

На правах рукописи
عبدالقهي الذيباني

Аль-Дайбани Абдулгани Мохаммед Салех

**ИССЛЕДОВАНИЕ МЕТОДОВ И РАЗРАБОТКА АЛГОРИТМОВ
ОБРАБОТКИ СИГНАЛОВ ДЛЯ СИСТЕМ АВТОМАТИЧЕСКОГО
РАСПОЗНАВАНИЯ ТЕЛЕФОННОЙ РЕЧИ В РЕСПУБЛИКЕ ЙЕМЕН**

Специальность 05.12.13 – Системы, сети и устройства телекоммуникаций

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Владимир 2019

Работа выполнена на кафедре радиотехники и радиосистем в ФГБОУ ВО
«Владимирский государственный университет
имени Александра Григорьевича и Николая Григорьевича Столетовых»

Научный руководитель **Левин Евгений Калманович,**
доктор технических наук, доцент, профессор кафедры
радиотехники и радиосистем Владимирского
государственного университета имени Александра
Григорьевича и Николая Григорьевича Столетовых

Официальные оппоненты: **Туляков Юрий Михайлович,**
доктор технических наук, доцент, доцент Волго-
Вятского филиала Московского технического
университета связи и информатики (ВВФ
МТУСИ), г. Нижний Новгород.

Кучин Сергей Игоревич,
кандидат технических наук, ведущий специалист,
АО "КОБРА", г. Владимир.

Ведущая организация: ФГБОУ ВО Ковровская государственная
технологическая академия имени В.А. Дегтярева.

Защита диссертации состоится «25» декабря 2019 г. в 16.00 часов на заседании диссертационного совета Д212.025.04 при Владимирском государственном университете имени Александра Григорьевича и Николая Григорьевича Столетовых по адресу: г. Владимир, пр-т Строителей, 3/7, ауд. 301-3.

С диссертацией можно ознакомиться в научной библиотеке Владимирского государственного университета имени Александра Григорьевича и Николая Григорьевича Столетовых и на сайте <http://diss.vlsu.ru>

Автореферат разослан «24» октября 2019 г.

Отзывы на автореферат, заверенные печатью, просим направлять по адресу: 600000, г. Владимир, ул. Горького, 87, ВлГУ, ФРЭМТ.

Ученый секретарь диссертационного совета
доктор технических наук, профессор



А. Г. Самойлов

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы

Развитая телефонная сеть в республике Йемен и современный уровень вычислительной техники создают предпосылки использования систем автоматического распознавания речи (САРР) в телефонии. Использование САРР обеспечивает простой доступ широких слоев населения к автоматическим справочным и регистрационным системам. Однако беспокойная политическая обстановка в стране и отсутствие инженерных и научных кадров соответствующей квалификации не позволяют до сих пор построить соответствующие системы.

Следует учесть, что особенности арабского языка повышают сложность создания САРР по сравнению с аналогичными системами, используемыми, например, в Европе и США. В частности, разговорный арабский язык характеризуется множеством диалектов.

Наличие диалектов повышает степень изменчивости речи, что обуславливает увеличение числа ошибок распознавания. В ряде работ, показано, что достоверность распознавания диалектного арабского языка (ДАЯ) значительно повышается при использовании автоматической идентификации диалектов в составе САРР. Однако исследования в области идентификации арабских диалектов распознавания очень малочисленны и сориентированы на национальные диалекты отдельных стран. Отсутствуют какие-либо данные о расчете вероятности ошибки идентификации диалекта. Представлены лишь экспериментальные данные об идентификации конкретных диалектов. В Йемене исследования по созданию идентификаторов диалектов не проводились.

Снижение достоверности распознавания во многом обусловлено отличием частотной характеристики (ЧХ) канала связи, который использовался при создании звукозаписей, предназначенных для обучения САРР, от ЧХ каналов связи, которыми пользуются абоненты телефонных систем в процессе эксплуатации САРР. Указанное снижение достоверности обусловлено зависимостью параметров речевого сигнала (РС), которые используются при распознавании, от ЧХ канала связи. Для снижения зависимости используется нормализация параметров сигнала по их среднему значению. Однако отсутствуют исследования, связанные с оценкой влияния различных факторов на степень стабилизации значений нормализованных параметров РС при изменении ЧХ канала связи.

Автоматическое распознавание речи в телефонии осуществляется в присутствии разнообразных акустических помех, что снижает достоверность распознавания. Для подавления помех, присутствующих в речевых сигналах, применяется спектральное вычитание и фильтр Винера. Однако при подавлении помех искажаются сами сигналы, что снижает достоверность распознавания. Условия эффективного использования указанных методов подавления помех зависят от вида и уровня помех, а также от вида самого сигнала. Указанные условия можно определить, главным образом, на основе экспериментальных исследований САРР. Такие исследования в Йемене не проводились.

Для оценки эффективности методов обработки сигналов, поступающих на вход САРР, необходимо иметь соответствующие программные средства, а также наборы (выборки) звукозаписей, которые используются для обучения и тестирования САРР. Такие выборки в Йемене не создавались.

Большой вклад в решение проблемы повышения достоверности автоматического распознавания речи внесли следующие ученые: Болл С.Ф., Винцюк Т.К., Галунов В.И., Грей А., Маркел Дж.Д., Потапова Р.К., Прохоров Ю.Н., Рабинер Л.Р., Сапожков М.А., Скаларт П., Хуанг К, Шафер Р.В., Янг Б. Работы данных исследователей и их последователей позволили значительно снизить частоту ошибок распознавания. Однако специфика арабской речи требует проведения дополнительных исследований по оценке устойчивости САРР к воздействию помех и вариаций частотной характеристики канала связи.

Таким образом, **актуальной** является задача исследования существующих методов предварительной обработки РС, применяемых в системах САРР, оценки их возможностей, разработки алгоритмов обработки речевых сигналов и средств их исследования с целью создания САРР, предназначенной для использования в арабской республике Йемен.

Объектом исследования является телефонная система автоматического распознавания голосовых команд.

Предметом исследования являются алгоритмы обработки сигналов, повышающие устойчивость САРР к воздействию помех, изменению частотной характеристики канала связи, а также к смене диалекта пользователя системы.

Цель работы: разработка алгоритмов обработки речевых сигналов, обеспечивающих повышение достоверности автоматического распознавания голосовых команд, произносимых жителями республики Йемен - пользователями телефонных систем.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Исследовать существующие методы повышения устойчивости САРР к воздействию аддитивных помех и разработать алгоритмы оценки влияния аддитивных помех и средств их подавления на параметры РС, используемые при распознавании речи
2. Исследовать существующие методы повышения устойчивости САРР к изменению частотной характеристики (ЧХ) канала связи и разработать алгоритм оценки влияния ЧХ на параметры РС и средств его подавления, используемых при распознавании речи.
3. Разработать методику оценки влияния смены диалекта на достоверность распознавания голосовых команд.
4. Исследовать существующие методы идентификации диалектов и разработать алгоритм оперативной идентификации диалекта во время сеанса связи.
5. Разработать программное обеспечение экспериментальных исследований предложенных алгоритмов.
6. Создать выборки звукозаписей для обучения САРР, ее тестирования и провести экспериментальные исследования.

Методы исследования. Поставленные задачи решались с использованием теории вероятностей, теории цифровой обработки сигналов, математической статистики, имитационного моделирования.

Теоретическая значимость проведенных исследований.

- Получены выражения для анализа влияния вида оконной функции, используемой при дискретном преобразовании Фурье, на результат нормализации по среднему значению мел-частотных кепстральных коэффициентов (МЧКК).
- Получены выражения для оценки вероятности ошибки автоматической идентификации диалекта в разговорной речи жителей Йемена.

Практическая значимость проведенных исследований.

1. Разработаны методика и соответствующий алгоритм оценки эффективности применения спектрального вычитания и фильтра Винера для повышения помехоустойчивости САРР.
2. Разработаны методика и соответствующий алгоритм оценки эффективности нормализации МЧКК для снижения влияния ЧХ канала связи на достоверность распознавания голосовых команд.
3. Разработано программное обеспечение, реализующее разработанные алгоритмы, которое позволяет обеспечить оптимальную настройку средств подавления влияния помех и ЧХ канала связи на работу САРР.
4. Предложенный алгоритм идентификации диалектов обеспечивает относительную ошибку идентификации равную 0,24%. что позволяет повысить достоверность распознавания арабских названий цифр, как минимум, на 7%.
5. Составлены и обработаны выборки аудиозаписей для обучения и тестирования САРР.

Научная новизна

- Получены выражения для оценки вероятности ошибки идентификации диалекта, использующей акустические модели произнесений контрольных слов.
- Получены результаты экспериментальных исследований идентификаторов йеменских диалектов, использующих акустические модели произнесений контрольных слов.
- Получены выражения, определяющие зависимость значений мел-частотных кепстральных коэффициентов, нормализованных по среднему значению, от вида оконной функции, используемой при дискретном преобразовании Фурье, и неравномерности АЧХ канала связи.
- Получены результаты экспериментального исследования влияния различных оконных функции на значения нормализованных мел-частотных кепстральных коэффициентов.
- Получены результаты экспериментальных исследований возможностей спектрального вычитания и фильтра Винера по подавлению помех при автоматическом распознавании речи в Йемене.

Достоверность полученных результатов обусловлена использованием соответствующего математического аппарата и имитационного моделирования.

Внедрение результатов работы. Результаты диссертационной работы внедрены в ООО «Центр речевых технологий» (г. Санкт-Петербург), а также в учебный процесс на кафедре радиотехники и радиосистем Владимирского государственного университета имени Александра Григорьевича и Николая Григорьевича Столетовых (ВлГУ).

Положения, выносимые на защиту.

1. Предложенный алгоритм автоматической идентификации диалекта обеспечивает повышение достоверности распознавания голосовых команд.
2. Использование предложенной методики оценки эффективности нормализации МЧКК позволяет выделить из имеющегося перечня оконных функций ту функцию, которая обеспечивает наибольшее подавление влияния ЧХ канала связи на параметры РС.
3. Использование спектрального вычитания и фильтра Винера для подавления помех при автоматическом распознавании названий цифр произнесённых на диалектах Йемена повышает достоверность распознавания, если отношение сигнал-помеха меньше 35 дБ.

Апробация работы. Материалы диссертационной работы докладывались и обсуждались на:

- XII-й Международной научной конференции «Физика и Радиоэлектроника в Медицине и Экологии» ФРЭМЭ'2016 (г. Владимир, г. Суздаль 2016 г);
- XIII-й Международной научной конференции «Физика и Радиоэлектроника в Медицине и Экологии» ФРЭМЭ'2018 (г. Владимир, г. Суздаль 2018 г);
- XIII-й Международной научно-технической конференции «Перспективные технологии в средствах передачи информации» ПТСПИ (г. Владимир, 2019 г);
- 2019 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT). (25-26 April 2019, Yekaterinburg, Russia).

Публикации. По материалам диссертации опубликовано 10 работ, в том числе 3 статьи в журналах, рекомендованных ВАК, 7 - на международных конференциях (одна работа - в издании IEEE, индексируемом SCOPUS). Получено 4 свидетельства о государственной регистрации программ для ЭВМ.

Личное участие. Основные результаты диссертации были получены лично автором в период с 2015 по 2019 г.

Структура и объём диссертации. Диссертация состоит из Введения, четырёх глав, заключения, библиографического списка, включающего 100 наименований, списка сокращений и 7 приложений. Объём диссертации составляет 130 страниц машинописного текста, 48 рисунков и 27 таблиц. Объём приложений составляет 22 страницы.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обосновывается актуальность темы диссертационной работы, сформулированы цель исследования и решаемые задачи, показаны научная новизна и практическая значимость полученных результатов, определены положения, выносимые на защиту.

В первой главе представлены особенности и проблемы построения систем автоматического распознавания арабской речи. Арабский язык характеризуется большим разнообразием диалектов. Стандартизированным диалектом является современный диалект арабского языка (Modern Standard Arabic - MSA). Он является официальным языком арабского мира. При повседневном общении население использует диалектный арабский язык (DA). DA является основным языком для драматических, комедийных программ и во многих жанровых передачах. Из-за значительных отличий арабские диалекты можно рассматривать как разные языки при решении таких задач, как, например, автоматическая идентификация диалекта.

Большое отличие диалектов обуславливает высокую изменчивость произнесения одних и тех же слов, что увеличивает число ошибок распознавания. Поэтому возникает необходимость включения идентификатора диалекта в состав системы распознавания речи. Анализ существующих результатов и подходов к идентификации диалектов арабского языка позволяет сделать следующие выводы. Во-первых, нет данных по идентификации диалектов Йемена. Во-вторых, достаточно высокая точность идентификации достигается при использовании акустических моделей, на основе скрытых Марковских моделей и при использовании мел-частотных кепстральных коэффициентов (МЧКК) в качестве параметров речевых сигналов.

Во многих САРР собственно процесс распознавания основан на использовании акустических моделей звуков речи в виде скрытых Марковских моделей (СМП). В составе звуков арабской речи есть такие звуки, которые отсутствуют в других национальных языках. Поэтому акустические модели звуков, используемые при распознавании арабской речи, значительно отличаются от моделей, которые используются в составе существующих САРР. Следовательно, требуется провести большой объем экспериментальных исследований для оценки достоверности распознавания САРР диалектного языка Йемена. Следует отметить, что создание систем распознавания арабской речи сдерживается отсутствием объемных выборок звукозаписей, подготовленных для обучения и тестирования САРР.

Современные системы автоматического распознавания речи (САРР) включают в себя модуль предварительной обработки речевых сигналов (РС). В процессе предварительной обработки РС определяются паузы в сигнале, подавляются аддитивные помехи. Затем определяются параметры РС, которые поступают на вход алгоритма распознавания. Для повышения устойчивости МЧКК к изменениям частотной характеристики канала связи используют нормализацию коэффициентов по их среднему значению.

Большое количество ошибок, возникающих при автоматическом распознавании речи, обусловлено влиянием акустических помех, которые обычно сопровождают речевой сигнал. Помехи искажают речевой сигнал, что приводит к изменению значений параметров речевого сигнала (РС) по сравнению с теми значениями, которые использовались при создании моделей речевых сигналов на стадии обучения системы автоматического распознавания речи.

Для очистки зашумленной речи применяют спектральное вычитание (СВ) и фильтр Винера (ФВ). При использовании СВ оценивается спектральная плотность мощности помехи на интервале паузы РС, и полученная оценка вычитается из оценки зашумленного сигнала. Частотная характеристика фильтра Винера формируется так, чтобы минимизировать среднеквадратическое отклонение очищенного от помех сигнала от "чистого" сигнала. Подавление помех с помощью фильтра Винера осуществляется при прохождении суммы сигнала и помехи через указанный фильтр.

В диссертации исследованы возможности подавления помех с помощью СВ и ФВ на примере автоматического распознавания названий цифр на диалектном языке Йемена. Проведены экспериментальные исследования изменений параметров речевого сигнала МЧКК при воздействии помехи в виде модели белого шума и использовании указанных методов подавления помех. Блок-схема алгоритма эксперимента представлена на рисунке 1.

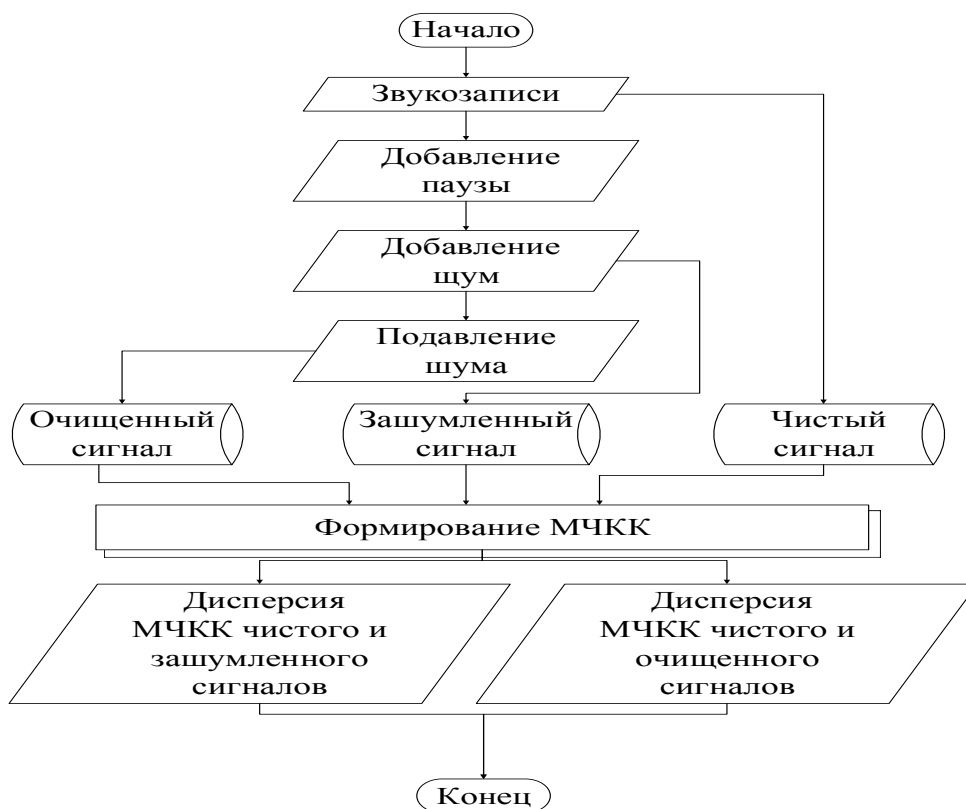


Рисунок 1. Алгоритм оценки эффективности использования СВ и ФВ на результаты параметров МЧКК

Использована следующая методика проведения эксперимента.

- К звукозаписям арабских названий цифр (0 – 9) добавляется пауза.
- К полученным звукозаписям добавляется белый шум. Уровень шума определяется заданным отношением сигнал-шум (signal noise ratio – SNR).
- Помеха подавляется с помощью СВ или ФВ.
- Определяются значения МЧКК для чистого, зашумлённого и очищенного от помехи сигналов.
- Определяются значения дисперсии разности между МЧКК чистого и зашумленного сигналов.
- Определяются значения дисперсии разности между МЧКК чистого и очищенного от помех сигналов.

Эксперимент показал, что, в основном, использование ФВ обеспечивает меньшие искажения параметров по сравнению с использованием СВ. (рисунок 2).

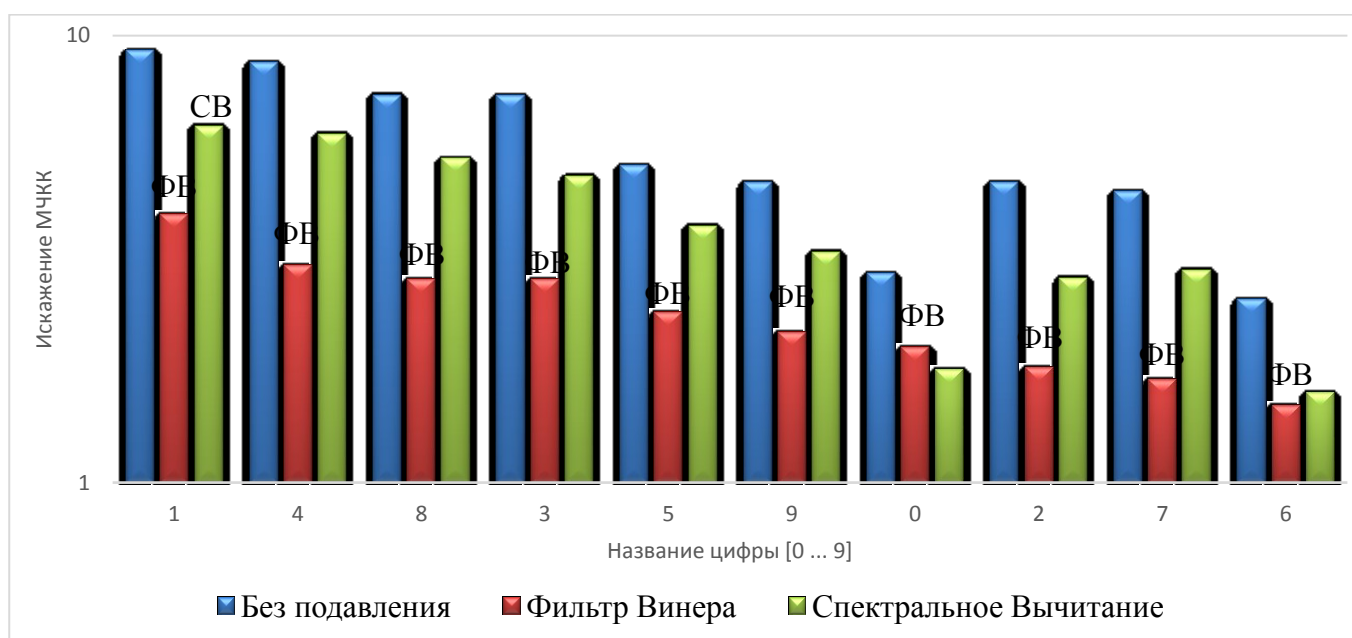


Рисунок 2. Зависимость изменений МЧКК от названия цифры (SNR= 20 дБ)

Результаты проведения эксперимента по оценке достоверности распознавания речи при использовании СВ и ФВ отражены графиками зависимости достоверности (точности) распознавания, усреднённой по всем названиям цифр от отношения сигнал-шум (рисунок 3).

Из графиков следует, что при ОСШ меньше 35 дБ применение СВ и ФВ повышает достоверность распознавания. Можно также сделать вывод о том, что применение ФВ и СВ практически приводит к одному результату. Однако, учитывая, что фильтр Винера по сравнению со спектральным вычитанием обеспечивает меньшие искажения параметров сигнала, целесообразно использовать именно ФВ для подавления помех. В среднем использование ФВ эквивалентно снижению уровня помех на (10 – 15) дБ. По результатам

эксперимента можно сделать вывод о целесообразности использования СВ и ФВ при ОСШ менее 35 дБ.

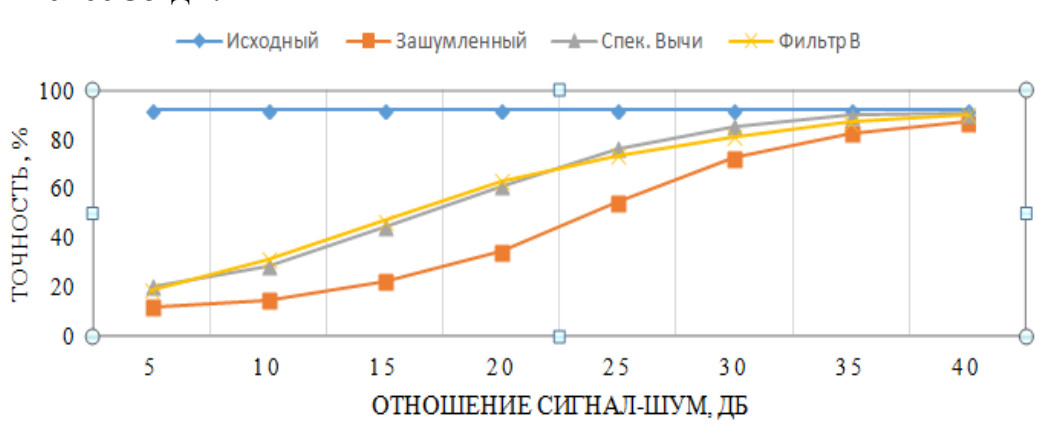


Рисунок 3. Зависимость достоверности распознавания от отношения сигнал-шум

Во второй главе рассмотрено использование идентификации диалекта для повышения достоверности распознавания.

Йеменский диалектный язык можно разделить на несколько основных диалектных групп, каждая из которых обладает своей лексикой и фонетикой. Наиболее значительными из этих групп являются: северный (СД), южный (ЮД) и западный (ЗД). Количество носителей СД в стране составляет 68,3% от общего населения, количество носителей ЮД - 13,4%, количество носителей ЗД - 12,7%.

Рассмотрена возможность использования акустических моделей произнесений тестовых слов для оперативной идентификации одного из трех йеменских диалектов в процессе обращения пользователя к автоматической телефонной справочной системе. Рассмотрен случай, когда САРР предназначена для распознавания произнесений арабских названий отдельных цифр.

Анализ транскрипций произнесений цифр для трех диалектов показал, что наиболее сильно различаются по диалектам произнесения цифр: 2, 3, 4, 5, 7, 9. Причем наибольшая степень отличий соответствует цифре 9.

Анализируя отличия указанных транскрипций по диалектам, можно прийти к выводу о возможности их использовании при построении классификаторов рассматриваемых йеменских диалектов.

Целесообразность идентификации диалектов оценивается экспериментально путем сравнения результатов автоматического распознавания названий цифр с учетом идентификации, когда для каждого названия цифры на каждом диалекте создается своя акустическая модель, и без ее учета, когда акустическая модель названия каждой цифры является общей для всех диалектов.

Результаты проведенного эксперимента показывают, что при отсутствии идентификации, когда при распознавании используются общие для всех диалектов акустические модели произнесений названий цифр, относительная частота правильного распознавания, усредненная по всем названиям цифр, равна 91,6 %. Если же для каждого диалекта используется своя совокупность моделей для распознаваемых произнесений (случай безошибочной идентификации диалекта),

то такая же относительная частота для ЮД составляет 98,2%, для СД - 98,1% и для ЗД - 98,3%. Следовательно, идентификация диалектов позволяет на (7 - 8) % повысить относительную частоту правильного распознавания.

Проанализирована возможность идентификации любого из рассматриваемых диалектов при произнесении названия цифры 9. Для каждого диалекта создается своя акустическая модель произнесения названия цифры 9. Для идентификации диалекта произносится контрольное название цифры 9. Результатом идентификации является диалект, которому принадлежит акустическая модель, которой с наибольшей вероятностью соответствует произнесение.

Определим вероятность ошибочной идентификации.

$$P_{\text{Класс3}} = P(C) \cdot P(\text{Ю}|C) + P(C) \cdot P(\text{З}|C) + P(\text{Ю}) \cdot P(C|\text{Ю}) + P(\text{Ю}) \cdot P(\text{З}|\text{Ю}) + P(\text{З}) \cdot P(\text{Ю}|\text{З}) + P(\text{З}) \cdot P(C|\text{З}) =$$

$$= P(C) \cdot [P(\text{Ю}|C) + P(\text{З}|C)] + P(\text{Ю}) \cdot [P(C|\text{Ю}) + P(\text{З}|\text{Ю})] + P(\text{З}) \cdot [P(\text{Ю}|\text{З}) + P(C|\text{З})]$$

Здесь $P(C)$, $P(\text{Ю})$, $P(\text{З})$ – вероятности использования абонентом телефонной сети северного, южного и западного диалектов; $P(i|j)$ – условная вероятность ошибочной идентификации j -го диалекта в качестве i -го диалекта. Если считать вероятности появления каждого из трех диалектов равными, то

$$P(C) = P(\text{Ю}) = P(\text{З}) = 0,33;$$

$$P_{\text{Класс3}} = 0,33(P(\text{Ю}|C) + P(\text{З}|C) + P(C|\text{Ю}) + P(\text{З}|\text{Ю}) + P(\text{Ю}|\text{З}) + P(C|\text{З})).$$

Рассмотрим результаты тестирования идентификатора с использованием названия цифры 9. В таблице 1. Указаны относительные частоты правильной и ошибочной идентификаций диалектов (произнесение названия цифры 9).

Используя данные таблицы 1 и выражения, приведенные выше, получаем оценку вероятности ошибочной идентификации в среднем по трем диалектам.

$$\hat{P}_{\text{Класс3}} = 0,33[0,08 + 0,04] + 0,33[0 + 0] + 0,33[0,04 + 0] = 0,053.$$

Наибольшая вероятность ошибки соответствует произнесению названия цифры на северном диалекте: $1 - P(C|C) = 1 - 0,88 = 0,12$.

Таблица 1. Результаты идентификации диалектов (произнесение названия цифры 9).

Результат идентификации диалекта	Идентифицируемый диалект		
	Северный	Южный	Западный
	Относительная частота классификации диалекта, %		
Северный диалект	88	0	4
Южный диалект	8	100	0
Западный диалект	4	0	96

Рассмотрим возможность использования для идентификации произнесения не одного, а двух названий цифр. Рассмотрим совместную работу "двоичного" (произнесение названия цифры 2), когда осуществляется классификация на две группы, и "троичного" (произнесение названия цифры 9), когда осуществляется классификация на три группы, классификаторов в случае идентификации южного диалекта. В таблице 2 указаны относительные частоты правильной и ошибочной идентификаций диалектов (произнесение названия цифры 2).

Таблица 2. Результаты эксперимента по классификации диалектов на две группы (произнесение названия цифры 2)

Результат классификации диалекта	Классифицируемые группы диалектов	
	Северный	Южный+Западный
	Относительная частота классификации диалекта, %	
Северный диалект	98	0
Группа (южный+западный диалекты)	2	100

Вероятность правильного срабатывания обоих классификаторов

$$P_{\text{прав.класф (2,9)}}(\text{ю}) = P_9(\text{ю}|\text{ю})P_2(\text{юз}|\text{ю}).$$

Тогда вероятность ошибки идентификации

$$P_{\text{ош.класф (2,9)}}(\text{ю}) = 1 - P_9(\text{ю}|\text{ю})P_2(\text{юз}|\text{ю}).$$

Если данные классификаторов противоречат друг другу, то можно обнаружить ошибку идентификации. Ее вероятность

$$P_{\text{ош.класф.обнаруж (2,9)}}(\text{ю}) = P_2(\text{юз}|\text{ю})P_9(\text{с}|\text{ю}) + P_2(\text{с}|\text{ю})[P_9(\text{ю}|\text{ю}) + P_9(\text{з}|\text{ю})].$$

Следовательно, вероятность необнаружения ошибки идентификации

$$P_{\text{ош.класф.необнаруж (2,9)}}(\text{ю}) = P_{\text{ош.класф (2,9)}}(\text{ю}) - P_{\text{ош.класф.обнаруж (2,9)}}(\text{ю}).$$

Аналогично получены выражения для северного и западного диалектов.

$$P_{\text{ош.класф (2,9)}}(\text{с}) = 1 - P_9(\text{с}|\text{с})P_2(\text{с}|\text{с}).$$

$$P_{\text{ош.класф.обнаруж (2,9)}}(\text{с}) = P_2(\text{юз}|\text{с})P_9(\text{с}|\text{с}) + P_2(\text{с}|\text{с})[P_9(\text{ю}|\text{с}) + P_9(\text{з}|\text{с})].$$

$$P_{\text{ош.класф.необнаруж (2,9)}}(\text{с}) = P_{\text{ош.класф (2,9)}}(\text{с}) - P_{\text{ош.класф.обнаруж (2,9)}}(\text{с})$$

$$P_{\text{ош.класф (2,9)}}(\text{з}) = 1 - P_9(\text{з}|\text{з})P_2(\text{юз}|\text{з}).$$

$$P_{\text{ош.класф.обнаруж (2,9)}}(\text{з}) = P_2(\text{юз}|\text{з})P_9(\text{с}|\text{з}) + P_2(\text{с}|\text{з})[P_9(\text{ю}|\text{з}) + P_9(\text{з}|\text{з})].$$

$$P_{\text{ош.класф.необнаруж (2,9)}}(\text{з}) = P_{\text{ош.класф (2,9)}}(\text{з}) - P_{\text{ош.класф.обнаруж (2,9)}}(\text{з})$$

Из эксперимента следует

$$P_{\text{ош.клсф.необнаруж (2,9)}(\text{ю})} = 0; P_{\text{ош.клсф.необнаруж (2,9)}(\text{с})} = 0,0024; P_{\text{ош.клсф.необнаруж (2,9)}(\text{з})} = 0;$$

Усредненная по всем диалектам оценка вероятности ошибки идентификации, которая не обнаруживается

$$P_{\text{ош.клсф.необнаруж (2,9)}(\text{сюз})} = (0 + 0 + 0,0024)/3 = 0,0008.$$

Видно, что использование совокупности "двоичного и "троичного" классификаторов значительно уменьшает ошибку классификации по сравнению с использованием одного лишь "троичного" классификатора.

Применение совокупности двух классификаторов создает ситуации, когда обнаруживается противоречие в работе классификаторов, и требуется повторное произнесение заданных названий цифр. Данная ситуация снижает привлекательность использования системы распознавания. Определим численные значения оценок вероятностей возникновения такой дискомфортной ситуации.

Используя данные таблиц 1 и 2, а также выражения для северного диалекта, получим значение оценки вероятности повторного произнесения контрольных названий цифр равное 0,13. При равной вероятности использования трех диалектов вероятность повторного произнесения заданных названий цифр, и, соответственно, численное значение оценки вероятности

$$P_{\text{клсф.повтор (2,9)}(\text{сюз})} = P(\text{ю})P_{\text{ош.клсф.обнаруж (2,9)}(\text{ю})} + P(\text{с})P_{\text{ош.клсф.обнаруж (2,9)}(\text{с})} + P(\text{з})P_{\text{ош.клсф.обнаруж (2,9)}(\text{з})} = \frac{0 + 0,13 + 0,04}{3} = 0,057.$$

Таким образом, относительная частота повторного произнесения последовательности названий двух цифр равна 5,7%.

Определим вероятность ошибки распознавания голосовой команды (ГК) при наличии идентификатора диалекта в составе системы распознавания. Вероятность ошибки распознавания ГК при произнесении команды с северным диалектом

$$P_{\text{ошГКиднт}}(\text{С}) = P_{\text{ошГКиднт}}(\text{С|С}) + P_{\text{ошГКиднт}}(\text{Ю|С}) + P_{\text{ошГКиднт}}(\text{З|С}),$$

где $P_{\text{ошГК}}(\text{С|С})$ – вероятность ошибки распознавания, когда идентификация диалекта произошла правильно; $P_{\text{ошГК}}(\text{Ю|С})$ – вероятность ошибки распознавания, когда идентификация диалекта произошла с ошибкой: вместо северного определен южный диалект; $P_{\text{ошГК}}(\text{З|С})$ – вероятность ошибки распознавания, когда идентификация диалекта произошла с ошибкой: вместо северного определен

западный диалект. Определим вероятность ошибки распознавания, когда идентификация диалекта произошла правильно

$$P_{\text{ошГКиднт}}(C|C) = P_{\text{иднт}}(C|C) * P_{\text{ошГК}}(C|C),$$

где $P_{\text{иднт}}(C|C)$ – вероятность правильной идентификации диалекта; $P_{\text{ошГК}}(C|C)$ - вероятность ошибки распознавания, когда в системе распознавания используются акустические модели, соответствующие северному диалекту.

Определим вероятность ошибки распознавания, когда идентификация диалекта произошла с ошибкой: вместо северного определен южный диалект;

$$P_{\text{ошГКиднт}}(Ю|C) = P_{\text{иднт}}(Ю|C) * P_{\text{ошГК}}(C|Ю),$$

где $P_{\text{иднт}}(Ю|C)$ – вероятность ошибочной идентификации диалекта: вместо северного определен южный диалект; $P_{\text{ошГК}}(C|Ю)$ - вероятность ошибки распознавания, когда в системе распознавания используются акустические модели, соответствующие южному диалекту.

Определим вероятность ошибки распознавания, когда идентификация диалекта произошла с ошибкой: вместо северного определен западный диалект.

$$P_{\text{ошГКиднт}}(З|C) = P_{\text{иднт}}(З|C) * P_{\text{ошГК}}(C|З),$$

где $P_{\text{иднт}}(З|C)$ – вероятность ошибочной идентификации диалекта: вместо северного определен западный диалект; $P_{\text{ошГК}}(C|З)$ - вероятность ошибки распознавания, когда в системе распознавания используются акустические модели, соответствующие западному диалекту.

С использованием приведенных выражений и с учетом данных эксперимента получены следующие оценки достоверности распознавания при произнесении одного контрольного слова. Для южного диалекта: 0,98; для западного - 0,96; для северного – 0,91. Данные значения незначительно отличаются от случая использования идеального безошибочного идентификатора.

При произнесении двух контрольных слов достоверность практически соответствует случаю работы идеального идентификатора.

В третьей главе исследована возможность применения нормализации параметров сигнала для подавления влияния частотной характеристики (ЧХ) канала связи на работу САРР. Рассмотрено влияние вида оконной функции и неравномерности амплитудно-частотной характеристики (АЧХ) канала связи (участка канала связи от диктора до аналого-цифрового преобразователя) на степень стабилизации нормализованных параметров сигнала при изменении АЧХ. Получено выражение для разности нормированных логарифмов спектров сигнала для случаев обучения и тестирования САРР, когда "звуковой" участок канала связи меняется.

$$\Delta(p, f_1) = Y_{LN1}(p, f_1) - Y_{TN1}(p, f_1) \approx k_{f_2}(f_1) \cdot \left[\frac{u(p, f_2)}{u(p, f_1)} - \left(\frac{u(p, f_2)}{u(p, f_1)} \right) \right] \cdot \left[\frac{m_1(f_2)}{m_1(f_1)} - \frac{m_2(f_2)}{m_2(f_1)} \right]$$

где $u(p, f_1)$ и $u(p, f_2)$ – модули коэффициентов БПФ на частотах f_1 и f_2 ; $m_1(f_1)$ и $m_1(f_2)$ – модули коэффициентов передачи микрофонов на частотах f_1 и f_2 ; $k_{f_2}(f_1)$ – коэффициент ослабления боковых лепестков оконной функции на частоте f_1 при применении БПФ к спектральному компоненту на частоте f_2 .

Здесь отношения: $\frac{m_1(f_2)}{m_1(f_1)}$ и $\frac{m_2(f_2)}{m_2(f_1)}$ – характеризуют неравномерности АЧХ (частоты f_1, f_2) звуковых трактов, используемых при обучении системы распознавания речи и при ее тестировании, соответственно. Видно, что, чем меньше отличаются ЧХ микрофонов и меньше уровень боковых лепестков оконной функции, тем меньше отличий в нормализованных коэффициентах для случаев обучения и тестирования SAPR. Если спектральный компонент $u(p, f_1) \ll u(p, f_2)$, то при малом изменении от сегмента к сегменту компонента $u(p, f_1)$ и при малой постоянной составляющей $u(p, f_2)$ разность нормализованных коэффициентов становится большой. На рисунке 4. Показана блок-схема алгоритма оценки влияния нормализации на параметры РС.

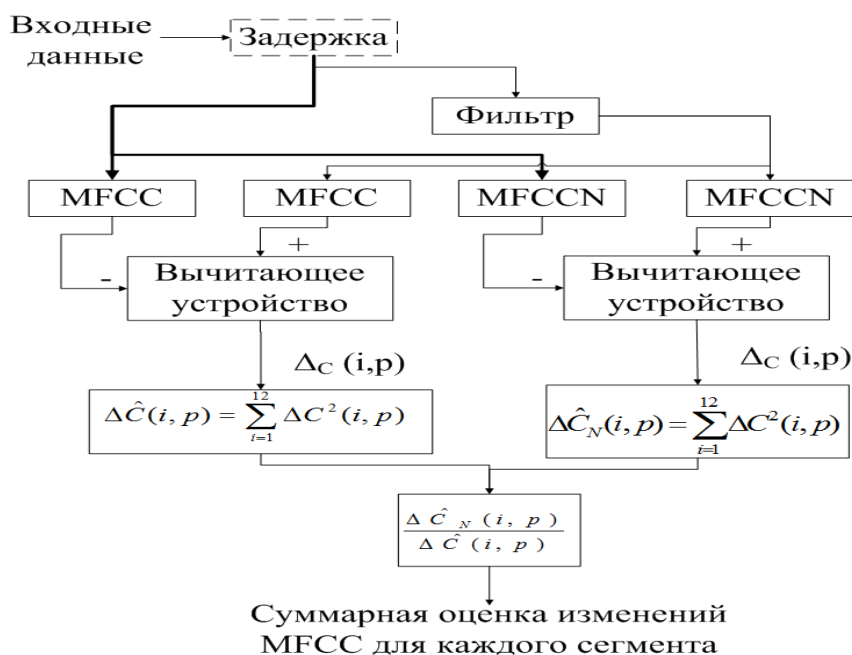


Рисунок 4. Алгоритм оценки влияния нормализации на параметры речевого сигнала

С целью проверки полученных выше соотношений проведено имитационное моделирование основных процессов формирования МЧКК на этапах создания обучающей и тестирующей выборки данных звука. Проведены две серии экспериментов. В первой серии использовался специально сформированный тестовый сигнал. А во второй серии – РС. На рисунке 5 представлены графики

зависимости достоверности распознавания от вида используемой оконной функции.

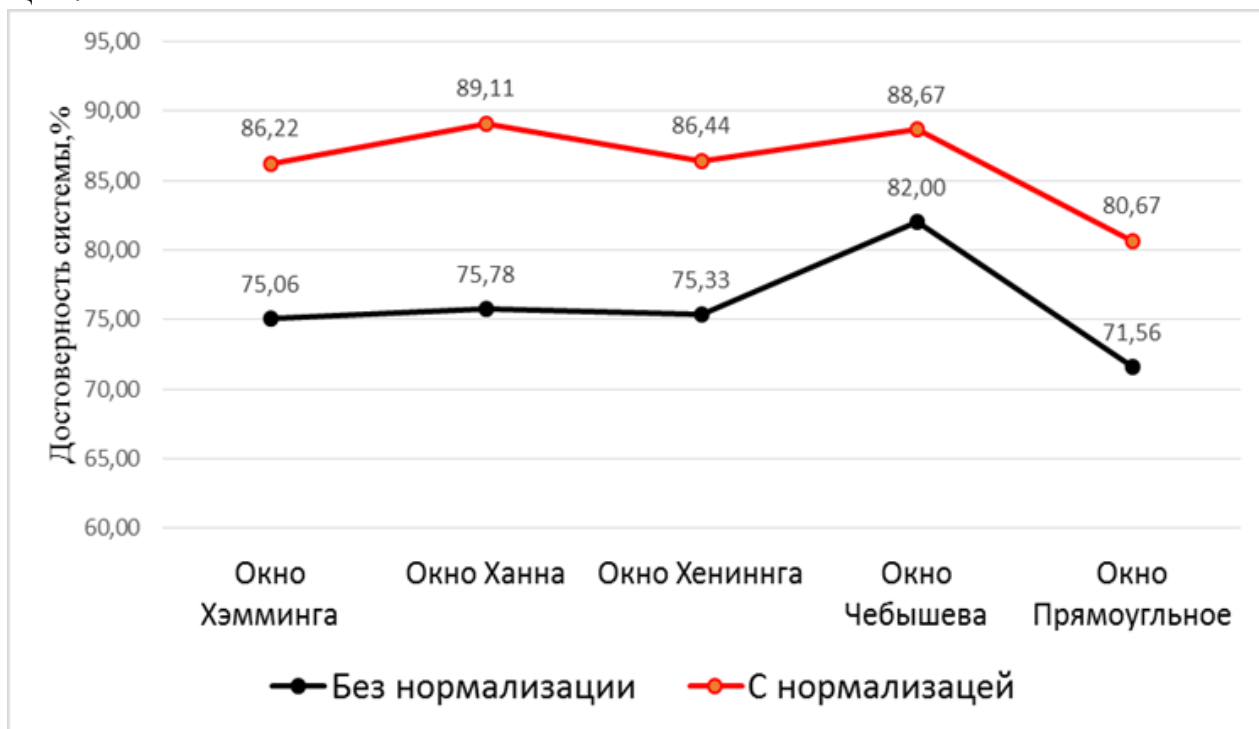


Рисунок 5. Достоверность распознавания при использовании нормализации параметров речевого сигнала

Видно из графиков, что точность распознавания увеличивается с использованием нормализации параметров РС. Использование окон Чебышева и Ханна, которые имеют малый уровень боковых лепестков, удаленных от основного лепестка, обеспечивают большую достоверность распознавания по сравнению с окном Хэмминга.

В четвертой главе описано разработанное программное обеспечение экспериментов. Представлено описание программного комплекса, предназначенного для оценки помехоустойчивости систем распознавания речи в телефонии. При проектировании средств повышения помехоустойчивости систем автоматического распознавания речи (САРР) необходимо оценивать возможности использования различных средств подавления помех. Программный комплекс позволяет оценить эффективность подавления аддитивных помех с помощью спектрального вычитания, а также путем использования фильтра Винера.

Комплекс позволяет также оценить возможности нормализации параметров речевого сигнала по их среднему значению во времени. Нормализация позволяет уменьшить влияние изменчивости частотной характеристики канала связи на точность распознавания. На рисунке 6. представлено изображение главного окна программного комплекса. Изображенные закладки позволяют выбрать режим работы комплекса.

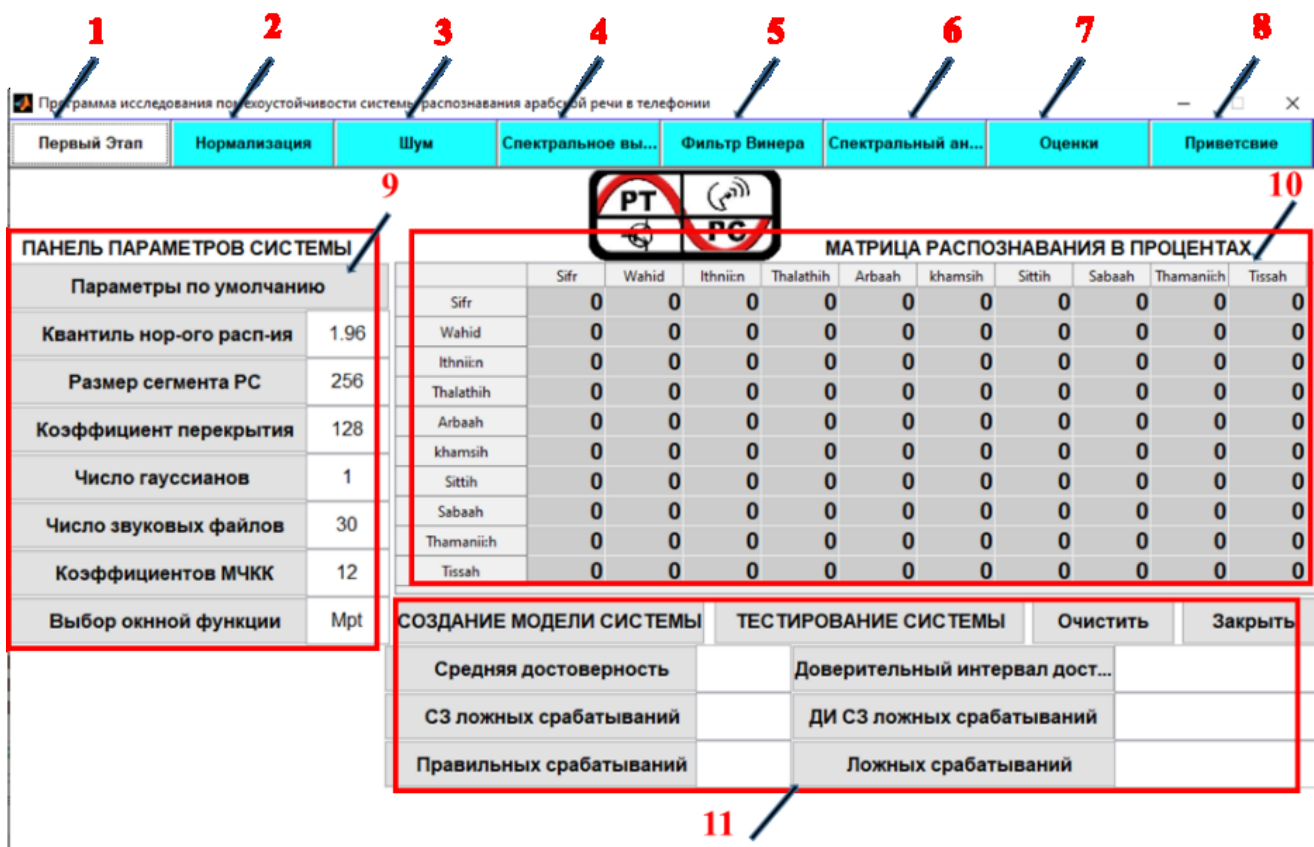


Рисунок 6. Изображение главного окна программного комплекса

Первая закладка. Обеспечивает ввод параметров САРР и проведение исследований в условиях, когда помехи отсутствуют, и нормализация не проводится.

Вторая закладка. Позволяет протестировать САРР и определить точность распознавания, когда для построения акустических моделей используются нормализованные параметры речевого сигнала.

Третья, четвертая, пятая и шестая закладки. Позволяют протестировать САРР и определить точность распознавания, когда на сигнал накладывается помеха, а также когда используются шумоподавление и нормализованные МЧКК. При выборе закладок "Спектральное вычитание" и "Фильтр Винера" появляется возможность настройки указанных средств подавления помех по критерию максимума точности распознавания.

Седьмая закладка. Обеспечивает оценку степени соответствия произнесения акустической модели.

Восьмая закладка "Приветствие". используется для вывода информации о текущей версии программного комплекса.

Панель параметров системы (на рис. 6, окно 9). Здесь можно задать параметр исследования и анализировать его влияние на достоверность распознавания при обучении и тестировании системы. К числу параметров САРР относятся такие, как: размер сегмента РС; вид оконной функции; величина

перекрытия сегментов; количество МЧКК, количество звуковых файлов, которые используются для тестирования САРР.

Область №10 главного окна программы. Результаты работы САРР отражаются таблицей "Матрица распознавания". В верхней строке таблицы указаны произнесенные названия цифр: 0 – 9, а в левом столбце указаны названия моделей голосовых команд. Число, стоящее на диагонали матрицы, является относительной частотой (в %) правильного распознавания ГК. Числа, находящиеся вне диагонали, являются относительными частотами ошибок распознавания.

Область №11 главного окна программы. Помимо точечных оценок вероятностей правильного и ошибочного распознавания возможно получение интервальных оценок.

На рисунке 7. Представлены экспериментальные результаты САРР при применении нормализации.

ПАНЕЛЬ ПАРАМЕТРОВ СИСТЕМЫ

Параметры по умолчанию	
Квантиль нор-ого расп-ия	1.96
Размер сегмента РС	256
Коэффициент перекрытия	128
Число гауссианов	1
Число звуковых файлов	150
Коэффициентов МЧКК	12
Выбор оконной функции	Mpt

МАТРИЦА РАСПОЗНАВАНИЯ В ПРОЦЕНТАХ

	Sifr	Wahid	lthniin	Thalathih	Arbaah	khamseh	Sittih	Sabaah	Thamaniih	Tissah
Sifr	99.333	0	0	0	0	0	0	0	0	0
Wahid	0	100	0	0	0	0	0	0	0	0
lthniin	0	0	100	0	0	0	0	0	0.6667	0
Thalathih	0.6667	0	0	95.333	0	0	0	0	0	0
Arbaah	0	0	0	0	99.333	0	0.6667	1.3333	0	0
khamseh	0	0	0	0	0.6667	100	0	0	0	0
Sittih	0	0	0	0	0	0	99.333	0.6667	0	1.3333
Sabaah	0	0	0	4.6667	0	0	0	98	0	0
Thamaniih	0	0	0	0	0	0	0	0	99.333	0
Tissah	0	0	0	0	0	0	0	0	0	98.667

СОЗДАНИЕ МОДЕЛИ СИСТЕМЫ **ТЕСТИРОВАНИЕ СИСТЕМЫ** Очистить Закрыть

Средняя достоверность	98.9333%	Доверительный интервал дост...	95.6796% <= CI <= 99.74
СЗ ложных срабатываний	1.0667%	ДИ СЗ ложных срабатываний	0.25677% <= CI <= 4.320
Правильных срабатываний	148	Ложных срабатываний	2

Параметры системы с использованием нормализации параметров РС: Коэффициент нормального распределения = 1.96, Длительность сегмента сигнала = 256, Коэффициент перекрытия = 128, Число гауссианов = 1, Число звуковых файлов = 150, Число коэффициентов МЧКК = 12, Вид оконной функции = Mpt

Рисунок 7. Составная часть программы, закладка нормализации

Видно, что в данном примере при использовании нормализации параметров РС средняя достоверность распознавания составила более 98 %

В Заключение приведены результаты решения поставленных в диссертации задач.

В Приложениях представлены результаты экспериментов классификации диалектов на три группы и на две группы, характеристики исследуемых шумов и результаты их влияния на достоверность системы распознавания, Результат тестирования САРР, акты внедрения, свидетельства о регистрации программ для ЭВМ.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

1. Анализ литературных источников выявил отсутствие разработок САРР для разговорного языка республики Йемен.
2. Установлено, что отличие звуков арабской речи от звуков речи населения развитых стран не позволяет в полной мере использовать опыт построения существующих САРР для построения САРР для арабской речи.
3. Разработанная методика и соответствующий алгоритм оценки уровня искажений параметров РС, обусловленных влиянием аддитивных помех, позволяет установить целесообразность использования определенного метода подавления помех при известном отношении сигнал-помеха.
4. В ходе экспериментального исследования установлено, что использование фильтра Винера для подавления помех при отношении сигнал-шум менее (25 – 30) дБ является более целесообразным по сравнению с применением спектрального вычитания.
5. Разработаны методика и соответствующий алгоритм оценки эффективности нормализации параметров (МЧКК) речевого сигнала для снижения влияния частотной характеристики канала связи на достоверность распознавания голосовых команд
6. Проведенные теоретические и экспериментальные исследования показывают, что использование оконных функций с большим уровнем подавления боковых лепестков, достаточно удаленных от основного лепестка, обеспечивает большую стабилизацию нормализованных параметров сигнала при прохождении его через канал связи с неравномерной АЧХ в полосе частот сигнала.
7. Проведенные эксперименты показывают, что использование идентификации диалектов разговорного языка республики Йемен в составе САРР приводит к повышению точности распознавания произнесений названий цифр на (7 – 8) %.
8. Для оперативной идентификации диалекта при автоматическом распознавании названий цифр целесообразно использовать произнесение одного или нескольких контрольных названий цифр. Целесообразно для построения идентификатора диалектов использовать акустические модели таких произнесений названий цифр, которые наиболее сильно отличаются по диалектам
9. В результате анализа причин ошибок идентификации диалекта получены выражения, позволяющие определить вероятность ошибки идентификации.
10. Предложенный алгоритм идентификации трех наиболее распространенных диалектов республики Йемен обеспечивает относительную ошибку идентификации равную 0,24%, что позволяет повысить достоверность распознавания арабских названий цифр, как минимум, на 7%.

11. Разработано программное обеспечение, реализующее разработанные алгоритмы, которое позволяет обеспечить оптимальную настройку средств подавления влияния помех и ЧХ канала связи на работу САРР.
12. Использование программного комплекса позволяет проводить детальное исследование устойчивости САРР к воздействию аддитивных помех различного вида, а также к изменению частотной характеристики канала связи. Графический интерфейс комплекса значительно сокращает затраты времени на проведение исследований.
13. Создана объемная выборка звукозаписей произнесений арабских (республика Йемен) названий цифр от нескольких дикторов, произношение которых характеризуется наиболее распространенными диалектами. Данная выборка позволяет проводить обучение и тестирование САРР в процессе их разработки.

Публикации автора по теме диссертации

Статьи в изданиях, входящих в перечень ВАК

1. **Аль-Дайбани А.М.** Подавление помех в системе распознавания арабской речи / А.М Аль-Дайбани, Е.К. Левин // Проектирование и технологии электронных средств. – 2016, №4. – С. 14 – 18.
2. **Аль-Дайбани А.М.** Анализ возможностей подавления влияния частотной характеристики канала связи на параметры речевого сигнала / А.М Аль-Дайбани, Е.К. Левин // Проектирование и технологии электронных средств. – 2018, №3. – С. 14 – 18.
3. **Аль-Дайбани А.М.** Идентификация диалектов арабской разговорной речи при автоматическом распознавании голосовых команд вы телефонии / А.М Аль-Дайбани, Е.К. Левин, Т.В. Левина // Проектирование и технологии электронных средств. – 2019, №1. – С. 35 – 40.

Прочие издания

4. **Аль-Дайбани А.М.** О возможности использования системы распознавания речи при регистрации заявок на получение медицинских услуг в Йемене / А.М Аль-Дайбани, Е.К. Левин // XII Международная научно-техническая конференция «Физика и радиоэлектроника в медицине и экологии – ФРЭМЭ'2016». - г. Суздаль, 05 – 07 июля 2016. – С. 301-303.
5. **Аль-Дайбани А.М.** Возможности использования автоматизированных систем регистрации заявок на получение медицинских услуг в Йемене / А.М Аль-Дайбани // XII-й Международная научно-техническая конференция «Физика и радиоэлектроника в медицине и экологии. – ФРЭМЭ–2016». – г. Суздаль, 05 – 07 июля 2016. – С.322–324.
6. **Аль-Дайбани А.М.** Классификация диалектов республики Йемен для повышения точности распознавания речи / А.М Аль-Дайбани // XIII-я

- Международная научно-техническая конференция «Физика и радиоэлектроника в медицине и экологии – ФРЭМЭ–2018». – г. Суздаль, 03 – 05 июля 2018. – С. 358–361.
7. **Аль-Дайбани А.М.** Зависимость результатов нормализации MFCC от вида используемой оконной функции/ А.М Аль-Дайбани // XIII-я Международная научно-техническая конференция «Физика и радиоэлектроника в медицине и экологии – ФРЭМЭ–2018». – г. Суздаль, 03 – 05 июля 2018. – С.361 – 365.
 8. **Аль-Дайбани А.М.** Анализ факторов, влияющих на результат нормализации параметров речевого сигнала по среднему значению / А.М Аль-Дайбани, Е.К Левин // XIII-я Международная научно-технической конференция «Физика и радиоэлектроника в медицине и экологии – ФРЭМЭ–2018». – г. Суздаль, 03 – 05 июля 2018. – С. 365 – 369.
 9. **Аль-Дайбани А.М.** Программный комплекс для оценки помехоустойчивости систем распознавания речи в телефонии / А.М Аль-Дайбани, Е.К Левин // XIII-я Международная научно-техническая конференция «Перспективные технологии в средствах передачи информации – ПТСПИ-2019». – г. Владимир, 03 – 05 июля 2019. – Т.2. – С.155–157.
 10. **Аль-Дайбани А.М.** Использование спектрального вычитания и фильтра Винера для подавления помех при автоматическом распознавании голосовых команд/А.М Аль-Дайбани, Е.К Левин // XIII-я Международная научно-техническая конференция «Перспективные технологии в средствах передачи информации - ПТСПИ-2019». – г. Владимир, 03 – 05 июля 2019. – Т.2. – С. 157-159.
 11. **Evgenii Levin** Research of Window Function Influence on the Result of Arabic Speech Automatic Recognition/ Levin Evgenii, Abdulghani Al-Dhaibani // 2019 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT)/ Publisher: IEEE/ Yekaterinburg. – Russia. – 2019. -pp. – 204-207.

Свидетельства о регистрации программ для ЭВМ

12. **Аль-Дайбани А.М.** Анализ влияния оконной функции на результат нормализации параметров речевого сигнала/А.М Аль-Дайбани, Е.К Левин // Владимирский государственный университет//Свидетельство о регистрации программы для ЭВМ, № 2019616903. – Владимир. – 2019.
13. **Аль-Дайбани А.М.** Комплексная программа анализа влияния оконной функции на результаты нормализации параметров речевого сигнала/А.М Аль-Дайбани, Е.К Левин //Владимирский государственный университет//Свидетельство о регистрации программы для ЭВМ, № 2019616650. – Владимир. – 2019.
14. **Аль-Дайбани А.М.** Оценка эффективности применения спектрального вычитания и фильтра Винера для подавления помех при автоматическом распознавании речи/А.М Аль-Дайбани, Е.К Левин//Владимирский

- государственный университет // Свидетельство о регистрации программы для ЭВМ, № 2019619052. – Владимир. – 2019.
15. **Аль-Дайбани А.М.** Оценка эффективности использования нормализованных параметров речевого сигнала в телефонных системах автоматического распознавания речи / А.М Аль-Дайбани, Е.К Левин // Владимирский государственный университет // Свидетельство о регистрации программы для ЭВМ, № 2019662082. – Владимир. – 2019.

Подписано в печать 22.10.2019 г. Формат 60x84/16.
Бумага для множительной техники. Гарнитура Таймс.
Печать на ризографе. Усл. печ. л. 1л. Тираж 100 экз.
Издательство

Владимирского государственного университета
имени Александра Григорьевича и Николая Григорьевича Столетовых.
600000, Владимир, ул. Горького, 87.