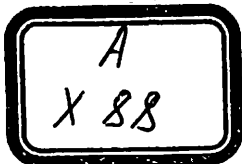


ТОШКЕНТ АХБОРОТ ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ
ҲУЗУРИДАГИ ИЛМИЙ ДАРАЖАЛАР БЕРУВЧИ
DSc.13/30.12.2019.T.07.01 РАҚАМЛИ ИЛМИЙ КЕНГАШ

ТОШКЕНТ АХБОРОТ ТЕХНОЛОГИЯЛАРИ УНИВЕРСИТЕТИ



ХУЖАЯРОВ ИЛЁС ШИРАЛИЕВИЧ

ИНТЕГРАЛЛАШГАН НЕЙРОН ТАРМОҚЛАР АСОСИДА ЎЗБЕК
ТИЛИ НУТҚИНИ ТАНИШ АЛГОРИТМЛАРИ ВА ДАСТУРИЙ
ВОСИТАЛАРИ

05.01.04 – Ҳисоблаш машиналари, мажмуалари ва компьютер
тармоқларининг математик ва дастурий таъминоти

ТЕХНИКА ФАҢЛАРИ БЎЙИЧА ФАЛСАФА ДОКТОРИ (PhD)
ДИССЕРТАЦИЯСИ АВТОРЕФЕРАТИ

Техника фанлари бўйича фалсафа доктори (PhD) диссертацияси мавзуси Ўзбекистон Республикаси Вазирлар Маҳкамаси ҳузуридаги Олий аттестация комиссиясида В2020.4.PhD/T901 рақам билан рўйхатга олинган.

A/2634

сахи
жойи

А
Х 88 | Хужаяров, И. Ш.
Интеграллашган нейрон
тармоқлари асосида уз-
бек тили нутқини таниш
алгоритмлари ва дасту-
рий воситалари : дис.
автореферати

1, 2021

A/2634

**ВОЗВРАТИТЕ КНИГУ НЕ ПОЗЖЕ
обозначенного здесь срока**

қарилган.
оме)) Илмий кенгаш веб
талида (www.ziyounet.uz)

стон Миллий

университети ҳузуридаги
» Феврал да соат 14⁰⁰ даги
мир Темур кўчаси, 108-уй.

рситетининг Ахборот-ресурс
) Манзил: 100202, Тошкент

қатилди.
аси).

Р.Х.Хамдамов
ий даражалар берувчи илмий
нгаш раиси, т.ф.д., профессор

Ф.М.Нуралиев
ий даражалар берувчи илмий
кенгаш илмий котиби, т.ф.д.

М.А.Рахматуллаев
Илмий даражалар берувчи
ий кенгаш ҳузуридаги илмий
минар раиси т.ф.д., профессор

КИРИШ (фалсафа доктори (PhD) диссертацияси аннотацияси)

Диссертация мавзусининг долзарблиги ва зарурати. Жаҳонда нутк сигналларига ишлов бериш, улардан муҳим белгиларни ажратиб олиш ҳамда интеграллашган нейрон тармоқлари асосида автоматик таниб олиш усуллари ривожлантиришга алоҳида аҳамият берилмоқда. Бунга асосий сабаблар сифатида, ҳозирда инсонни қуршаб олган қурилмаларнинг мураккаблашиши, улар билан ўзаро алоқа принципларининг қийинлашишига бўлган эҳтиёж тобора ортиб бормоқда. Инсон учун энг табиийларидан бири бу нутк интерфейсидир. Нуткни таниб олиш тизимлари кўплаб дастурий иловаларининг интерфейсларида, ахборот тизимларида овозли кидирув, телефон сўзлашувларини таниб олиш ва бошқа соҳаларда кенг қўлланилмоқда. Шунинг учун ҳам нуткни автоматик таниб олишнинг алгоритм ва усуллари ишлаб чиқиш ва ривожлантириш муҳим масалалардан бири бўлиб қолмоқда. Бу йўналишда ривожланган хорижий мамлакатларда, шу жумладан, АҚШ, Франция, Ҳиндистон, Россия Федерацияси, Хитой, Япония, Германия, Жанубий Корея, Англия ва бошқа мамлакатларда нуткни таниб олиш усуллари ривожлантиришнинг назарий ҳамда амалий масалаларини ечишга катта эътибор қаратилмоқда.

Жаҳонда нутк сигналларига ишлов бериш, уларни таниб олиш ва таҳлил қилиш учун автоматлаштирилган тизимлар яратишнинг мавжуд усул ва алгоритмларини такомиллаштириш ҳамда янги ҳисоблаш алгоритмларини ишлаб чиқишга йўналтирилган кенг қамровли илмий-тадқиқот ишлари оlib борилмоқда. Бу борада, нутк корпусини шакллантириш, нутк сигналларини характерловчи белгиларни ажратиш усуллари такомиллаштириш ва яратиш ҳамда ушбу характерловчи белгилар асосида таниб олиш алгоритмларини ишлаб чиқиш, улар асосида нуткни автоматик таниб олишнинг дастурий воситаларини яратиш ахборот технологияларини ривожлантиришнинг муҳим вазифаридан бири ҳисобланади.

Республикамизда нутк сигналларига ишлов бериш усуллари ривожлантириш ва ўзбек тили нутқини автоматик таниб олишнинг интеллектуал тизимларини яратиш муаммоларини дастурий воситалар ёрдамида ҳал этиш юзасидан илмий тадқиқотлар олиб борилмоқда. 2017-2021 йилларда Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегиясида, жумладан «...илғор ахборот-коммуникация технологияларини жорий этиш ва улардан фойдаланиш, илмий ва инновацион ютуқларини амалиётга жорий этишнинг самарали механизмларини яратиш...»¹ вазифарлари белгиланган. Мазкур вазифарларни амалга оширишда жумладан, нуткни таниб олиш усулларида фойдаланиб ахборот тизимларида овозли сўровларни амалга ошириш, техник қурилмаларни овозли буйруқлар асосида бошқариш, овозли кидирув, эшитиш ва нутқда муаммоси мавжуд болалар

¹ Ўзбекистон Республикаси Президенти 2017 йил 7 февралдаги ПФ-4947-сон “Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегияси тўғрисида”ги Фармони

реабилитацияси ҳамда уларни ўқитиш муҳим масалалардан бири ҳисобланади. Шунинг учун нутқга ишлов бериш, нутқ корпусини шакллантириш, уларни тавсифловчи характерли белгиларни ажратиш ва таниб олиш масалалари, интеллектуал ишлов бериш усул ва алгоритмларини ривожлантириш ҳамда уларни нутқни автоматик таниб олиш тизимларда қўллаш долзарб муаммо ҳисобланади.

Республикамызда бу соҳани тубдан такомиллаштириш бўйича Ўзбекистон Республикаси Президентининг 2017-йил 7-февралдаги ПФ-4947-сон «Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегияси тўғрисида»ги, 2018 йил 19 февралдаги ПФ-5349-сон «Ахборот технологиялари ва коммуникациялари соҳасини янада такомиллаштириш чора-тадбирлари тўғрисида»ги, 2020 йил 28 апрелдаги ПҚ-4699-сон «Рақамли иқтисодиёт ва электрон ҳукуматни кенг жорий этиш чора тадбирлари тўғрисида»ги Фармонлари, ҳамда мазкур фаолиятга тегишли метёрий-ҳуқуқий ҳужжатларда белгиланган вазифаларни амалга оширишда ушбу диссертация тадқиқоти муайян даражада хизмат қилади.

Тадқиқотнинг республика фан ва технологиялари ривожланишининг устувор йўналишларига мослиги. Мазкур тадқиқот фан ва технологиялари ривожланишининг IV. «Ахборотлаштириш ва ахборот-коммуникация технологияларини ривожлантириш» устувор йўналиши доирасида бажарилган.

Муаммонинг ўрганилганлик даражаси. Жаҳонда нутқ корпусини шакллантириш, нутқ сигналларига интеллектуал ишлов бериш ва уларни автоматик таниб олиш усулларини ишлаб чиқиш ва такомиллаштириш масалалари ҳамда уларнинг амалий қўлланилиши бўйича бир қатор тадқиқотлар олиб борилган жумладан, Rabiner L.R, Graves A, Zhang Y, LeCun Y, Bengio Y, Abdel-Hamid O, Graz`ina Korvel, Patel I, И. С. Кипяткова, Н. М. Марковников, А.А.Карпов, М.Н. Гусев, А.Л.Ронжин, Солонина А.И, О.Мамырбаев, С.Рустамов ва бошқа хорижий олимларнинг бу борадаги илмий ишлари киёсий ўрганиб чиқилди.

Ўзбекистонда сигналларга ишлов бериш ва нутқни таниб олишнинг назарий асосларини ривожлантиришга М.М.Камилов, Т.Ф.Бекмуратов М.М.Мусаев, М.Арипов, Н.А.Игнатъев, Х.Н.Зайнутдинов, Ў.Р.Хамдамов, Н.С.Маматов, Ф.А.Рахматов ва бошқалар ўзларининг катта ҳиссаларини қўшганлар.

Ҳозирги кунда нутқ сигналларига интеллектуал ишлов бериш асосида нутқни автоматик таниб олиш тизимлари жадал суръатлар билан ривожланмоқда. Ушбу йўналиш бўйича ўтказилган тадқиқотлар тахлили шуни кўрсатадики, нутқ корпусини шакллантириш, барқарор, яхши ифодаланган ва юқори информативли нутқни характерловчи белгилар тўпламини ҳосил қилиш ҳамда уларни таниб олишни таъминловчи нейрон тармоқларига асосланган алгоритмларни ишлаб чиқиш етарлича тадқиқ этилмаган.

Диссертация тадқиқотининг бажарилган олий таълим муассасасининг илмий-тадқиқот ишлари режалари билан боғлиқлиги. Диссертация тадқиқоти Тошкент ахборот технологиялари университетининг

илмий-тадқиқот ишлари режасининг №А5-026 «Кўп ядроли ҳисоблаш аппарат платформалар таҳлили ва мультимедиа тизимлари учун оқимли ишлов бериш» (2012-2014), №А-5-015 «Спектрал ўзгартиришлар асосида нутқли сигналларнинг таҳлили ва параллел методлар, алгоритмлар ва дастурлар яратиш» (2014-2016), БЕОА-5-005 «Инсон нутқини таниш тизимлари қурилмаларининг таҳлили ва нутқий сигналларни қайта ишловчи аппарат-дастурий воситалар яратиш» (2016-2018), БҒО-А5-007 «Видеоматбулотларни автоматик таниб олиш масаласини ечишнинг дастурий воситаси, усул ва алгоритмлари ишлаб чиқиш» (2016-2018) ва ЁБВ-Атех-2018-134 «Нутқий сигналларни таниш алгоритмларининг рақамли сигнал процессорларида таҳлили ва дастурий таъминотини яратиш» (2017-2019) мавзуларидаги лойиҳалари доирасида бажарилган.

Тадқиқотнинг мақсади ўзбек тили нутқ сигналлари таҳлили асосида нутқ корпусини шакллантириш, чекланган луғатдаги дикторга боғлиқ бўлмаган нутқни автоматик таниб олишнинг нейрон тармоғига асосланган алгоритмлари ва дастурий воситаларини ишлаб чиқишдан иборат.

Тадқиқотнинг вазибалари:

нутқ сигналларига ишлов берувчи аппарат ва дастурий воситаларни таҳлил қилиш ва имкониятларини ўрганиш;

турли аудио манбалардан нутқ корпусини шакллантириш алгоритмлари ва уларнинг сифатини текшириш имкониятини берувчи қулай интерфейс ишлаб чиқиш;

ўзбек тили нутқ сигналларига ишлов бериш жараёнларида қўлланиладиган моделларни, усулларни таҳлил қилиш ва улар асосида тасвир каби ифодаланган нутқни характерловчи белгилар картасини ҳосил қилиш алгоритмларини ишлаб чиқиш;

чуқур ўқитишга асосланган нейрон тармоқлари ёрдамида тасвир каби ифодаланган нутқни таниб олишнинг алгоритмларини ишлаб чиқиш;

чекланган луғатдаги (1000 та сўзгача) ўзбек тили нутқини автоматик таниб олишнинг интеграллашган нейрон тармоқлар архитектурасини ва таниб олиш алгоритминини яратиш;

нутқни автоматик таниб олиш учун ишлаб чиқилган алгоритмлар асосида яратилган дастурий воситани амалиётда қўллаш юзасидан тавсиялар ишлаб чиқиш.

Тадқиқотнинг объекти сифатида турли шароитларда ёзиб олинган ўзбек нутқ сигналлари ва улардан ҳосил қилинган икки ўлчовли белгилар картаси (спектрограмма), интеграллашган нейрон тармоқларинини ўқитиш жараёни қаралган.

Тадқиқотнинг предметини ўзбек тили нутқини таниб олиш, нутқ корпусини шакллантириш усуллари, алгоритмлари ва услубиятлари ташкил этади.

Тадқиқотнинг усуллари. Диссертация иши тадқиқотларни олиб бориш қуйидаги усуллардан фойдаланилди: эҳтимоллар назарияси; тимсолларни таниш; маълумотларга интеллектуал ишлов бериш; математик статистика;

чизикли алгебра; сигналларга рақамли ишлов бериш; объекта йўналтирилган дастурлаш.

Тадқиқотнинг илмий янгилиги қуйидагилардан иборат:

аудио манбалардан нутқ корпусини шакллантиришни ҳисобга олган ҳолда аудио файл ва унга мос келувчи матни сегментациялаш алгоритмлари ҳамда сегментланган аудиофайл ва унинг матни мослигини текширувчи дастурий интерфейси ишлаб чиқилган;

дискрет косинус алмаштириш ёрдамида нутқ сигналидан «вақт-частота» соҳасидаги белгилар картасини ҳосил қилиш алгоритми ишлаб чиқилган;

график процессор ёрдамида нутқ сигналидан белгилар картасини ҳосил қилишнинг тезкор алгоритми ишлаб чиқилган;

сверткали нейрон тармоқлари ёрдамида тасвир каби ифодаланган нутқни таниб олишнинг алгоритми ишлаб чиқилган;

чекланган луғатдаги ўзбек тили нутқини ҳисобга олган ҳолда автоматик таниб олишнинг интеграллашган нейрон тармоқлари архитектураси яратилган.

Тадқиқотнинг амалий натижаси қуйидагилардан иборат:

турли аудио манбалардан нутқ корпусини шакллантириш алгоритмлари асосида дастур ишлаб чиқилди ва чекланган луғатдаги ўзбек тили нутқини таниш тизими учун қўлланди;

нутқ сигналларини яхши ифодалай оладиган, турли ҳалақитларга барқарор белгилар картаси ҳосил қилиш алгоритми ва «Нутқ сигналларидан икки ўлчовли белгилар картасини шакллантириш» дастури яратилди;

икки ўлчовли белгилар картаси ёрдамида тасвир каби ифодаланган ўзбек тилидаги нутқ буйруқларини таниб олиш алгоритми ишлаб чиқилди;

чекланган луғатдаги ўзбек тили нутқини автоматик таниб олиш дастурлари шаҳарсозлик кадастри геоахборот тизимига тадбиқ қилинди.

Тадқиқот натижаларининг ишончлилиги. Тадқиқот натижаларининг ишончилиги тақдим этилган алгоритмларнинг тўғрилиги, тестлаш натижалари, диссертациядаги таклиф қилинган алгоритм ва усулларнинг апробацияси босма нашрлар ва халқаро илмий конференциялар маърузаларида, тажрибавий тадқиқотлар ва яратилган дастурий воситани амалиётта тадбиқ этиш билан тасдиқланади.

Тадқиқотлар натижаларининг илмий ва амалий аҳамияти. Нутқ сигналларини таҳлил қилиш ва таниб олиш асосида Ўзбекистонда биринчилар қаторида чекланган луғатдаги ўзбек тили нутқини автоматик таниб олиш технологияларининг назарий асосларининг истиқболли ривожланишига ишлаб чиқилган алгоритмларнинг ҳисса қўшиши билан изоҳланади.

Тадқиқот натижаларининг амалий аҳамияти ишлаб чиқилган усул ва алгоритмлар ахборот тизимларда фойдаланувчиларга хизмат кўрсатиш сифати ва автоматлаштириш босқичларини ошириш, call-центрлар операторлари ишини автоматлаштириш, овозли бошқариш дастурлари, «нутқ-матн» тизимлари, лингвистик реабилитация ва тилни ўргатувчи дастурлар, овозли қидирув, имконияти чекланганлар учун ёрдамчи дастурий воситаларда қўлланилиши билан изоҳланади.

Тадқиқот натижаларининг жорий қилиниши. Ишлаб чиқилган алгоритмлар ва дастурий воситалар асосида:

нутқ корпусини шакллантириш, икки ўлчовли белгилар картасини ҳосил қилиш ва нутқни таниб олиш алгоритмлари асосида яратилган дастурий воситалар «ЎзГАШҚЛИТИ» ДУК нинг геоахборот тизимида регионал ҳудудлар номлари, шаҳарсозлик регламенти, шаҳарсозлик ҳужжатларининг объектларини овозли қидирувларни амалга ошириш мақсадида жорий қилинган (Ахборот технологиялари ва коммуникацияларни ривожлантириш вазирлигининг 2020 йил 2 декабрдаги 33-8/7343-сон маълумотномаси). Илмий тадқиқот натижасида геоахборот тизимдан объектларни қидириб топишга боғлиқ жараён 15-20 фоизга тезкор амалга ошириш имконини берган.

свёрткили нейрон тармоқлари ёрдамида тасвир каби ифодаланган нутқни таниб олишнинг алгоритми асосида яратилган дастурий восита «ЎЗБЕКТЕЛЕКОМ» АК Қашқадарё филиалида телефон қақирувларининг бошқарув тизимида (call-марказ) сўзлашувлар ёзувларидаги калит сўзларни қидириш, калит сўзлар бўйича қирувчи кўнғироқларни синфлаштириш мақсадида жорий қилинган (Ахборот технологиялари ва коммуникацияларни ривожлантириш вазирлигининг 2020 йил 2 декабрдаги 33-8/7343-сон маълумотномаси). Илмий тадқиқот натижасида call марказида қақирувларга ишлов бериш операцион харажатлари сезиларли қисқарган ва алоқа каналининг юктамаларини 15 % гача қисқартириш имконини берган.

Тадқиқот натижаларининг апробацияси. Мазкур тадқиқот натижалари 5 та халқро ва 2 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 9 та мақола, жумладан 3 таси хорижий ва 6 таси Республика журналларида нашр қилинган ҳамда 5 та ЭҲМ учун яратилган дастурий воситаларини қайд қилиш гувоҳномалари олинган.

Диссертациянинг ҳажми ва тузилиши. Диссертация таркиби қириш, тўртта боб, умумий хулосалар, фойдаланилган адабиётлар рўйхати ва иловалардан иборат. Диссертациянинг ҳажми 120 бетни ташкил этган.

ДИССЕРТАЦИЯНИНГ АСОСИЙ МАЗМУНИ

Қириш қисмида диссертация мавзусининг долзарблиги ва зарурияти асосланган, тадқиқотнинг Ўзбекистон Республикаси фан ва технологиялар тараққиётининг устувор йўналишларига мослиги кўрсатилган, тадқиқотнинг мақсад ва вазибалари белгилаб олинган ҳамда тадқиқот объекти ва предмети аниқланган, олинган натижаларнинг ишончлилиги асослаб берилган, уларнинг назарий ва амалий аҳамияти, тадқиқот натижаларини амалда жорий қилиш ҳолати, нашр этилган ишлар ва диссертация тузилиши бўйича маълумотлар келтирилган.

Диссертациянинг «Нутқ сигналларини таниб олишнинг замонавий усуллари, моделлари ва алгоритмлар таҳлили» деб номланган биринчи боби, олтита параграфдан иборат бўлиб, унда замонавий нутқни таниш тизимларига қўйиладиган талаблар, уларнинг бошланғич структуралари, нутқни таниш тизимларининг классификацияси, нутқ сигналларига дастлабки ишлов беришнинг мавжуд усуллари изоҳланган. Нутқ сигналларини спектрал тақдим этиш усуллари, нутқни таниш масалаларини ечишнинг акустик моделлаштириш алгоритмларига асосланган ёндашувлар бўйича хорижий давлатлар ва Республикамизда олиб борилаётган илмий ва амалий тадқиқот ишлари атрофлича таҳлил қилинган ва ушбу таҳлиллар натижасида нутқ сигналларини автоматик таниб алгоритмлар кетма-кетлиги танлаб олинган.

Нутқни таниб олиш масаласини ечишда эътиборга олиш керак бўлган қўшлаб омиллар мавжуд буларга, микрофон ва унинг атрофидаги акустик муҳит параметрлари, нутқни узатиш каналининг хусусияти, луғат ҳажми, нутқнинг ўзгарувчанлиги, нутқ атрофидаги ҳалақитлар даражаси, кирувчи нутқнинг типи (ажратилган/узлуксиз). Ушбу омиллар асосида диссертация ишида замонавий нутқни таниш тизимларига қўйиладиган асосий талаблар ва уларнинг структуралари таҳлили олиб борилган. Нутқ сигналларига дастлабки ишлов беришнинг мавжуд усул ва алгоритмлари таҳлили натижасида нутқ чегараларини ажратиш, филтрлаш, спектрал ишлов бериш ҳамда уларнинг ҳисоблаш вақтини қисқартириш ва аниқлигини ошириш мақсадида мавжуд мос алгоритмларни танлаш зарурлиги асосланган.

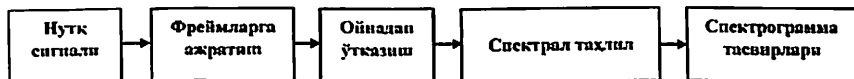
Бундан ташқари нутқ сигналларини спектрал тақдим этиш учун Фурье ва Вейвлет таҳлилларининг усуллари ўрганиб чиқилган. Уларни нутқ сигналларини спектрал тақдим этиш, турли ҳалақитлардан тозлаш, сегментациялаш, ҳалақитларга бардошли белгилар векторини ҳосил қилишдаги афзалликлари ҳамда уларни қайси ҳолларда қўллаш зарурияти келтирилган. Шунингдек, диссертациянинг ушбу бобида нутқни акустик моделлаштириш алгоритмларининг таҳлиллари олиб борилган. Ушбу алгоритмларни ўқитиш ва таниб олиш аниқлиги кўрсаткичлари бўйича афзалликлари ва камчиликлари келтирилган. Бундан ташқари, тезкор ва аниқлик даражаси юқори бўлган акустик моделлаштириш алгоритмларини танлаш ушбу йўналишдаги асосий муаммолардан бири эканлиги қайд этилган.

Олиб борилган таҳлиллар натижасидан келиб чиққан ҳолда ўзбек тили нутқини таниб олиш босқичлари шакллантирилди.

Диссертациянинг «Нутқга акустик ишлов бериш, параметрик тақдим этиш ва интеллектгуал ишлов бериш алгоритмлари» деб номланган иккинчи боби 3 та параграфдан иборат бўлиб, нутқ сигналининг белгилар таснифи, сунъий нейрон тармоқларига асосланган нутқ сигналларини акустик моделлаштириш усуллари ва нутқни таниш масалаларини график процессорларда самарали ишлов беришга бағишланган.

Самарали нутқни таниб олиш тизимларини яратишда энг муҳим вазифалардан бири таҳлил қилинаётган сигналнинг мазмунига етарли даражада адекват ва шу билан бирга дикторларнинг овозларига инвариант нухасини танлашдан иборат. Одатда белгилар тизимида қуйидаги талаблар

қўйилади: информативлик; кичик ҳажм; дикторга боғлиқ бўлмастик; параметрлари нормаллашган бўлиши; оддий метрикадан фойдаланиш имконияти. Сигналларнинг спектрал хусусиятларидан барқарор, яхши ифодаланган ва юқори информативли нутқни тасифловчи белгилар картасини (спектрограмма) ҳосил қилиш алгоритми тақлиф қилинди. Ушбу белгилар картасини ҳосил қилиш жараёни 1-расмда кўрсатилган.



1-расм. Нутқ спектрограммасини ҳосил қилиш босқичлари

Ушбу A1 алгоритм қуйидагича амалга оширилади:

1-қадам. Нутқ сигнали 16 кГц дискретлаш частотасида рўйхатга олинади ва сигнал филтрланади. Дастлабки филтр сигнални вақт кўринишида кириш/чикиш муносабатига асосланган бўлиб тенглама қуйида келтирилган.

$$y(n) = x(n) - \alpha x(n-1),$$

бунда α дастлабки филтр коэффициенти $0,9 < \alpha < 1,0$.

2-қадам. Нутқ сигнали 16 мс фреймларга ажратилади. Фреймлар ўртасидаги силжиш қадами (overlap) 10 мс қилиб ўрнатилади. Нутқ сигналнинг дискретлаш частотаси 16 кГц бўлганлиги сабабли фрейм узунлиги $N=256$, силжитиш узунлиги $M=160$ қийматлардан иборат. Силжитиш қадами фрейм узунлигининг 62.5% ни ташкил этади.

3-қадам. Ажратилган фрейм силлиқлаш ойнасидан ўтказилади, амалиётда вазн ойнаси сифатида кенг тарқалгани Хэмминг ойнасидир.

$$w_n = 0.54 - 0.46 \times \cos\left(2\pi \frac{n}{N-1}\right), \quad n=0, \dots, N-1,$$

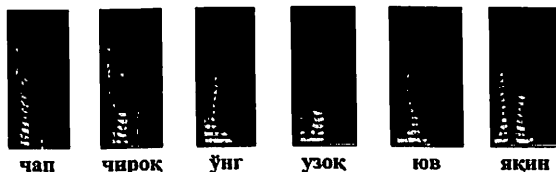
бу ерда N - ойна узунлиги, $N > 1$.

4-қадам. DCT амали бажарилади,

$$L(0) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n), \quad L(k) = \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} x(n) \cos\left(\frac{2n-1}{2N} k\pi\right) \quad k=1, 2, \dots, N-1$$

бунда $n=0, N-1$.

5-қадам. 4-қадамда ҳосил қилинган спектрал коэффицентлар орқали тасвир кўринишдаги спектрограмма ҳосил қилинади. Яъни биз ҳосил қилган ҳар бир фреймдаги коэффицентлар 0 дан 255 ораликқа келтирилади. Бу эса кулранг тасвирни қиймати сифатида қаралади ва қуйидаги 2-расм кўринишга келади.

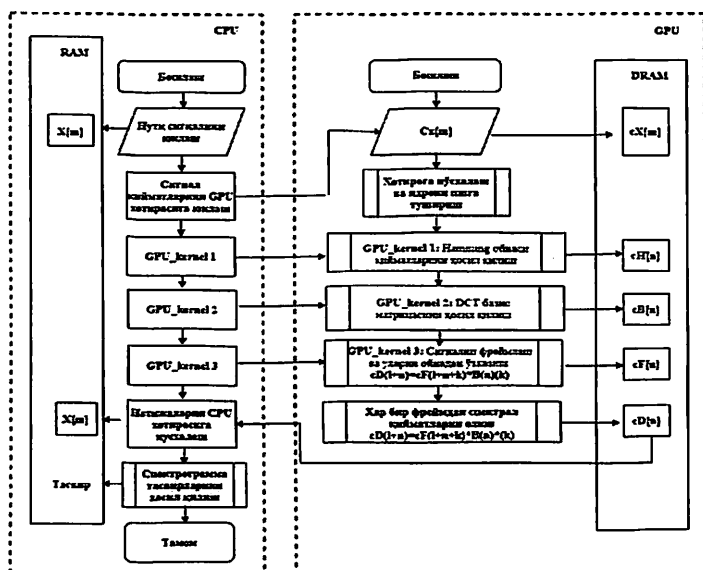


2-расм. Белгилар картасига мисоллар

6-қадам. Тамом.

Нутқ сигналларини сунъий нейрон тармоқларига асосланган акустик моделлаштириш усуллари устида тадқиқотлар олиб борилди. Ҳар бир акустик моделлаштириш усули бўйича ажратилган сўзларни танишни амалга ошириш босқичлари, афзалликлари ва камчиликлари, ҳамда нутқни таниш тизимларида қўллаш юзасидан тақлифлар ишлаб чиқилди.

Нутқни таниш тизимлари катта ҳажмдаги нутқ маълумотларидан фойдаланган ҳолда нейрон тармоқларида ўқитилади. Агарда нутқ сигналдан ҳар 10 мс (битта фрейм) деб ажратиб оладиган бўлсак, унда ҳар 10 соатлик нутқ маълумотлари 3.6 миллион фрейм, ўртача катталиқдаги 100 соатлик нутқ маълумотлари эса 36 миллион фреймларга тенг бўлади. Бу ҳолатда нейрон тармоқни ўқитиш вақти ошади. Ўқитиш тезлигини график процессорнинг (Graphics processing unit - GPU) параллел ишлов бериш имкониятларидан фойдаланиш орқали яхшилаш мумкин. GPU ресурсларидан нутқни таниб олишнинг параметрлаштириш ва тармоқни ўқитиш босқичларида фойдаланиш самаралидир. Диссертацияда нутқ сигналини параметрлаш босқичидаги спектрограмма тасвирларини ҳосил қилишни GPU да параллел алгоритмлар ёрдамида амалга ошириш босқичлари келтирилган (3-расм).



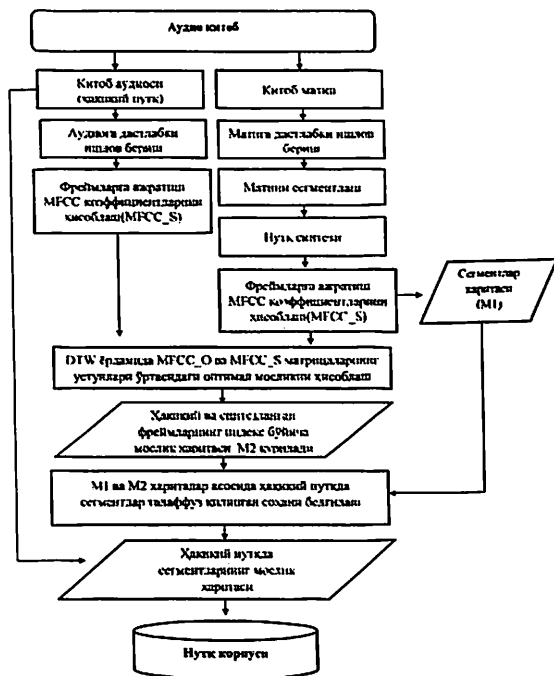
3-расм. GPU да спектрограмма тасвирларини ҳосил қилиш

GPU да сигналларига самарали ишлов бериш учун CPU ва GPU нинг функционал вазифалари аниқлаштирилган, мос GPU архитектура танланган, дастурлаш тамойиллари ҳамда оптимал алгоритмлар келтирилган.

Диссертациянинг «Интеграллашган нейрон тармоқлари ва яратилган нутқ корпуси асосида нутқни автоматик таниб олиш

алгоритмларни ишлаб чиқиш» деб номланган учинчи бобида нутқ корпусини яратиш алгоритмлари, CNN ва рекуррент тармоқларини бирлашган структурасига асосланган интеграллашган нейрон тармоқларида тасвир каби ифодаланган нутқни таниб олишнинг алгоритми, ҳамда улар ёрдамида чекланган луғатдаги сўзларни автоматик таниб олишнинг алгоритмлари келтирилган. Нутқ корпусини ҳосил қилиш учун уларга қўйилган асосий талаблардан келиб чиқган ҳолда нутқ корпусини шакллантириш усули таклиф қилинди. Ушбу усул турли аудио манбалардан нутқ корпуси шакллантиришга асосланган. Бунинг учун аудио манба ва унга мос келувчи матнни синхрон равишда сегментациялаш алгоритмлари ишлаб чиқиш зарурияти туғилди.

Аудио манбадан нутқ корпусини шакллантириш босқичлари ишлаб чиқилди ва бу жараён 4-расмда келтирилган.

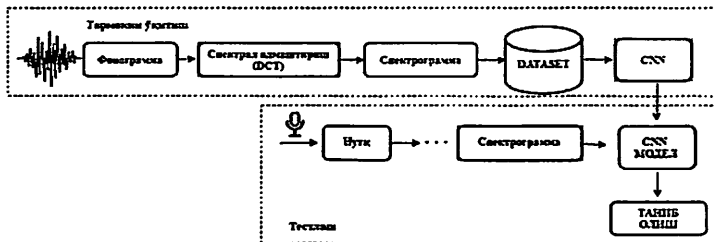


4-расм. Аудио китобдан нутқ корпусини ҳосил қилиш жараёни

Ушбу жараённинг бошланғич қисмида аудио китобнинг матни ва унинг аудиосига дастлабки ишлов берувчи процедуралар бажарилади. Матнга ишлов бериш давомида нутқ матни сегментланади ҳамда уни синтезлаш асосида сегментлар харитаси (M1) ҳосил қилинади. Ушбу харитада матн сегментларининг синтезланган нутқдаги жойлашув оралиқлари келтирилган. Кейинги қадамда аудио китобдаги нутқ сигнали ва синтезланган нутқ сигналларидан MFCC коэффицентлари олинади.

Ушбу аудио ва синтезланган нуткнинг MFCC коэффицентлари ўртасидаги мослик динамик дастурлаш алгоритми (Dynamic time warping - DTW) асосида ҳисоблаб чиқилади. Шундан сўнг аудио ва синтезланган нутк фреймларининг индекс бўйича мослик харитаси (M2) ҳосил қилинади. Ҳосил қилинган M1 ва M2 хариталардан фойдаланиб, аудио файлда нуткда сегментлар талаффуз қилинган соҳа белгиланади ва алоҳида аудио файл ва унга мос келувчи матн транскрипцияларига ажратилади.

CNN да нуткни таниб олишнинг тасвир каби ифодаланган ёндашуви 5-расмда келтирилган.



5-расм. Спектрограмма тасвирларига асосланган таниш жараёни

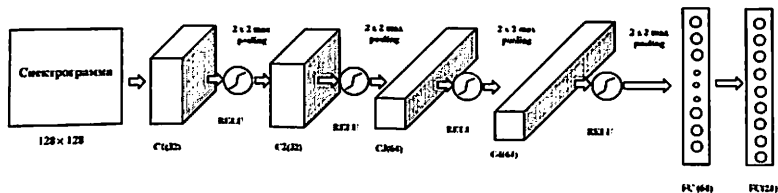
Ушбу ёндашувни амалга ошириш учун A2 алгоритм ишлаб чиқилди ва у қуйидагича амалга оширилади:

1-қadam. Фонограмма босқичи бўлиб, бунда нутқ сигналига дастлабки ишлов бериш процедуралари бажарилади.

2-қadam. Нутқ сигналдан спектрограмма тасвирлари A1 алгоритм ёрдамида ҳосил қилинади.

3-қadam. Ҳар бир нутқ сигналдан олинган 128x128 ўлчамдаги спектрограмма тасвирлари асосида ўргатувчи танланма яратилади.

4-қadam. CNN тармоқ лойиҳалаштирилади. Энг юқори таниб олиш аниқлигини берувчи тармоқ архитектурасининг қатлам ва тартибга солиш (регуляризация) параметрлари танлаб олинади. Тармоқ архитектураси эса қуйидаги 6-расмда келтирилган.



6-расм. Нутқ буйруқларини танишнинг CNN архитектураси

Ушбу архитектуранинг C1 қатламида қадами 1 га тенг бўлган 3x3 свёртка ядроси ишлатилган ва 32 та белгилар картаси ҳосил қилинган. Ундан кейин rectified linear units (ReLU) активация функцияси ва 2x2 ўлчамдаги max pooling амали бажарилади. C2 қатламда ҳам худди C1 қатламда келтирилган параметрдаги амаллар такрорланган. C3 ва C4 қатламларда 3x3 ўлчамдаги

ядро свёртки ишлатилган ва 64 та белгилар картаси ҳосил қилинган. Ҳар бир свёртка қатламларидан кейин ReLU активация функцияси қўлланилган. С4 қатламидан кейин мос равишда 150 та нейронлардан иборат тўлиқ боғланишли (Fully-connected - FC) қатлами ва 24 та нейронлардан иборат Softmax қатлами келади. Меъёридан ортиқ ўрганиш (overfitting) муаммосига учрамаслик учун FC ва Softmax қатламлари орасида 50% нисбатда dropout амали қўлланилган.

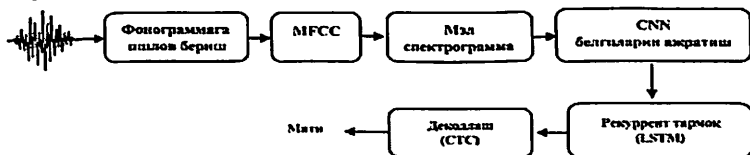
5-қadam. Ўргатувчи танлама танланган архитектурада ўқитилади ва CNN модел ҳосил қилинади.

6-қadam. CNN модел кирувчи янги маълумотлар устида тестланади.

7-қadam. Тамом.

Ушбу ёндашув асосида ўзбек тилидаги 24 та нутқ буйруқларини 94.5% аниқликда таниб олишга эришилди.

Чекланган луғатдаги ўзбек тилидаги сўзларни интеграллашган нейрон тармоқлари асосида таниб олиш алгоритми таклиф қилинган (7-расм).



7-расм. Интеграллашган нейрон тармоқларида нутқни таниш қадамлари

1-қadam. Фонограммага ишлов бериш босқичида сигнални филтрлаш, фреймлаш ва Хэмминг ойнасидан ўтказиш амаллари бажарилади.

2-қadam. Ўргатувчи маълумотлар тўпламидан MFCC коэффицентларини ҳисоблаш амалга оширилади.

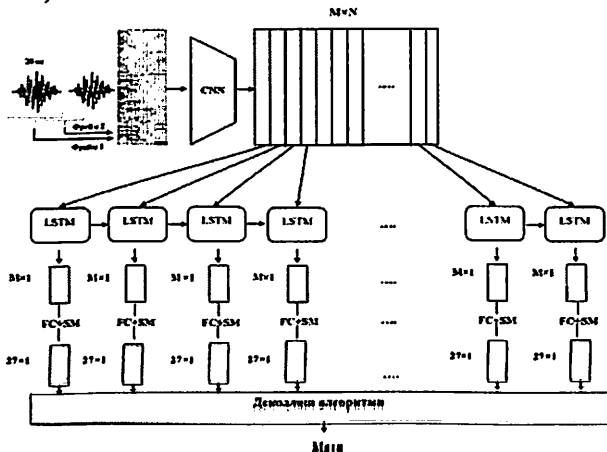
3-қadam. Мэл спектрограммалари ҳар бир фреймдан 40 та MFCC коэффицентларидан фойдаланилган ҳолда яратилади. Бунда фрейм узунлиги 20 мс, силжитиш қадами 10 мс га тенг. Мэл спектрограмма тасвири ўлчами $40 \times N$ га тенг, бунда N – фреймлар сони. Шу тартибда ўргатувчи танламадаги барча аудио файллардан Мэл спектрограмма тасвирлари ҳосил қилинади (8-расм).



8-расм. Ҳосил қилинган MFCC спектрограмма намуналари

4-қadam. Мэл спектрограммалардан локал фазовий белгиларни ажратиш CNN асосида, A2 алгоритмдаги кетма-кетлик бўйича амалга оширилади. Бунда A2 алгоритмдан фарқли равишда тармоқ архитектурасининг параметрлари ва кирувчи тасвир ўлчамлари ўзгаради.

5-қадам. CNN тармоқ $M \times N$ ўлчамда Мэл спектрограммадан олинган белгилар матричаси ҳосил қилади. Бу ерда N - фреймлар сони, M - ҳар бир фреймдаги белгилар сони. Ҳар бир фреймдан $M \times 1$ ўлчамдаги белгилар вектори LSTM тармоқга кирувчи маълумотлар сифатида берилади. LSTM дан чиқувчи белгилар вектори тўлиқ боғланишли (FC) ва softmax қатламидан ўтказилгандан сўнг, ҳар бир вақт қадамидаги N та 27×1 ўлчамдаги алфавит бўйича эҳтимолликлари тақсимланган векторлар декодлаш алгоритмига узатилади (9-расм).



9-расм. Рекуррент тармоқда акустик моделлаштириш босқичлари

б-қадам. Декодлашни жараёни CTC алгоритми асосида амалга оширилади. Бунинг учун ёрликни ифодаловчи барча мумкин бўлган йўллар эҳтимоллиги ҳисобланади:

$$P(\pi | x) = \prod_{t=1}^T u_{\pi_t}, \forall \pi \in L^T$$

бу ерда π_t - бу π йўлидаги t вақт қадами кузатувидаги ёрликдир. L ёрликнинг шартли эҳтимоллигини барча мос йўллар эҳтимолликлари йиғиндиси сифатида баҳолашимиз мумкин. Йўллар B операцияси ёрдамида кетма-кет ёрлик бўйича тасвирланиб, такрорланувчи ёрликлар ва бўш жойлар кетма-кет олиб ташланади.

$$P(l | x) = \sum_{\pi \in B^{-1}(l)} P(\pi | x)$$

Кирувчи кетма-кетликни ёрликлашни эҳтимоллигини ҳисоблаш ва уни моделлаштириш DTW алгоритми асосида ечилади. Бунда белгилашга мос келадиган йўллар йиғиндисини ушбу белгининг префиксларига мос келадиган йўллар бўйлаб итератив йиғиндиларга ажратилади. Сўнгра итерацияларни тўғри ва тескари йўналтирилган рекурсив ўзгарувчилар билан самарали ўқитиш мумкин. Бунинг учун $\alpha_t(s)$ ўзгарувчи киритилади, бунда $\alpha_t(s)$ - t ёрликни башоратлаш учун кетма-кет вақт моментиде қайсидир

нуқтадан бошланувчи ҳамда s – чи жойлашув ҳолатида турган символ перфликсидаги барча кўйи йўлларнинг умумий эҳтимоллиги ҳисобланади.

$$\alpha_t(s) = \sum_{\substack{x \in N^T \\ B(x_t) = s_t}} \prod_{r=1}^t y'_r$$

Тескари йўналтирилган ўзгарувчилар $\beta_t(s)$ ҳам худди $\alpha_t(s)$ каби t вақт momentiда $I_{x,t}$ тўлик эҳтимоллик асосида аниқланади.

$$\beta_t(s) = \sum_{\substack{x \in N^T \\ B(x_t) = I_{x,t}}} \prod_{r=1}^t y'_r$$

Ушбу $\alpha_t(s)$ ва $\beta_t(s)$ ёрдамида исталган t вақт momentiда декодланаётган сўзни ёрлиқлашнинг тўлик эҳтимоллигини куйидагича ифода орқали топилади.

$$p(I|x) = \sum_{s=1}^{|I|} \frac{\alpha_t(s)\beta_t(s)}{y'_t} \quad (1)$$

Юқоридаги (1) ифодани y'_k бўйича дифференциаллаш учун фақат k ёрликдан t вақт momentiда ўтадиган йўллар кўриб чиқилади.

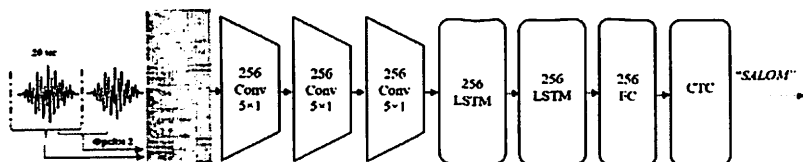
$$\frac{\partial p(I|x)}{\partial y'_k} = -\frac{1}{y'_k{}^2} \sum_{s:seq(s)=k} \alpha_t(s)\beta_t(s)$$

7-қadam. Тамом.

Юқорида келтирилган алгоритмни дастурий амалга ошириш учун тажрибавий тадқиқотлар олиб борилди ва энг самарали деб топилган тармоқ архитектура параметрлари танлаб олинди. Тармоқ архитектурасидаги қатламлар сони ва уларнинг параметрлари 1-жадвалда келтирилган. Таклиф қилинган тармоқ архитектурасининг схематик кўриниши 10-расмда келтирилган.

1-жадвал. Таклиф қилинган тармоқнинг қатламлар параметрлари

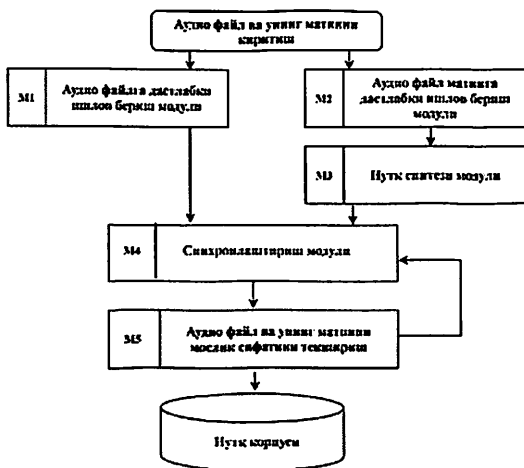
№	Қатламнинг номи	Қатлам параметрлари	Қатламлар сони
1	Кирувчи	40×1	1
2	Свёрткали	Фильтр ўлчами=5×1, сони=256, қадами=1, Dropout=20%, Активация = Relu	3
3	Рекуррент қатлам	RNN тури=LSTM, қатлам ўлчами=256 Batch size=32, оптимизатор=Adam Ўқитиш тезлиги=0.001, Dropout=20% Активация функцияси=Relu	2
4	Тўлик боғланишли қатлам	Қатлам ўлчами = 256, Dropout = 20% Активация функцияси = Softmax	1
5	Транскрипция қатлами	CTC Loss функция	1



10-расм. Таклиф қилинган тармоқ архитектураси

Диссертациянинг «Ишлаб чиқилган алгоритмлар ва дастурий воситалар асосида мажмуа яратиш, уларни тажрибавий тадқиқ қилиш ва амалга ошириш» тўртинчи боби нутқ корпусини шакллантирувчи, чекланган лугатдаги нутқни таниб олиш модуллари ҳамда ушбу модулларини амалга оширувчи дастурий мажмуани баён этиш ва давлат шахарсозлик кадастрига жорий этишга бағишланган.

Нутқ корпуси шакллантириш дастури Python дастурлаш муҳитида амалга оширилган бўлиб, аудио файлга ва матнга ишлов берувчи бир нечта модуллардан ташкил топган. 11-расмда модуллик тамойили асосида ишлаб чиқилган аудио файллардан нутқ корпусини автоматик шакллантириш дастурий воситасининг умумий ишлаш схемаси келтирилган.

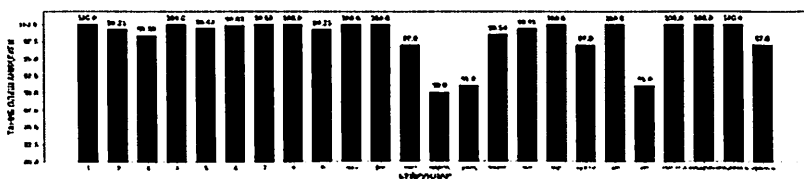


11-расм. Нутқ корпусини шакллантириш дастури ишлаш схемаси

Таниб олиш модулларини яратиш ва тестлаш дастурий воситаси яратилди. Дастурий восита икки қисмдан иборат. Фойдаланувчи қисми - қулай интерфейс яратилган, таниб олинishi керак бўлган аудио файлларни юклаш, учуриш, танлаш, танланган файлни таниб олиш учун мўлжалланган созланмалар амалга оширилади ҳамда нуктни эшитиб кўриш, таниб олиш натижасида ҳосил қилинган матни ўзгартириш(хато билан таниб олганда) имкониятлари мавжуд.

Таниб олиш қисми –ўргатувчи танламадан нутқ белгиларини ажратиш, нутқ буйруқлари ва чекланган лугатдаги ўзбек тили нутқини автоматик таниб олиш учун таклиф қилинган нейрон тармоқ архитектурасини ўқитиш ва таниб олиш жараёнининг барча босқичлари амалга оширади.

CNN да тасвир каби ифодаланган нутқ буйруқларини таниб олиш ва интеграллашган нейрон тармоқлар асосида ўзбек тили нутқини автоматик танишнинг алгоритмлари асосида тажрибавий тадқиқотлар олиб борилди. Тасвир каби ифодаланган нутқни таниб олиш алгоритми асосида ўзбек тилидаги 24 та нутқ буйруқлари 0 да 9 гача санок сўзлар ва 14 та турли буйруқ сўзлари «чап», «ўнг», «чиroy», «ўзою», «якин», «юр», «тухта», «оч», «ён», «пастга», «юкорига», «олдинга», «орқага», «юв» каби жами 24 та буйруқ сўзлар тестланди. Ушбу буйруқ сўзларнинг нутқ намуналари тайёрлашда 30 та эркак ва 30 та аёл киши иштирок этишган. Ҳар бир синфга 190 тадан узунлиги 1 секунд бўлган нутқ намунаси тайёрланган. Тестлаш натижаси 12-расмда келтирилган. Спектрограмма асосида ўқитиш тўпламида синфлаштириш аниқлиги 98.2% , текширувчи тўпلامда 96% ҳамда тестловчи тўпلامда 94.5% эришилди.



12-расм. Тестланган нутқ буйруқларининг таниб олиш аниқлиги

Интеграллашган нейрон тармоқларида ўзбек тили нутқини автоматик таниб олиш алгоритми 300 дан ортиқ Ўзбекистон Республикаси Давлат шаҳарсозлик кадастрининг геоахборот тизимидаги регионал ҳудудлар, шаҳарсозлик регламенти, шаҳарсозлик ҳужжатларининг объектлар номларидан иборат сўзларни таниб олишда қўлланилди. Ўргатувчи танланмани ҳосил қилишда жами 50 та (20 та эркак ва 30 та аёл) киши диктор сифатида иштирок этган. Ўргатувчи танланманинг 80% ўқитишга, 10% тармоқ параметрларини сошлашга ҳамда 10% тестлашга мўлжалланган қисмларга ажратилган. Тестлаш натижасини баҳоланганда, тестлаш жараёнида нотўғри таниб олинган символлар сони бўйича баҳолаш метрикаси (Character Error Rate - CER) $CER=6.15\%$ ни ташкил қилди.

ХУЛОСА

Диссертация иши бўйича олиб борилган тадқиқотлар натижасида куйидаги хулосалар олинди.

1. Нутқ сигналларига дастлабки ишлов беришнинг мавжуд замонавий усуллари, моделлари ва алгоритмлари, нутқни таниб олишнинг интеллектуал

алгоритмларга асосланган ёндашувлар тахлили амалга оширилди ҳамда уларнинг кучли томонлари ҳамда камчиликлари аниқланди.

2. Нутқ сигналларини параметрлаштиришда қўлланиладиган усуллар таснифланди. Нутқ сигналидан белгилар векторини ҳосил қилишда спектрал соҳадаги усуллардан фойдаланиш етарлича таниб олиш самарадорлигини, ҳалақитларга барқарорлиги ҳамда улардаги сонли усулларнинг алгоритм хусусиятлари параллел ҳисоблаш усулларидан самарали фойдаланиш имкониятини мавжудлиги аниқланди. Ушбу афзалликларни инobatта олган ҳолда нутқни таниш учун нутқ сигналидан спектрограмма тасвирларини яратиш алгоритми ишлаб чиқилиб, ундан икки ўлчамли белгилар картаси сифатида фойдаланиш таклиф қилинди.

3. Белгилар картасини ҳосил қилишнинг тезкорлигини ошириш мақсадида замонавий видео карталарда қўлланиладиган ва нутқ сигналларига ишлов бериш самарадорлигини оширувчи график процессор танланди. Ушбу процессорда ечиладиган масалалар синфи ва асосий процессор билан ўзаро функционал вазифалари белгиланди.

4. Турли аудио маълумотлардан нутқ корпусини шакллантириш учун аудио ва унинг матнини синхронлаштириш алгоритмлари таклиф қилинди. Сегментланган аудио ва унинг матни сифатини текширувчи дастурий восита ишлаб чиқилди. Ушбу алгоритмлар асосида ўзбек тилининг дастлабки 50 соатлик миллий нутқ корпуси яратилди.

5. Свёрткали нейрон тармоқларида нутқ сигналларининг спектрограммасини ўқитиш ва спектрограмма тасвирлари асосида нутқни таниб олиш алгоритми таклиф қилинди. Натижада, ўзбек тили нутқ бўйруқларини таниб олишнинг аниқлиги 94,5 % эришилди.

6. Интеграллашган нейрон тармоқлари асосида чекланган лугатдаги (1000 тагача) ўзбек тили сўзларини таниб олишнинг алгоритми ишлаб чиқилди. Ушбу алгоритм нейрон тармоқга кириш ва чиқиш кетма-кетлигини бараварлаштирмасдан сўзлар кетма-кетлигини таниб олишни рекуррент тармоқларда ўрганишга имкон беради. Ушбу таниб олиш алгоритми асосида чекланган лугатдаги сўзларни нотўғри таниб олинган символлар сони бўйича тил моделисиз баҳолаш CER=6.15 % га эришилди.

7. Таклиф қилинган алгоритмлар асосида нутқ корпусини шакллантириш, нутқ бўйруқларини, чекланган лугатдаги ўзбек тили нутқини таниб олиш учун мўлжалланган нейрон тармоқ архитектуралари ва таниб олиш модуллари ҳамда уларнинг дастурий мажмуаси ишлаб чиқилган.

8. Чекланган лугатдаги ўзбек тили нутқини автоматик таниб олиш учун таклиф этилган алгоритмларни синовдан ўтказиш ва амалий масалаларни ечиш учун “ЎЗГАШКЛИТИ” ДУК нинг Давлат шаҳарсозлик кадастри геоахборот тизимида регионал ҳудудлар номлари, шаҳарсозлик регламенти, шаҳарсозлик ҳужжатларининг объектларини овозли кидириб топиш учун амалий фаолиятига жорий этилди.

9. Яратилган дастурий мажмуа “Ўзбектелеком” АК Қашқадарё филиали call марказида тадбиқ этилди. Тадбиқ натижасида қирувчи кўнғироқлар IVR тизим операторлари орасида классификацияланди ва тақсимланди.

**НАУЧНЫЙ СОВЕТ DSc.13/30.12.2019.Т.07.01 ПО ПРИСУЖДЕНИЮ
УЧЕНЫХ СТЕПЕНЕЙ ПРИ ТАШКЕНТСКОМ УНИВЕРСИТЕТЕ
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ**

**ТАШКЕНТСКИЙ УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ**

ХУЖАЯРОВ ИЛЁС ШИРАЛИЕВИЧ

**АЛГОРИТМЫ И ПРОГРАММНЫЕ СРЕДСТВА РАСПОЗНАВАНИЯ
УЗБЕКСКОЙ РЕЧИ НА БАЗЕ ИНТЕГРИРОВАННЫХ НЕЙРОННЫХ
СЕТЕЙ**

05.01.04 – Математическое и программное обеспечение вычислительных машин,
комплексов и компьютерных сетей

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

Ташкент – 2021

Тема диссертации доктора философии (PhD) зарегистрирована в Высшей аттестационной комиссии при Кабинете Министров Республики Узбекистан за В2020.4.PhD/Т901.

Диссертация выполнена в Ташкентском университете информационных технологий. Автореферат диссертации на трех языках (узбекский, русский, английский (резюме)) размещен на веб-странице (www.tuit.uz) и на Информационно-образовательном портале «Ziyouet» (www.ziyouet.uz).

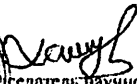
- Научный руководитель:** Мусаев Мухаммаджон Махмудович
доктор технических наук, профессор
- Официальные оппоненты:** Муминов Баходир Болтаевич
доктор технических наук
Давронов Рифкат Рахимович
кандидат технических наук
- Ведущая организация:** Национальный университет Узбекистана им. Мирзо Улугбека

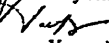
Защита диссертации состоится «04» Февраль 2021 г. в 14⁰⁰ часов на заседании научного совета DSc.13/30.12.2019.T.07.01 при Ташкентском университете информационных технологий. (Адрес: 100202, г. Ташкент, ул. Амира Темура, 108. Тел.: (99871) 238-64-43; факс: (99871) 238-65-52; e-mail: tuit@tuit.uz).


С диссертацией можно ознакомиться в Информационно-ресурсном центре Ташкентского университета информационных технологий (регистрационный номер №2/173). (Адрес: 100202, г. Ташкент, ул. Амира Темура, 108. Тел.: (99871) 238-65-44).

Автореферат диссертации разослан «22» Январь 2021 года.
(протокол рассылки №2 от «22» Январь 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.


М.А.Рахматуллаев
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 особое внимание уделяется обработке речевых сигналов, выделению из них важных информативных признаков и развитию методов применения интегрированных нейронных сетей для автоматического распознавания речи. Основной причиной является многообразие современных устройств, окружающих человека, а также сложность принципов взаимодействия с ними. В этой связи растет потребность в альтернативных, естественных методах управления. Одним из наиболее естественных является речевой интерфейс. Системы распознавания речи широко используются в интерфейсах многих программных приложений, голосового поиска в информационных системах, распознавания телефонных разговоров и других сферах. Поэтому разработка алгоритмов и методов автоматического распознавания речи остается одним из важных вопросов. В этом направлении большое внимание уделяется решению теоретических и практических задач распознавания речи в развитых зарубежных странах, в таких как в США, Франции, Индии, Российской Федерации, Китае, Японии, Германии, Южной Корее, Англии и др.

Во всем мире ведутся обширные исследования по совершенствованию существующих методов и алгоритмов для создания автоматизированных систем обработки, распознавания и анализа речевых сигналов, а также по разработке новых вычислительных алгоритмов. В связи с этим одной из важных задач развития информационных технологий является автоматическое формирование речевого корпуса, совершенствование и разработка методов по выявлению свойств и параметров, характеризующих речевой сигнал, а также алгоритмов распознавания речевых сигналов на основе данных характеристик.

В Республике Узбекистан в настоящее время проводятся научные исследования по созданию методов обработки речевых сигналов и решению задач создания интеллектуальных систем автоматического распознавания узбекской речи. В Стратегии действий по дальнейшему развитию Республики Узбекистан на 2017-2021 годы определены задачи, в том числе «...внедрение передовых информационно-коммуникационных технологий и их использование, разработка эффективных механизмов внедрения научных и инновационных достижений в практику...»². Одним из важных моментов в реализации этих задач является реализация голосовых опросов в информационных системах с использованием методов распознавания речи, управление техническими устройствами с помощью голосовых команд, реабилитация и обучение детей с дефектом речи и слуха. В связи с этим является актуальной разработка методов и алгоритмов интеллектуальной обработки речевых сигналов и их применение в системах автоматического

² Указ Президента Республики Узбекистан №УП-4947 от 7 февраля 2017 г. «О Стратегии действий по дальнейшему развитию Республики Узбекистан»

распознавания, а также формирование речевого корпуса, выделение информативных признаков.

Данное диссертационное исследование в определенной степени служит выполнению задач, изложенных в Указе Президента Республики Узбекистан № УП-4947 от 7 февраля 2017 года «О стратегии действий по дальнейшему развитию Республики Узбекистан» в Постановлениях Президента № ПФ-5349 от 19 февраля 2018 года «О мерах по дальнейшему совершенствованию сферы информационных технологий и коммуникаций», и № ПФ-4699 от 28 апреля 2020 года «О мерах по широкому внедрению цифровой экономики и электронного правительства».

Соответствие исследования приоритетным направлениям развития науки и технологий республики. Данное исследование выполнено в соответствии с приоритетным направлением развития науки и технологий Республики Узбекистан IV. «Развитие информатизации и информационно-коммуникационных технологий».

Степень изученности проблемы. В мире был проведен ряд исследований по формированию речевого корпуса, разработке и совершенствованию методов интеллектуальной обработки речевых сигналов и их автоматическому распознаванию, а также их практическому применению. Научные труды таких учёных, как Rabiner L.R, Graves A, Zhang Y, LeCun Y, Bengio Y, Abdel-Hamid O, Grazina Korvel, Patel I, С. Кипяткова, Н. М. Марковников, А.А. Карпов, М. Гусев, А.Л. Ронжин, Солонина А.И., О.Мамырбаев, С.Рустамов и других, были изучены в рамках исследования.

Среди узбекских ученых значительный вклад в разработку теоретических основ обработки сигналов и распознавания речи внесли Камиллов М.М., Бекмуратов Т.Ф., Мусаев М.М., Арипов М.А., Игнатъев Н.А., Зайнутдинов Х.Н., Хамдамов У.Р., Н.С. Маматов, Ф.А. Рахматов и другие ученые.

В настоящее время стремительно развиваются системы автоматического распознавания речи, основанные на интеллектуальной обработке речевых сигналов. Анализ исследований в этой области показывает, что формирование речевого корпуса, создание набора характеристик, характеризующих устойчивую, четко сформулированную и высокоинформативную речь, разработка алгоритмов на основе нейронных сетей, позволяющих их распознавать, изучены недостаточно.

Связь диссертационного исследования с планами научно-исследовательских работ высшего образовательного учреждения, где выполнена диссертация. Диссертационное исследование было выполнено в рамках научно-исследовательского плана Ташкентского университета информационных технологий по проектам: №А5-026 «Анализ многоядерных вычислительных аппаратных платформ и потоковой передачи мультимедийных систем» (2012-2014 гг.), №А-5-015 «Создание параллельных методов, алгоритмов и программ анализа речевых сигналов на базе спектральных преобразований» (2014-2016), БЕОА-5-005 «Анализ устройств систем распознавания речи человека и разработка аппаратно-программных

средств обработки речевых сигналов» (2016-2018), БЁО-А5-007 «Разработка программного обеспечения, методов и алгоритмов для решения задачи автоматического распознавания видеоданных» (2016-2018) и ЁБВ-Атех-2018-134 «Анализ алгоритмов распознавания речи на цифровых сигнальных процессорах и разработка программного обеспечения» (2017-2019).

Целью исследования является формирование речевого корпуса на основе анализа речевых сигналов узбекского языка, разработка алгоритмов и программных средств дикторонезависимого распознавания речи с ограниченным словарем на основе нейронных сетей.

Задачи исследования:

анализ и изучение возможностей аппаратных и программных средств при обработке речевых сигналов;

разработка алгоритмов формирования речевого корпуса из различных аудио источников и удобного интерфейса, позволяющего проверить их качество;

анализ моделей и методов, применяемых при обработке речевых сигналов узбекского языка, разработка на их основе алгоритмов создания карты признаков речевых сигналов;

разработка алгоритмов распознавания речи, представляемой в виде изображений, с использованием нейронных сетей глубокого обучения;

разработка программных средств распознавания узбекской речи с ограниченным словарем (до 1000 слов);

разработка архитектуры интегрированных нейронных сетей и рекомендаций по практическому применению программных средств, автоматического распознавания речи на основе разработанных алгоритмов.

Объектом исследования являются записанные в различных условиях сигналы узбекской речи, построенная на их основе двумерная карта признаков (спектрограмма), процессы обучения интегрированных нейронных сетей.

Предметом исследования являются методы и алгоритмы распознавания узбекской речи, методика формирования речевого корпуса.

Методы исследования. При проведении диссертационных исследований использовались следующие методы: теория вероятностей; распознавание образов; интеллектуальная обработка данных; математическая статистика; линейная алгебра; цифровая обработка сигналов; объектно-ориентированное программирование.

Научная новизна исследования заключается в следующем:

разработаны алгоритмы сегментации аудиофайла и соответствующего текста с учетом формирования речевого корпуса из аудио источников, а также разработан программный интерфейс, проверяющий совместимость сегментированного аудиофайла и его текста;

на основе дискретно косинусного преобразования разработан алгоритм генерации карты признаков в области «время-частота» речевого сигнала;

разработан ускоренный алгоритм генерации карты признаков речевого сигнала на графическом процессоре;

разработан алгоритм распознавания речи, представляемой в виде изображений, с использованием сверточных нейронных сетей;

разработана архитектура интегрированных нейронных сетей для автоматического распознавания узбекской речи с учетом ограниченного словаря.

Практические результаты исследования заключаются в следующем:

разработана программа на основе алгоритмов формирования речевого корпуса из различных аудио источников для системы распознавания речи узбекского языка с ограниченным словарем;

разработан алгоритм создания двумерной карты признаков, которая представляет речевые сигналы, устойчивые к различным помехам и создана программа «Формирование двумерных карт признаков речевых сигналов»;

разработан алгоритм распознавания представленных в виде изображений речевых команд узбекского языка с использованием двумерной признаковой карты;

введены программы в геоинформационную систему городского кадастра автоматического распознавания узбекской речи с ограниченным словарем.

Достоверность результатов исследования. Достоверность результатов исследования подтверждается корректностью предложенных моделей, результатами тестирования, а также апробацией основных теоретических положений диссертации в печатных трудах и докладах на международных научных конференциях, экспериментальными исследованиями и практическим применением разработанных программных средств.

Научная и практическая значимость результатов исследования.

Научная и практическая значимость исследований подтверждается созданной впервые в Узбекистане технологий автоматического распознавание узбекской речи с ограниченным словарем.

Практическое значение результатов исследований заключается в возможности повышения качества услуг, автоматизации работы операторов call-центров, создании программ речевого управления и систем «речь-текст», в решении задач лингвистической реабилитации и обучения, речевого поиска для людей с ограниченными возможностями.

Внедрение результатов исследования. На основе научных результатов, связанных с решением задачи автоматического распознавания узбекской речи с ограниченным словарем:

разработанные на основе алгоритмов формирования речевого корпуса, создания двумерной карты признаков и распознавания речи программные средства внедрены в ГУП «УзГАШКЛИТИ» (Справка Министерства по информационным технологиям и развитию коммуникаций от 2 декабря 2020 года № 33-8 /7343). В результате исследования в геоинформационной системе Государственного градостроительного кадастра Республики Узбекистан сформирован речевой корпус и разработаны программы, предназначенные для распознавания названий регионов, регламентов градостроительства, объектов

градостроительной документации. По результатам исследований сокращено время поиска геоинформационных объектов на 15-20%;

разработанное на базе сверточных нейронных сетей и алгоритма распознавания речи программные средства внедрены в call-центре Кашкадарынского филиала АК «Узбектелеком» (Справка Министерства по информационным технологиям и развитию коммуникаций от 2 декабря 2020 года № 33-8/7343). В результате внедрения программных средств речевого корпуса и модуля распознавания, сокращены операционные затраты обработки вызовов, а также сокращена нагрузка каналов связи на 15%.

Апробация результатов исследования. Результаты данного исследования были представлены и обсуждены на 5 международных и 2 республиканских научно-практических конференциях.

Публикация результатов исследования. Всего по теме исследования опубликовано 21 научных работ, из них 9 в журнальных изданиях, рекомендованных Высшей аттестационной комиссией Республики Узбекистан, в том числе в 3 иностранных и в 6 республиканских журналах. Получено 5 свидетельств об официальной регистрации программы для ЭВМ.

Структура и объем диссертации. Содержание диссертации состоит из введения, четырех глав, общих выводов, списка использованной литературы и приложений. Объем диссертации 120 страниц.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении обоснована актуальность и востребованность темы диссертации, обосновано соответствие исследования приоритетным направлениям развития науки и технологий Республики Узбекистан, сформулированы цель и задачи, а также определены объект и предмет исследования, научная и практическая значимость результатов исследования, приводится информация о внедрении результатов на практике, об опубликованных работах, а также о структуре диссертации.

Первая глава диссертации под названием «Анализ современных методов, моделей и алгоритмов распознавания речевых сигналов» состоит из шести параграфов, в которых анализируются требования к современным системам распознавания речи, их структуры, классификация систем распознавания речи, существующие методы предварительные обработки речевых сигналов. Подробно проанализированы методы спектрального представления речевых сигналов, подходы к решению задач распознавания речи на основе алгоритмов акустический моделирование, проанализированы научные и практические исследования в зарубежных странах и в нашей стране, а также выделен ряд алгоритмов автоматического распознавания речевых сигналов.

При решении проблемы распознавания речи следует учитывать множество факторов, таких как параметры микрофона и окружающей акустической среды, характеристики канала передачи речи, размер словаря, изменчивость речи, уровень помех вокруг речи, тип речи

(раздельная/непрерывная). На основе этих факторов в диссертации анализируются основные требования к современным системам распознавания речи и их структурам.

Анализ существующих методов и алгоритмов предварительной обработки речевых сигналов основан на необходимости выбора соответствующих алгоритмов с целью разделения границ речевых сигналов, фильтрации, спектральной обработки и сокращения времени их вычисления и повышения точности.

Кроме того, были изучены методы Фурье и Вейвлет-анализа для спектрального представления речевых сигналов. Рассмотрены их преимущества в спектральном представлении речевых сигналов, очистке от различных помех, сегментации, формировании вектора помехоустойчивых признаков и эффективности их использования. Также, в этой главе диссертации проводится анализ алгоритмов акустического моделирования речи. Представлены преимущества и недостатки этих алгоритмов с точки зрения точности обучения и распознавания. Отмечено, что выбор алгоритмов акустического моделирования с высокой скоростью и точностью является одной из основных проблем в этой области.

Исходя из результатов проведенного анализа сформулированы этапы распознавания узбекской речи.

Вторая глава диссертации «Акустическая обработка речи, параметрическое представление и интеллектуальные алгоритмы обработки», состоит из 3 параграфов, посвященных классификации речевых сигналов, методам акустического моделирования на основе искусственных нейронных сетей и эффективной обработке речи на графических процессорах. Одной из важнейших задач при создании эффективных систем распознавания речи является выбор образца, который достаточно адекватен содержанию анализируемого сигнала и в то же время инвариантен к голосам дикторов. Как правило, к вектору признаков предъявляются следующие требования: информативность, малый объем, независимость от диктора, нормализация параметров, применение простых метрик. На основе спектральных свойств речевых сигналов была предложена алгоритм создания карты признаков (спектрограмм), характеризующая стабильную, компактно выраженную и высокоинформативную речь. Процесс создания карты признаков показан на рис. 1.



Рис. 1. Этапы формирования речевых спектрограмм

Данный алгоритм (А1) реализуется следующим образом.

Шаг 1. Речевой сигнал записывается с частотой дискретизации 16 кГц и фильтруется. Фильтр основан на соотношении ввода/вывода во временной области, где уравнение имеет вид:

$$y(n) = x(n) - ax(n-1),$$

где a - коэффициентов фильтра ($0.9 < a < 1.0$).

Шаг 2. Речевой сигнал делится на кадры по 16 мс. Перекрытие между кадрами установлено в 10 мс. Так как частота дискретизации речевого сигнала составляет 16 кГц, длина кадра равна $N = 256$, а величина сдвига $M = 160$. Шаг сдвига составляет 62,5% длины кадра.

Шаг 3. На практике сглаживающим весовым окном является окно Хэмминга.

$$w_n = 0.54 - 0.46 \times \cos\left(2\pi \frac{n}{N-1}\right), n = 0, \dots, N-1,$$

где N - длина окна.

Шаг 4. Выполняется операция DCT:

$$L(0) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n), \quad L(k) = \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} x(n) \cos\left(\frac{2n-1}{2N}k\pi\right) \quad k = 1, 2, \dots, N-1$$

где $n = 0, N-1$.

Шаг 5. Используя спектральные коэффициенты, сгенерированные на шаге 4, формируется спектрограмма, представляемая в виде изображения. Коэффициенты в каждом сгенерированном кадре находятся в диапазоне от 0 до 255. Это рассматривается как значение серого изображения и выглядит следующим образом (рис. 2).

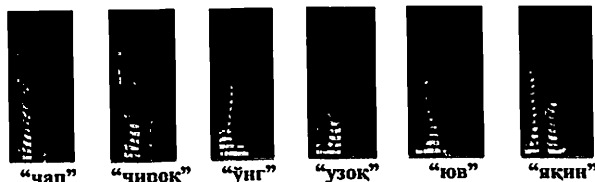


Рис. 2. Примеры карты признаков

Шаг 6. Конец алгоритма.

Проведены исследования методов акустического моделирования речевых сигналов на основе искусственных нейронных сетей. Для каждого метода акустического моделирования разработаны рекомендации по этапам реализации распознавания слов, достоинствам и недостаткам, а также по использованию в системах распознавания речи.

Системы распознавания речи обычно обучаются на нейронной сети с использованием речевого корпуса в большем объеме. Если разделить речевой сигнал по 10 мс (один кадр), то каждые 10 часов речевые данные преобразуются в 3,6 миллиона кадров. При таком расчете объем 100 часов речевых данных составит 36 миллионов кадров. Это увеличивает время обучения.

Скорость обучения предложено повысить за счет использования параллельной обработки на графическом процессоре (Graphics processing unit - GPU). Использование ресурсов GPU в системах распознавания речи эффективно на этапах параметризации и обучения сети. В диссертации

показаны этапы выполнения формирования изображений спектрограммы на этапе параметризации сигнала на графическом процессоре (рис. 3).

Для эффективного формирования изображений спектрограмм и ускорения обработки определены функциональные задачи CPU и GPU, выбрана архитектура GPU, даны оптимальные алгоритмы.

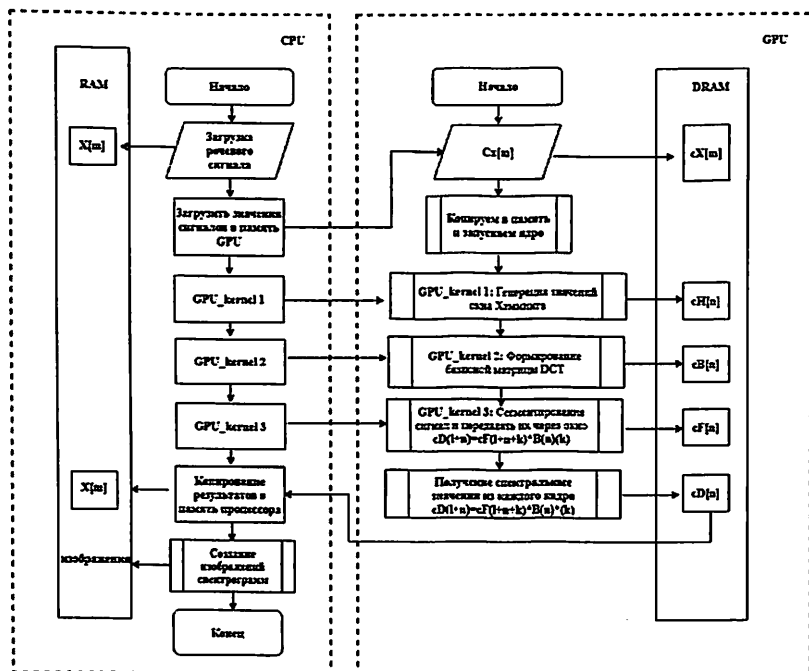


Рис. 3. Формирование изображений спектрограмм на GPU

В третьей главе диссертации «Разработка алгоритмов автоматического распознавания речи на основе интегрированных нейронных сетей и созданного речевого корпуса» приведены алгоритмы создания речевого корпуса, алгоритмы представления спектрограмм в виде изображения для распознавания речи на CNN и алгоритмы автоматического распознавания ограниченного словаря с помощью интегрированных нейронных сетей.

Метод формирования речевого корпуса был выбран с учетом предъявляемых к нему требований. Этот метод основан на формировании речевого корпуса из различных источников звука. Для этого разработаны алгоритмы синхронной сегментации аудио источника и соответствующего текста, этапы формирования речевого корпуса из аудио источника (рис. 4).

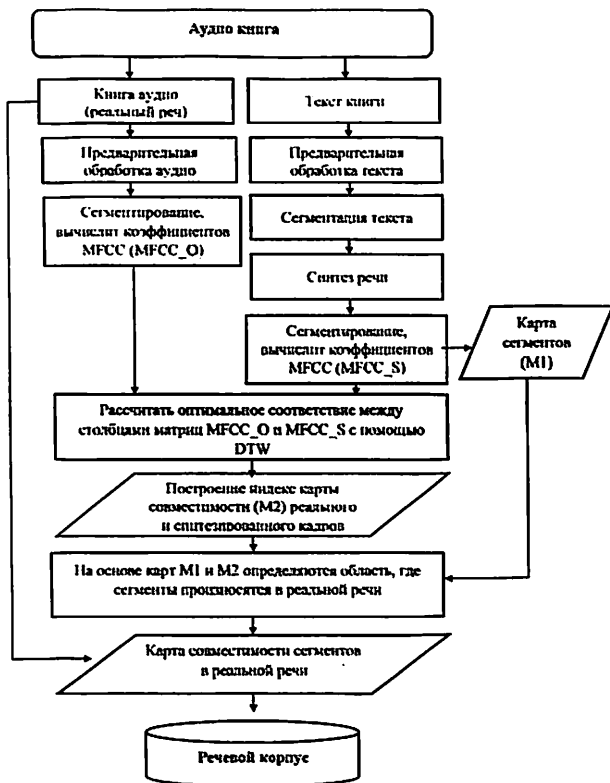


Рис. 4. Процесс создания речевого корпуса из аудиокниги

На начальном этапе этого процесса обрабатывается текст аудиокниги реализуется процедуры предварительной обработки речи. Во время обработки речевой текст сегментируется, и на основе процедуры синтеза речи создается карта сегментов (M1). На этой карте обозначаются интервалы размещения текстовых сегментов в синтезированной речи. Следующим шагом является получение коэффициентов MFCC (Mel-frequency cepstral coefficients) из речевых сигналов аудиокниги и синтезированных речевых сигналов. Соответствие между коэффициентами MFCC аудио и синтезированной речи определяется на основе алгоритма динамического программирования (Dynamic time warping).

После этого формируется индексная карта совместимости (M2) аудио и синтезированных речевых кадров. Используя сгенерированные карты M1 и M2, сегменты аудио речи размечаются в выделенной области и делятся на отдельный аудиофайл и соответствующие текстовые расшифровки.

Подход к распознаванию речи в CNN на основе изображений показан на рис. 5.

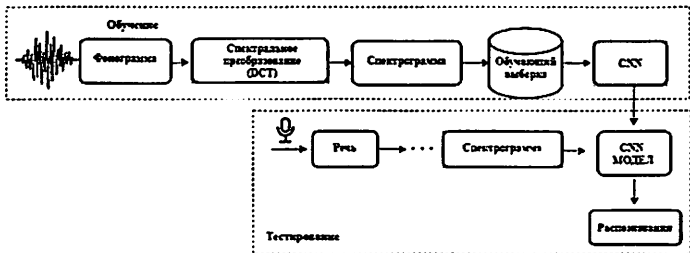


Рис. 5. Процесс распознавания речи на основе изображений спектрограмм

Для реализации этого подхода разработан алгоритм А2, который реализуется следующим образом.

Шаг 1. Это этап формирования фонограммы, над которой выполняются процедуры начальной обработки речевого сигнала.

Шаг 2. Формирование изображения спектрограммы речевого сигнала на основе алгоритма DCT с использованием вышеуказанного алгоритма А1.

Шаг 3. Обучающие выборки создаются на основе изображений спектрограмм 128×128 , взятых из каждого слова речевого сигнала.

Шаг 4. Разрабатывается сеть CNN, выбирается архитектура сети и параметры регуляризации, обеспечивающие максимальную точность распознавания. Сетевая архитектура показана на рис. 6.

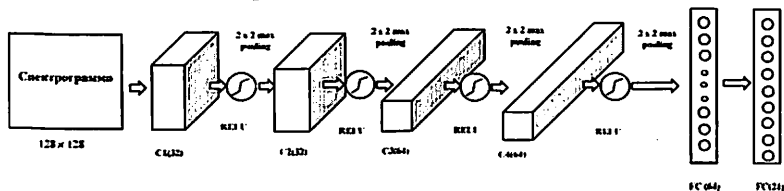


Рис.6. Архитектура CNN для распознавания речевых команд.

На уровне C1 этой архитектуры используется ядро свертки 3×3 с шагом 1 и создаются 32 карты признаков. Затем выполняется функция активации выпрямленных линейных блоков (ReLU) и операция max pooling 2×2 . На уровне C2 повторяются те же шаги, что и на уровне C1. В слоях C3 и C4 используются свертки размером 3×3 и создаются 64 карты признаков. Функция активации ReLU применяется после каждого слоя свертки. За слоем C4 следует слой полносвязный слой (Fully-connected - FC) из 150 нейронов и слой Softmax из 24 нейронов соответственно. Между слоями FC и Softmax используется операция исключения 50% нейронов чтобы избежать проблемы переобучения.

Шаг 5. Выбираются признаки изображений спектрограммы, они обучаются и таким образом формируется модель распознавания CNN.

Шаг 6. Модель CNN тестируется на новых данных.

Шаг 7. Конец работы.

На основе этого подхода 24 речевые команды на узбекском языке были распознаны с точностью 94,5%.

Рассмотрим алгоритм распознавания слов с ограниченным словарем на основе интегрированных нейронных сетей (рис.7).

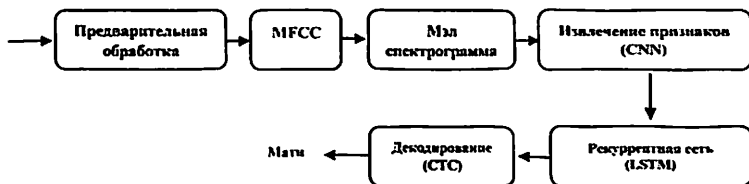


Рис.7. Шаги распознавания речи на основе интегрированных нейронных сетей

Шаг 1. Обработка фонограммы: фильтрация, разделение на фреймы и обработка окном Хэмминга.

Шаг 2. Вычисление коэффициентов MFCC.

Шаг 3. Создание мелспектрограммы с использованием 40 коэффициентов MFCC для каждого фрейма. Длина фрейма 20 мс, шаг сдвига 10 мс. Размер Мэл-спектрограммы $40 \times N$, где N -число фреймов. В том же порядке изображения Мэл-спектрограммы генерируются из всех аудиофайлов в обучающей выборке (рис.8).



Рис.8. Образцы полученных MFCC спектрограмм.

Шаг 4. Выделение пространственных локальных признаков из Мэл-спектрограмм выполняется на основе CNN последовательно алгоритмом A2. В данном случае, в отличие от алгоритма A2, меняются параметры сетевой архитектуры и размер входящего изображения.

Шаг 5. Сеть CNN генерирует матрицу признаков, полученную из Мэл-спектрограммы размером $M \times N$, где N - количество кадров, M - количество признаков в каждом кадре. Вектор символов размером $M \times 1$ из каждого кадра передается в рекуррентную сеть LSTM в качестве входных данных. После того, как вектор признаков из LSTM передается в слой (FC) и в слой softmax. Затем векторы с вероятностями, распределенными по N алфавиту 27×1 , передаются в алгоритм декодирования (рис. 9).

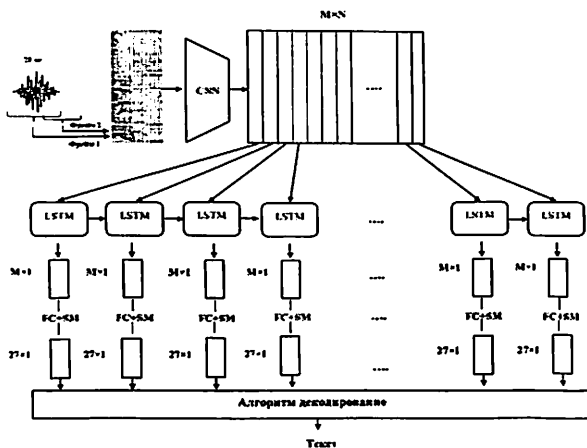


Рис. 9. Этапы акустического моделирования в рекуррентной сети
Шаг 6. Процесс декодирования выполнялся на основе алгоритма STC.
 Для этого вероятность всех возможных путей, представляющих метку, равна:

$$P(\pi | x) = \prod_{i=1}^T y_{\pi_i}^i, \quad \forall \pi \in L^T$$

где π_i - это представляет собой метку в позиции i последовательности. Мы можем оценить условную вероятность метки L как сумму вероятностей всех соответствующих путей. Пути описываются последовательно метками с помощью операции \vee , а повторяющиеся метки и пробелы последовательно удаляются.

$$P(l | x) = \sum_{\pi \in B^{-1}(l)} P(\pi | x)$$

Расчет вероятности меток входящей последовательности и ее моделирование решается на основе алгоритма DTW. Ключевая идея состоит в том, что сумму по путям, соответствующим маркировке, можно разбить на итеративную сумму по путям, соответствующим префиксам этой маркировки. Затем итерации можно эффективно обучать с рекурсивными переменными вперед и назад. Для этого вводится переменная $\alpha_i(s)$, где $\alpha_i(s)$ - суммарная вероятность всех подпутей. Префикс этих подпутей заканчивается символом в s -й позиции в последовательности в момент времени i .

$$\alpha_i(s) = \sum_{\substack{\pi \in N^T: \\ B(\pi_i) = l_s}} \prod_{r=1}^i y_{\pi_r}^r$$

Точно так же обратные переменные $\beta_i(s)$ определяются как полная вероятность $l_{s|i}$ в момент времени i .

$$\beta_i(s) = \sum_{\substack{\pi \in N^T: \\ B(\pi_i) = l_{s|i}}} \prod_{r=i}^T y_{\pi_r}^r$$

С помощью этих $\alpha_t(s)$ и $\beta_t(s)$ определяется полная вероятность метки декодируемого слова в любой момент времени t :

$$p(l|x) = \sum_{s=1}^n \frac{\alpha_t(s)\beta_t(s)}{y_t^l}. \quad (1)$$

Чтобы дифференцировать приведенное выше выражение (1) через y_k^t , рассматриваются только пути, проходящие через k -метку в момент времени t .

$$\frac{\partial p(l|x)}{\partial y_k^t} = -\frac{1}{y_k^t{}^2} \sum_{s:seq(s)=k} \alpha_t(s)\beta_t(s).$$

Шаг 7. Конец работы алгоритма.

Были проведены многочисленные эксперименты по программной реализации описанного выше алгоритма и выбраны наиболее эффективные параметры сетевой архитектуры. Количество уровней в сетевой архитектуре и их параметры приведены в таблице 1.

Таблица 1. Параметры слоя предлагаемой сети

№	Название слоя	Параметры слоев	Количество слоев
1	Входящей	40×1	1
2	Сверточный	Размер фильтра = 5 × 1, количество = 256, шаг = 1, Dropout = 20%, активации = Relu	3
3	Рекуррентной слой	Тип RNN = LSTM, размер слоя = 256, Batch_size=32, оптимизатор=Adam, Скорость обучения = 0,001, Dropout = 20%, активации = Relu	2
4	Полно связанные слои	Размер слоя = 256, Dropout = 20%, Активации = Softmax	1
5	Слой транскрипции	Функция потерь CTC	1

Схематическое изображение предлагаемой сетевой архитектуры показано на рис. 10.

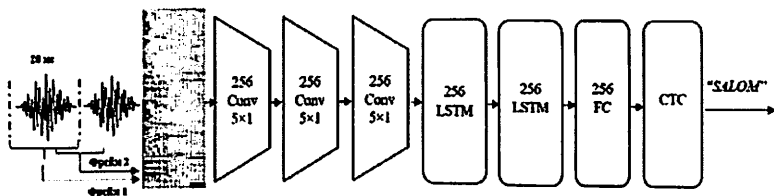


Рис. 10. Предлагаемая сетевая архитектура

В четвертой главе диссертации «Экспериментальное исследование и практическая реализация программного комплекса, разработанного на базе созданных алгоритмов и программ» описаны основанные на созданном речевом корпусе предлагаемые модули распознавания речи с ограниченным словарем и представлены программные интерфейсы, использующие эти модули. Представлены результаты внедрения разработанного программного комплекса на геоинформационной системе градостроительной организации (государственном кадастре).

Программа формирования речевого корпуса реализована в среде программирования Python и состоит из нескольких модулей, обрабатывающих аудиофайлы и текст. На рис.11 представлена общая схема работы программы автоматического формирования речевого корпуса из аудиофайлов, созданных на принципе модульности.

В процессе выполнения работы создан программный инструмент для реализации и тестирования модулей распознавания. Программный инструмент состоит из двух частей.

Пользовательская часть представляет собой удобный интерфейс, который позволяет загружать, удалять и выбирать аудиофайлы. Данный интерфейс позволяет в удобной форме распознавать файлы, настраивать их параметры распознавания и прослушивать речь, а также изменять (в случае нахождения ошибок) текст, сгенерированный в результате распознавания.

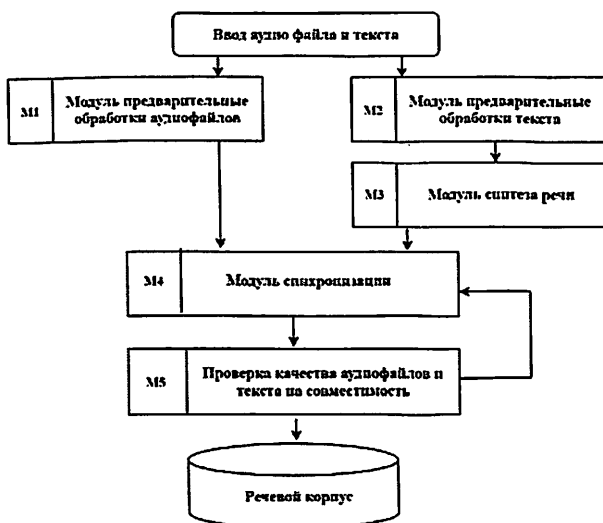


Рис. 11. Схема работы программы формирования речевого корпуса

Часть распознавания. В этой части рассматриваются все этапы процесса обучения и распознавания с помощью предлагаемой архитектуры нейронной сети для распознавания команд и слов узбекской речи.

Экспериментальные исследования проводились в сетях CNN при распознавании команд с помощью спектрограмм в виде изображений. Кроме того, реализовались алгоритмы автоматического распознавания узбекской речи с помощью интегрированных нейронных сетей. На основе алгоритма распознавания речи 24 команды на узбекском языке (пронумерованные от 0 до 9) и 14 различных командных слов «чай», «ўнг», «чирок», «ўзюк», «якин», «юр», «стухта», «оч», «ёп», «пастга», «юкорига», «олдинга», «орқага», «ов» были протестированы. В подготовке речевых образцов этих командных слов участвовали 30 мужчин и 30 женщин.

Для каждого класса было подготовлено 190 речевых образцов длительностью в 1 секунду. Результат теста показан ниже на рис.12.

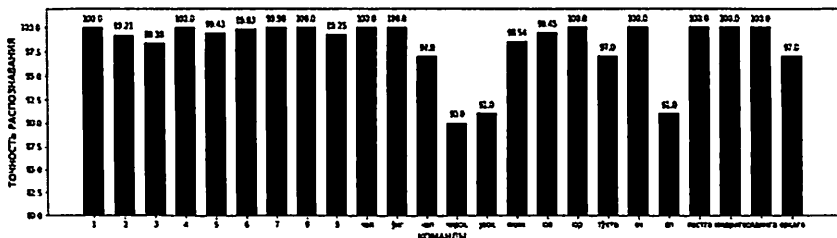


Рис.12. Точность распознавания проверенных речевых команд

Точность классификации в обучающей выборке по спектрограмме составила 98,2%, в контрольной - 96% и в тестовой - 94,5%.

Алгоритм автоматического распознавания узбекской речи в интегрированных нейронных сетях был использован для распознавания более 300 слов в геоинформационной системе Государственного городского кадастра Республики Узбекистан.

В состав распознаваемых слов входили: названия регионов, регламенты градостроительства, названия объектов градостроительной документации. В создании обучающей выборки приняли участие 50 дикторов, в том числе 20 мужчин и 30 женщин. 80% отобранных выборок предназначены для обучения, 10% - для настройки параметров сети во время обучения и 10% - для тестирования.

При оценке результата теста количество символов, неправильно идентифицированных в процессе тестирования, составило CER = 6,15%.

ЗАКЛЮЧЕНИЕ

В результате выполнения диссертации были сделаны следующие выводы.

1. Проведен анализ существующих современных методов, моделей и алгоритмов предварительной обработки речевых сигналов и подходов, основанных на интеллектуальных алгоритмах распознавания речи а также выявлены их сильные и слабые стороны.

2. Классифицированы методы, используемые при параметризации речевых сигналов. Установлено, что методы, используемые в спектральной области при распознавании сигналов, могут быть использованы для формирования признаков. Выявлено, что они обеспечивают высокую эффективность, устойчивость к помехам и параллельную обработку. Учитывая эти преимущества, для распознавания речи был разработан алгоритм формирования спектрограмм в виде изображений и было предложено использовать их в качестве двумерной карты признаков.

3. Для ускорения процесса создания карты признаков применялись графические процессоры, работающие в составе видеокарт компьютера. В работе дано разделение вычислительных функций между основным и графическим процессорами.

4. Предложены алгоритмы синхронизации аудиоданных и соответствующего текста для формирования речевого корпуса из различных типов аудиисточников. Был разработан программный инструмент проверяющий качество сегментированных аудиоданных и соответствующего текста. На основе этих алгоритмов был создан 50-часовой национальный речевой корпус узбекского языка.

5. Предложен алгоритм распознавания речи на основе изображений спектрограмм и обучении спектрограмм речевых сигналов в сверточных нейронных сетях. В результате достигнута точность распознавания речевых команд узбекского языка в 94,5%.

6. Разработан алгоритм распознавания узбекских слов с ограниченным словарем (до 1000) на основе интегрированных нейронных сетей. Такой подход позволяет распознавать слова в рекуррентных сетях без выравнивания входных и выходных последовательностей. На основе этого алгоритма для ограниченного словаря при оценке (без языковой модели) количества неправильно распознанных символов было достигнуто значение CER = 6,15%.

7. На основе предложенных алгоритмов разработаны нейросетевые архитектуры и модули распознавания для формирования речевого корпуса, распознавания речевых команд, распознавания узбекской речи с ограниченным словарем, а также создан программный комплекс для оценки точности распознавания.

8. Для тестирования предложенных алгоритмов автоматического распознавания узбекской речи с ограниченным словарем, прикладные программы распознавания были включены в геоинформационную систему наименования регионов, градостроительные регламенты, объекты градостроительной документации Государственного проектного научно-исследовательского института инженерных изысканий в строительстве, геоинформатики и градостроительного кадастра (ГУП «УЗГАШКЛИТИ»).

9. Разработанный программный комплекс внедрен в call-центре Кашкадарьинского филиала АК «Узбектелеком». В результате внедрения выполнена классификация входящих вызовов и их распределение по операторам IVR-системы.

**SCIENTIFIC COUNCIL AWARDING SCIENTIFIC DEGREES
DSc.13/30.12.2019.T.07.01 AT TASHKENT UNIVERSITY OF
INFORMATION TECHNOLOGIES**

TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES

KHUJAYAROV ILYOS SHIRALIYEVICH

**ALGORITHMS AND SOFTWARE TOOLS FOR RECOGNITION OF
UZBEK SPEECH BASED ON INTEGRATED NEURAL NETWORKS**

05.01.04 – Mathematical and software of computers, complexes and computer networks

**ABSTRACT OF DOCTOR OF PHILOSOPHY (PhD)
DISSERTATION 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 B2020.4.PhiD/T901.

The dissertation has been prepared at 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: **Musaev Muhammadjon Makhmudovich**
Doctor of Technical Sciences, Professor

Official opponents: **Muminov Bakhodir Boltaevich**
Doctor of Technical Sciences

Davronov Rifkat Rahimovich
Candidate of Technical Sciences


Leading organization: **National University of Uzbekistan named after Mirzo Ulugbek**


The defense will take place "04" February 2021 at 14⁰⁰ at the meeting of Scientific council No. DSc.13/30.12.2019.T.07.01 at Tashkent University of Information Technologies (Address: 100202, Tashkent city, Amir Temur street, 108. Ph.: (+99871) 238-64-43, fax: (+99871) 238-65-52, e-mail: tuit@tuit.uz).

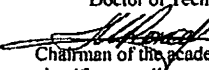
The dissertation can be reviewed at the Information Resource Centre of Tashkent University of Information Technologies (is registered under No. 01/173). (Address: 100202, Tashkent city, Amir Temur street, 108. Ph.: (+99871) 238-64-43, fax: (+99871) 238-65-52).

Abstract of dissertation sent out on "22" January 2021 y.
(Dispatching protocol No. 2 on "22" January 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. Rakhmatullaev
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 the formation of a speech corpus based on the analysis of speech signals of the Uzbek language, the development of algorithms and software for speaker-independent speech recognition with a limited vocabulary based on neural networks.

The object of the research work is the speech signals of the Uzbek language recorded under various conditions and a two-dimensional map of features (spectrogram) built on their basis.

The scientific novelty of the research work is as follows:

algorithms for segmentation of an audio file and the corresponding text have been developed to form a speech corpus from audio sources, and a software interface has been developed that checks the compatibility of a segmented audio file and its text;

an algorithm for generating a feature map in the "time-frequency" domain of the speech signal has been developed;

an accelerated algorithm for generating a map of speech signal features on a graphics processor has been developed;

a speech recognition algorithm has been developed, presented in the form of images, using convolutional neural networks;

an architecture of integrated neural networks for automatic recognition of Uzbek speech with a limited vocabulary is proposed.

Implementation of the research results. Based on scientific results related to solving the problem of automatic recognition of Uzbek speech with a limited vocabulary:

the developed software was implemented in the "State Design Research Institute of Engineering Surveys in Construction, Geoinformatics and Urban Cadastre" (The Ministry of Information Technologies and Communications Development of the Republic of Uzbekistan Reference № 33-8 / 7343 of December 2, 2020). As a result of research in the geoinformation system of the State Urban Cadaster of the Republic of Uzbekistan, a speech corpus was formed and programs were developed for recognizing the names of regions, urban planning regulations, objects of urban planning documentation. This recognition program has been integrated into the geographic information system and tested. During testing, the time for searching for geoinformation objects was reduced by 15-20%;

the developed software tool was implemented in the call-center of the Kashkadarya branch of "Uzbektelecom" JSC (The Ministry of Information Technologies and Communications Development of the Republic of Uzbekistan Reference № 33-8/7343 of December 2, 2020). As a result of the implementation, the process of searching for keywords, the classification of incoming calls and their integration into the voice menu system (IVR) were studied. As a result of the implementation of the speech corpus software and the recognition module, the operational costs of call processing have been reduced, and the load on communication channels has also been reduced by 15%.

Structure and volume of the dissertation. The dissertation consists of an introduction, four chapters, conclusion, references and appendices. The volume of the dissertation is 120 pages.

ЭЪЛОН ҚИЛИНГАН ИШЛАР РЎЙХАТИ
СПИСОК ОПУБЛИКОВАННЫХ РАБОТ
LIST OF PUBLISHED WORKS

I бўлим (I часть; I part)

1. Musaev M.M., Khujayorov I.Sh., Buriboev A.Sh. Accelerate the Solution of Problems of Digital Signal Processing Technology Based INTEL CILK PLUS // Asian Journal of Computer and Information Systems. Vol.9, Issue 02, 2015. PP. 48-51. (№ 5; Global Impact Factor; IF=0.654).

2. Khujayorov I.Sh., Islomov Sh.Z., Nuriev S.A., Ochilov M.M. Parallel Solving Tasks of Digital Image Processin // International Journal on Recent and Innovation Trends in Computing and Communication,-India, Vol.4, Issue 06, 2016. PP. 335 – 339 (№ 5; Global Impact Factor; IF=0.676).

3. Хужаяров И.Ш., Очилов М.М. CUDA технологияси ёрдамида тасвирларни спектрал усулда қайта ишлаш. // ТАТУ Хабарлари. № 1(41), 2017.-Б. 54-63. (05.00.00; №31).

4. Хужаяров И.Ш. Замонавий график процессорлар архитектураси ва технологиялари // ТАТУ Хабарлари. №1(45), 2018. -Б. 22-41. (05.00.00; №31).

5. Хужаяров И.Ш., Очилов М.М. Нутқ товушларини таниш алгоритмини ишлаб чиқиш // ТАТУ Хабарлари. №3(51), 2019. -Б. 18-27. (05.00.00; №31).

6. Мусаев М.М., Хужаяров И.Ш., Очилов М.М. Машинали ўқитиш алгоритмлари асосида ўзбек тили фонемаларини таниб олиш // Журнал Проблемы информатики и энергетики. №6, 2019. -Б. 67-77. (05.00.00; №5)

7. Ilyos Khujayorov., Ochilov Mannon. Parallel Signal Processing Based-On Graphics Processing Units // International Conference on Information Science and Communications Technologies. ICISCT, Tashkent, 2019. (05.00.00;30.09.2019 №269/8-сон раёсат қарори).

8. Хужаяров И.Ш., Очилов М.М., Ҳамзаев Ж.Ф. Автоматик нутқни таниш тизимлари учун нутқ корпусини ҳосил қилиш // ТАТУ хабарлари. №1(53), 2020.-Б. 14-25. (05.00.00; №31).

9. Хужаяров И.Ш., Очилов М.М. Нейрон тармоқларига асосланган нутқ сигналларини акустик моделлаштириш усуллари таҳлили // ТАТУ хабарлари. №2(54), 2020.-Б. 2-15. (05.00.00; №31).

II бўлим (II часть; II part)

10.Хужаяров И.Ш., Берданов У.А. Параллельное решение задач цифровой обработки сигналов. Международная конференция «Актуальный проблемы развития инфокоммуникаций и информационного общества» сборник докладов, Ташкент, 2012. -Б. 173-176.

11.Мусаев М.М., Хужаяров И.Ш. График процессор архитектуралари ва уларда дастур бажарилишининг ўзига хос хусусиятлари. "Иқтисодийнинг реал тармоқларини инновацион ривожланишида ахборот коммуникация

технологияларининг ахамияти". Халқаро илмий-амалий ва маънавий-маърифий анжумани. Тошкент-2018. -Б. 280-284.

12.Хужаяров И.Ш., Очилов М.М. Нутқни қайта ишлаш масалаларини график процессорда амалга ошириш. Халқаро илмий-амалий ва маънавий-маърифий анжумани. Тошкент-2018. Б.139-142.

13. Khujayorov I.Sh. Speech classification using machine learning algorithms. «Modern problems of applied mathematics and information technologies- al- Khorezmiy 2018». Tashkent-2018. -P. 46-47.

14. Muhammadjon Musaev., Ilyos Khujayorov., Mannon Ochilov. 2019. Image Approach to Speech Recognition on CNN. In Proceedings of the 2019 3rd International Symposium on Computer Science and Intelligent Control (ISCSIC 2019). Association for Computing Machinery. USA, 2019, Article 57.-PP. 1–6.

15. Musaev M., Khujayorov I., Ochilov M. The use of neural networks to improve the recognition accuracy of explosive and unvoiced phonemes in Uzbek language. 2020 Information Communication Technologies Conference (ICTC). China, 2020. -PP. 231-234.

16.Berdanov U.A., Khujayorov I.Sh., Abdurashidova K.T., Salimova Kh.R.,Musadjanova D.A. Using Artificial Intelligence Algorithms for Speech Therapy Systems // International Journal of Innovative Technology and Exploring Engineering. India.Vol. 9, Issue 12, 2020. PP. 821-825. (№3; Scopus; IF=0.6).

17.Хужаяров И.Ш., Бойназаров И.М., Худояров Ш.Ш., Раҳманов Д.К., Хужаяров Т.Ш. Программное обеспечение для сжатия и фильтрации сигналов. // Свидетельство об официальной регистрации программы для вычислительных машин. № DGU 03556. 24.12.2015.

18.Хужаяров И.Ш., Худояров Ш.Ш., Очилов М.М., Хужаяров Т.Ш. Программа для фильтрации модельных сигналов во временной области. // Свидетельство об официальной регистрации программы для вычислительных машин. № DGU 03555. 24.12.2015.

19.Раҳмонов Х.Э, Хужаяров И.Ш. Сегментация и подсчет динамических объектов по видеопоследовательности, используя библиотеку OpenCV. // Свидетельство об официальной регистрации программы для вычислительных машин. № DGU 04769. 24.08.2017.

20. Хужаяров И.Ш, Бойназаров И.М, Очилов М.М. Тасвирларга спектрал ишлов беришни параллеллаштириш дастури. // Свидетельство об официальной регистрации программы для вычислительных машин. № DGU 04766. 24.08.2017.

21.Хужаяров И.Ш, Очилов М.М. Нутқ сигналларидан икки ўлчовли белгилар картасини шакллантириш дастури. // Свидетельство об официальной регистрации программы для вычислительных машин. № DGU 09381. 17.11.2020.

**Автореферат «ТАТУ Хабарлари» илмий-техника ва ахборот-таҳлилий
журнали таҳририяти таҳриридан ўтказилди ва ўзбек, рус тилларидаги
матнларини мослиги текширилди.**

**Бичими: 84x60 1/16. «Times New Roman» гарнитура рақамли босма усулида
босилди.**

Шартли босма табағи: 3. Адади 100. Буюртма №18.

« _____ » босмаҳонасида чоп этилди.
