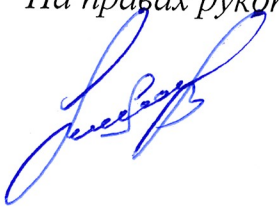
МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РЕСПУБЛИКИ ТАДЖИКИСТАН
ТАДЖИКСКИЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
ИМЕНИ АКАДЕМИКА М.С. ОСИМИ

УДК:004.934.2

На правах рукописи



АШУРЗОДА Бахром Хайриддин

МЕТОДЫ И МОДЕЛИ ПОИСКА КЛЮЧЕВЫХ СЛОВ В РЕЧИ НА ТАДЖИКСКОМ ЯЗЫКЕ (СПЕКТРАЛЬНЫЙ АНАЛИЗ – ОСОБЕННОСТИ)

на соискание ученой степени кандидата технических наук
по специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей»

Научный руководитель: Худойбердиев Хуршед Атохонович
кандидат физико-математических наук, доцент

Душанбе - 2022

СОДЕРЖАНИЕ

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ 4
ВВЕДЕНИЕ 5
ГЛАВА 1. ОБЗОР МЕТОДОВ И МОДЕЛЕЙ ПОИСКА КЛЮЧЕВЫХ СЛОВ В РЕЧИ 12
1.1. Обзор моделей представления речевых сигналов 12
1.2. Анализ методов поиска фрагментов в слитной речи 22
1.3. Обзор существующих программных систем поиска ключевых слов в речи 27
1.4. Постановка задачи 33
1.5. Выводы по главе 36
ГЛАВА 2. РАЗРАБОТКА МОДЕЛЕЙ И МЕТОДОВ ОБРАБОТКИ РЕЧЕВЫХ СИГНАЛОВ НА ТАДЖИКСКОМ ЯЗЫКЕ 40
2.1. Особенности обработки речевых сигналов на таджикском языке 40
2.2. Разработка структуры и состава системы поиска ключевых слов 45
2.3. Модель представления речевого сигнала в системе 51
2.4. Разработка метода поиска ключевых слов 57
2.5. Выводы по главе 66
ГЛАВА 3. РАЗРАБОТКА И ИССЛЕДОВАНИЕ АЛГОРИТМОВ РЕАЛИЗАЦИИ МЕТОДА 68
3.1. Разработка способа расчёта параметров вероятностной графической модели (скрытая Марковская модель, условные случайные поля) 68
3.2. Разработка алгоритма обучения модели 76
3.3. Разработка алгоритма вывода на модели 81
3.4. Исследование алгоритмов 82
3.5. Выводы по главе 85
ГЛАВА 4. РАЗРАБОТКА КОМПЛЕКСА ПРОГРАММ ПОИСКА КЛЮЧЕВЫХ СЛОВ В РЕЧИ 88
4.1. Архитектура программного комплекса 88

4.2. Проектирование вычислительных модулей 98
4.3. Экспериментальное исследование программной системы 106
4.4. Выводы по главе 110
ЗАКЛЮЧЕНИЕ 113
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 115
ПРИЛОЖЕНИЯ 129

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

ANN – Artificial Neural Network

ТЯ – Таджикский язык

ГК – глагольная комбинатсия

СММ – скрытая Марковская модель

HMM – Hidden Markov Model

DTW – Dynamic Time Warping

ООП – объектно-ориентированное программирование

ЧОТ – частота основного тона

MFCC – Mel-frequency cepstrum coefficients

ЭВМ – Электронно-вычислительная машина

HCRF – Hunter Children's Research Foundation

ГМС – Гауссовской модели смеси

ТСС – трифонной модели со связанными состояниями

TF-IDF – Term frequency–inverse document frequency

СГК – специальные глагольные конструкций

LPC – Linear prediction coding

ВВЕДЕНИЕ

Актуальность темы. Настоящее исследование посвящено анализу и разработке моделей и методов распознавания речи на таджикском языке. Актуальность исследования заключается в необходимости разработки новой модели распознавания речи, направленной на практическую точность систем распознавания слитной речи.

Вопросы взаимодействия человека и машины являются одним из самых важных при создании новых компьютеров. Многие из наиболее эффективных средств взаимодействия человека с машиной были бы для него естественными: через визуальные образы и речь.

В очередном Послании Основателя мира и национального единства — Лидера нации, Президента Республики Таджикистан уважаемого Эмомали Рахмона Парламенту страны 26 декабря 2019 года, указывается, что мы должны усилить контроль за освоением учениками современных знаний, побуждать подростков и молодежь на чтение художественных и научных книг, укреплять их творческие способности. В связи с этим предлагаю с целью большего улучшения налаживания изучения естественных, точных и математических наук, а также развития технического мышления подрастающего поколения 2020-2040 годы объявить: «...двадцатилетием изучения и развития естественных, точных и математических наук».

В этом Послании Основатель мира и национального единства - Лидер нации, Президент Республики Таджикистан, уважаемый Эмомали Рахмон также подчеркнул, что: «Мы должны как свою мать и свою Родину любить наш сладкозвучный и поэтичный таджикский язык и беречь как бесценный жемчуг, как основу нашего бытия».

С этой целью в Республике Таджикистан созданы благоприятные условия, принимаются и внедряются Государственные стратегии и программы.

В свете указаний и рекомендаций главы государства и в соответствии с требованиями современного периода, научно-техническим прогрессом и технологий выбранная тема диссертации обладает особой актуальностью. Распознавание и приведение в соответствие письменной речи к сложным процессам вычислительных машин, создание благоприятных условий для изучения знаний и умений на родном языке в современный период признается важной научной проблемой и актуальным направлением в познавательной сфере.

В связи с вышеизложенными обстоятельствами тема диссертации является актуальной. Стоит отметить, что особые трудности вызывает распознавание речи на языке, который является производным от другого языка или содержит многочисленные его элементы. К примеру, древний таджикский язык, имея свои особенности, под которые можно создать систему распознавания речи, однако современный таджикский язык характеризуется существованием в нем русских заимствований, в связи с чем не представлялось до сегодняшнего дня возможным обозначить распознаватель только по отношению к таджикскому языку, поскольку возникало много ошибок при выделении ключевых слов из слитной речи.

Степень изученности и разработанности темы исследования. Проблемы внедрения системы распознавания речи в информационные устройства с ограниченными вычислительными ресурсами в мобильных телефонах и планшетах. В результате было предложено решение проблемы в виде переноса процесса распознавания ресурсоёмких аккаунтов из маломощных пользовательских устройств на облачные мощные сервера, где и будет происходить их распознавание. Пользовательское голосовые запросы отправляются на эти облачные сервера, а ответы после обработки запроса получают по средствам Интернет-соединения. По этой схеме успешно работают системы Siri от Apple и Google Voice Search от Google. Однако, такая схема реализации диалога требует возможности постоянного доступа к сети Интернет, которую не всегда можно реализовать. Так же возникает необходимость в

создании компактного и надёжного, автономно работающего устройства, способного на месте обеспечить все диалоговые процессы человека с машиной. Проблема создания таких устройств является актуальной задачей не только гражданской, но и военной сферы деятельности. В аспекте решения такой задачи израильским концерном Aerospace Industries был создан робот REX. REX занимался перевозкой боеприпасов, продовольствия и эвакуацией личного состава. При этом робот постоянно следует за человеком, который его ведет, и полностью управляется его голосовыми командами.

Сегодня многие ведущие учёные в известных научных центрах и вузах мира, в частности: В.Н. Сорокин в Институте проблем передачи информации РАН; Ю. И. Журавлев и В. Я. Чучупал в Вычислительном центре РАН; Н.Г. Загоруйко и В.М. Величко в Институте математики РАН и Новосибирском государственном университете; О.Ф. Кривнова в МГУ им. Ломоносова; Ю.Н. Жигулевцев в МГТУ им. Н. Е. Баумана; Р.К. Потапова в Московском государственном университете языкознания; Ю.Н. Прохоров в Московском техническом университете связи и информатики; А.И. Евсеев в Московском энергетическом институте, а также ученные из университета Иллинойса (США), университета Карнеги-Меллона (США), Орегонский институт науки и технологий (США), Петербургский институт информатики и автоматизации Российской академии наук, компании IBM, Philips, Dragon Systems, Cognitive Technologies, Istrasoft, Sacrament и другие, проводят активные исследования в этой области, что свидетельствует о ее актуальности.

Связь исследования с программами (проектами) или научными темами: Данное исследование выполнено в рамках реализации перспективного плана научно-исследовательской работы кафедры технологий программирования и компьютерной техники института технологий и инновационного менеджмента в городе Куляб.

Цель работы. Целью диссертационной работы является разработка модели системы распознавания речи с выделением ключевых слов из слитой речи на таджикском языке.

Идея работы заключается в использовании возможностей искусственного интеллекта и машинного обучения с учетом грамматических правил таджикского языка и на их основе разработки алгоритмов распознавания ключевых слов из речи.

Для достижения данной цели в работе поставлены **следующие задачи**:

- рассмотреть существующие модели систем распознавания речи;
- разработать методы и модель системы распознавания речи на таджикском языке;
- разработать алгоритмы реализации метода.

Объект исследований. Объектом исследования является система распознавания ключевых слов, основанная на скрытых Марковских моделях.

Предмет исследования. Методы, алгоритмы и способы распознавания ключевых слов таджикской речи.

Методы исследования. Для решения целей и задач, поставленных в исследовании, используются следующие методы:

- методы анализа научно-исследовательских источников в сфере речевых технологий;
- методы статистического анализа с использованием возможностей математических и компьютерных моделей;
- методы анализа и моделирования на основе технологий искусственного интеллекта и машинного обучения;
- экспериментальные методы Скрытой Марковской Модели и случайных полей, теории информации и обработки звуковых сигналов;
- методы объектно-ориентированного программирования и обработки реляционных базы данных.

Научная новизна работы. В ходе исследования предлагается новый подход к созданию акустической модели ключевых слов с использованием акустических моделей фонем, отличающихся от известных моделей, в том числе и в языковом направлении. Впервые решена задача качественного и точного

распознавания таджикских слов на основании сравнительного фонемного анализа:

- реализован новый метод распознавания ключевых слов на таджикском языке;
- реализован новый метод представления ключевых слов с применением скрытой Марковской модели и случайного поля;
- проведён сравнительный анализ результатов работы предлагаемых методов на коллекции из 20 дикторов 300 слов, подтверждающий их эффективность;
- создан комплекс алгоритмов и программ для обработки базы данных большого объема, реализующий описанные в данной работе методы.

Теоретическая и научно-практическая значимость заключается в возможности применения созданной модели распознавания ключевых слов на таджикском языке с минимальной вероятностью ошибок. А также, исследованы некоторые способы представления речевого сигнала такие как, простейщая цифровая модель, упрощенная дискретная модель, коэффициент линейного предсказания, скрытая Марковская модель, N-граммные модели.

На защиту выносятся следующие положения:

1. Метод построения системы распознавания ключевых слов на таджикском языке.
2. Новый метод представления ключевых слов с применением скрытой Марковской модели и случайные поля.
3. Сравнительный анализ результатов предложенных методов по коллекции, состоящему из 20 дикторов и 300 слов, что доказывает эффективность компьютерной программы.
4. Программные средства, входящие в состав системы распознавания ключевых слов в речи на таджикском языке.

Достоверность и обоснованность научных результатов подтверждается корректным использованием известных научных методов обоснования

полученных результатов, выводов и рекомендаций. Были изучены и критически проанализированы известные достижения и теоретические положения других авторов. Обоснованность результатов основывается на воспроизводимости и согласованности данных компьютерного моделирования и научных выводов. Полученные научные результаты основываются на известных достижениях фундаментальных и прикладных научных дисциплин, таких как математика, математическое моделирование, теория вероятности, теория системы, нейронные сети, нечеткие системы и вейвлеты.

Соответствие диссертации паспорту специальности. В диссертации присутствуют оригинальные результаты одновременно из трех областей: математическое и программное обеспечение вычислительных машин, комплексы и компьютерные сети, что соответствует паспорту специальности 05.13.11 – Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей, по пунктам 2, 4 и 9.

Личный вклад автора заключается в определении основных задач исследования и определении основного направления исследований под руководством научного руководителя. Алгоритмы и программные продукты, а также окончательные результаты исследования были получены автором самостоятельно.

Апробация диссертации. Основные результаты диссертации обсуждены в следующих семинарах и конференциях:

- семинар кафедры технология программирования и компьютерный техника в институте технологий и инновационного менеджмента в городе Куляб (Куляб 2018-2022 гг)
- совместный семинар кафедры факультета информационных технологий и искусственного интеллекта Технологического университета Таджикистана под руководством к.ф.-м.н. доцента Исмоилова М.А. (Душанбе, 2018-2021 гг.)
- семинар кафедры программирования и информационных систем Политехнического института Технического университета Таджикистана имени

академика М.С.Осими под руководством кандидата технических наук Косимова А.А. (Худжанд, 2019 г.)

- совместный семинар кафедр факультета информационно-коммуникационных технологий Технического университета Таджикистана имени академика М.С.Осими под руководством кандидата технических наук, доцента Гафурова М.Х. (Душанбе, 2019-2022 гг.)
- международная научно-практическая конференция «Роль интеграции науки, инновации и технологии в экономическом развитии стран» (Куляб, 2020г.);
- международная научно-практическая конференция «Проблемы информационной лингвистики, учебные и инновационные технологии», ТТУ имени акажемика М.С.Осими (Душанбе, 2019г.).

Публикации. По результатам проведённых исследований опубликовано 8 статей, в том числе 4 статьи в журналах из перечня, рекомендованного ВАК для публикации результатов диссертационных работ. Получено свидетельство об официальной регистрации программы для ЭВМ в отделе по правам и защиты прав автора Министерства Культуры Республики Таджикистан. А также, по результатам исследований получено авторское свидетельство о государственной регистрации информационного ресурса в патентном центре при Министерстве экономического развитияи и торговли Таджикистан.

Объем и структура диссертации. Диссертационная работа состоит из введения, 4 глав, заключения, списка использованной литературы и приложений. Основной текст размещен на 132 страницах, включает 18 таблиц, 32 рисунков. Список литературы, включает 124 наименований.

ГЛАВА 1. ОБЗОР МЕТОДОВ И МОДЕЛЕЙ ПОИСКА КЛЮЧЕВЫХ СЛОВ В РЕЧИ

1.1. Обзор моделей представления речевых сигналов

Голосовая система человека представляет собой особую акустическую систему, состоящую из нескольких каналов: ротового и носового. Эта система возбуждает поток квазициркулярных импульсов — колебаниями голосовых связок и турбулентным звуком.

Турбулентный звук возникает, при проталкивании воздуха через узкий путь звука в некоторых местах. Звуковое устройство, воздействующее на элементы, действует как записывающий фильтр, параметры которого со временем меняются. В результате генерируется речевой сигнал.

На коротких интервалах приближение возбуждаемого сигнала можно предсказать по импульсным характеристикам звукового тракта.

Упрощенный пример звукового сигнала на рис. 1.1. описан. Согласно этой модели, звонкие звуки (голоса) генерируются генератором последовательных импульсов, а фрикционные звуки генерируются генератором случайных чисел.

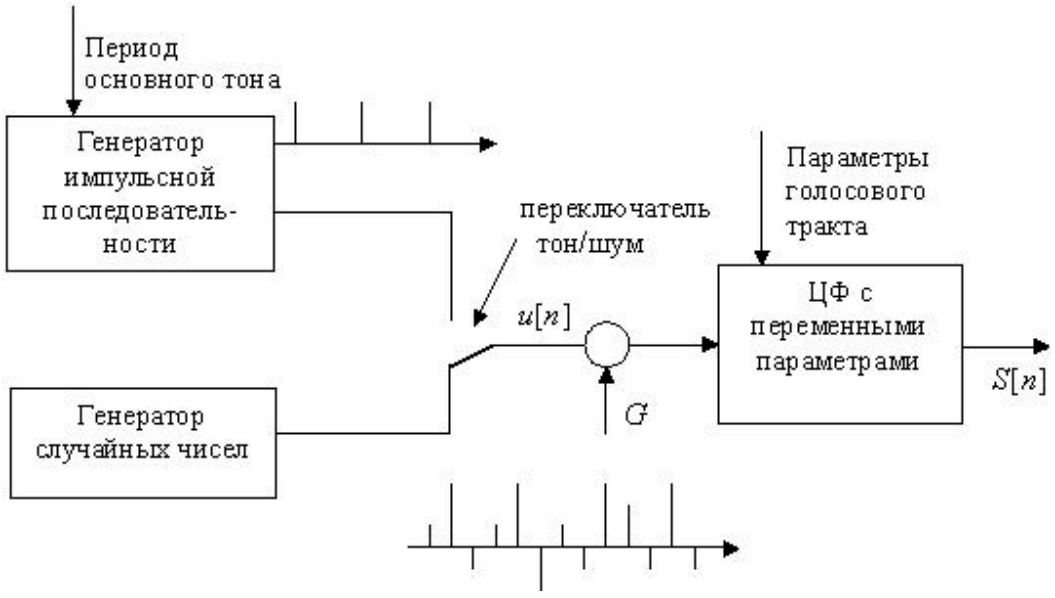


Рисунок 1.1. - Цифровая модель формирования речевого сигнала

Период повторения импульсов совпадает с выходом импульсной последовательности в соответствии с основным периодом стимуляции голосовых связок. Генератор создаеть звукового сигнала с той же спектральные плотности. Цифровой фильтр с переменными данными приблизительно совпадает с передаточной особенности голосового тракта. Вид голосовой связки не меняется на интервалы около 3-20 мс. Таким образом, характеристика цифрового фильтра в этом диапазоне будет постоянными. Амплитуда входного сигнала u(n) цифровой фильтр определяется с увеличением G.

Рассмотренная модель генерация речевого сигнала лежит в основе многих способов описания речевых сигналов: из периодического выделения речевого сигнала до оценивания параметра модели, которые описаны на рисунке 1.1.

В работе автор рассматривает, выбор правильного способа описания речевого сигнала определяет ту задачу, которую важно решить [100]:

- анализ речи – является составная часть любой систем распознавания речевых сигналов и систем идентификации;

- синтез речи – потребность синтез речи возникает в рамках информационной и справочной системы;

- анализ системы сжатия речевого сигнала – для достижения к целям отправка речи по компьютерной и телекоммуникационных сетей или по классическим каналами связи.

Рассмотрим более подробно классификацию речевых сигналов [100].

Известно, что звуки речи разделяются на гласные и согласные. Гласные звуки представляет из себя некую геометрическую форму, которая формируются в результате трансформации вокализованного сигнала при ее прохождении через речевой тракт. Изменение геометрии речевого тракта приводит к изменению акустических резонансных свойств, в результате чего некоторые частоты увеличиваются, а некоторые – ослабляются. Эти зоны называют форматными частотами или форматами. Отличие гласных звуков состоит именно в различиях форматной структуры (рис. 1.2).

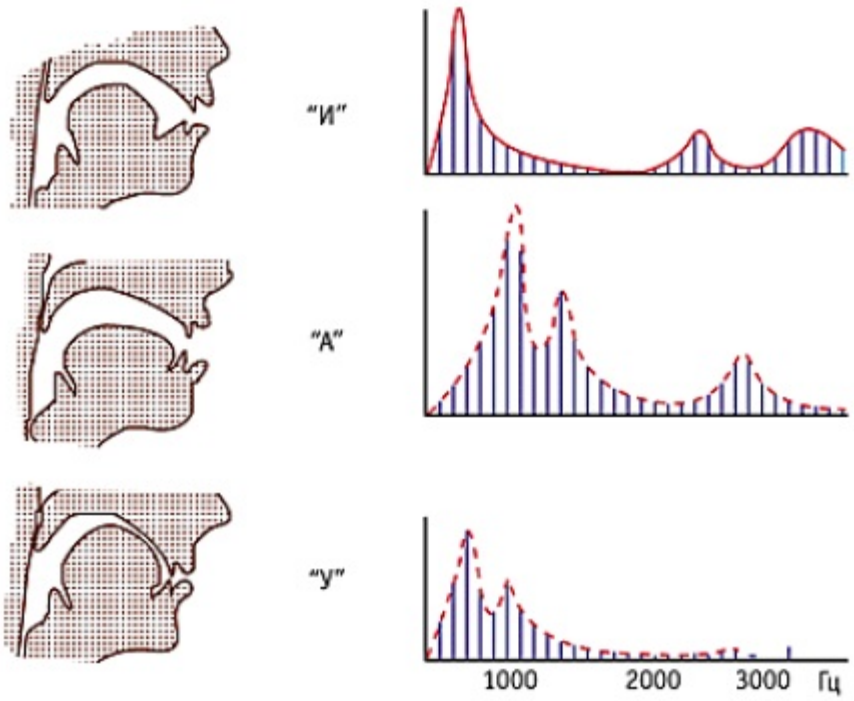


Рисунок 1.2. - Различия гласных по форматной структуре

Согласные звуки представляют собой три условных групп (поскольку единая классификация затруднена по причине наличия у феноменов признаков разных классов):

- фрикативные;
- смычные;
- сонорные.

Фрикативные согласные формируются посредством «трения» потока воздуха при сужении речевого тракта, которое создаются языком, зубами губами и т.д.

Смычные согласные формируются посредством полного перекрытия речевого тракта органами артикуляции при открытой голосовой щели. Воздух, поступающий в открытую голосовую щель под сильным давлением, создает «взрывной» звук при быстром размыкании препятствий.

Сонорные согласные образуют особенную группу. Они не содержат сильного турбулентного шума, поскольку при произношении для воздуха формируется дополнительный проход.

Целесообразно упомянуть переходные звуки и призвуки. Они формируются в случае, когда органы артикуляции в слитной меняют свое положение плавно во времени. В фонетике выделяют следующие стадии произнесения звука: экскурсия (начальное положение органов), выдержка (произношение) и рекурсия (перестройка к произношению следующего звука) [60].

Данная коартикуляция порождает многочисленные призвуки, которые не входят в алфавиты, но могут быть классифицированы.

Стоит отметить также, что гласные несут количество энергии больше, чем согласные. Это можно увидеть на спектрограмме слова «Вошел» (рис. 1.3).

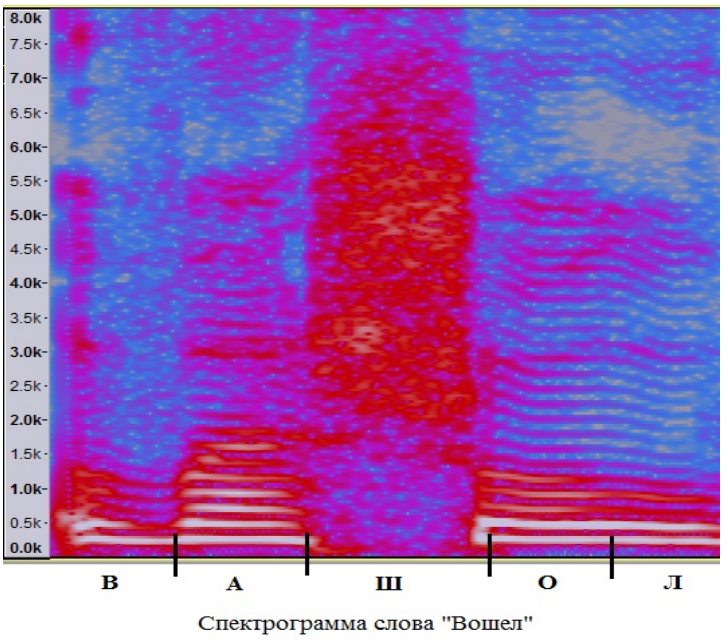


Рисунок 1.3. - Спектрограмма слова «Вошел»

Так же можно рассмотреть, согласно целям исследования, спектрограммы слова «даромад» (рис. 1.4) и предложения «ман ба хона даромадам» (рис. 1.5).

Здесь можно отметить, что на гласный звук «А» приходится некоторое количество энергии, но уже на звук «О» приходится энергии больше, чем на все остальные. Таким образом, ударные гласные несут больше всего энергетического потенциала [2–А].

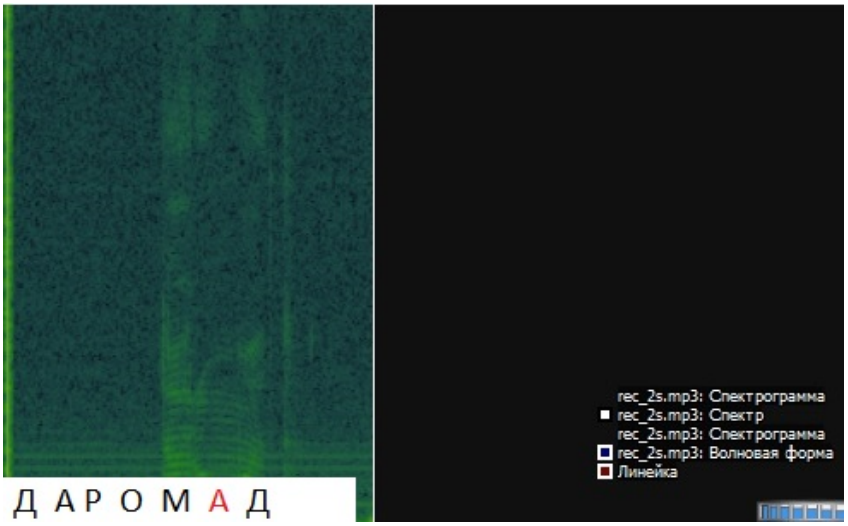


Рисунок 1.4. - Спектрограмма слова «Даромад»

На спектрограмме (рис. 1.4) можно увидеть, что слово, произнесенное на таджикском языке, имеет ударную гласную, энергия которой больше остальных неударных гласных. В произнесенном предложении «ман ба хона даромадам» на таджикском языке, наблюдается такая же ситуация, как и с отдельными словами. Однако, как можно отметить из представленной спектрограммы, показатели увеличиваются на каждом слове, то есть мы видим несколько всплесков на ударной гласной каждого слова [2–А].

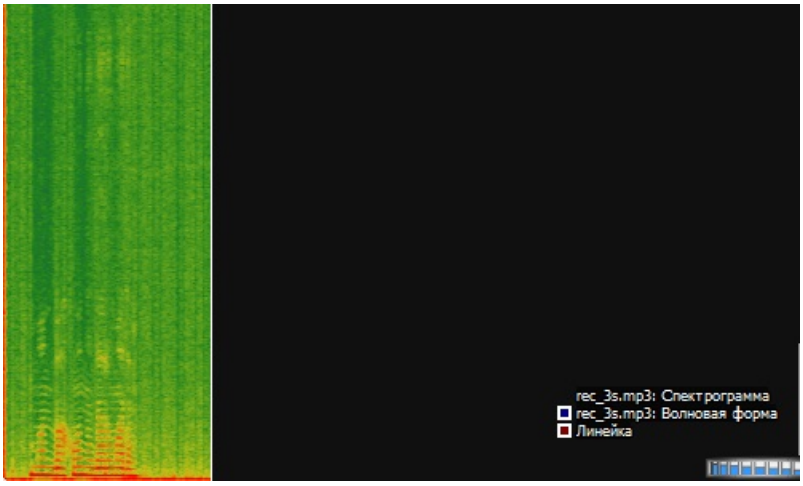


Рисунок 1.5. - Спектрограмма предложения «ман ба хона даромадам»

Существует упрощенная дискретная модель речевого сигнала, суть которого заключается в предположение что дискретный сигнал y(n) является

выходом линейного цифрового фильтра h, через которого проходит и «возбужденный» сигнал x(n).

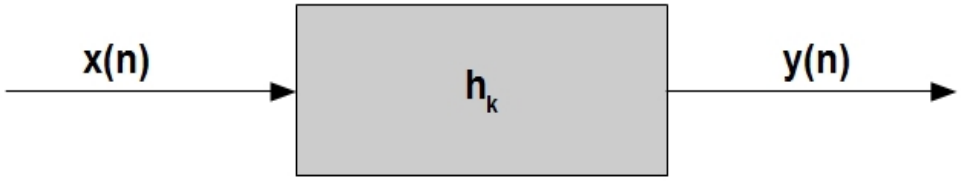


Рисунок 1.6. - Упрощенная дискретная модель

Достаточно подробно автором изложены [62] что, если изменить коэффициенты фильтра h_k и в некоторых случаях сам сигнал возбуждения, то можно добиться того, что звук на выходе будет совершенно другим.

Однако, на практике возникает много неопределенностей в фазе распределения сигнала на голосовой и на безмолвной сегменты. Это требует слишком сложной обработки сигнала через математические устройства. Кроме того, большие скорости изменения h (n), x (n) стимулируют не стационарность и сложность характера процесса. С учётом этого, для повышения результативность оценки используемых в рассматриваемом модели параметров, необходимо проводить обработки сигнала на промежутках времени длительность которых во много раз больше, чем периода основного исходного сигнала. Модель очень ограничена для практического использования и для описания фрикционных и «взрывных» звуков.

От данной модели можно перейти к модели, которая дает возможность оценить параметры речевого сигнала (спектральный анализ) – LPC (коэффициенты линейного предсказания).

Коэффициенты в данной модели непосредственно могут описывать речевой тракт. Несмотря, что такое описание не является исчерпывающим, оно является достаточным для многих случаев.

Для того, чтобы получить описание состояния речевого тракта на каком-либо сегменте речи, требуется решить задачу оценки коэффициентов.

Таким образом, совокупность численных коэффициентов характеризующих полярных фильтров составляют основу алгоритмы расчета LPC. Вытяжки коэффициентов даёт возможность выхода фильтра в временной области (формула 1.1), и к общему виду характеристики z (формула 1.2):

$$V(z) = \frac{G}{1 - \sum_{k=1}^{p} 1\, a_k \cdot z^{(-k)}} \qquad (1.1)$$

$$V(n) = Gg\,(n) + \sum_{k=1}^{p} 1\, a_k \,.\, v\,(n\text{-}k) \qquad (1.2)$$

где V — звуковая линия, G — некоторый комплексный полином, зависящий от коэффициента отражения r, k — некоторые зависящие от r действительные коэффициенты, P — количество труб в описаном выше модели.

Из разложения z-характеристики в знаменателе можно получить значения тех частот, которые относятся к полюсам этого фильтра. Значения этих же частот позволяют аппроксимировать форматные частоты речевого тракта в конкрентном сегменте речи.

LPC сигнал полученный при помощи возбуждения дискретного фильтра выходит похожим на белый шум или на дельта-функцию смещенной во времени.

LPC-анализ в общем, является очень эффективный метод. Однако модель содержит ряд связанных с особенностью прохождении звукового сигнала от разных людей и от протеканиием сигнала в области губ, что не очень соответствует моделью речевого тракта в виде дискретного фильтра.

Однако можно восстановить некоторую функцию, которая повторяет площадь речевого такта, но отличается от нее на какой-то масштабирующий множитель.

На участках, где имеется сигнал, модель показывает отличные результаты, однако на переходящих участках модель менее эффективна.

В работе авторов [38] рассматриваемой на практике этот метод часто используется как основной при определении частотных значений форматов, с помощью которого уже можно восстановить функцию звукового поля.

Метод банка фильтров [24] также относится к категории, к которой принадлежит предыдущий метод – сжатия спектров. Суть заключается в том, что если обозначить входной сигнал как s (t), тогда на выходе из фильтра сигнал будет представлять собой краткосрочный спектральный сигнал в момент времени t. Очевидно, что в данной модели фильтр обрабатывает сигнал независимо. Однако перед началом анализа речевой сигнал должен быть обработан предварительно. Для этого убирается шум, долгосрочные спектральные тренды и происходит выравнивание сигнала в спектральной области.

Особо следует подчеркнуть, что в более общем виде структура распознавания речи через модели представления речевого сигнала в себе включает методы формирования акустических единиц речевого потока: непараметрические (методы основаны от формальных грамматик и метрик на множестве речевых сигналов) и параметрические (вероятностные, на основание скрытых Марковских моделей, нейросетевые) [4].

При акустической обработке происходит сегментация (формирование последовательности перекрывающихся участков исходного сигнала), выделение признаков (сопоставление каждому речевому сегменту вектора признаков; выбор вектора зависит от решаемой задачи – условий записи, языка и т.д.) и моделирование акустических единиц (сопоставление последовательности векторов признаков и последовательности акустических единиц).

При непараметрическом методе сохраняется копия каждой последовательности векторных признаков для каждого выражения словаря, и в дальнейшем производится сравнение неизвестного выражения с сохраненными копиями.

Параметрический метод предполагает обучение параметрической модели для каждого выражения, в дальнейшем происходит сравнение неизвестного выражения с сохраненными моделями.

Скрытые Марковские модели (HMM или CMM) – это мощный инструмент, который позволяет распознавать речевой сигнал с высоким качеством.

Основными часть HMM является:

- ненаблюдаемая цепь Маркова с конечным числом случаев, матрицей переходных вероятностей и вектором вероятностей начальных условий

- функция плотности связана с каждым случаем

Таким образом, HMM представляет из себя модель, состояние которого может меняется в любой выделенный (дискретный) промежуток времени t (рис. 1.7). Вероятность a_{ij} перехода из одного состояния (s_i) в другое состояние (s_j) совершается случайным образом. В каждый дискретный момент времени модель наблюдает вектор свойств, получаемый в преобразователе сигналов с вероятностью bj (ot).

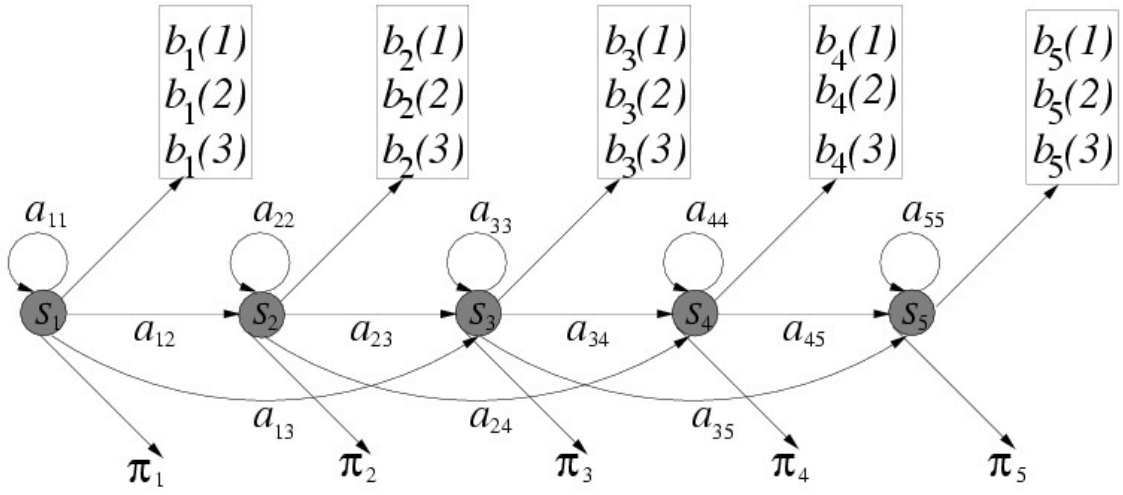


Рисунок 1.7. - Представление скрытой Марковской модели

Разность плотностей вероятностей наблюдений по итоговой системе Гаусса состоит из двух компонентов. Каждый из этих узоров является признаком одного из звуков языка или отсутствия звука.

Существуют различные алгоритмы распознавания речи, основанные на Марковской модели.

N-граммные модели основываются на предположении, что вероятность появления очередного слова в предложении имеет зависимость от n-слов.

Более популярными на сегодняшний день становятся биграммные и триграммные модели языка. Поиск осуществляется по корпусу. Однако, несмотря на то, что алгоритм в данных моделях работает быстро, сами модели не могут улавливать синтаксические и семантические связи, если слова находятся на расстоянии пяти слов друг от друга, а использование n-граммных моделей, где $n \geq 5$, является довольно сложным алгоритмом и требует больших мощностей [20].

Заслуживает внимания работа [121] и модель HNM основное содержание, которых опирается на концепции определении максимума частоты вокализованного сигнала, для каждого сегмента речи. В сегментах вокализованного и переходного сигнала статус гармоники приобретают компоненты до определенной частоты, а компоненты с более высокими частотами воспринимаются как шум. Формирование структуры сигнала осуществляется перекрытием и накоплением компонент. Перекрывающиеся компоненты концентрируются относительно сегменту для формирования когерентной фазы. Шумовые компоненты оцениваются в интервале анализа, отмечаются в виде окрашенного фона белого гауссова шума кратковременным фильтром (LPC).

HNM применялась и для синтезатора текста в речь (TTS), и для системы конверсии голоса. Исследования показали, что для TTS HNM превосходит TD-PSOLA (алгоритм изменения ЧОТ) по всем параметрам, кроме математической сложности. Максимальная частота вокализованного сигнала ограничивается значением 4 кГц, при этом речевой сигнал оцифровывается с частотой дискретизации 16 кГц [114].

Таким образом, мы обозначили не только базовые модели, на которых могут строится более прогрессивные методы анализа и синтеза речи, но и

некоторые модели, которые на данный момент используются, как самостоятельные. Однако, если взять во внимание достаточность каждой из моделей, на основании спектрального анализа или с основанием скрытых Марковских моделей, стоит сказать, что в каждой из них имеются недостатки, которые не могут позволить в полной мере представить речевой поток.

1.2. Анализ методов поиска фрагментов в слитной речи

Для обработки речевого потока и его представления в виде некоторой системы звуков необходимо выделить эти звуки. Человеческая речь является слитным потоком звуков, которые распознаются системами и на основании этого можно воссоздать фразы и произносимые тексты. Для того, чтобы обработать таким образом звуки, необходимо провести сравнительный анализ обученной модели и речевого потока.

Более подробно распознавание слитной речи происходит условно в три этапа. После того, как предварительно будет обработан речевой сигнал и из него будут выделены информативные признаки, производится выделение лексических фрагментов. Второй уровень предполагает выделение слогов и морфем, третий – слов, предложений и сообщений (рис.1.7).

Поиск ключевых слов является одной из наиболее сложных задач в обработке речи. С помощью решения данной задачи можно реализовать аудиоиндексацию и поиск информации.

Проанализируем базовые методы, которые могут составлять основу любых сложных методов в системах распознавания речи:

- распознаватель со словарем;
- методы на основе HMM;
- метод на основании применения решеток фрагментов.

Самый простой метод поиска ключевых слов применяет распознаватель с большим словарем для перевода слитной речи в текст. Для того, чтобы

осуществить поиск ключевого слова, производится поиск в полученном тексте с применением классических алгоритмов поиска текста (см. рис. 1.8).

Рисунок 1.8. - Этапы распознавания слитной речи

Проблемный момент в этом методе заключается в ограниченном наборе слов в распознавателе, в связи с чем не представляется возможным распознать слова [33], которые отсутствуют в словаре, к примеру, акронимы, иностранные слова или имена.

Второй метод основывается на скрытых Марковских моделях (HMM), которые используются для каждого ключевого слова, с применением одной модели «мусора» для остальных слов. Данный метод практически не содержит ограничений при условии, что установлено множество ключевых слов, требуемых для поиска. Однако для каждого нового ключевого слова требуется не только обучить новую HMM, но и модель «мусора».

Третий метод [61] является наиболее распространённым решением в поиске ключевых слов в слитной речи, где применяются акустико-фонетическая HMM и расчет апостериорных вероятностей фонемной решетки, где каждый узел ассоциирован с моментом времени в рамках слитной речи.

Поиск ключевых слов на основе клетки речевых фрагментов имеет то преимущество, что даже если фонема ключевого слова не является лучшей гипотезой между точками клетки, она остаётся результатом распознавания. Результат поиска не зависит от словаря, т. к. поиск может производиться для любой последовательности фонем искомых слов.

Метод определения концов слова используется для фильтрации речи от помех и уменьшения количества арифметических операций, так как обрабатываются только те отрезки, которые имеют звуковой сигнал. Для этого можно использовать метод Рабинера-Самбура, который основан на расчете энергии фрейма и частоты переходов через ноль.

Для вычисления значения энергии может быть применен метод нахождения евклидовой нормы. Частота переходов через ноль определена как количество раз, когда исходный сигнал изменяет свой знак и его значение устанавливается выше порога шума.

Энергия фрейма – нормированная сумма абсолютных значений амплитуд дискретных отсчетов сигнала:

$$E = \frac{1}{\mathrm{K}} \sum_{\mathrm{n}=1}^{\mathrm{N}} 1\mathrm{A}_{\mathrm{n}}. \qquad (1.3)$$

где К- коэффициент нормировки, N – длина фрейма.

Коэффициент нормировки выбирается равным длине фрейма.

Содержание метода Рабинера-Самбура рассмотрим как показано на рисунке 1.9.

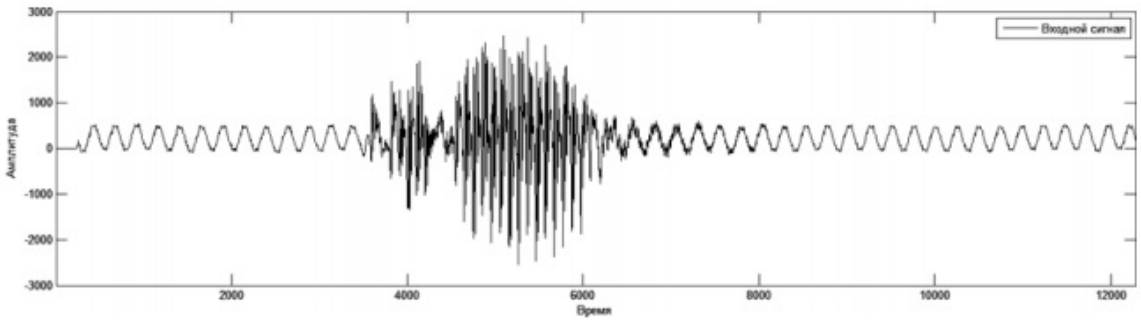


Рисунок 1.9. Временная диаграмма слова «дар»

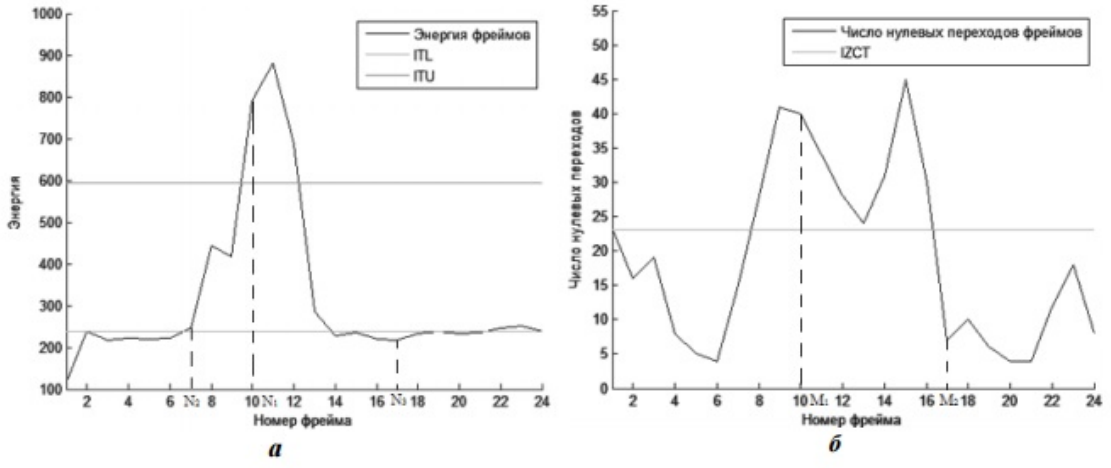


Рисунок 1.10. - Изменение значений:
***а* – энергии фреймов слова «раз», *б* – нулевых переходов слова «раз»**

При вводе сигнала подразумевается, что начальные фреймы (~86 мс) не содержит в себе полезной информации. Обрабатывая этот участок можно определить статистические особенности шума. Затем, выполняется поиск фрейма, в котором количество переходов через ноль и количество средней энергии и превышают пороги. В случае, если такой фрагмент найден (знач. N_1 на рис. 1.10 (а) и M_1 на рис. 1.10 (б)), то это означает, что фрейм содержит полезный сигнал и начало слова находится вне пределов.

При записи сигнала в память определяется конечный момент слова. Поиск конечного момента похож на принцип поиска начала и содержит точку тогда, когда выполняется условие:

$$\begin{cases} \mathrm{E_i} < ITU \\ \mathrm{Zi} < IZCT \end{cases} \quad (1.4)$$

Если выполнено всего одно условие, это еще не означает, что найдена конечная точка. Существуют слова, которые содержат моменты тишины между фонемами. В этом случае должна быть определена максимальная длительность тишины – отрезок времени, в пределах которого значения параметров могут превысить соответствующие уровни.

Данный метод используется при обработке входящего речевого сигнала и в дальнейшем, после выделения фрагмента, обрабатывается в соответствии с применяемой системой.

Однако рассмотренные нами методы являются случаями частного порядка. Они могут использоваться как в структуре систем обработки речевого потока совместно или по отдельности, дополняясь другими методами, разработанными на их основе.

Можно рассматривать различные системы поиска фрагментов и ключевых слов – усовершенствованные авторские модели призваны устранить несовершенство базовых методов, на основе которых были сделаны разработки. Например, в работе авторов [73] обращают внимание на то, что, модель для поиска ключевых слов на основе слоговой клетки, которая состоит из трех этапов: распознавание обучение через языковой и акустической базы данных; поиск в решетке возможных слов; проверка ключевых слов на основе шкалы достоверности. Соответственно, поиск ключевых слов выполняется на основе решетки и усовершенствованной меры достоверности, чтобы исключить ложного отказа и ложного срабатывания. Однако, это уже не называется метод, а комплекс методов, которая может быть полностью применена в какой-то отрасли.

Здесь также можно отметить, что в основе моделей лежат модели описания звуков речи – акустическая и языковая. При этом языковая модель более простая, акустическая – более сложная. Языковая модель задает вероятность слова, которое следует за каждым другим словом, то есть определяет последовательность. Акустическая модель содержит доказанную гипотезу, согласно которой во всех человеческих языках используются звуки, набор которых ограничен, то есть набор фонем. В связи с этим акустическую модель можно разделить на две составляющих. Первая касается произношения и задает для конкретного слова распределение вероятностей по множеству последовательностей фонем. Поскольку фонемы не являются наблюдаемыми, речевой поток может быть представлен как скрытая Марковская модель,

переменная состояния которой X_T определяет, какая именно фонема произносится в определенный момент времени T. Вторая часть акустической системы относится к методу, при помощи которого фонемы могут быть реализованы в виде акустических сигналов. То есть, переменная E_T для СММ задает наблюдаемые признаки акустического сигнала в момент времени T, а акустическая модель устанавливает вероятность $P(E_T \mid X_T)$, где X_T – текущая фонема. Данная модель дает возможность учитывать ударение, скорость и другие характеристики речи, и основана на способах обработки речевого потока, которые позволяют создавать представления сигнала и являются довольно устойчивыми к различным влияниям.

Другие существующие модели будут рассмотрены далее в рамках анализа систем для распознавания речи.

1.3. Обзор существующих программных систем поиска ключевых слов в речи

В предыдущих параграфах мы уже отметили некоторые системы, которые базируются на ключевых методах представления речевого потока и поиска ключевых слов и лексических единиц. В этом разделе анализируем современные методы решения проблемы распознавания речи и поиска ключевых слов в слитной речи. Однако граница между данными задачами является условной: распознать слитную речь – это и значит выделить какие-то сегменты, при совокупности и сложении которых на выходе система распознает речь как одно целое. Все системы, в первую очередь, подразумевают поиск ключевых слов или лексических элементов в слитной или раздельной речи. Поиск зависит от различных факторов и условий, и должен определяться несколькими составляющими для качественного поиска ключевого слова. Если при создании систем принимается во внимание только один фактор, и при этом остальные не учитываются, но существуют в системе в любом случае, система не будет

работать на качество и не сможет использоваться по назначению. Элементы и признаки любой системы, следующие:

1. Размер словаря – чем он объемней, тем больше появляется ошибок при распознавании слов. Также учитывается уникальность слов словаря. Повышается уровень системных ошибок, если в словаре есть похожие слова.

2. Дикторозависимость – существуют дикторозависимые и дикторонезависимые системы. Дикторозависимая предназначается для работы с ограниченным количеством человек, чаще – с одним. Дикторонезависимая система работает с любым количеством. На данном этапе развития систем распознавания речевого потока дикторонезависимая система имеет больше ошибок, чем дикторозависимая.

3. Раздельная или слитная речь – раздельная речь подразумевает произношение слов с определенным промежутком, слитная является более естественной речью, но при этом ее распознавание более сложное из-за нечеткости границ слов.

4. Структурные единицы – могут выступать фонемы, дифоны, аллофоны, слова и фразы. Те системы, которые распознают речь, по целым словам, или фразам являются дикторозависимыми. Создание таких систем обычно менее трудоемко, в отличие от систем, где использует поиск слов по фонемным решеткам или по другим единицам речи.

5. Принцип выделения структурных элементов речи – в системах используются различные подходы к распознаванию и выделению единиц:

- преобразование Фурье - преобразует исходный сигнал из временного амплитудного пространства в частотно-временной интервал. Между тем, во временной области преобразование Фурье представляет собой линейное предсказание речи, которое может описывать речевые сигналы с помощью авторегрессионной модели. Однако у преобразования Фурье есть свои недостатки, заключающиеся в том, что теряется информация о временных характеристиках речевых сигналов.

- вейвлет-анализ – предусматривает разложение сигнала в базис функций, которые характеризуют и частоту, и время. При помощи вейвлет-анализа можно проводить анализ свойств в физическом и в частотном пространстве. Кроме того, каждому вейвлету соответствует сове преобразование, поэтому можно подобрать такую функцию, где определены свойства частотно-временной локализации;

- кепстральный анализ [65] – создание систем с таким элементом является довольно трудоемким процессом и требует от создателя очень высокой квалификации. В общем виде можно представить кепстральный анализ в виде обратное преобразование Фурье от логарифма прямого преобразования. Вообще кепстр – это спектр логарифма спектра исходного сигнала. Кепстральная обработка:

- в отличие от спектра, нет понятия усреднения, потому что данные вводятся один раз;

- при вводе данных необходимо сделать 80 отсчетов. Если вводится меньше данных, часть информации теряется, если больше — происходит дублирование данных;

- направление чтения не имеет значения;

- частота опроса также не имеет значения.

Важно понимать в этом случае, какие отличия имеет кепстральный анализ от спектрального, то есть необходима выяснить в чём заключается отличия частотных компонент в этих двух видах анализа. В традиционном спектральном анализе любая частотная компонента имеет физический смысл, в кепстральном анализе наличие частотных гармоник не всегда может являться обоснованием периодичности исходного сигнала. Кепстральный анализ дает основание говорит только о некоторой степени островершинности дискретных составляющих исходного спектра. Таким образом, применение кепстрального анализа требует не только квалификации, но и внимательности при интерпретации результатов и сформулирование выводов.

6. Алгоритмы распознавания – производится вероятностная оценка принадлежности элементов речи к элементам словаря. Вероятность перехода между транзакциями можно определят через матрицу вероятностей A = {aij}, где aij — вероятность перехода из позиция i на позиция j. Вероятность выпадения из M значений параметра в каждом N случае может быть определена с помощью вектора B = {bj(k)}, где bj(k) — вероятность выпадения из k-. значение параметра в позиция j. Вероятность возврата в исходное состояние задается вектором π = {π1}, где π1 — вероятность того, что в первый момент система может находиться в i-м положении. Таким образом, можно представить скрытую модель Маркова следующим образом: λ = {A, B, π}.

Кроме СММ можно использовать и динамическое программирование, и нейронные сети. На последних можно формировать системы, которые способны как к обучению, так и к самообучению. Системы, использующие данные элементы, должны отвечать следующим условиям: разработка заключается только в построении архитектуры (рассмотрим далее); есть возможность контролировать свои действия и их корректировать; есть возможность накопления базы знаний о рабочих объектах; система должна быть автономной.

В отличии от традиционного программирования, использование нейронных сетей предполагает возможность динамического изменения алгоритма при изменении архитектуры, что невозможно в традиционном программировании. Однако качество распознавания в системах, построенных на нейронных сетях, намного ниже, чем в системах на СММ.

7. Назначение – определяет требуемый уровень абстракции распознавания речевого потока. К примеру, распознавание по шаблону используют командные системы. Система диктовки, в отличие от командной, требует распознавания по лексическим элементам, в том числе не только по звучащему элементу, но и по элементу, который звучал до этого. Кроме того, в такую систему может встраиваться грамматическое правила или набор правил. Чем строже правила, тем проще внедрить систему, но тем меньше количество предложений, которые

можно распознать. Ранее было отмечено, что разработка систем заключается в построении архитектуры. Типичная архитектура выглядит следующим образом:

- шумоочистка и выделение полезного речевого сигнала;

- акустическая модель – дает возможность оценить распознавание речевого сегмента на звуковом уровне;

- языковая модель – позволяет определить наиболее вероятностные последовательности слов. Сложность зависит от конкретного языка. К примеру, для английского языка достаточным будет применение статических моделей (N – граммы), а для высокорефлективных (в том числе и русского) используют гибридные модели;

- декодер – предназначен для совмещения данных, которые получены в процессе поиска и выделения элементов речевого потока, и определения наиболее вероятной последовательности слов, которая является итогом распознавания.

Кроме того, могут быть вариации в архитектуре системы, это зависит от разработчика и тех задач, которые необходимо решить в процессе.

Как отмечают разработчики компании «Центр Речевых Технологий», чем больше фонемная база, тем лучше. На рынке присутствует продукт, который можно назвать «пофонемным распознаванием». На основе данной технологии были разработаны системы для распознавания речи (CSREgine) и поиска ключевых слов (VoiceDigger).

Относительно поиска слов системы могут подразделяться по задачам следующего характера:

выявить и определить позиция ключевых слов или словосочетания (в фонограммы и речи, основы речи, в реальном время);

- распознавать команды в непрерывной речи (например, навигация по меню);

- понять смысл полной речи через поиск ключевых слов для систем общения.

Таким образом, состояние системы имеет зависимость и от области применения.

Системы поиска ключевых слов применяются в call-центрах, службах безопасности, радиокомпаниях, телекомпаниях, телекоммуникационных организациях, службах безопасности и других компаниях, которые используют большой поток информации или архивные базы. Кроме того, что программные системы поиска ключевых слов используются для речевого потока, они применимы и для аудио и видеопотоков.

Указанный программный модуль поиска ключевых слов VoiceDigger работает на основе синтеза и распознавания фонем. Система SL-SpeechAnalytics [65], разработанная компанией «Гран-при», работает на основе технологий распознавания ключевых слов и изменения эмоционального фона говорящего. Таким образом, можно отметить, что использование технологий в системах поиска ключевых слов в речи представляют собой применение базовых моделей в одной системе. Кроме того, существуют авторские модули, предложенные молодыми исследователями, создание и применение которых не поставлено на коммерческий поток.

К примеру, можно рассмотреть систему, основанную на решетке слогов и усовершенствованной мере достоверности (отмечена нами ранее) [75]. С целью повышения точности выражения сигнала с помощью последовательности фонем был разработан полный алгоритм MMR на основе сложной матрицы, что позволило учесть статистику ошибок предсказанного акустического моделирования. Создается гипотетическая последовательность слов на первом этапе распознавание для реализации MMP. Для каждого слова вероятностное и последовательность икоты W' слогов, и рассчитывается MMR для фактической и приблизительной строк. Затем межстрочное состояние сравнивается с окончательным значением. Таким образом, при использовании усовершенствованной MMP получается набор возможных слов, включающий большое количество замен и сгенерированных входных данных. Для каждого предоставленного ключевого слова проводится проверка с использованием

шкалы достоверности. Исследования, проведенные учеными, показали, что наилучший результат можно получить при определении меры достоверности на основе одного признака, используя следующие вероятности, которые рассчитываются на основе данных, содержащихся в решетке.

Поскольку алгоритм MMP анализирует измененную последовательность фонем с точки зрения вставки и замены, использование вероятности пастеризации в ее классической форме становится ненадежным для проверки. Таким образом, в связи с резким увеличением количества вариантов строк следует использовать доверительную меру, учитывающую предыдущую информацию о желаемой последовательности слогов в ключевом слове. Для тестирования рекомендуется использовать полное измерение достоверности MMR с учетом схожести строк:

$$\mathrm{CM1}\ (\mathrm{W}) = \mathrm{k}_1\mathrm{P}(\mathrm{w}|\mathrm{O}) + \mathrm{k}_1\mathrm{P}_{\mathrm{cmed}}(\mathrm{W}) + \mathrm{k}_3\mathrm{Length}\ (\mathrm{W}) + \mathrm{k}_4\mathrm{AS}\ (\mathrm{W}), \qquad (1.5)$$

где k_i (i = 1-4) — const, P (w | O) — вероятностная мера гипотетического ключевого слова, P_{cmed} — улучшенная MMP, Length (W) — количество слогов в ключевом слове W, AS (B) — акустическая устойчивость.

Реализация измерения достоверности CM1 позволяет вычислить измерение достоверности для вероятных слов и уменьшить количество ложных срабатываний, возникающих в результате использования алгоритма MMP.

1.4. Постановка задачи

В настоящее время наука вкладывает большое количество средств в разработку систем распознавания человека и речи. При рассмотрении систем, разработанных в этой области, программных модулей, представляющих большую сами по себе, дают большой ценность и развивающих распознавание речи как науку. Все математические методы распознавания потока речи, обработки и сравнения его с элементами словаря в совокупности дают результат.

Неопытность пользователя, спонтанная неправильная с точки зрения грамматики, стиля и произношения, наличие помех и искажений и речевые помехи — это тех условия, которые, сильно влияющие на качество распознавания речи.

В сегодня существует системы, которые не в полной мере отвечают вышеуказанным условиям, то есть не предназначены для работы в этом направлении. В этом случае разработчики этих системы могут добиться результата 98%.

Разрешить системы большего масштаба с определенным процентом распознавания не позволяет использовать такие недостатки[1–А]:

- предварительная настройка системы может занимать некоторое время;

- некоторые проверки дают, в лучшем случае, 5% ошибок;

- являются словами, которые находятся вне грамматики нашего языка, поэтому количество ошибок увеличивается, при этом механизм исправления этих ошибок недостаточно исправлен, чтобы говорить о высоком качестве системы;

- обработка больших словарей в системах распознавания речи занимает много времени.

Системы, основанные на HMM, предполагают возможные модели человеческого поведения, но это предположение не может быть очевидным. Поэтому при обработке речевого потока следует учитывать фактор звуковой невосприимчивости, т. е. обеспечение достаточной четкости передачи речевого смысла при различных вариантах нарушений речеобразования и восприятия (ситуативных или патологических расстройств). Помехозащищенность может быть обеспечена следующими механизмами [14]:

- параллельно работающие способы выделения одних и тех же элементов речевого потока на основании анализа акустического сигнала (применение форматных признаков и полосных признаков для определения фонетических элементов в речи;

- параллельное использование фонемного и целостного способа восприятия слов в речевом потоке.

В этих механизмах главное не то, что признаки систем повторяются, а то, что применение отдельных слов или словосочетаний не требует наличия всего набора признаков. При этом набор признаков может быть определен по семантическому, шумовому, прагматическому и другим контекстам.

Системы, работающие на распознавание речи, решают задачу понимания смысла речи в большей степени за счет распознавания элементов речи, получения, распознаваемого базы по семантическому модулю [46]. Во многих случаях сигнал на воде в семантическом блоке представляет собой матрицу векторов с вероятностью распознавания каждого сегмента речевого потока, соответствующего слову или словоформе (при успешной сегментации). Семантический блок строит список предложений из векторов вероятностей, ограниченных минимальной вероятностью. Слитная речь иногда настолько аграмматична, что кажется «проглатываемой» с учетом языка говорящего, а также того факта, что могут «проглатывать» суффиксы. Поэтому ответ семантического модуля часто используется с модулем распознавания. Таким образом, повторяя цикл поиска, можно получить более высокий процент правильных утверждений.

Другой способ к пониманию речи включает в себя настройку модуля распознавания перед обработкой входного сигнала. Этот метод будет оправдан только в тех случаях, когда требуется не столько передача информации, сколько выявление общей «семиотической ситуации». В этом случае продуктивность имеет место, если учитывается весь текст, а не только речевой элемент (окружающий диалог), а также если система обладает интеллектом, то есть имеет встроенную базу знаний об окружающем мире, и может понимать и анализировать новую информацию. Проблема разработки такой системы заключается в правильном построении базы знаний, системном обучении и возможностях системного анализа.

Наиболее простым способом реализации этого метода являются модели семантического фреймов, в которых подключается та или иная модель диалога. При этом распознавание ограничивается поиском ключевых слов, по которым из числа «смысловых» моделей выбирается наиболее подходящая.

Вопрос анализа акустических систем также следует учитывать при разработке систем распознавания речи. Это влияет на способность слушателя различать поток речи в присутствии громкого шума. Системы устранения помех и относительного ослабления посторонних звуков при разработке систем распознавания речи нуждаются в дальнейшем развитии и являются одним из основных направлений современной науки.

Мы пришли к выводу, что в настоящее время для таджикского языка должна быть разработана система, отвечающая всем требованиям, а кроме этого, вопрос поиска ключевых слов с учетом особенностей таджикского языка должен быть решен в лучшем виде.

1.5. Выводы по главе

Таким образом, было выяснено, что существует несколько моделей представления речевого сигнала:

- простейшая цифровая модель;
- упрощенная дискретная модель;
- коэффициент линейного предсказания;
- метод банка фильтров;
- СММ (скрытая марковская модель);
- N-граммные модели;
- HNM.

Таким образом, были обозначены не только базовые модели, на которых могут строится более прогрессивные методы анализа и синтеза речи, но и некоторые модели, которые на данный момент используются, как самостоятельные. Однако, если взять во внимание недостаточность каждой из

моделей, на основании спектрального анализа или с основанием скрытых Марковских моделей, стоит сказать, что в каждой из них имеются факторы, которые не могут позволить в полной мере представить речевой поток.

Методы поиска фрагментов в слитной речи определяются полностью системой, в которой они используются и не могут применятся по отдельности. Базовые методы, которые могут составлять основу любых сложных методов в системах распознавания речи:

- распознаватель со словарем;
- методы на основе HMM;
- метод на основании применения решеток фрагментов;
- метод определения конечных точек в слове.

Каждый из указанных методов имеет достоинства и недостатки.

Проблемный момент в методе распознавателя со словарем заключается в ограниченном наборе слов в распознавателе, в связи с чем не представляется возможным распознать слова, которые отсутствуют в словаре, к примеру, акронимы, иностранные слова или имена.

Второй метод базируется на скрытых Марковских моделях (HMM), которые используются для каждого ключевого слова, с применением одной модели «мусора» для остальных слов. Данный метод практически не содержит ограничений при условии, что установлено множество ключевых слов, требуемых для поиска. Однако для каждого нового ключевого слова требуется не только обучить новую HMM, но и модель «мусора».

Поиск ключевых слов на основе клетки речевых фрагментов имеет то преимущество, что даже если фонема ключевого слова не является лучшей гипотезой между точками клетки, она остается результатом распознавания. Результат поиска не зависит от словаря, т. к. поиск может производиться для любой последовательности фонем искомых слов.

Метод определения концов слова используется для фильтрации речи от помех и уменьшения количества арифметических операций, так как обрабатываются только те отрезки, которые имеют звуковой сигнал. Для этого

можно использовать метод Рабинера-Самбура, который основан на расчете энергии фрейма и частоты переходов через ноль.

Современные системы, в которых применяются модели поиска фрагментов или ключевых слов подразумевают поиск слов или элементов в слитной или раздельной речи.

Поиск зависит от различных факторов и условий, и должен определяться несколькими составляющими для качественного поиска ключевого слова:

- размер словаря;

- диктора зависимость;

- слитное или раздельное произношение;

- структурные единицы (преимущественные единицы речи, которые выделяются в конкретной системе);

- принцип выделения структурных единиц: преобразование Фурье, вейвлет-анализ или кепстральный анализ;

- алгоритм распознавания;

- назначения.

Таким образом, от перечисленных условий будет зависеть принцип работы всей системы, механизм представления речевого потока и поиска ключевых элементов. Для каждого потребителя, который использует систему, является главным тот или иной принцип, а поэтому разработка должна производится с учетом конечного потребителя, в зависимости от той информации, которую необходимо получить на выходе.

Неопытность пользователя, спонтанная неправильная с точки зрения грамматики, стиля и произношения, наличие помех и искажений и речевые помехи – это тех условия, которые, сильно влияющие на качество распознавания речи.

На сегодняшний день в системах распознавания речи и выделения ключевых элементов речевого потока главной проблемой являются помехоустойчивость систем и обеспечение понимания смысла речи.

Помехоустойчивость системы означает устойчивость к шумам и искажениям речи, выделение звукового потока в правильном направлении, без потери информации. Смысл речи должен пониматься системами как можно более адекватно, то есть модели выделения ключевых моментов должны быть разработаны таким образом, чтобы учитывались не только особенности «живой» речи в целом, но и язык говорящего.

ГЛАВА 2. РАЗРАБОТКА МОДЕЛЕЙ И МЕТОДОВ ОБРАБОТКИ РЕЧЕВЫХ СИГНАЛОВ НА ТАДЖИКСКОМ ЯЗЫКЕ

2.1. Особенности обработки речевых сигналов на таджикском языке

Для того, чтобы определить, в чем заключаются особенности обработки речевого сигнала, который поступает на таджикском языке, следует рассмотреть общие отличия таджикского языка от русского, на основании которого построены многие рассмотренные нами модели распознавания речи в таблицы 2.1.

Таблицы 2.1. Общие отличия таджикского языка от русского

Основание	Отличия или затруднения
Звуки и их сочетание	Ж, Ц, Ы, Щ. Есть затруднения при произношении мягких гласных, а иногда и твердых согласных
Особенности ударения	В таджикском, в отличии от русского, ударение закрепляется и обычно падает на последний слог слова: хона – «дом», одам – «человек», талаба – «учащийся», хонаҳо – «дома», коргар – «рабочий», давлатманд – «богатый». В заимствованных из русского языка словах ударение может падать на разные слоги.
Грамматические особенности	В таджикском языке нет рода
Особенности алфавита	35 букв – 6 гласных и 25 согласных; Й не является отдельной фонемой, а является графической условностью для ударного конечного И; Йотированные гласные *у, я, ю, е,* после согласных не требуют разделительных Ь, Ъ; Противоречивость буквы Е. После согласных это обычная гласная, в начале слова для [е] используется Э, а буква Е обозначает /йе/
Графические особенности	Каждая буква имеет прописную и строчную буку, имеется курсив; Буквы имеют цифровые обозначения; ғ, ӣ, қ, ӯ, ҳ, ҷ. Это группы партнеры (например, г, ғ; к, қ). Согласные ҷ, қ, ғ, ҳ не имеют соответствий в русском языке; Непроизносимых согласных нет.
Грамматические особенности	Для связи именных частей речи в предложении используется изафетная связь; Прилагательные и числительные, а также большинство местоимений не принимают показателя множественного числа.

Таким образом, в таджикском языке имеются отличия от русского языка, в первую очередь, в количестве букв в алфавите, шесть из них не имеют аналога в русском языке. Для распознавания речи имеет значение и тот факт, что отсутствие в таджикском языке родов и падежных окончаний затрудняет распознавание речи и ее воспроизведение, поскольку при распознавании необходимо опираться только на контекст, из которого должно следовать, о каком объекте идет речь – женского или мужского рода (в соответствии с русским языком).

Вот как выглядит предложение на таджикском языке («Дар асоси амсилаи сохташуда дастабарномаҳое таҳия шудаааааст, ки он имкон медиҳад мушкилоти марбут табдили нутқ ба матнро ҳал кунад. Дастгирии иттилоотии барномаи таҳияшудаи Speech Recognizer як пойгоҳи додаҳоест, ки калимаҳоро тавсиф мекунад.»)

Вот как выглядит предложение на таджикском языке в переводе: («На основе созданной модели разработаны программы, позволяющие решать задачи, связанные с преобразованием речи в текст. Информационным обеспечением разработанной программы Speech Recognizer является база данных, описывающая слова.»):

Так как в таджикском языке нет падежей и рода, то соединение слов осуществляется с помощью изафет (в именных словосочетаниях с прилагательными, числительными, местоимениями), приставок и суффиксов. Дополнительная связь создается с помощью буквы -и, именно она связывает определяемое слово с определительным словом. Например:

«мошини ман» - «моя машина», «бародари ту» - «твой брат», «қалами ту» - «твой карандаш».

Как правило, определяемое слово в таджикском языке стоит перед определением, а в русском языке наоборот – перед определяемым словом. Кроме того, в таджикском языке большинство местоимений, числительные и прилагательные части речи не принимают множественного значения:

«девори баланд» - «высокая стена», «деворҳои баланд» - «высокие стены».

Интерес представляет также (вне зависимости от языка) темперамент говорящего.

Акустические фонетические рационы обычно принимаются во внимание при рассмотрении вербальных признаков, которые связаны с силой и равновесием между возбуждением и ингибированием (по словам И.П. Павлова), и с психологическими симптомами темперамента (согласно дальнейшей систематизации Гиппократа и И. Канта). Акустические фонетические соотношения состоят из следующих типов личности как индикатор достижения языка, введенный Гиппократом:

- сильный тип GNA (sanguine) - Сангвинический темперамент обычно основан на сильном, сбалансированном, мобильном типе более высокой нервной активности, быстром темпе речи, быстрых умственных процессах, однообразных "намерениях", среднем типе громкости, лучше в аспекте беглости, нейтральных слогах. Речь в четко структурированных форматах
- сильный тип HNA (холерик) - Холерный темперамент определяется дисбалансом нервных процессов, гиперактивной, высокой реактивностью и отсутствием подвижности. Изменение средней продолжительности слога происходит из-за отсутствия постоянного темперамента. Среднее значение основного тона (F 0) сильно отличается. Это скорее "тон отставания". Что касается речи, то происходит увеличение вариации (F 0), ударные и безударные слоги не равны, а также изменение объема речевого сигнала и структуры обычно сжатых форматов.
- сильный тип GNA (флегматика) - флегматические темпераменты противоположны сангинским темпераментам. Он в основном инертен. Медленная речь с увеличением продолжительности среднего слога. Он неэмоционален и имеет уменьшение акустической разницы между ударными и безударными слогами. Он также имеет снижение общей насыщенности спектральной энергии, при этом сохраняя четкие структурные особенности форматов, увеличивая продолжительность вокализма и т.д.;

• слабый тип HNA (меланхолический) - У слабого типа HNA (меланхолический) темперамент, есть уменьшение общего объема речи и увеличение паузы. Она состоит из повышенной чувствительности, медленного развития как возбуждающих, так и тормозящих процессов. Существует неопределенность в речи. Обычно существует концентрация спектральной энергии в низкочастотных областях. Есть отсутствие акустических плуга с аудио, а вместо этого заменить их на соответствующую скорость, медленность (увеличение средней длины слога) и так далее.

Речь человека относится к общей функции моторного комплекса (FDC). Процесс формирования речи связан с динамическими характеристиками его умственной активности (темп, ритм, интенсивность психических процессов и настроение), общей деятельностью человека, его двигательными способностями и эмоциональной степенью.

Потапова Р.К., в своей работе [44], указала пару словесных знаков, имеющих свои собственные акустические пропорции и связанные с эмоционально-волевой регуляцией. Положительные и отрицательные эмоциональные состояния различаются по принципу "активного состояния" и "пассивного состояния". Эти два состояния, в свою очередь, связаны с процессами возбуждения и ингибирования, ослабления самоконтроля и подавления реакций.

Гласные и буквы, появляющиеся в речи при активации эмоциональных и ментальных процессов, имеют следующие акустические отношения:

▪ гласные имеют более выраженную гармоническую структуру звуков.

▪ продолжительность спектральной энергии в ударном слоге (гласный) увеличивается.

▪ расширены полосы частот команд F2, F3, F4.

▪ увеличение на 20-30% общей спектральной энергии во всех командных областях.

▪ появление высших командных частот.

Гласные, воспринимаемые в речи при деактивации эмоциональных и ментальных процессов, имеют следующие акустические характеристики:

▪ гласные имеют менее выраженную гармоническую структуру звуков

▪ сокращается длительность спектральной энергии в ударном слоге (гласный)

▪ полосы частот команд F2, F3, F4 сужены

▪ Консонанс имеет определённые звуковые отношения.

Особенно показательны проектировочные параметры, представляющие собой полный набор взаимосвязей на уровне секций.

Для реализации синеза и распознавания речи могут быть использованны следующие типи пауз:

P_s – это пауза между слогами, когда мы произносим слово;

P_w – это пауза между словами, когда мы читаем предложение;

P_i – это пауза, которая отмечает внутренний знак пунктуации;

P_e – это пауза, которая обозначает внешнюю пунктуацию;

P_a – это пауза, которая обозначает конец абзаца.

Символ W обозначает слово с определенной последовательности букв. Когда выражаем гласные буквы через 1, а согласные буквы через 0 (буква «ӯ» воспринимается как согласный), тогда слово W называется слова с упорядоченной последовательностью выражается в виде W^*_0,1 - нулей и единиц. Такое преобразование называется кодированием слова W, а полученный результат- W^*_0,1 слоговой структурой слова W.

Число букв или число двоичных символов в слово W, определяет размер структуры $W^*_{0.1}$.

Если, выражения слов в двоичной форме идентичны и их размерность аналогичны, структура обоих слов считается одинаковыми, в других случаях они считаются разными по структуре. Для каждого слова W связано только одно изображение с W*0.1.

В коллекции $W^*_{0.1}$ было обнаружено 274 различных слоговых структуры таджикских слов, которые составляют минимальный и максимальный размер 1

и 14 слов. Определена связь слоговой структуры слов с частотой их появления в текстах на таджикском языке, выявлено закономерность статистического распределения структур слов.

Анализ выполнение эмпирического соотношения **ν = ν (n)** в каждой группе из 3259 слогов расположенных по порядке убывания их частоты встечаемости показало, статистическое распределение слогов соответствуют определенному числу с частотой ν (в %) для соответствующих слогов. Исследования поеазали, что представленные в таблице 2.2 41 составляют 50% текста.

Таблица 2.2 Частота встречаемости таджикских слогов

n	**слог**	**ν**	**n**	**слог**		**ν**	**n**	**слог**	**ν**
1	и	4,212	6	ди		4,212	11	ми	4,212
2	да	2,452	7	ки		2,452	12	би	2,452
3	ро	2,351	8	о		2,351	13	то	2,351
4	ба	2,239	9	мо		2,239	14	я	2,239
5	ҳо	2,0243	10	до		2,0243	15	ии	2,0243

Вместе с этим было выявлено, что 148 из 3259 слогов составляют 75% текста таджикского языка, 204 слога 80%, 418 слогов 90% и 683 слога 95% текста, а доля остальных - от 684 до 3259 слогов составляют всего 5% текста. Следовательно, вероятность появление каждого отдельного слога из этого набора —редкое событие. В таджикском языке много слов, которые произошли от русских основ. Таким образом, распознаватель настроен на распознавание только таджикской речи, в ситуации по слогового фрагментирования, может столкнуться с проблемой, когда распознавание русского слова осуществить не удастся.

2.2. Разработка структуры и состава системы поиска ключевых слов

Современные распознаватели речи состоят из двух блоков: блок лингвистической обработки текста, при помощи которого происходит построение полной фонетической транскрипции синтезируемого текста, и блок

акустического синтеза, при помощи которого генерируется речевой сигнал, если имеется в виду еще и синтез речи.

Блок лингвистической обработки текста представляет собой довольно сложную структуру, так как создание транскрипции состоит из нескольких этапов: определение языка входной речи, устранение потенциальных орфографических ошибок, осуществление морфологического анализа словоформ для правильной постановки ударения. Самый трудный этап при лингвистической подготовке текста – создание интонации и просодических характеристик фразы. В большинстве случаях для подготовки текста на данном этапе требуется более сложный синтаксический и семантический анализ фразы. Последняя стадия подготовки текста – формирование фонетической транскрипции. На данном этапе используются стандартные правила чтения. Трудоемкость и сложность на данной стадии представляют собой отношения между произношением и орфографией.

После того, как была создана фонетическая транскрипция, вступает в работу второй блок – акустический синтезатор. Задача данного блока состоит в переводе транскрипции в цифровой сигнал, преобразующийся в звуковые колебания с помощью цифро-аналогового преобразователя.

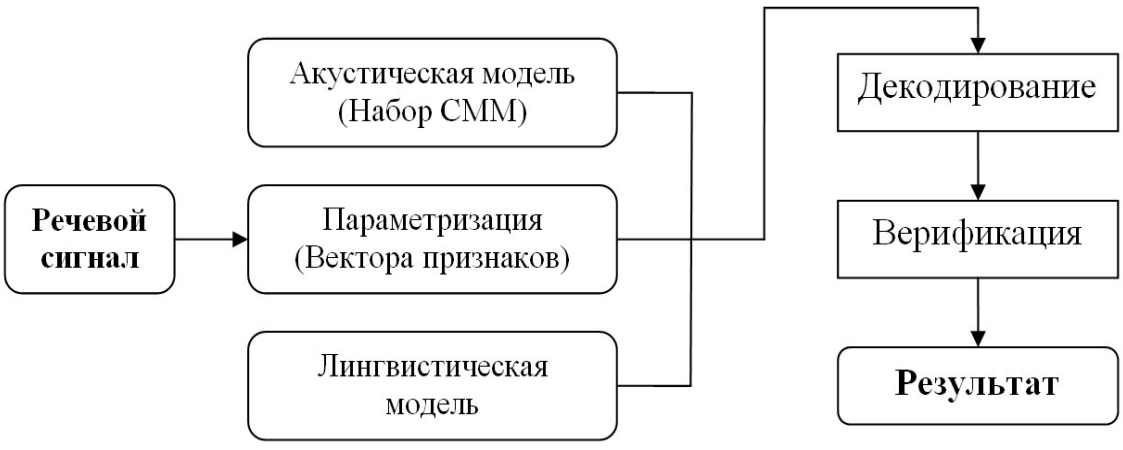


Рисунок 2.1. – Общая схема поиска ключевых слов

Для решения вопроса автоматического распознавания ТЯ (таджикского языка) и анализа отдельных слов является процесс моделирования и алгоритмизации процедуры показа специальных глагольных конструкций (СГК).

Информационные основы ИТ-схема глагольных парадигм и набор структурных типов глаголов таджикского языка.

Обратите внимание, что в результате этого процесса могут использоваться пробелы в таджикских словоформах. В работе под глагольной структурой понимается особая последовательность словоформ в неразрывной форме, представляющая собой структуру сказуемого-глагола или связь с другими частями речи. В работе авторы ограничиваются распознованием простых и сложных глаголов. Проводят исследование структурных типов глаголов в которых выделяются[1]:

- наличие в структуре глагола словоформы, из простых глаголов (V), причастия (Part), существительного, прилагательного и местоимения (H) и приставки (Pr);

- ограничении глагола шестью словоформами;

- отсутствие знаки препинания между словоформами глагола;

- наличие предлога только в первой словоформе;

- последняя словоформа является простым глаголом

Таджикское слово разделяется на три части - префикс, корень и постфикс, которые называются морфемами. Порядка деления слов на корни и аффиксы составляет основу процесса автоматического распознавания в модели GL, а также учитывает информацию о речевой компоненты слове. (POS-тегер, частичный речевой тег) [19, 69].

В таджикском языке простые глаголы имеют только один корень. Глаголы в таджиксуом языке добалением дабавление префиксов и суфиксов выражаются в настоящем или будушщем времени. Процесс узнавания простого глагола происходит через лексические морфы.

[1] Собиров Д.Д, Гращенко Л.А., Усманов З.Д. – Изв. АН РТ. Отд. физ.-мат., хим., геол. и техн. Н., 2011, №3, с. 41-46.

Если, основной корень глаголов образованы из других глаголов, он считается прошедшем времени (19 и 3). Глаголы могут иметь различные формы прошедшего времени. Для упрощения процедуры расчётов пред морфологическим анализом используется POS-теггер.

В работе [66] на основе автоматической фильтрации проведен экспериентальный анализ 80 000 словарей предыдущих экспертов и показано, что: 746 слов являются простыми инфинитивными глаголами; 398 именными глаголами; 348 неспецифическими и 214 инфинитивными, отличающими от прошедшего времени глаголами. На основании этого был составлен подходящий список устного общения. Множество {-", -а, -ағӣ, - ағист} может быть объединено на основе настоящего и будущего времени для целей спряжения. За множеством следует личный суффикс или короткий суффикс. В сочетании их комбинация приводит к множеству {", -ам, -ӣ, -ад, -ем, -ед, -и, -аст}. Ожидаемое количество исправлений - 31. Исходя из постфиксов -ид или -онид, в форме производных глаголов можно рассматривать широкий спектр постфиксов от 95. Система глагольных префиксов простая 13, сложная 35, всего 48.

Распознавание ключевые слова гораздо сложнее и вероятнее, потому что закономерности взаимосвязаны.

Теггер POS имеет большого количества обращений на базу данных (словарь) по количеству слов в предложении правильнее уменьшать количество его употребления.

Разделение предложений, собственные имена, аббревиатуры, наречия, союзы и причастия проводились с целю создания кумулятивной лексики, с целью создания основы данных из гравиметрического анализатора. Внутрисегментный анализ реестра проводился на длину, не превышающую минимальное значение, а внутрисегментный анализ регистра по длине максимального значения выборки ГК (6), но не превышающего минимального значения показателей длины сегмента.

Суть использованного метода поиска, по ключевым словам, заключается в том, что при последовательном применении процедур фильтрации с постепенно

возрастающей вычислительной сложностью удаляются элементы ключевых слов, не имеющих определяющие роли. Применяются правила используемые при графматическом анализе информации. Осуществляется процесс поиска простых глаголов и определения словоформ перед этими глаголами. Последний шаг, наиболее затратный в вычислительном отношении, — это процесс использования правил, учитывающих морфологические особенности последовательности.

Представляем следующий алгоритм поиска таджикских глаголов в предложении «*Ҳалима метавонист хушбахт бошад, валекин ҳама вақт барои бадбахтиҳои ӯ баҳона меёфт*» (анализ приведен на рис. 2.2):

1. Знаки препинания использованы как разделители главное предложения. (шаг 1 – Графематический анализ на рис. 2.2) Исключением из анализа: слова с большой буквы, наречия, аббревиатуры, союзы, наречия, местоимения, производные части предложения делятся на подгруппы, (шаг 2 на рис. 2.2), составляются фразеологизмы (из стоп-списка) по характеристикам, полученным на графическом этапе анализа. В этом предложение исключаются из рассмотрения союз - «*валекин*», предлог - «*барои*» и местоимения - «*ҳама*», «*ӯ*».

Отметим, что словосочетание «*бадбахтиҳо ӯ*» в таджикском языке эквивалентно словоформе «*бадбахтиҳояш*», следовательно при реализации данного этапа необходимо руководствоваться внутренними языковыми нормами.

2. В каждом фрейме в строгом порядке, с конца к началу, выполняется процедура POS-теггера для поиска простого глагола. Если глагол не простой, то он считается частью, не имеющей ГК (шаг 3 на рис. 2.2).

3. В части с простым глаголом, чтобы глубина части не превышала пяти, части речи отделяются перед простым глаголом (этап 4. рис. 2.2). Если часть речи следующей словоформы не соответствует дереву решений, весь процесс останавливается (рис. 2.2). В результате этого шага будет составлен список правильных форм слов, отсортированный по длине.

4. При наличии более одной записи для возможных ГК необходимо их сократить на основе дополнительных правил для информации о запрещенных аффиксах и вспомогательных глаголах (шаг 5, рис. 2.2).

5. Запись с максимальным длиной остается на целевом ГК после процесса диссимиляции.

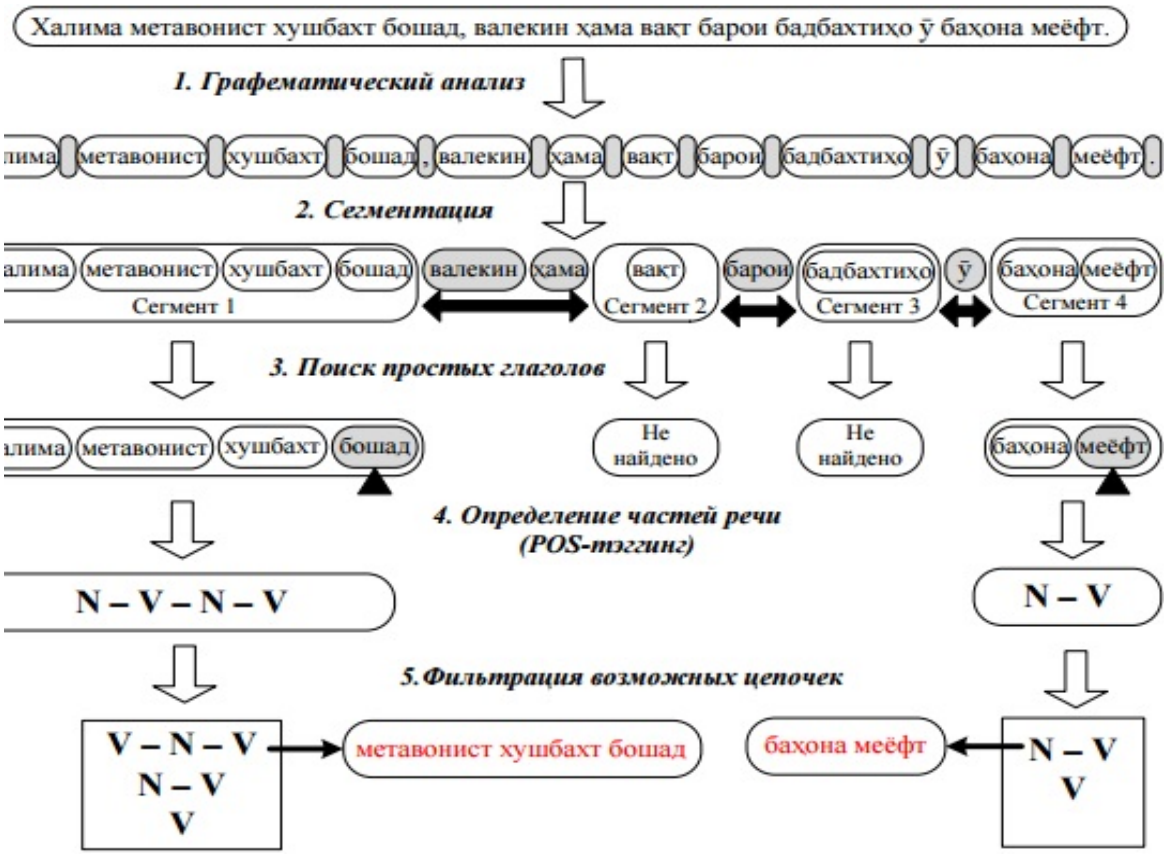


Рисунок 2.2. – Порядок распознавания слов в таджикском языке

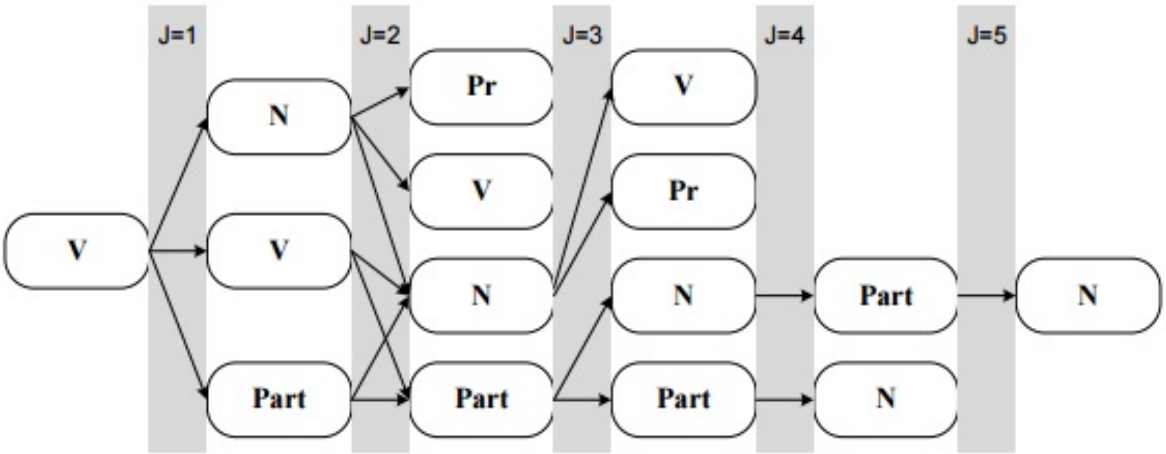


Рисунок 2.3. - Дерево решений

Настоящий алгоритм составлен с целью определения частоты встречаемости различных типов глагольных сочетаний в таджикском языке и уточнения правил процесса фильтрации ложных ГК в виде проблемно-ориентированного программного стенда. После завершения наладки данного этапа алгоритм и база данных будут доступны практическому использованию.

2.3. Модель представления речевого сигнала в системе

Нельзя не отметить актуальность разработки адекватных аналитических моделей речевого сигнала для слитной речи. Задачи, стоящие перед специалистами, определяют большой интерес к разработке систем автоматизированной и автоматической обработки речевых сообщений достаточно эффективных для обеспечения возложенных на них задач. Реализация таких систем в свою очередь невозможна без эффективного моделирования речи, учитывающего все ее нюансы. В рамках научного исследования была разработана аналитическая модель речевого сигнала, которая представлена на рисунках 2.5-2.7.

В ходе исследования ведущих достижений в области распознавания речи [40], был выявлен ряд недостатков. Их анализ позволил выдвинуть гипотезу:

1) Речевой сигнал образуется неизвестным преобразованием, в общем случае нелинейным, (смесью) двух сигналов: семантическим (смысловым) и шумовым (индивидуальным, определенным особенностями речеобразующего тракта диктора).

2) Применив некоторое базисное преобразование речевого сигнала, переводящее нелинейное соотношение семантической и шумовой составляющей в их суперпозицию, становится возможным применение методов калмановской фильтрации к процедуре разделения этих составляющих (выделению семантики), что и является целью распознавания речи.

Рисунок 2.5. – Блок-схема формирования базы индивидуальных шумов

Рисунок 2.6. – Блок-схема формирования семантического сигнала

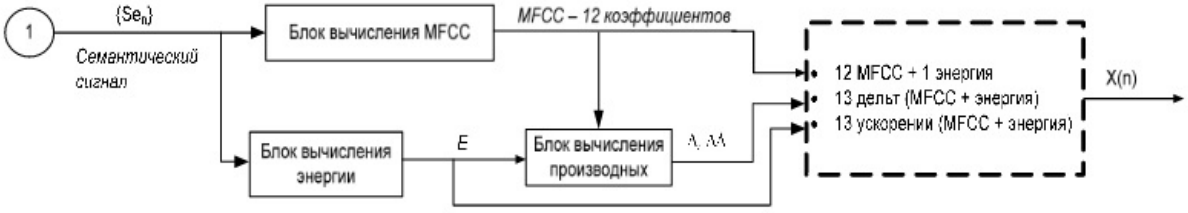


Рисунок 2.7. – Блок-схема извлечения информативных признаков семантического сигнала

Таким образом, предлагается использовать сингулярное разложение матриц параметров обучающих выборок в качестве базисного преобразования речевого сигнала. В качестве информативных параметров речевого сигнала были выбраны его частотные характеристики: результаты дискретно-косинусного преобразования, Мел-коэффициенты и их производные, как наиболее эффективные с позиции соотношения скорости обработки к качеству

распознавания [94]. В итоге аналитическая модель речевого сигнала представляется в следующем виде:

Предварительная обработка речевого сигнала:

исходный речевой сигнал в дискретном виде (1); предыскажение (2); вектор (3); выделение пауз (4); сингулярное разложение (5), где (6) – семантический сигнал, (7) – индивидуальный шум диктора, (8) - сигнал коррекции.

1. S_n ,$0 \le n < N_x$,
2. $S_n = S_n - 0.95 * S_n - 1, 0 \le n < N_x$,
3. $\{\bar{S}\} = \zeta(S_n) = \bar{S_0}, \bar{S_1}, \bar{S_2}, \ldots, \bar{S_k},$
4. $\bar{S_i} = (S_0, S_1, \ldots, S_m), 0 \le i \le k$
5. $\overline{S'_n} = \phi(Se_n, N_n, C_n),$
6. $Se_n = f(\bar{S}'_n, C_n)$
7. $N_n = k(\bar{S}'_n, C_n)$
8. $C_n = u(Se_n, N_n)$

Процесс формирования базы индивидуальных характеристик дикторов (шумов): (1), если (2):

1. $\Delta U = \lambda\_noise(Se_n, N_n, C_n)$
2. $\Delta U \le U_{nopoz} \implies \{N_i\} = g\_noise(N_n)$

Процесс выделения семантического сигнала:

$N'_n = g\_sem(\bar{S}'_n),$

$N_n = h\_sem(N'_n, \{N_i\}),$

$Se_n = d\_sem(\bar{S}'_n, N_n).$

Процесс вычисления информативных признаков речевого сигнала и исходный семантический сигнал в дискретном виде:

$x[n], 0 \le n < N;$

$E = \log \sum_{n=0}^{N-1} x^2[n];$

Преобразование Фурье [76]

$X[k] = \sum_{n=0}^{N-1} x[n]\, e^{\frac{-2\pi i}{n} kn}, 0 \le k < N.$

Гребенка фильтров, на основе оконной функции [120]

$$H[m,k] = \begin{cases} 0, \text{если} k < f[m-1], \\ \dfrac{k - f[m-1]}{f[m] - f[m-1]}, \text{елсе} f[m-1] \le k \le f[m], \\ \dfrac{f[m-1] - k}{f[m+1] - f[m]}, \text{если} f[m] \le k \le f[m+1] \\ 0, \text{если k} > \text{f[m} + 1], \end{cases}$$

$\sum_{m=1}^{M} H[m,k] = 1,\ 0 \le k < N$, M – количество вильтров,

$f(m) = \left(\frac{N}{F_s}\right) B^{-1}\left(B(f_1) + m\frac{B(f_h) - B(f_i)}{M+I}\right)$,

$B(f) = 1125ln\left(1 + \frac{f}{700}\right), B^{-1}(b) = 700\left(e^{\frac{b}{1125}} - 1\right)$,

Отфильтрованный речевой сигнал:

$S[m] = ln\{\sum_{k=0}^{N-1} |X[k]|^2 H[m,k]\}, 1 \le m \le M.$

Дискретно-косинусное преобразование:

$c(n) = \sum_{m=0}^{M-1} S[m] \cos\frac{\pi n(m-0.5)}{M}, 1 \le n \le M.$

Мел-кепстральные коэфициенты:

$C = \{c(n)\}, 1 \le n \le M,$

$\Delta(n) = \frac{d(n+1) - d(n-1)}{2}, 1 \le n \le M,$

$\Delta\Delta(n) = \Delta(\Delta(n)), 1 \le n \le M.$

Таким образом, в результате процедуры параметрического отображения речевой сигнал трансформируется в последовательность признаковых векторов, которая состоит из следующих параметров:

$$\{MFCC, E, \Delta MFCC, \Delta E, \Delta\Delta MFCC,\ \Delta\Delta E\}.$$

Оптимальным значением числа Мел-коэффициентов является 12, а вектор информативных признаков речевого сигнала будет состоять из 39 параметров.

Необходимо отметить, что в предложенной модели для систем автоматической обработки слитной речи использован новый подход к представлению слитной речи. Он основан на моделировании речевого сигнала как смеси семантической и индивидуальной (шумовой) составляющих. В общем случае проведенные исследования по анализу и аппроксимации данного соотношения известными функционалами, а также анализ моделей речевого сигнала, опубликованных специалистами в данной области, показали на сложную нелинейную зависимость. Однако применение сингулярного разложения определенного представления речевого сигнала позволяет, с некоторыми допущениями, представить смесь семантического и шумового сигналов как линейную, и, тем самым, обосновать корректность применения методов Калмановской фильтрации, для выделения семантической составляющей.

В качестве направления дальнейших исследований представляется расширение языковых моделей; поиск оптимального, с точки зрения линейности аппроксимирования зависимости семантической и шумовой составляющих, представления речи в виде признакового вектора; типологизация шумовых составляющих речи, с целью построения высокоэффективных систем автоматической аутентификации диктора.

В работе [4] автор рассматривает общий вид структуры распознавания речи через модели представления речевого сигнала и включения метода формирования акустических единиц речевого потока: непараметрические (методы основаны от формальных грамматик и метрик на множестве речевых сигналов) и параметрические (вероятностные, на основание скрытых Марковских моделей, нейросетевые).

Акустический механизм речеобразования, является универсальным и устанавливает правила формирования речи независимо от языка.

На акустическом этапе исследования слитной звуковой речи реализуется спектральный (микроанализ) структура звука. Имеются возможности фонетического сопоставления и определения как внутриязыкового, так и

межязыкового фонетического заимствования. Для проведения сравнительного анализа уровня речевого материала служебного использования голосов, на каждый гласный звук отбирается три (не менее 15). Сопоставление гласных должны проводится при одним и тем же шумовом фоне. Русские слова могут быть использованы в спектральном микроанализе, если они записаны в эталоне образце спектров и фонограмме.

Для гласных раздельного произношения в таджикском языке установлены следующие две формы возрастания частотных периодов:

Таблица 2.3 – Частоты по форматам

Гласный звук	**Значение F1 (Гц)**	**F2 (Гц)**
/и/	328	2275
/е/	429	2107
/а/	656	1277
/о/	443	808
/у/	309	729
N	428	1162

Порог (изменчивость) формирующего отношения определяется следующим образом:

$$Fn \pm Пn/Fk \pm Пk = Fn/Fk \pm Пk \pm Пn/Fk \pm Пk$$

где, Пn и Пk, пороги соответствующих формат Fn и Fk. Выражение: Пn/Fk±Пk определяет порог отношения формат.

Отсюда, предельное значения изменчивости соотношения форматов Fn/Fk определяются средним значениям Пn/Fk. Аартикуляционные особенности согласных звуков нализируются по закономерностям перехода форматного спектра, которые формируются в результате коартикуляционного эффекта, формирующегося в результате последовательного воздействия жестов.

Скорость изменения частоты основного тона, можно оценить увеличением или уменьшением RF в Герцах (ГЦ) за один мс (10^{-3} с). Величина скорости

вычисляется со знаком «+» или «-» в Гц/мс, при увеличение или уменьшение частоты основного фона, соответственно. Важным моментом определения скорости изменения основного тона, является удачный подбор материала. Здесь, дольжно учитываться уровень словесного ударения и эмоциональность гворящего. Все эти особенности речи должны соответствовать фонетической позицией говорящего.

Другой особенностью можно сказать сложность акустического анализа эмоционального состава разговорной речи заключается в том, что изменение эмоционального состава речи происходить за макроскопически большие масштабы времени, длительность фонограмм не могут их правильно отразить. Акустический анализ речи в нашей работе имеет 105 признаков, в связи с этим при акустическом анализе системы каждому индивидууму речи присваиваются конкретные индивидуальные признаки речи (I_1, I_2...I_{105}).

Пусть, признак P_k^1 проявляется у половины населения, т.е. вероятность появления этого признака p (I_k) равна 0,5. Распространяя это предположение на все символы с учётом свойства вероятности приходим к теореме умножения вероятностей:

P = p(I1) * p(I2) * p(I3) *…*p(I105) = $0{,}5^{105}$ приблизительно равно $2{,}46 * 10^{32}$

2.4. Разработка метода поиска ключевых слов

Разработка методов поиска ключевых слов в слитной речи является актуальным процессом, поскольку в последнее время число задач, которые должны быть решены на основании таких методов, многократно увеличилось. Для того, чтобы обработать речевой сигнал, требуется получить вектор его признаков. Вектор признаков должен состоять из 39 коэффициентов (13 кепстральных, 13 дельта и 13 перегрузочных коэффициентов). Размер окна, определяющего один вектор признаков, равен 25 миллисекундам. К данному окну применима весовая функция Хэмминга с порядком фильтра 26. Дистанция

между окнами выборок в отношении соседних векторов признаков составляет 10 миллисекунд.

На начальном этапе первым шагом является создание гауссовского монофона СММ использованием прототипа модели и набора векторов наблюдения. Каждый СММ имеет пять состояний. Первое и последнее состояние – это модель молчания. Каждое состояние представлено одним гауссовым, состоящим из среднего, дисперсионного и смешанного веса. Для системы фонем может быть применена левая-правая типология трёх состояний с гауссовым распределением вероятностей матрицы наблюдения. Для каждого монофона создается СММ. Монофоническая модель скомпилирована для компиляции всех словарных слов в виде определенной последовательности монофонов.

Трифон СММ построены путем преобразования монофонных транскрипций в трифон транскрипций. Набор моделей трифона создается путем копирования монофонов и переоценки. В модели связанного состояния трифона (TCS) трифоны с идентичными векторами могут быть заменены на общее акустическое состояние.

Таким образом, для каждой модели при обучении СММ векторы признаков для каждого акустического состояния уточняются.

Соколов А. Н. [44] в своей статье "Внутренняя речь и понимание" инициировал первую теоретическую попытку решить проблему извлечения ключевых слов ("опорных", " обобщающих").

В работах Сахарного Л.В., Сиротко-Сибирского С. [57] и Штерна А.С. дается интересная и современная интерпретация ключевых слов. По их определение: - «Порядок ключевых слов в наборе ключевых слов, является несжатая тематика текста, одна из минимальных вариаций текста, и этот тип «текста» имеет целостность «оригинала» с максимальной связью» [11].

Понятие набора ключевых слов, репрезентирующих весь текст, впервые стало понятием «первичный текст» в психолингвистических исследованиях.

Официальных стандартов длины или грамматической точности исходного текста не существует. В ее основе лежат задания, учитывающие «человеческий фактор», создающий или понимающий текст и коммуникативную ситуацию. В данном подходе под ключевым словом понимается слово или словосочетание, имеющее относительно него смысловую нагрузку в расширенном тексте.

«Ключевое слово — это слово, взятое из текста, которое вместе с другими ключевыми словами может представлять текст. Примером поиска по документу является набор ключевых слов в документе. Набор ключевых слов близок к аннотации, плану и абстракции, которые также представляют документ с меньшей детализацией, но не имеют синтаксической структуры».

Ключевые слова могут быть получены лингвистическими и вычислительными методами для оказания помощи в поиске информации из ключевых слов больших данных.

Ключевые слова имеют разное значение в разных контекстах.

- слово, взятое из текста, который в сочетании с другими ключевыми словами может представлять текст.
- словарная группа из одной части и нескольких частей, отражающая содержание документа [78];
- некоторые слова из текста, представляющие собой наиболее важные слова для группировки и поиска статей;
- важные термины документа используются для объяснения содержания документа читателю на высоком уровне;
- важные термины в документе используются для разъяснения содержания документа читателю на высоком уровне;
- случайные слова и словосочетания в документах, рассматриваемых как образец (вопрос) в рамках общего набора документов;
- слова, наиболее важные для решения задач в инструкции.

Для того чтобы определить ключевые слова в текстах, необходимо пройти следующие этапы. Обработайте текст и удалите элементы кода, слова, удалите

строгие слова, поставьте слова в виде словаря (приставки, союзы, причастия, местоимения, суффиксы и т.д.). Далее, фильтруя слово-кандидат по требованиям к ключевому слову обрабатывая характерные особенности каждого кандидата, составили набор слов кандидатов. Ключевое слово выбирался из числа слов кандидатов.

В [78] представлен универсальный алгоритм извлечения ключевых слов. Здесь, на основе метода извлечения слов фильтруются кандидаты в ключевые слова Статистическая фильтрация включает в себя выбор определенного количества реалистичных лексем.

К. Д. Мэннинг [35] для филтрации кандидатов в кллючевые слова использует расчет веса их информативности. Такой подход дает восможность оценить относительную важность слов, заметим, что появляется известная метрика TF-IDF.

Н.Е. Ефремова и др. [25] говорят о *лингвистических* и *статистических* критериях выбора ключевых слов. В настоящее время используются эти две модели и метода. По мнению автора, наилучшее качество достигается за счет сочетания лингвистического и статистического подходов.

В связи с увеличением объема и сложности информации извлечение ключевых слов из текста становится проблемой. Для автоматизации извлечения ключевых слов, позволяющей своевременно и адекватно обрабатывать тексты, разработаны автоматические инструменты извлечения ключевых слов.

В работе [31, 11] для поиска ключевых слов использован алгоритм быстрого преобразования Фурье и теории спектрального анализа.

Для извлечения ключевых слов и фраз из словосочетаний используются следующие методы: лингвистический, статистический, спектральный и гибридный. Рассмотрим подробно каждый из этих методов.

Существуют многочисленные попытки создания универсального лингвистического метода выделения ключевых слов, но заметные успехи не были достигнуты.

Данные онтологии и семантики слова используются в качестве основы лингвистических методов. Они происходят от лингвистических знаний/особенностей. Они используют лингвистические особенности слов, прежде всего предложения и документы. Лингвистический подход включает лексический анализ, синтаксический анализ, дискурсный анализ и другие. Эти подходы требуют значительных вычислительных затрат. Они требуют специальных знаний языка и домена. К тому же на начальном этапе эти методы требуют очень длительного времени. Разработка онтологий очень трудоёмкий процесс. Особенно, при ручном лингвистическом анализе допускаются ошибки, которые осложняют процесс анализ документов. Реальный возможность решения задачи — это программное обеспечение и автоматизации процесса анализа текста документов [9].

Автоматический метод извлечения ключевых терминов из текстовых документов, представлен и в работе М. Гринева [21]. Здесь, подход основан на алгоритме Гривана-Ньюмана с использованием базы данных Википедии, который измеряет семантическое сходство терминов. По мнению автора в этом подходе анализ проводится непосредственно в базе данных, правильнее и полнее извлекать основных терминов из текста.

При обработки естественного языка, широко используются и графические лингвистические методы анализа, которые также являются один из разнообразии существующих методов и алгоритмов. Здесь, ключевые слова извлекаются методами обработки графов. Подбор ключевых слов проводятся на основе статистических параметров, морфологического, синтаксического или семантического анализ [78].

В работе [30] автор использует основанные на маркерном анализе лингвистические методы. Такой подход как правила используется для установления связей между интуитивно определенными словами и символами. Это позволит в обучении о том, как автоматически выделить ключевые слова. Предлагаемая методология достигается за счет использования имеющегося

программного обеспечения. Можно сделать вывод, что предложенный алгоритм вполне пригоден для централизованного анализа научных текстов.

В [25]. на основе экспериментального исследования, предложено способы автоматически на основе лексико-синтаксических моделей языка LSPL, определить термины в текстах.

Статистические методы, основанные на извлечении ключевых слов, считаются простыми и не требуют изучения информации в тексте или документах. В основе статистического метода лежит частотность слов и языковых структур. По мнению С.О. Шереметевой, преимуществом статистических методов является то, что этот метод может использовать разные алгоритмы для извлечения ключевых слов. Метод не требует трудоемких процедур по созданию языковой базы знаний и прост в применении. Результаты, полученные статистическим методом, не всегда удовлетворительны. Следует также отметить, что этот метод эффективно используется в медленных морфологических языках. Таким образом, появляются задания для естественных языков с богатой морфологией.

Метрика TF-IDF является классическим статистическим подходом и широко используется для расчета информационных весов и оценки важности слов в текстовых документах . (Количество слов, используемых в документе, и частота встречаемости этого слова в других документах считается весом слова.) Имеются расширенные версии модели TFIDF, - Okapi, BM25. Вовремя проведении расчётов данные не должны изменятся, при проведении расчётов в реальном времени, вычисления осложняются.

Существующие статистические методы определения состава слов с устойчивым уровнем в работах В.П. Захаров [28] и М.В. Хохловой [73]. Самый простой способ определить общность в тексте — это составить список частот слов, словоформ и сочетаний их частот.

Существуют такие статистические показатели, как MI, t- балл, Log-Likelihood, логарифмическая вероятность, z-балл и т.д., которые используются для измерения силы ассоциации. Эти меры служат устройством для связывания

случайных и условных источников слов. Для определения согласованности этих показателей используются методы маематической статистические.

Существуют различные подходы статистического определения структуры терминов, например, искать фразы из n слов на основе заданных частотных характеристик, в том числе по значениям абсолютных или относительных частот определенных словосочетаний.

О.Н. Камшилова [29] на основа метода координатного индексирования, где содержание документа представляется через набора ключевых слов, предложила способ выделения ключевых слов информационного значения Для определения частоты используется термин «плотность ключевых слов», который выражается в процентах.

В работе [12] П.И. Брославский анализирует 5 способов автоматического выбора структур произвольной длины. Здесь, словари, тезаурусы и другие семантические ресурсы не используются, определение структура и содержание терминов проводиться на этапе отбора кандидатов. по определенной теме. Подход среди специалистов получил название «с чистого листа».

Ю. Малинина [34] для улучшения извлечения терминов и получения более согласованных результатов (уменьшение информационного шума), считает необходимым выполнение следующих условий: лингвистическое исследование семантических отношений терминов; условия давности терминологических единиц в конкретном регионе и в данном тексте; сочетания статистических и лингвистических методов, и использование более одной стратегии. Считается полезным составление общей таблицы для сопоставления и оценки качества извлеченных терминов.

Спектр документа представляет собой набор пар основной формы слов или словосочетания в тексте. Из спектрограммы текста можно легко переходить в простую компьютерную программу. В дальнейшие проводятся работы по подборе ключевых слов и анализе спектрограмм различных текстов.

В работе [31]. на основе алгоритмов мгновенного преобразования Фурье и теории спектрального анализа предложено спектральный метод поиска

ключевых слов полнотекстового документа. Исследование спектрограмм текстов, по теме диссертации, составляет основного содержания дальнейших работ

Спектральный анализ на основе использования преобразование Фурье и «карт усреднения для поиска ключевых слов в текстовых файлах был разработан и в работе К. Я. Кудрявцевой [31].

При реализации процедуры быстрого поиска ключевых слов из текстовых документов спектральный метод показывает более лучшие результаты чем другие методы. Здесь, нет необходимость в проведении сложных процедур лексико-морфологического анализа, создания и использования сложных указателей, словари, поисковых систем B+, суффикс или GiST.

С помощью вейвлет-преобразования получаются аналоги частотно-временного выражения. В этом методе использованием текстовых спектрограмм качественно записывается содержание текстового файла. Описан спектральный алгоритм поиска ключевых слов.

По мнение С.О. Шереметева [78] основных методов и моделей автоматического выделения ключевых слов можно разделить на две группы: на статистические и на гибридные методы. Разница в методах, по его мнение, определяется порядком обработки текста на каждом этапе процесса и объемом лингвистических знаний для этого. По слова С.О. Шереметева: - «Большие возможности имеют гибридные методы, в которых статистические методы обработки документов дополняются одной или несколькими лингвистическими процедурами и лингвистической базой знаний разной глубины».

О результатах различных исследований в области гибридных лингвистических и статистических методов см. также С.В. Попова.

Как видно, анализы показывают, что последние годы учеными и специалистами разработаны различные методы автоматической обработки текста, извлечения ключевых слов и словосочетание. Разработка эффективных инструментов – программ, является один из основных направлений компьютерной обработки текста.

Программа NTC может быть использована в качестве основы для реализации всех разработанных систем. Проведен эксперимент по распознаванию цифровых последовательностей в непрерывной речи с целью изучения характеристик методов поиска ключевых слов. В качестве критерия надежности было использовано значение апостериорной вероятности существования гипотезы признанного слова. Если слово превышает значение, то слово считается распознанным правильно. В противном случае слово неверно. Во время обработки результатов правильно распознанные слова должны фиксироваться их именами, взятыми из словаря. Неправильные исправлены как "UNKNOWN".

Транскрипция тестовых файлов сравнивается с транскрипциями результатов непрерывного распознавания речи для определения эффективности процесса поиска ключевых слов.

Языковая модель разработанного системного комплекса априори дана и является циклическим графом, состоящим из чисел от 0 до 9. Словарные основы для разработанного системного комплекса были построены на основе латинских букв.

Группы акустических состояний для разработанного системного комплекса указаны в таблице 2.4.

Таблица 2.4. Транскрипция слова из фонемы

Слово	**Транскрипция слова из фонемы**	**Новые фонемы**
Як	y |a| k	y’. a. k
Ду	d |u|	d.u
Се	s |e|	s.e’
Чор	ch’|o| r’	ch’.r
Панҷ	p’ |a| n |j|’	p’.j
Шаш	sh’ |a| sh’	sh’.sh’
Ҳафт	h’|a| |f| t’	h’t
Ҳашт	h|a|sh’|t’	h’sh’t
Нӯҳ	n’|u’|h| t’	n’h’
Сифр	s |i| |f| r’	s’fr’
Общее количество акустических состояний		**22**

В базе данных зарегистрировано 20 диктора (8 женщин и 12 мужчин). Паузы были добавлены в начале и конце каждого элемента речи, длительностью 0,3 секунды. Система обучалась с использованием последовательности чисел или отдельных элементов. Транскрипция каждого элемента известна. Тестовыми данными являются отдельные числа или их последовательности в присутствии дополнительной спонтанной речи. Транскрипция слов для тестовых данных устанавливается путем записи их имен, взятых из словаря, а остальные слова фиксируются как " UNKNOWN".

2.5. Выводы по главе

Таким образом, в предложенной модели для систем автоматической обработки слитной речи использован новый подход к представлению слитной речи. Он основан на моделировании речевого сигнала как смеси семантической и индивидуальной (шумовой) составляющих. Однако применение сингулярного разложения определенного представления речевого сигнала позволяет с некоторыми допущениями, представить смесь семантического и шумового сигналов как линейную, и, тем самым, обосновать корректность применения методов Калмановской фильтрации, для выделения семантической составляющей.

В области информатики активно используется понятие ключевых слов как носителя важной информации о тексте, особенно в задачах информационного поиска. Шаблон поиска документа задается как набор ключевых слов. Близко к сведению, набросок, резюме.

Для того чтобы определить ключевые слова в текстах, необходимо пройти следующие этапы. Обработайте текст и удалите элементы кода, слова, удалите строгие слова, поставьте слова в виде словаря (приставки, союзы, части, местоимения, суффиксы и т.д.). Подбираем слова для ключевых слов-кандидатов. Мы фильтруем каждое слово-кандидат по ключевому слову и

анализируем важные особенности каждого кандидата. Ключевое слово выбирается из числа кандидатов.

В таджикском языке в отличия от русского языка, в первую очередь, в количестве букв в алфавите, шесть из них не имеют аналога в русском языке. Для распознавания речи имеет значение и тот факт, что отсутствие в таджикском языке родов и падежных окончаний затрудняет распознавание речи и ее воспроизведение, поскольку при распознавании необходимо опираться только на контекст, из которого должно следовать, о каком объекте идет речь – женского или мужского рода (в соответствии с русским языком).

Интерес представляет также (вне зависимости от языка) темперамент говорящего.

Современные распознаватели речи состоят из двух блоков: блок лингвистической обработки текста, при помощи которого происходит построение полной фонетической транскрипции синтезируемого текста, и блок акустического синтеза, при помощи которого генерируется речевой сигнал, если имеется в виду еще и синтез речи.

Таким образом, при акустическом анализе речи учитывают 105 признаков, т. е. с точки зрения акустики речь каждого человека характеризуется системой признаков (I1, I2...I105). Разработка методов поиска ключевых слов в слитной речи является актуальным процессом, поскольку в последнее время число задач, которые должны быть решены на основании таких методов, многократно увеличилось. Для того, чтобы обработать речевой сигнал, требуется получить вектор его признаков. Вектор признаков должен состоять из 39 коэффициентов (13 кепстральных, 13 дельта и 13 перегрузочных коэффициентов). Размер окна, определяющего один вектор признаков, равен 25 миллисекундам. К данному окну применима весовая функция Хэмминга с порядком фильтра 26. Дистанция между окнами выборок в отношении соседних векторов признаков составляет 10 миллисекунд.

ГЛАВА 3. РАЗРАБОТКА И ИССЛЕДОВАНИЕ АЛГОРИТМОВ РЕАЛИЗАЦИИ МЕТОДА

3.1. Разработка способа расчёта параметров вероятностной графической модели (скрытая Марковская модель, условные случайные поля)

В начале 1990-х годов существование двух всеобъемлющих подходов привело исследователей к идее объединения HMM и ANN в новой модели – гибридной модели HMM/ANN. Эта модель позволяет более эффективно интегрировать положительные аспекты Марковских моделей и нейронных сетей. Таким образом, HMM позволяет моделировать долговременные зависимости и нейронные сети – универсальную непараметрическую аппроксимацию, используя вероятностную оценку, алгоритмы дискриминантного обучения, уменьшая количество оцениваемых параметров в стандартных случаях HMM.

Для того чтобы использовать HMM в системе распознавания речи, важно послать речевой сигнал со следующих позиций:

- последующие наблюдения статистически независимы. В результате вероятность последовательности является результатом вероятностей конкретных наблюдений. Несмотря на то, что речь является динамическим процессом, она моделируется с использованием последовательности наблюдаемых векторов, которые частично являются стационарными процессами;
- это настоящая гипотеза Маркова. Другими словами, это вероятностная гипотеза в определенном состоянии времени t, в зависимости от случая, когда процесс был на момент t - 1.

Применение графических моделей в разработке представляется следующим образом.

Последовательная временная классификация используется для решения когнитивных задач распознования речи.

Математическим образом признаки в речи можно реализовать по следующем формуле:

$$\bar{x} = \{x_1, x_2, \dots, x_t \dots, x_T\}, x_t \in R^d. \quad (3.1)$$

где T – длина последовательности, d – размерность пространства признаков

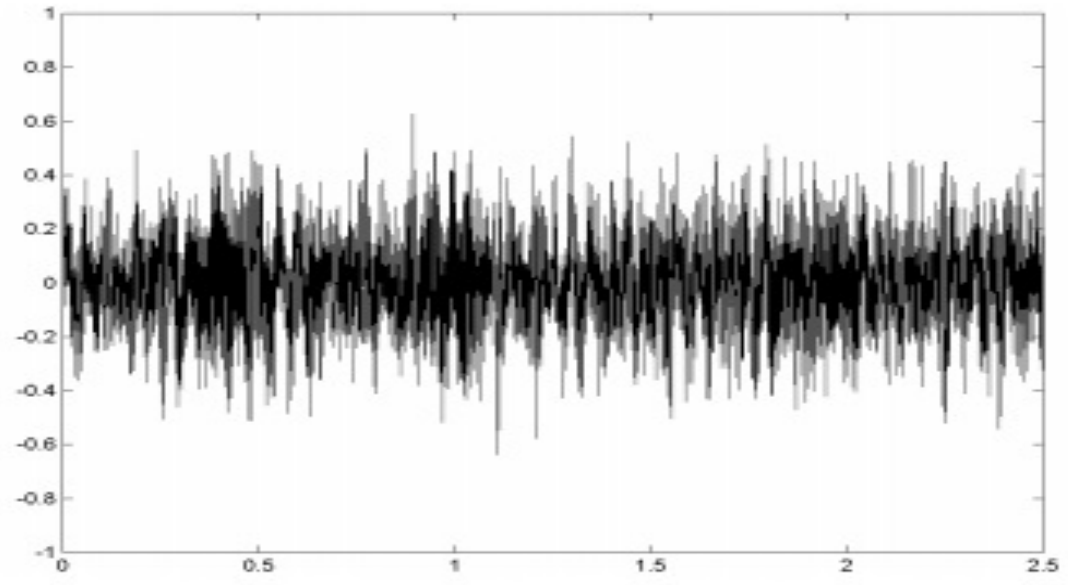


Рисунок 3.1. - Объект классификации

Для имитации временной последовательности вводится дополнительное дискретное случайное значение h (скрытый режим):

$$p(\bar{h},\ \bar{x}) = \prod_{t-1}^{T} p(\ h_t|h_{t-1})\mathrm{p}(x_t|h_t). \quad (3.2)$$

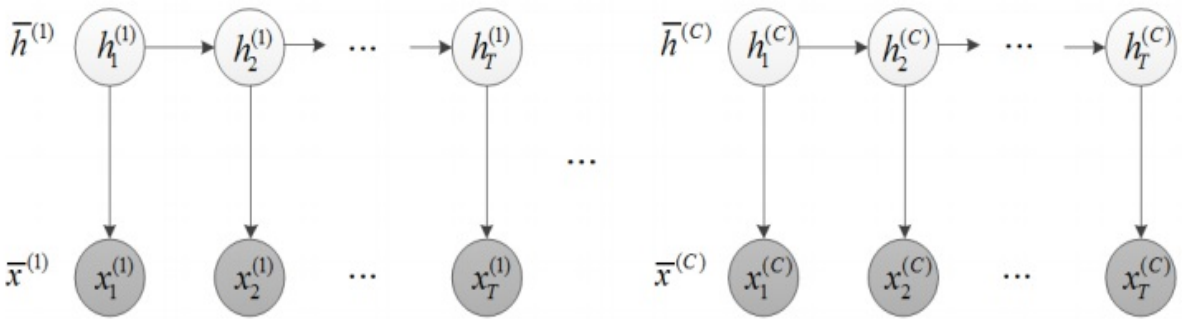


Рисунок 3.2. - Графическое представление Скрытой Марковской Модели

Однако, несмотря на широкое распространение, HMM имеет свои недостатки:

- необходимо выбрать плотность вероятности вида fh(x);

- необходимость зависимости от компонент вектора свойства (в случае Гауссовой плотности вероятности наличие линейной зависимости между компонентами приводит к деградации ковариационной матрицы | Σ | = 0).

Плотность многостороннего распределения Гаусса:

$$f_h(x) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^{\top}\Sigma^{-1}(x-\mu)}, x, \mu \in R^d. \quad (3.3)$$

Некоторые недостатки Скрытой марковской модели были устранены в предложенное A. Gunawardana и др. дискриминантной модели «Условные случайные поля со скрытым состоянием», (Hidden Conditional Random Fields - HCRF), в 2005 году в работе “Hidden conditional random fields for phone classification”.

Недостатки Hidden Conditional Random Fields:

- целевая функция оптимизации не является выпуклой;

- дискриминантной модели требуется больше данных для исследования, чем модели производителя, чтобы получить минимальную асимптотическую ошибку классификации.

Цель: разработка обобщенной графической модели, обеспечивающей хорошее качество классификации при небольшом объеме данных исследования и не требующей оценки параметров плотности вероятности.

Максимизируйте вероятность (MLE):

$$L(X\,|\Sigma, \mu) = \sum_{i-1}^{n} lnf(x_i|\Sigma, \mu) \rightarrow \max_{\Sigma, \mu}. \quad (3.4)$$

где n – размер обучающей выборки; X – обучающая выборка; Σ, μ – параметры модели.

Минимизация эмпирического риска (ERM):

$$Q(X, W) = \sum_{i-1}^{n} E(x_i, \mathrm{W}) . \quad (3.5)$$

где W – параметры модели

Эквивалентность MLE и ERM

$$lnf(x_i|\Sigma, \mu) = \mathrm{E}(x_i, \mathrm{W}). \quad (3.6)$$

Проверка данных по нелинейным коллекторам

Для каждого k = 0,1, . . ., d − 1 среди всех k-мерных нелинейных многообразий в Rd найти такое Mk ⊂ Rd, что сумма квадратов уклонений xi, i = 1 . . . n от Mk минимальна (рис. 3.3):

$$Q(X, M_k) = \sum_{i=1}^{n} dist^2 x_i , M_k) \rightarrow \min_{M_k} . \quad (3.7)$$

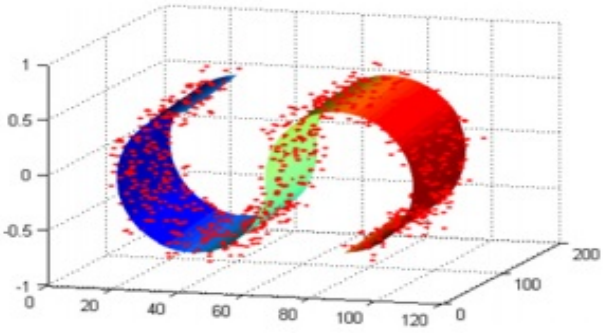


Рисунок 3.3. - Аппроксимация нелинейным многообразием

Построение такого коллектора трудоемко, поэтому аппроксимация его точек производится с помощью таких алгоритмов, как самоорганизующиеся карты Кохонена, упругие карты, алгоритм растущего «нервного газа» и так далее

$$Q(X, W) = \sum_{i=1}^{n} \sum_{j=1}^{k} ||x_i - w_j||_2^2 \rightarrow \min_{W} . \quad (3.8)$$

где wj –- значение j-ого узла сетки в Rd , wj ∈ Mk; k – количество узлов в сетке.

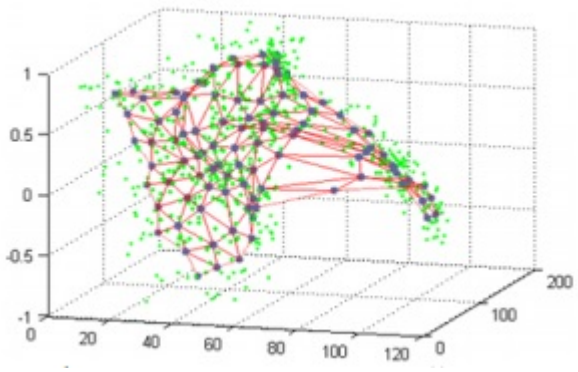


Рисунок 3.4. - Аппроксимация нелинейным многообразием в виде сетки узлов

Исходя из того, что: - MLE эквивалентно ERM: - ln f (xi | Σ, μ) = E (xi, W); - количество узлов, близких к сети, соответствует количеству латентных состояний, т.е. h = {1,. . . , k} получаем общее распределение случайных величин модели NPMPGM:

$$P(\bar{h}, \bar{x}) = \prod_{t-1}^{T} p(h_t|h_{t-1})\Psi(x_t|h_t) \qquad (3.9)$$

$$\Psi(x_t, h_t) = \frac{exp\left(-\| x_t - w_{h_t} \|_2^2\right)}{\sum_{j=1}^{k} exp\left(-\| x_t - w_j \|_2^2\right)}, p(h_t|h_{t-1}) = \frac{a_{h_t},\ h_{t-1}}{\sum_{j=1}^{k} a_{h_{t,j}}} \qquad (3.10)$$

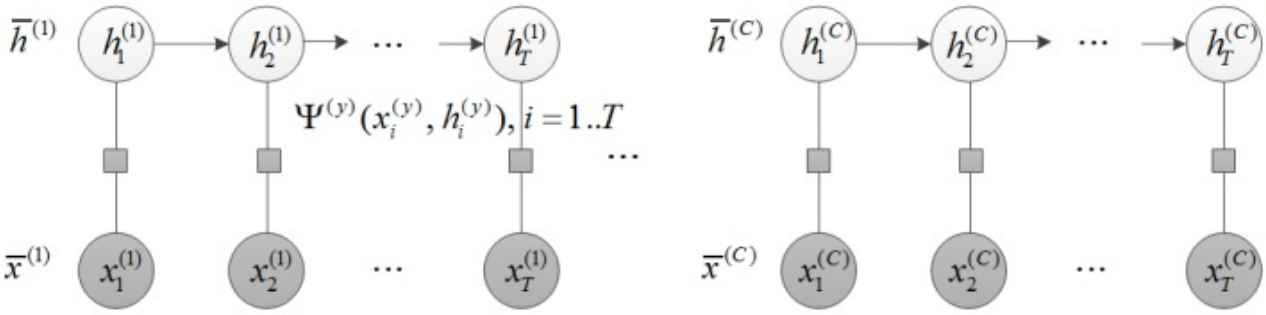


Рисунок 3.5 — Графическое представление модели NPM-PGM (C – количество классов)

Вычислите вероятность p (ql | xn) вероятности пастеризации состояния HMM ql для акустического вектора xn, наблюдаемого в начале 1990-х гг. Берлард и др. предложили использовать многослойные рецепторы.

Оптимальным значением выхода МП является распределение вероятности в дискретном состоянии CMM, определяемое входным вектором.

$$g_k\left(x_n, \Theta_{MLP}^{opt}\right) = p(S_k|x_n, \Theta_{HMM}). \quad (3.11)$$

где, Θ opt MLP — совокупность полученных при обучении МП, параметры. Показано простой способ расширения предложенной модели для применения контекстной информации, то есть последовательное использование 2c + 1 акустических векторов в качестве входных данных для персептрона $x_{n-c}^{n+c} = \{x_{n-c}, \dots, x_C, \dots, x_{n+c}$.

Тогда (3.11) переписывается

$$g_k\left(x_n, \Theta_{MLP}^{opt}\right) = p\left(q_n = S_k\middle|x_{n-c}^{n+c}, \Theta_{HMM}\right). \quad (3.12)$$

В предложенной модели прводится корреляция акустических векторов и их ограничения, связанные со статистических наблюдаемых векторов. А также, в [71] Х. Берланд и Н. Морган отмечают, что в качестве исходного параметра СММ следует использовать условие, рассчитанное на следующем шаге:

$$g_k\left(x_n, \Theta_{MLP}^{opt}\right) = p(q_k^n|q_k^{n-1}, \Theta_{HMM}) \; \forall \mathrm{k} = 1, \dots, \mathrm{K}. \quad (3.13)$$

Таким образом, эта модель использует сеть Time Delay Neural Network.

Вот как работает предлагаемая вычислительная структура. В каждый момент времени n последовательность акустических векторов x_{n-c}^{n+c} и состояние Скрытой марковской модели на предыдущем шаге q_k^{n-1} передаются на входной слой МП. Слой МФ и выходной слой — это распределение вероятностей текущего состояния Скрытой марковской модели, которое обусловлено x_{n-c}^{n+c} и q_k^{n-1}

Поскольку выходным вектором МП является приближенная вероятность последней, то $g_k\left(x_n, \Theta_{MLP}^{opt}\right)$, приближенное

$$p(q_k|x_n) = \frac{p(x_n|q_k)\,p(q_k)}{p(x_n)}. \quad (3.14)$$

которое является косвенным.

Излучение включает в себя вероятность $p(x_n|q_k)$ и априорную вероятность состояния HMM $p(q_k)$. Поскольку вероятность участвует в работе в виде мультипликативного элемента, это позволяет заменить такую величину, как априорная вероятность, в процессе классификации без дополнительного исследования персептрона для стабилизации речезависимой вероятности используется корпус данных.

Для использования вероятности $p(x_n|q_k)$ в качестве вероятности разряда для необходимо чтобы при исследуемой выборе HMM разделил выход персептрона $g_k(x_n)$ на относительную частоту возникновения, с условия S_k таким образом, что в итоге можно было оценить выражения типа $p(x_n|q_k)\,p(q_k)$. При распознавании член $p(x_n)$ остается постоянным для всех случаев и не влияет на классификацию.

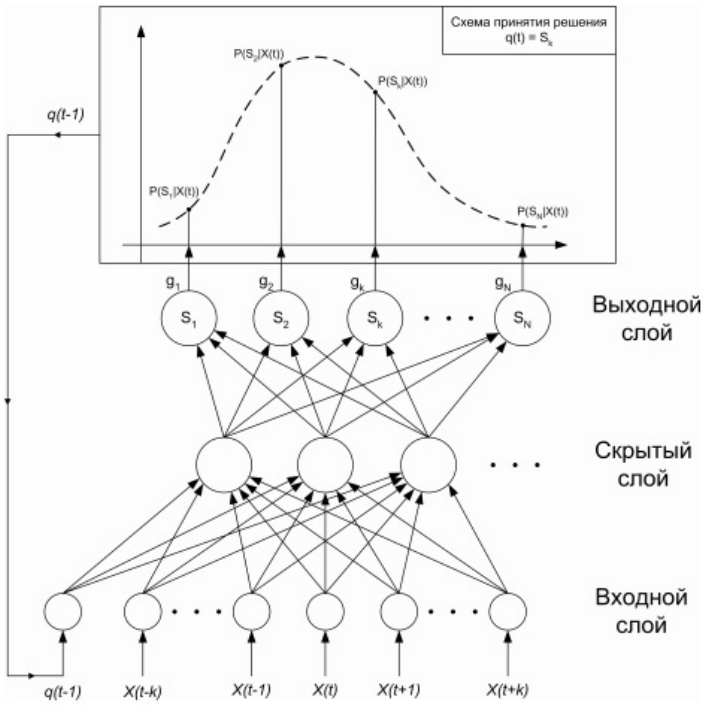


Рисунок 3.6. - Оценка вероятности с помощью TDNN сети

Подобная модель была предложена Робинсоном и другими, которые использовали повторяющуюся сеть вместо сети TDNN, а также для расчета

вероятности выбросов HMM. Предложили заменить линейные матричные операторы в основных уравнениях линейных динамических систем нелинейной обратной сетью, получив расчетную структуру.

Текущий акустический вектор x_n заносится в сети вместе с текущим вектором состояния u_n. Эти векторы передаются через стандартную разомкнутую сеть для получения выходного вектора g_n и следующего вектора состояния u_{n+1}. Если мы определим входной вектор как z_n и матрицу весов сетевой матрицы как W и V, то

$$z_n = \begin{bmatrix} 1 \\ x_n \\ u_n \end{bmatrix}. \tag{3.15}$$

$$g_n^k = \frac{\exp\,(W_k z_n)}{\sum_j exp W_j z_n} \tag{3.16}$$

Включение 3.1 в (3.15) позволяет создать компенсацию нелинейного питания. Как и в модели Боурларда, использующей сеть TDNN, результатом сети итераций является приблизительная вероятность условия HMM q_k^n в момент времени n:

$$g_n^k = P(q_n^k | X_1^n, u_0). \tag{3.17}$$

Авторы в работе [84] дают теоретическую базу для такой интерпретации. При использовании сети повторителей для расчета вероятности излучения в гибридной модели можно получить очень большой акустический контекст, используя вектор внутреннего состояния un. Как упоминалось выше, наблюдения статистически независимы, и при использовании HMM существуют предположения, а процесс маркировки относится к первой степени, т.е.

$$p(x_n | Q_1^n, X_1^{n-1}) = p(x_n | q_k^n). \tag{3.18}$$

где Q_1^n = {q_1, q_2, ..., q_n} – последовательность СММ состояний в моменты времени t = 1, 2, ...n. Использование рекуррентной сети позволяет сократить число предположений, т.е.

$$p(x_n|Q_1^n, X_1^{n-1}) = p(x_n|p_n,\ X_1^{n-1})\ . \tag{3.19}$$

что позволяет учитывать акустический контекст модели локального наблюдения. Тогда, пересматривая (3.5) с учетом (3.19) для модели M_i, получаем

$$p\left(X_1^L\ \middle|Q_1^L, M_i, \Lambda_i\right) = \prod_{l=1}^{L} p\ (x_l|X_1^{l-1}). \tag{3.20}$$

Поскольку множитель $p(x_l|X_1^{l-1})$ не зависит от фоновой последовательности, его можно не учитывать на этапе распознавания. Поскольку рекуррентная сетка используется для вычисления $P(q_l|x_l)$, необходимо вычислить оставшийся член $P(q_l|X_1^{l-1})$P. Один из самых простых методов вычисления — предположить, что текущее состояние не зависит от наблюдаемого контекста [111], т.е.

$$P\left(q_l\middle|X_1^{l-1}\right) = P(q_l)\ . \tag{3.21}$$

где, $P(q_l)$ можно определить, как относительную частоту появления состояния q_l в модели обучения, т.е. мы получим результат, аналогичный модели Берларда.

3.2. Разработка алгоритма обучения модели

Обучение гибридной модели включает в себя оценку параметров марковской цепи и весов нейронной сети. Алгоритма, позволяющего одновременно вычислять оба набора параметров для НММ и нейронной сети, до сих пор не существует. Кроме того, контролируемые исследования используются для нейронных сетей. В результате требуется большое количество отображаемых вручную данных, которых сейчас нет. Процедура повторного

обучения была предложена Берларом. Она началась с первоначальной маркировки данных обучения акустике. Сеть обучается на основе этой информации. Данные исследования выводятся заново с помощью тренировочной сети для определения вероятности потерь и проверки алгоритма Витерби. Результат записывается, и сеть переобучается.

В рамке стандартной ГМС можно обозначить начальную сегментацию или провести разделение последовательность акутичеких наблудений на одинаковые сегменты. Каждый сегмент состоит из соответствующей HMS позиции. Аналогичный метод используется в работе [93].

Алгоритм Витерби в качестве учебного варианта для расчета параметров системы с использованием гибридных моделей с итерационными сетями предложил Т. Робинсон [111]. Алгоритм Витерби также часто используется на этапе обучения и является основным инструментом для задачи распознавания. В этой ситуации параметры обновляются, чтобы увеличить вероятность последовательности наиболее вероятных состояний.

Мартин и др., предложил двухэтапный метод улучшения распознавания отдельных слов. В этом случае на этапе обучения собирается информация о конкурирующих моделях (включая информацию о продолжительности). На втором этапе информация используется для устранения неясностей.

Последовательность векторов параметров, подлежащих определению в соответствии с состоянием основанной на HMM системы ASR, является следующей:

- топология модели, которая относится к числу состояний и допустимых переходов;
- максимизация вероятности перехода;
- максимизация выбросов.

Для получения достоверных оценок этих вероятностей требуется большой объем данных об обучении (на единицу речи).

Метод уменьшения уклона используется для максимизации вероятности выбросов. Вероятность передачи максимизируется с помощью экстремальных моделей оценки продолжительности. Учебный период метода состоит из 3-х шагов:

Первым шагом является размещение фоновых меток на каждом кадре тренировочных данных, что обычно делается вручную экспертом. На втором шаге создается модель фонового континуума с целью использования результатов преобразования итеративной сети для оценки вероятности, ручной разметкой вычисляется существующий вероятность фона. Последний шаг аналогичен рекуррентной сети, которая обучается на основе ручной маркировки.

Берлард и его коллеги в 1988-1994 годах провели серию удачных экспериментов по использовании гибридных моделей для распознавания слитной речи [90]. Ими было показано что для обучения достаточно четырех итераций очень большого числа систем [111].

Система DECIPHER является независимой от динамиков системой непрерывного распознавания речи на основе телефона, которая построена на скрытых марковских моделях. Автор на конкретных примерах доказывает, что DECIPHER [90] используется для управления ресурсами проекта DARPA. Эксперименты проводились на независимой от спикера базе данных DARPA Resource Management. Размер словаря в базе данных составил 998 слов. При использовании грамматики пары слов, разброс был 60 и без грамматики, недоумение 998.

Кроме того, в случае множественной плотности использовались различные варианты расшифровки произношения слов, моделирования фоновых и акустических слов и контекстно-зависимых фоновых моделей. В системе DECIPHER используются как контекстно-зависимые, так и контекстно-независимые модели.

Многослойный перцептрон использовался в начальной гибридной системе для вычисления контекстно-независимой модели. В базовой системе MPL использовалось 69 контекстно-независимых моделей телефона. Каждое слово в

словаре использует один транскрипт произношения. CMM представляют собой фоновые модели, состоящие из двух или трех случаев, включающих параметр вероятностной плотности. Этот гибрид был сопоставлен с системой DECIPHER CMM, в которой вероятность выбросов была смоделирована с помощью гауссовых смесей. В то же время DECIPHER использовался в качестве основной системы для получения начального фонетического знака в первой итерации исследования многослойного перцептрона.

На основе этих экспериментов было достигнуто значительное улучшение качества распознавания по сравнению с контекстно-независимой системой XMM. Так, в одном из тестовых множеств (Februar91) контекстно-независимая гибридная модель показала частоту ошибок 5,8%. Этот коэффициент ошибок значительно лучше, чем независимая модель HMM, имеющая коэффициент ошибок 11% [105]. Кроме того, в одном эксперименте гауссовские смеси использовались для оценки комбинированной вероятности выбросов. Для комбинирования MLP и стандартных оценок вероятности выхода состояний были опробованы две эвристики. В первых взвешенных бревнах MLP и Gaussian оценки вероятности смеси были использованы:

$$\log\left(P\left(x\middle|q_j\right)\right) = \lambda_1 \log\left(\frac{P_{mlp}\left(x\middle|q_j\right)}{P\left(q_j\right)}\right) + \lambda_1 \log\left(P_{gm}\left(x\middle|q_j\right)\right). \qquad (3.22)$$

где, Pmlp представляет MLP оценку вероятности, а Pgm - оценку гауссовой смеси.

Во всех штатах использовался единый набор λ_i. Этот метод оценки показал лучшее качество с частотой ошибок около 5,5%.

Робинсоном в ноябре 1993 года был протестирован аналогичный эксперимент, в проекте ARPA Wall Street Journal Test, а также в Европейской оценке качества речи SQALE для лингвистической инженерии) [122].

Некоторые из ведущих мировых систем распознавания сравниваются с помощью SQALE. Примерами являются:

1. Cu-Con и Cu-HTK разработаны инженерным факультетом Кембриджского университета (Великобритания),

2. ЛИМСИ реализуется Лабораторией информатики для механики и науки Ингеньера (Франция)

3. Группа PHILIPS Man-Machine Interface (Германия).

Системы Cu-Con, LIMSI и PHILIPS основаны на HMM, которые используют постоянную плотность для акустического моделирования. Для акустического моделирования Cu-Con использовал четыре сети [96] повторяющихся нейронов. Каждая сеть состояла из одного слоя, и её выход через каждый временной интервал был возможным вектором фоновой оценки, а в качестве возврата на входной уровень использовался вектор 256-мерного состояния.

Полученные результатов по каждой сети (четыре вероятности), были затем объединены в одну вероятность для каждого фона в каждом входном кадре. Кроме того, периодические сети были использованы для оценки контекстных классов каждого фона. Затем они обучались на основе вектора состояния каждой сети. Результаты такой оценки были объединены с контекстно-независимыми фоновыми вероятностями для получения контекстно-зависимых умноженных фоновых вероятностей.

Для выбора контекстов использовалась древовидная процедура принятия решений [98]. Триграммы и большие числа использовались в качестве языковой модели системы. Для американского английского языка был проведен сравнительный эксперимент с использованием триграмм и биграмов. Результаты приводятся ниже.

Хеннеберг и его коллеги [95] представили глобальному пастору модели сложность теоретических основ, разработанных Боурлардом и Морганом, обобщив локальный пастор, который был разработан как новый алгоритм обучения для гибридной модели. Это расширение основано на работе Франко и коллег [92], в которой CMM был заменен автономным контекстом с моделью,

которая позволяет включать акустический контекст в себя. Введение зависимости от контекста может быть факторизовано теоремой Байеса.

$$P(x|q_j, c_k) = \frac{P(q_j|X, c_k)P(X|c_k)}{P(q_j|c_k)}. \quad (3.23)$$

где X — вектор акустических наблюдений, q_j — состояние СММ, c_k (k = 1, 2, ..., K) — рассматриваемый контекст.

В уравнении (3.23) виличина P (q_j | X, c_k) вычисляется с использованием многослойного персептрона. В модели Боурларда для оценки полной вероятности состояния P (q_j | X), $q_j \in S$ и K в работе [92] использовался персептрон, где K — длина рассматриваемого контекста. Таким образом, k-й персептрон обучен вычислять P (qj | X, ck) для всех qj ∈ S. Для обучения используется стандартный алгоритм обратной связи. Величину P(X|ck) также можно разделить на множители

$$P(X|c_k) = \frac{P(c_k|X)P(X)}{P(c_k)}. \quad (3.24)$$

где P (X | ck) рассчитывается с использованием многослойного персептрона, аналогичного модели, используемой Боурлардом. Величину P(ck) рассчитывают, как частоту появления k-го контекста в учебном корпусе акустических данных.

3.3. Разработка алгоритма вывода на модели

Расчеты вероятности выполняются с использованием алгоритма прямого-обратного хода (Р. Л. Стратонович, 1960), такого как модели HMM и HCRF. Временная сложность алгоритма O (k 2T).

Поскольку количество узлов в приближенной сетки k может быть очень большим, для ускорения вычислительного процесса можно использовать алгоритмы сэмплирования:

Таблица 3.1 UCI: Spoken Arabic Digit Data Set

Model	HMM(1 mix)	HMM(16 mix)	HCRF(BFGS)	HCRF(CG)	NPM-PGM
States	5	5	5	5	1024
F-score	0.6018	0.5374	0.6420	0.6274	0.8043

Таблица 3.2 UCI: Character Trajectories Data Set

Model	HMM(1 mix)	HMM(16 mix)	HCRF(BFGS)	HCRF(CG)	NPM-PGM
States	7	7	7	7	256
F-score	0.9329	0.9288	0.9651	0.9516	0.9809

Размер обучающей выборки: 800 экземпляров каждого класса

Таблица 3.3 UCI: Spoken Arabic Digit Data Set

Model	HMM(1 mix)	HMM(16 mix)	HCRF(BFGS)	HCRF(CG)	NPM-PGM
States	5	5	5	5	1024
F-score	0.8778	0.8368	0.9525	0.9344	0.9349

Список преимуществ предлагаемой алгоритма:

- значение F-score получает наибольшее, чем HMM во всех вариантах;

- значение F-score оцивается, чем у HCRF, наибольшее количестве данных.

Недостатки предлагаемой алгоритма:

- сложный выбор количества узлов в аппроксимационной сетки k;

- набольшое число скрытых случаев k, влияющих на время классификации (сложность времени работы алгоритма «прямого-обратного» хода O(k 2T)).

3.4. Исследование алгоритмов

Для качественного анализа речи, ее нужно отобразить в понятное для в вычислительной системе: аналоговой, цифровой, спектральной, оптической и другой форме.

При моделировании и исследования речи на персональном компьютере предусмотрена только один способ -представление звукового сигнола в цифровой форме

Для преобразования акустического сигнала в цифровую информацию используют модуляцию звуковых упругих волн, где звуковые сигналы превращаются в электрические (например, с помощью микрофона (рис. 3.8).

Ввод звука в компьютер

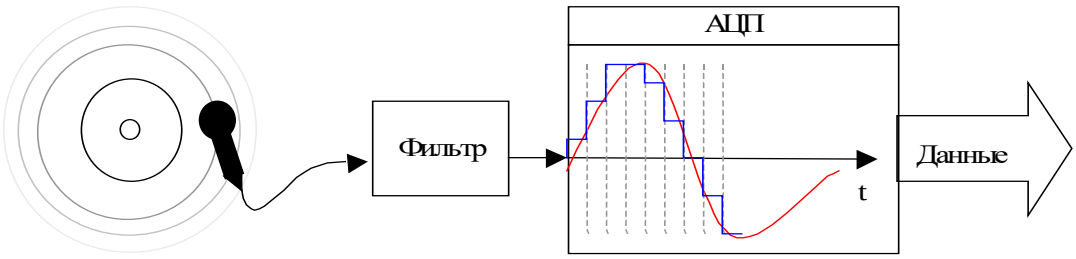


Рисунок 3.7. - Ввод звука в компьютер

Далее этот сигнал необходимо отфильтровать и преобразовать с помощью аналого-цифрового преобразователя с определенной частотой *fd* (*частота дискретизации*) в цифровой форме и и записать т.е., проводит процедуру квантования. По теореме Колесникова

$$f\ \max < \frac{fd}{2} \qquad (3.25)$$

Определяющим качество квантования сигнала параметром, являются *частота дискретизация (fd)* и глубина преобразования битов (количество единиц информации, кодирующих отсчет). Максимальная частота сигнала определяется выбранной частотой, эта частота может быть записана в виде (3.25) Типичные значения 11025, 22050, 44100 Гц.

Степень точности кодирование звука при преобразовании из аналогового в цифровую зависит от типа разрядности. Типичные значения: 4 бита, 8 бит, 16 бит для одной выборки.

Предварительная обработка и выделение первичных признаков. Речевой сигнал, который поступает для распознавания, предварительно обрабатывается для компенсации погрешностей при вводе звука и определения специфики сигнала. Такая обработка состоит в шумоочистке сигнала, фильтрации и приведение к норме до определенного уровня.

Далее требуется выделить наиболее информативные признаки сигнала, то есть наиболее полно описывающие сигнал в краткой форме. Совершенно очевидно, что эффективность данного процесса обусловливает результативность обработки сигнала в дальнейшем и его распознавание. Временное представление сигнала довольно неэффективно, прежде всего по причине того, что не

учитывается периодичность звука, и, кроме того, по причине изменчивости речевого сигнала, когда один и тот же звук, который произнесен одним и тем же диктором, может сильно отличаться во временной интерпретации.

Более информативно представление речи в виде спектра. Для его получения применяют систему полосовых фильтров, которая настроена на выявление частот или дискретное преобразование Фурье. Далее спектр преобразовывается, к примеру, логарифмически изменяется масштаб (в пространстве амплитуд и частот), сглаживается спектр для выделения огибающей, проводится кепстральный анализ (обратное преобразование Фурье). Это дает возможность учитывать определенные особенности звука уменьшение информативности участков спектр с более высокими частотами, логарифмическую чувствительность слухового аппарата человека и т.д.

Однако визуализировать образ речевого сигнала, используя только спектр, невозможно. Необходима также информация о речевой динамике. Для получения данных используются дельта-параметры, являющиеся производными от основных параметров по времени.

Параметры полученного сигнала должны учитывать его основные характеристики, которые представляют сигнал при дальнейшей обработке.

Выделение примитивов речи

Примитивы речи – фонемы, которые составляют сложную речь (по поводу числа фонем нет единого мнения: по некоторым данным, в русском языке 43 фонемы, по другим – 64, по третьим – более 100). Процесс выделения и распознавания фонем является первым этапом при распознавании во многих системах. От результатов зависит процесс распознавания на следующих уровнях.

При применении нейросетей обучение процессу выделения заключается в создании нейронных ансамблей, с наличием ядер, соответствующих наиболее частым формам каждого из примитивов. Создание ансамблей представляет собой обучения системы без учителя, когда статистическая обработка поступающих сигналов осуществляется на входе нейронной сети. Редкие

сигналы система запоминает позже. Это требует участия механизма фокусировки или какой-либо другой формы контроля.

Распознавание сложных звук, слов, фраз и многое другое. Для распознавания слитной речи наиболее понятной является система уровней: первый уровень – распознавание фонем, второй – слогов, и далее слов, фраз и т.д. На каждом из уровней происходит кодировка сигнала элементами предыдущих уровней. Во время перехода по уровням, кроме представителей сигнала, передаются и их дополнительные признаки, зависимости по времени и отношения между сигналами. При аккумуляции сигналов с предыдущих уровней, более высокие уровни имеют большой объем информации и могут производить управление теми процессами, которые происходят на уровнях ниже, к примеру, с привлечением механизма внимания.

3.5. Выводы по главе

Гибридная СММ/ИНС модель. Данная модель дает возможность наиболее эффективно аккумулировать положительные характеристики СММ и нейронных сетей, то есть СММ может обеспечить возможность моделирования в отношении длительных зависимостей, а нейросети обеспечивают возможность моделирования таких величин, как универсальная непараметрическая аппроксимация, оценка вероятности, алгоритм дискриминантного обучения, уменьшая количество критериев оценки.

Вероятности максимизируются с помощью градиентного уменьшения, а вероятности перехода максимизируются с помощью моделей оценки экстремальной продолжительности. Период обучения состоит из следующих этапов:

Этап 1. Поместить метки на каждый фрейм данных исследования, то есть первоначальная разметка обычно выполняется экспертом вручную.

Этап 2. На основе ручной разметки строится модель длительности фона и рассчитывается априорная вероятность фона, которая используется для преобразования рекуррентной сети в оценку вероятности.

Этап 3. Аналогичным образом на основе ручной разметки обучается рекуррентная сеть.

Этап 4. Используя параметры, рассчитанные на этапе 2, и сеть рекуррентной, обученную на этапе 3, мы демонстрируем обучающую информацию и переходим к этапу 2.

Расчеты вероятности выполняются с использованием алгоритма прямого-обратного хода (Р. Л. Стратонович, 1960), такого как модели HMM и HCRF. Временная сложность алгоритма O (k 2T).

Алгоритмы распознавания речи:

1. Для отображения акустического сигнала в цифровом виде почти все системы, имеющие дело со звуком, используют импульсную модуляцию.

2. Речевой сигнал, который поступает в распознавательную систему, предварительно обрабатывается для компенсации погрешностей при вводе звука и с учетом особенностей сигнала. Гораздо более информативно спектральное представление речи.

3. Примитивы речи – фонемы, которые составляют сложную речь (по поводу числа фонем нет единого мнения: по некоторым данным, в русском языке 43 фонемы, по другим – 64, по третьим – более 100). Процесс выделения и распознавания фонем является первым этапом при распознавании во многих системах. От результатов зависит процесс распознавания на следующих уровнях.

4. Для распознавания слитной речи наиболее понятной является система уровней: первый уровень – распознавание фонем, второй – слогов, и далее слов, фраз и т.д.

Описанная нами модель имеет достоинства и недостатки:

Преимущества предлагаемой модели:

- Лучше F-score показатель, чем HMM во всех тестах;

- Лучшая F-score оценка, чем у HCRF, при небольшом количестве данных исследования.

Недостатки предлагаемой модели:

- сложность отбора оптимального количества узлов сетки k, дающего наилучшее качество классификации;

- большое количество скрытых случаев k, влияющих на время классификации (сложность времени работы алгоритма «прямого-обратного» хода O (k 2T)).

ГЛАВА 4. РАЗРАБОТКА КОМПЛЕКСА ПРОГРАММ ПОИСКА КЛЮЧЕВЫХ СЛОВ В РЕЧИ

4.1. Архитектура программного комплекса

Разработка интерфейса приложения является одним из ключевых компонентов объектно-ориентированного программирования и программирования в среде многозадачной операционной системы. Интерфейс присутствует практически во всех приложениях, разработанных для операционной системы Windows.

Для разработки интерфейса приложения можно использовать различные методы. Искомый интерфейс состоит из большого набора программных кодов, описывающих характеристики объекта, расположенного на интерфейсе. Существует множество языков программирования, с помощью которых можно создать приложение в среде Windows. Каждый язык программирования предлагает специальные методы программирования и инструменты для подготовки интерфейса программы.

С момента зарождения первого программного интерфейса и до полного появления концепции объектно-ориентированного программирования всем программистам и разработчикам компьютерных программ предлагалось разрабатывать программный интерфейс на основе предварительно созданных объектов. Однако с появлением необычных проблем при разработке программного интерфейса этот метод создания интерфейса имеет серьезные недостатки. Например, из-за графических возможностей любого языка программирования невозможно создать форму круга, треугольника и так далее. По этой причине профессиональные программисты рекомендуют вместо RAD использовать закодированную форму. Язык программирования Delphi поддерживает разработку интерфейса форм и других объектов как через код, так и через среду RAD.

Проект выполнен с использованием объектно-ориентированного подхода («ООП»). Delphi — еще один программный продукт Borland, который используется для мгновенного проектирования и основан на языке Pascal. До появления VB язык Delphi был одним из немногих языков, в которых существовала форма. Этот простой язык использовался для создания COM и ActiveX-объектов. В Delphi начало и конец блоков помечаются BEGIN…END, а в C# круглыми скобками — «{» и «}».

Диаграмма классов

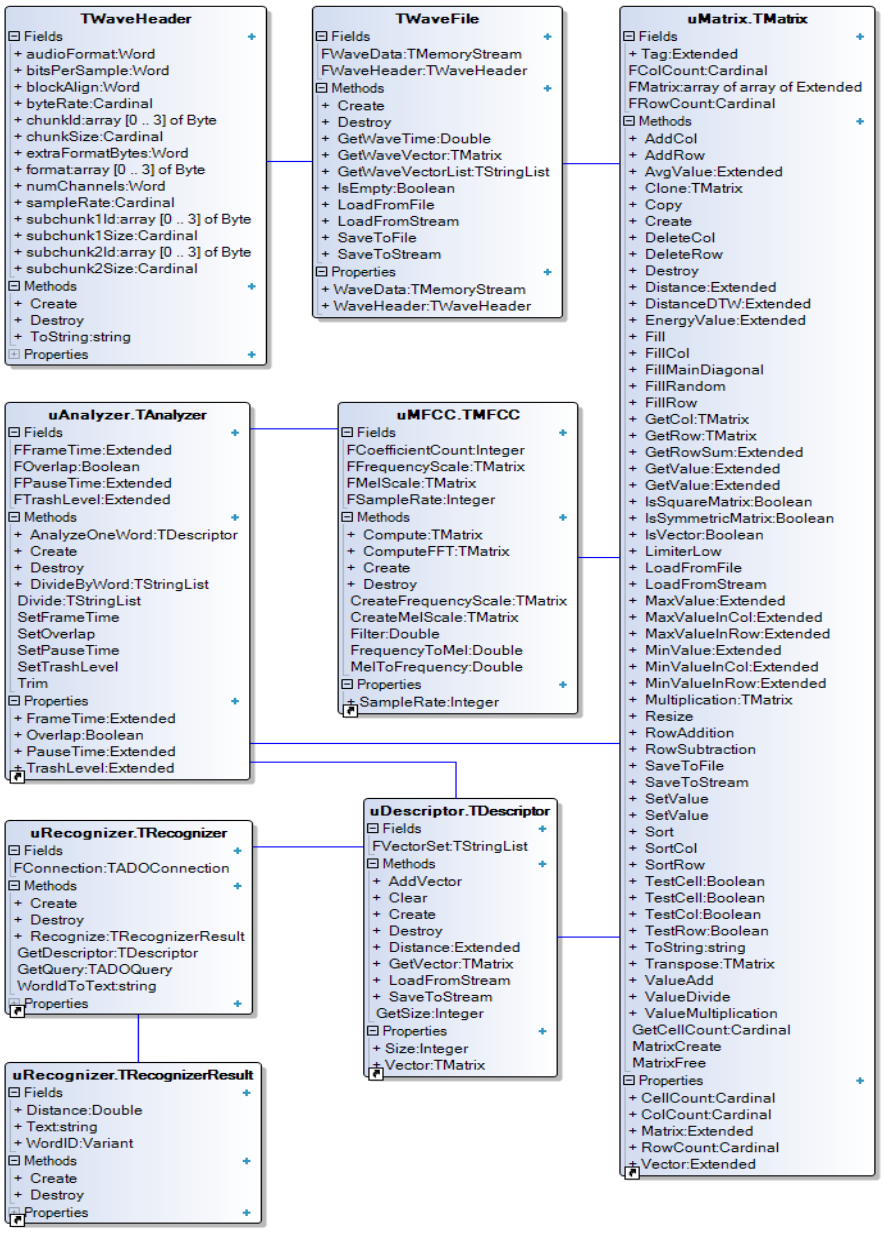


Рисунок 4.1. - Схема диаграмма классов в комплекс программ

Класс «TMatrix» для работы с матрицами и векторами. На основе данного класса выполняются все операции над векторами и матрицами, которые используются в проекте. Также данный класс реализует вычисление расстояния DTW.

Таблица 4.1. Конструктор класса «TMatrix»

Field Summary	
internal Cardinal	FColCount - Число столбцов матрицы
internal array of array of Extended	FMatrix - Матрица
internal Cardinal	FRowCount - Количество строк матрицы
public Extended	Tag - Переменная для нужд прикладных программистов
Property Summary	
public Cardinal	CellCount - Количество ячеек (только для векторов)
public Cardinal	ColCount - Количество столбцов
public Extended	Matrix - Значения матрицы
public Cardinal	RowCount - Количество строк
public Extended	Vector - Значение вектора
Method Summary	
public Sub	AddCol() Метод добавления столбца матрицы
public Sub	AddRow() Метод добавления строки матрицы
public function Extended	AvgValue() Функция подсчета среднего значения матрицы
public function TMatrix	Clone() Метод клонирования матрицы
public Sub	Copy(Source: TMatrix) Метод копирования значений из другой матрицы
public Sub	DeleteCol(Col: Cardinal) Метод удаления столбца матрицы
public Sub	DeleteRow(Row: Cardinal) Метод удаления строки матрицы
public Sub	Destroy() Деструктор класса
public function Extended	Distance(M: TMatrix) Метод расчета расстояния между векторами

public function Extended	DistanceDTW() Метод расчета дистанции алгоритмом динамической трансформации временной шкалы (DTW)
public function Extended	EnergyValue() Вычисление среднеквадратичной энергии вектора (матрицы)
public Sub	Fill(Value: Extended) Метод заполнения матрицы
public Sub	FillCol(Col: Cardinal; Value: Extended) Метод заполнения столбца матрицы
public Sub	FillMainDiagonal(Value: Extended) Метод заполнения главной диагонали каким-либо значением (только для квадратных матриц)
public Sub	FillRandom(AFrom: Integer; ATo: Integer; Symmetric: Boolean) Метод заполнения матрицы случайными числами
public Sub	FillRow(Row: Cardinal; Value: Extended) Метод заполнения строки матрицы
internal function Cardinal	GetCellCount() Количество ячеек (только для векторов)
public function TMatrix	GetCol(Col: Cardinal) Метод извлекает столбец матрицы
public function TMatrix	GetRow(Row: Cardinal) Метод извлекает строку матрицы
public function Extended	GetRowSum(Row: Cardinal) Метод вычисления суммы в строке
public function Extended	GetValue(Row: Cardinal; Col: Cardinal) Метод чтения значения ячейки матрицы
public function Extended	GetValue(Cell: Cardinal) Метод чтения значения элемента вектора
public function Boolean	IsSquareMatrix() Возвращает True, если матрица квадратная
public function Boolean	IsSymmetricMatrix() Метод проверки матрицы на симметрию
public function Boolean	IsVector() Возвращает True, если матрица - вектор
public Sub	LimiterLow(Limit: Extended; Value: Extended) Метод заменяет все значения матрицы значением "Value" если они ниже "Limit".
public Sub	LoadFromFile(FileName: string) Метод загрузки матрицы из файла
public Sub	LoadFromStream(S: TStream) Метод загрузки матрицы из потока

Класс «TWaveFile». Класс для работы с файлами формата Wave. Внутри данного класса реализованы функции преобразования сигнала в вектор (метод «GetWaveVector» класса «TMatrix»), над которым и будут проводиться все дальнейшие операции.

Таблица 4.2. Конструктор класса «TWaveFile»

Field Summary	
internal TMemoryStream	FWaveData - Тело файла
internal TWaveHeader	FWaveHeader - Заголовок файла
Property Summary	
public TMemoryStream	WaveData - Тело файла
public TWaveHeader	WaveHeader - Заголовок файла
Method Summary	
public Sub	Destroy() - Деструктор класса
public function Double	GetWaveTime() - Метод вычисления времени воспроизведения
public function uMatrix.TMatrix	GetWaveVector() - Метод выгружает вектор данных для тела
public function TStringList	GetWaveVectorList(MainVector: uMatrix.TMatrix ; SegmentSize: Integer; Overlap: Boolean) - Метод деления исходного вектора на сегменты
public function Boolean	IsEmpty() - Метод проверяет сигнал на наличие полезной информации
public Sub	LoadFromFile(FileName: string) - Метод загрузки сигнала из файла
public Sub	LoadFromStream(Stream: TStream) - Метод загрузки сигнала из потока
public Sub	SaveToFile(FileName: string) - Метод сохранения сигнала в файл
public Sub	SaveToStream(Stream: TStream) - Метод сохранения сигнала в поток

Класс, описывающий заголовок Wave файла.

Таблица 4.3. Конструктор класса «TWaveHeader»

Field Summary	
public Word	audioFormat Для PCM = 1 (то есть, Линейное квантование).

public Word	bitsPerSample Так называемая "глубиная" или точность звучания. 8 бит, 16 бит и т.д.
public Word	blockAlign numChannels*bitsPerSample/8 Количество байт для одного сэмпла, включая все каналы.
public Cardinal	byteRate sampleRate * numChannels * bitsPerSample/8
public array [0 .. 3] of Byte	chunkId Содержит символы "RIFF" в ASCII кодировке
public Cardinal	chunkSize Это оставшийся размер цепочки, начиная с этой позиции. 36+subchunk2Size, или более точно:4+(8+subchunk1Size)+(8+ subchunk2Size Иначе говоря, это размер файла-8, то есть, исключены поля chunkId и chunkSize.
public Word	extraFormatBytes Резерв
public array [0 .. 3] of Byte	format Содержит символы "WAVE"
public Word	numChannels Количество каналов. Моно = 1, Стерео = 2 и т.д.
public Cardinal	sampleRate Частота дискретизации. 8000 Гц, 44100 Гц и т.д.
public array [0 .. 3] of Byte	subchunk1Id Содержит символы "fmt "
public array [0 .. 3] of Byte	subchunk2Id Содержит символы "data"
public Cardinal	subchunk2Size numSamples * numChannels * bitsPerSample/8 Количество байт в области данных.
Method Summary	
public Sub	Destroy() Деструктор класса
public function string	ToString() Метод переводит заголовок файла в строку

Класс «TAnalyzer». Класс анализа звуковых последовательностей (звуковых векторов). Данный класс реализует функции, связанные с делением

исходного сигнала на слова, делением слова на фреймы, вычисление дескриптора для слова. Основным методом данного класса является «AnalyzeOneWord», который используется для вычисления дескриптора для одного слова.

Таблица 4.4. Конструктор класса «TAnalyzer»

Field Summary	
internal Extended	FFrameTime Продолжительность фрейма
internal Boolean	FOverlap Определяет будет ли дескриптор вычисляться с перекрытием (или без него). Параметр "перекрытие".
internal Extended	FPauseTime Продолжительность паузы между словами
internal Extended	FTrashLevel Пограничное значение для средней энергии фрейма. Фрейм считается "тишиной", если его средняя энергия ниже уровня TrashLevel. Параметр "уровень тишины".
Property Summary	
public Extended	FrameTime Свойство "продолжительность фрейма"
public Boolean	Overlap Свойство "перекрытие"
public Extended	PauseTime Свойство "продолжительность тишины"
public Extended	TrashLevel Свойство "уровень тишины"
Method Summary	
public function uDescriptor.TDescriptor	AnalyzeOneWord(MainVector: uMatrix.TMatrix ; SampleRate: Integer) Метод вычисления дескриптора для одного слова
public Sub	Destroy() Деструктор класса
internal function TStringList	Divide(MainVector: uMatrix.TMatrix ; SegmentSize: Integer; Overlap: Boolean) Метод разделения исходного вектора на фреймы
public function TStringList	DivideByWord(MainVector: uMatrix.TMatrix ; SampleRate: Integer) Метод разделения исходного вектора на слова (если исходный сигнал состоит из нескольких слов)
internal Sub	SetFrameTime(Value: Extended) Задает значение параметра "Продолжительность фрейма"

Класс «TDescriptor». Дескриптор сигнала (так же «дескриптор слова» - набор векторов, которые описывают фреймы сигнала (слова)). Является результатом работы класса «TAnalyzer».

Таблица 4.5. Конструктор класса «TDescriptor»

Field Summary	
internal TStringList	FVectorSet - Набор фреймов
Property Summary	
public Integer	Size - Размер дескриптора (количество векторов)
public uMatrix.TMatrix	Vector - Вектора из дескриптора
Method Summary	
public Sub	AddVector(Vector: uMatrix.TMatrix) Метод добавляет
public Sub	Clear() Метод очистки дескриптора
public Sub	Destroy() Деструктор класса
public function Extended	Distance(Descriptor: TDescriptor) Метод расчета расстояний между дескрипторами
internal function Integer	GetSize() Размер дескриптора (количество векторов)
public function uMatrix.TMatrix	GetVector(Index: Integer) Метод выгрузки вектора из дескриптора
public Sub	LoadFromStream(S: TStream) Метод загрузки дескриптора из потока
public Sub	SaveToStream(S: TStream) Метод сохранения дескриптора в поток

Класс «TRecognizer». Класс, выполняющий поиск наиболее близкого дескриптора. После анализа звукового сигнала и разделения его на отдельные слова, для каждого слова классом «TAnalyzer» вычисляется дескриптор, после этого каждый дескриптор передается в метод «Recognize» класса «TRecognizer» и для этого дескриптора производится поиск наиболее близкого для него дескриптора (наиболее близкого слова из словаря). Метод «Recognize» класса «TRecognizer» возвращает объект класса «TRecognizerResult», который

содержит результаты сравнения (распознавания). В процессе поиске класс активно использует базу данных, выгружая из неё дескрипторы для словарных слов. Основным методом данного класса является «Recognize».

Таблица 4.6. Конструктор класса «TRecognizer»

Field Summary	
internal TADOConnection	FConnection - Соединение с базой данных
Method Summary	
public Sub	Destroy() - Деструктор класса
internal function uDescriptor.TDescriptor	GetDescriptor(PronunciationID: Integer) - Метод получения дескриптора варианта произношения по его идентификатору
internal function TADOQuery	GetQuery(SQL: String) - Метод отвечающий за выполнение запросов к базе данных
public function TRecognizerResult	Recognize(Descriptor: uDescriptor.TDescriptor) - Метод выполняющий распознавание (поиск наиболее близкого дескриптора).
internal function string	WordIdToText(WordID: Variant) - Метод определяющий текст слова по идентификатору слова

Класс «TRecognizerResult». Класс для хранения результатов распознавания (результатов сравнения дескрипторов).

Таблица 4.7. Конструктор класса «TRecognizerResult»

Field Summary	
public Double	Distance - Расстояние между дескрипторами
public string	Text - Текст (слово) соответствующий наиболее близкому варианту произношения
public Variant	WordID - Идентификатор слова
Method Summary	
public Sub	Destroy() - Деструктор класса

Класс «TMFCC». Класс для вычисления мел-частотных кепстральных коэффициентов (Mel-frequency cepstral coefficients). Функциональность данного

класса используется внутри класса «TAnalyzer» при вычислении дескрипторов. Основным методом данного класса является «Compute».

Таблица 4.8. Конструктор класса «TMFCC»

Field Summary	
internal Integer	FCoefficientCount - Количество вычисляемых коэффициентов
internal uMatrix.TMatrix	FFrequencyScale - Частотная шкала для заданного количества коэффициентов
internal uMatrix.TMatrix	FMelScale - Mel-шкала для заданного количества коэффициентов
internal Integer	FSampleRate - Частота дискретизации сигнала (8000 Гц, 44100 Гц и т.д.)
Property Summary	
public Integer	SampleRate - Частота дискретизации сигнала (8000 Гц, 44100 Гц и т.д.)
Method Summary	
public function uMatrix.TMatrix	Compute(Frame: uMatrix.TMatrix) Вычисление вектора признаков (MFCC)
public function uMatrix.TMatrix	ComputeFFT(Frame: uMatrix.TMatrix) Метод вычисления быстрого преобразования Фурье
internal function uMatrix.TMatrix	CreateFrequencyScale(MelScale: uMatrix.TMatrix) Метод вычисления частотной шкалы
internal function uMatrix.TMatrix	CreateMelScale(Size: Integer; MinFrequency: Integer; MaxFrequency: Integer) Метод вычисления мел-шкалы
public Sub	Destroy() Деструктор класса
internal function Double	Filter(CoefficientIndex: Integer; Frequency: Double; FrameSize: Integer) Треугольный фильтр
internal function Double	FrequencyToMel(Value: Double) Преобразование частоты в мел
internal function Double	MelToFrequency(Value: Double) Преобразование мел в частоту

4.2. Проектирование вычислительных модулей

Чтобы сохранить звуковой сигнал в цифровой среде, он должен быть распределен на несколько расстояний, и на каждом должно быть получено «усредненное» значение.

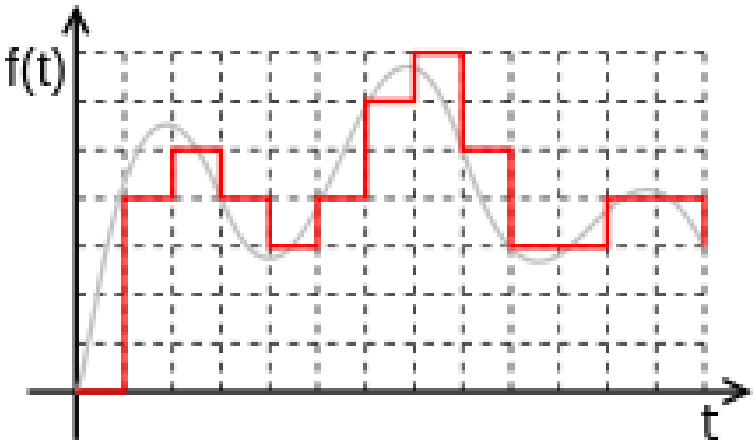


Рисунок 4.2. - Структура WAVE файла в заданное время t

Обычно использует стандартный формат файлов WAV (Windows PCM) для обработки сигналов в памяти компьютера. Поток данных в файле формата WAV (windows audio voice) состоит из двух частей:

1. Заголовок файла в этом разделе хранится следующая информация: размер файла, количество каналов, частота дискретизации и количество битов в сэмпле (глубина звука).

2. Области данных, амплитуда волн в цифровом формате.

Чтобы лучше понять смысл, в названии следует упомянуть поле данных и нумерацию звука. Звук — это изменение, происходящее при оцифровке звука. Причина этого в том, что компьютер имеет возможность воспроизводить звук определенной амплитуды за короткий промежуток времени. Отметим, что этот промежуток не может быть очень коротким. Его длительность определяется скоростью выброса. Например, если в файле с частотой 44,1 кГц, длительность этого промежутка времени составляет 1/44100 секунды. Современные

устройства с поддержкой звуковых карт могут поддерживать частоту дискретизации до 192 кГц.

Точность аудиосигнала, возможность выделения его от шумового фона зависит и от амплитуды сигнала (громкости звука). Амплитуда выражается как число, которое занимает 8, 16, 24, 32 бита в памяти (файле) (теоретически возможно). Следовательно, одна из амплитуд может занять в файле 1, 2, 3, 4 байта в течение короткого времени. В монозвуковом варианте значения амплитуд располагаются в последовательном порядке, а в стереозвуковом варианте сначала идет значение левого канала, затем правого и так далее.

Совокупность амплитуд в заданном периоде времени называется паттерном.

Таблица 4.9. Сочетание амплитуды

Местоположение	**Поле**	**Описание**
0..3 (4 байта)	chunkId	“RIFF” в ASCII кодировке (0x52494646 в big-endian представлении). Представляет собой начало RIFF-цепочки.
4..7 (4 байта)	chunkSize	Остаток цепи с данной позиции - размер файла – 8, т.е. исключены поля chunkId и chunkSize.
8..11 (4 байта)	format	Содержит символы «WAVE» (0x57415645 в big-endian представлении)
12..15 (4 байта)	subchunk1Id	“fmt “ (0x666d7420 в big-endian представлении)
16..19 (4 байта)	subchunk1Size	16 для формата PCM. Остаток подцепи с данной позиции.
20..21 (2 байта)	audioFormat	Для PCM = 1 (Линейное квантование). Другие значения обозначают формат сжатия.
22..23 (2 байта)	numChannels	Число каналов. Моно = 1, Стерео = 2 и т.д.
24..27 (4 байта)	sampleRate	Частота дискретизации. 8000 Гц, 44100 Гц и т.д.
28..31 (4 байта)	byteRate	Число байт, которые переданы за секунду воспроизведения.
32..33 (2 байта)	blockAlign	Число байт для одного сэмпла, в том числе все каналы.
34..35 (2 байта)	bitsPerSample	Число бит в сэмпле - “глубина” или точность звучания. 8 бит, 16 бит и т.д.
36..39 (4 байта)	subchunk2Id	“data” (0x64617461 в big-endian представлении)
40..43 (4 байта)	subchunk2Size	Число байт в области данных.
44..	data	WAV-данные.

При анализе сигнал разбивается на небольшие пространства - фреймам. Фреймы не следуют друг за другом, не перекрываются. То есть легче проводить между волнами на расстоянии, а не в конкретных точках.

Размещение фреймов описанным выше способом позволяет сгладить результаты анализа и представить фреймы в виде окна, перемещающегося по исходной функции.

Было установлено, что длина каждого фрейма должна соответствовать интервалу в 10 мс, а перекрывающиеся фреймы – 50%.

Если средняя длина слова 500 мс, это дает нам примерно 500/(10*0,5)=100 фреймов на слово.

MFCC является характеристикой речевого сигнала (вычисляется для каждого фрейма). MFCC – спектральная энергия сигнала. Положительными сторонами являются:

- применяется спектр сигнала (разложение на основе ортогональных [ко]синусоидальных функций), что позволяет обратить внимание на характер волновом сигнала;
- спектр переводится в специальную mel-шкалу, что позволяет выделить наиболее важные для восприятия частоты;
- количество вычисляемых коэффициентов ограничено любыми значениями (например, 12), что позволяет сжимать фрейм и количество информации при обработке.

Необходимо рассмотреть порядок расчета коэффициентов для абстрактного фрейма.

В первую очередь нужно рассчитать спектр сигнала с помощью дискретного преобразования Фурье (по сути, его «быстрая» реализация FFT).

$$X[k] = \sum_{n=0}^{N-1} \mathrm{x[n]} * e^{-2*\pi*i*k*n/N}, 0 \leq k < N,$$

Следует отметить, что после этого преобразования ось X — это частота (Гц) сигнала, а ось Y — величина (как способ отойти от комплексных значений):

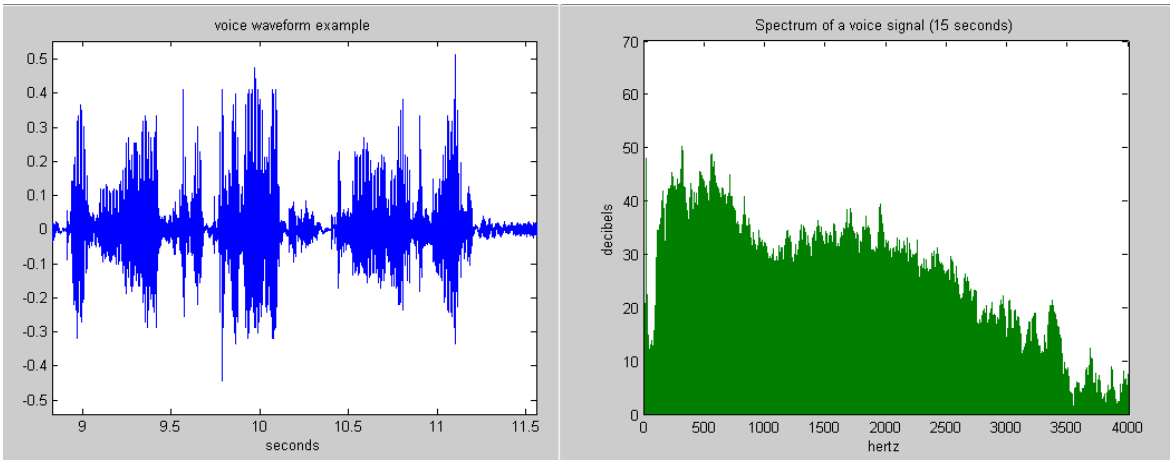


Рисунок 4.3. - Сигнал звук и спектр звук

Mel зависит от частоты сигнала, то есть тембра и амплитуды. Это значение, которое показывает, звук определенной частоты.

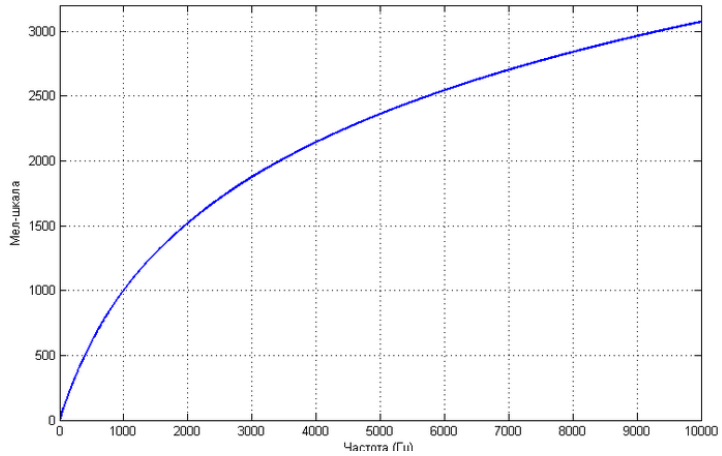


Рисунок 4.4. - График зависимости тембра и амплитуды в Mel-шкале

Например, для 256 элементов в некотором кадре, частота звука записывается в диапазоне 16000 Гц. Следует отметит, что, человеческая речь находится в диапазоне [300; 8000] Гц. При этом, количество Mel-коэффициентов равняется M = 10.

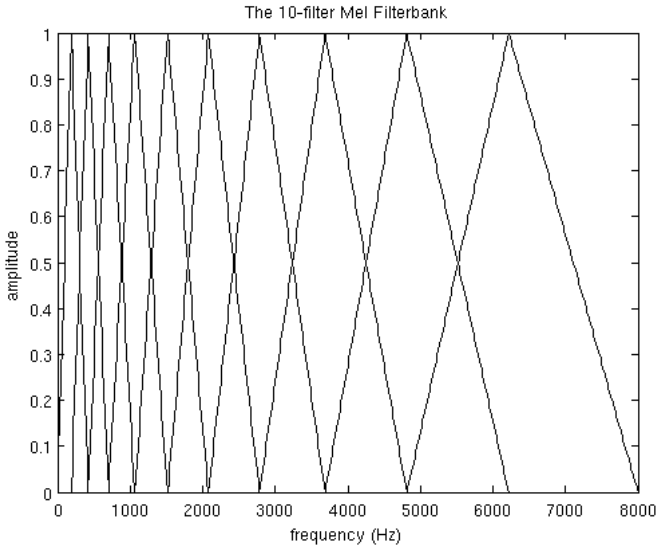


Рисунок 4.5. - Mel-фильтр

Для большого порядка Mel-коэффициентов, необходимо реализовать большое основание фильтра. В связи с этим, распределение частотного диапазона по диапазонам, обрабатываемым фильтрами, производится по шкале.

Для масштабирования спектра рассматривается абстрактный фрейм с частотой, расположенной на оси X с длиной спектра 256 элементов ичастотой 16000 Гц.

Данный фильтр для умножения значений фильтра и спектра на две части. В результате этого процесса должен быть получен mel-коэффициент. Так как у нас есть M фильтров, то и коэффициентов у нас будет равного количества.

Результаты фильтра Mel к спектральным значениямп полученные в результате логарифмических, показателей снижает чувствительность и шума.

Теперь имеется набор M коэффициентов-mfcc для каждого фрейма, которые можно использовать для дальнейшего анализа.

Допустим, у нас имеется эталон звукового сигнала для слова «Салом». Для этого звукового сигнала выполнена процедура анализа, т.е. сигнал разбит на фреймы, для каждого фрейма рассчитан набор MFCC векторов. Также у нас имеется запись звукового сигнала, содержащая это же слово, выполнено деление сигнала на фреймы и рассчитаны MFCC вектора. Но в этом случае слово произносилось медленнее, чем в случае с эталонным сигналом и логично

заметить, что эта особенность при анализе даст набор MFCC векторов больший, чем в случае эталона. Возникает вопрос, как выполнять сравнение сигналов? В этом случае приходит на помощь DTW-алгоритм [4–A].

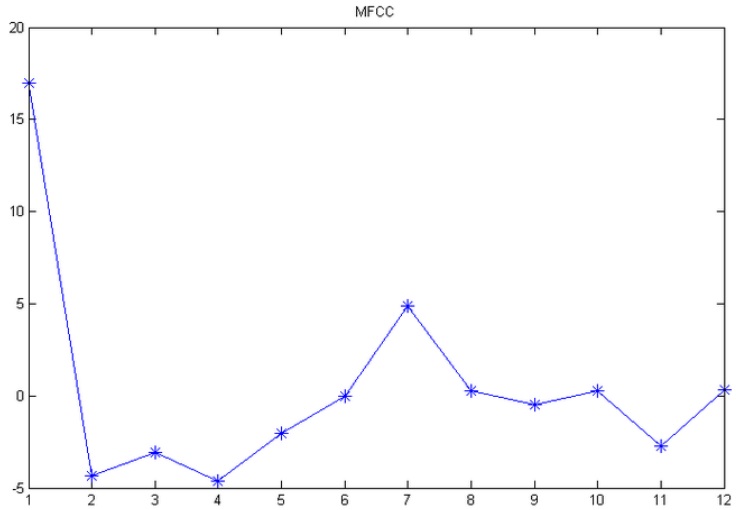


Рисунок 4.6. - Результаты анализа сигнала в MFCC

Алгоритм динамического искажения времени (DTW-алгоритм, от англ. dynamic time warping) — это алгоритм, позволяющий найти наилучшее соответствие между временными последовательностями. Впервые он был использован в распознавании речи, чтобы определить, как два речевых сигнала представляют исходную произнесенную фразу. Впоследствии были запросы и в других областях [4–A].

Временные ряды — это широко используемый тип данных, существующий практически во всех областях науки, и сравнение двух последовательностей является нормальным явлением. Для расчета отклонения достаточно простого измерения расстояния между двумя последовательными компонентами (евклидово расстояние). Однако часто две последовательности имеют примерно одну общую форму, но эти формы не совпадают по оси X. Чтобы определить сходство между этими последовательностями, нам нужно «деформировать» временную ось одной (или обеих) последовательностей и достигнуть лучшего выравнивания.

Приложение Tajik Speech Recognizer. Для реализации распознавания ключевых слов в звуковом потоке на таджикском языке на основе алгоритма динамической трансформации временной шкалы разработано приложение **Tajik Speech Recognizer**. Программные модули написаны с использованием языка программирования Delphi Embarcadero Dev. Ключевая функциональная часть приложения составляет модуль, определяющий расстояние динамической трансформации временной шкалы звуковых потоков, записанных в разное время с одинаковой громкостью.

В процессе тестирования участвовали три диктора, разного пола и возраста. Используя одни и те же звуковые технические средства, дикторы вводили голосовой поток на основе 5 ключевых слов «Тоҷикистон», «Истиқлолият», «Модар», «Ватан», «Душанбе» на таджикском языке.

Таблица 4.10 Значения расстояний у дикторов в DTW

№	Диктор	Пол	Возраст	Слово	DTW
1.	Диктор 1	м	29	Тоҷикистон	902,077
				Истиқлолият	1106,3097
				Модар	562,7853
				Ватан	481,2121
				Душанбе	637,6289
2.	Диктор 2	м	25	Тоҷикистон	919,7111
				Истиқлолият	1075,7343
				Модар	583,0894
				Ватан	584,1706
				Душанбе	957,9518
3.	Диктор 3	ж	19	Тоҷикистон	811,5779
				Истиқлолият	1425,2341
				Модар	778,9961
				Ватан	729,5987
				Душанбе	819,6407

В результате опытной эксплуатации и практических испытаний с ключевыми словами при одинаковой громкости и времени произношения наименьшее расстояние равняется 481,2121. Значения расстояний у первого диктора, который является мужчиной, составляет минимум 481,2121, максимум

1106,3097. А значения расстояний у третьего диктора (женщина) значение расстояния получилось намного больше, то есть 1425,2341. Отклонение в полученных значениях расстояния доказывает, что алгоритм наиболее оптимальным для распознавания звукового потока (см. нижеследующую таблицу) [4–А].

На основе проведенных исследований можно утвердить следующее: на сколько меньше полученное расстояние, тем введенный звуковой поток будет похож на заранее записанный звук. Исходя из этого, следует утверждать, что можно получить эффективное распознавание ключевых слов в речи. Но необходимо учитывать, ряд исключений, а именно время ввода звукового потока и его громкость записи.

Алгоритм динамической трансформации временной шкалы является наиболее простым методом обучения для предварительной подготовки шаблонов звуковых потоков для автоматических систем распознавания речи. Используя возможности алгоритма проводились практические исследования с использованием собственной приложением для распознавания ключевых слов в заранее подготовленной небольшим объемом словарём. Разработанное приложение распознаёт слова с разными длинами с точностью до 87%.

Предположим, что имеются две числовые последовательности (a1, a2,...,an) и (b1,b2,...,bm). Длина последовательности может варьироваться. Выполнение алгоритма начинается с вычисления отклонения между элементами этих двух групп чисел, в которых заключены разные виды отклонений. Наиболее распространённым считается способ, который рассчитывает величину абсолютного отклонения. Путь деформации представляет собой минимально возможную дистанцию между элементами матрицы a_{11} и a_{nm}, которые состоят из a_{ij} элементов, выражающих дистанцию до a_{nm}.

Глобальные деформации включают две последовательности и могут быть определены по формуле:

$$GC = \frac{1}{p}\sum_{i=1}^{p} W_i$$

где w_i – элементы, принадлежащие пути деформации; p – их число. Осуществление расчетов производилось в отношении двух последовательностей. Итоговые значения показаны в таблице 4.11, где отмечена последовательность деформации.

Таблица 4.11 Последовательность деформации сигнала

	-2	10	-10	15	-13	20	-5	14	2
3	5	12	25	37	53	70	78	89	90
-13	16	28	15	43	37	70	78	105	104
14	32	20	39	16	43	43	62	62	74
-7	37	37	23	38	22	49	45	66	71
9	48	38	42	29	44	33	47	50	57
-2	48	50	46	46	40	55	36	52	**54**

Есть несколько условий для алгоритма быстрой сходимости:

1. Монотонность - путь никогда не меняется на обратный, индексы i и j, которые используются последовательно, никогда не уменьшаются.

2. Продолжительность – прогрессивно: за один шаг показатели i и j могут увеличиваться не более чем на 1 балл.

3. Ограничение - начало последовательности находится в левом нижнем углу, а заканчивается в правом верхнем.

4.3. Экспериментальное исследование программной системы

Добавление слова в словарь. Пользователь задает написание слова и привязывает, к слову, варианты произношения, которые анализируется (вычисление дескриптора) и сохраняются системой в базу данных.

Запись варианта произношения слова. Данный этап обработки выполняется при помощи функций класса «TWaveFile».

Пользователь произносит слово, слово записывается в формате WAVE.

Преобразование звукового сигнала в вектор. Данный этап обработки выполняется при помощи функций класса «TWaveFile».

На данном этапе тело Wave-файла переводится в вектор (экземпляр класс «TMatrix»). Эта процедура необходима для того, чтобы исключить неэффективное многократное неудобное чтение (разбор) структуры Wave-файла, а выполнять операции с гибким по функциональности объектом (объектом «TMatrix»). Преобразование выполняется методом «GetWaveVector» класса «TWaveFile». В конце данного этапа мы получаем «Исходный вектор».

Нормализация вектора. Данный этап обработки выполняется при помощи функций класса «TMatrix».

Исходный вектор нормализуется (значения амплитуды сигнала не будут выходить за пределы диапазона [-1; 1]), т.е. находится максимальное значение амплитуды (максимальный элемент), после чего, все элемент вектора делятся на это значение. Это делается для того, чтобы иметь возможность для вычисления относительных оценок и введения пороговых значений.

Удаление тишины в начале и в конце вектора (сигнала). Данный этап обработки выполняется при помощи функций класса «TAnalyzer».

Исходный вектор разделяется на фреймы без перекрытия. Для каждого фрейма рассчитывается его средняя энергия. Те фреймы, энергия которых ниже порогового значения (0,05) считаются фреймами «тишины». Фреймы «тишины» в начале и в конце вектора удаляются.

Вычисление дескриптора для вектора. Данный этап обработки выполняется при помощи функций класса «TAnalyzer».

Дескриптор для исходного вектора (отдельного слова) вычисляется следующим образом: вектор разделяется на фреймы с перекрытием на 50 %, после чего, для каждого фрейма вычисляется MFCC характеристика.

Дескриптор представляет из себя набор MFCC векторов. Дескриптор описывается классом «TDescriptor».

Запись дескриптора в базу данных. Дескриптор записывается в таблицу «Pronunciation» (произношение) в поле «Descriptor» и привязывается к соответствующему слову из таблицы «Word» по ключу.

Запись звукового сигнала. Данный этап обработки выполняется при помощи функций класса «TWaveFile».

Пользователь произносить определенный набор слов, набор слов записывается в формате WAVE. За работу с Wave-фалами в нашем проекте отвечает класс «TWaveFile».

Преобразование звукового сигнала в вектор. Данный этап обработки выполняется при помощи функций класса «TWaveFile».

На данном этапе тело Wave-файла переводится в вектор (экземпляр класс «TMatrix»). Эта процедура необходима для того, чтобы исключить неэффективное многократное неудобное чтение (разбор) структуры Wave-файла, а выполнять операции с гибким по функциональности объектом (объектом «TMatrix»). Преобразование выполняется методом «GetWaveVector» класса «TWaveFile». В конце данного этапа мы получаем т.н. «Исходный вектор».

Нормализация вектора. Данный этап обработки выполняется при помощи функций класса «TMatrix».

Исходный вектор нормализуется (значения амплитуды сигнала не будут выходить за пределы диапазона [-1; 1]), т.е. находится максимальное значение амплитуды (максимальный элемент), после чего, все элемент вектора делятся на это значение. Это делается для того, чтобы иметь возможность для вычисления относительных оценок и введения пороговых значений.

Удаление тишины в начале и в конце вектора (сигнала). Данный этап обработки выполняется при помощи функций класса «TAnalyzer».

Исходный вектор разделяется на фреймы без перекрытия. Для каждого фрейма рассчитывается его средняя энергия. Фреймы, энергия которых ниже

порогового значения (0,05) считаются фреймами «тишины». Фреймы «тишины» в начале и в конце вектора удаляются.

Разделение исходного вектора (сигнала) на слова. Данный этап обработки выполняется при помощи функций класса «TAnalyzer».

Исходный вектор разделяется на фреймы без перекрытия. Для каждого фрейма рассчитывается его средняя энергия. Фреймы, энергия которых ниже порогового значения (0,05) считаются фреймами «тишины». Если участок тишина по продолжительности более 200 миллисекунд, то данный участок является разделителем слов (паузой между словами). После этой операции исходный вектор делится на «сегменты» или «подвектора» (каждый «сегмент» соответствует отдельному слову, т.е. при записи сигнала необходимо выдерживать паузы между словами).

Вычисление дескрипторов для полученных сегментов. Данный этап обработки выполняется при помощи функций класса «TAnalyzer».

Напоминаю, что сегмент – это «подвектор» из «исходного вектора». Дескриптор для конкретного сегмента (отдельного слова) вычисляется следующим образом: сегмент разделяется на фреймы с перекрытием на 50 %, после чего, для каждого фрейма вычисляется MFCC характеристика.

Дескриптор представляет из себя набор MFCC векторов. Дескриптор описывается классом «TDescriptor». Вычисление дескриптора выполняется для каждого сегмента. После того, как для каждого слова вычислен дескриптор, приступаем к поиску эквивалентов вычисленных дескрипторов (наиболее близкого дескрипторов из словаря системы).

Поиск эквивалентов вычисленных дескрипторов в словаре. Данный этап обработки выполняется при помощи функций класса «TRecognizer».

Метод «Recognize» класса «*TRecognizer*» принимает в качестве параметра вычисленный дескриптор и последовательно выполняет его сравнение со всеми дескрипторами, которые хранятся в БД при помощи алгоритма DTW (дескрипторы могут быть различного размера, и евклидово расстояние в этом

случае будет неприменимо). Напоминаю, что алгоритм DTW возвращает дистанцию между дескрипторами.

База данных. База данных представлена следующей схемой:

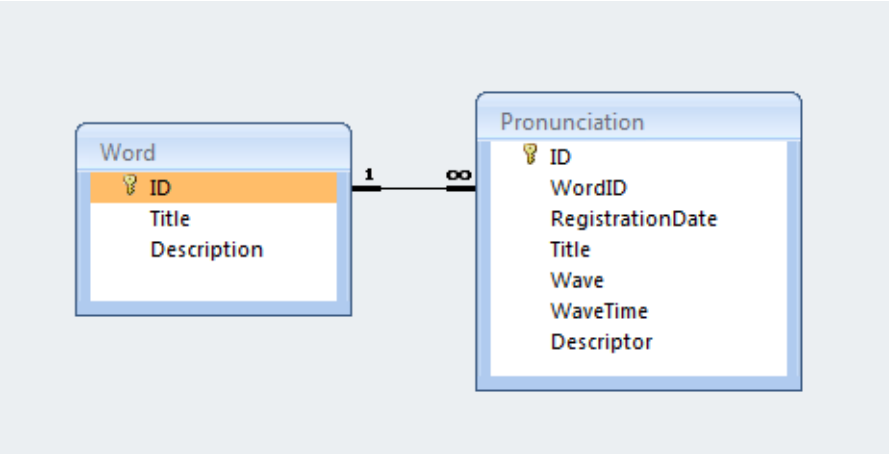


Рисунок 4.7. - Логическая схема базы данных

База данных состоит из следующих таблиц:

- таблица «Word» используется для хранения слов в текстовом формате;
- таблица «Pronunciation» для записи характеристики озвученных слов и дескрипторов.

Из приведенной схемы следует, что для одного слова может быть несколько вариантов произношений и соответственно, несколько дескрипторов.

4.4. Выводы по главе

Разработка комплекса программ поиска ключевых слов в речи происходит в несколько этапов с использованием объектно-ориентрованного программирования. Представленно архитектура программного комплекса, которая включает следующие классы: Класс «TMatrix», Класс «TWaveFile», Класс «TWaveHeader», Класс «TAnalyzer», Класс «TDescriptor», Класс «TRecognizer», Класс «TRecognizerResult», Класс «TMFCC».

Проектирование происходит в первую очередь с представления структуры WAVE файла (системы анализируемых сигналов). Он состоит из двух четко

разделяющихся частей: заголовка файла и области данных. При анализе сигнал разделяется по небольшим промежуткам – фреймам. Фреймы не идут друг за другом, не перекрываются. То есть конец одного фрейма совпадает с началом следующего. Мел-кепстральные коэффициенты (MFCC) является характеристикой речевого сигнала (вычисляется для каждого фрейма). MFCC представляет собой спектральную энергию сигнала.

Рисунок 4.8. - Блок – схема добавления звука в базу данных

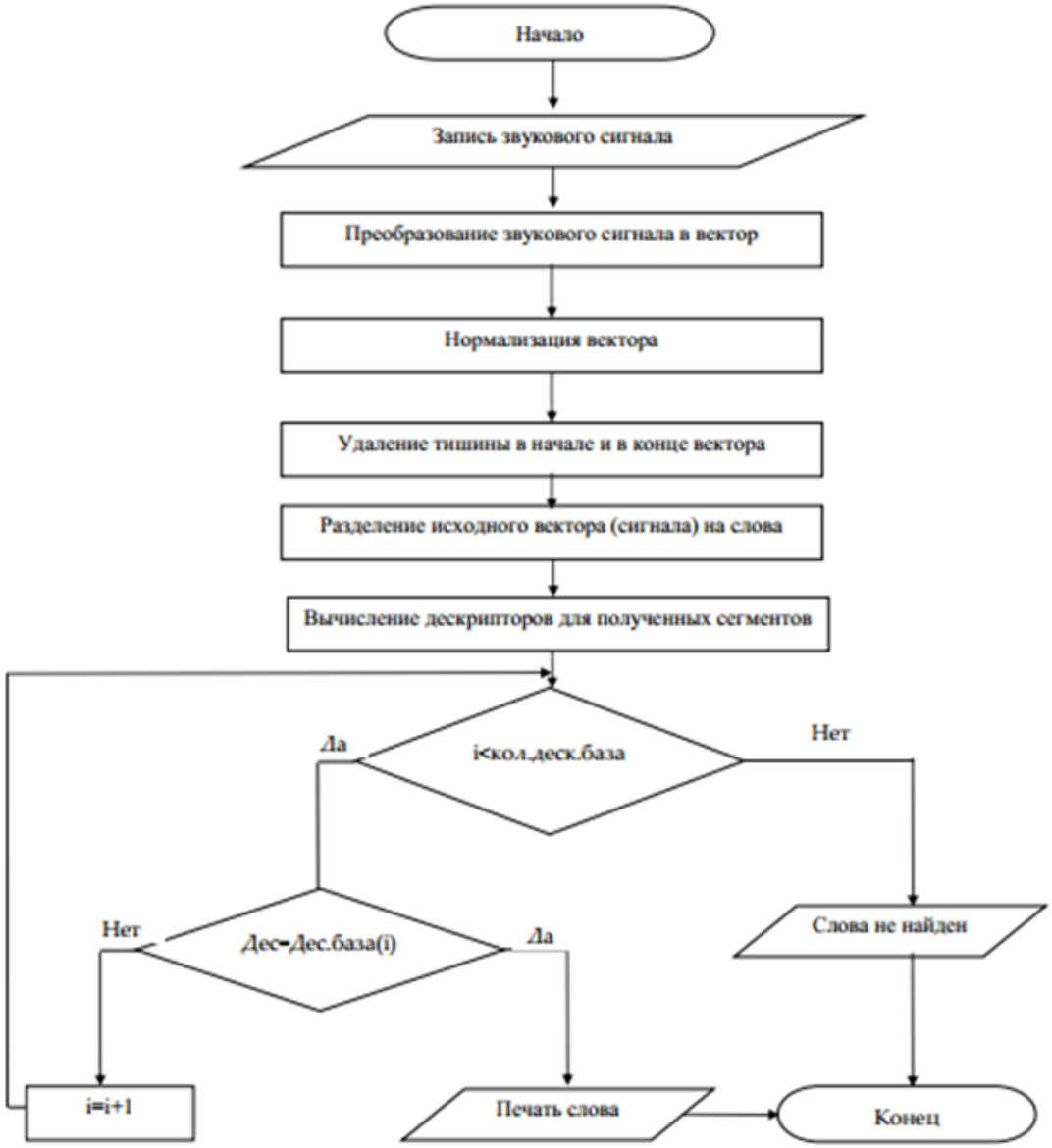


Рисунок 4.9. – Блок-схема при распознавании речи

Таким образом, упорядоченный процесс распознавания речи дает возможность использовать программный комплекс для распознавания таджикской слитной речи, а также воспользоваться данным комплексом и для синтеза речи на таджикском языке.

ЗАКЛЮЧЕНИЕ

Основными результатами диссертации являются:

1. Создание методов, алгоритмов и комплекса программных обеспечений входящих в состав автоматической системы распознавания ключевых слов в речи на таджикском языке, путём обработки речевых сигналов и разделения их на фонемы.

2. Проведен сравнительный анализ результатов разработанных методов распознавания ключевых слов в коллекции голосовых фрагментов из 300 слов озвученных 20 дикторами, что подтвердило эффективность разработанных программных обеспечений.

3. Разработаны математические и компьютерные модели на базе которых получены методы построения системы распознавания ключевых слов на таджикском языке.

4. Разработка новых методов представления ключевых слов с применением скрытой Марковской модели и случайных полей на основе машинного обучения с описанием звуковых сигналов в виде дескрипторов и векторов.

5. Формализация правил распознавания ключевых слов в слитной речи на таджикском языке.

6. Разработка комплекса программ автоматического распознавания ключевых слов в речи на таджикском языке. Результаты диссертации в сочетании с методикой машинного обучения послужили основой для создания программных пакетов распознавания ключевых слов и получены свидетельства об официальной регистрации программы для ЭВМ в отделе по правам и защите прав автора Министерства Культуры Республики Таджикистан (№14, 14.02.2017г.) и свидетельство о государственной регистрации информационного ресурса в Национальном патентно-информационном центре Министерства экономического развития и торговли Республики Таджикистан (№3202200502 от 13.10.2022г).

Программное обеспечение получило применение в реабилитационном центре для детей с ограниченными возможностями и в учебном процессе для повышения качества обучения дисциплин в сфере информационных технологий.

Разработанные модели и программное обеспечение составляют научно-практическую основу для решения актуальных задач компьютерной лингвистики таджикского языка. Результаты представляют широкие возможности применения для решения, как теоретических, так и практических проблем в области речевых технологий.

.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Абрамов Е.Г. Подбор ключевых слов для научной статьи [Текст] / Е.Г.Абрамов // Научная периодика: проблемы и решения. – 2011. – № 2. – С. 35−40.

2. Авсентьев А.О., Применение скрытых марковских моделей для распознавания речи диктора [Текст] / Авсентьев А.О., Лукьянов А.С. // Воронежский институт МВД России, г. Воронеж. –2015. –№2.

3. Агашин О.С., Методы цифровой обработки речевого сигнала в задаче распознавания изолированных слов с применением сигнальных процессоров // О.С. Агашин, О.Н. Корелин. - Труды Нижегородского государственного технического университета им. Р.Е. Алексеева № 4(97). - Радиотехника, системы телекоммуникаций, антенны и устройства СВЧ, 2012. – С. 32-44.

4. Аграновский А. В., Метод текстонезависимой идентификации диктора на основе индивидуальности произношения гласных звуков /А.В. Аграновский, Д.А. Леднов, С.А. Репалов // Акустика и прикладная лингвистика: ежегодник РАО. - 2002. - Вып. 3. - С. 103–115.

5. Алексеев И. В., Митрохин М. А., Кольчугина Е. А. Программное средство оценки эффективности технологий распознавания речи /И.В. Алексеев, М.А. Митрохин, Е.А. Кольчугина // Известия ВУЗов. Поволжский регион. Технические науки. 2018. №3 (47).

6. Алимурадов А. К. Обзор и классификация методов обработки речевых сигналов в системах распознавания речи. / А.К. Алимурадов и [др]. // Измерение. Мониторинг. Управление. Контроль. - 2015, №2. – С. 27-35.

7. Астраханцев Н.А. Автоматическое извлечение терминов из коллекции текстов предметной области с помощью Википедии / Н.А. Астраханцев //Труды Института системного программирования РАН. – 2014. –Т. 26, –№ 4. – С. 7-20.

8. Баканова Н.Б., Исследование ключевых слов как инструмент оптимизации управления электронными документами [Электронный ресурс] /Н.Б. Баканова, И.В. Усманова //Современные проблемы науки и образования:

электрон. науч. журн. – 2014. – № 2. – Режим доступа: http://www.science-education.ru/116-12387

9. Бек К. Экстремальное программирование: разработка через тестирование / К. Бек // СПб.: Питер, 2003.

10. Беленко М.В., Сравнительный анализ систем распознавания речи с открытым кодом /М.В. Беленко, П.В. Балакшин // МНИЖ. –2017. –№4-4 (58)

11. Большакова Е.И. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. пособие /Е.И. Большакова, Э.С. Клышински, Д.В. Ландэ [и др.]. - М.: МИЭМ, 2011. - 272 с.

12. Браславский П.И., Сравнение пяти методов извлечения терминов произвольной длины /П.И. Браславский, Е.А. Соколов //Компьютерная лингвистика и интеллектуальные технологии: по материалам ежегод. Междунар. конф. «Диалог‘2008». Вып. 7(14). – Москва: Изд-во РГГУ, –2008. – С. 67–74.

13. Брауде Э. Технология разработки программного обеспечения /Э. Брауде СПб.: Питер, –2004.

14. Галунов В.И. Помехоустойчивость как системообразующий фактор речи / В.И. Галунов //Проблемы и методы экспериментально-фонетических исследований. – М, –2002. – с.205-300

15. Галяшина Е.И. Речь под микроскопом /Е.И. Галяшина // Компьютерра. – 1999. – №15. – С. 16-24.

16. Гефке Д. А. Применения скрытых марковских моделей для распознавания звуковых последовательностей. / Д.А. Гефке, П. М. Зацепин. // Известия Алтайского государственного университета. – 2012. – С. 72-76

17. Голубинский А. Н. Методика расчета параметров модели речевого сигнала в виде импульса АМ-колебания с несколькими несущими частотами, для случая модуляции суммой гармоник /А.Н. Голубинский // Системы управления и информационные технологии. –2008. № 4.1. С. 156–161.

18. Горшков Ю.Г. Анализ и засекречивание речевого сигнала /Ю.Г. Горшков //Учебное пособие. – М.: Изд. МГТУ им. Н.Э. Баумана, –2007. – 37 с.

19. Гращенко Л.А. Математические основы автоматизированной таджикско-персидской конверсии графических систем письма /Л.А. Гращенко // Автореф. ... дис. к.физ.-мат.н. – Душанбе: ИМ АН РТ. –2010. –19 с.

20. Гребнов С.В. Разработка эффективных методов и комплексов программ распознавания речи в системах человеко-машинного взаимодействия / С.В. Гребнов, канд.техн.наук, дисс. – Иваново, 2010.

21. Гринева М., Анализ текстовых документов для извлечения тематически сгруппированных ключевых терминов /М. Гринева, М. Гринев //Труды Института системного программирования РАН. Т.16. - 2009. - С. 155-165.

22. Гультяева, Т. А. Классификация последовательностей, порожденных близкими скрытыми марковскими моделями, при наличии шума, распределенного по закону Коши / Т. А. Гультяева, А. А. Попов // Материалы российской науч.-технич. конф. Информатика и проблема телекоммуникаций. – 2011. – Т. I. – С. 60-63.

23. Данков Н.И. Исследование возможностей нейросетевых технологий в области идентификации голоса /Н.И. Данков // Экономика и качество систем связи. –2018. –№3

24. Ермилов А.В. Методы, алгоритмы и программы решения задач идентификации языка и диктора / А.В. Ермилов, // дисс. канд. физ.-мат. наук. – М., 2014. – 127 с.

25. Ефремова Н.Э., Терминологический анализ текста на основе лексико-синтаксических шаблонов / Н.Э. Ефремова, Е.И. Большакова, А.А. Носков, В.Ю. Антонов //Компьютерная лингвистика и интеллектуальные технологии: по материалам ежегод. Междунар. конф. «Диалог‘2010». Вып. 9(16). - Москва: Изд-во РГГУ, –2010. – С. 124-129.

26. Жиляков, Е.Г. Сегментация речевых сигналов на основе анализа распределения энергии по частотным интервалам [Текст] / Е.Г. Жиляков, Е.И. Прохоренко, А.В. Болдышев, А.А. Фирсова, М.В. Фатова // Научные ведомости

Белгородского государственного университета. Серия: История. Политология. Экономика. Информатика, Том 18. – 2011. – №7-1 (102). – С. 187-196

27. Захаров В.П., Автоматическое выявление терминологических словосочетаний /В.П. Захаров, М.В. Хохлова //Структурная и прикладная лингвистика. Вып.10. - СанктПетербург: Изд-во С.-Петерб. ун-та, 2014. – С. 182-200.

28. Захаров В.П., Анализ эффективности статистических методов выявления коллокаций в текстах на русском языке /В.П. Захаров, М.В. Хохлова //Компьютерная лингвистика и интеллектуальные технологии: по материалам ежегод. Междунар. конф. «Диалог‘2010». Вып. 9(16). – Москва: Изд-во РГГУ, 2010. – С. 137-143.

29. Камшилова О.Н., Малые формы научного текста: ключевые слова и аннотация (информационный аспект) /О.Н. Камшилова //Известия Российского государственного педагогического университета им. А.И. Герцена. – 2013. – № 156. – С. 106-117.

30. Кретов А.А., Маркемы и ключевые слова в научных текстах /А.А. Кретов //Мир лингвистики и коммуникации: электрон. науч. журн. – 2012. – Т. 1, № 27. – С. 1-13.

31. Кудрявцев К.Я., Спектральный метод поиска ключевых слов в полнотекстовых базах данных /К.Я. Кудрявцев //Информационные технологии. - 2010. - N 4. - С. 2-8.

32. Лукашевич Н.В., Комбинирование признаков для автоматического извлечения терминов /Н.В. Лукашевич, Ю.М. Логачев //Вычислительные методы и программирование. Т.11. - 2010. - С. 108-116.

33. Лучинкина О.И., Карпов О.Н. Моделирование лингвистических уровней системы распознавания слитной речи / О.И. Лучинкина, О.Н. Карпов // Т.15. – Д., 2011. – С. 4.29-4.31.

34. Малинина Ю.В., Автоматическое выявление таксономии в области преобразований программ на основе анализа семантических связей в

публикациях /Ю.В. Малинина //Конструирование и оптимизация параллельных программ. - С. 156-163.

35. Маннинг К.Д., Введение в информационный поиск: пер. с англ. /К.Д. Маннинг, П. Рагхаван // - М. [и др.]: Вильямс, 2011.

36. Матвеев Ю.Н. Цифровая обработка сигналов: учебное пособие. / Ю.Н. Матвеев, К.К. Симончик, А.Ю. Тропченко, М.В. Хитров. //– СПб.: Изд. СПбНИУ ИТМО, –2013. С. 166.

37. Модели и методы распознавания речи: сборник. М.: Вычислительный центр им. А. А. Дородницына РАН. –2010. С. 135.

38. Музычук Д.С. Сегментация, шумоподавление и фонетический анализ в задаче распознавания речи [Текст] / Музычук Д.С., Медведев М.С. // Молодой ученый. – 2013 – №6. – С. 86-96.

39. Назаров М. В., Методы цифровой обработки и передачи речевых сигналов / М. В. Назаров, Ю. Н. Прохоров // М.: Радио и связь, 1985. – 176 с.

40. Насыпный В.В. Распознавание речи на основе интеллектуальных систем. М.: МПГУ, –2010. С. 58.

41. Патент РФ № 2230375: МПК G 10 L 15/00, G 10 L 17/00. Метод распознавания диктора и устройство для его осуществления / П. В. лабутин, А. Н. Раев, С. л. Коваль – № 2002123509/09; заявл. 03.09.02; опубл. 10.06.04.

42. Попова С.В. Извлечение ключевых словосочетаний /С.В. Попова, И.А. Ходырев // Научно-технический вестник Санкт-Петербургского государственного ун-та информационных технологий, механики и оптики. – 2012. – № 1(77). – С. 68-72.

43. Послание Президента Республики Таджикистан, Лидера нации, уважаемого Эмомали Рахмона Маджлиси Оли [Электронный ресурс] URL: http://prezident.tj/ru/node/21977

44. Потапова Р.К. Речь: коммуникация, информация, кибернетика. М.: Радио и связь, 1997 (2 доп. изд. 2001; 3 доп. изд. 2003; 4 доп. изд. 2007).

45. Потапова Р.К., Потапов В.В. Язык, речь, личность /Р.К. Потапова, В.В. Потапов // М.: Языки славянской культуры, –2006

46. Программа DETware. Национальный институт стандартов США, NIST, http://www.nist.gov.

47. Разумихин Д.В. Разработка системы понимания устной речи в диалоге / Международная конференция по компьютерной лингвистике "Диалог 2001", – 2001. – с.323-329.

48. Рассказова С. И., Метод форматного анализа на основе вейвлет-преобразования в системах распознавания речи /С. И. Рассказова, А. И. Власов // IX Научно-техническая конференция «Наукоемкие технологии и интеллектуальные системы»: Сборник трудов. Москва: МГТУ им. Н. Э. Баумана, –2007. С. 38-43.

49. Ролдугин С. В. Модели речевых сигналов для идентификации личности по голосу / С. В. Ролдугин, А. Н. Голубинский, Т. А. Вольская, // Радиотехника. –2002. № 11. С. 79–81.

50. Ронжин А.Л., Автоматическое распознавание русской речи / А.Л. Ронжин, И. В. Ли //Вестник Российской академии наук, –2007. том 77. –№ 2. –С. 133-138.

51. Рубцова Ю.В., Методы автоматического извлечения терминов в динамически обновляемых коллекциях для построения словаря эмоциональной лексики на основе микроблоговой платформы Twitter / Ю.В. Рубцова, //Доклады ТУСУРа. – Томск, –2014. - № 3(33). – С. 140-144.

52. Савченко В.В. Результаты натурных испытаний метода фонетического декодирования слов в задачах распознавания и диаризации разговорной русской речи / В.В. Савченко // Информационные системы и технологии. Известия Орел ГТУ. –2013. № 1 (75). С. 12-21.

53. Савченко Л. В. Алгоритм пофонемного распознавания устной речи на основе метода нечеткого фонетического кодирования-декодирования слов // Информационно-управляющие системы. –2014. №1 (68).

54. Сагациян, М.В. NN-SCG speech recognition – научноисследовательская программа по изучению алгоритмов нейросетевого дикторонезависимого распознавания речевых команд / М.В. Сагациян, Г.С. Тупицин // Свидетельство

о государственной регистрации программы для ЭВМ № 2015616920 от 30 апреля –2015г.

55. Саймиддинова Д., Таджикско-русский словарь, т.2 /Под ред. Д. Саймиддинова, С.Д. Холматовой, С. Каримова. – Душанбе: Дониш, 2005, 461 с.

56. Сахарный Л.В., Набор ключевых слов как тип текста /Л.В. Сахарный, А.С. Штерн //Лексические аспекты в системе профессионально-ориентированного обучения иноязычной речевой деятельности. – Пермь: Перм. политехн. ун-т, –1988. – С. 34-51.

57. Сиротко-Сибирский, С.А. К измерению качества работы предметизатора / С.А. Сиротко-Сибирский, А.С. Штерн //Предметный поиск в традиционных и нетрадиционных информационно-поисковых системах. Вып.8. - Л.: ГПБ, –1988. - С. 201-219.

58. Созыкин А. В. Обзор методов обучения глубоких нейронных сетей // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. –2017. №3

59. Соколов А.Н. Внутренняя речь и понимание //Ученые записки государственного научно-исследовательского ин-та психологии. – М., –1941. – Т.2. –С. 99-146.

60. Солонина А.И. Основы цифровой обработки сигналов. Курс лекций: Учебное пособие, 2-е изд. // А.И. Солонина, Д. Улахович, С. Арбузов, Е. Соловьева. – СПб.: БХВ-Петебург, –2012.

61. Сорокин В. Н. Синтез речи / В. Н. Сорокин // М.: Связь, –1992. – С. 392 с.

62. Сорокин В. Н. Фундаментальные исследования речи и прикладные задачи речевых технологий /В. Н. Сорокин // Речевые технологии. –2008. –№1. С. 18-48.

63. Сорокоумова Д. А., Корелин О. Н., Сорокоумов А. В. Построение и обучение 56 нейронной сети для решения задачи распознавания речи / Д. А. Сорокоумова, О. Н. Корелин, А. В. Сорокоумов // Труды НГТУ им. Р.Е. Алексеева. –2015. –№3 (110).

64. Статистика упоминания ключевого слова “hidden Markov models” между 1800 и 2008 годами, полученная с помощью сервиса Google Ngram Viewer [Электронный ресурс]. – Режим доступа: https://books.google.com/ngrams/graph?content=hidden+Markov+models&year_start=1800&year_end=2008&corpus=15&smoothing=3&share=&direct_url=t1%3B%2Chidden%20Markov% 20models%3B%2Cc0

65. Суворов В.Н. О кепстральном анализе в популярной форме // В.Н. Суворов. – СПб.: «Ви Тэк». –№4, 2006. – С. 52-53.

66. Таджикско-русский словарь, т.1 /Под ред. С.Д.Холматовой, С. Солехова, С. Каримова. – Душанбе: Дониш, 2004, 388 с. ;

67. Тимофеев Е.Н. Применение автоматизированной системы «диалект» на базе компьютерной речевой лаборатории CSL (США) при решении задач идентификации дикторов / Е.Н. Тимофеев //Учебное пособие. – М.: Изд. ЭКЦ МВД России, –2000. – 120 с.

68. Тупицин, Г.С. Повышение качества закрытой текстонезависимой идентификации диктора в условиях шумов с помощью бинарных масок / Г.С. Тупицин, М.В. Сагациян // Докл. 12-й междунар. научно-технической конф. «Оптико-электронные приборы и устройства в системах распознавания образов, обработки изображений и символьной информации». – Курск, – 2015.

69. Усманов З.Д., Довудов Г.М. – ДАН РТ, 2010, т. 53, № 4, с. 257-262.

70. Фаулер М. UML / М. Фаулер Основы, 3-е издание. СПб.: Символ-Плюс, –2004.

71. Фирсова А.А. О различиях распределения энергии звуков русской речи и шума / А.В. Болдышев, А.А. Фирсова // Материалы 12-ой Международной конференции и выставки "Цифровая обработка сигналов и её применение. – "DSPA'2010". – Москва, 2010. – С. 204–207.

72. Фролов А., Фролов Г., Синтез и распознавание речи. Современные решения [Электронный ресурс] / Александр Фролов, Григорий Фролов. – Электрон. журн. – 2003. – Режим доступа: http://www.frolov-lib.ru

73. Хохлова М.В., Экспериментальная проверка методов выделения коллокаций // Инструментарий русистики: корпусные подходы. - Хельсинки, 2008. – С. 343-356.

74. Худобердиев Х.А. Комплекс программ синтезирования таджикской речи / Х.А. Худобердиев, автореф. дисс.кандидата физ.-мат.наук. – Душанбе, 2009.

75. Цзинбинь Янь, Поиск ключевых слов в слитной речи на основе усовершенствованной меры достоверности /Янь Цзинбинь, У Ши, А.В. Ткаченя, И.Э. Хэйдоров. – Вестник БГУ. – Сер.1 2009. - №3. – С. 44-48.

76. Черник, Н. Н. Сегментация спонтанной речи в языках различных типов/ Н.Н. Черник// Вестник Белорусского государственного экономического университета. - 2009 - N 4 - С. 101-107.

77. Чистович Л. А., Физиология речи. Восприятие речи человеком / Л. А. Чистович, А. В. Венцов, М. П. Грамстрем и др. // М.: Наука, –1976. –С 388.

78. Шереметьева, С.О. Методы и модели автоматического извлечения ключевых слов / С.О. Шереметьева, П.Г. Осминин //Вестник Южно-Уральского государственного ун-та. –2015. – № 1, т.12. – С. 76-81.

79. Электронный ресурс http://archive.ics.uci.edu

80. Электронный ресурс http://archive.ics.uci.edu/ml/datasets/Character+Trajectories

81. Электронный ресурс режим доступа URL: http://www.sl-systems.ru/ - официальный сайт компании «Гран При».

82. Якобсон А., Унифицированный процесс разработки программного обеспечения / А. Якобсон, Г. Буч, Дж. Рамбо //СПб.: –Питер, –2003.

83. Якушев Д. И., Скляров О. П. Моделирование гласных звуков / Д.И. Якушев, О. П. Скляров // Акустический журнал. –2003. –Т. 49. № 4. – С. 567–569.

84. «Bill Gates predictions about speech recognition a historical review.htm» in Matthew Paul Thomas blog at http://mpt.net.nz/archive/2005/12/30/gates.

85. A. Egozi, IAI's REX robot to face first operational trial, [Электронный ресурс] // Israel Defence, 2012. URL: http://www.israeldefense.com/?CategoryID=411&ArticleID=1515.

86. B. Johnson, How Siri works, [Электронный ресурс] // How Stuff Works. URL: http://electronics.howstuffworks.com/gadgets/high-techgadgets/siri2.htm.

87. Bazzi, I, Glass, J. Modeling out of vocabulary words for robust speech recognition / I. Bazzi, J. Glass. // Proc. ICASSP –2000, Beijing, China, Vol. 1, pp.401-404.

88. Bourlard H., Connectionist Speech Recognition. A Hybrid Approach / H. Bourlard, N. Morgan // The Kluwer International Series in Engineering and Computer Science, Vol. 247, Kluwer Academic Publishers, Boston, –1994.

89. Casale P. Personalization and user verification in wearable systems using biometric walking patterns / P. Casale, O. Pujol, P. Radeva // Personal and Ubiquitous Computing. – 2012. – Vol. 16, –№5. – pp. 563-580.

90. Cohen M., The DECIPHER speech recognition system / M. Cohen, H. Murveit, H. Bernstein, P. Price, M. Weintraub // IEEE ICASSP, Albuquerque, –1990. pp. 77-80.

91. Englund C., Speech recognition in the JAS 39 Gripen aircraft – adaptation to speech at different G-loads. Master degree thesis / C. Englund, // Royal Institute of Technology, Stockholm, –2004.

92. Franco H., Context-dependent connectionist probability estimation in a hybrid hidden Markov model-neural net speech recognition system / H. Franco, M. Cohen, N. Morgan, D. Rumelhart, V. Abrash // Computer Speech and Language. – 1994. –8. –pp. 211- 222.

93. Franzini M.A., Connectionist Viterbi training: a new hybrid method for continuous speech recognition / M.A. Franzini, K.F. Lee, A. Waibel // IEEE ICASSP –1990, pp. 425-428.

94. Gaurav D., Development of Application Specific Continuous Speech Recognition System in Hindi / D. Gaurav, D. Shakina, K.S. Gopal, B. Mahua // Journal of Signal and Information Processing. –2012. No. 3. P. 394-401.

95. Henneberg J., Estimation of global posteriors and forward-backward training of hybrid HMM/ANN systems /J. Henneberg, C. Ris, H. Bourlard, S. Renals, N. Morgan // Proceedings of EUROSPEECH, –1997. Vol. 4, Rhodi, pp. 1951-1954.

96. Hochberg M., ABBOT: the CUED hybrid connectionist-HMM large vocabulary recognition / M. Hochberg, S. Renals & A. Robinson.

97. Joel Pinto, H.N. V. Sitaram. Confidence measures in speech recognition based on probability distribution of Likelihoods. http://www.hpl.hp.com/techreports/2005/HPL-2005-144.pdf

98. Kershaw D. J., Context dependent classes in a hybrid recurrent network-HMM speech recognition system / D.J. Kershaw, M.M. Hochberg, A.J. Robinson // Cambridge University Engineering Department, Technical Report, CUED/FINFENG. TR. 217. –1995.

99. Khreich, W. A survey of techniques for incremental learning of HMM parameters / W. Khreich, E. Granger, A. Miri, R. Sabourin // Inf. Sci. – 2012. – Vol. 197. – pp. 105-130.

100. Lawrence R. Rabiner and Biing-Hwang Juang. Fundamentals of Speech Recognition. — Prentice Hall, –1993

101. Lindasalwa M. Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient and Dynamic Time Warping Techniques / M. Lindasalwa // Journal of Computing. –2010. Vol. 2, Issue 3. P. 138-143.

102. Lou Boves, Johan Koolwaaij. Weighting phone confidence measures for automatic speech recognition / Boves Lou, Koolwaaij Johan. // Proc. COST 249, Gent, Belgique, May 2000

103. M. Pinola, Speech recognition through the decades: how we ended up with Siri, [Электронный ресурс] // PCWorld, 2011. URL: http://www.techhive.com/article/243060/speech_recognition_through_the_decades_how_we_ended_up_with_siri.html?page=0

104. Mc Cullagh P., Generalized Linear Models / P. Mc Cullagh, J. A. Nelder // –London. Chapman and Hall. –1983.

105. Morgan N., Hybrid neural network/ hidden Markov model system for continuous speech recognition / N. Morgan, H. Bourlard // Intl. Journal of Pattern Recognition and Arificial Intelligence, Special Issue on Advances in Pattern Recognition Systems.

106. Morist M.U. Emotional speech synthesis for a radio dj: corpus design and expression modeling: master thesis MTG-UPF dissertation / M.U. Morist. – Barcelona –2010.

107. Munteanu, D. P. Automatic speaker verification experiments using HMM / D. P. Munteanu, S. A. Toma // International Conference on Communications (COMM). – 2010. – pp. 107-110.

108. Nguyen Q.C., Pham Thi N.Y., Castelli E. Shape vector characterization of Vietnamese tones and application to automatic recognition // ASRU, 2001. P. 231-247.

109. Prudnikov A. Improving Acoustic Models For Russian Spontaneous Speech Recognition / A. Prudnikov, I. Medennikov, V. Mendelev, M. Korenevsky, Y. Khokhlov // Speech and Computer, Lecture Notes in Computer Science. — 2015. — Vol. 9319. — P. 234–242

110. Rachna Vijay Vargiya. Keyword spotting using normalization of posterior probability of confidence measures. Ms. Thesis in Computer Science, 2005, USA.

111. Robinson T., Hochberg M., Renals S. The use of recurrent neural networks in continuous speech recognition // In: C.H. Lee, F.K. Soong, K.K. Paliwal (Eds), Automatic Speech and Speaker Recognition: Advanced Topics, The Kluwer International Series in Engineering and Computer Science, Kluwer Academic Publishers. –Boston. –USA. –1996.

112. Rossia, A. Volatility estimation via hidden Markov models / A. Rossia, G. M. Gallob // Journal of Empirical Finance. – 2006. – Vol.13, №2. – pp. 203-230.

113. Schmitt A., Speech recognition for mobile devices / A. Schmitt, D. Zaykovskiy, W. Minker, // International Journal of Speech Technology, Springer, – 2008. Vol. 11, Pp. 63-72.

114. Stylianou Y. // 3rd ESCA Speech Synthesis Workshop, –Nov. –1998.

115. Stylianou Y. Apply the harmonic plus noise model in concatenative speech synthesis // IEEE Trans. on Speech and Audio Process. –2001. –Vol. 9. –№ 1. –P. 21–29.

116. Sui. M, Gish, H. Evaluation of word confidence for speech recognition systems / M. Sui, H. Gish, //Computer Speech and Language, –Vol. 13. pp. 299- 319.

117. Timothy J. Hazen, Stephanie Seneff and Joseph Polifroni. Recognition confidence scoring and its use in speech understanding systems / J. Timothy, S. Stephanie and P. Joseph // Computer Speech and Language. –2002. –Vol. 16. –pp. 49-67.

118. Vimala C. A Review on Speech Recognition Challenges and Approaches // World of Computer Science and Information Technology Journal. –2012. –Vol. 2, No. 1. –pp. 1-7.

119. Vu Q., Demuynck K., Compernolle D.V., Vietnamese automatic speech recognition: the FlaVoR approach. Singapore: ISCSLP Kent Ridge, –2006. –pp. 155.

120. Xuedong H., Alex A., Hsiao-Wuen H. Spoken Language Processing: A Guide to Theory, Algorithm and System Development. Prentice Hall, –2001.

121. Yannis Styliano // IEEE Transactions on Speech and Audio Processing. January –2001. –Vol. 9, No. 1,

122. Young S.J., Multilingual large vocabulary speech recognition: the European SQALE project / S.J. Young, M. Adda-Dekker, X. Aubert // Computer Speech and Language, –1997, 11, –pp. 73-89.

123. Young S.j., The HTK BOOK. Ver. 2.1. Cambridge University, –1997.

124. Zavarehei E., Vaseghi S., Yan Q. Noisy speech enhancement using harmonic-noise model and codebook-based post-processing / E. Zavarehei, S. Vaseghi, Q. Yan // IEEE Trans. on Speech and Audio Process. –2007. –Vol. 15. –№ 4. –pp. 1194–1203.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в научных журналах Перечня ВАК при Президенте республики Таджикистан:

[1–А] **Б.Х. Ашурзода.** Проблемы распознавания слитной речи и поиска ключевых слов / **Б.Х. Ашурзода** // Вестник таджикского национального Университета Серия естественных наук. - Душанбе. - 2018. № 2 (33). - С. 53-57.

[2–А] **Б.Х. Ашурзода**. О проблемах формирования речевой базы для системы распознавания речи на таджикском языке/ **Б.Х. Ашурзода** // Политехнический вестник. Серия Интеллект. Инновации. Инвестиции. - Душанбе. - 2021. №3 (55) - С. 74-76.

[3–А] **Б.Х. Ашурзода**. Моделирование процесса распознавания речи в контексте таджикской язычной речи/ **Б.Х. Ашурзода**, Х.А. Худойбердиев // Политехнический вестник. Серия Интеллект. Инновации. Инвестиции. - Душанбе. - 2022. № 2 (58) - С. 39-42. (на таджикском языке)

[4–А] **Б.Х.Ашурзода**. Применение алгоритма динамической трансформации временной шкалы для распознавания ключевых слов в звуковом потоке на таджикском языке / **Б.Х. Ашурзода** // Вестник технологического университета Таджикистана. - Душанбе. - 2022. № 3 (50). - С. 132-136.

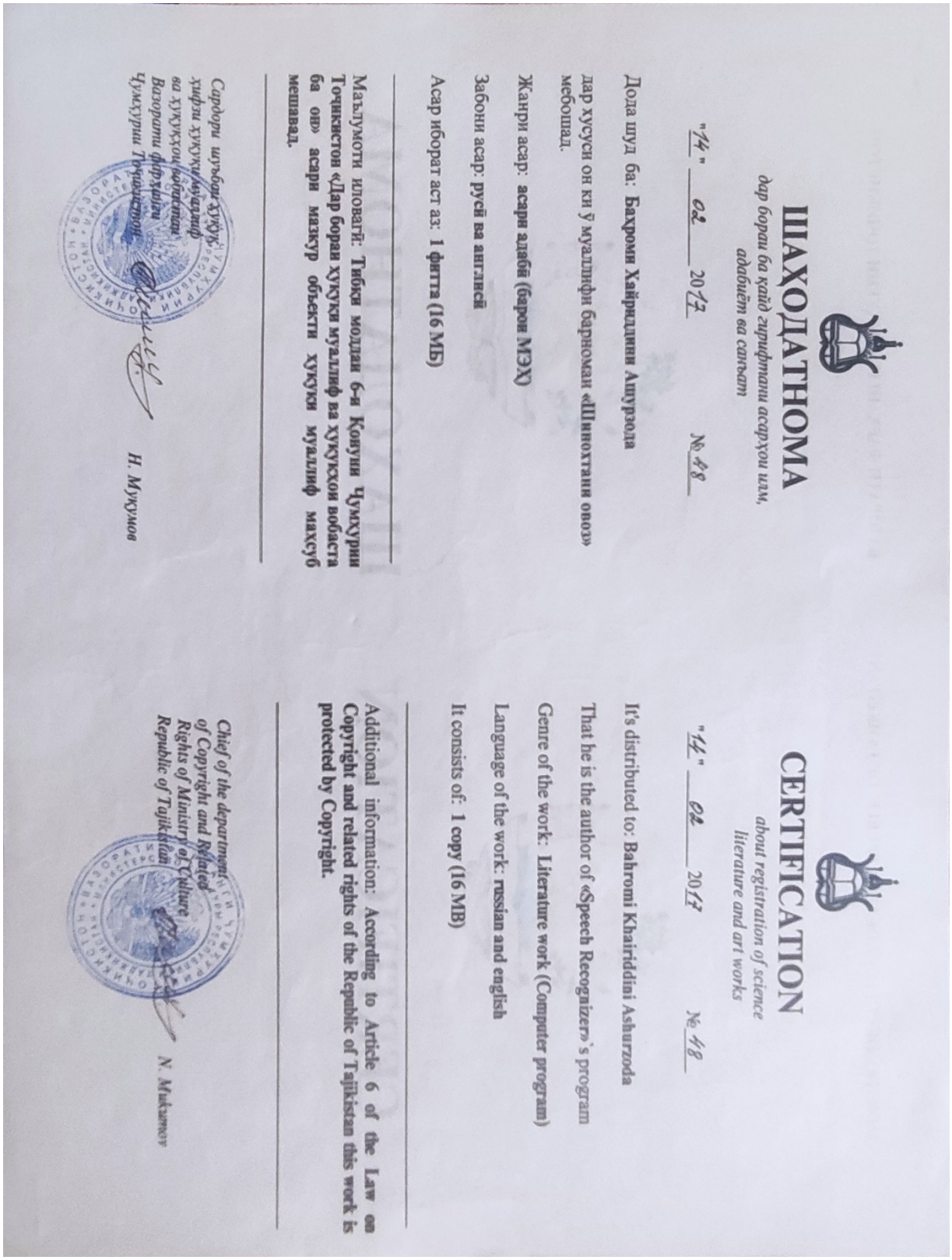
Авторские права и свидетельство:

[5–А] **Б.Х. Ашурзода.** Свидетельство. Распознавание речи / **Б.Х. Ашурзода** // Свидетельство об официальной регистрации программы для ЭВМ в отделе по правам и защиты прав автора Министерства Культуры Республики Таджикистан; свидетельство №14 (зарегистрирован 14 февраля 2017г.).

[6–А] **Б.Х. Ашурзода**. Свидетельство. Автоматическая распознавание речи человека в таджикском языке / **Б.Х. Ашурзода** // Свидетельство о государственной регистрации информационного ресурса №3202200502 от 13.10.2022г. внесен в реестр информационных ресурсов Республики Таджикистан.

ПРИЛОЖЕНИЯ

Свидетельство. Распознавание речи / Ашурзода Бахром Хайриддин // Свидетельство об официальной регистрации программы для ЭВМ в отделе по правам и защиты прав автора Министерства Культуры Республики Таджикистан; свидетельство №14 (зарегистрирован 14 февраля 2017 г.).



ШАҲОДАТНОМА

дар бораи ба қайд гирифтани асарҳои илм, адабиёт ва санъат

"14" 02 2017 № 48

Дода шуд ба: **Бахроми Хайриддини Ашурзода**

дар хусуси он ки ӯ муаллифи барномаи **«Шиносхтани овоз»** мебошад.

Жанри асар: **асари адабӣ (барои МЭҲ)**

Забони асар: **русӣ ва англисӣ**

Асар иборат аст аз: **1 файл (16 МБ)**

Маълумоти иловагӣ: **Тибқи моддаи 6-и Қонуни Ҷумҳурии Тоҷикистон «Дар бораи ҳуқуқи муаллиф ва ҳуқуқҳои вобаста ба он» асари мазкур объекти ҳуқуқи муаллиф маҳсуб мешавад.**

Сардори шуъбаи ҳуқуқ, ҳифзи ҳуқуқи муаллиф ва ҳуқуқҳои вобаста Вазорати фарҳанги Ҷумҳурии Тоҷикистон *Н. Мукумов*

CERTIFICATION

about registration of science literature and art works

"14" 02 2017 № 48

It's distributed to: **Bahromi Khairiddini Ashurzoda**

That he is the author of **«Speech Recognizer»**'s program

Genre of the work: **Literature work (Computer program)**

Language of the work: **russian and english**

It consists of: **1 copy (16 MB)**

Additional information: **According to Article 6 of the Law on Copyright and related rights of the Republic of Tajikistan this work is protected by Copyright.**

Chief of the department of Copyright and Related Rights of Ministry of Culture Republic of Tajikistan *N. Mukumov*

Свидетельство. Автоматическая распознавание речи человека в таджикском языке / Ашурзода Бахром Хайриддин // Свидетельство о государственной регистрации информационного ресурса №3202200502 от 13.10.2022г внесен в реестр информационных ресурсов Республики Таджикистан.

ВАЗОРАТИ РУШДИ ИКТИСОД ВА САВДОИ ЧУМХУРИИ ТОҶИКИСТОН
МУАССИСАИ ДАВЛАТИИ «МАРКАЗИ МИЛЛИИ ПАТЕНТУ ИТТИЛООТ»

МИНИСТЕРСТВО ЭКОНОМИЧЕСКОГО РАЗВИТИЯ И ТОРГОВЛИ РЕСПУБЛИКИ ТАДЖИКИСТАН
ГОСУДАРСТВЕННОЕ УЧРЕЖДЕНИЕ «НАЦИОНАЛЬНЫЙ ПАТЕНТНО- ИНФОРМАЦИОННЫЙ ЦЕНТР»

ШАҲОДАТНОМА

дар бораи бақайдгирии давлатии захираи иттилоотӣ

СВИДЕТЕЛЬСТВО

о государственной регистрации информационного ресурса

Номгӯй Наименование **Автоматическая система распознавания речи человека в таджикском языке**
Сарзамин Страна **Республика Таджикистан**
Доранда Владелец **Ашурзода Бахром Хайриддин**
Тахиягар Разработчик **Ашурзода Бахром Хайриддин**
№ қайди давлатӣ № государственной регистрации **№ 3202200502**

Ба Феҳристи давлатии захирахои иттилоотии Чумхурии Тоҷикистон дохил карда шудааст
Внесен в Государственный реестр информационных ресурсов Республики Таджикистан **13 октября 2022 г.**

Директор **Исмоилзода М.Х**

«УТВЕРЖДАЮ»
Директор института технологий
и инновационного менеджмента
в г. Куляб, к.т.н., доцент
_____ Шоев А.Н.
«16» 05 2022

СПРАВКА
о внедрении в учебный процесс результатов диссертационного исследования

Результаты диссертационной работы Ашурзода Бахром Хайриддин на тему «Методы и модели поиска ключевых слов в речи на таджикском языке (спектральный анализ – особенности)» внедрены в учебный процесс Института технологий и инновационного менеджмента в г. Куляб, в частности:

1. При изучении студентами факультета экономики и информационных технологий обучающихся по специальности 1-400101 – «Программное обеспечение информационных технологий» и 1-40010202 – «Системы и информационных технологий (в экономики)».

2. Внедрены в учебно-методический комплекс и рабочие программы дисциплина «Основы алгоритмов и программирования».

Использование указанных результатов позволяет повысить качество освоения вышеуказанных дисциплин с учетом современных научных и практических требований.

Предложенные при участии автора «Разработанные методы и модели поиска ключевых слов в речи на таджикском языке» нашло отражение также и в курсах переподготовки и повышения квалификации специалистов, выполнении курсовых работ, при подготовке ВКР бакалавров и диссертации магистрантов, в рамках получения второго образования, в форме тренинг-курсов, а также при выполнении научно-исследовательских работ студентами.

Заместитель директора по учебной работе
ИТИМК, кандидат сельскохозяйственных
наук, доцент Раджабов И.Х,

Начальник учебного отдела ИТИМК
кандидат технических наук, доцент Курбонов Б.Д.

«Согласовано»
Директор Институт технологий
и инновационного менеджмента
в городе Куляб(ИТИМК)
_____ Шоев А.Н.
06 2022

«Утверждаю»
Директор ОО Заршедабону в
городе Куляб
_____ Курбонова З.Ф.
06 2022

АКТ

о внедрении результатов диссертационной работы

Ашурзода Бахром Хайриддин на тему

«Методы и модели поиска ключевых слов в речи на таджикском языке (спектральный анализ – особенности)» и свидетельство №17 от 14.02.2017г. в отделе по правам и защиты прав автора Министерства Культуры Республики Таджикистан

Мы, нижеподписавшимся комиссия в составе: Менеджер центра Каримова Ибодат с одной стороны и кандидат физика-математический наук доцент Худойбердиев Хуршед Атохонович, соискатель кафедры Ашурзода Бахром Хайриддин, заведующей кафедры технология программирования и компьютерный техника кандидат педагогический наук И.О доцент Шарипов Бегиджон Рамазонович и начальник технический центр и ремонт компьютер Усмонов Фирдавс Намозалиевич составили настоящий акт о том, что в период апрель – мая месяцев 2022г. проводили испытания комплекс программу в Реабилитационный центр для детей с ограничениями возможностями Парасту и утверждаем что это комплекса можно применят в нашу работах с детьми ограничениями возможностями.

Председатель ОО Заршедабону
Курбонова З.Ф., _____
Менеджер центр Парасту
Каримова И._____

Представители: от Института технологий и инновационного менеджмента в городе Куляб
Худойбердиев Х.А., _____
Ашурзода Б.Х., _____
Шарипов Б.Н., _____
Усмонов Ф.Н., _____