффициенттерінің әдісі (ағылшын тілінде Perceptual Linear Prediction – PLP) болып табылады. Сөйлеуді тану есебін шешудің негізгі тәсілдері адам қабылдауға тән шаблонмен бір мезгілде және озыңқы салыстыру (бүркемелеу) шудан үлкен орнықтылықты қамтамасыз ететін қасиеттерді бөлу үшін модельдеуі және пайдаланылуы мүмкін. Осы мақсатта кадрлардың өлшемін түрлендіру әдісі құрылды. Есту рецепторына жауапты жүйке жасушаларының жұмыс ерекшелігін ескере отырып диапозонды автокорреляция әдісі ұсынылды.

Басқа тәсіл дыбыстық сигналдардың талдауына негізделген. Жүйеге келіп түсетін шулы сигналдардың оқыту барысында алынған "таза" сигналдарды шаблондардан айырмашылығы тану жүйелері жұмысының тұрақсыздығының негізгі себебі болып табылады. Тәсілдің мақсаты бұл

айырмашылықты азайту болып табылады. Дыбыс сигналдарындағы шу аддитивті және стационарлық деп болжанады. Орташаланған шудың орташа мәнін бағалау кепстрден немесе спектрден (Spectral Subtraction, SS, Virag, 1999) шегеріледі. Мұндай әдістердің кейбір модификациялары спектральды қисықты пайдаланатын бейсызық спектральды азайтуды қамтиды. Мұндай техникалар шуды жақсы бағалауды талап етеді, оны іс жүзінде алу қиын, әсіресе стационарлық емес фондық шу жағдайында.

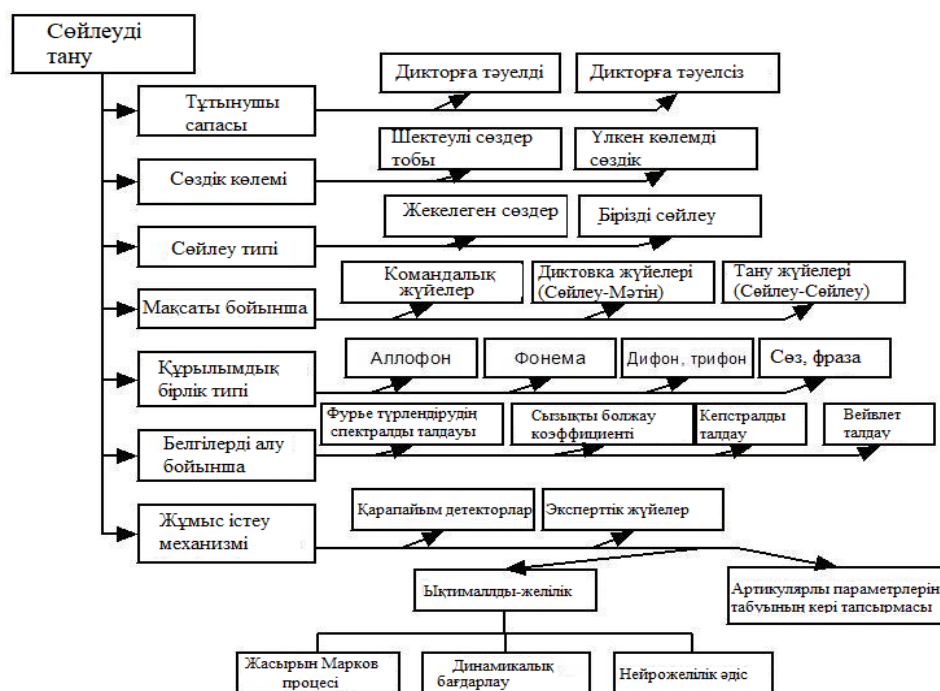
Шуылдан және таза сигналдардан алынған қасиеттер арасындағы айырмашылыққа қарсы күрестің тағы бір жолы жоғары жиілікті сүзгіні пайдалану болып табылады. Сигналда шу тұрақты емес, баяу өзгереді болады деп болжанады. RASTA әдісінде салыстырмалы спектральды өзгерістер бекітілетіндей түрде берілген. Шудан туындаған баяу өзгерістер жойылады. Бұл жағдайда шуды анық бағалау қажеттілігі болмайды.

Үшінші тәсіл көп өлшемді кеңістіктерді (Ephraim & Trees, 1994) пайдалануға негізделген. Бұл тәсілдің негізгі идеясы-бағалау функциясын азайтатын сызықтық бейнелеуді табу. Көп жағдайда мұндай бейнелеу ретінде сипаттар векторын түрлендіру матрицасына көбейту алынады. Бұл тәсілдің мысалдары ретінде негізгі компоненттік талдау (Principal Component Analysis, PCA) және тәуелсіз компоненттік талдау (Independent Component Analysis, ICA, Koscor, 2000), сондай-ақ көп өлшемді ішкі кеңістікті жобалауды (Gales, 2002) айтуға болады.

1.6 Сөйлеуді тану жүйесінің және әдістерінің жіктелуі

Сөйлеуді тану жүйелерінің жіктелуін қарастыру кезінде оның әр түрлі параметрлеріне байланысты жіктелетінін атап өту керек. Сөйлеуді тану жүйесінің әдеби көздеріне сүйенсек, оны келесідей жіктеуге болады:

- сөздіктің мөлшеріне байланысты: сөздікті тану жүйесі шектеулі сөздер жиынтығы; үлкен көлемдегі сөздікпен жүйелер;
- дикторға тәуелділікке байланысты: дикторға тәуелді және дикторға тәуелсіз;
- танылатын сөздің түріне байланысты: бірізді сөйлеумен жұмыс істейтін жүйелер немесе бөлшек сөйлеу;
- жүйенің мақсатына байланысты жүйені бөлу қабылданған: диктовкалар және командалық жүйелер;
- тану жүйесінде қолданылатын алгоритмге байланысты: нейрондық желілер, Жасырын Марков модельдері, динамикалық бағдарламалау; Жүйеде қолданылатын құрылымдық бірліктердің түріне байланысты сөздер, дифондар, фонемалар, аллофондар, фразалар қолданылуы мүмкін;
- тану жүйесінің құрылымдық бірліктері алынатын принциптерге байланысты шаблон арқылы тану жүйесі және лексикалық элементтерді алу жүйелеріне бөлінеді (Сурет 1.7).



Сурет 1.7 – Сөйлеуді тану жүйесінің әртүрлі жіктелуі

Сөйлеуді автоматты тану жүйелері үшін, кедергіден қорғалу ең алдымен екі механизмдерді пайдаланумен қамтамасыз етіледі:

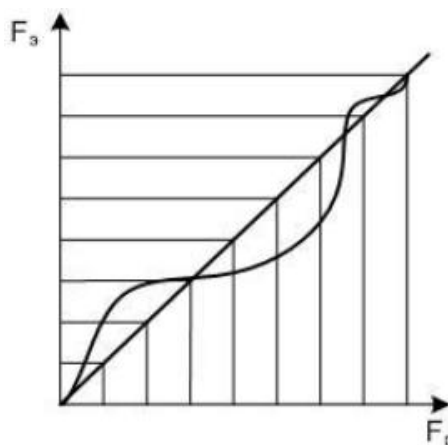
- акустикалық сигналды талдау негізінде сөйлеу сигналының бір элементтерін бөлу тәсілдерін бірнеше параллель жұмыс істеуші әдісті қолдану;
- сөйлеу ағымында сөздерді сегментті (фонемдік) және тұтас қабылдауды параллель тәуелсіз қолдану.

Сөйлеуді тану жағдайында ең күрделісі екі тілдік элементтерді салыстыру процедурасын жүзеге асыру болып табылады, ол уақыт бойынша созылуын да сипаттайды. Қазіргі уақытта мұндай процедуралар мен әдістер жеткілікті (сурет 1.8).



Сурет 1.8 – Сөйлеуді тану әдістері

Сөйлеуді танудың сызықтық модельдері бәрінен бұрын пайда болды. Сөйлеуді танудың сызықтық модельдерінде сөйлеу сигналын эталонмен салыстыру жұмыстары үшін уақыт бойынша қарапайым масштабтау жеткілікті. Алайда, сөзде уақыттың елеулі сызықсыз бұрмалануы орын алады. Басқаша айтқанда, сызықтық модель іске асыруды сызықтық заң бойынша эталонмен салыстыруды болжады (сурет 1.9), ал іске асырудағы өзгерістер сызықты емес бұрмалауға ұшырайды. Әлбетте, мұндай модель қанағаттанарлықсыз жұмыс істеді.



Сурет 1.9 – Сөйлеуді танудың сызықтық моделінің жұмыс істеу сұлбасы

Кейінірек, сызықтық модельдің кемшіліктерін ескере отырып, уақыт бойынша динамикалық теңестіру әдісі пайда болды. Уақыт бойынша динамикалық теңестіру әдісінде іске асыруды эталонмен салыстыру траекториясы сызықтық әдістен ерекшеленуі мүмкін. Динамикалық бағдарламалау (ДБ) әдісі деп те аталатын бұл әдіс тану нәтижелерін айтарлықтай жақсартуға мүмкіндік берді. ДБ әдісті құрудың алғышарты қарқынның және айтылу мәнерінің еріксіз өзгеруі салдарынан пайда болатын айтылатын команданың уақытша деформациясын қалыпқа келтіруге ұмтылу болды. Бірақ бұл әдістің кейбір кемшіліктері әлі де қалды. Атап айтқанда, жақсы тану үшін үлкен есептеу ресурстары және эталондарды сақтау үшін жадының үлкен көлемі талап етілді (әрбір диктор үшін – өзінің эталондар жиынтығы). Жасырын Марков модельдері әдістерінде эталон сигналды есептеу тізбегі түрінде емес, бір күйден екіншісіне өту мүмкіндігі түрінде ұсынылады. Жоғарыда көрсетілген әдістердің басты артықшылығы – олар аз шығынмен танудың сенімді нәтижелерін береді. Олардың негізгі кемшілігі – өте күрделі оқыту процедураларының болуы. Осыған байланысты нейрожелілік модельдер негізінде белсенді дамып келе жатқан әдіс пайда болды. Мұндай модельдер оқыту механизмін неғұрлым тиімді түрде формалдауға мүмкіндік береді. Тағы бір әдіс – сараптамалық білім негізінде құрылған – құрылымдық-сараптамалық әдістер деп аталады. Бұл тәсіл тек формальды математикалық модельдерді ғана емес, сонымен қатар кейбір

тілдік және тілдік модельдерді қалыптастыратын білім базасын да пайдалануды көздейді. Бұл әдіс қазіргі уақытта тілдік тану үшін ерекше перспективалы, оның болашағы зор. Қалған әдістер жеке түрде қолданудан қалуда. Сондықтан іс жүзінде әзірлеушілер бір жүйе шеңберінде бірнеше әдістерді пайдалануға жиі жүгінеді. Мұндай гибридті модельдерде жаңа әдістердің барлық оң қасиеттері қолданылады.

1.7 Сөйлеуді автоматты танудың мақсаты, негізгі міндеті, жұмыс істеу принципі

Ауызша сөйлеуді автоматты түрде тану — жасанды интеллекттің дәстүрлі міндеті. Онымен информатика пайда болған кезде де автоматты түрде бір тілден екінші тілге аударудың міндеті сияқты ынта-ықыласпен айналыса бастады. Алайда бірнеше онжылдықтар өткеннен кейін көптеген зерттеу топтарының нәтижелері сол және басқа облыста өте қарапайым болып қала береді.

Сонымен сөйлеуді автоматты тану (САТ) сондай-ақ компьютерлі сөйлеуді тану компьютерлік бағдарлама ретінде жасалған алгоритм арқылы сөйлеу сигналын сөздер тізбегіне, басқа да лингвистикалық бірліктер түріне айналдыру процесін айтамыз. САТ негізгі мақсаты – бұл сөйлеуді компьютерлерге кіріс деректер ретінде тануға мүмкіндік беретін әдістер мен жүйелерді жүзеге асыру.

Сөйлеуді танудың екі негізгі міндеті – командалардың шектеулі жиынтығында жүз пайыз дәлдікке қол жеткізу және диктордан тәуелсіз еркін сөйлеуді қолайлы сапамен тану-оларды әзірлеудің жарты ғасырлық тарихына қарамастан шешілмеген.

Тану жүйелері мәтінге тәуелді және мәтінге тәуелсіз болып бөлінуі мүмкін. Бұл пайдаланушы айтатын және жүйе тануы тиіс мәтін жүйеге белгілі болу болмауына байланысты. Мәтінге тәуелді тану кезінде бекітілген фразалар да, жүйемен жасалған және пайдаланушыға ұсынылған фразалар да қолданылуы мүмкін. Мәтінге тәуелсіз жүйелер еркін сөйлеуді өңдеуге арналған.

Әдетте сәйкестендіру жүйелері үш кезеңде жұмыс істейді:

1. Сигналды өңдеу кезеңі. Бұл кезеңде сигнал тану міндеті үшін маңызды белгілерді анықтау үшін өңделеді. Сигналды өңдеу кезеңінде сөйлеу сигналы белгілер векторларының реттілігі арқылы беріледі;
2. Модельдерді құру кезеңі. Пайдаланушыны тіркеу кезінде бұл кезең сигналды өңдеу кезеңінен алынған модельді құру үшін белгілер векторларының бірізділігін пайдаланады. Моделдеу белгілердің векторларын қарапайым көшіруде де, ықтималдық модельдерді немесе басқа да құрылымдарды құруда да болуы мүмкін. Содан кейін осы белгілерде белгілер мен сақталған модель арасындағы ұқсастық дәрежесін есептеу мүмкін болады;

3. Шешім қабылдау кезеңі. Шешім қабылдау функциялары дәстүрлі түрде жеке кезеңге бөлінеді, бірақ ол тривиалды функцияларды орындай алады немесе егер модельдер кезеңінде соңғы шешімдер есептелсе болмайды. Шешім қабылдау үшін модельдер кезеңінде есептелген ұқсас дәрежелер және қажет болса, берілген шектеулер қолданылады

1.8 Сөйлеуді тану кезеңдері

Сөйлеуді тану жүйесінде біз сөзді мәтінге, яғни сөйлеуді тану жүйесіндегі шығыс дерегіне жататын мәтінге түрлендіреміз және ол танылған сөйлеуге эквивалентті. Сөйлеуді тану қолданбалары соңғы бірнеше жылда дамытылды. Бұл қолданбаларда дауыстық іздеу, қоңырауды маршрутизациялау пәрмені және басқару, құралды дауыспен басқару, дауыс арқылы теру, компьютер арқылы тілді оқыту, робототехника және т.б. Қазіргі сөйлеуді тану жүйелері ЖММ-не, яғни Жасырын Марков Моделіне немесе жоғарыда ағылшын тілімен аталғандай НММ негізделген. ЖММ-нің кеңінен қолданылуының негізгі себебі – ЖММ автоматты түрде оқылатын немесе оқытылатын параметрлері бар және оқыту үшін қолданылатын әдістер қарапайым және есепті түрде пайдалануға орынды. Сөйлеуді автоматты тануда көп жетістіктерге қол жеткізілді, алайда осы уақытқа дейін кез келген ортада көптеген адам тілін түсінетін машинаны жасап шығару мүмкін болмады [9].

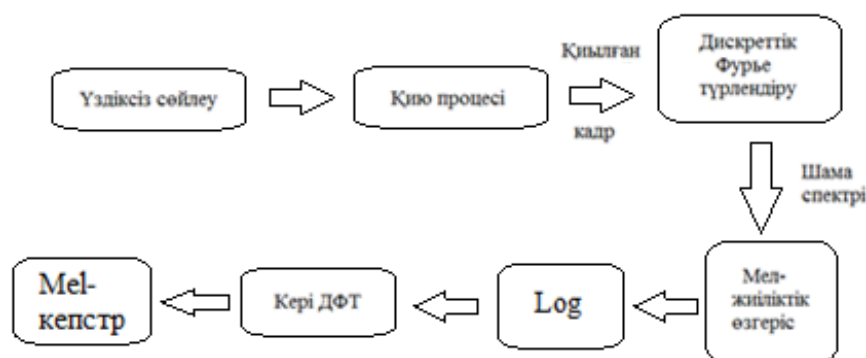
Сөйлеуді тану жүйесінің негізгі мақсаты – айтылған ақпараттың негізінде тыңдау, түсіну, содан кейін әрекет ету қабілеті болып табылады. Сөйлеуді тану жүйесі төменде суретте көрсетілгендей қосымша жіктелетін төрт негізгі кезеңді қамтиды.



Сурет 1.10 – Сөйлеуді тану кезеңдері

1. Талдау кезеңі: Сөйлеуші сөйлегенде, сөйлеуінде сөйлеушіні анықтауға көмектесетін ақпараттың әрбір түрі қамтылған. Ақпарат вокал жолы, ынталандыру көзі сияқты әртүрлі мінез-құлық ерекшеліктеріне орай әр-алуан. Жоғарыда көрсетілгендей, сөйлеуді талдау кезеңін үш санатқа бөлуге болады: сегменттеу талдауы, субсегментті талдау, жоғары сегментті талдау.

2. Ерекшеліктерді алу кезеңі: Ерекшеліктерді алу сөйлеуді тану жүйесінің негізгі бөлігі болып табылады. Ол жүйенің жүрегі болып саналады. Оның жұмысы сөйлеушіні тануға көмектесетін кіріс сөзінен (сигналдан) функцияларды алу болып табылады. Алу функциясы сөйлеу сигналының қуатына ешқандай зиян келтірместен кіріс сигналының (вектордың) шамасын сығады. Ерекшеліктерді алудың көптеген кезеңдері бар.

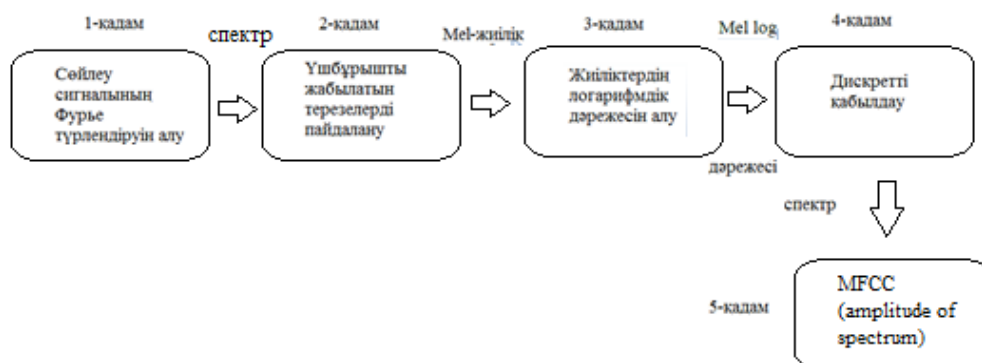


Сурет 1.11 –Ерекшеліктерді алу сұлбасы

Жоғарыда келтірілген суретте ерекшеліктерді алу сұлбасы көрсетілген. Бұл, бір жағынан біз үздіксіз сөйлеу сигналдарын қию процесі үшін енгіземіз. Қию процесінде кадрдың басында да соңында да болатын жаңылыстар барынша төмен. Осыдан кейін үздіксіз сөйлеу сигналы қиылған кадрларына түрлендіріледі. Бұл қию кадрлары дискретті Фурье түрлендіргішіне беріледі. Сөйтіп қию кадрлары шама спектрлерге бөлінеді. Енді келесі кезеңде спектралды талдау мел-шама өндіретін мел-жиіліктер шкаласы болып табылатын жиіліктердің субъективті шкаласы бойынша тіркелген рұқсатпен орындалады. Содан кейін бұл шама Log-ға беріледі, содан кейін соңғы нәтиже – Mel-Cepstrum-жіберетін кері дискретті Фурье түрлендіруі жүзеге асады. Mel-Cepstrum дикторды анықтау үшін қажетті сипаттамалардан тұрады. Ерекшеліктерді алу әдістерінің бірнеше сипаттамасы келесілерді қамтиды:

- сызықтық болжаммен кодтау: СБК-сөйлеуді өңдеу үшін қолданылатын құрал. Модель болжамға негізделген. Сөйлеу мысалдарының топтамасында біз ұсынылуы мүмкін n-ші топтаманы болжай аламыз. Бұл топтама мақсатты сигналдың (k) алдыңғы топтамаларының қосындысы түрінде берілуі мүмкін. Кері сүзгі өндірісі осылай жасалуы керек. Сондықтан бұл сөйлеу топтамаларының формантты аймақтарына сәйкес келеді. Осылайша осы сүзгілерді топтамалар ішінде қолдану СБК процесі болып табылады.

- Мел-жиілік кепстраль (МЖК): Мел-жиілік кепстраль коэффициенттер 1000 Гц төмен жиіліктегі адам құлағы жолағының сыни енінің белгілі вариацияларына негізделген. МЖК процессорының негізгі мақсаты – адам құлағының мінез-құлығын көшіру. МЖК келесі ретпен жүзеге асырылады:



Сурет 1.12 – МЖК жүзеге асу процесі

- салыстырмалы спектральды сүзгілеу: Бұл шулы ортада жазба кезінде сөйлеуді жақсарту үшін қолданылатын әдіс. Сөйлеу сигналдары көрінісінің уақыттық траекториялары Салыстырмалы спектральды сүзгілеудің жолақты сүзгісінен өтеді. Бастапқыда салыстырмалы спектральды сүзгілеу тек сөйлеу сигналыдағы шудың әсерін азайту үшін қолданылды, енді, сонымен қатар, ол сигналды тікелей күшейту үшін пайдаланылады. Келесі суретте Салыстырмалы спектральды сүзгілеу кезеңдері көрсетілген. Мұндағы басты ой -тұрақты факторларды бағындыру.



Сурет 1.13 – Салыстырмалы спектральды сүзгілеу кезеңдері

- ықтималдық сызықтық ажырату талдауы (ЫСАТ): бұл әдіс сызықтық ықтималдық талдаудың (СЫТ) кеңейтілген түрі болып табылады. Бастапқыда бұл әдіс адамдардың бет-әлпетін тану үшін пайдаланылды, бірақ қазір ол сөйлеуді тану үшін қолданылады.

3. Модельдеу кезеңдері: модельдеу кезеңінің мақсаты – ерекшеліктерді алуды қолдана отырып (ерекшелік векторы) сөйлеуші модельдерін жасау болып табылады. Сурет 1.10-да көрсетілгендей, модельдеу әдістері одан әрі сөйлеушіні тану мен идентификациясы деп бөлінеді. Сөйлеушіні тануды сөйлеушіге тәуелді және тәуелсіз деп бөлуге болады. Сөйлеушінің идентификациясы – бұл сөйлеу сигналынан алынған ақпарат негізінде сөйлеушіні сәйкестендіруге қабілетті процесс.

4. Сәйкестендіру кезеңі: табылған сөз келесі әдістердің бірін пайдалана отырып белгілі сөз үшін сөзді тану механизмімен пайдаланылады. Олар толық сөзді сәйкестендіру және жартылай сәйкестендіру.

1.9 Matlab көмегімен сөйлеу сигналының талдауына кіріспе

Жоғарыда баяндалған элементтерді практикалық аспектілерде көрсету мақсатында сөйлеу сигналына қысқаша талдау жүргізейік. Талдау соңында:

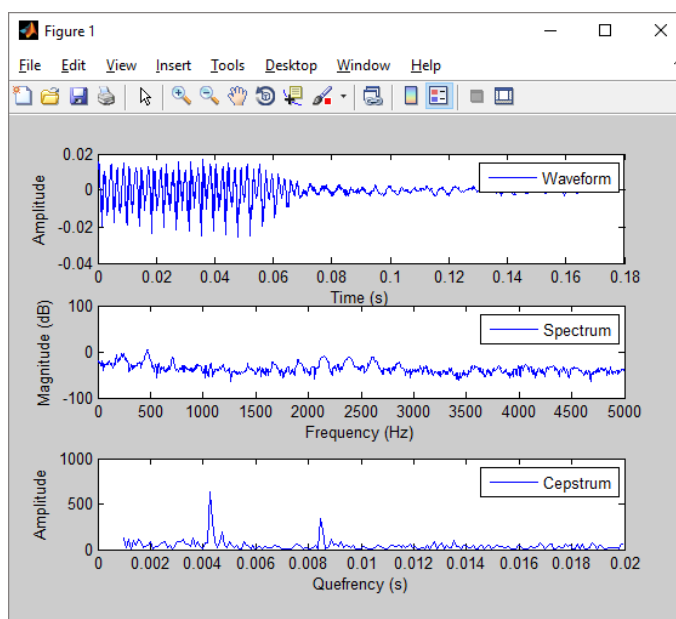
- сөйлеу сигналының учаскесінен негізгі жиілігін оның спектрінен бағалаудың бір тәсілін;
- сөйлеу сигналының учаскесінен негізгі жиілігін оның толқын формасынан бағалаудың бір тәсілін;
- сөйлеу сигналы секциясынан жиілік формантын бағалаудың бір тәсілін;
- негізгі жиілік пен формант жиілігінің кейбір мәселелерін бағалауды қарастырамыз.

Негізгі жиілікті бағалаудың жалпы міндеті сигналдың бір бөлігін алу және қайталанатын бөліктердің басым жиілігін табу болып табылады. Қиындықтар барлық сигналдар периодты еместігінен және периодты сигнал бөлігі қажетті уақыт ішінде негізгі жиілікте өзгеруі мүмкін, сигналдар шумен, тіпті басқа негізгі жиіліктердің периодты сигналдарымен бұрмалануы мүмкін, Т интервалымен периодты болып табылатынымен байланысты туындайды. Сондықтан ең аз периодты интервалды немесе ең жоғары негізгі жиілікті табу керек; және тіпті тұрақты негізгі жиіліктегі сигналдар қажетті интервал ішінде басқа тәсілдермен де өзгеруі мүмкін.

Ұзын, таза, тұрақты сөйлеу сигналдары үшін басым фундаменталды жиілік бағасын алудың сенімді тәсілі cepstrum пайдалану болып табылады. Cepstrum-логарифмдік амплитудалық спектр сигналының Фурье-талдауы. Егер логарифмдік амплитуданың спектрінде үнемі таратылған гармониктер көп болса, онда спектрдің Фурье-талдауы гармониктер арасындағы қашықтықты көрсетеді: яғни негізгі (фундаменталды) жиілік. Тиімділік үшін сигнал спектрін басқа сигнал ретінде қарастырамыз, содан кейін спектрдің өзінде периодтылықты іздейміз.

Cepstrum (кепстр) деп аталу себебі ол спектрді кері айналдырады. Кепстрдің жиілік бірліктері бар, ал кепстр шыңдары (спектрдегі кезеңділікке қатысты) рамоника деп аталады.

Кепстр фундаменталды жиілік бағасын алу үшін, біз типтік тілдік негізгі жиіліктер аймағындағы шынды іздейміз.

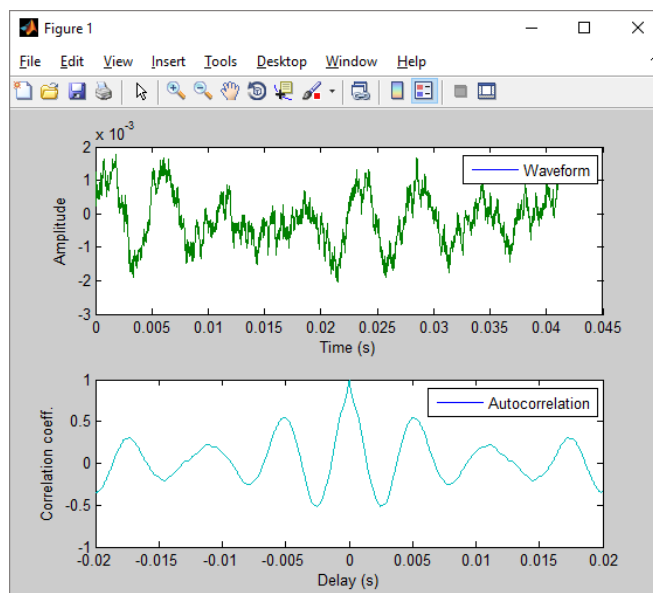


Сурет 1.14 – Негізгі жиілікті бағалау-жиіліктік аймақ

Кепстр негізгі жиілік тез өзгермегенде, негізгі жиілік жоғары болмағанда және сигнал шудан салыстырмалы таза болғанда жақсы жұмыс жасайды. Кепстрдің кемшілігі – жиілік аймағының өңделуі қымбатқа түсуінде.

Негізгі жиілікті бағалау-уақыттық аймақ – кепстр сигналдың лог-спектрінде периодтылықты іздейді, ал адамда тонның биіктігін қабылдау толқынның өз формасындағы периодтылықпен тығыз байланысты. Толқын пішінінен негізгі жиілікті бағалау құралы автокорреляцияны пайдалану болып табылады. Сигналдың аумағы үшін автокорреляция функциясы сигналдың формасы әртүрлі кідірістер кезінде өздігімен қаншалықты жақсы корреляцияланатындығын көрсетеді. Біз мерзімді сигнал тангаждың қысқа кезеңдеріне сәйкес келетін өте қысқа Периодты сигнал өзімен әртүрлі кідірісіздерде жақсы корреляцияланады деп күтеміз.

Бұл код сөйлеу аймағы үшін автокорреляция функциясын көрсетеді:

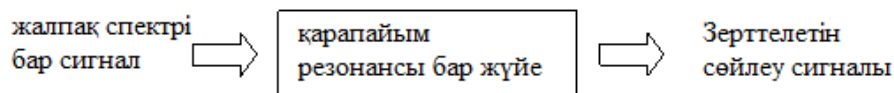


Сурет 1.15 – Негізгі жиілікті бағалау-уақыттық аймақ

Автокорреляциялық тәсіл сигналдың төмен, тұрақты қадамы болғанда және сигналдың спектрлік мазмұны тым тез өзгермеген кезде жақсы жұмыс істейді. Автокорреляция әдісі қадамның екі периодының қате таңдауында қадамды теңдей бөлу қателіктеріне бейім. Оған форманттық резонанстардан туындаған сигналдың кезеңділігі әсер етуі мүмкін, әсіресе әйелдер дауыстары үшін F1 FX жиілігінен төмен болуы мүмкін.

Форманттың жиілігін бағалау әдетте, негізгі жиілікті бағалауға қарағанда қиын. Мәселе мынада, формантты жиіліктер дауыс жолы жүйесінің қасиеттері болып табылады және тек қана өлшеніп қана қоймай, сөйлеу сигналынан шығарылуы тиіс. Дауыс трактісінің қозуының спектралды түрі байқалатын спектралды өршітуге қатты әсер етеді, сондықтан біз дауыстық трактінің барлық резонанстары байқалатын спектралды аумақта шыңдарды шақыратынына кепілдік бере алмаймыз.

Форманттың жиілігін бағалаудың басым әдісі сөйлеу сигналын модельдеуге негізделген, егер ол белгілі бір ақпарат көзден және сүзгі түрімен генерацияланса:

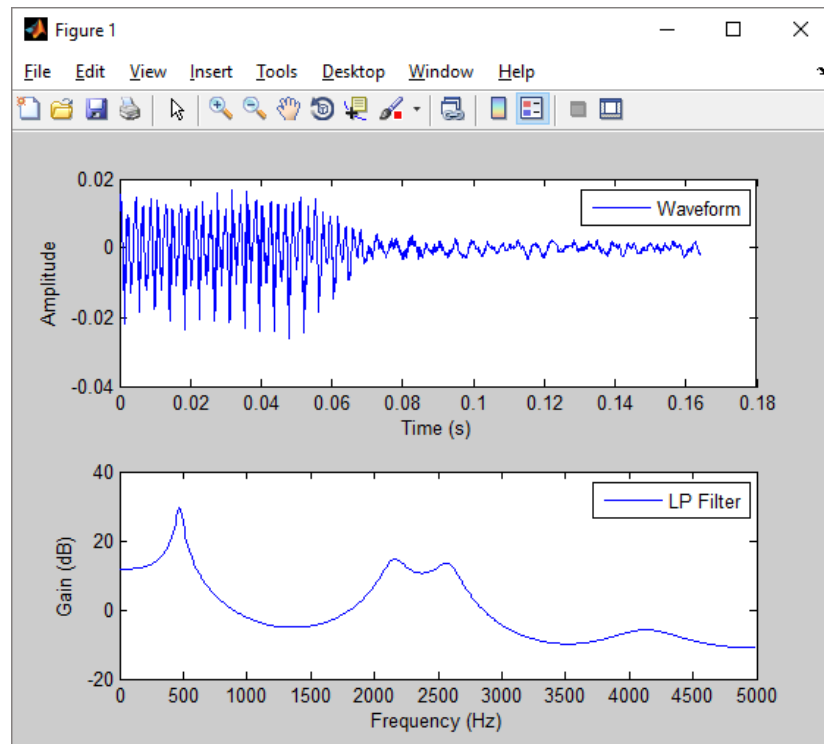


Сурет 1.16 – Сөйлеу сигналының генерациялану сұлбасы

Талдаудың бұл түрі ақпарат көзі-сүзгіні бөлу деп аталады, және форманттың жиілігін бағалаған жағдайда бізді тек модельдеуші жүйе мен оның резонанстарының жиілігі қызықтырады. Ең жақсы сәйкестік жүйесін табу үшін біз сызықтық болжау деп аталатын талдау әдісін пайдаланамыз.

Сызықтық болжам сигналды модельдейді, ол таза рекурсивті сүзгі ІІР арқылы өтетін ең аз энергия сигналымен жасалады.

Біз бұл идеяны LPC пайдалана отырып, сөйлеу сигналынан ең жақсы ІІР сүзгісін табу, содан кейін сүзгінің жиілік сипаттамасын құру үшін көрсетеміз.



Сурет 1.17 – Осциллограммадағы формант жиіліктері

Сүзгіден формант жиілігін табу үшін сүзгіні құрайтын резонанстардың орнын табу керек. Бұл полинома ретінде сүзгі коэффициенттерін өңдеуді және полином түбірін шешуді қамтиды. Мұнда LP сүзгісінен формант жиілігін табу коды:

Шығатын нәтиже:

Formant 1 Frequency 471.6

Formant 2 Frequency 2147.5

Formant 3 Frequency 2316.8

Formant 4 Frequency 2585.5

Formant 5 Frequency 4131.5

2. Сөйлеудің фонема бойынша сигналдық ерекшеліктерін зерттеу

Осы бөлімде сөйлеу сигналын алдын ала өңдеу әдістері сипатталады. 24 элементті спектралды векторды алу процесі зерттеледі, өйткені спектралдық талдау сөйлеудің жиі қолданылатын параметрлік көріністерінің бірі болып табылады. Сонымен қатар Кохонен алгоритмі бойынша нейрондық желі негізінде сөйлеу сигналын тану моделі ұсынылады. Практикалық зерттеу Matlab, CoolEditPro, Praat бағдарламалық жабдықтар көмегімен орындалады. Сөйлеу сигналының негізгі көрсеткіштері бойынша салыстырмалы практикалық талдауы жүргізіледі. Оның ішінде сигналды сыртқы факторлардың әсерінен кейін параметрлерінің өзгерісін, фонемдік және акустикалық талдауы орындалады.

2.1 Сөйлеу сигналын өңдеуде Нейрон желілерін қолдану

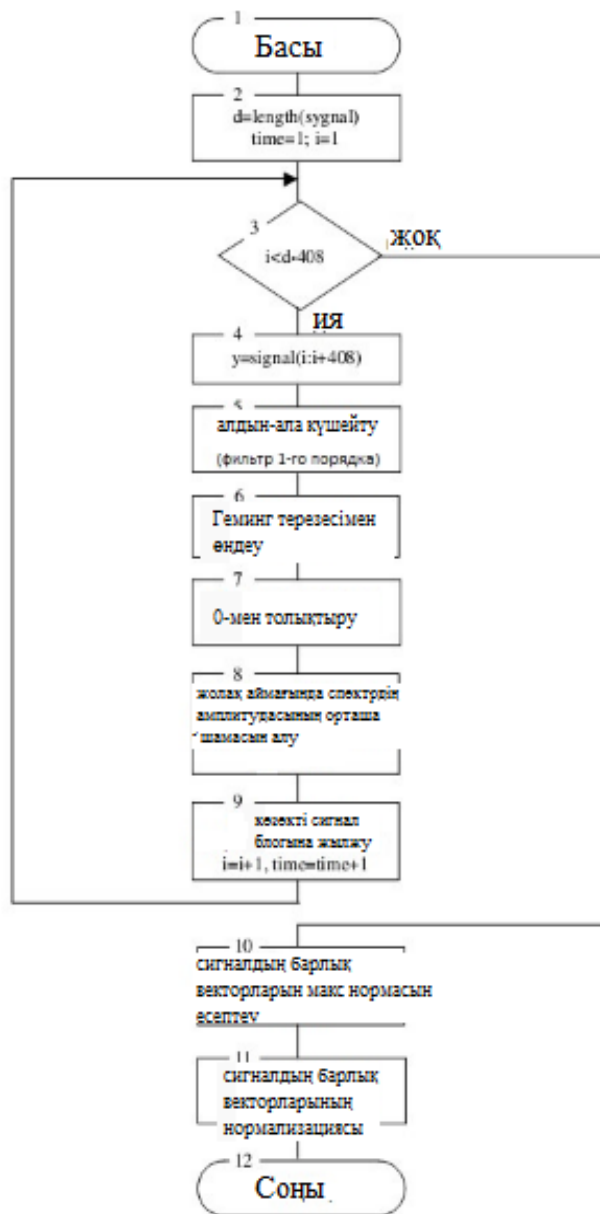
Ақпараттық-телекоммуникациялық жүйелердің заманауи әлемінде өзекті мәселелердің бірі болып сөйлеу деректерін өңдеу міндеті соңына дейін шешілмей тұр. Олардың арасында дауыс бойынша сәйкестендіру жүйесі, сөйлеуді мәтінге түрлендіру, мәтін бойынша синтез және дауыстық басқару жатады. Дыбыстар энергиясын тарату ерекшеліктерін зерттеу кезінде барлық дыбыстардың жиілік аралықтары бойынша жеке энергия таратылуы бар екені анықталды. Сонымен қатар, энергияны тарату айтушының орналасқан жеріне, оның эмоциялық жағдайы мен интонациясына байланысты. Сөйлеу сигналдарының сипаттамаларының арасында барлық дыбыс бойы шамалы өзгертіндерді бөліп көрсетуге болады. Дауыстық тербелістердің жиілігін сипаттайтын параметрлердің бірі негізгі тон жиілігі болып табылады. Негізгі тонның жиілігі адам әңгімесі кезінде өзгереді және оның салыстырмалы өзгеруі 15% - ға жетуі мүмкін. Еуропалық тілдерде негізгі тон сөйлеудің эмоционалдық құрамдас бөлігін, ал кейбір шығыста – мағыналық бөлігін береді. Эксперименталды анықталғандай, әйел дауыстары үшін негізгі тонның кезеңі орташа есеппен 220-дан 350 Гц-ке дейін, ал ерлер үшін 100-ден 220 Гц-ке дейін. Байлам тербелісі – бұл сөйлеу жолының дауыстық қозу көзінің негізгі параметрі болып табылады. Олар дауысқа дыбыс беріп, оның биіктігін сипаттайды. Негізгі тонның жиілігі байлам ұзындығына, олардың салмағына және тартылуына байланысты. Сөйлеу сигналының алдын-ала өңдеу сигналды сипаттайтын көптеген спектралды векторларды алу үшін жасалады.

Қазіргі заманғы тану құралда жасалатын принцип сөйлеу сигналы бірнеше ондаған миллисекунд интервалында тұрақты (яғни оның спектралды сипаттамалары салыстырмалы тұрақты) деп қарастырылады. Сондықтан алдын ала өңдеудің негізгі функциясы кіріс сөйлеу сигналын интервалдарға бөлу және әрбір интервал үшін тегістелген спектралды бағалар алу болып табылады.

Бір интервалдың типтік шамасы-25,6 мс. Көрші аралықтар алдыңғы аралыққа қатысты жылжумен алынады. Аралық жабудың қолданылатын

шамасы 10 мс тең. Көрсетілген нтервалдардың әрқайсысын алдын ала қарастыру нәтижесінде бірнеше ондаған спектралды мәндерден вектор аламыз.

Сөйлеу сигналын алдын ала өңдеу алгоритмінің блок-схемасы сурет 2.1 -де көрсетілген.



Сурет 2.1 – Сөйлеу сигналын алдын-ала өңдеуінің блок-схемасы

Сөйлеу сигналының әрбір аралығын алдын ала пысықтау үшін орындалуы қажет қадамдар бұдан әрі егжей-тегжейлі сипатталған.

1.16 КГц жиілігімен және 16 бит разрядтығымен дискреттелген сөйлеу үлгілерін қарастырамыз. Дискретизацияланған сөйлеу сигналы ұзақтығы 25,6 мс, яғни 409 санау аралығына бөлінеді. Аралықтар 10 мс жылжумен жабылады (160 есептеу).

1. Еріндердің шөгуі сигналдың әлсіреуіне әкеледі және оны компенсациялау үшін сигналды 1-ші ретті сүзгіден өткізу арқылы жоғары жиілікті күшейту жүргізіледі

$$S(1)=0; S(n)=y(n)-y(n-1), n=2,\dots,409, \quad (2.1)$$

мұндағы $y(n)$ -блоктың n -ші санақ.

2. Терезелі Хэмминг функциясымен талданатын сигнал учаскенің орау мүмкіндігі қабықта спектралды құраушылардың өтуін жою үшін жасалады. Ол келесідей анықталады

$$D(n)=(0,54-0,46 \cdot \cos(2\pi \cdot (n-1)/408)) \cdot S(n), n=1,\dots,409 \quad (2.2)$$

3. Спектралды бағаларды алу үшін Фурье түрленуі қолданылады

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j2\pi nk/N}, k = \overline{0, N-1}, \quad (2.3)$$

мұндағы $x_n, n=\overline{0, N-1}$ -дискретті сигнал. N -түрлену периоды.

Оның оң жағында қажетті нөлдермен толықтыру арқылы блоктың ұзындығы 512 элементке дейін артады. Кейін ұзындығы 512 нүктесі бар жылдам Фурье түрлендіруі (ЖФТ) қолданылады және шығыста 512 спектрлік кешенді мәндерді аламыз. Себебі, Фурье түрлендіру қолданылатын 512 мәндер жарамды болып табылады, онда алынған спектралды кешенді мәндер қосарлы түйінделеді, яғни 512-н екінші мәнмен, үшіншісі-511-н және тағы басқалай өйткені соңғы 256 кешенді мәндер алдыңғылармен кешенді байланысқан және жаңа ақпаратты көтермейді, оларды түрлендіру елемейді.

4. Бастапқы 256 кешенді спектралды мәндер үшін олардың амплитудасы анықталады. «Үшбұрышты» жиілік жолағының айналысында Фурье амплитуда спектрі сызықсыз Mel-шкалада орналасқан спектралды коэффициент амплитудасымен қосылу арқылы тегістеледі (орташаланады). 16кГц-ке тең тілдің шекті жиілігі үшін 24 осындай жиілік жолағы алынады.

Mel-шкала әдісі (MFCC) адамның есту мүшелерінің жұмыс істеу моделіне негізделген және мел жиілік шкаласын пайдаланады. Бұл мел шкала 1000 Гц дейін сызықтық болып табылатын адам құлағының жиілік сезімталдығын және логарифмдік мәндерде 1000 Гц-тан астам жиілікті модельдейді.

Бірінші амплитудалық коэффициент-спектрдің тұрақты құрамдас бөлігі еленбейді, ал қалған 255 спектрлік мәндердің амплитудасы орташаланады. Орташаландыру процесі 24 үшбұрышты жолақ өткізгіш сүзгілер ретінде іске асырылады.

Matlab-тан алынған сұлбалар топтамасының бірінші суретінде Asel.wav сөйлеу сигналы көрсетілген, ол 16 кГц жиілігімен және 16 разрядты разрядтылығымен дискреттелген.

Екінші суретте ұзақтығы 25,6 мс сөйлеу сигналының бір блогы (интервал) көрсетілген. Бұл интервалға 409 санау сәйкес келеді.

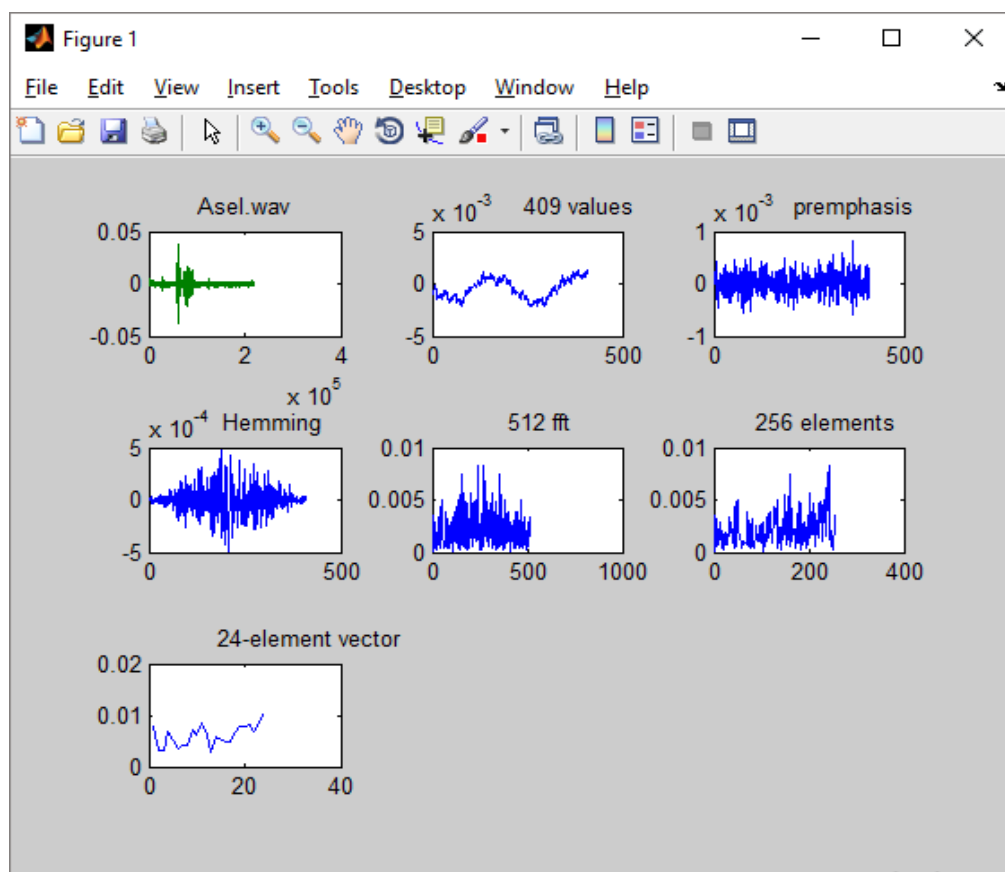
Үшінші суретте бірінші ретті сүзгімен өңдегеннен кейін сөйлеу сигналының бір блогын көруге болады.

Төртінші модель Хэмминг тересезін қолданғаннан кейінгі бір блокты көрсетеді.

Осы 1 блоктың жылдам Фурье түрлендірудің 512 амплитудалық мәнін 5-сурет береді.

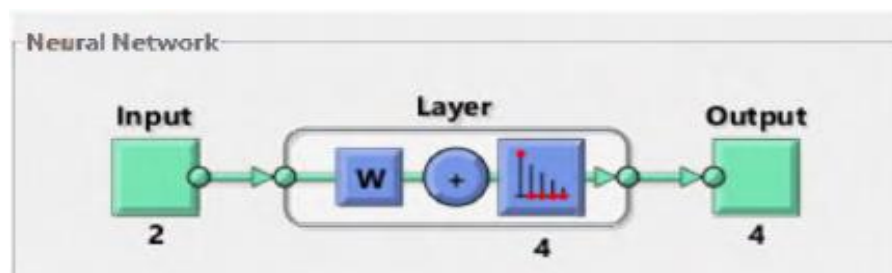
ЖФТ амплитудалық мәндері жоғарыда аталғандай қос-қостан сәйкес болғандықтан, 6-суретте көрсетілген алғашқы 256 амплитудалық мәндер ғана алынды.

Жетінші сурет 24-элементтік вектордың мәнін береді, оның компоненттері 24 "үшбұрышты" жиілік жолақтары шегінде 256 амплитудалық мәндерді орташалағаннан кейін алынған.



Сурет 2.2 – Алдын-ала өңдеудің кезеңдері

Жасалған жұмыс нәтижесінде Кохоненнің нейрондық желісі негізінде сөйлеу сигналын тану моделі ұсынылды, оның мысалы сурет 2.3-те көрсетілген. Кохоненнің нейрондық желілері аса күрделі жүйелерді жүзеге асыруға көмектеседі, өйткені бір жүйеде әр түрлі құрылыммен желілерді бір уақытта қолдану кезінде дикторға тәуелсіз жүйені жүзеге асыруға қол жеткізуге болады.

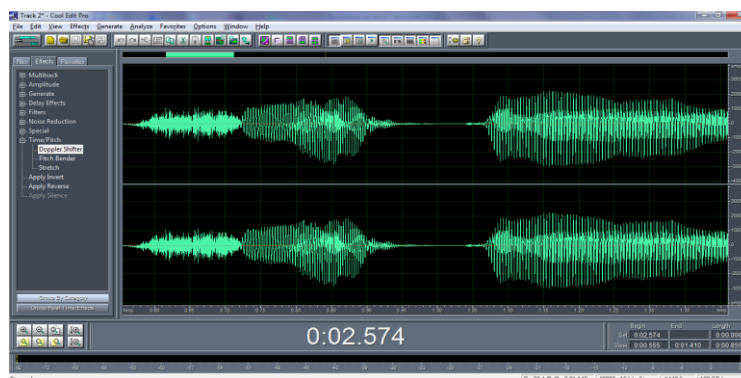


Сурет 2.3 – Кохененнің нейрон желісінің моделі

Жасанды нейрондық желілер негізінде құрылған сөйлеу сигналын айырудың осындай жүйелері өнеркәсіптік және тұрмыстық электрониканың түрлі салаларында қолданылуы мүмкін. Нейрондық желілер тану міндетін әлдеқайда жылдам атқарғандықтан және сөйлеу сигналын тану жүйесінің дәлдігін айтарлықтай арттыруға мүмкіндік береді.

2.2 Фонемалық дыбыстық сигнал кезінде өтпелі мәндерді талдау

Сөйлеу сигналының пішінін визуалды анықтау және оның амплитудасы, жиілігі, импульстердің ұзақтығы және олардың жүру кезеңі және тағы басқа параметрлерін талдау үшін компьютерге арнайы бағдарлама орната отырып, оны микрофоннан алынған электр тербелістерін зерттеу үшін сандық осциллограф ретінде пайдалануға болады. Микрофонға жеке дыбыстар мен сөздер айтыла отырып, осциллограммада сөйлеуді құрайтын жекелеген элементтерді қарап, сөйлеу сигналының амплитудасының өзгеруін және кідіріс ұзақтығын бақылауға болады. Зерттеу жүргізу үшін, ең алдымен, дыбыстық сигналдардың осциллограммаларын көруге мүмкіндік беретін бағдарламалық қамтамасыз ету қажет. Дауысты бағдарламада жазу үшін алдымен компьютер немесе ноутбукта аудиоплатаға қосылған микрофонның болуы қажет. Керекті параметрлерді таңдаған соң дауысты жаза беруге болады. Мысал ретінде "жеті" сөзі таңдалды [26].



Сурет 2.4 – "Жеті" сөзіне тән визуализация

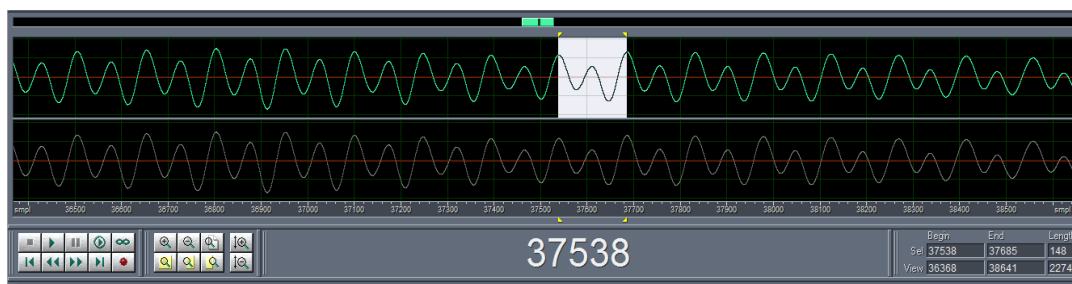
Белгілі болғандай, әрбір сөзді фонема деп аталатын шағын бөліктерге бөлуге болады. Сурет 2.5-те сөздің әр түрлі төрт фонемасын табуға болады. Байқау бойынша, фонемалар әр түрлі ұзақтыққа ие. Мысалы "і" фонемасының ұзақтығы аз. Олар сондай-ақ түрлі амплитуда мәніне ие. Сонымен қатар "е" және "і" фонемаларының "ж" және "т" фонемаларына қарағанда үлкен амплитудасы бар.

Егер сөйлеуді фонемдік элементтерге сегменттеу орындалса сөйлеуді тану немесе тану процестерімен байланысты міндеттерді шешу мақсатында фонетикалық элементке арналған сегмент шектерін анықталатын объектінің акустикалық дыбысталуына қарамастан фонем бейнесін дұрыс анықтау дәрежесін көтеретіндей етіп белгілеу қажет. Екінші жағынан, фонемдердің шекаралары сөйлеудің фонемалық құрылымдық элементтерінің базасын алу мақсатында оны сегменттеу кезінде тілдің құрылымдық элементі мәтінді сөзге түрлендіру немесе сөйлеу немесе сөзге ұқсас тізбектерді синтездеу үшін мүлдем басқа болады. Олар синтезделген сөйлеудің жоғары сапасын қамтамасыз етуі тиіс. Сонымен қатар синтезделген сөйлеу сапасына сөйлеушінің жеке ерекшеліктерін ескере отырып, сөйлеу просодикасын қалыптастыру әдісі үлкен әсер етеді.

Адамның сөйлеу аппараты сөйлегенде оның пішіні мен динамикасын орнатуға болмайды. Бір фонемалар нақты көрсетілген шекараларсыз басқаларына өтеді. Бір фонемадан екінші фонемаға көшкен кезде сөйлеу аппараты қайта құрылуы және келесі фонеманы құру үшін орналасу қажет, сондықтан бір фонемдік құрылымнан екіншісіне дәл өту шегін орнату мүмкін емес.

Сөйлеуді сегменттеу әдістерін талдау сөйлеуді компиляциялық әдіспен синтездеу үшін тілдің фонемдік құрылымдық бірліктерінің базасын қалыптастыру үшін динамикалық бағдарламалауды пайдаланатын сегменттеу әдісі неғұрлым ыңғайлы екенін көрсетті. Ол үшін біртұтас сөйлеудің базаға қажетті барлық фонемдік құрылымдық бірліктері бар фонетикалық жазбасы болуы керек. Әдетте бұл орыс, қазақ, белорус тілдеріне арналған 300-400 аллофондар және қытай тілі үшін шамамен 1200 фонемдік құрылымдық тіл бірлігі, өйткені ол тоналды.

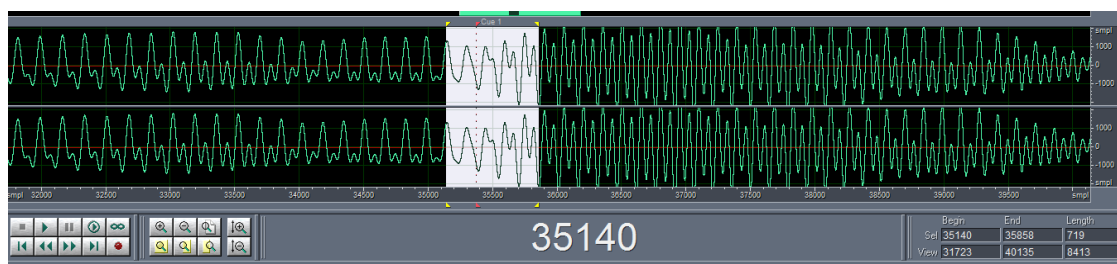
Дыбысталатын фонемаларды табу оңай. Себебі оларда қайталанатын негізгі кезеңдері болады. Бұл кезеңдер айтушының жынысы мен жасына тәуелді болатын ұзақтыққа (2-ден 10 мс-ге дейін) ие. Ерлерде (әдетте 7 ... 10 мс) әйелдер мен балаларға (әдетте 2..5 мс) қарағанда негізгі кезең ұзақтықтары көбірек болады. Енді «е» фонемасының уақыттық сипаттамасының үлкейтілген сұлбасына назар аударайық (Сурет 2.5). Мұнда сигналдың бірдей кескіні және ұқсас ұзақтықтары бар көптеген негізгі кезеңдер келтірілген (шамамен 3,7 мс). Ұзақтығы 3,7 мс-ге тең болғандықтан, бұл әйел не баланың дауыс жазбасы екені белгілі болады [2].



Сурет 2.5 – "Е" фонемасының уақыттық сипаттамасының үлкейтілген бөлігі

Көрші фонемалар арасындағы сигнал түрінің бірқалыпты өзгеруі дауыстың бір ерекшелігі болып табылады. Сөйлеудің фонетикалық бірліктерінің құрылымын талдауы көрсеткендей, фонеманың нақты және бір мәнді шекарасын белгілеу мүмкін емес, өйткені фонема анық көрінетін және өтпелі аймақ бар. Бұл аймақтар фонеманың алдында да, одан кейін де бар. Қоршаған ортаға байланысты өтпелі аймақтардың ұзындығы әртүрлі болуы мүмкін.

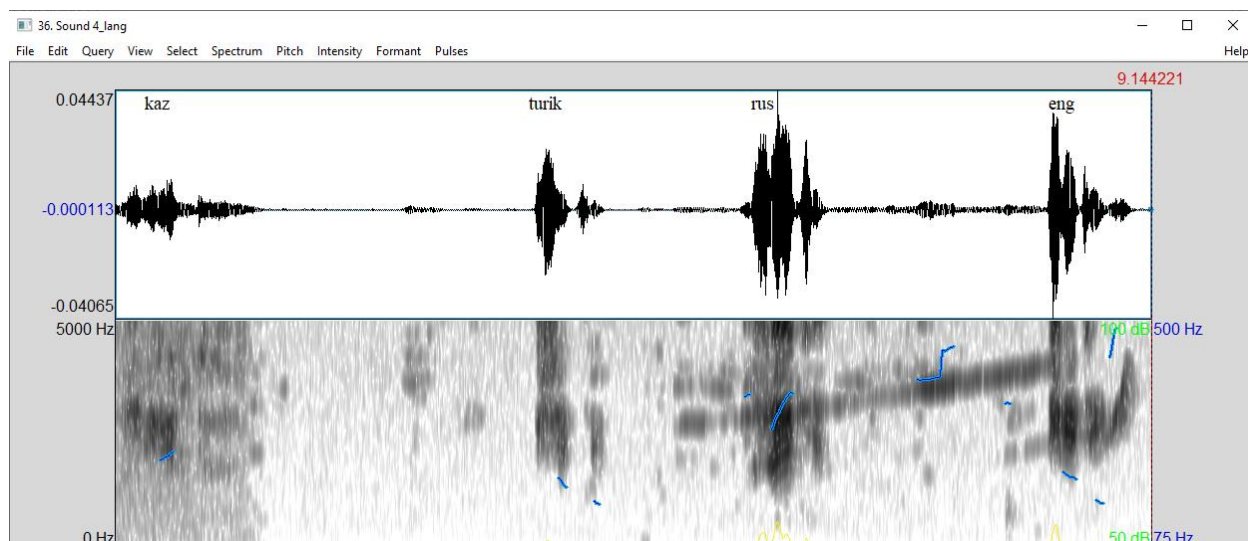
Сурет 2.6-да «жеті» сөзіндегі «ж» және «е» фонемалары арасындағы өтпелі аймақ көрсетілген. Өтпелі аймақтың ұзақтығы бұл жағдайда шамамен 14 мс құрайды.



Сурет 2.6 – «Жеті» сөзіндегі «ж» және «е» фонемалары арасындағы өтпелі аймақ

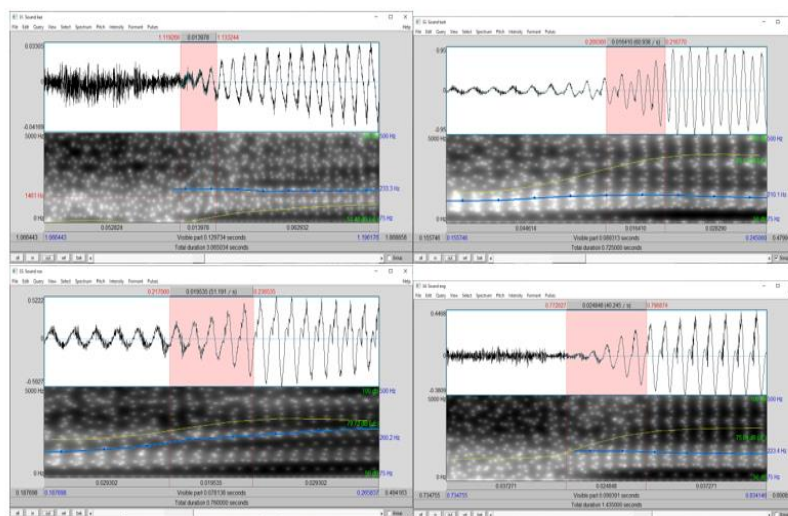
Бір фонеманың өтпелі аймақтарының шамасы қатты дәрежеде нақты диктор үшін сөйлеудің айтылу қарқынына, байланысқан фонемдердің түріне және 10-нан 50 мс-ге дейін құрайтыны анықталды.

Тәжірибе үшін осы «жеті» сөзін қазақ, түрік, орыс, ағылшын тілді дикторлардың айтуымен салыстырып қарастырайық. Рет-ретімен алынған осцилограмма сурет 2.7-де.

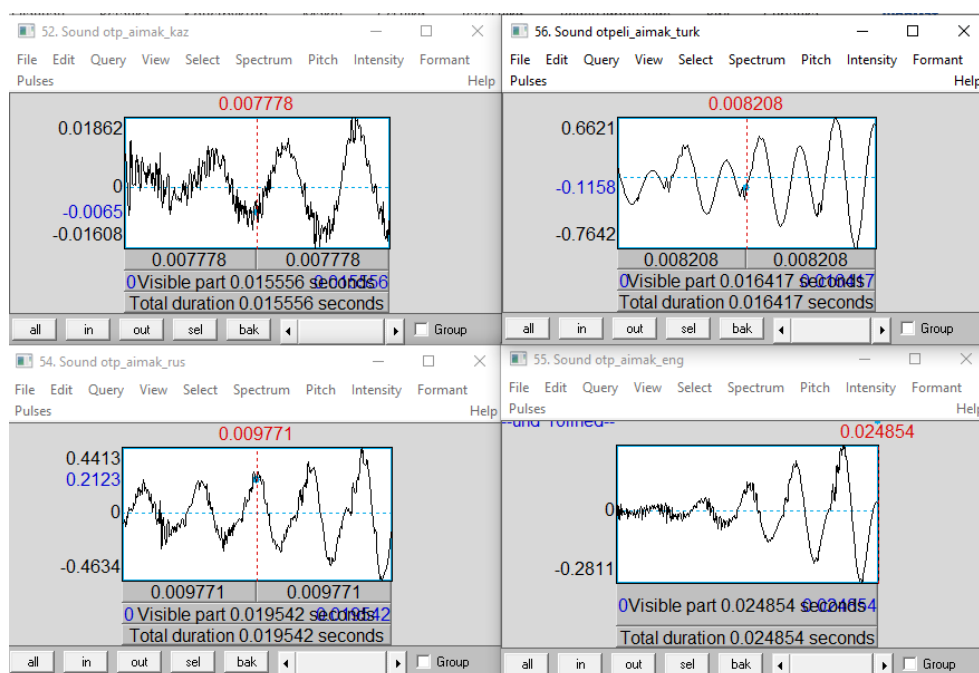


Сурет 2.7 – «Жеті» сөзін қазақ, түрік, орыс, ағылшын тілді дикторлардың айтылу осцилограммасы және олардың динамикалық спектограммасы

Суретте әр диктордың айтқан сөзінің спектрограммасын зерттеп қана қоймай, сонымен қатар қарқындылық (сары түспен) және негізгі тон жиілікті (көк түспен) байқауға болады. Осы жерден «ж» және «е» фонемалар арасында өтпелі аймақты 4 тілді дикторларда қалай өзгертетінін қарастырайық.



Сурет 2.8 – Бірдей сөзді айту кезіндегі фонемалар арасындағы өтпелі аймақ (жоғарыда солдан оңға қазақ және түрік, астыңғыда солдан оңға орыс және ағылшын)



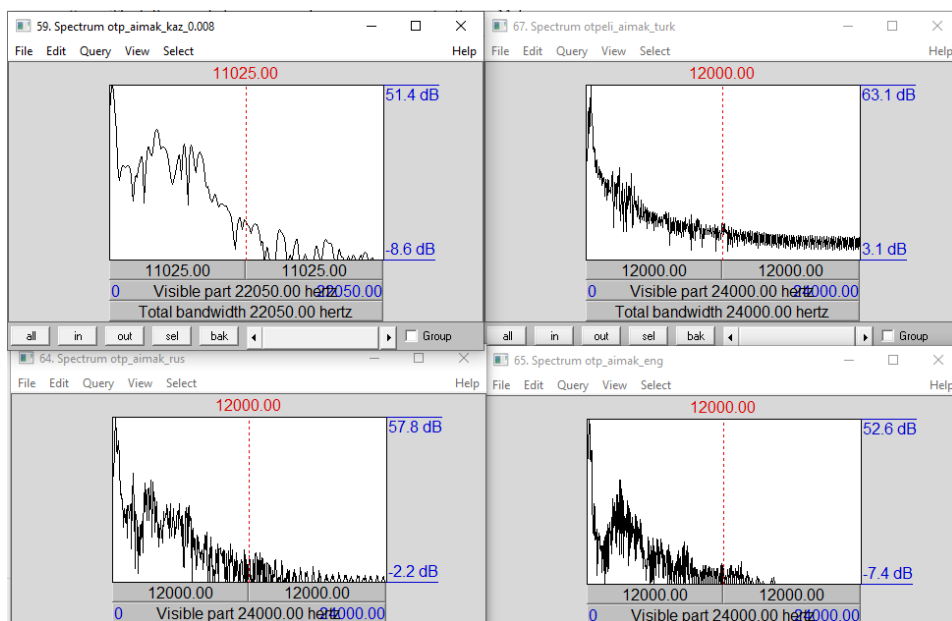
Сурет 2.9 – Бірдей сөзді айту кезіндегі фонемалар арасындағы өтпелі аймақтың жақындатылған бөлігі

Байқағанымыздай өтпелі аймақтардың әр дикторда сипаттамалры әртүрлі. Оның ішінде жалпы энергиясы да түрліше. Оны кесте құру арқылы байқауға болады.

2.1-кесте – Фонемалар арасындағы өтпелі аймақтардың 4 диктор үшін сипаттамалары

	қазақ	түрік	орыс	ағылшын
Энергиясы, $\text{Па}^2 \cdot \text{с}$	0.0024	0.0015	0.00065	0.00023
Қуаты, Па^2	0.15	0.09	0.03	0.009
Қарқындылығы, дБ	85.86	83.58	79.2	73.6

Әр сұлбадан алынған сипаттамаларды қарастыра отырып қарқындылықтың жиілікке тәуелділік графигін тұрғызайық.



Сурет 2.10 – Өтпелі аймақтың ФДТ сұлбалары

Ендігі кезекте шулы фонемалар өзінің уақыттық сипаттамаларына негізгі кезеңдерді енгізбейді, бірақ көптеген тұрақты емес сигналдарды қамтиды. Қазақ тілінде мұндай фонемалар бар. Шулы фонеманың уақыттық сипаттамасының мысалы Сурет 2.11-де көрсетілген.



Сурет 2.11 – Шулы фонема және оның жақындатылған формасы (оң жақта)

Байқағанымыздай, бірдей негізгі кезеңдер қайталанбайды, сигнал амплитудасын жиі өзгертеді, кейде күшті сигнал кездейсоқ пайда болады.

2.3 Ағзаның жынысы мен құрылысына байланысты сөйлеу жүйесінің параметрлері

Сөйлеу процесінде шамамен 10-15 % өкпе көлемі пайдаланылады. Көмей VI-VII мойын омыртқаларының шекарасында кеңірдекке өтеді, ерлерде бұл деңгей төмен, әйелдерде-жоғары. Кеңірдектің ұзындығы 17 см жетуі мүмкін. Кеңірдектің қимасының ауданы 2-ден 3 см²-ге дейін жатыр. Кеңірдектің орташа ұзындығы 9-13 см, ал диаметрі 15-18 мм.

Ерлерде әйелдерге карағанда кеңірдектің орташа мөлшері көп.

Кеңірдектің диаметрінің жоғарғы шегі ерлер үшін-25-27 мм, әйелдер үшін - 21-23 мм. Төменгі шегі ерлер үшін - 13 мм, әйелдер үшін-10 мм, әйелдерде кеңірдектің ені орта есеппен 10% - ға аз, кеңірдектің ұзындығы 6-12% - ға аз. Осылайша, кеңірдектің ұзындығындағы айырмашылығы шамамен 10 мм-ге тең. Кейбір жағдайларда кеңірдектің ұзындығы жынысына байланысты емес деп есептеледі.

Ерлерде дауыстық қатпарларда ұзындық бар. Ерлерде дауыс қатпарларының ұзындығы-20-24 мм, сондықтан дауыс төмен; әйелдерде-18-20 мм, және дауыс жоғары. Мысалы, тенор-қатпарлардың ұзындығы-15-17 мм, дыбыстық тербелістер саны – секундына 580, немесе баритон - 18-21 мм - 426. Бұл, өз кезегінде, дыбыстарды айту кезінде түрлі резонанстық жиілікке ие болады. Ерлерде орта есеппен олар 91-145 Гц, ал әйелдерде-115-145 Гц құрайды. Сондай - ақ, дауыстық қатпарлардың сапалылығы-2,1-5,1 ерлерде және 2,6-7,1 әйелдерде айтарлықтай ерекшеленеді. Ерлердің дауыстық қатпарларының ені әйелдерден көп және әйелдерде 2,5-3 мм және ерлерде 3-3,5 мм тең.

Нәрестенің көмейі үлкен өлшемдері бар; ол қысқа, кең, сүзгі тәрізді, ересектерге қарағанда жоғары (II-IV омыртқа деңгейінде) орналасқан. 7 жасқа қарай көмейдің төменгі шеті VI мойын омыртқасының жоғарғы шеті деңгейінде орналасқан. Ересек адамдарда көмей IV-VI-VII мойын омыртқаларының деңгейінде орналасқан.

Ендігі кезекте «жеті» сөзінің ер адам, әйел және бала дауысымен салыстырып қарайық.



Сурет 2.12 – Жыныстық ерекшелікке байланысты визуализацияның өзгерісі

Қоршауларда жазбаның төрт түрлі бөлігі бар. Біріншісінің амплитудасы аз, екіншісінде көбірек, үшіншісі аздау және төртінші көбірек амплитудаға ие. Бұл жазбалар әртүрлі адамдардан келіп түссе де, олардың ішінде әртүрлі бөліктердің бірдей саны бар. Әр бөліктің максималды ұзақтықтары өзгеруі мүмкін, бірақ ол өзгерістерді анықтау қиын. Алайда ғалымдардың зерттеулерінде дәлелденген ең аз ұзақтықты табу мүмкіндігі бар.

Ендігі кезекте дыбыстық редакторда спектрлік анализ жасау барысында бағдарламаның DeviceControls терезесінде сигналдың формасын және спектрін аламыз.

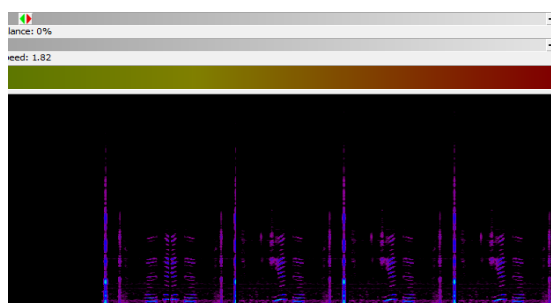
Сурет 2.13 - те байқағанымыздай, «жеті» сөзінің айтылу кезінде дыбыстық сигналдың спектрі графикалық түрде және диаграмма түрінде берілген. Диаграмма төртбұрыштары жоғары болған сайын ондағы сәйкесті жиіліктердің соншалықты жоғары екенін көрсетеді.



Сурет 2.13 – Графикалық және диаграмма түріндегі спектр

Дыбыстардың параметрлерін анықтау үшін сарапшы сөйлеу сигналдарын сегменттеуді жүргізеді және дисплей экранында бейнеленген сөйлеу толқынында, дауысты және олардың үнсіз дыбыстарын қоршаған ортаны белгілейді. Сегменттеудің дұрыстығын сөйлеушінің сөйлеудің бөлінген фрагменттерін тыңдау арқылы және спектрограммада, сондай-ақ дисплей экранында бейнеленген сөйлеу спектрін талдау арқылы бақылайды.

Спектрограмма (сонограмма) — бұл "уақыт-жиілік-қарқындылық" координаталарындағы сөйлеу сигналының бейнесі. Спектрограмма белгілі бір уақыт аралығында сөйлеу сигналының фрагментінің спектрлік құрамын көрсетеді. Графикте ордината осі бойынша маңызы бар спектрлік жиілік жинақталады, ал абсцисса осіне — уақыт жинақталады. Қарқындылық түстің жарықтық градациясымен сипатталады.



Сурет 2.14 – Дыбыстық сигналдың спектограммасы

Сурет 2.14-те байқағанымыздай, дыбыс сигналының спектрі уақыт бойынша өзгереді. Бұл ретте көлденең ось бойынша уақыт, ал тігінен сигнал жиілігі алынады. Қарқындылыққа келетін болсақ, ол түспен белгіленеді. Кіші қарқындылығы бар жиілік қара түспен бейнеленеді, содан кейін күлгін, көк, циан, жасыл, сары және қызыл түсті болады. Қызыл түс максималды қарқындылыққа сәйкес келеді.

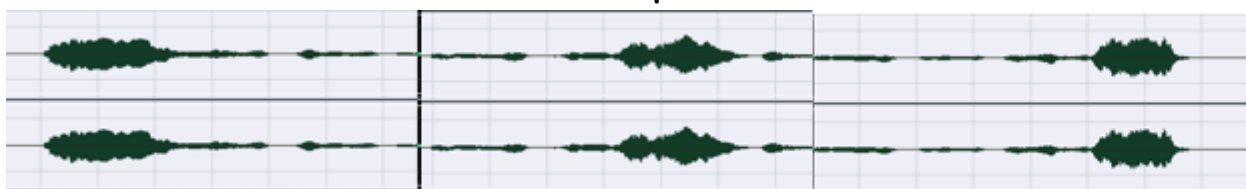
Келесі ретте осциллограммада интонацияның өзгерісі қалай байқалатынын қарастырайық. Ол үшін кемінде 3 сөзден құралатын сөйлем қажет. Мысал ретінде «Айнала қандай тамаша» сөйлемін алайық. Енді интонацияны бірінші сөзге түсіргенде, одан екіншісі, одан үшіншіге қоямыз:

[Айнала] қандай тамаша!

Айнала [қандай] тамаша!

Айнала қандай [тамаша]!

Нәтижесінде интонацияның қай сөзге түсірілгеніне байланысты амплитуда өзгерісін байқауға болады.



Сурет 2.15 – Әр сөзге қойылған интонацияның сигналға әсері

Бірінші жағдайда қай жердің тамаша екенін ерекшелеп, айналаның қандай тамаша екенін анықтаймыз.

Екінші жағдайда күшейтпелі шырайды қолдана отырып айналаның сондай тамаша екенін ерекшелейміз.

Үшінші жағдайда айналаның қандай екенін сұрау арқылы оның сондай тамаша екенін белгілейміз. Әр жағдайда түсірілген интонацияға байланысты сәйкесінше сөздердің амплитудасының күрт өсуін байқауға болады.

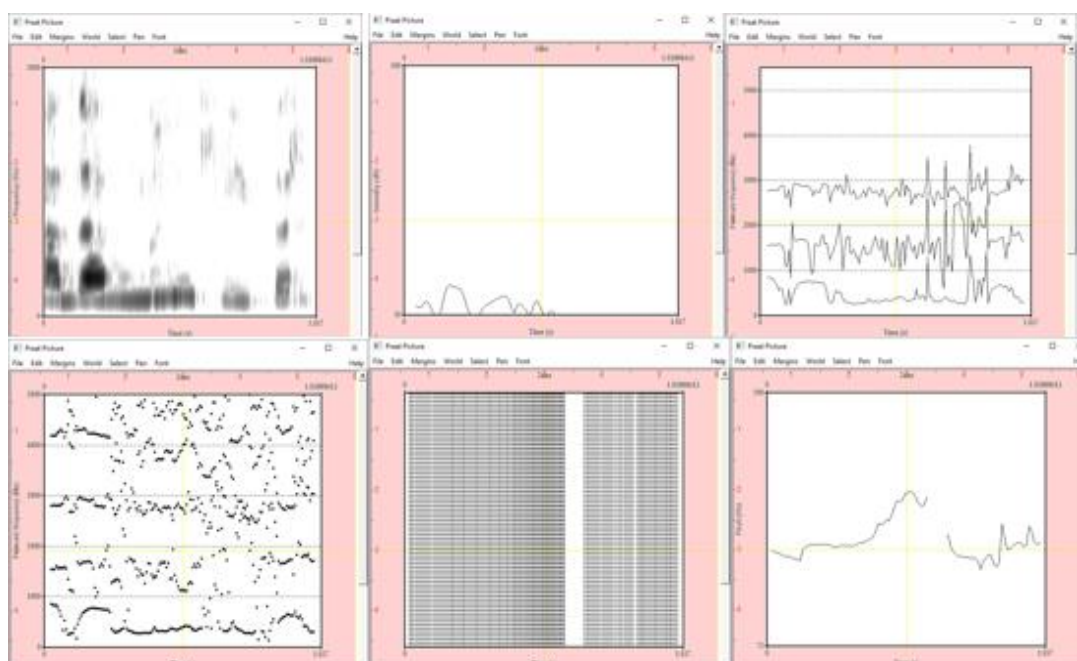
Сөйлеу синтезаторларын жобалау кезінде интонацияны ескеру маңызды.

2.4 Қазақ дауысты дыбыстарының акустикалық мәндерін зерттеу

Негізінен фонетикада жиі қолданылатын акустикалық талдауы қазіргі таңда лингвистикада да қажетті салалардың біріне айналды. Осыған орай бірнеше мысал келтіруге болады: фонологтар кейде теориялық дәлелдерді негіздеу үшін акустикалық деректерге жүгінеді, социолингвистер дауыстардың акустикалық қасиеттері тұрғысынан жылжуын және бірігуін сипаттауға бейім, ал психолингвистер эксперименттер үшін ынталандыруларды құру үшін акустикалық талдау әдістерін жиі қолданады.

Толқын формасының астындағы сұр реңктердегі сурет әр уақытта толқын формасының жиіліктік мазмұнын білдіретін спектрограмма болып табылады. Көру ауқымы 0-ден 5000 Гц-ке дейін орнатылған. Спектрограммаға салынған көк қисық уақыттың әр сәтінде толқын нысанында кезеңділікті білдіретін негізгі тон жиілігінің қисығы болып табылады. Көру ауқымы 75-тен 500 Гц дейін орнатылған, бұл сөйлеу үшін стандартты диапазон.

Дыбыс терезесінде көрінетін басқа талдаулар автоматты түрде форманттар, қарқындылық қисығы және дауыс импульстері өлшенеді. Осы сипаттамалардың жекелеген графикалық көрінісін әр командадағы Draw, яғни сызу батырмасын басу арқылы аламыз.



Сурет 2.17 – Сипаттамалардың графикалық көрінісі

Эмоцияларды тану саласында дауыс – эмоциялық деректердің екінші көзі. Дауысты бірнеше параметр бойынша сипаттауға болады. Дауыстың биіктігі-осындай негізгі сипаттамалардың бірі, алайда акустикалық технологиялар саласында бұл параметрді негізгі тонның жиілігімен атау дұрыс.

Негізгі тонның жиілігі интонациямен тікелей қатысы бар. Ал интонация, мысалы, дауыстың эмоциялық-экспрессивті сипаттамаларымен байланысты. Негізгі тонның жиілігі сөйлеушінің дауыстық байламдарының жеке анатомиялық сипаттамаларымен, атап айтқанда, байлам ұзындығымен, олардың қалыңдығы мен икемділігімен байланысты. Дауысты дыбыстардағы негізгі тонның жиілігінің вариативтілік шамасын анықтау үшін F0 мәні тиісті дауыстың басында, оның ортасында және соңында есептеледі [8].

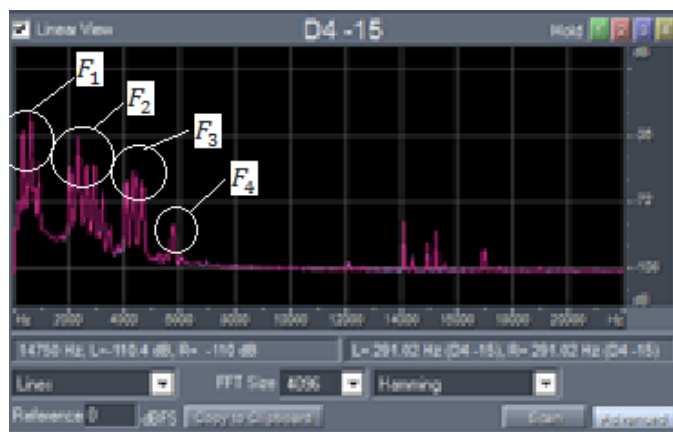
2.5 Сөйлеу сигналын құраушы фонемдердің форманттарын алу

Форманттар сөйлеу сигналының қуат спектрінің айқын максимумдары деп аталады және сөйлеу дыбыстарының акустикалық сипаттамасын білдіретін резонанстық жиіліктері болып табылады. Ол мәтін мазмұнын анықтауға мүмкіндік береді. Олар резонаторлық қуыстар арқылы өткен кезде дыбыс толқынының кейбір гармониктері күшейіп, басқалары әлсірейді [1].

Формантты сипаттамалар сөйлеушінің сөйлеуші органдарының жеке параметрлерін көрсетеді (атап айтқанда, олар ауыздың, тілдің, тістердің формасымен және өлшемдерімен байланысты) және ауызша сөйлеуді түсіну үшін сөйлеу дыбыстарының параметрлерімен ең маңызды болып саналады.

Дыбыстың формантты құрылымы нақты тілде дыбысты айтуда белгілі бір индивидтің артикуляторлық дағдыларын көрсетеді және дыбыстың лингвистикалық табиғаты (тіл аясы) туралы ақпаратты қамтиды.

Формант спектрлік графикте амплитудалық қозу ретінде қарастырылады (Сурет 2.18). Әр дыбысты айтқанда шамамен 5 формант пайда болады, олардың әрқайсысы сөйлеудің әр түрлі аспектілерін сипаттайды. Форманттар F әрпімен белгіленеді; олардың нөмірленуі нөлдік форманттан басталады – негізгі тонның жиілігі (F0), одан әрі бірінші (F1), екінші (F2), үшінші (F3) және төртінші (F4) форманттар жүреді.



Сурет 2.18 – «Ж»дыбысының форманттары

Мысал ретінде орташа дауысты дыбыстың жиілік форманттарын есептеуді қарастырайық. Ол үшін келесі формуланы қолданамыз:

$$F_n = \frac{(2n-1)c}{4l} \quad (2.4)$$

мұндағы l- дауыстық тракт, ер адамда l=17,5см, әйелде l=15 см және сәбиде l=8-10 см;

C-дыбыс жылдамдығы:C=350 м/с; n=1,2,3.

Ер адамда:

$$F_1 = \frac{(2 \times 1 - 1) \times 350}{4 \times 17 \times 0,01} = 500 \text{ Гц}; F_2 = 1500 \text{ Гц}; F_3 = 2500 \text{ Гц}$$

Ары қарай әйел және сәбидің сөйлеу форманттарының жиіліктерін алып, 2.2-кестедегідей мәндерді аламыз.

2.2-кесте – Есептелген жиілік форманттар

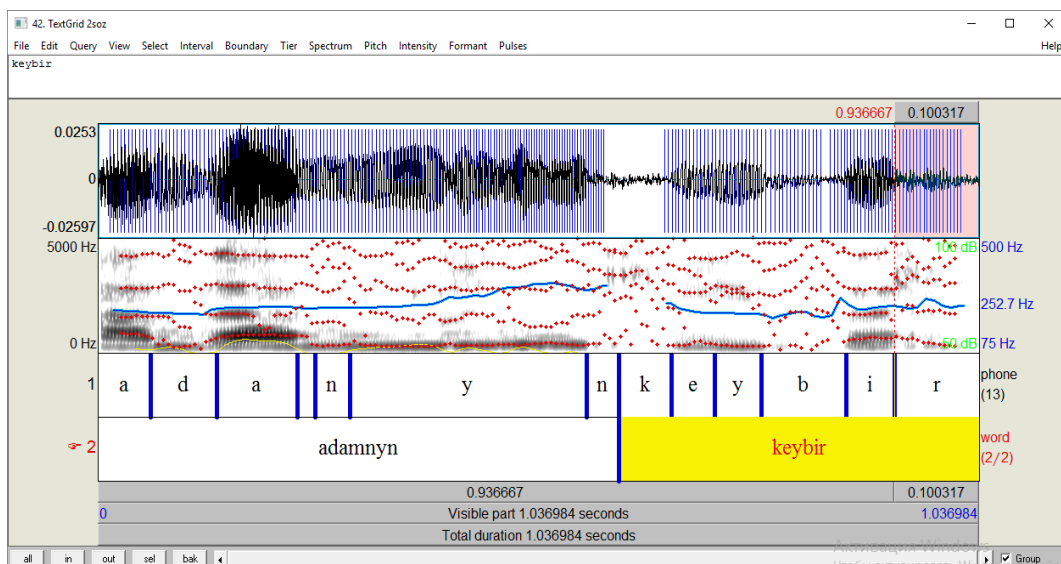
Нейтралды дауысты дыбыс	Сөйлеу форманттарының жиіліктері		
	F1	F2	F3
Ер адам	500	1500	2500
Әйел	583	1750	2917
Сәби	1094	3281	5469

Сонымен, сөйлеу дыбысы – сөйлеу жолында пайда болатын, резонанс әсерінен туындайтын спектралды құрамдастардың түрленуі. Жоғарыда қарастырғандай, сөйлеу жолының әр түрлі ұзындықтағы және формалардағы резонанстары форманттар болып табылады. Дауысты дыбыстылардың ең маңызды сипаттамалары 1,2,3 –ші форманттар [3]. Бұл жиіліктерде дауысты энергетикалық спектрдің басым бөлігі шоғырланады. Формант мәнін анықтау дауысты дыбыстың ортасында - оның квазистационарлық бөлігінде жүзеге асырылады.

Бағдарламалық сегменттеу нәтижелері бойынша дауысты T_d және дауыссыз $T_{дз}$ ұзақтығын, негізгі тонның F_0 жиілігінің мәндерін, $F_1 - F_4$ формантты жиіліктерді есептеу жүргізіледі. Параметрлердің мәндері белгіленеді, содан кейін олардың негізінде арнайы бағдарламалық модульдің көмегімен дыбыстардың микроанализінің белгілері есептеледі.

Сөйлеу бірліктерінің уақыт параметрлері оның жеке құраушыларының ауызша мәтініндегі ұзақтығы мен таралуын сипаттайды. Зерттеушілердің айтуынша, салыстырмалы көлемдерде көрсетілген тілдің ырғақты ұйымдастырылуына айтылу қарқынының өзгеруі әсер етпейді.

Қайта зерттеу нысанына оралатын болсақ, бағдарламада сигналдың құраушы фонемдерін жеке-жеке және форманттарын алуды қарастырайық.



Сурет 2.19 – Сигнал және оны құраушы фонемдер

2 сөзден 13 дыбыс енгізілген соң, Praat Objects терезесінде Praat - New Praat Script командасы арқылы ашылған терезеде белгілі, алдын-ала жазылып дайындалған сценарийді енгіземіз. Ол сценарий бойынша сөйлеуді құраушы фонемдердің форманттарын аламыз.

```

untitled script (modified)
File Edit Search Convert Font Run

writeInfoLine: "Extracting formants..."

# Extract the names of the Praat objects
thisSound$ = selected$("Sound")
thisTextGrid$ = selected$("TextGrid")

# Extract the number of intervals in the phoneme tier
select TextGrid 'thisTextGrid$'
numberOfPhonemes = Get number of intervals: 1
appendInfoLine: "There are ", numberOfPhonemes, " intervals."

# Create the Formant Object
select Sound 'thisSound$'
To Formant (burg)... 0 5 5000 0.025 50

# Create the output file and write the first line.
outputPath$ = "C:\Users\Dinara\Desktop\formanty.txt"
writeFileLine: "'outputPath$'", "file,time,word,phoneme,F1,F2,F3"

# Loop through each interval on the phoneme tier.
for thisInterval from 1 to numberOfPhonemes
  appendInfoLine: thisInterval

  # Get the label of the interval
  select TextGrid 'thisTextGrid$'
  thisPhoneme$ = Get label of interval: 1, thisInterval
  appendInfoLine: thisPhoneme$

  # Find the midpoint.
  thisPhonemeStartTime = Get start point: 1, thisInterval
  thisPhonemeEndTime = Get end point: 1, thisInterval
  duration = thisPhonemeEndTime - thisPhonemeStartTime
  midpoint = thisPhonemeStartTime + duration/2

  # Extract formant measurements
  select Formant 'thisSound$'
  f1 = Get value at time... 1 midpoint Hertz Linear
  f2 = Get value at time... 2 midpoint Hertz Linear
  f3 = Get value at time... 3 midpoint Hertz Linear

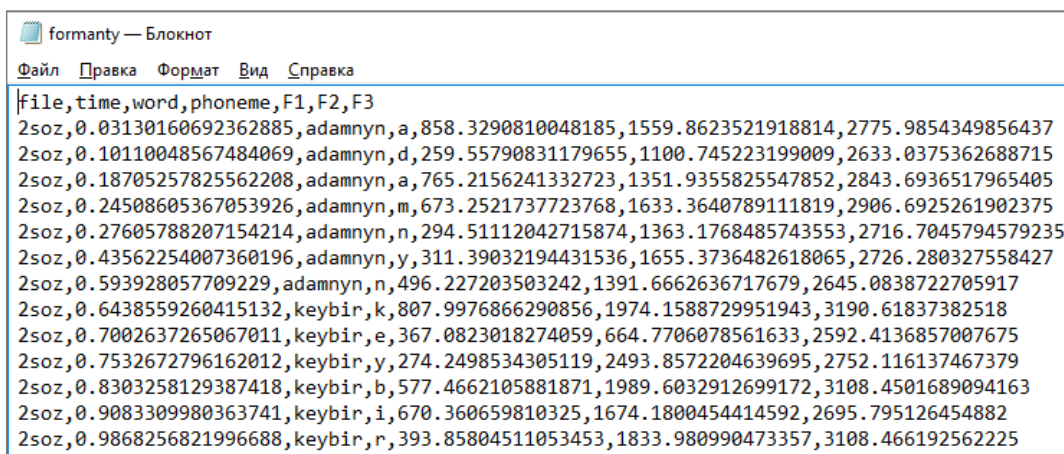
  # Get the word interval and then the label
  select TextGrid 'thisTextGrid$'
  thisWordInterval = Get interval at time: 2, midpoint
  thisWord$ = Get label of interval: 2, thisWordInterval

  # Save to a spreadsheet
  appendFileLine: "'outputPath$'",
    ...thisSound$, ",",
    ...midpoint, ",",
    ...thisWord$, ",",
    ...thisPhoneme$, ",",
    ...f1, ",",
    ...f2, ",",
    ...f3

```

Сурет 2.20 – Орындаушы скрипт мазмұны

Нәтижесінде алдын-ала сақталған бос txt файлды ашқанда алынатын мәндер келесі суретте көрсетілген.



Сурет 2.21 – Алынған нәтиже

Мұнда файл аты, әр фонема тіркелген уақыт, сөз, фонема және 1–ші, 2–ші, 3–ші формант жиіліктері рет-ретімен және мәндерімен келтірілген.

2.6 Дауысты дыбыстар форманттарының шығыс қуыс, жұтқыншақ және ауыз резонатор ұзындығына тәуелділігі

Дауысты дыбыстың бірінші формантының мәні шығыс (ауыз) қуыстың мөлшеріне тікелей пропорционалды (қуыс кең болса, 1–ші формант та жоғары) және жұтқыншақ қуысының көлеміне кері пропорционалды (жұтқыншақ қуысы үлкен болса, соғұрлым F1 төмен). Артикуляциялық тұрғыдан бұл екі параметр де дауысты артикуляциялау кезінде тілдің көтерілуімен анықталады. Екінші форманттың мәні ауыз резонаторының ұзындығына кері пропорционалды, ол дауыстының қатарына байланысты келеді. Мысал үшін [a], [ы], [e] дауыстыларды салыстырайық. Тілдің төмен көтерілуінен пайда болатын [a] дауыстының шығыс ауыз қуысы максималды үлкен, ал жұтқыншақ көлемі минималды аз, сондықтан F1 үлкен мәнге ие. Керісінше [ы], [e] дауыстылардың шығыс ауыз қуысы минималды аз, ал жұтқыншақ көлемі максималды үлкен, сондықтан F1 аз мәнге ие.

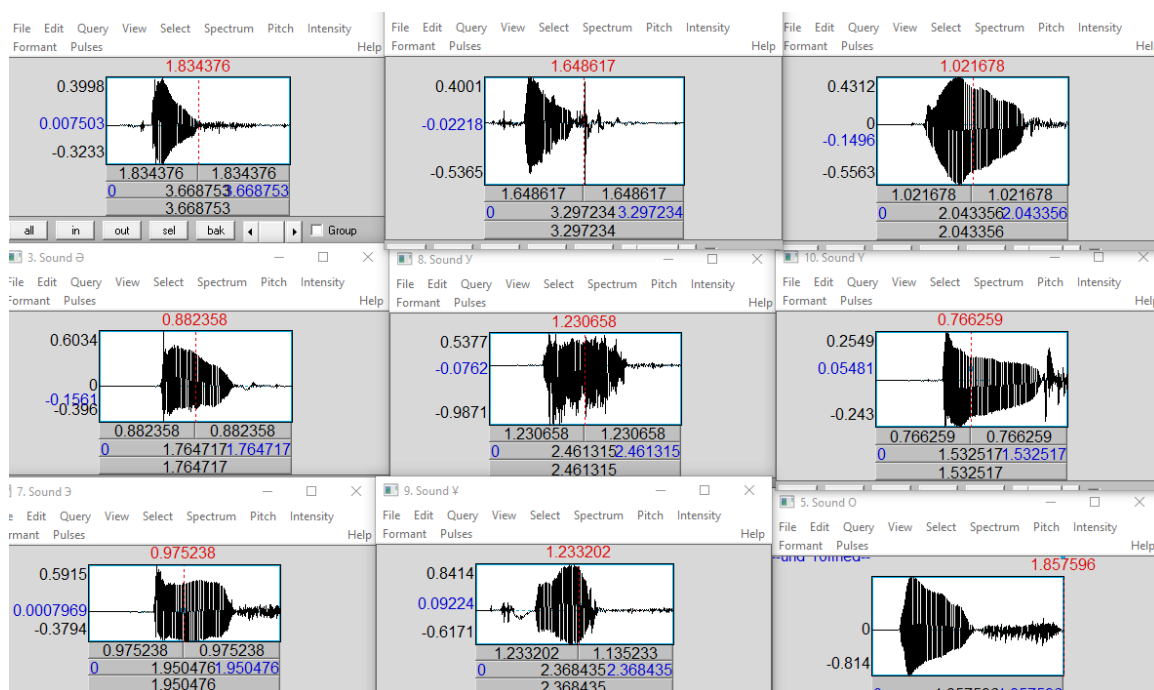
Ал ауыз резонаторы ұзындығына келер болсақ (ауыздың ең тар бөлігінен ерінге дейінгі қашықтық), ол алдыңғы дауысты дыбыста [ы] минималды аз болады (және оның екінші форманты мәні үлкен) және артқы дауысты дыбыста максимум мәнде [мысалы у], ортаңғы дауысты дыбыс [a] аралық позицияны алады (2.3-кесте). Осыған орай алынған сөздердегі дауысты дыбыстарды айта отырып және форманттарына қарай отырып келесідей кесте құрамыз.

2.3-кесте – Дауысты дыбыстар форманттарының шығыс қуыс, жұтқыншақ және ауыз резонатор ұзындығына тәуелділігі

	[a]	[ы]	[e]	[i]
Шығыс қуыстың мәні	үлкен	аз	аз	орташа
Жұтқыншақ қуысының көлемі	аз	үлкен	үлкен	орташа
F1	жоғары	төмен	төмен	орташа
Ауыз резонаторының ұзындығы	орташа	аз	үлкен	орташа
F2	орташа	жоғары	аз	орташа

Нәтижесінде форманттардың әр адамда түрліше болатынын, алайда жалпы форманттық құрылымда [a] дыбысы үшін жоғары F1 мен орташа F2, [ы] дыбысы үшін төмен F1 мен жоғары F2, [e] дыбысы үшін төмен F1 мен аз мәнді F2, [i] дыбысы үшін орташа F1 мен орташа F2 сақталатыны белгілі болды. Барлық дауысты дыбыстардың нөлдік форманты болатынын есте ұстаған жөн – яғни негізгі тон жиілігі, бірақ ол дауысты дыбысты айту кезінде дауыстық трактінің қасиеттеріне тәуелсіз.

Ендігі қазақ тілі дауыстылардың 1-ші және 2-ші форманттарының моделін аламыз. Стандартты график бар болғанымен, әр адам түрліше сөйлейтін болғандықтан модель де әркімде түрліше болады. Осы тұрғыда өз даусыммен айтылған дыбыстардан алынған форманттар бойынша зерттейік.

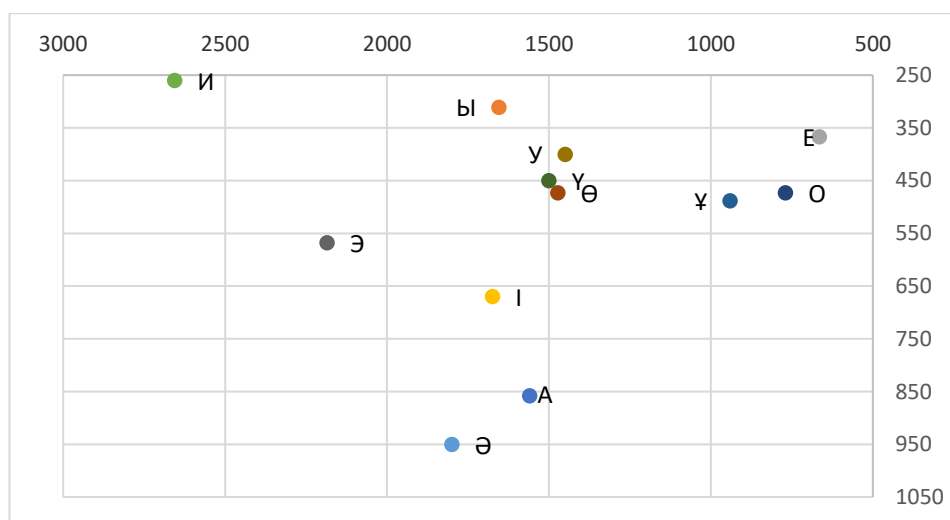


Сурет 2.23 – Зерттелген дауысты дыбыстылар осцилограммасы

Форманттарды кестеге Excel бағдарламасына жаза отырып, вертикаль бағытта F1-дің, горизонталь бағытта F2-нің қиылысынан алынған нәтижені көруге болады.

2.4-кесте – Қазақ тілінің дауысты дыбыстарының форманттары

	А	Ы	Е	І	Ә	И	О	Ө	Э	У	Ұ	Ү
F1	858	311	367	670	950	260	473	473	568	400	488	450
F2	1560	1655	665	1674	1800	2656	770	1473	2185	1450	941	1500



Сурет 2.24 – Алынған нәтиже

Жалпы резонатор ұзындығының ұлғаюы және шығыс қуыстың төмендеуі барлық форманттардың азаюына әкеледі, ал оның төмендеуі – жоғарылауға әкеледі. Ауыз қуысының алдыңғы бөлігінің тарылуы F2 ұлғаюымен F2 және F3 конвергенциясына әкеледі, ал артқы жағындағы тарылу - F2 төмендеуімен F2 мен F1 конвергенциясына әкеледі. Жиілік бойынша екі формант арасындағы қашықтықтың азаюы олардың амплитудасын арттырады, бұл кейде жалпы максимумның пайда болуына әкеледі (бұл F1 және F2 [y] дауысты дыбысына өте тән.

3. Қазақ тілі фонемдерін тануды бағдарламалық іске асыру

Бұл тарауда сөйлеудің компьютерлік тануының жалпы сұрақтары қарастырылады, Windows жүйесінде дыбысты бағдарламалау, MS Visual Studio 2017-де C# бағдарламалау тілі көмегімен және орнатылуы тиіс Microsoft Speech Platform SDK қосымшасы арқылы сөйлеуді тану іске асырылады. Нәтижесінде практикалық қолданыс мысалы ретінде компьютерді өшіру бұйрығы орындалады.

3.1 Бағдарламаның жұмыс сипаттамасы

Бұл жасалатын жұмыста «Сөйлеу» қызметінің SDK дестесі көмегімен микрофоннан алынған сөйлеуді интерактивті тану және аудиожаубадан тексттік жазбаны алуды қарастырамыз. Бұл функцияны әңгімелесулерді тану сияқты жалпы тану тапсырмаларын орындау үшін қолданбаларға немесе құрылғыларға оңай біріктіруге болады. Ол сондай-ақ дауыс көмекшілерін жасау үшін "сөйлеу" қызметі үшін SDK пакетімен Bot Framework пайдалану сияқты күрделі интеграциялық үшін жарамды.

Компьютерге микрофон арқылы сөйлеуді тану үшін алдымен Microsoft ұсынған дауыс компоненттерді жүктеп орнату қажет. Олар:

- Microsoft Speech Platform-Runtime v.11.0;
- Microsoft Speech Platform-SDK;
- Microsoft Speech Platform-Runtime Languages;
 - MS Visual Studio Express.

Microsoft Speech Platform — (Microsoft сөйлеу платформасы) - бұл әзірлеушілерге сөйлеу технологияларын қолдайтын қосымшалар мен сервистерді (сөйлеуді тану және мәтін бойынша сөйлеуді синтездеу) жасауға, ал пайдаланушыларға осындай қосымшалармен өзара әрекеттесуге мүмкіндік беретін бағдарламалық модульдер мен өңдеу құралдарының жиынтығы. Ресми анонстарда сөйлеу платформасы Microsoft Speech API (MS SAPI) жалпы атауымен белгілі Microsoft сөйлеу технологиясының дамуы белгілі, бірақ бұл MS SAPI бар-жоқтығына қарамастан жүйеде орнатуға болатын өте дербес бағдарламалық өнім.

Microsoft Speech Platform-SDK— бұл сөйлеуді танумен жұмысты бастау үшін қажет платформа және құрал-саймандары бар қажетті десте.

Сөйлеу платформасы Microsoft Speech Platform Runtime орындау уақытының құрамдастарын қамтиды, олар платформаны қолданғысы келетіндерге орнатылуы қажет; тек бағдарламалық жасақтаманы жасаушыларға қажет сөйлеу бағдарламалары мен Microsoft Speech Platform SDK қызметтерін әзірлеу құралдары; сондай-ақ пайдаланушының таңдауы бойынша орнатылатын Language Runtime тілдік модульдері бар. Қазір, 11.0 нұсқасының сөйлеу платформасында 26 тіл, олардың ішінде орыс тілі де бар.

MS Visual Studio Express - Microsoft компаниясы әзірлеген Microsoft Visual Studio жеңілдетілген нұсқасы, ASP.NET технология негізінде Web-қосымшаларды, сайттарды, әртүрлі веб-сервистерді құру құралы.

Бір айта кетерлігі Microsoft Speech Platform-Runtime Languages орнату барысында тілдік базадан MS_SR-ru-ru инсталляторын таңдау қажет, яғни өзімізге жақын тілді қолдану ұсынылады.

Кілт пен жазылым аймағы бар Speech Config нысанын жасаңыз.

Speechrecognizer нысанын жоғарыда келтірілген SpeechConfig нысанын пайдаланып жасаңыз.

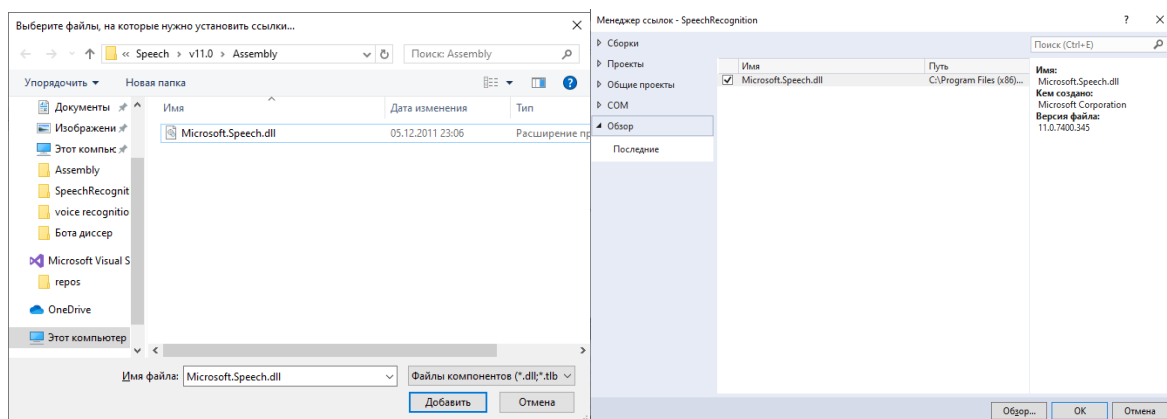
SpeechRecognizer көмегімен бір сөйлеу үзіндісі үшін тану процесін іске қосыңыз.

Қайтарылған нәтижені тексеріңіз SpeechRecognitionResult.

3.2 Visual Studio-дағы сөйлеу базасына байланысты жобаны іске асыру қадамдары

Компьютерге MS Visual Studio Express 2017 орнатылған. Онда жобаны бастау үшін Windows Forms қосымшасын таңдап, жобаны «SpeechRecognition» деп атап, алдын-ала сақтап қойдық.

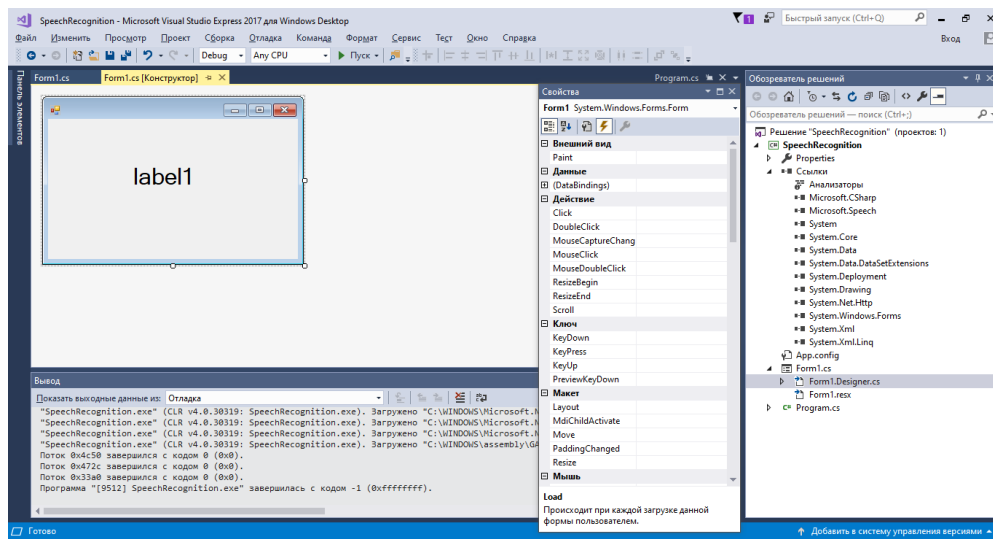
Ары қарай сілтемеден дауысты тану үшін қажет MS.Speech.dll кітапханасын жүктеп қою қажет. Ол үшін сілтеме батырмасын басып сілтеме қосу командасын таңдаймыз. Ары қарай шолуды басып Windows(C:) кіріп бағдарламалық файлдар ішінен Microsoft SDKs одан ары Speech содан v.11.0 таңдай келе Assembly ішінен MS.Speech.dll кітапханасын басамыз.



Сурет 3.1 – MS.Speech.dll кітапханасын орнату

Басты беттегі шешімдерді шолу терезесінде сілтеме батырмасында Microsoft.Speech қосымшасы пайда болады.

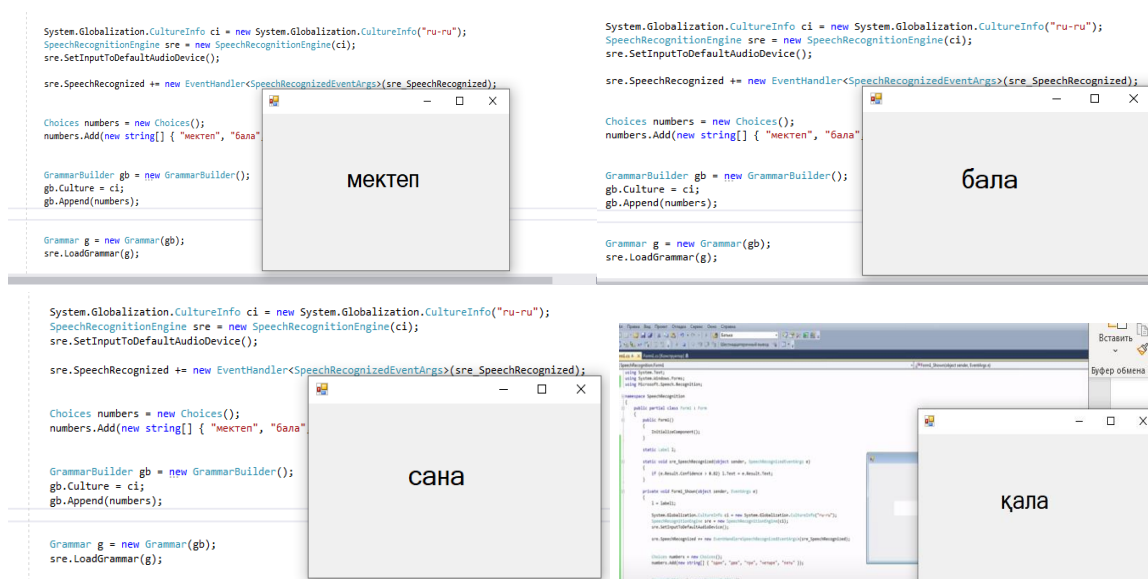
Form1.cs (Конструктор) терезесінде өңдеу жұмыстарын жүргіземіз: қаріп, Shown функцияларын өзгертеміз. Талапқа сай қадамдардан кейін кодты жазуға кірісеміз.



Сурет 3.2 – Талапқа сай форма терезесі

Өңдеу терезесіне кодты жазамыз. Ары қарай дауыс өңдеуге қажет тілді орнату үшін тіл платформасының кодын қосу қажет. Ал ары қарай дауысты микрофоннан жазу үшін `sre.SetInputToDefaultAudioDevice();` жазамыз.

Кодтағы 0,82-сәйкестік дәрежесін білдіреді. Яғни сәйкестік дәрежесі 0,82-ден асса ол терезеде сөйлеуді танып көрсетеді. Ал Multiple функциясы бірнеше тану қызметін білдіреді. Ендігі қателер болса түзеп жобаны іске қосамыз. Кезек-кезек сөздерді айта отырып, терезеге танылған сөздерді аламыз. Бастапқы 4 сөз еш қисынсыз танылып, ал соңғы «әдемі» сөзі 0,82 сәйкестік дәрежесінен төмен болған себептен танылмады. Ол тілдік базада қазақ тілін тану жүйесінің жоқтығынан болды. Біріктірілген нәтижелер келесідей:

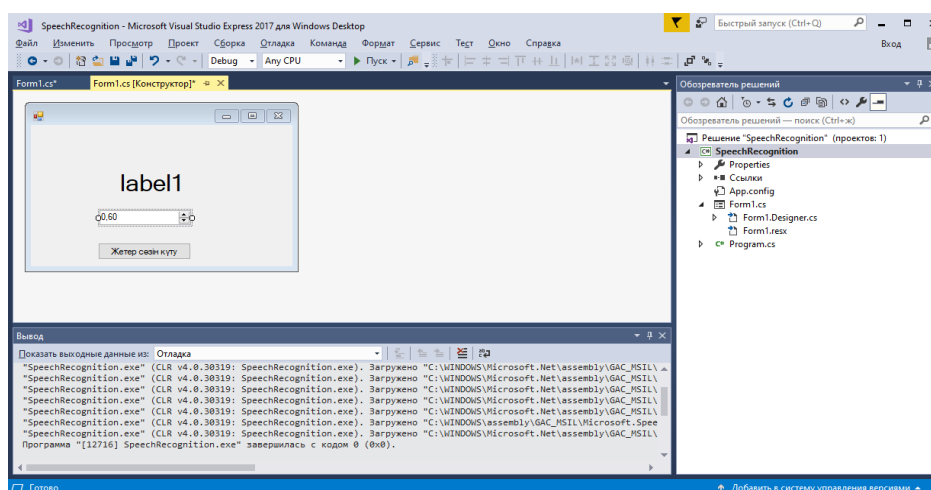


Сурет 3.3 – Нәтиже терезелері

3.3 C# бағдарламалау тілі арқылы сөйлеу командасын орындау

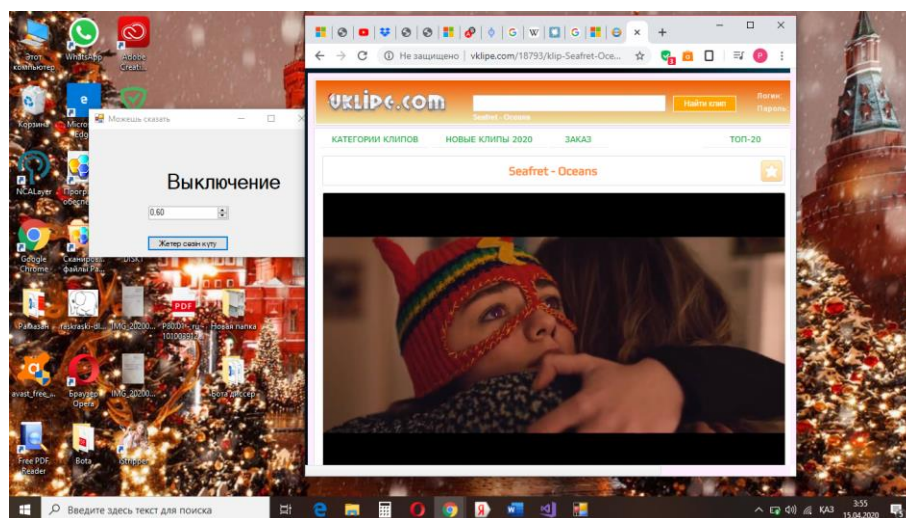
Келесі тәжірибеде C# бағдарламалау тілінде сөйлеу арқылы компьютерді өшіру бұйрығын орындайық. Бұл жағдай үйде, демалыс уақытында, отырған орыннан тұрмастан өшіру бұйрығын беру арқылы компьютерді өшіру үшін ыңғайлы. Осы сияқты мысалдар өте көп құрастыруға болады: олардың ішінде «ақылды үй» жүйесіндегі қызметтерді еске түсіруге болады. Әзірге осы мысалды бағдарламалап көрсек:

Ол үшін ең алдымен кодты әзірлеу қажет. Кодты енгізген соң, Пуск батырмасын басқаннан кейін Forms терезесі шығады.



Сурет 3.4 – Пуск батырмасын басқан соң бағдарлама жауабы

Қажетті параметрлерін өңдеу жұмыстарынан соң, кодты құрастырамыз. Кодтың ішінде өз бұйрығымызды енгіземіз. Осы жұмыста талап бойынша жетер бұйрығы айтылса, компьютер өшуі қажет.



Сурет 3.5 – Бұйрықты күту деңгейіндегі дайын бағдарлама

Ал енді қалаған бейнежазбаны қосып «жетер сөзін күту» батырмасын басып, орнымызға жайғасуға болады. Бейнежазбадан жалығып енді компьютерді өшіру үшін «жетер» сөзін дауыстап айту қажет. Содан соң компьютер сөзді танып, компьютерді өшіру жұмысына кіріседі. Ал пайдаланушы емін-еркін орнынан тұрмай-ақ демала береді. Керегі де осы тапсырма!

3.4 Сөйлеу технологиясы заманауи нарықта

Қазіргі заманда сөйлеу технологиялары өзінің ыңғайлылығы және оңайлығымен заманауи адамның өміріне енді. Олардың көмегімен кітаптарды, sms-хабарларды, құжаттарды және бүкіл веб-сайттарды дыбыстауға, іздеу жүйелерінде пернетақтаның көмегінсіз сұраныстар, тілдерді үйрену, құпия сөзді пайдаланбай жеке шоттан ақпарат алу тіпті жеке автомобильге нұсқау беруге болады [12].

Сөйлеу технологиялары саласындағы әзірлемелер көптеген салаларда өсіп келе жатқан сұранысқа ие:

- мемлекеттік сектор;
- қаржы;
- денсаулық сақтау;
- құқықтық және сот жүйелері;
- медиа коммуникация;
- әскери өнеркәсіп.

Бұл бағытта ғылыми-зерттеу институтының және бизнестегі қызуғышылықтың негізгі себебі өндіріс пен бизнестегі рутинді процестерді оңтайландыру үшін шешімдерге сұраныстың өсуі болып табылады.

Үкімет және мемлекеттік құрылымдар

Сөйлеу технологиялары нарығының өсу драйверлерінің бірі мемлекеттік және күш құрылымдары әрекет етеді. Шудан тазарту құрылғыларынан және сөйлеу сигналдарын жазу жүйелерінен басқа дауыспен және бет бойынша криминалистикалық есеп жүйелері енгізілуде. Кейбір заң шығарушы және атқарушы билік органдарда қазір тілді біріктіріп тану технологиясын пайдалана отырып, отырыстардың стенограммаларын дайындау жүйесін енгізуде.

Байланыс орталықтары

Сөйлеу технологиялары нарығында осы "бум" түрлі компаниялардың байланыс орталықтарына және аутсорсингте жұмыс істейтін байланыс орталықтарына белсенді енгізілетін "өзіне-өзі дауыстық қызмет көрсету жүйелері" (IVR) деп аталатын бастан кешуде.

Дауыстық өзіне-өзі қызмет көрсету жүйесі негізінде жатқан технологиялар үнемі даму үстінде: анықтамалық ақпарат беру және автоматты режимде типтік сұраныстарды өңдеуден басқа, байланыс орталығы жасанды

интеллект мүмкіндігі бар виртуалды кеңесшілерді құру бойынша міндеттер қояды.

Байланыс орталықтарына сөйлеу талдау жүйелерін, операторлар жұмысының сапасын басқару жүйелерін енгізу және клиенттердің қанағаттануын бағалау бойынша жобалар саны өсуде. Бұл жүйелерді пайдалану операторлардың күйзеліске төзімділік деңгейін және психофизикалық жай-күйін талдауын, клиенттердің қайта жүгіну себептерін талдау, олардың адалдығы мен қанағаттану деңгейін анықтау мүмкіндігін ашады.

Қаржы секторы

Соңғы бірнеше жылда алаяқтық несиелер санының елеулі өсуі банктердің алаяқтық тәуекелдерін төмендету, қолданыстағы клиенттерді қорғау және банкке деген сенімді арттыру үшін сөйлеу технологиялары базасында шешімдер қабылдауды белсенді түрде енгізуіне алып келді.

Клиенттерге қашықтан қызмет көрсету қажеттілігіне қарай, мысалы, байланыс орталығы операторының жеке басын растағанда, сондай-ақ сайтта, жеке кабинетке кірген кезде немесе мобильді қосымшада банктер дауыстық биометрия технологияларын пайдалануға көбірек бейім келеді. Алыстан қызмет көрсету жағдайында адам дауысы пайдаланушыны верификациялаудың ең сенімді тәсілі болып табылады, өйткені оны кез келген басқа ақпаратқа қарағанда ұрлауға немесе қолдан жасауға болмайды.

Денсаулық сақтау

Сөйлеу технологиялары нарығында денсаулық сақтау үшін де шешімдер бар. Дәрігердің сөзін тану және медициналық жүйеге ақпаратты автоматты түрде енгізу технологиясы картаны толтыру кезінде пациентті қарау кезінде, сондай-ақ операциялық жұмыс кезінде қолданылады. Ссы технологияны енгізуден зерттеу санының өсуі, медбике қызметінен бас тарту мен дәрігерлердің уақытын үнемдеу негізінде ақша қаражатының үнемделуі байқалады.

Қауіпсіздік қызметтері

Коммерциялық ұйымдарда, сондай-ақ өнеркәсіптік және азаматтық мақсаттағы объектілерде, қалалық қоғамдық көлікте, білім беру мекемелері мен ойын-сауық саласындағы мекемелерде байланыс арналары бойынша жауапты тұлғаларды жаппай бейнебақылау мен автоматты құлақтандыруды ұйымдастыру есебінен периметрді тиімді қорғау және эвакуациялау жүйелерін қауіпсіздік қызметіне енгізу жөніндегі жобалар іске асырылуда.

Ақпарат таралып кетуінің алдын алу және оқиғаларды талдау үшін қауіпсіздік қызметтері сөйлеуді және эмоциялық жағдайды талдауды және диспетчерлер, операторлардың келіссөздер тіркеуінің орталықтандырылған жүйесін қолданады.

Сот жүйесі

Үкімет сондай-ақ сот құрылымдарында бірнеше мәжілістермен бір мезгілде стенограммаларды дайындау жылдамдығы мен ұтқырлығын арттыру үшін стенографиялау жүйесі пайдаланылады. Заманауи сөйлеу

технологиялары сөйлеушінің сөзін оның бейне үлгісімен синхронды жазуды жүзеге асыру, сондай-ақ хаттама дайындауды автоматтандырылған режимде жүргізуге мүмкіндік береді.

Интернет және теледидар

Интернет пен теледидарды күнделікті пайдалану кезінде сөйлеу технологияларын қолдану арқылы соқтығысудың үлкен ықтималдығы бар.

Мысалы, субтитрлермен спорттық ойындарды онлайн трансляциялауды ұйымдастыру сөйлеуді тану технология негізінде құрылады

Сонымен қатар, интернет-ресурстар мен корпоративтік порталдар үшін жеке командаларды тану технологиясы веб-сайт бойынша дауыстық навигация сервисін енгізу үшін қолданылады, ал жеке тұлғаны дауыспен тексеру ойын-сауық интернет-ресурстары мен корпоративтік порталдар үшін қолданылады.

Автомобиль өнеркәсібі

Автомобиль өнеркәсібінде сөйлеу технологияларын қолдану бұған дейін автоөндірушілерге қол жетпейтін ойын-сауық және сервистік функцияларды пайдаланудың жаңа мүмкіндіктерін ашады. Автокөлік үшін әмбебап дауыстық шешімдер жүргізуші мен жолаушылардың мультимедиа және навигациялық жүйемен өзара іс-қимыл жасау тәсілдерін жеңілдетеді, осылайша, авариялық қауіпті төмендетіп, жоғары назар аударуды талап ететін әдеттегі борттық компьютерлерге қарағанда жүргізушіні автокөлікті басқарудан алаңдатпайды. Жүргізушінің дауысты биометриялық аутентификациясын қолдану автокөлікті айдап кету қаупін азайтуға мүмкіндік береді.

Сөйлеу технологиялары нарығындағы қазіргі жағдайдың ерекшеліктері:

- сөйлеу технологиялары мен сөйлеуді компьютерлік өңдеу құралдары нарығы-бүгінгі күні ең жылдам өсетін бағыттардың бірі;
- қазіргі заманғы сөйлеу шешімдерін пайдалану компанияның ішкі процестерін оңтайландыруға және іс жүзінде барлық салаларда шығындарды азайтуға мүмкіндік береді;
- негізінен компаниялар сөзді автоматты түрде тану, мәтінді сөзге түрлендіру технологиясы мен спикерді верификациялау жүйесіне қаржы салады;
- әзірлеушілер арасындағы көшбасшылар АҚШ, Ұлыбритания, Жапония, Израиль және Ресей болып табылады, бірақ болжам бойынша Азия-Тынық мұхит елдері өңірінде өсу қарқыны алда тұр.

ҚОРЫТЫНДЫ

Сөйлеуді тану саласында көптеген зерттеулер бар, бірақ әлі күнге дейін сөйлеуді тану жүйесі жүз пайызға дәл емес. Әзірленген жүйелер бірнеше пайдаланушылардың сөзін сәйкестендіре алмайды, сөйлеушінің сөзін анық тану үшін шудан тыс болуы қажет, одан басқа акценті және айтылу мәселесі туындайды. Болашақта сөйлеуді тану жүйелері жүз пайыздық нәтиже беру үшін осы шектеулерді болдырмау тиіс. Бұл жұмыста көптеген жылдар бойы сөйлеуді тану шеңберінде әзірленген негізгі жүйелер қарастырылды. Ең басты жаңалық қазақша сөйлеу фонемдерінің өтпелі аймақтарындағы өзгерістер талданды. Нәтижесінде фонем арасындағы сипаттамалар салыстырылып, тәуелділік графиктері тұрғызылды.

Негізгі бөлімде сөйлеу сигналындағы дыбысты фонемалардың бірдей кезеңдерден тұратынын, ал дыбыссыз фонемалардың тұрақсыз сигналдардан құралатынын анықталды. Жыныстық белгіге байланысты сөйлеу сигналының 3 түрлі уақыттық сипаттамасы салыстырылды. Таңдалған сөйлемде әр сөзге қойылған интонацияның әсерін қарастырдық. Соңында форманттарды есептеу мысалын келтірдік және нейтралды дауысты дыбыс үшін ер адам, әйел, сәби үшін форманттар жиілігінің кестесін құрдық. Ал акустикалық талдауда формант алуды есептеуді қарастырдым. Бағдарламада жиілік форманттарын скриптке енгізілген сценарий бойынша автоматты табу алгоритмі қарастырылды. Нәтижесінде форманттардың әр адамда түрліше болатынын, алайда жалпы форманттық құрылымда белгілі-бір ереженің сақталатыны белгілі болды.

Бұл зерттеулер сөйлеу сигналының параметрлері туралы біраз түсінік береді. Сонымен қатар сигналдың спектрі, тоны, қарқындылығы, формант және импульсін графикалық бейнеде және мәндерін алу жұмысы жүргізілді. Қазақ тіліндегі дыбыстардың сөйлеу сигналдарының параметрлерінің уақыттық, спектралды талдауы меңгеріледі. Ол Cool Edit Pro бағдарламасы арқылы жүзеге асырылады. Акустикалық зерттеуде әртүрлі дыбыстарды манипуляциялау үшін жалпы құралдар жиынтығы біріктірілген компьютерлік бағдарламасы, мүмкіндіктері көп PRAAT (Boersma & Weenink 1992-2013) іске асырылды.

Диссертациялық жұмыста C#.net және Matlab сөйлеу сигналдарын тануды модельдеудің берілген алгоритмін іске асыратын жоғары деңгейдегі бағдарламалау тілдерінде бағдарлама жасалды. Нәтижесінде сөйлеу командалары орындалды. Алынған нәтижелер сөйлеуді тану үшін сөйлеу сигналдарының бөлінетін параметрлерін пайдалану мүмкіндігін көрсетті.

ҚЫСҚАРТЫЛҒАН СӨЗДЕР ТІЗІМІ

АЦТ – аналогты-цифрлы түрлендіру
СЦӨ – сигналдарды цифрлық өңдеу
ФДТ – Фурьенің дискретті түрленуі
НТЖ – негізгі тон жиілігі
НТП – негізгі тон периоды
СТ – сөйлеуді тану
СБК – сызықтық болжаммен кодтау
МЖК – Мел-жиілік кепстраль
ЫСАТ – ықтималдық сызықтық ажырату талдауы
ЖММ – жасырын Марков моделі
MFCC – Mel frequency cepstral coefficients
PLP – Perceptual linear prediction
VFR– Variable Frame Rate analysis
SBCOR– Subband-Autocorrelation
CMS – Cepstral Mean Subtraction
RASTA – Relative Spectral Analysis
NSS – non-linear Spectral Subtraction,
PCA – Principal Component Analysis
ICA – Independent Component Analysis
САТ – сөйлеуді автоматты тану
ДБ – динамикалық бағдарламалау

ҚОЛДАНЫЛҒАН ӘДЕБИЕТТЕР ТІЗІМІ

1. Dr.Shaila, D.Apte.Speech and Audio Processing.Дели, 2012-438 б.
- 2.JanuszDulas. Speech recognition based on the Grid Method and Image Similarity. OpoleUniversityofTechnology, Польша, №16-320-339 б.
3. В.Н. Позднин, М.Г.Хохлов. Расчет формант по участку речевого сигнала. Технические науки, 2017, №1-6 б.
- 4.В.Е. Петеляк. Система поэлементного (фонемного) распознавания речевых сигналов. –Москва, 1995. – С. 107;
5. И.В. Бочарев, Д.Ю. Акатьев. Распознавания речевых сигналов на основе корреляционного метода. – М.: Электронный журнал «Исследовано в России». 2003– 131 с;
6. К.К. Томчук. Сегментация речевых сигналов для задач автоматической обработки речи. –Санкт-Петербург, 2017 –С.196.
7. Joey Stanley A tutorial on extracting formants in Praat, How-to Guides, Phonetics, 2019;
8. Paul Boersma Chapter 17:Acoustic analysis, 2013, 21 б;
9. Taabish Gulzar, Anand Singh. Comparative Analysis of LPCC, MFCC and BFCC. –International Journal of Computer Applications, 2014. — № 101(12). — 22–27.
10. Рабинер Л. Р., Шафер Р. В. Цифровая обработка речевых сигналов. — М.: Радио и связь, 1981. — 496 с.
11. Смит С. Цифровая обработка сигналов. Практическое руководство для инженеров и научных работников. — М.: Додэка-XXI, 2012. — 720 с.
12. Дырмовский Д., Матвеев Ю., Балыкина Л. Современный рынок речевых технологий. Control engineering Россия №1 (55), 2015
13. Искусственный интеллект. Системы общения и экспертные системы. Кн. 1 / Под ред. Э.В.Попова. - М.: Радио и связь, 1990. - 461 с.
14. Оппенгейн А.В., Шафер Р.В. Цифровая обработка сигналов, М.: Радио и связь, 1979 ., 347 с.
- 15.Рабинер Л.Р. Шафер Р.В. Цифровая обработка речевых сигналов, М.: Радио и связь, 1981 ., 258 с.
- 16.Литюк В.И. Методическое пособие № 2231 часть 3 «Методы расчета и проектирование цифровых многопроцессорных устройств обработки радиосигналов», Таганрог, 1995, 48 с.
17. Кузнецов В., Отт А. Автоматический синтез речи. - Таллинн: Валгус, 1989. - 135 с.
18. Методы автоматического распознавания речи / Под ред. У.Ли. - М.: Мир, 1983. - 716 с.
19. Зиндер Л.Р. Общая фонетика. - М.: Высшая школа, 1979. - 312 с.
20. Златоустова Л.В., Потапова Р.К., Трунин-Донской В.Н. Общая и прикладная фонетика. М.: МГУ, 1986. - 304 с.
21. Линдсей П., Нордман Д. Переработка информации у человека. - М.: Мир, 1974. - 550 с.

22. Потапова Р.К. Речевое управление роботом. - М.: Радио и связь, 1989. - 248 с.

23. Малых Д. А., Система управления устройствами «умного дома» с использованием голосовых команд / Д. А. Малых, Ю. С. Кириллова. — Текст: непосредственный, электронный // Молодой ученый. — 2017. — № 19 (153). — С. 60-64.

24. Бакаева Т.Н. Безопасность жизнедеятельности. Часть 2: Безопасность в условиях производства: Учебное пособие. Таганрог: ТРТУ, 1997, 318 с.

25. Фрумкин Г.А. «Расчет и конструирование РЭА», Москва: Высшая школа, 1997, 289 с.

26. Б.П.Раманкулова, Л.Б.Илипбаева. Сөйлеу сигналының параметрлеріне әсер ету факторларын зерттеу. Сатбаев окуларының еңбектері II Том, Алматы 2019, 133 б..

27. Б.П.Раманкулова, Л.Б.Илипбаева. Сөйлеу сигналының негізгі көрсеткіштері бойынша салыстырмалы талдау. Вестник КазННТУ №2 (138), 2020, 601 б.