

**ТОШКЕНТ АХБОРОТ ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ
ҲУЗУРИДАГИ ИЛМИЙ ДАРАЖАЛАР БЕРУВЧИ
DSc.13/30.12.2019.Т.07.01 РАҚАМЛИ ИЛМИЙ КЕНГАШ
АСОСИДАГИ БИР МАРТАЛИК ИЛМИЙ КЕНГАШ**

**ТОШКЕНТ АХБОРОТ ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ
ҲУЗУРИДАГИ АХБОРОТ-КОММУНИКАЦИЯ ТЕХНОЛОГИЯЛАРИ
ИЛМИЙ-ИННОВАЦИОН МАРКАЗИ**

ЮЛДОШЕВ ЮСУФ ШЕРАЛИЕВИЧ

**НУТҚ СИГНАЛЛАРИГА ДАСТЛАБКИ ИШЛОВ БЕРИШ
АЛГОРИТМЛАРИ**

05.01.03 – Информатиканинг назарий асослари

**ТЕХНИКА ФАНЛАРИ БЎЙИЧА ФАЛСАФА ДОКТОРИ (PhD)
ДИССЕРТАЦИЯСИ АВТОРЕФЕРАТИ**

Тошкент – 2021

**Техника фанлари бўйича фалсафа доктори (PhD) диссертацияси
автореферати мундарижаси**

**Оглавление автореферата диссертации
доктора философии (PhD) по техническим наукам**

**Contents of Dissertation Abstract of the Doctor of Philosophy (PhD) on
Technical Sciences**

Юлдошев Юсуф Шералиевич

Нутқ сигналларига дастлабки ишлов бериш алгоритмлари 3

Юлдошев Юсуф Шералиевич

Алгоритмы предварительной обработки речевых сигналов 21

Yuldoshev Yusuf Sheralievich

Pre-processing algorithms for speech signals 39

Эълон қилинган ишлар рўйхати

Список опубликованных работ

List of published works 43

**ТОШКЕНТ АХБОРОТ ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ
ХУЗУРИДАГИ ИЛМИЙ ДАРАЖАЛАР БЕРУВЧИ
DSc.13/30.12.2019.Т.07.01 РАҚАМЛИ ИЛМИЙ КЕНГАШ
АСОСИДАГИ БИР МАРТАЛИК ИЛМИЙ КЕНГАШ**

**ТОШКЕНТ АХБОРОТ ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ
ХУЗУРИДАГИ АХБОРОТ-КОММУНИКАЦИЯ ТЕХНОЛОГИЯЛАРИ
ИЛМИЙ-ИННОВАЦИОН МАРКАЗИ**

ЮЛДОШЕВ ЮСУФ ШЕРАЛИЕВИЧ

**НУТҚ СИГНАЛЛАРИГА ДАСТЛАБКИ ИШЛОВ БЕРИШ
АЛГОРИТМЛАРИ**

05.01.03 – Информатиканинг назарий асослари

**ТЕХНИКА ФАНЛАРИ БЎЙИЧА ФАЛСАФА ДОКТОРИ (PhD)
ДИССЕРТАЦИЯСИ АВТОРЕФЕРАТИ**

Тошкент – 2021

Техника фанлари бўйича фалсафа доктори (PhD) диссертацияси мавзуси Ўзбекистон Республикаси Вазирлар Маҳкамаси ҳузуридаги Олий аттестация комиссиясида В2020.2.PhD/Т894 рақам билан рўйхатга олинган.

Диссертация Тошкент ахборот технологиялари университети ҳузуридаги ахборот-коммуникация технологиялари илмий-инновацион марказида бажарилган.

Диссертация автореферати уч тилда (ўзбек, рус, инглиз (резюме)) Илмий кенгаш веб-саҳифасида (www.tuit.uz) ва «Ziyonet» Ахборот таълим порталида (www.ziyonet.uz) жойлаштирилган.

Илмий раҳбар:

Бабомурадов Озод Жўраевич

техника фанлари доктори

Расмий оппонентлар:

Мухамедиева Дилноз Тулкуновна

техника фанлари доктори, профессор

Мирзаев Олимжон Намозович

техника фанлари бўйича фалсафа доктори

Етакчи ташкилот:

Ўзбекистон миллий университети

Диссертация ҳимояси Тошкент ахборот технологиялари университети ҳузуридаги DSc.13/30.12.2019.Т.07.01 рақамли Илмий кенгашнинг 2021 йил «___» _____ соат _____ даги мажлисида бўлиб ўтади. (Манзил: 100202, Тошкент шаҳри, Амир Темур кўчаси, 108-уй. Тел.: (99871) 238-64-43, факс: (99871) 238-65-52, e-mail: tuit@tuit.uz).

Диссертация билан Тошкент ахборот технологиялари университети Ахборот-ресурс марказида танишиш мумкин (_____ рақам билан рўйхатга олинган.). (Манзил: 100202, Тошкент шаҳри, Амир Темур кўчаси, 108-уй. Тел.: (99871) 238-65-44).

Диссертация автореферати 2021 йил «___» _____ да тарқатилди.
(2021 йил «___» _____ даги _____ рақамли реестр баённомаси.)

Р. Х. Хамдамов

Илмий даражалар берувчи илмий кенгаш раиси, т.ф.д., профессор

Ф. М. Нуралиев

Илмий даражалар берувчи илмий кенгаш илмий котиби, т.ф.д., доцент

М.А.Исмаилов

Илмий даражалар берувчи илмий кенгаш қошидаги илмий семинар раиси, т.ф.д., профессор

КИРИШ (фалсафа доктори (PhD) диссертацияси аннотацияси)

Диссертация мавзусининг долзарблиги ва зарурати. Жаҳонда нутқ сигналларига рақамли ишлов бериш, нутқ сигналлари асосида шахсни таниб олиш, нутқни матнга ўтказиш ҳамда тимсолларни таниб олиш усул ва алгоритмларини такомиллаштириш, ишлаб чиқиш ва жорий қилишга катта эътибор қаратилмоқда. «Сўнгги йилларда нутқга асосланган интеллектуал тизимларни яратиш жадал ривожланаётган йўналишлардан бири бўлиб, у анъанавий таниб олиш усуллари ютуқларидан фойдаланиш билан бирга улар орқали ечиб бўлмайдиган масалаларни ечиш имкониятига эга»¹. Ушбу йўналишда рақамли сигналларни интеллектуал таҳлил қилишнинг асосий функционал масалалари синфлаштириш, таснифлаш, таниб олиш, башпоратлаш ва баҳолаш, қонуниятларни аниқлаш, қарор қабул қилиш усул ва алгоритмларини ишлаб чиқиш ҳамда иловаларини яратиш муҳим масалалардан бири бўлиб қолмоқда. Ҳозирги кунда нутқни таниб олиш технологияларига асосланган кўплаб иловалар ишлаб чиқилган. Хорижий мамлакатларда, жумладан Хитой, Россия Федерация, АҚШ, Англия, Германия, Ҳиндистон, Франция ва бошқа давлатларда нутқ сигналларига дастлабки ишлов бериш ҳамда нутқни таниб олиш йўналишларининг назарий ва амалий масалаларини ечишга катта эътибор қаратилмоқда.

Жаҳонда нутқ сигналларини қайта ишлаш, нутқ сигнали белгиларини ажратиш, нутқни таниб олиш усул ва алгоритмларини такомиллаштириш ва ишлаб чиқиш ҳамда ҳисоблаш алгоритмларини яратишга йўналтирилган илмий тадқиқот ишлари олиб борилмоқда. Бу борада, жумладан, нутқ сигналлари сифатини ошириш, нутқ сигналдаги фойдали нутқ жойлашган соҳасини топиш ва уни характерловчи белгиларини аниқлаш, таниб олиш алгоритмларини ишлаб чиқиш, такомиллаштириш ҳамда нутқ моделлари ва автоматлаштирилган таниб олиш тизимларини яратиш муҳим вазифалардан бири ҳисобланади.

Республикамизда мазкур йўналишда ўзбек тилидаги нутқ сигналларига дастлабки ишлов бериш, интеллектуал таҳлил қилиш ва таниб олишга мўлжалланган автоматлаштирилган тизимларни ишлаб чиқишга алоҳида эътибор қаратилмоқда. 2017-2021 йилларда Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегиясида, жумладан «... иқтисодиёт, ижтимоий соҳа ва бошқарув тизимида ахборот-коммуникация технологияларини жорий этиш, ... ахборот хавфсизлигини таъминлаш ва ахборотни ҳимоя қилиш тизимини такомиллаштириш, ахборот соҳасидаги таҳдидларга ўз вақтида ва муносиб қарши ҳаракатларни ташкил этиш»² вазифалари белгиланган. Мазкур вазифаларни амалга оширишда жумладан, нутқ сигналларидан фойдаланиб шахсни идентификациялаш, нутқни матнга ўтказиш ҳамда нутқ орқали бошқариш тизимларини яратиш ахборот-коммуникация технологияларини ривожлантиришнинг муҳим масалаларидан

¹ https://www.tadviser.ru/index.php/Статья:Речевые_технологии:_на_пути_от_распознавания_к_пониманию

² Ўзбекистон Республикаси президентининг 2017 йил 7 февралдаги ПФ-4947-сон “Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегияси тўғрисида”ги Фармони

бири ҳисобланади. Бу борада, нутқ сигналларига дастлабки ишлов бериш усул ва алгоритмларини, шахсни таниб олиш ва нутқ сигналлари информатив белгиларини аниқлашни инобатга олган ҳолда ривожлантириш шунингдек, кириш-чиқишни назорат қилиш ҳамда овозли мурожаатларни таҳлил қилиш тизимларида қўллаш муҳим вазифалардан бири ҳисобланади.

Ўзбекистон Республикаси Президентининг 2017 йил 7 февралдаги ПФ-4947-сон «Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегияси тўғрисида»ги Фармони, 2017 йил 18 апрелдаги ПҚ-2898-сон «Ички ишлар органларининг жиноятларни тергов қилиш соҳасидаги фаолиятини тубдан такомиллаштириш чора-тадбирлари тўғрисида»ги, 2018 йил 21 ноябрдаги ПҚ-4024-сон «Ахборот технологиялар ва коммуникацияларини жорий этилишини назорат қилиш, уларни ҳимоя қилиш тизимини такомиллаштириш чора-тадбирлари тўғрисида»ги Қарорлари ҳамда мазкур фаолиятга тегишли бошқа меъёрий-ҳуқуқий ҳужжатларда белгиланган вазифаларни амалга оширишга ушбу диссертация тадқиқоти муайян даражада хизмат қилади.

Тадқиқотнинг республика фан ва технологиялари ривожланишининг устувор йўналишларига мослиги. Мазкур тадқиқот республика фан ва технологиялар ривожланишининг IV. «Ахборотлаштириш ва ахборот-коммуникация технологияларини ривожлантириш» устувор йўналишига мос равишда бажарилган.

Муаммонинг ўрганилганлик даражаси. Нутқ сигналига дастлабки ишлов бериш, қайта ишлаш ва таниб олиш модел, усул ва алгоритмларини ишлаб чиқиш ва такомиллаштириш масалаларини ечиш ҳамда уларни амалиётга жорий этиш бўйича хорижий олимлардан Михайлов В.Г., Сапожков М.А., Назаров М.В., Прохоров Ю.Н., Ланнэ А. А., Грэй А. Х., Маркел Дж.Д., Рабинер Л. Р., Ситняковский И.В., Шафер Р.В., Шалимов И.А., Петровский А.А., Ковалгин Ю.А., Итакура Ф., Рейдер Ч., Макхоул Дж ва бошқаларнинг илмий ишлари диққатга сазовор.

Ўзбекистонда таниб олиш ва нутқ сигналларини таҳлил қилиш назарий асосларини ривожлантиришга М.М.Камилов, Ш.Х.Фозилов, Э.М.Алиев, Р.Ҳ.Ҳамдамов, М.М.Мусаев, Н.А.Игнатъев, С.С.Содиқов, Н.С.Маматов, Н.М.Мирзаев, С.С.Раджабов, Ш.Е.Туляганов ва бошқалар ўзларининг ҳиссаларини қўшиб келмоқдалар.

Ҳозирги кунда нутқ сигналларига ишлов бериш асосида шахсни таниб олиш тизимлари ва нутқни таниб олиш технологиялари жадал суръатлар билан ривожланмоқда. Ушбу йўналиш бўйича ўтказилган тадқиқотлар таҳлили шуни кўрсатадики, нутқ сигналларини таҳлил қилиш асосида нутқни таниб олиш технологиялари энг яхши, истиқболли технологияларидан бири ҳисобланади. Бироқ, бундай технология асосида шахсни ва нутқни таниб олиш автоматлаштирилган тизимларини яратишда вужудга келадиган муаммолар ҳозирги кунгача етарли даражада ҳал этилмаган. Бундан ташқари, дастлабки ишлов бериш асосида шахсни ва нутқни таниб олиш ихчам, ишончли, мустақил ва юқори тезликни таъминловчи усул ва алгоритмларини ишлаб чиқиш муаммоси етарли даражада ўрганилмаган.

Диссертация тадқиқотининг диссертация бажарилган олий таълим ва илмий-тадқиқот муассасасининг илмий-тадқиқот ишлари режалари билан боғлиқлиги. Диссертация тадқиқоти Муҳаммад ал-Хоразмий номидаги Тошкент ахборот технологиялари университети хузуридаги Ахборот-коммуникация технологиялари илмий-инновацион маркази ва Муҳаммад ал-Хоразмий номидаги Тошкент ахборот технологиялари университетининг илмий-тадқиқот ишлари режасининг №БВ-Атех-2018 «Юз тасвирларини оқимли қайта ишлаш асосида шахсни идентификация қилиш алгоритмлари ва дастурий таъминотини ишлаб чиқиш» (2018-2020), №М-Ф4-09 «Норавшан муҳитда мураккаб тузилмали ахборотларга ишлов бериш, таниб олиш ва башоратлашнинг баҳоларини интеллектуал тизимлари» (2015-2020), №ПЗ-201906202 «Ўзбек тилидаги матнли ахборотларни таҳлил қилиш ва таснифлаш тизимини ишлаб чиқиш» (2019-2020) мавзулардаги илмий лойиҳалари доирасида бажарилган.

Тадқиқотнинг мақсади турли шароитларда ёзиб олинган нутқ сигналларини халақитларини камайтириш ва силлиқлаш, частотали соҳаларига ишлов бериш, нормаллаштириш ва сегментлаш алгоритмлари ҳамда дастурий мажмуасини ишлаб чиқишдан иборат.

Тадқиқотнинг вазифалари:

нутқ сигналидаги халақитларни камайтириш ва силлиқлаш алгоритминини ишлаб чиқиш;

нутқни частотали соҳаларига ишлов бериш усулини такомиллаштириш; кировчи сигналдан нутқ жойлашган соҳани аниқлаш алгоритминини ишлаб чиқиш;

нутқ сигналини нормаллаштириш ва сегментлаш алгоритмларинини ишлаб чиқиш;

шахсни овози асосида идентификациялаш ва нутқни таниб олиш масалаларинини ечиш учун белгилар фазосинини шакллантириш;

нутқ сигналларига дастлабки ишлов беришнинг алгоритмик ва дастурий мажмуасинини яратиш.

Тадқиқотнинг объекти сифатида турли шароитларда ёзиб олинган нутқ сигналлари, уларга ишлов бериш усул ва алгоритмлари олинган.

Тадқиқотнинг предметини сигналларга дастлабки ишлов бериш, қайта ишлаш, тимсолларни таниб олиш усул ва алгоритмлари ҳамда дастурий таъминоти ташкил этади.

Тадқиқотнинг усуллари. Ишнинг назарий тадқиқотлари тизимли таҳлил, имитацион моделлаштириш, эҳтимоллар назарияси, математик статистика, дискрет математика, сигналларга ишлов бериш ва тимсолларни таниб олиш усулларидадан фойдаланилган.

Тадқиқотнинг илмий янгилиги қуйидагилардан иборат:

халақитнинг спектрал зичлигинини инобатга олган ҳолда нутқ сигналидаги халақитларни камайтириш ва силлиқлаш алгоритминини ишлаб чиқилган;

экстремумларни тарқалиш қонуниятлари асосида нутқ жойлашган соҳани аниқлаш алгоритминини ишлаб чиқилган;

сигнал-халақит муносабатларини аниқлаш орқали нутқ частотали соҳаларига ишлов бериш усули такомиллаштирилган;

нутқ сигнали сифатига таъсир қилувчи омилларни инобатга олган ҳолда нормаллаштириш ва сегментлаш алгоритмлари ишлаб чиқилган;

нутқ сигналлини дискретлигини инобатга олган ҳолда шахсни овози асосида идентификациялаш ва нутқни таниб олиш масалаларини ечиш учун белгилар фазосини шакллантириш алгоритми ишлаб чиқилган.

Тадқиқотнинг амалий натижалари қуйидагилардан иборат:

нутқ сигналидаги халақитларни камайтириш, силлиқлаш, нормаллаштириш, сегментлаш ва таниб олиш масалаларини ечиш учун белгилар фазосини шакллантириш алгоритмлари ҳамда нутқнинг частотали соҳаларига ишлов бериш усули ишлаб чиқилган;

мавжуд ҳамда ишлаб чиқилган усул ва алгоритмларга асосланган нутқ сигналларига дастлабки ишлов бериш ва шахсни нутқи асосида таниб олиш дастурий мажмуаси ишлаб чиқилган.

Тадқиқот натижаларининг ишончлилиги алгоритмларни ишлаб чиқишда нутқ сигналларига ишлов бериш ва тимсолларни таниб олишнинг математик аппаратининг тўғри қўлланилиши ҳамда тажрибавий тадқиқотларнинг ижобий натижалари билан тасдиқланади.

Тадқиқот натижаларининг илмий ва амалий аҳамияти. Тадқиқот натижаларининг илмий аҳамияти нутқ сигналларига дастлабки ишлов бериш асосида шахсни биометрик идентификациялаш технологияларининг назарий асосларининг истиқболли ривожланишига ишлаб чиқилган алгоритмларнинг ҳисса қўшиши билан изоҳланади.

Тадқиқот натижаларининг амалий аҳамияти ишлаб чиқилган дастурий мажмуа шахсни ва нутқни таниб олиш автоматлаштирилган тизимларини яратишда қўлланилиши билан изоҳланади.

Тадқиқот натижаларининг жорий қилиниши. Нутқ сигналларига дастлабки ишлов беришнинг мавжуд ҳамда таклиф этилган усул ва алгоритмлари асосида яратилган дастурий таъминот асосида:

нутқ сигналидаги халақитларни камайтириш ва силлиқлаш, нутқнинг частотали соҳаларига ишлов бериш ва таниб олиш масалаларини ечиш учун белгилар фазосини шакллантириш усул ва алгоритмлари асосида яратилган дастурий мажмуа Қашқадарё вилояти Шаҳрисабз шаҳар ҳамда Қамаш туман ҳокимиятларида жорий қилинган (Ахборот технологиялари ва коммуникацияларни ривожлантириш вазирлигининг 2020 йил 22 декабрдаги 33-8/7838-сон маълумотномалари). Натижада аудио маълумотларни таҳлил қилиш учун кетадиган вақтни ўртача 15% қисқартириш имконини берган;

нутқ сигналлини нормаллаштириш ва сегментлаш, нутқ сигналларидаги халақитларни камайтириш, таниб олиш масалаларини ечиш учун белгилар фазосини шакллантириш, силлиқлаш усул ва алгоритмлари асосида яратилган дастурий мажмуа Қашқадарё вилояти Шаҳрисабз туманлараро фуқаролик суди ҳамда Шаҳрисабз туман ички ишлар бошқармасида жорий қилинган (Ахборот технологиялари ва коммуникацияларни ривожлантириш вазирлигининг 2020 йил 22 декабрдаги 33-8/7838-сон маълумотномалари).

Натижада нутқ сигналларига дастлабки ишлов бериш, силлиқлаш, филтрлаш, турли халақитлардан тозалаш орқали нутқни юқори аниқликда ифодаланилишини таъминлаш ва аудиомаълумотларни таҳлил қилиш учун кетадиган вақтни 15% га қисқартириш ва уни тушуниш самарадорлигини ўртача 20% га ошириш имконини берган.

Тадқиқот натижаларининг апробацияси. Мазкур тадқиқот натижалари 5 та халқаро ва 6 та республика илмий-амалий анжуманларида маъруза қилинган ва муҳокамадан ўтказилган.

Тадқиқот натижаларининг эълон қилинганлиги. Тадқиқот мавзуси бўйича жами 21 та илмий иш чоп этилган бўлиб, шулардан Ўзбекистон Республикаси Олий аттестация комиссияси томонидан тавсия этилган илмий нашрларда 10 та мақола, жумладан 3 таси хорижий ва 7 таси республика журналларида нашр қилинган ҳамда 2 та ЭҲМ учун яратилган дастурий воситаларини қайд қилиш гувоҳномалари олинган.

Диссертациянинг тузилиши ва ҳажми. Диссертация кириш, тўртта боб, хулоса, фойдаланилган адабиётлар рўйхати ва иловалардан иборат. Диссертациянинг ҳажми 118 бетни ташкил этади.

ДИССЕРТАЦИЯНИНГ АСОСИЙ МАЗМУНИ

Киришда диссертация мавзусининг долзарблиги ва зарурияти асосланган, тадқиқотнинг республика фан ва технологиялари тараққиётининг устувор йўналишларига мослиги кўрсатилган. Тадқиқотнинг мақсад ва вазифалари белгилаб олинган ҳамда тадқиқот объекти ва предмети аниқланган, олинган натижаларнинг ишончлилиги асослаб берилган, уларнинг назарий ва амалий аҳамияти кўрсатилган, тадқиқот натижаларини амалиётга жорий қилиш ҳолати, нашр этилган ишлар ва диссертация тузилиши бўйича маълумотлар келтирилган.

Диссертациянинг «**Нутқ сигналларига дастлабки ишлов бериш муаммолари**» деб номланган биринчи боби, нутқ сигналларига дастлабки ишлов беришнинг замонавий ҳолатини ўрганишга бағишланган. Унда нутқ сигналларига қўйиладиган талаблар ва ишлов бериш усуллари таҳлиллари ҳам келтирилган бўлиб, нутқ сигналларига ишлов беришнинг оммалашган усулларининг ютуқ ва камчиликлари ҳамда нутқ сигналларига ишлов бериш соҳасидаги асосий муаммолар баён этилган. Бундан ташқари, нутқ сигналларига дастлабки ишлов бериш асосида шахсни таниб олиш муаммолари ўрганилиб, уларни бартараф этиш йўллари ва тадқиқот иши масаласининг қўйилиши келтирилган.

1.1-параграфда нутқни узатиш каналининг хоссаси, луғат ўлчами, нутқнинг ўзгарувчанлиги, атрофдаги халақит даражаси, нутқ тури каби кўплаб параметрлар таҳлили асосида нутқни автоматик таниб олиш тизимларига кирувчи дастлабки нутқ сигналларига қўйиладиган талаблар келтирилган.

1.2-параграфда фойдаланувчи нутқ сигналининг ўзига хос хусусиятлари ва нутқни таниб олиш тизимларини яратишдаги муаммолар таҳлили асосида

нутқ сигналларини ишлов бериш тизимларига бўлган асосий талаблар ишлаб чиқилган.

1.3-параграфда нутқ сигналларига дастлабки ишлов бериш усуллари таҳлил этилган бўлиб, унда нутқ табиати ташкил этувчилари таснифланган, нутқ сигнаliga ишлов беришнинг мавжуд тизимлари камчиликлари аниқланган, шовқинларни камайтириш алгоритмларини ривожлантириш зарурлиги кўрсатиб берилган. Бундан ташқари, нутқ ва халақитларни моделлаштириш, спектрни созлаш функцияси, нутқнинг стационарлик даврини инобатга олган ҳолда мавжуд алгоритмларни такомиллаштириш зарурлиги асосланган.

1.4-параграфда суҳандонни идентификация қилиш тизимларининг халақитларга нисбатан барқарорлик даражасини ошириш босқичлари таклиф этилган бўлиб, улар асосида диссертация тадқиқотининг мақсад ва вазифалари шакллантирилган.

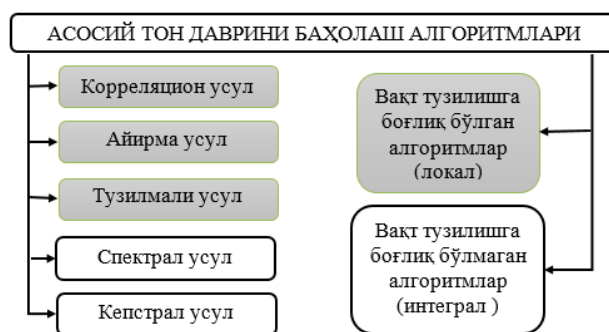
Диссертациянинг «**Нутқ сигналларига дастлабки ишлов бериш усул ва алгоритмлари таҳлили**» деб номланувчи иккинчи боби 3 та параграфдан иборат бўлиб, нутқ сигналини нормаллаштириш ва фойдали нутқ жойлашган соҳани ажратиш, халақитларни аниқлаш, сегментлаш ва характерловчи параметрларини баҳолашнинг мавжуд усул ва алгоритмларни баён этишга бағишланган.

2.1-параграфда нутқ сигналини дискретлаш, нормаллаштириш усул ва алгоритмлари таҳлили асосида уларнинг ютуқ ва камчиликлари аниқланган.

Таниб олиш масаласи учун тенг бўлмаган вақт давомийлик фрагментларига автоматик сегментлашни амалга ошириш учун сегментлаш алгоритмида қуйидагиларни бўлиши талаб этилади:

1. нутқ сигнали сегментларининг детектирланган турлари (нутқ сигнал фрагментларини тўхталишлар, халақитли, портловчи товушларига таснифлаш);
2. сегментларнинг вақт чегараларини аниқлаштириш (қайта ўтиш алгоритми);
3. нутқ сигнал сегментларини параметрлаш;
4. нутқ сигнал сегментларининг таркибини таҳлил қилиш ва созлаш (алоҳида фонемаларга мос келадиган сегментларга ажратиш мумкин бўлган қуйи сегментларни батафсил таҳлил қилиш);
5. танланган фонемаларни таснифлаш.

2.2-параграфда нутқ сигналларини сегментларга ажратиш усуллари таҳлил қилиниб, турли математик ёндашувларга асосланган алгоритмлар таҳлил этилган бўлиб, унда сегментлаш масаласини ечишда асосий тон даврини балаҳолаш алгоритмларининг қуйидаги таснифи тавсия этилган.



1-расм. Нутқ сигнали асосий тон даврини баҳолаш алгоритмлари таснифи

2.3-параграфда нутқ сигналлари белгиларини шакллантиришнинг мавжуд ва таклиф этилган алгоритмлари баёни келтирилган бўлиб, унда нутқ сигнали белгилар векторини шакллантириш учун *А1 алгоритм* таклиф этилган ва у қуйидаги қадамлардан иборат.

1-қадам. Кириш сигналини нормаллаштириш, бунда унинг амплитудаси текисланиб, юқори частоталар кучайтирилади. Қуйи частота форманталари юқори частотага нисбатан катта амплитудага эга бўлади, шунга қарамай юқори частота форманталари ҳам таниб олиш учун муҳим ахборотни ўз ичига олади. Шунинг учун кириш сигналга (якуний импульс характеристика) ЯИХ-фильтри қўлланилади:

$$s'(n) = s(n) - 0.95 \cdot s(n-1).$$

2-қадам. Сигналнинг қисқа вақтли (~ 32 мс) қисмини ажратиб олиш ва спектр тушиб қолишини минималлаштириш учун ойна функцияси $w(k)$ ни қўллаш. Ҳосил бўлган бўлак K узунликдаги намунадан иборат фрейм деб номланиб, навбатдаги амаллар фрейм доирасида бажарилади:

$$x_t(k) = s'(k + t \cdot K) \cdot w(k), 0 \leq k \leq K - 1.$$

Шундай қилиб, кириш сигнали $s'(n)$ ҳар бир фрейми учун $0 \leq t \leq T$ ораликда белгилар вектори шакллантирилади. Ойна функцияси сифатида одатда Хемминг функциясидан фойдаланилади, яъни:

$$w(k) = 0.54 - 0.46 \cdot \cos\left(\frac{2\pi k}{K-1}\right)$$

3-қадам. Фрейм учун дискрет Фурье алмаштиришни ҳисоблаш:

$$X_t(m) = \sum_{k=0}^{K-1} x_t(k) \cdot e^{-j \frac{2\pi mk}{K}}$$

Дискрет Фурье алмаштиришни ҳисоблаш учун тез Фурье алмаштириш алгоритми қўлланилади.

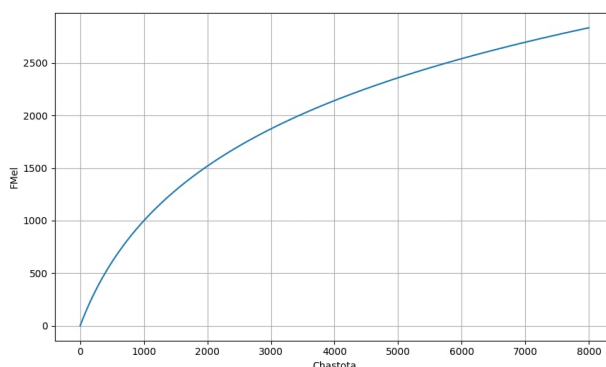
4-қадам. Частота соҳаси бўйича Q мел-фильтрлар тўпламини фреймга қўллашда ҳар бир q –фильтр чиқишида ўзининг ўтказиш полосаси бўйича $Y_t(q)$ энергия олинади. Шундай қилиб, шахснинг нутқи моделлаштирилади: спектр бўйлаб паст частотадан юқори частотага ўтишда эшитиш аниқлиги ошади. Мел-фильтрларининг марказий частоталари F_q одатдаги логарифмик

қонуниятга боғлиқ бўлган мел – шкаласи деб номланган шкала бўйича танланади (2-расм):

$$F_{mel} = 2095 \cdot \log_{10} \left(1 + \frac{F_{Hz}}{700} \right).$$

Одатда Мел-фильтрлари 0 дан $F_{Hz}^{\max} = 8kHz$ гача бўлган частота оралиқда жойлашган бўлади. Агар F_{mel}^{\max} мел – шкалада F_{Hz}^{\max} га мос келувчи частота аниқланиб, 0 дан F_{Hz}^{\max} гача бўлган оралиқ Q тенг бўлақларга бўлиниб, ҳар бир қиймат учун унга мос F_{Hz} частота мавжуд бўлса, у ҳолда Q частота қийматлари аниқланади ва улар орасидаги масофа логарифмик қонуниятга мувофиқ ўсади. Мел-шкаладан анъанавий шкалага ўтиш қуйидаги формула орқали амалга оширилади:

$$F_{Hz} = 700 \cdot \left(10^{\frac{F_{mel}}{2095}} - 1 \right).$$



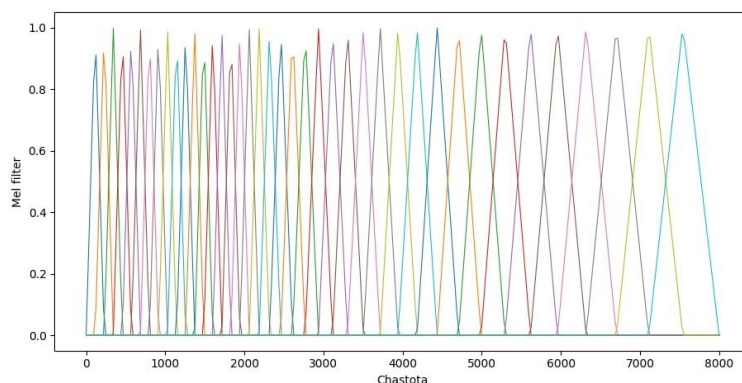
2-расм. Мел-шкаласи частоталарига оддий шкала частоталарининг мослиги

Марказий частоталар аниқлангандан сўнг, учбурчак шаклга эга бўлган филтрларнинг амплитуда-частота характеристикаларини ҳисоблаш талаб этилади (3-расм):

$$U_q(k) = \begin{cases} 1 - \frac{|k - F_q|}{\Delta_q}, & |k - F_q| < \Delta_q, \\ 0, & |k - F_q| \geq \Delta_q. \end{cases}$$

Филтрларнинг ўтказиш полосалари ҳам логарифмик қонуниятга мувофиқ ўсади, кейинги филтрнинг ўтказиш полосаси олдинги ҳар бирининг ўтказиш полосаси ўртасигача "сурилади". Бунда мел-фильтрларни устма-уст қўйиш орқали Q фреймнинг энергия қиймати ҳисобланади:

$$Y_i(q) = \sum_{k=0}^{K-1} |x_i(k)| \cdot U_q(k).$$



3-расм. Q=24 мел-фильтрларнинг АЧХси

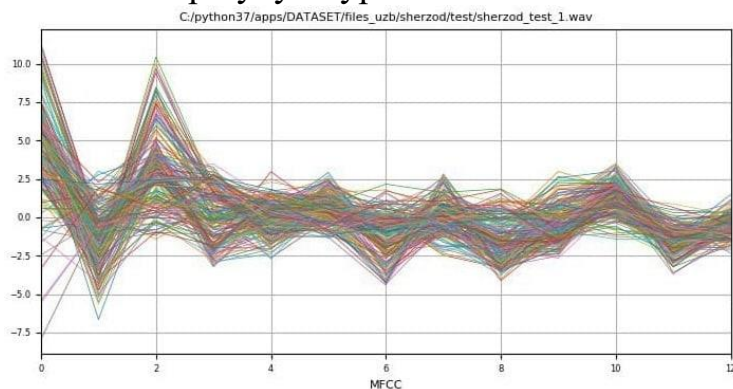
5-қadam. $Y_t(q)$ ни логарифмлаш ва гомоморфик алмаштириш амалга оширилади. Қўзғалиш сигналини фильтр характеристикасидан ажратиб олиш учун $Y_t(q)$ модули логарифми ҳисобланади:

$$L_t(q) = \log|Y_t(q)|$$

6-қadam. Тескари дискрет Фурье алмаштириши (ТДФА) амалга оширилади. Бунда $L_t(q)$ учун ТДФАни ҳисоблаш орқали кепстрал коэффицентлар олинади ва $L_t(q)$ ҳақиқий ва симметрик бўлганлиги сабабли, тескари дискрет Фурье алмаштириши дискрет косинус алмаштиришига эквивалент бўлади:

$$o_t(p) = \sum_{q=1}^Q L_t(q) \cdot \cos\left(p \cdot \left(q - \frac{1}{2}\right) \cdot \frac{\pi}{Q}\right), p = 0, \dots, P.$$

Сўнгра P коэффицентлар олинади, P ва Q тенг бўлиши мумкин бўлса-да, одатда MFCC (Mel Frequency Cepstral Coefficients) қийматларининг фақат ярмидан фойдаланилади. Бунинг сабаби, қўзғалиш сигналнинг кепструми вокал трактининг кепструмининг "ўнг томонида" бўлишидир. Шундай қилиб, t – чи фрейм учун кепстрал коэффицентларини ўз ичига олган p -ўлчамли вектор $o_t(p)$ таниб олувчига келади. 4-расмда 32 мс ли фрейм учун MFCC /a/ унлисининг бешта мисоллари учун кўрсатилган.



4-расм. /a/ унлисининг бешта мисоллари учун бир фрейм MFCCси

4-расмдан битта талаффуз қисми учун MFCC қийматлари ҳар хил талаффуз мисоллари учун яқин эканлигини кўриш мумкин.

7-қadam. Тамом.

Таклиф этилаётган белгилари сонини камайтириш ҳисобига тизим ишлаши учун сарфланадиган вақтни камайтириш **A2 алгоритми** қуйидагича амалга оширилади.

1-қадам. Нутқ сигналини қабул қилиш. Бунинг учун 16000 Гц ва 16 битда ёзилган *.wav форматдаги файл яратиш.

2-қадам. Нутқ сигналдан қуйидаги қонунният асосида белгилар ажратиб олиш.

$$s_i = \frac{1}{L} \sum_{j=i-L}^{(i+1)\cdot L} x_j$$

бу ерда x_j – берилган нутқ сигнали белгиларидан иборат вектор. L – қадам.

3-қадам. Белгиларни ажратиш яъни ҳар бир s учун Мел-частотали кепстрал коэффициентлар (MFCC) ҳисобланади.

4-қадам. Шахс нутқи моделини қуришни ва у асосида нутқ сигналини таниб олишни амалга ошириш.

5-қадам. Тамом.

Диссертациянинг «**Нутқ сигналларига дастлабки ишлов бериш алгоритмларини ишлаб чиқиш**» деб номланувчи учинчи боби нутқ сигналларини нормаллаштириш, нутқли соҳаларни аниқлаш, нутқ сигналнинг фойдали ва тўхталиш қисмини ажратиш ҳамда халақитлардан тозалаш алгоритмларини ишлаб чиқишга бағишланган.

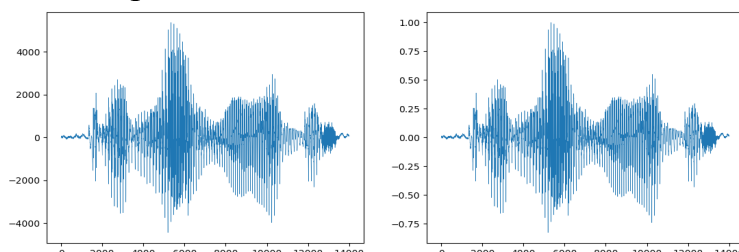
3.1-параграфда нутқ сигналларини нормаллаштириш ва нутқ сигналдан фойдали қисмини ажратиш алгоритмлари баёни келтирилган.

Овоз баландликдаги кучли ўзгаришларни йўқотиш учун одатда амплитуда бўйича нормаллаштириш процедураси қўлланилади. Ушбу усул ёрдамида сигнал амплитудаси $\left[\frac{\Delta}{2}, -\frac{\Delta}{2} \right]$ ораликқа келтирилади (5-расм).

Нормаллаштирилган сигнал $\hat{s}[n]$ намуналарини олиш қуйидаги формула асосида амалга оширилади:

$$\hat{s}[n] = \frac{\Delta}{\max_m |s[m]| + 1} \cdot s[n]$$

бу ерда Δ – нормаллаштириш бўлаги кенглиги, абцисса ўқига нисбатан симметрик (масалан, 5-расмда $\Delta = 1$).



а)

б)

5-расм. Нормаллаштиришгача (а) ва нормаллаштиришдан сўнг (б) рақамлаштирилган нутқ сигнали

Овоз баландлиги ўзгарувчанлигини баҳолаш учун бир ёки бир нечта сўзлар талаффузининг Q та тўплами кўриб чиқилади q –чи намуна учун $M(q)$, баландликнинг ўртача қийматини N намуналар узунлиги билан ва Q намуна учун $M(q)$ ўртача қиймати қуйидагича аниқланади:

$$M(q) = \sum_{n=1}^N |s[n]|, \quad q = 1, \dots, Q; \quad (1)$$

$$M_Q = \frac{1}{Q+1} \sum_{q=1}^Q M(q). \quad (2)$$

Ҳар бир намуна баландлиги ва ўрта қиймат орасидаги нисбий фарқ $D(q)$ ҳисобланади:

$$D(q) = \left| 1 - \frac{M(q)}{M_Q} \right|.$$

(1), (2) формулалардан натижага намунанинг абсолют қиймати ва уларнинг умумий сони ҳам таъсир қилишини кўриш мумкин. Шунинг учун битта синфга тегишли бўлган намуналар тўплами баландлигини ўзгарувчанлигини, унинг узунлиги тахминан бир хил бўлиши ва тўлиқ базадаги овоз баландлиги ўзгарувчанлигини баҳоланиши талаб этилади.

Экстремумлар тақсимотининг хусусиятидан фойдаланиб, кириш сигналдан нутқ сигнални ажратиш олиш **АЗ алгоритми** ишлаб чиқилган.

Бу алгоритм қуйидагича амалга оширилади.

1-қadam. Нормаллаштириш модули – микрофон орқали олинган сигнал амплитудаси 10 мс кенгликдаги сурилувчи ойна орқали нормаллашади. Ушбу процедура овоз ва қурилманинг индивидуал хусусиятларига кўра сигнал баландлиги ўзгарувчанлигини бартараф этиш учун зарур. Барча қисмлар бир хилда овоз баландлиги қийматига кўтарилиши учун ойна ичида нормаллаштириш амалга оширилади.

2-қadam. Сигнални қамраб олувчи эгри чизикни қуриш модули қўлланилади. Бунда паст амплитудали тебранишларни текисланади.

3-қadam. Нормаллаштириш модули қўлланилади. Бунда қамраб олувчи эгри чизик қурилиши туфайли сигнални такрорий нормаллаштириш керак.

4-қadam. m -қисм учун экстремумлар тақсимотини топиш модули қўлланилади.

5-қadam. Ҳал қилувчи мантиқ модули қўлланилади. Бунда сигналнинг m - қисми нутқ ёки халақитга эга эканлиги бўйича қарор қабул қилинади. Ҳал қилувчи мантиқ сифатида уч қатламли тўлиқ боғланган тўғри тарқалувчи нейрон тармоғи танланди. Шу билан бирга, тармоқни ишлатишдан олдин, у ўқитилиши керак, унинг учун ўқитиш модули жавоб беради.

3.2-параграфда ишлаб чиқилган сегментлаш алгоритмларининг баёни келтирилган. Сигналнинг сўнувчи қисмини ажратиш алгоритмларини ҳисоблаш мураккаблигини камайтириш учун биринчи қadamда нутқ сигнали

локал максимумлари тўплами M қуйидаги қоида бўйича шакллантирилади, бунда дастлаб $M = \emptyset$ деб олинади.

Агар $(s(i) > s(i-1) \text{ ва } s(i) > s(i+1))$ бўлса, у ҳолда $M = M \cup i, i = \overline{1, N}$, бу ерда i – нутқ сигналининг бошланғич саноклари сони, $s(i)$ – нутқ сигнали қийматлари. Кейинчалик сўнувчи қисмнинг санок рақамлари итератив кўринишда M тўплам элементларидан олинади. Агар $s(n)$ -да сўнувчи қисмларнинг жорий санок сонлари сифати қабул қилинган бўлса, у ҳолда кейинги $s(n+1)$ сўнувчи қисм саноклари қуйидагича аниқланади:

$$\begin{cases} K = \{m \in M \mid n < m < n + L\}, \\ s(n+1) \in K, \\ \forall k \in K \mid s(n+1) \geq s(k), \\ \{k \in K \mid (s(n+1) = s(k) \wedge (n+1) < k)\} = \emptyset. \end{cases}$$

Бунда L параметрдан сўнувчи қисм саноклар ўртасидаги масофасини ошиб кетмаслигини таъминлаш мақсадида фойдаланилади.

Таклиф этилаётган нутқ сигналнинг сўнувчи қисмларини олиб ташлаш **A4 алгоритми** қуйидагича амалга оширилади.

1-қadam. Нутқ файли асосида f вектор шакллантирилади.

2-қadam. ε га қиймат ўзлаштирилади.

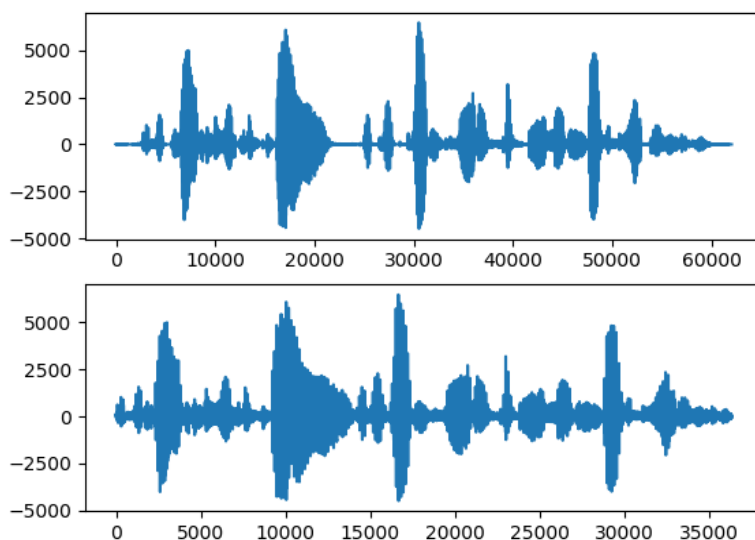
3-қadam. унинг қиймати ҳисобланади:

$$y = \begin{cases} 0, & f(i) < \varepsilon, \\ 1, & f(i) \geq \varepsilon. \end{cases}$$

4-қadam. f вектор ўлчамини камайтириш амалга оширилади, яъни $y=0$ бўлса, у ҳолда унга мос элемент f вектордан чиқариб юборилади.

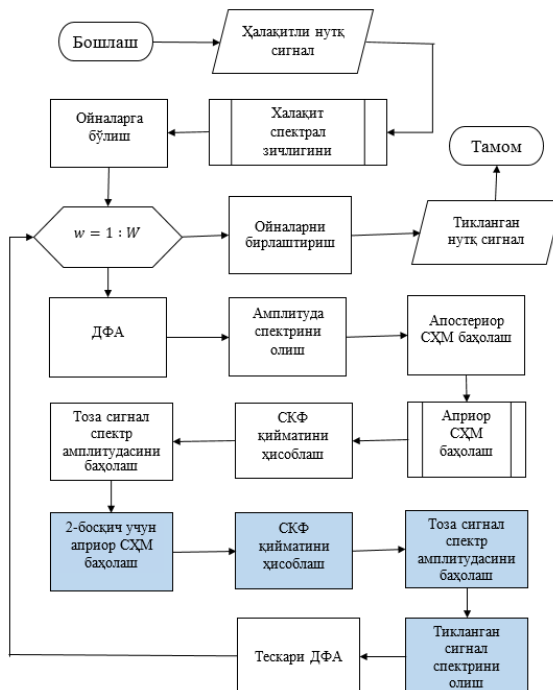
5-қadam. Тамом.

Масалан, $\varepsilon = 50$ учун натижа 6-расмда келтирилган:



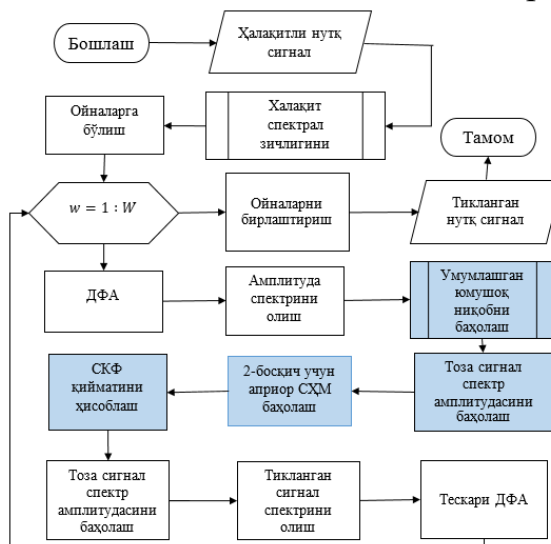
6-расм. Сўнувчи қисмли ва сўнувчи қисмсиз нутқ сигнали

3.3-параграфда халақитлардан тозалашнинг такомиллаштирилган алгоритмлари баёни келтирилган бўлиб, у ишда **A6** алгоритм деб номланган. Халақитни камайтириш икки босқичли **A6 алгоритмининг** блок-схемаси 7-расмда келтирилган.



7-расм. Халақитни камайтириш икки босқичли алгоритми блок схемаси

Таклиф этилган юмшоқ ниқобга асосланган икки босқичли **A8 алгоритми** қуйидаги блок-схема асосида амалга оширилади.



8-расм. Юмшоқ ниқобга асосланган халақитни пасайтириш алгоритмининг блок-схемаси

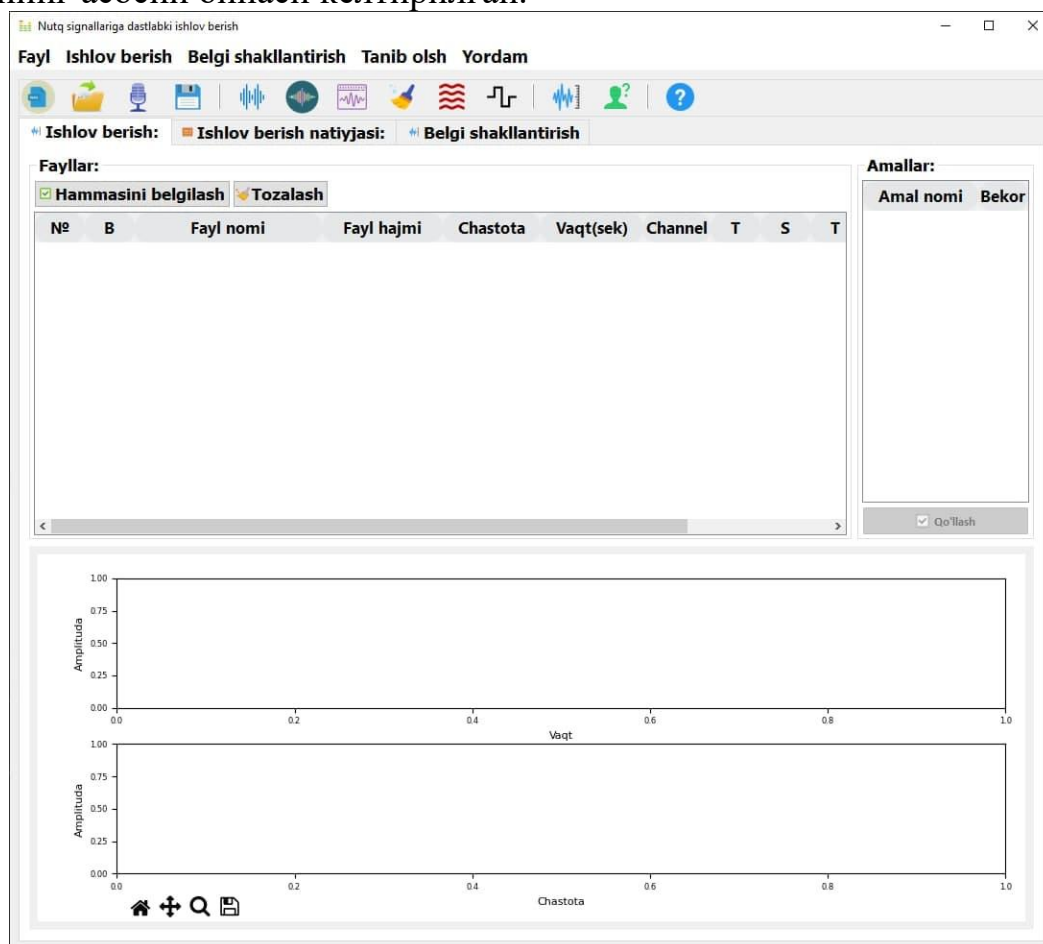
Халақитни камайтириш алгоритмини модификацияланган икки босқичли халақитни пасайтириш алгоритмининг биринчи босқичи сифатида қўлланилиши мумкин. Иккинчи босқич учун спектрни тўғрилаш функцияси

сифатида қисқа муддатли спектр амплитудасининг минимал ўртаквadratic хатолиги спектрни коррекциялаш функциясини танлаш таклиф этилади.

Диссертациянинг «**Дастурий мажмуа ва унинг амалий масалаларда қўлланилиши**» деб номланувчи тўртинчи боби нутқ сигналларига дастлабки ишлов бериш дастурий мажмуасини яратиш, тажрибавий тадқиқотлар натижаларини баён этиш ва амалий масалаларни ечишга бағишланган.

4.1-параграфда таклиф этилган, мавжуд ва такомиллаштирилган алгоритмлар асосида ишлаб чиқилган дастурий мажмуа баёни келтирилган бўлиб, у нутқ сигналларига дастлабки ишлов бериш, нутқ сигналлари белгиларини ажратиш олиш ва шахсни таниш олишга мўлжалланган.

9-расмда модуллилик тамойили асосида ишлаб чиқилган дастурий мажмуанинг асосий ойнаси келтирилган.



9-расм. Дастурий мажмуанинг асосий ойнаси

4.2-параграфда тажрибавий тадқиқот натижалари келтирилган. MFCC ва LPC (Linear Predictive Coding) коэффицентларини ҳисоблаш алгоритмларини модификация натижаларини тадқиқот қилиш учун Pythonда оддий шароитда ёзилган чекланган ҳар хил матнларининг товуш сўзларни таниш олиш тизимидан фойдаланилган. Ушбу ишда нутқ фонограммаларининг етарлича катта товуш матнлари базасидан фойдаланиш зарурати туфайли, https://www.dropbox.com/s/87v8jxxu9tvbkns/development_set.zip?dl=0 ва ўзбек тилидаги маълумотлар базасидан дискретизация частотаси 16кГц бўлган 170та ҳар хил матн фонемаларидан фойдаланишга қарор қилинди.

Базани шакллантиршда 34та суҳандон овозидан фойдаланилган. Синов ва тестлаш гуруҳларнинг суҳандонлари кесишмаган. Натижалар қуйида 1-жадвалда келтирилган.

1-жадвал

Такомиллаштирилган MFCC+VQ, LPC+VQ ва MFCC+GMM алгоритмлари асосида шахсни овози орқали таниб олиш аниқлиги

	Асосий	2->1	4->1	8->1	16->1
LPC+VQ	73%	82%	81%	79%	72%
MFCC+VQ	84%	89%	85%	78%	73%
MFCC+GMM	86%	92%	87%	80%	84%

1-жадвалда кўриб чиқилган модификацияланган алгоритмларининг таниб олишдаги натижалари MFCC алгоритмига нисбатан яхшиланганлиги кўрсатилган. R нинг нисбий яхшиланиши қуйидаги формула бўйича ҳисобланади:

$$RI = \frac{RR_A - RR_{MFCC}}{100 - RR_{MFCC}} \times 100\%.$$

Такомиллаштирилган MFCC+VQ ва LPC+VQ коэффициентларини ҳисоблаш алгоритмларини 5 та суҳандон томонидан студияда бирхил матндан ёзиб олинган товушларини 15 та файли таниб олиш тадқиқот натижалари 2-жадвалда келтирилган. Такомиллаштирилган MFCC+VQ

2-жадвал

Такомиллаштирилган MFCC+VQ, LPC+VQ ва MFCC+GMM алгоритмлари асосида шахсни овози орқали таниб олиш аниқлиги

	Асосий	2->1	4->1	8->1	16->1
LPC+VQ	87%	93%	93%	90%	86%
MFCC+VQ	85%	85%	83%	90%	89%
MFCC+GMM	98%	99%	99%	97%	96%

4.3-параграфда дастурий мажмуадан фойдаланиб амалий масалаларни ечиш натижалари келтирилган.

Мавжуд ва диссертация ишида таклиф этилган усул ва алгоритмлар асосида нутқ сигналларидаги ҳалақитларни камайтириш ва силлиқлаш, нутқ сигналени нормаллаштириш ва сегментлаш, нутқ жойлашган соҳани аниқлаш, белгилар фазосини шакллантириш ва таниб олиш масаласи ҳал этилган. Бу аудиомаълумотларни таҳлил қилиш учун сарфланадиган вақтни ўртача 15% га қисқартиришга ва уни тушуниш самарадорлигини ўртача 20% га оширишга имконини берди.

Олинган натижалар мазкур соҳа мутахассислари томонидан яхши қабул қилинди. Хулосалар эса кейинги тадқиқотларда нутқни матнга ўтказиш ва шахсни таниб олиш самарадорлигини ошириш ҳамда информатив белгиларга нисбатан олиб борилиши орқали таниб олиш даражасини ошириш мақсадга мувофиқ деб топилди. Таклиф этилган усул ва алгоритмлар асосида нутқ технологияларини ривожлантириш орқали амалий тадқиқотларда келажакдаги ишларни режалаштириш кўплаб соҳа мутахассисларига қулайликлар туғдирди ва ортиқча сарф-харажатларни камайтирди.

ХУЛОСА

“Нутқ сигналларига дастлабки ишлов бериш алгоритмлари” мавзусида олиб борилган диссертация тадқиқотининг асосий натижалари қуйидагилардан иборат:

1. Нутқ сигналларига дастлабки ишлов бериш мавжуд алгоритмлари таҳлил қилинди ва уларнинг кучли томонлари ҳамда камчиликлари аниқлаб олинди. Бу мавжуд камчиликларни бартараф этиш имконини беради.

2. Нутқ сигналини сегментларга ажратишнинг турли халақитларга турғун ва тезкор бўлган алгоритмлари ишлаб чиқилди. Ушбу алгоритмлар турли хил шароитлардан ёзиб олинган нутқ сигналларини қайта ишлаш имконини беради.

3. Частоталар соҳасида ва вақтли соҳада белгилар векторини шакллантиришда турли хил ёндашувлар таҳлил қилинди ва белгилар векторини шакллантириш усули сифатида такомиллаштирилган мел кепстрал коэффициентлар усули мақбул эканлиги аниқланди. Бу нутқни таниб олиш самарадор тизимларини ишлаб чиқиш имконини беради.

4. Тўпланган экспериментал нутқ базасида кўплаб омиллар туфайли юзага келадиган нутқ сигналлари баландлигининг ўзгарувчанлиги ўрганилди: қурилмаларнинг сезгирлиги, овознинг доимий равишда ўзгариб турадиган баландлиги ва бошқалар. Нутқ баландлиги ўзгарувчанлигини камайтириш имконини берадиган нормаллаштириш алгоритми таклиф этилди.

5. Халақитли нутқ сигнали чегараларини аниқлашга имкон берадиган кириш сигналнинг локал экстремумлари тақсимотини таҳлил қилиш асосида нутқни ажратиб олишнинг янги усули таклиф этилди. Нормаллаштириш, ҳал қилувчи мантиқ, қамраб олувчи эгри чизиқни топиш ва экстремумлар тақсимотини топиш блокларидан иборат бўлган нутқни ажратиб олиш алгоритми таклиф этилди. Мазкур алгоритм нутқни таниб олиш учун нисбатан информатив нутқ сигналини олиш имконини беради.

6. Частоталар соҳасида халақитни камайтириш усуллари кўриб чиқилди. Нутқ ва халақитларни моделлаштириш, спектр коррекция функциясини танлаш, сигнал/халақит муносибатининг априор баҳолаш каби масалалар кўриб чиқилди. Нутқ сигналида халақит ва халақитларни камайтириш учун бинар ниқоблардан фойдаланиш, шу билан бирга халақитларни камайтиришдаги янги йўналиши юмшоқ ниқоблар усуллари кўриб чиқилди. Бундан ташқари такомиллаштирилган частотали соҳада халақитни камайтириш икки босқичли алгоритми ва юмушоқ ниқобга асосланган икки босқичли алгоритми таклиф этилди.

7. Нутқ сигналларига дастлабки ишлов беришнинг мавжуд ва таклиф этилган алгоритмларни ягона интерфейсга бирлаштирувчи дастурий мажмуаси ишлаб чиқилди. Дастурий мажмуани ишлаб чиқишда унинг ишончлилиги билан бирга тезкорлик масаласи ҳам эътиборга олинди.

**РАЗОВЫЙ НАУЧНЫЙ СОВЕТ ПРИ НАУЧНОМ СОВЕТЕ
DSc13/30.12.2019.Т.07.01 ПО ПРИСУЖДЕНИЮ УЧЕНЫХ СТЕПЕНЕЙ
ПРИ ТАШКЕНТСКОМ УНИВЕРСИТЕТЕ ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ**

**НАУЧНО-ИННОВАЦИОННЫЙ ЦЕНТР ИНФОРМАЦИОННО-
КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ ПРИ ТАШКЕНТСКОМ
УНИВЕРСИТЕТЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ**

ЮЛДОШЕВ ЮСУФ ШЕРАЛИЕВИЧ

**АЛГОРИТМЫ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ РЕЧЕВЫХ
СИГНАЛОВ**

05.01.03 - Теоретические основы информатики

**АВТОРЕФЕРАТ ДИССЕРТАЦИИ
ДОКТОРА ФИЛОСОФИИ (PhD) ПО ТЕХНИЧЕСКИМ НАУКАМ**

Ташкент – 2021

Тема диссертации доктора философии (PhD) по техническим наукам зарегистрирована в Высшей аттестационной комиссии при Кабинете Министров Республики Узбекистан за В2020.2.PhD/T894.

Диссертация выполнена в Научно-инновационном центре информационно-коммуникационных технологий при Ташкентском университете информационных технологий.

Автореферат диссертации на трех языках (узбекский, русский, английский (резюме)) размещен на веб-странице научного совета (www.tuit.uz) и на Информационно-образовательном портале «ZiyoNet» (www.ziyo.net).

Научный руководитель: **Бабомурадов Озод Жураевич**
доктор технических наук

Официальные оппоненты: **Мухамедиева Дилноз Тулкуновна**
доктор технических наук, профессор

Мирзаев Олимжон Намозович
доктора философии по техническим наукам

Ведущая организация: **Национальный университет Узбекистана**

Защита диссертации состоится «__» _____ 2021 г. в __ часов на заседании Научного совета DSc.27.06.2017.T.07.01 при Ташкентском университете информационных технологий. (Адрес: 100202, г. Ташкент, ул. Амира Темура, 108. Тел.: (99871) 238-64-43; факс: (99871) 238-65-52; e-mail: tuit@tuit.uz).

С диссертацией можно ознакомиться в Информационно-ресурсном центре Ташкентского университета информационных технологий (регистрационный номер №__). (Адрес: 100202, г.Ташкент, ул. Амира Темура, 108. Тел.: (99871) 238-65-44).

Автореферат диссертации разослан «__» _____ 2021 года.
(протокол рассылки №__ от «__» _____ 2021 г.).

Р.Х.Хамдамов

Председатель научного совета по присуждению
учёных степеней, д.т.н., профессор

Ф.М.Нуралиев

Ученый секретарь научного совета по присуждению
учёных степеней, д.т.н., доцент

М.А.Исмаилов

Председатель научного семинара при научном совете по
присуждению ученых степеней д.т.н., профессор

ВВЕДЕНИЕ (аннотация диссертации доктора философии (PhD))

Актуальность и востребованность темы диссертации. В мире большое внимание уделяется усовершенствованию, разработке и внедрению методов и алгоритмов цифровой обработки речевых сигналов, идентификации личности на основе речевых сигналов, преобразования речи в текст и распознавания образов. «В последние годы создание речевых интеллектуальных систем стало одним из самых быстрорастущих направлений, использующим методы искусственного интеллекта для решения проблем, которые невозможно решить с помощью традиционных методов»¹. В этом направлении основными функциональными задачами интеллектуального анализа речевых сигналов являются классификация, кластеризация, распознавание образов, прогнозирование и оценка, определение закономерностей, разработка методов и алгоритмов принятия решений, а также разработка приложений. К настоящему времени разработано множество приложений, основанных на технологии распознавания речевых сигналов. В зарубежных странах, в том числе в Китае, России, США, Англии, Германии, Индии, Франции и др., большое внимание уделяется решению теоретических и практических задач предварительной обработки речевых сигналов и распознавания речи.

В мире ведутся исследовательские работы по усовершенствованию и созданию методов и алгоритмов обработки речевых сигналов, выделения признаков речевых сигналов и распознавания речи, а также созданию вычислительных алгоритмов. В этой связи, важной задачей является разработка и усовершенствование алгоритмов распознавания, повышения качества речевых сигналов, определения области полезной речи в речевом сигнале и выделения его характерных признаков, а также создание речевых моделей и автоматизированных систем распознавания.

В Республике особое внимание уделяется разработке автоматизированных систем распознавания, интеллектуального анализа и предварительной обработки речи на узбекском языке. В Стратегии действий по дальнейшему развитию Республики Узбекистан в 2017-2021 гг. определены, в частности, задачи по «... внедрению информационно-коммуникационных технологий в экономику, социальную сферу, систему управления, ... обеспечению информационной безопасности и усовершенствованию систем защиты информации, организацию своевременного и адекватного реагирования на угрозы в сфере информации»². Для реализации данных задач важными проблемами развития информационно-коммуникационных технологий являются, в частности, создание систем идентификации личности на основе речевых сигналов, преобразования речи в текст и голосового управления. В этой связи, одним из важных задач является разработка методов и алгоритмов предварительной обработки речевых сигналов с учетом распознавания личности и определения

¹ https://www.tadviser.ru/index.php/Статья:Речевые_технологии:_на_пути_от_распознавания_к_пониманию

² Указ Президента Республики Узбекистан от 7 февраля 2017 года №УП-4947 «О стратегии по дальнейшему развитию Республики Узбекистан»

информативных признаков речевых сигналов, а также их применение в системах контроля доступа и анализа голосовых обращений.

Настоящее диссертационное исследование в определенной степени служит для реализации задач, предусмотренных указом Президента Республики Узбекистан №УП-4947 от 7 февраля 2017 года “О стратегии действий по дальнейшему развитию Республики Узбекистан”, постановлениями Президента №ПП-2898 от 18 апреля 2017 года “О мерах по коренному совершенствованию деятельности органов внутренних дел в сфере расследования преступлений”, №ПП-4024 от 21 ноября 2018 года «О мерах по совершенствованию системы контроля за внедрением информационных технологий и коммуникаций, организации их защиты», а также другими нормативно-правовыми актами, принятыми в данной сфере.

Соответствие исследования с приоритетными направлениями развития науки и технологий республики. Данное исследование выполнено в соответствии с приоритетным направлением развития науки и технологий IV. “Развитие информатизации и информационно-коммуникационных технологий”.

Степень изученности проблемы. Заслуживают внимания научные работы решения задач по разработке и усовершенствованию методов, алгоритмов и моделей предварительной обработки и распознавания речевого сигнала таких зарубежных учёных, как Михайлов В.Г., Сапожков М.А., Назаров М.В., Прохоров Ю.Н., Ланнэ А. А., Грэй А. Х., Маркел Дж.Д., Рабинер Л. Р., Ситняковский И.В., Шафер Р.В., Шалимов И.А., Петровский А.А., Ковалгин Ю.А., Итакура Ф., Рейдер Ч., Макхоул Дж и др.

В Узбекистане в развитие теоретических основ распознавания образов и анализа речевых сигналов внесли свой вклад М.М.Камилов, Ш.Х.Фазылов, Э.М.Алиев, М.М.Мусаев, Н.А.Игнатъев, С.С.Садыков, Н.С.Маматов, Н.М.Мирзаев, Ш.Е.Туляганов и другие учёные.

На сегодняшний день теоретические основы идентификации личности и технологии распознавания речи развиваются стремительно. Анализ исследований в этой области показывает, что технологии распознавания речи являются одной из наиболее перспективных технологий. Однако проблемы, возникающие при создании автоматизированных систем идентификации личности и распознавания речи на основе таких технологий, в настоящее время в достаточной степени не решены. Кроме того, недостаточно изучена проблема разработки методов и алгоритмов, обеспечивающих надежную, быструю идентификацию личности и дикторнезависимое распознавание речи на основе предварительной обработки речевого сигнала.

Связь диссертационного исследования с планами научно-исследовательских работ научно-исследовательского учреждения, где выполнена диссертация. Диссертационное исследование выполнено в рамках проектов плана научно-исследовательских работ Научно-инновационного центра информационно-коммуникационных технологий при Ташкентском университете информационных технологий имени Мухаммада ал-Хоразмий и Ташкентского университета информационных технологий

имени Мухаммада ал-Хоразмий по следующим темам: № БВ-Атех-2018 «Разработка алгоритмов и программного обеспечения идентификации личности на основе потоковой обработки изображений лиц» (2018-2020), № М-Ф4-09 «Интеллектуальные системы обработки сложноструктурированной информации, распознавания и прогнозирования основанные на классе логико-эвристических алгоритмов вычисления оценок в нечеткой среде» (2015-2020), № ПЗ-201906202 «Разработка системы анализа и классификации текстовой информации на узбекском языке» (2019-2020).

Целью исследования является разработка алгоритмов и программного комплекса предварительной обработки речевых сигналов.

Задачи исследования:

разработка алгоритма сглаживания и шумоподавления в речевом сигнале; усовершенствование метода обработки речевого сигнала в частотной области;

разработка алгоритма выделения участков с речью в исходном сигнале; разработка алгоритмов нормализации и сегментации речевого сигнала; формирование пространства признаков речевого сигнала для решения задач идентификации личности по голосу и распознавания речи; создание алгоритмического и программного обеспечения предварительной обработки речевого сигнала.

Объектом исследования являются цифровые речевые сигналы.

Предмет исследования составляют методы, алгоритмы и программное обеспечение предварительной обработки сигналов, распознавания образов.

Методы исследования. Теоретические исследования в работе базировались на методах системного анализа, имитационного моделирования, теории вероятностей, математической статистики, дискретной математики, обработка сигналов и распознавания образов.

Научная новизна исследования заключается в следующем:

разработан алгоритм сглаживания и шумоподавления в речевом сигнале, учитывающий спектральную плотность этого шума;

разработан алгоритм выделения участков с речью в исходном сигнале на основе закономерностей распространения экстремумов;

модифицирован метод обработки речевого сигнала в частотной области на основе определения отношения сигнал-шум;

разработаны алгоритмы нормализации и сегментации, учитывающие влияющие факторы на качество речевого сигнала;

разработан алгоритм формирования пространства признаков речевого сигнала, учитывающий его дискретность, для решения задач идентификации личности по голосу и распознавания речи.

Практические результаты исследования заключаются в следующем:

разработаны алгоритмы сглаживания, шумоподавления, нормализации, сегментации, формирования пространства признаков речевого сигнала, а также метод обработки речевого сигнала в частотной области для решения задач идентификации личности по голосу и распознавания речи;

разработан программный комплекс “NID” на основе известных и

предложенных методов и алгоритмов, предназначенный для предварительной обработки речевых сигналов и идентификации личности по голосу.

Достоверность результатов исследования обосновывается корректностью применения математического аппарата предварительной обработки речевых сигналов и распознавания образов при разработке алгоритмов, а также положительными результатами экспериментальных.

Научная и практическая значимость результатов исследования. Научная значимость результатов исследования обосновывается вкладом разработанных алгоритмов в перспективное развитие теоретических основ технологий биометрической идентификации по голосу.

Практическая значимость результатов исследования заключается в том, что разработанный программный комплекс можно использовать при создании автоматизированных систем идентификации личности по голосу и распознавания речи.

Внедрение результатов исследования. На основе программного обеспечения, созданного на базе известных, а также предложенных в работе методов и алгоритмов предварительной обработки речевых сигналов:

программный комплекс, разработанный на основе методов и алгоритмов формирования пространства признаков речевого сигнала для решения задач сглаживания, шумоподавления, обработки речевого сигнала в частотной области и распознавания речи внедрен в хокимияты города Шахрисабз и Камашинского района Кашкадарьинской области (справка Министерства по развитию информационных технологий и коммуникаций №33-8/7838 от 22 декабря 2020 г.). На основе использования результатов научного исследования время предварительной обработки, сглаживания, фильтрации, очистки речевых сигналов от различных шумов для обеспечения более четкого звучания речи и анализа аудиоданных сократилось на 15% и точность восприятия речи повысилась в среднем на 20%;

программный комплекс, разработанный на основе методов и алгоритмов сглаживания, формирования пространства признаков речевого сигнала для решения задач нормализации, сегментации, уменьшения помех речевых сигналов и распознавания речи внедрен в Шахрисабзском межрайонном суде по гражданским делам и в Шахрисабзском РОВД Кашкадарьинской области (справка Министерства по развитию информационных технологий и коммуникаций №33-8/7838 от 22 декабря 2020 г.). На основе использования результатов научного исследования время анализа аудиоданных сократилось в среднем на 15%.

Апробация результатов исследования. Теоретические и практические результаты диссертационного исследования докладывались и обсуждались на 5 международных и 6 республиканских научно-практических конференциях.

Опубликованность результатов исследования. По теме диссертации опубликовано 22 научные работы, в том числе 11 статей в научных изданиях, рекомендованных Высшей аттестационной комиссией Республики Узбекистан для публикации основных научных результатов докторских

диссертаций, из них 8 в республиканских и 3 в зарубежных журналах, получено 2 свидетельства об официальной регистрации программы для ЭВМ.

Структура и объем диссертации. Диссертация состоит из введения, четырёх глав, заключения, списка использованной литературы и приложений. Объем диссертации составляет 118 страниц.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении обоснованы актуальность и востребованность темы диссертации, определены цель и задачи, объект и предмет исследования, приводится соответствие исследования приоритетным направлениям развития науки и технологий Республики Узбекистан, изложены научная новизна, практические результаты исследования, обоснована достоверность полученных результатов, раскрывается теоретическая и практическая значимость результатов исследования, приведены сведения о внедрении результатов исследования, об опубликованности результатов и структуре диссертации.

Первая глава диссертации **“Проблема предварительной обработки речевых сигналов”** посвящена изучению современного состояния проблемы предварительной обработки речевых сигналов. Также в данной главе приведены требования к исходным речевым сигналам, проведен анализ известных методов обработки этих сигналов, приведены их достоинства и недостатки, а также основные проблемы в области обработки речевых сигналов. Кроме того, рассмотрены вопросы распознавания личности на основе анализа речевых сигналов, приведены способы их решения, а также постановка задачи диссертационной работы.

В параграфе 1.1 на основе анализа таких параметров как свойства канала передачи речи, размер словарного запаса, изменчивость речи, уровень окружающего шума, тип речи, приведены требования к речевым сигналам, являющимся входными данными для систем автоматического распознавания речи.

В параграфе 1.2 приведены основные требования к системам обработки речевого сигнала на основе анализа особенностей речевого сигнала пользователя и проблем при создании систем распознавания речи.

В параграфе 1.3 проведен анализ методов предварительной обработки речевых сигналов, классифицированы компоненты речи, приведены выявленные недостатки существующих систем обработки речевых сигналов, обоснована необходимость разработки алгоритмов шумоподавления и усовершенствования существующих алгоритмов с учетом моделирования речи и шума, функции коррекции спектра и периода стационарности речи.

В параграфе 1.4 предложены этапы повышения устойчивости систем идентификации диктора к помехам, на основе которых формулируются цель и задачи диссертационного исследования.

Вторая глава диссертации **“Анализ методов и алгоритмов предварительной обработки речевых сигналов”** состоит из 3 параграфов и

в данной главе рассмотрены известные методы и алгоритмы нормализации речевого сигнала, выделения на нем участков, в которых находится речь, обнаружения помех, сегментации и оценки характерных параметров этого сигнала.

В параграфе 2.1 проведен анализ методов и алгоритмов дискретизации и нормализации речевых сигналов, приведены их достоинства и недостатки.

При решении задачи распознавания речи алгоритм автоматической сегментации на фрагменты неравной длительности для задачи распознавания, общая последовательность работы алгоритмов сегментации должна быть следующей:

1. детектированные типы сегментов речевого сигнала (классификация фрагментов речевого сигнала на паузы, шумные, взрывные звуки);
2. уточнение временных границ сегментов (алгоритм повторения);
3. параметризация сегментов речевого сигнала;
4. анализ и корректировка содержания сегментов речевого сигнала (детальный анализ нижних сегментов, которые можно разделить на сегменты, соответствующие отдельным фонемам);
5. классификация выбранных фонем.

В параграфе 2.2 проведен анализ методов и алгоритмов сегментации речевых сигналов, основанных на различных математических подходах, а также предложена следующая классификация алгоритмов оценки периода основного тона при решении задачи сегментации речевого сигнала.

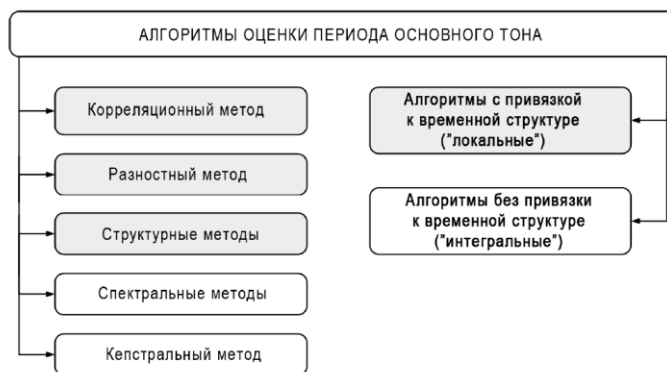


Рис.1. Классификация алгоритмов оценки периода основного тона речевого сигнала

В параграфе 2.3 рассмотрены известные алгоритмы формирования признаков речевого сигнала и предложен алгоритм, обозначенный через А1 и состоящий из следующих шагов.

Шаг 1. Осуществляется нормализация входного сигнала, при которой его амплитуда выравнивается, а высокие частоты усиливаются. Низкочастотные форманты имеют большую амплитуду, чем высокочастотные, хотя высокочастотные форманты также содержат важную информацию для распознавания речи. Поэтому для входного сигнала применяется фильтрация конечной импульсной характеристики (КИХ):

$$s'(n) = s(n) - 0.95 \cdot s(n-1).$$

Шаг 2. Применяется оконная функция $w(k)$ для выделения кратковременной части сигнала (~ 32 мс.) и минимизации падения спектра. Полученный отрезок называется фреймом отсчета длины K , и последующие операции выполняются в рамках фрейма:

$$x_t(k) = s'(k + t \cdot K) \cdot w(k), 0 \leq k \leq K - 1.$$

Таким образом, для каждого фрейма входного сигнала $s'(n)$ выбирается вектор признаков $0 \leq t \leq T$. Функция Хемминга обычно используется как оконная функция:

$$w(k) = 0.54 - 0.46 \cdot \cos\left(\frac{2\pi k}{K-1}\right).$$

Шаг 3. Вычисляется дискретное преобразование Фурье (ДПФ) для фрейма:

$$X_t(m) = \sum_{k=0}^{K-1} x_t(k) \cdot e^{-j2\pi mk/K}.$$

Для вычисления ДПФ используется алгоритм быстрого преобразования Фурье (БПФ).

Шаг 4. Применяется набор Q мел-фильтров к фрейму в частотной области и на выходе каждого фильтра q по его полосе пропускания получается энергия $Y_t(q)$. Таким образом, моделируется восприятие человеческой речи: четкость звучания речи увеличивается по мере перехода от низкой частоты к высокой по всему спектру. Центральные частоты F_q мел-фильтров выбираются по шкале, называемой мел-шкалой и зависимой от общего логарифмического закона (рисунок 2):

$$F_{mel} = 2095 \cdot \log_{10}\left(1 + \frac{F_{Hz}}{700}\right).$$

Мел-фильтры изменяются в диапазоне частот от 0 до $F_{Hz}^{\max} = 8kHz$. Таким образом, если соответствующая частота F_{Hz}^{\max} найдена на мел-шкале F_{mel}^{\max} , то интервал от 0 до F_{Hz}^{\max} делится на равные Q частей, и для каждого значения определяется соответствующая частота F_{Hz} . В результате мы получаем значения частоты Q , расстояние между которыми увеличивается по логарифмическому закону. Переход от мел-шкалы к традиционной шкале осуществляется по следующей формуле:

$$F_{Hz} = 700 \cdot \left(10^{\frac{F_{mel}}{2595}} - 1\right).$$

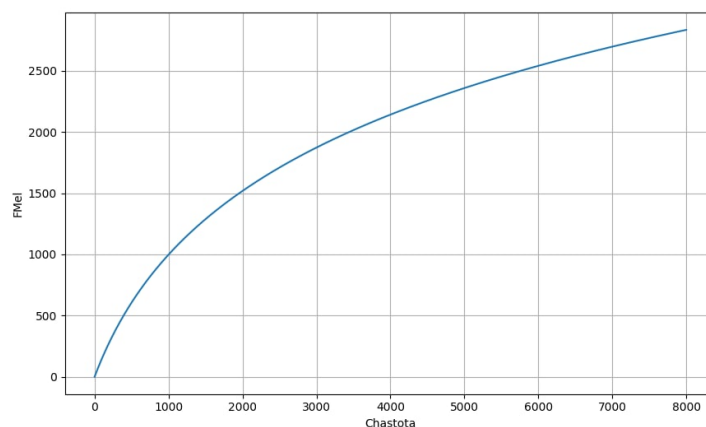


Рис.2. Соотношение частот обычной шкалы к частотам мел-шкалы

После определения центральных частот необходимо построить амплитудно-частотные характеристики (АЧХ) фильтров треугольной формы (рисунок 3):

$$U_q(k) = \begin{cases} 1 - \frac{|k - F_q|}{\Delta_q}, & |k - F_q| < \Delta_q, \\ 0, & |k - F_q| \geq \Delta_q. \end{cases}$$

Полоса пропускания фильтров также увеличивается по логарифмическому закону, при этом полоса пропускания следующего фильтра «подталкивается» к середине полосы пропускания предыдущего. В этом случае значение энергии фрейма Q вычисляется путем наложения мел-фильтров:

$$Y_t(q) = \sum_{k=0}^{K-1} |x_t(k)| \cdot U_q(k).$$

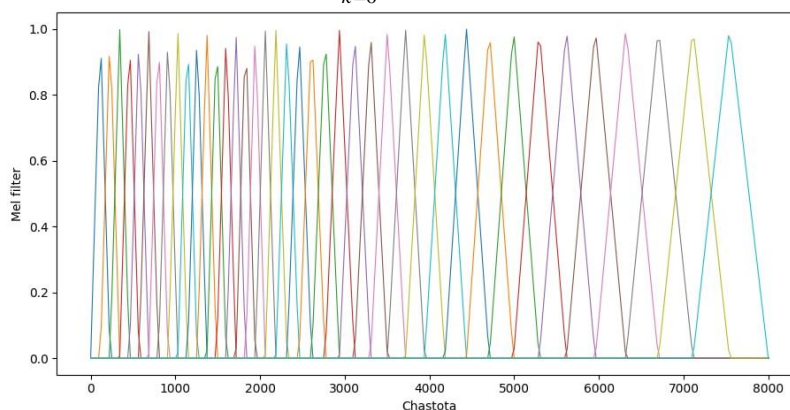


Рис.3. АЧХ мел-фильтров $Q=24$

Шаг 5. Производится логарифмизация и гомоморфная замена $Y_t(q)$. Для отделения сигнала возбуждения от характеристики фильтра, вычисляется логарифм модуля $Y_t(q)$:

$$L_t(q) = \log |Y_t(q)|.$$

Шаг 6. Производится обратное дискретное преобразование Фурье (ОДПФ). Кепстральные коэффициенты получают путем вычисления ОДПФ

для $L_t(q)$, и, поскольку $L_t(q)$ является действительным и симметричным, ОДПФ эквивалентен дискретному косинусному преобразованию:

$$o_t(p) = \sum_{q=1}^Q L_t(q) \cdot \cos\left(p \cdot \left(q - \frac{1}{2}\right) \cdot \frac{\pi}{Q}\right), p = 0, \dots, P.$$

В результате получают P коэффициентов. При этом, обычно используется только половина значений мел-частотных кепстральных коэффициентов (MFCC), хотя P может быть равно Q . Это потому, что кепстр сигнала возбуждения находится «правее» кепстра речевого тракта. Таким образом, в распознаватель поступает p -мерный вектор $o_t(p)$, содержащий кепстральные коэффициенты для t -го фрейма. На рисунке 4 приведены пять примеров MFCC гласного /а/ для фрейма длиной 32 мс.

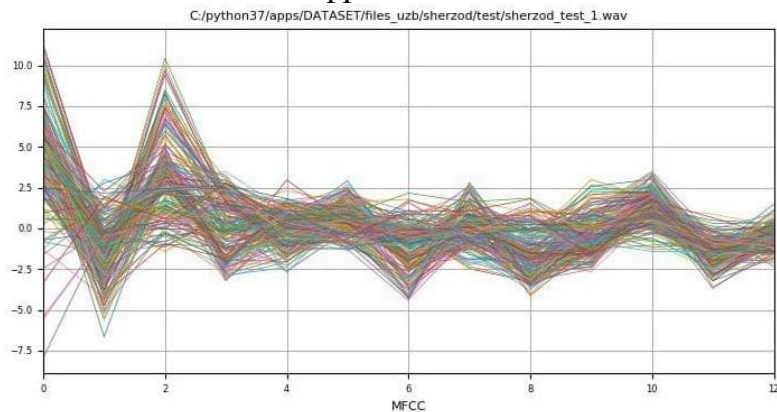


Рис.4. MFCC одного фрейма для пяти примеров гласного /а/

На рисунке 4 показано, что значения MFCC для одного участка произношения близки для разных примеров произношения.

Шаг 7. Конец.

Предлагаемый алгоритм, обозначенный через A2, сокращает время, затрачиваемое на работу системы, за счет уменьшения числа выделенных признаков и состоит из следующих шагов.

Шаг 1. Вводится речевой сигнал. Для этого создается файл в формате *.wav, записанный в 16000 Гц и 16 бит.

Шаг 2. Выделяются признаки из речевого сигнала на основе следующей закономерности

$$s_i = \frac{1}{L} \sum_{j=i \cdot L}^{(i+1) \cdot L} x_j,$$

где x_j – вектор, состоящий из признаков входного речевого сигнала, L – шаг.

Шаг 3. Выделяются признаки, т.е. для каждого s вычисляются MFCC.

Шаг 4. Строится модель человеческой речи и на ее основе осуществляется распознавание речевого сигнала.

Шаг 5. Конец.

Третья глава диссертационной работы **“Разработка алгоритмов предварительной обработки речевых сигналов”** посвящена разработке алгоритмов нормализации речевых сигналов, выделения на них участков с

речью, определения на них полезной части и участков останова, а также шумоподавления.

В параграфе 3.1 приведены алгоритмы нормализации речевых сигналов и выделения полезной части на речевом сигнале.

Для устранения больших изменений громкости обычно используется процедура нормализации сигнала по амплитуде. С ее помощью амплитуда сигнала сводится к интервалу $\left[\frac{\Delta}{2}, -\frac{\Delta}{2}\right]$ (рис. 5). Образец нормализованного сигнала $\hat{s}[n]$ получается по следующей формуле:

$$\hat{s}[n] = \frac{\Delta}{\max_m |s[m]| + 1} \cdot s[n],$$

где Δ – ширина отрезка нормализации, симметричного относительно оси абсцисс (например, на рис.5 $\Delta = 1$).

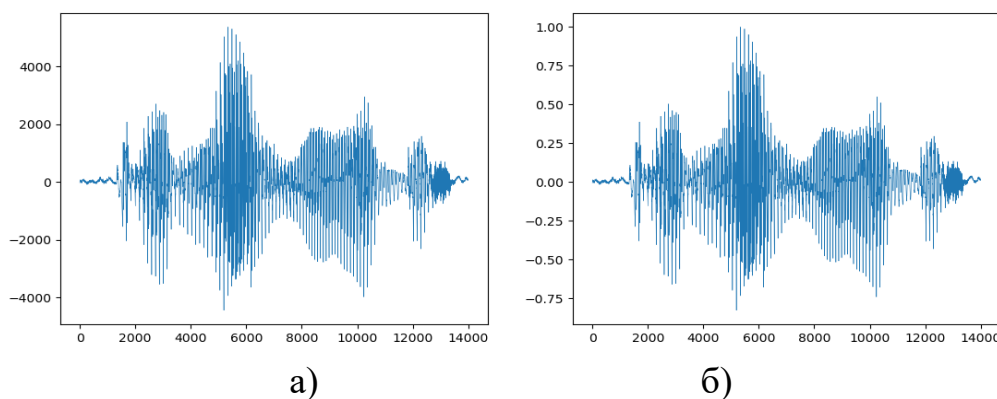


Рис.5. Оцифрованный речевой сигнал до (а) и после (б) нормализации

Для оценки вариативности громкости речи рассматривается набор из Q образцов произношения одного или нескольких слов. Среднее значение громкости $M(q)$ для q -го образца длиной N отсчетов и среднее значение M_Q для Q образцов определяются следующим образом:

$$M(q) = \sum_{n=1}^N |s[n]|, \quad q = 1, \dots, Q;$$

$$M_Q = \frac{1}{Q+1} \sum_{q=1}^Q M(q). \quad (1)$$

Вычисляется относительное отклонение $D(q)$ громкости каждого образца от среднего:

$$D(q) = \left| 1 - \frac{M(q)}{M_Q} \right|. \quad (2)$$

Из (1) и (2) видно, что на результат влияют абсолютное значение отсчетов и их количество в образце. Поэтому необходимо оценить вариативность громкости набора образцов одного класса, длина которых примерно одинакова, и варьирование громкости базы в целом.

Предложен алгоритм выделения участков с речью на исходном сигнале с

использованием функции распределения экстремумов, обозначенный через АЗ. Данный алгоритм состоит из следующих шагов.

Шаг 1. Используется блок нормализации – записанный сигнал нормализуется по амплитуде скользящим окном шириной 10 мс. Данная процедура необходима для исключения вариативности громкости сигнала, зависящую от индивидуальных характеристик устройства. Нормализация выполняется в пределах окна для того, чтобы все участки были выровнены до одинакового значения громкости.

Шаг 2. Используется блок построения огибающей, сглаживающей колебания малой амплитуды речевого сигнала.

Шаг 3. Применяется блок нормализации. Сигнал необходимо повторно нормализовать из-за построения огибающей.

Шаг 4. Используется блок нахождения распределения экстремумов для m -го участка сигнала.

Шаг 5. Применяется решающий логический блок. В нем принимается решение, содержит ли в m -ый участок сигнала речь или шум. В качестве решающей логики была выбрана трехслойная полносвязная нейронная сеть прямого распространения, которую перед применением необходимо обучить. Обучение нейронной сети производится в блоке обучения.

В параграфе 3.2 приведено описание разработанных алгоритмов сегментации речевых сигналов. Для снижения вычислительной сложности алгоритмов выделения затухающей части сигнала на первом этапе формируется множество локальных максимумов M речевого сигнала, которое вначале инициализируется как пустое – $M = \emptyset$, по следующему правилу:

Если $(s(i) > s(i-1) \text{ и } s(i) > s(i+1))$, то $M = M \cup i$, $i = \overline{1, N}$,

где i – число отсчетов речевого сигнала, $s(i)$ – значения отсчетов речевого сигнала. Затем отсчеты затухающей части итеративно удаляются из множества M . Если принять текущий номер отсчета затухающей за $s(n)$, то номер $s(n+1)$ следующего отсчета затухающей определяется следующим образом:

$$\left\{ \begin{array}{l} K = \{m \in M \mid n < m < n + L\}, \\ s(n+1) \in K, \\ \forall k \in K \mid s(n+1) \geq s(k), \\ \{k \in K \mid (s(n+1) = s(k) \wedge (n+1) < k)\} = \emptyset. \end{array} \right.$$

Здесь параметр L используется для обеспечения того, чтобы затухающая часть не превышала расстояние между отсчетами.

Предложен алгоритм удаления затухающих частей речевого сигнала, обозначенный в диссертации через А4. Данный алгоритм состоит из следующих шагов.

Шаг 1. Формируется вектор f на основе речевого файла.

Шаг 2. Присваивается значение ε .

Шаг 3. Вычисляется значение y :

$$y = \begin{cases} 0, & f(i) < \varepsilon, \\ 1, & f(i) \geq \varepsilon. \end{cases}$$

Шаг 4. Осуществляется уменьшение размера вектора f , т. е. если $y=0$, то соответствующий элемент исключается из вектора f .

Шаг 5. Конец.

Например, результат реализации данного алгоритма при $\varepsilon = 50$ приведен на рисунке 6.

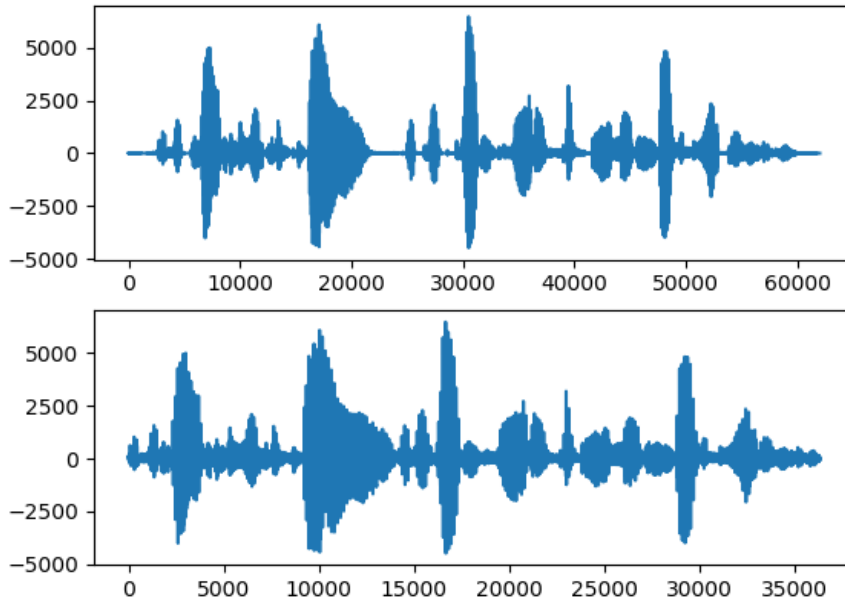


Рис.6. Речевой сигнал до и после удаления затухающих частей

В параграфе 3.3 приводится описание модифицированных алгоритмов шумоподавления в речевом сигнале. Блок-схема алгоритма, обозначенного в диссертации через Аб, приведена на рисунке 7.

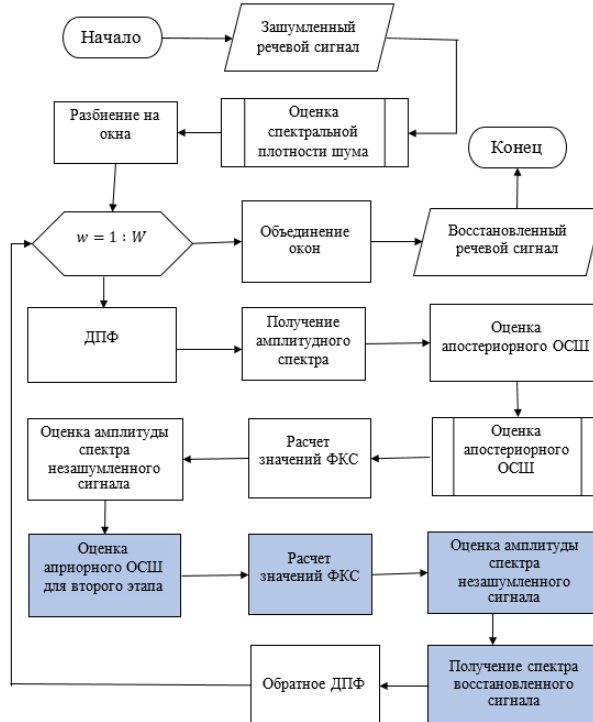


Рис.7. Блок-схема двухэтапного алгоритма шумоподавления

Блок-схема предлагаемого двухэтапного алгоритма шумоподавления в речевом сигнале на основе мягкой маски, обозначенного в диссертации через А8, приведена на рисунке 8.



Рис.8. Блок-схема алгоритма шумоподавления на основе мягких масок

Данный алгоритм может быть использован как первый этап модифицированного двухэтапного алгоритма шумоподавления. В качестве функции коррекции спектра для второго этапа предлагается выбрать функции коррекции спектра минимальной среднеквадратичной ошибки кратковременной амплитуды спектра.

Четвёртая глава диссертации **“Программный комплекс и его практическое применение”** посвящена разработке программного комплекса предварительной обработки речевых сигналов, описанию результатов экспериментальных исследований и решению практических задач.

В параграфе 4.1 дается описание программного комплекса, разработанного на основе предложенных, существующих и усовершенствованных алгоритмов предварительной обработки, выделения признаков речевых сигналов и распознавания речи.

На рисунке 9 приведен вид главного окна программного комплекса, разработанного по модульному принципу.

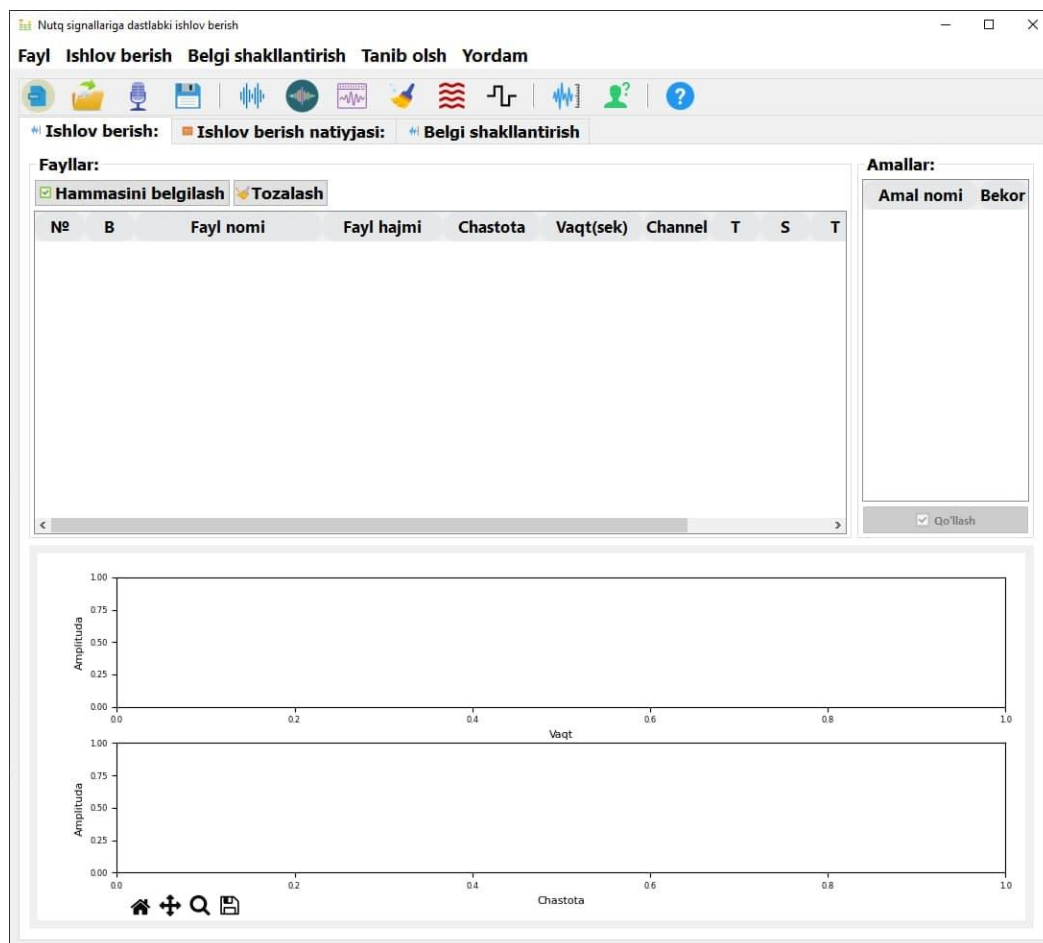


Рис.9. Главное окно программного комплекса

В параграфе 4.2 представлены результаты экспериментального исследования. Для изучения результатов модификации алгоритмов вычисления коэффициентов MFCC и LPC использовалась реализованная на языке Python система распознавания ограниченного числа слов из сигналов, записанных в обычных условиях. Для этого, в связи с необходимостью использования достаточно большой базы речевых фонограмм, была использована база из https://www.dropbox.com/s/87v8jxxu9tvbkns/development_set.zip?dl=0 и 170 различных текстовых фонем с дискретной частотой 16кГц из базы данных узбекского языка.

База речевых сигналов формировалась на основе записей речи 34 дикторов. При этом дикторы групп обучения и контроля не пересекались. Результаты экспериментов представлены в таблице 1.

Таблица 1

Точность идентификации личности по голосу из базы данных модифицированными алгоритмами MFCC + VQ, LPC + VQ и MFCC + GMM

	Основной	2->1	4->1	8->1	16->1
LPC+VQ	73%	82%	81%	79%	72%
MFCC+VQ	84%	89%	85%	78%	73%
MFCC+GMM	86%	92%	87%	80%	84%

Из этой таблицы видно, что результаты распознавания на основе предложенных модифицированных алгоритмов улучшены по сравнению с алгоритмом MFCC. Относительное улучшение распознавания R вычисляется по формуле

$$RI = \frac{RR_A - RR_{MFCC}}{100 - RR_{MFCC}} \times 100\%.$$

Результаты исследования распознавания речевых сигналов, записанных в студии при чтении одного и того же текста 5 дикторами по 3 раза, с помощью модифицированных алгоритмов расчета коэффициентов MFCC + VQ и LPC + VQ представлены в таблице 2.

Таблица 2

Точность идентификации личности по голосу, записанных в студии, модифицированными алгоритмами MFCC+VQ, LPC+VQ и MFCC+GMM

	Основной	2->1	4->1	8->1	16->1
LPC+VQ	87%	100%	93%	90%	86%
MFCC+VQ	85%	85%	83%	90%	89%
MFCC+GMM	98%	99%	99%	97%	96%

В параграфе 4.3 приведены результаты решения практических задач с использованием разработанного программного комплекса.

На основе известных и предложенных в диссертационной работе методов и алгоритмов решаются задачи сглаживания, шумоподавления, нормализация, сегментации речевого сигнала, выделения на нем участков с речью, формирования признакового пространства речевого сигнала и распознавания речи. Это позволило сократить время анализа аудиоданных в среднем на 15% и повысить точность восприятия речи в среднем на 20%.

Полученные результаты были одобрены экспертами в этой области.

В последующих исследованиях было признано целесообразным повышать уровень распознавания за счет преобразования речи в текст и повышения эффективности распознавания личности, а также за счет проведения его в отношении информативных признаков. Планирование работ в области прикладных исследований посредством развития речевых технологий на основе предложенных методов и алгоритмов принесло удобство многим специалистам в этой области и снизило накладные расходы.

ЗАКЛЮЧЕНИЕ

Результаты проведенного диссертационного исследования на тему «Алгоритмы предварительной обработки речевых сигналов» сводятся к следующим основным выводам:

1. Проведен анализ существующих алгоритмов предварительной обработки речевых сигналов, на основе результатов которого выявлены преимущества и недостатки этих алгоритмов. Это позволило сформулировать цель и задачи данной диссертации.

2. Разработаны эффективные в смысле быстродействия и устойчивости к различным помехам алгоритмы сегментации. Эти алгоритмы обеспечивают возможность обработки записанных при различных условиях речевых сигналов.

3. Проведен анализ различных подходов к формированию вектора признаков речевых сигналов в частотном и временном областях, на основе результатов которого выбран наиболее эффективный (в смысле точности распознавания) модифицированный метод вычисления мел-коэффициентов этого сигнала. Применение этого метода позволяет разрабатывать системы распознавания речи с высокой точностью.

4. Предложен алгоритм нормализации речевого сигнала, позволяющий снизить вариативность громкости речи, на основе результатов экспериментальных исследований, проведенных на сформированной базе речевых сигналов, по определению степени влияния на амплитуду сигнала таких факторов как чувствительность звукозаписывающего устройства, громкость голоса диктора, его удаленность от микрофона и т. д. Применение данного алгоритма позволяет снизить влияние громкости речи на точность его распознавания.

5. Предложен метод выделения участков речи на исходном сигнале, основанный на анализе распределения локальных экстремумов этого сигнала. На основе данного метода разработан алгоритм выделения участков речи на исходном сигнале, состоящий из блоков нормализации, решающей логики, нахождения огибающей кривой и распределения локальных экстремумов. Данный алгоритм позволяет получить более информативный речевой сигнал для его распознавания.

6. Предложены алгоритмы, подавляющие шум в частотной области и на основе мягких масок. Эти алгоритмы обеспечивают повышение точности распознавания речи.

7. Создан программный комплекс на основе разработанных, усовершенствованных, а также существующих методов и алгоритмов предварительной обработки речевых сигналов. Данное программное обеспечение может быть использовано при создании программно-аппаратных комплексов идентификации личности человека по голосу и распознавания речи.

**AD HOC SCIENTIFIC COUNCIL AT THE SCIENTIFIC COUNCIL
AWARDING SCIENTIFIC DEGREES DSc13/30.12.2019.T.07.01 AT
TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES**

**SCIENTIFIC AND INNOVATION CENTER OF INFORMATION AND
COMMUNICATION TECHNOLOGIES AT THE TASHKENT
UNIVERSITY OF INFORMATION TECHNOLOGIES**

YULDOSHEV YUSUF SHERALIYEVICH

PRE-PROCESSING ALGORITHMS FOR SPEECH SIGNALS

05.01.03 – Theoretical basis of computer science

**DISSERTATION ABSTRACT
OF THE DOCTOR OF PHILOSOPHY (PhD) ON TECHNICAL SCIENCES**

Tashkent-2021

The theme of doctor of philosophy (PhD) on technical sciences was registered at the Supreme Attestation Commission at the Cabinet of Ministers of the Republic of Uzbekistan under number B2020.2.PhD/T894.

The dissertation has been prepared at Scientific and Innovation Center of Information and Communication Technologies at the Tashkent University of Information Technologies.

The abstract of the dissertation is posted in three languages (Uzbek, Russian, English (resume)) on the website www.tuit.uz and on the website of «Ziyonet» Information and educational portal www.ziyonet.uz.

Scientific adviser: **Babomuradov Ozod Jurayevich**
Doctor of Technical Sciences

Official opponents: **Muxamedieva Dilnoz Tulkunovna**
Doctor of Technical Sciences, Professor

Mirzayev Olimjon Namozovich
of the doctor of philosophy on technical sciences

Leading organization: **National University of Uzbekistan**

The defense will take place “ ____ ” _____ 2021 at _____ on the meeting of Scientific council No. DSc.27.06.2017.T.07.01 at Tashkent University of Information Technologies (Address: 100202, Tashkent city, Amir Temur street, 108. Tel.: (+99871) 238-64-43, fax: (+99871) 238-65-52, e-mail: tuit@tuit.uz).

The dissertation is available at the Information Resource Centre of the Tashkent University of Information Technologies (is registered under No. ____). (Address: 100202, Tashkent city, Amir Temur street, 108. Tel.: (+99871) 238-64-43, fax: (+99871) 238-65-52).

Abstract of dissertation sent out on “ ____ ” _____ 2021 y.
(mailing report No. ____ on “ ____ ” _____ 2021 y.).

R.Kh. Khamdamov
Chairman of the scientific council
awarding scientific degrees,
Doctor of Technical Sciences, Professor

F.M. Nuraliev
Scientific secretary of scientific council
awarding scientific degrees,
Doctor of Technical Sciences, Docent

M.A. Ismailov
Chairman of the academic seminar under the
scientific council awarding scientific degrees,
Doctor of Technical Sciences, Professor

INTRODUCTION (abstract of PhD dissertation)

The aim of the research work is to develop algorithms and software for smoothing and reducing noise, processing frequency domains, normalizing and segmentation of speech signals recorded in various conditions of preliminary processing of speech signals.

The object of research work is digital signals recorded in various conditions, as well as methods and algorithms for processing speech signals.

The scientific novelty of the research work is as follows:

an algorithm for smoothing and reducing noise in a speech signal has been developed, taking into account the spectral density of the noise;

an algorithm for determining the area of speech location based on the patterns of distribution of extrema has been developed;

improved method of processing frequency domains of speech based on determination of noise ratios;

an algorithm for normalization and segmentation was developed, taking into account the factors affecting the quality of the speech signal;

an algorithm for the formation of the feature space for solving recognition problems has been developed, taking into account the discreteness of the speech signal.

Implementation of the research results. Based on the software created on the basis of existing, as well as proposed in the work, methods and algorithms for preprocessing speech signals:

a software package developed on the basis of methods and algorithms for the formation of a space of features for solving problems of smoothing and reducing noise of a speech signal, processing frequency domains of speech and recognition has been introduced into the khakimiyats of the city of Shakhrisabz and Kamashi district of the Kashkadarya region (Certificate of the Ministry for the Development of Information Technologies and Communications of the Republic of Uzbekistan No. 33-8/783 of December 22, 2020). Based on the use of the results of scientific research, the time for preliminary processing of speech signals, smoothing, filtering, cleaning from various noises to ensure high-precision speech expression and analysis of audio data was reduced by 15% and the understanding efficiency increased by an average of 20%;

a software package developed on the basis of methods and algorithms for smoothing and forming the feature space for solving problems of normalization and segmentation, reducing noise of speech signals, recognition was introduced in the Shakhrisabz interdistrict court for civil cases and in the Shakhrisabz district police department of the Kashkadarya region (Certificate of the Ministry for the Development of Information Technologies and Communications of the Republic of Uzbekistan No. 33-8/783 of December 22, 2020). Using the results of scientific research, the audio data analysis time has been reduced by an average of 15%.

Structure and volume of the dissertation. The dissertation consists of an introduction, four chapters, conclusion, the list of used literature and appendix. The volume of the dissertation is 118 pages.

ЭЪЛОН ҚИЛИНГАН ИШЛАР РЎЙХАТИ
СПИСОК ОПУБЛИКОВАННЫХ РАБОТ
LIST OF PUBLISHED WORKS

I бўлим (I часть; I part)

1. Бабомурадов О.Ж., Маматов Н.С., Юлдошев Ю.Ш., Тўракулов О.Х., Отахонова Б.И. Нутқни таниб олиш тизимларини ишлаб чиқиш ёндашувлари // Мухаммад ал – Хоразмий авлодлари. – Тошкент, № 3(5)/2018. – Б. 36-40. (05.00.00; №10)

2. Маматов Н.С., Юлдошев Ю.Ш. Нутқни таниб олиш тизимлари // Журнал «Информатика ва энергетика муаммолари». – Тошкент, 2018. – № 3. Б. 38-43. (05.00.00; №5)

3. Mamatov N.S., Yuldoshev Yu. Sh., Abdullaev Sh.Sh., Abdurazzaqov F.B. Methods and algorithms for the analysis of speech signals // Problems of computational and applied mathematics. – Toshkent, № 1(19) 2019. pp. 90-100. (05.00.00; №23)

4. Маматов Н.С., Юлдошев Ю.Ш., Туракулов О.Х. Нутқни автоматик таниб олиш талаблари ва ёндашувлари // Мухаммад ал – Хоразмий авлодлари. – Тошкент, № 4(10)/2019. – Б. 23-26. (05.00.00; №10)

5. Маматов Н.С., Абдуллаев Ш.Ш., Юлдошев Ю.Ш. Нутқни автоматик таниб олиш муаммолари ва ечимлари // Тошкент ахборот технологиялари университети «ТАТУ хабарлари». – Тошкент, 2019. – № 4. – Б. 105-118. (05.00.00; №31)

6. Ниёзматова Н.А., Самижонов А.Н., Юлдошев Ю. Ш., Алланов Н.М., Исмаилова А.С. Нутқ сигналлари белгиларини иажратиб олиш усуллари // Журнал «Информатика ва энергетика муаммолари». – Тошкент, 2019. – № 1 – Б. 80-90. (05.00.00; №5)

7. Юлдошев Ю.Ш. Нутқ сигналларини нормаллаштириш ва соҳаларини ажратиб олиш алгоритми // Мухаммад ал – Хоразмий авлодлари. – Тошкент, 2020. – № 3(13). – Б. 24-30. (05.00.00; №10)

8. Samijonov A., Mamatov N., Niyozmatova, N., Yuldoshev Yu. and Asraev M. Gradient method for determining non-informative features on the basis of a homogeneous criterion with a positive degree // IOP Conference Series: Materials Science and Engineering. – Russia, 2020.– pp. 1-5. (№3; Scopus; IF=0.6).

9. Mamatov N., Samijonov A., Niyozmatova N., Yuldoshev Yu., Mamadalieva N. Algorithm For Identification Based On Voice // International Journal of Innovative Technology and Exploring, Volume-9 Issue-2. – India, December 2019. – pp. 3431-3434. (№3; Scopus; IF=0.6).

10. Mamatov N., Payazov M., Niyozmatova N., Yuldoshev Yu., Mamadalieva N. Definition of Speech Intelligibility of the Uzbek Language //International Journal of Innovative Technology and Exploring Engineering, Volume-9 Issue-3. – India, January 2020. – pp. 2863-2865. (№3; Scopus; IF=0.6).

II бўлим (II часть; II part)

11. Mamatov N.S., Samijonov A.N., Yuldoshev Yu.Sh. and Rashidov Kh.Sh. Selection the Informative Features on the Basis of Interrelationship of Features // «Techno-Societal 2018», Proceedings of the 2nd International Conference on Advanced Technologies for Societal Applications, Volume 2. – India, 2020. pp. 121–129.

12. Маматов Н.С., Юлдошев Ю.Ш., Тўрақулов О.Х. Нутқни таниб олиш тизимлари // «Инновация-2018», Халқаро илмий-амалий анжуман илмий мақолалар тўплами. – Тошкент, 2018 й. – Б-225.

13. Бабомурадов О.Ж., Маматов Н.С., Юлдошев Ю.Ш., Тўрақулов О.Х. Нутқни таниб олиш тизимларини ишлаб чиқиш ёндашувлари // «Инновация-2018», Халқаро илмий-амалий анжуман илмий мақолалар тўплами. – Тошкент, 2018 й. – Б-229.

14. Ниёзматова Н.А., Нуримов П.Б., Юлдошев Ю.Ш., Абдуллаев Ш.Ш. Автоматическая идентификация диктора по голосу с использованием векторного квантования // Международная научно-практическая конференция «Инновационные идеи, разработки и современные проблемы их применения в производстве а также в обучении». – Андижон, 2019 й. – Б. 69-71.

15. Маматов Н.С., Юлдошев Ю.Ш., Самижонов А.Н. Нутқ сигналларини нормаллаштириш ва соҳаларни ажратиб олиш алгоритмлари // «Амалий математика ва информатсион технологияларнинг долзарб муаммолари» мавзусидаги халқаро анжуман тезлари тўплами. 14-15 ноябр 2019 й. Тошкент, 2019 й. – Б. 250-251.

16. Маматов Н.С., Юлдошев Ю.Ш., Тўрақулов О.Х., Самижонов А.Н. Нутқни автоматик таниб олиш талаблари ва ёндашувлари // «Математик моделлаштириш, алгоритмлаш ва дастурлашнинг долзарб муаммолари» мавзусидаги Республика миқёсидаги илмий-амалий конференция. 17-18 сентябрь 2018. – Тошкент, 2018. – Б. 544-549.

17. Абдуразақов Ф.Б., Юлдошев Ю.Ш., Нуримов П.Б. Нутқ сигналларига рақамли ишлов бериш назарияси ва технологияси // «Иқтисодиётнинг тармоқларини инновацион ривожланишида ахборот-коммуникация технологияларининг аҳамияти» мавзусидаги Республика илмий-техник анжуманининг маърузалари тўплами. 12-15 март 2019 й. – Ташкент, 2019. – Б. 16-18.

18. Маматов Н.С., Юлдошев Ю.Ш., Абдуллаев Ш.Ш., Нуримов П.Б. Нутқ сигналлари асосида шахсни таниб олиш дастури // «Ахборот коммуникация технологиялари ва дастурий таъминот яратишда инновацион ғоялар» мавзусидаги Республика илмий-техник анжуманининг маърузалар тўплами 1-қисм. 16-17 апрел 2019 й. – Самарқанд, 2019. – Б. 373-377.

19. Маматов Н.С., Юлдошев Ю.Ш., Абдуллаев Ш.Ш., Нуримов П.Б. Нутқни автоматик таниб олишда марков модели // Докладов Республиканской научно технической конференции «Современное состояние и перспективы применения информационных технологий в управлении». 5-6 сентября 2019 г. – Самарканд, 2019. С – 305-312.

20. Ниёзматова Н.А., Юлдошев Ю.Ш., Абдуллаев Ш.Ш., Нуримов П.Б. Нутқ сигнали белгиларини ажратиб олиш усуллари // «Ахборот коммуникация технологиялари ва дастурий таъминот яратишда инновацион ғоялар» мавзусидаги Республика илмий-техник анжуманининг маърузалар тўплами 1-қисм. 16-17 апрел 2019 й. – Самарқанд, 2019. – Б. 388-391.

21. Самижонов А.Н., Юлдошев Ю.Ш., Абдураимов Ж.Ф. Предварительная обработка речевых сигналов // «Ўзбекистон радиотехника, электроника ва алоқа илмий-техника жамияти иқтисодиётнинг тармоқларини инновацион ривожланишида ахборот-коммуникация технологияларининг аҳамияти» мавзусидаги Республика илмий-техник анжуманининг маърузалар тўплами 2-қисм. 14-15 март 2019 й. – Тошкент, 2019. – Б. 20-21.

22. Маматов Н.С., Юлдошев Ю.Ш., Самижонов А.Н., Паязов М.М. Нутқ сигналларига дастлабки ишлов бериш дастури // Ўзбекистон Республикаси адлия вазирлиги ҳузуридаги интеллектуал мулк агентлиги, электрон ҳисоблаш машиналари учун яратилган дастурнинг расмий рўйхатдан ўтказилганлиги тўғрисида гувоҳнома DGU 07063, 23.09.2019.

23. Маматов Н.С., Нуримов П.Б., Юлдошев Ю.Ш., Самижонов А.Н., Ниёзматова Н.А. MFCC ва VQ алгоритмларига асосланган шахсни таниб олиш дастури // Ўзбекистон Республикаси адлия вазирлиги ҳузуридаги интеллектуал мулк агентлиги, электрон ҳисоблаш машиналари учун яратилган дастурнинг расмий рўйхатдан ўтказилганлиги тўғрисида гувоҳнома DGU 07064, 23.09.2019.

Автореферат “Ҳисоблаш ва амалий математика муаммолари” илмий журнали таҳририятида таҳрирдан ўтказилди ҳамда ўзбек, рус ва инглиз тилларидаги матнларини мослиги текширилди.

Бичими 60x841/16. Рақамли босма усули. Times гарнитураси.
Шартли босма табағи: 3,5. Адади 100. Буюртма № 90.

Гувоҳнома reestr № 10-3719
“Тошкент кимё технология институти” босмаҳонасида чоп этилган.
Босмаҳона манзили: 100011, Тошкент ш., Навоий кўчаси, 32-уй.