Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
**«Владимирский государственный университет
имени Александра Григорьевича и Николая Григорьевича Столетовых»
(ВлГУ)**

Аль-Дайбани Абдулгани Мохаммед Салех

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

05.12.13 – Системы, сети и устройства телекоммуникаций

Диссертация
на соискание учёной степени
кандидата технических наук

Научный руководитель:
доктор технических наук доцент
Левин Евгений Калманович

Владимир – 2019

СОДЕРЖАНИЕ

Введение

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 неспокойная 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 [32, 36, 39, 84, 87].

Наличие диалектов повышает степень изменчивости речи, что обусловливает увеличение числа ошибок распознавания. В ряде работ [32, 36, 39, 85, 88] 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 [32, 36, 39, 84, 87, 95]. 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 [63]. 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 [89, 94, 44]. 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. [25,44,63,76,94]. 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.

Внедрение результатов работы

Результаты диссертационной работы внедрены в учебный процесс на кафедре радиотехники и радиосистем Владимирского государственного университета имени Александра Григорьевича и Николая Григорьевича Столетовых (ВлГУ) а также в центре речевых технологий ООО ЦРТ «Центр речевых технологий».

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 (USBEREIT). (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.

Структура и объём диссертации.

Диссертация состоит из введения, четырёх глав, заключения, библиографического списка, включающего 100 наименований, списка сокращений и 7 приложение. Объём диссертации составляет 128 страниц машинописного текста, 47 рисунков и 26 таблиц. Объём приложений составляет 22 страницы.

ГЛАВА 1 Особенности автоматического распознавания арабской речи

1.1 Особенности арабского языка

На арабском языке говорят более 350 миллионов человек (по оценкам 2017 года), его используют более чем в 22 странах [18, 32, 33, 36, 38, 39, 43, 58, 84, 86, 88]. Язык характеризуется большим разнообразием диалектов. Стандартизированным диалектом является современный диалект арабского языка (Modern Standard Arabic - MSA) [58, 96]. Он является официальным языком арабского мира. MSA преподается в школах и является основным языком в новостных передачах, парламенте и официальной речи в целом. Этот язык чаще используется при письме, чем в устной речи.

В арабских текстах (порядок написания – справа налево), как правило, используются буквы, соответствующие согласным звукам. Для указания гласных звуков используются диакритические знаки, которые обычно не указываются. Произнесения на арабском языке начинаются всегда с согласного звука. MSA включает в себя 28 согласных звуков, три коротких и три длинных гласных, а также два дифтонга. Для отображения арабских текстов исследователи, не владеющие арабским языком, пользуются транслитерацией Buckwalter [36]. Она соответствует привычному для них порядку написания текстов слева направо.

При повседневном общении население использует диалектный арабский язык (DA). DA является основным языком для драматических, комедийных программ и во многих жанровых передачах. Арабские диалекты могут рассматриваться как настоящие формы родного языка. Стандартных систем диалектного правописания нет. Из-за значительных отличий арабские диалекты можно рассматривать как разные языки при решении таких задач, как, например, автоматическая идентификация диалекта [36, 37, 83].

Различают следующие основные группы диалектов [3, 4, 37, 39, 42, 47, 62, 61]. Египетский арабский диалект (EGY) - охватывает диалекты долины Нила: Египет и Судан. Левантийский диалект (LAV) - включает в себя диалекты Ливана, Сирии, Иордании, Палестины. Диалекты стран арабского залива (GLF) - включают в себя диалекты Кувейта, Объединенных Арабских Эмиратов, Бахрейна и Катара. Североафриканский диалект (NOR) - охватывает диалекты Марокко, Алжира, Туниса и Мавритании. Различают также Иракский (IRQ) и Йеменский (Yem) диалекты. Таблица 1.1. показывает степень отличия произнесения одних и тех же фраз на разных диалектах Йемена и на MSA.

Таблица 1.1. Примеры диалектного произношения

MSA	**СД**	**ЮД**	**ЗД**	**English**
tissʕah tissaah	tissʕah tissaah	tissʕih tissaih	tissʔih tissiih	Nine

Из приведенного примера видно, что имеются существенные различия в произношении одних и тех же слов на разных диалектах, распространяющихся в Йемене, северный диалект (СД), южный диалект (ЮД), западный диалект (ЗД) и стандартный арабский язык (MSA). Большое отличие диалектов обусловливает высокую изменчивость произнесения одних и тех же слов, что увеличивает число ошибок распознавания. Поэтому возникает необходимость включения идентификатора диалекта в состав системы распознавания речи [36, 69, 68, 70].

Диалекты можно различать, используя их отличия на разных уровнях: на фонетическом, фонотаксическом, лексическом [39, 41, 67]. Использование машинного обучения при создании искусственных нейронных сетей позволяет осуществить идентификацию диалектов арабского языка. Однако обучение нейронных сетей требует наличия большого объема заранее подготовленных

аудиозаписей. Например, использование нейронных сетей при различии таких арабских диалектов, как египетский и диалект MSA обеспечивает точность идентификации 85,5%, [61].

В работе [47] решалась задача идентификации иорданских и египетских диалектов. В качестве параметров речевых сигналов использовались мел-частотные кепстральные коэффициенты (МЧКК - MFCC) и коэффициенты, полученные в результате вейвлет анализа. Была достигнута точность идентификации 80%.

В работе [32] рассмотрено два подхода, которые используют универсальную фоновую модель (UBM) в системе автоматической идентификации пяти арабских диалектов: Магриба: марокканского, тунисского и трех алжирских диалектов, которые относятся к западным, центральным и восточным районы Алжира. Получена точность идентификации равная 80,49%.

В работе [82] рассматривается классификатор, использующий метод опорных векторов. По результатам эксперимента достигнута точность классификации равная 93%.

В работе [55] исследовались возможности использования скрытых марковских моделей (Hidden Markov Models - HMMs) для построения моделей арабских диалектов с целью их последующей идентификации. Кроме того, используется гауссова смесь распределений (GMM). Если использовать в качестве параметров речевых сигналов MFCC совместно с их первыми и вторыми производными, то достигается точность идентификации равная 96,7%.

В работе [43] исследовался определитель MSA с использованием HMM. Монофонные акустические модели построены с использованием трех состояний. Плотность распределения для каждого состояния описывается смесью из 12 гауссианов. Используются МЧКК-MFCC. Длительность каждого кадра - 25 мс, со сдвигом кадра 10 мс. Каждый вектор признаков имеет 39 коэффициентов: 12

MFCC, энергия, 13 первых и 13 вторых производных. Параметры речевого сигнала нормализуются по среднему значению. Рассматривалась идентификация MSA, ливийского, египетского, иракского диалектов, а также диалект арабского залива. Достигнутая точность идентификации в зависимости от вида диалекта находилась в пределах (68 – 98) %

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 в качестве параметров речевых сигналов.

1.2 Состав системы автоматического распознавания речи

Автоматическое распознавание речи определяется как процесс преобразования речевого сигнала (РС) в соответствующую (наиболее вероятную) последовательность W_h слов. Речевые данные (данные наблюдения) на входе алгоритма распознавания представляют собой последовательность O наборов (векторов) параметров РС [36, 64]. Отсюда следует, что

$$W_h = arg \max_{\omega} P(\omega|o)P(\omega) \quad (1.1)$$

где ω – слово из соответствующего тематического словаря; $P(\omega)$ - вероятность появления слова – определяется моделью национального языка; $P(\omega|o)$ - условная вероятность слова, соответствующая данным o наблюдения.

Из выражения (1.1) видно, что система распознавания речи использует акустическое и языковое моделирование. Модель национального языка определяет вероятность появления слова по последовательности предыдущих

слов. Для создания модели широко используются искусственные нейронные сети (ИНС) [29, 53, 84].

Целью акустического моделирования является обучение модели, которая может сопоставить вектор $\boldsymbol{o}$ наблюдения с наиболее вероятной последовательностью знаков транскрипции, которую можно далее преобразовать в последовательность букв. Из-за временной и тембральной изменчивости РС наиболее приемлемыми для акустического моделирования оказались скрытые марковские модели (HMM) [36, 63].

Вектор наблюдения $\boldsymbol{o} = [o_1, o_2, \ldots, o_T]$ является результатом предварительной обработки входного сигнала. При формировании вектора на стадии предварительной обработки РС стараются обеспечить его независимость от особенностей произношения диктора, от помех акустического окружения диктора, от влияния канала связи.

На стадии предварительной обработки РС определяются паузы, подавляются аддитивные помехи. Затем определяются параметры РС, которые поступают на вход алгоритма распознавания. В качестве таких параметров широкое распространение получили Мел частотные Кепстральные Коэффициенты – МЧКК (Mel-Frequency Cepstral Coefficients – MFCC) [36, 63]. Рассмотрим основные этапы предварительной обработки сигнала при формировании MFCC, которые отображены на рисунке 1.1.

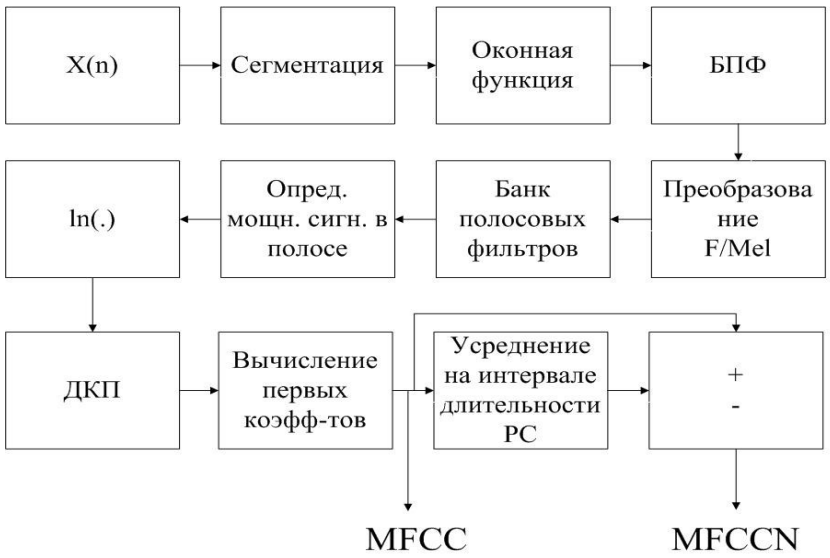


Рисунок 1.1. Схема обработки речевого сигнала при формировании MFCC

Сначала данные сигнала сегментируются, затем каждый сегмент РС взвешивается оконной функцией, и осуществляется быстрое преобразование Фурье (БПФ) – формируется кратковременный спектр сигнала. Для учета особенностей человеческого слуха частотная шкала преобразуется в мел-шкалу согласно выражению $\mathrm{m} = 1125 \ln\left(1 + \frac{\mathrm{f}}{700}\right)$ [64]. Далее мел-частотный спектр каждого сегмента равномерно разбивается на отдельные полосы набором полосовых фильтров, и определяется мощность сигнала на выходе каждого фильтра.

Полученный набор значений мощностей P_{sf} логарифмируется. Затем к результату логарифмирования каждого сегмента применяется дискретное косинусное преобразование (ДКП) – формируется кепстр РС. Несколько первых коэффициентов ДКП оставляются, остальные коэффициенты удаляются – формируется набор (вектор) MFCC.

Для подавления влияния частотной характеристики канала связи на параметры РС производится нормализация МЧКК следующим образом. По полученной временной последовательности наборов (векторов) МЧКК

определяется среднее значение вектора во времени. Среднее значение вычитается из каждого вектора – формируется набор МЧКК, нормализованных по среднему значению (МЧККН). Для учета динамических характеристик речи полученный набор параметров РС дополняется первыми и вторыми производными МЧКК [63].

Рассмотренная выше нормализация МЧКК по среднему значению используется при распознавании коротких произнесений. Если же распознается слитная речь, то нормализация осуществляется с помощью фильтра (метод RASTA) [63, 56, 66], который удаляет постоянную составляющую во временной последовательности MFCC.

Следует отметить, что в составе звуков арабской речи есть такие звуки, которые отсутствуют в европейских 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.

В работах [4, 5, 36, 39, 84] показано, что создание систем распознавании арабской речи является сложной задачей даже при отсутствии помех. Во-первых, разработку САРР сдерживает отсутствие объемных выборок звукозаписей, подготовленных для обучения и тестирования САРР. При подготовке наборов аудиозаписей необходимых для обучения системы и ее тестирования необходимо провести большую работу по простановке диакритических знаков и провести затем транскрибирование произнесений на различных диалектах. Особенно актуально наличие указанных выборок в настоящее время, когда в составе САРР используются искусственные нейронные сети [36]. Для их обучения требуются очень объемные выборки звукозаписей.

Во-вторых, наличие диалектов, обусловливает необходимость использования идентификатора диалекта. Ошибки идентификации увеличивают число ошибок распознавания.

Проведенные исследования, в основном, направлены на создание систем распознавания речи для MSA. Исследования, сориентированные на распознавание диалектной речи очень малочисленны [77, 36].

1.3 Подавление помех на стадии предварительной обработки сигналов

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 [9, 10, 13, 14, 15, 50].

Для очистки зашумленной речи применяют спектральное вычитание (СВ) и фильтр Винера (ФВ) [1, 30, 31, 44, 89, 94]. При использовании СВ оценивается спектральная плотность мощности помехи на интервале паузы РС, и полученная оценка вычитается из оценки зашумленного сигнал.

Однако на практике можно получить лишь оценки спектральной плотности мощности на ограниченных интервалах времени. Оценки могут значительно отличаться от значения спектральной плотности мощности. Данный факт приводит к появлению отрицательных значений разностей, когда отношение сигнал-помеха невелико. Для устранения данного явления результаты вычитания корректируют, что приводит к неполному подавлению

помехи и искажает параметры речевого сигнала. Обычно используется следующее правило коррекции разности [44, 89, 94].

$$\hat{S}_x(\omega)^2 = max\{\hat{S}_y(\omega)^2 - \alpha\hat{S}_n(\omega)^2, \lambda\}. \quad (1.2)$$

Здесь $\lambda \geq 0$ настраиваемый порог, α - коэффициент, корректирующий оценку спектральной плотности шума, $\hat{S}_y(\omega)$ - модуль оценки спектра зашумленного сигнала, $\hat{S}_x(\omega)$ - модуль оценки спектра очищенного от шума сигнала, $\hat{S}_n(\omega)$ - модуль оценки спектра шума. Если разность меньше порога, то в качестве оценки спектральной плотности мощности очищенного сигнала принимается указанное значение порога.

Необходимость коррекции разности приводит к появлению дополнительной помехи в очищенном сигнале. Помеха получила название "музыкальный шум", так как при прослушивании очищенного сигнала она воспринимается как что-то, похожее на музыку.

"Музыкальный шум" отсутствует, когда для подавления помехи используется ФВ. Подавление помех с помощью фильтра Винера осуществляется при прохождении суммы сигнала и помехи через указанный фильтр. Частотная характеристика *G(f)* фильтра формируется так, чтобы минимизировать среднеквадратическое отклонение очищенного от помех сигнала от "чистого" сигнала [44, 89, 94].

$$G(f) = \frac{S_p(f)}{S_p(f) + N_p(f)}, \quad (1.3)$$

где $S_p(f)$, $N_p(f)$ – спектры плотности мощности сигнала и шума соответственно. Путем деления каждого члена выражения (1.3) на $N_p(f)$ получаем:

$$G(f) = \frac{SNR(f)}{1 + SNR(f)}, \quad (1.4)$$

где SNR (Signal Noise Ratio) – отношение спектральных мощностей сигнала и помехи. Спектр сигнала после очистки

$$\hat{S}(f) = X(f) * G(f), \tag{1.5}$$

где *X(f)=S(f)+N(f).* – спектр зашумленного сигнала.

При цифровой обработке речевой сигнал разбивается на отдельные кадры (сегменты). Длительность сегмента выбирается из условия квазистационарности речевого сигнала (считают, что параметры речевого сигнала на интервале длительности сегмента не изменяются). Для каждого сегмента выполняется быстрое преобразование Фурье (БПФ) – определяется кратковременный спектр каждого сегмента. Для каждого сегмента определяется частотная характеристика фильтра Винера.

$$G(p,k) = \frac{SNR(p,k)}{1 + SNR(p,k)}, \tag{1.6}$$

где p – номер кадра; k – номер спектральной составляющей (индекс коэффициента БПФ).

Однако непосредственно использовать фильтр Винера для очистки зашумленного сигнала нельзя, потому что не известен спектр чистого сигнала. Вторая проблема заключается в том, что на практике можно использовать лишь оценки спектральной плотности мощности сигнала (так как собственно спектральная плотность мощности определяется путем усреднения на бесконечно большом интервале времени, что практически не реализуемо).

Так как в речевом сигнале имеются паузы, то на интервале ее длительности можно получить оценку спектральной мощности помехи (шума) $\hat{\gamma}_n(p,k)$. В этом случае становится доступной оценка отношения мощности зашумленного сигнала к мощности шума – апостериорная *(a posteriori)* оценка [89]:

$$\widehat{SNR}_{post}(p,k) = \frac{\left|\hat{X}(p,k)\right|^2}{\hat{\gamma}_n(p,k)}. \tag{1.7}$$

Однако для построения фильтра Винера требуется априорная (*a priori)* оценка отношения мощности чистого сигнала к мощности шума $\widehat{SNR}_{prio}(p,k)$. Для ее нахождения используется метод прямого решения (DD), который характеризуется нижеприведенным алгоритмом:

$$\widehat{SNR}^{DD}{}_{prio}(p,k) = \beta \frac{\left|\hat{S}(p-1,k)\right|^2}{\hat{\gamma}_n(p,k)} + (1-\beta)P\left[\widehat{SNR}_{post}(p,k) - 1\right], \tag{1.8}$$

где $\beta \approx 0{,}98$; $\hat{S}(p-1,k)$ – оценка спектра речевого сигнала в предыдущем сегменте, P – оператор «полуволнового выпрямления» - отрицательные значения разности заменяются малым положительным числом α

$$P\left[\widehat{SNR}_{post}(p,k) - 1\right] = \max_{\alpha \geq 0}\left\{\widehat{SNR}_{post}(p,k) - 1, \alpha\right\}. \tag{1.9}$$

Найденная оценка может использоваться для построения фильтра Винера (в дальнейшем и для очистки зашумленного сигнала), частотная характеристика которого определяется следующим выражением:

$$G_{DD}(p,k) = \frac{\widehat{SNR}^{DD}{}_{prio}(p,k)}{1 + \widehat{SNR}^{DD}{}_{prio}(p,k)}; \tag{1.10}$$

$$\hat{S}(p,k) = \hat{X}(p,k) * G_{DD}(p,k). \tag{1.11}$$

Из приведенных выражений следует, что оценка отношения сигнал-помеха, в основном, использует данные оценки спектра сигнала не на текущем, а на

предыдущем кадре, поэтому при прослушивании очищенного сигнала наблюдается небольшое эхо – эффект реверберации.

Чтобы устранить эффект реверберации, применяется двухступенчатый алгоритм определения частотной характеристики фильтра Винера (Two step noise reduction - TSNR) [44, 89, 94]. Первая стадия алгоритма полностью совпадает с определением вышеописанных формул. На втором этапе найденная частотная характеристика фильтра Винера используется для получения уточненной оценки отношения сигнал-помеха следующего вида:

$$\widehat{SNR}^{TSNR}{}_{prio}(p,k) = \frac{|X(p,k) * G_{DD}(p,k)|^2}{\hat{\gamma}_n(p,k)}. \quad (1.12)$$

Полученное выражение используется для построения частотной характеристики:

$$G_{TSNR}(p,k) = \frac{\widehat{SNR}^{TSNR}{}_{prio}(p,k)}{1 + \widehat{SNR}^{TSNR}{}_{prio}(p,k)}. \quad (1.13)$$

Затем сигнал пропускается через фильтр.

$$\hat{S}_{TSNR}(p,k) = \hat{X}(p,k) * G_{TSNR}(p,k) \quad (1.14)$$

Очищенный сигнал свободен от музыкального шума и эффекта реверберации.

Так как при подавлении помех появляются искажения сигналов, что ведет к появлению ошибок распознавания, то возникает задача экспериментального определения условий использования методов подавления помех, когда число ошибок распознавания находится в допустимых пределах.

На рисунке 1.2 показаны блок-схемы алгоритмов, реализующие методы спектрального вычитания и фильтра Венера [30, 44, 89].

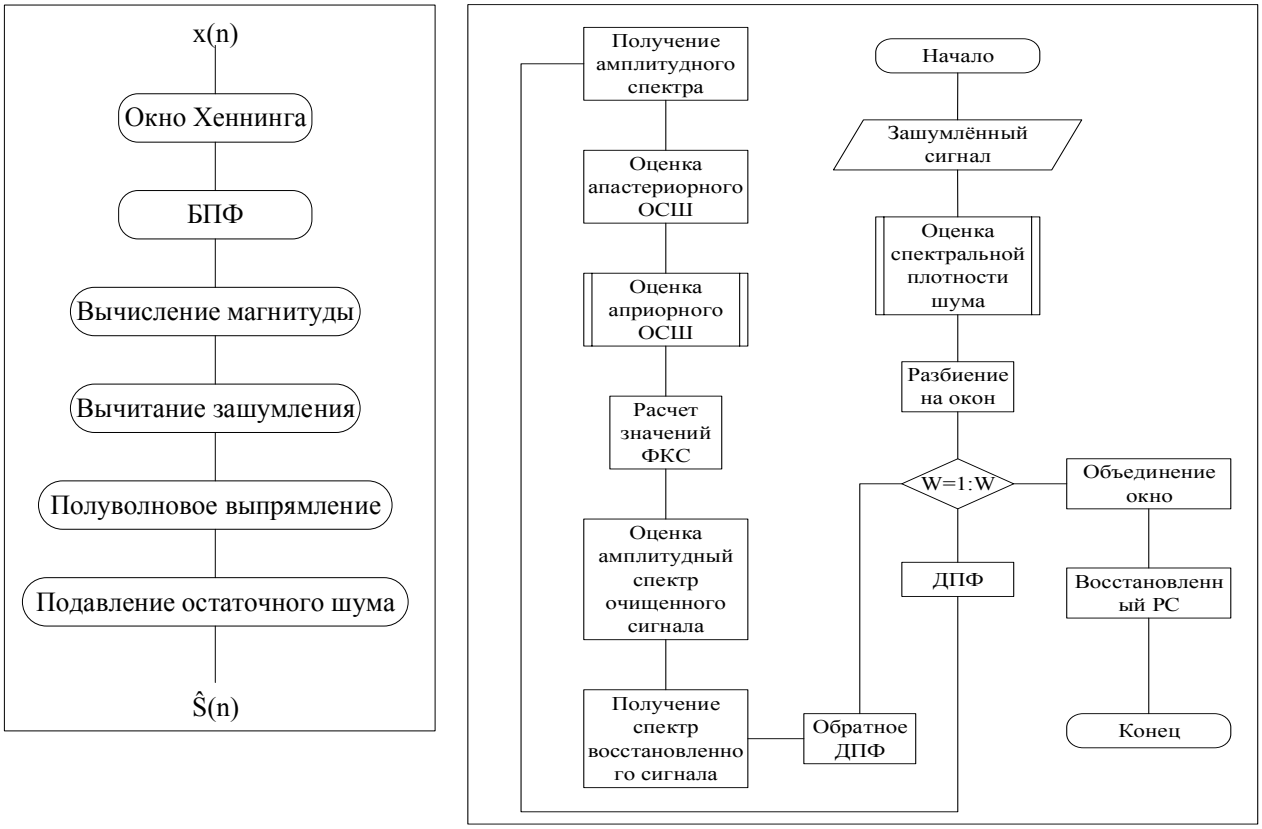


Рисунок 1.2. Алгоритмы подавления помехи, (а) алгоритм основан на СВ, (б) алгоритм основан на ФВ

Рассмотрим возможность использования спектрального вычитания (СВ) и фильтра Винера (ФВ) в составе САРР, предназначенной для распознавания произнесений арабских названий цифр. В таблице 1.2. показаны транскрипции названий десяти цифр в MSA [38, 93].

Рассмотрим сначала изменения МЧКК при подавлении помех. Используем следующую методику 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.

Блок-схема соответствующего алгоритма приведена на рисунке 1.3.

Таблица 1.2. Названия арабских цифр

Цифра	Арабское написание	Произношение
1	واحِد (wahid)	wâ-hĕd
2	إثْنَيْن (ithniyn)	‘aâth-nāyn
3	ثَلاثَة (thalathah)	thâ-lă-thâh
4	أَرْبَعَة (Arbaah)	‘aâr-bâ-‘aâh
5	خَمْسَة (khamsah)	khâm-sâh
6	سِتّة (Sittah)	sĕt-tâh
7	سَبْعَة (Sabaah)	sûb-‘aâh
8	ثَمانِيَة (Thamāniyah)	Thâ-mă-nĕ-yĕh
9	تِسْعَة (Tisaah)	tĕs-âh
0	صِفْر (Sifr)	sĕfr

Результаты эксперимента представлены Таблицей 1.3. Из данных таблицы следует, что использование СВ и ФВ для подавления помехи, как правило, уменьшает влияние помехи на изменение МЧКК. Однако степень уменьшения влияния снижается при увеличении отношения сигнал-шум. В частности, при SNR =35 дБ параметры РС при произнесении названия цифры "6" меняются в большей степени при использовании СВ и ФВ по сравнению с отсутствием средств подавления помехи. Следовательно, при указанном отношении сигнал-шум использование СВ и ФВ становится нецелесообразным.

Рисунок 1.3. Алгоритм оценки эффективности использования СВ и ФВ на результаты параметров МЧКК

Следует отметить, что использование ФВ при малых отношениях сигнал-шум является более целесообразным по сравнению с применением СВ.

Количественно эффективность применения средств подавления помехи можно оценить, сравнивая значения SNR для случаев наличия и отсутствия помехи при равных искажениях МЧКК. Например, при SNR=20дБ использование ФВ при произнесении названия цифры 1 обусловливает степень искажений равную 3,98. Такая же степень искажений при отсутствии средств подавления помехи соответствует SNR= (30 – 35) дБ. Можно говорить, что

использование ФВ эквивалентно снижению уровня помехи на (10 - 15) дБ. Использование СВ соответствует увеличению сигнал-шум примерно на 5 дБ.

Таблица 1.3. Зависимость изменения параметров РС от отношения сигнал-шум при использовании СВ и ФВ для подавления помехи

SNR, дБ	**Значения дисперсий разностей (без подавления помехи)**									
	Арабские названия цифр									
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
5	8,20	18,43	12,14	16,12	19,77	12,17	7,67	10,70	16,05	11,35
10	6,13	14,30	9,28	13,11	15,26	9,41	5,48	8,32	12,40	8,80
15	4,16	11,67	6,17	9,78	11,42	7,16	3,78	5,93	9,44	6,56
20	2,94	9,26	4,69	7,33	8,70	5,12	2,57	4,48	7,37	4,70
25	1,50	6,71	3,27	5,36	6,37	3,43	1,50	2,88	5,33	3,41
30	1,06	4,78	2,09	3,53	4,63	2,49	0,78	1,74	3,80	2,33
35	0,59	3,24	1,13	2,33	3,16	1,43	0,43	1,18	2,77	1,64
Использование спектрального вычитания										
5	5,36	14,21	8,88	12,42	15,06	9,11	5,85	7,71	12,50	8,52
10	4,45	10,54	6,58	9,36	11,52	6,84	3,48	5,90	8,98	6,78
15	2,61	8,66	4,28	7,20	8,52	5,06	2,57	4,37	6,81	4,65
20	1,79	6,28	2,87	4,86	6,02	3,75	1,59	2,99	5,31	3,28
25	0,99	4,66	1,84	3,75	4,28	2,36	1,01	1,89	3,98	2,46
30	0,84	3,27	1,40	2,37	3,16	1,73	0,69	0,97	2,66	1,73
35	0,54	2,21	0,70	1,62	2,10	1,14	0,48	0,77	2,22	1,24
Использование фильтра Винера										
5	3,63	7,79	4,86	7,35	8,07	5,22	2,80	3,95	6,94	4,45
10	2,65	6,85	2,91	5,42	5,97	3,33	2,12	3,09	5,03	3,57
15	2,26	4,39	1,86	4,07	4,10	2,99	1,92	2,12	3,75	2,91
20	2,01	3,98	1,81	2,85	3,06	2,40	1,49	1,70	2,85	2,17
25	1,98	3,03	1,50	2,44	2,48	2,00	1,62	1,56	2,29	1,93
30	1,90	2,32	1,27	2,13	1,78	1,72	1,30	1,42	1,78	1,66
35	1,93	1,94	1,37	1,91	1,50	1,49	1,17	1,40	1,58	1,47

Эффективность использования ФВ и СВ для подавления помех в сильной степени зависит от вида РС, что на рисунке 1.4 проиллюстрировано графиками зависимости изменений параметров РС от названия цифры. Из рисунка следует, что использование ФВ, как правило, обеспечивает меньшие искажения МЧКК по сравнению с использованием СВ. Исключение в данном случае соответствует произнесению названия цифры 0.

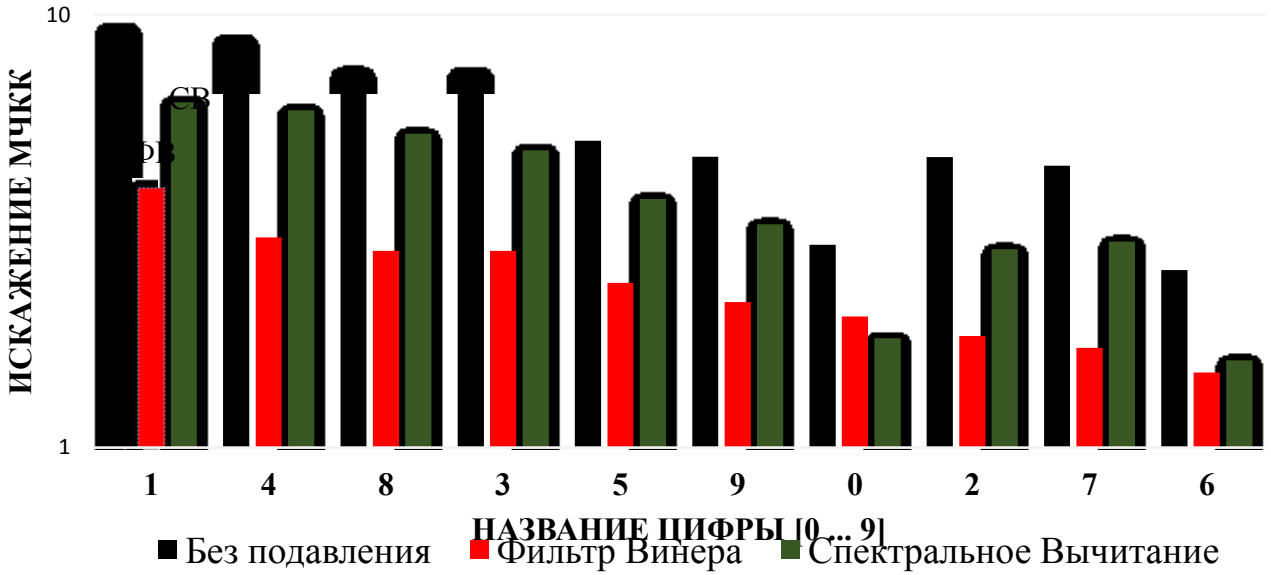


Рисунок 1.4. Зависимость изменений МЧКК от названия цифры (SNR= 20 дБ)

Рассмотрим теперь влияние используемых методов подавления помехи на результаты распознавания. Блок-схема алгоритма проведения эксперимента представлена на Рисунке 1.5.

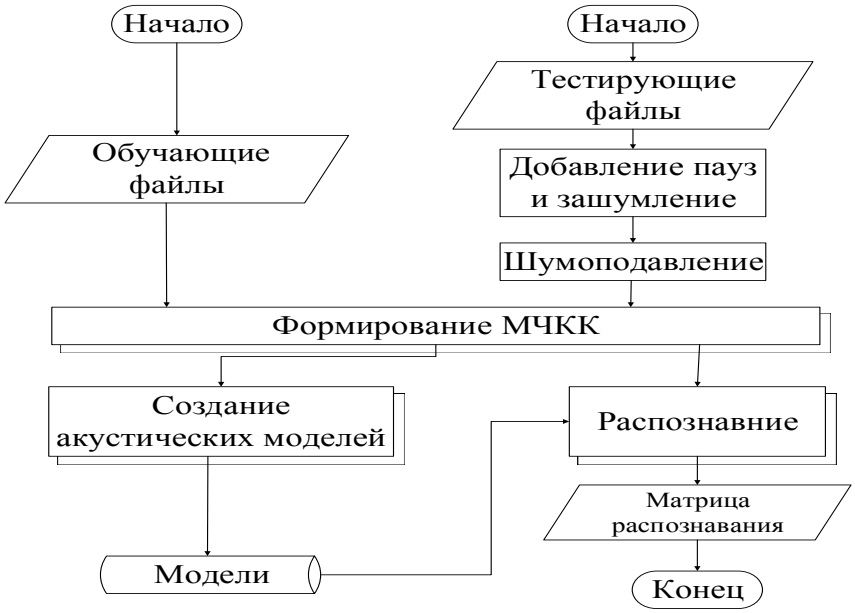


Рисунок 1.5. Алгоритм оценки влияния СВ и ФВ на результаты распознавания

Результаты экспериментов представлены в Таблицах 1.4 -1.5 для случая тестирования САРР при отсутствии идентификации диалекта. Результаты относятся к ситуации, когда в качестве помехи использовался аддитивный белый гауссов шум (АГБШ) и шум дождя.

Таблица 1.4. Достоверность (точность) распознавания при SNR =20 дБ при отсутствии идентификации диалекта и наличии АГБШ

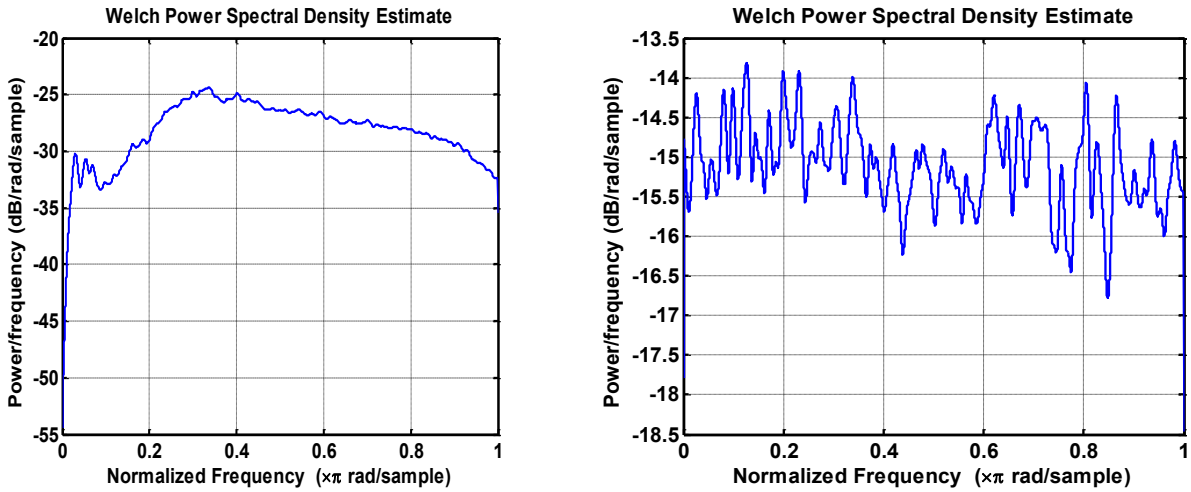
Условия эксперимента	**Значения достоверности распознавания, %**									
	Арабские названия цифр (ВсеД)									
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Без подавления помехи	92,47	23,53	24,47	35,06	14,35	18,35	15,29	35,06	7,53	81,18
СВ	97,18	48,00	46,82	65,41	55,06	57,65	63,06	54,12	30,12	94,12
ФВ	55,76	47,29	58,59	59,06	66,59	96,24	51,53	58,35	50,59	85,18

Таблица 1.5. Достоверность (точность) распознавания при SNR =20 дБ в случае отсутствии идентификации диалекта и наличия реального шума (шум дождя)

Условия эксперимента	**Значения достоверности распознавания, %**									
	Арабские названия цифр (ВсеД)									
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Без подавления помехи	91,06	30,82	39,29	37,18	7,06	30,35	19,29	35,76	18,35	89,88
СВ	89,18	48,94	72	74,35	61,41	67,06	52,94	54,59	50,82	98,35
ФВ	29,65	74,82	76,24	56,71	76,94	89,18	33,65	54,12	67,29	93,88

Из данных таблиц следует, что достоверность распознавания значительно зависит от вида речевого сигнала – произнесения названия цифры и вида помехи. Обозначения: (ВсеД) – при выполнении эксперимента использованы все диалекты; Результаты экспериментов при воздействии шума автобуса и шума офиса приведены в приложении 6.

На рисунке 1.6. приведены графики спектральных плотностей мощности аддитивного белого гауссова шума и реального шума - шума дождя.



Шум дождя АБГШ

Рисунок 1.6. Спектральная плотность мощности шума

Рассмотрим теперь зависимость достоверности (точности) распознавания от отношения сигнал-шум. На рисунке 1.7. приведена зависимость значения достоверности (точности) распознавания, усредненного по всем названиям цифр от отношения сигнал-шум для случая АБГШ.

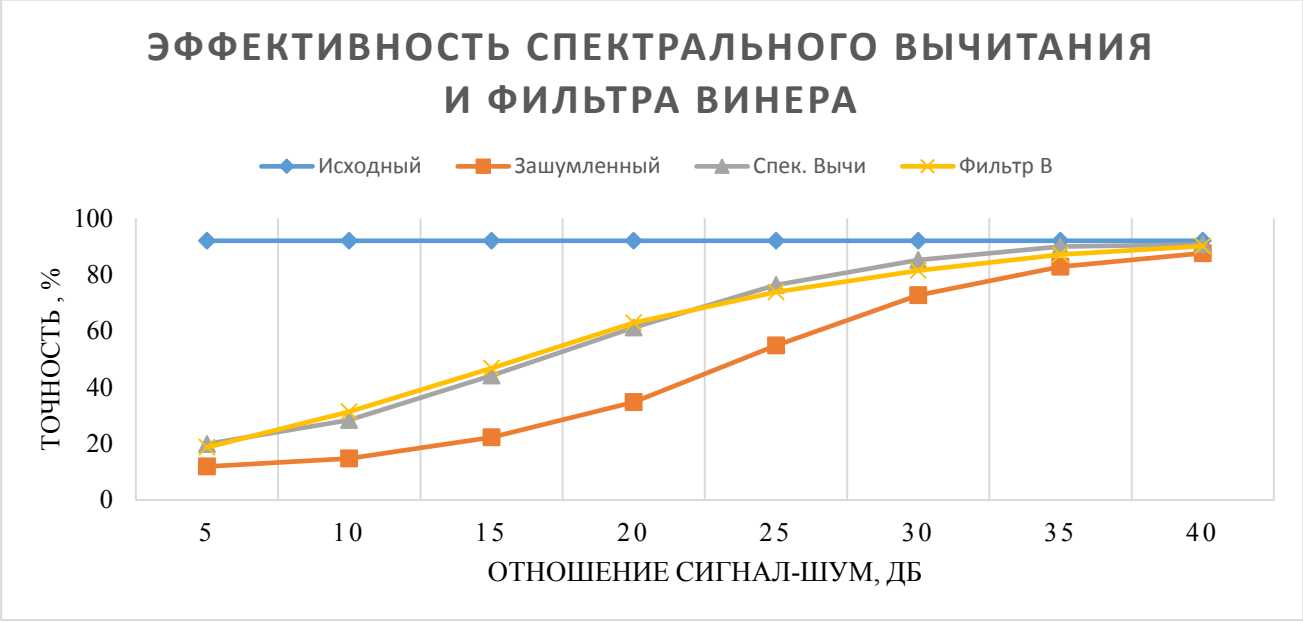


Рисунок 1.7. Зависимость достоверности распознавания от отношения сигнал-шум

Из графиков следует, что при отношении сигнал-шум менее 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.

Выводы по разделу 1

1. Особенностью арабской речи является большое разнообразие диалектов, что ведет к росту числа ошибок распознавания. Из обзора литературы следует, что проведенные исследования, в основном, направлены на создание систем распознавания речи для стандартного арабского языка MSA. Исследования, сориентированные на распознавание диалектной речи, очень малочисленны.
2. Для снижения числа ошибок распознавания в состав САРР необходимо включать идентификатор диалектов.
3. В проведенных исследованиях приводятся лишь данные эксперимента об ошибках идентификации, анализ вероятности ошибки идентификации отсутствует.
4. Создание систем распознавания арабской речи сдерживается отсутствием объемных выборок звукозаписей, подготовленных для обучения и тестирования САРР.
5. Для снижения влияния канала связи на достоверность распознавания используется нормализация параметров речевого сигнала по среднему значению. В литературных данных отсутствуют результаты исследований причин нестабильности нормализованных параметров.
6. Для подавления помех в речевых сигналах используется спектральное вычитание и фильтр Винера. Данные методы наряду с подавлением помех

меняют значения параметров сигнала, что может привести к увеличению числа ошибок распознавания.

7. Проведенный эксперимент показывает, при отношении сигнал-шум менее 35 дБ применение спектрального вычитания и фильтра Винера для подавления помех повышает достоверность распознавания.

ГЛАВА 2 Использование идентификации диалекта при распознавании голосовых команд в телефонии

2.1 Характеристика исследуемого метода идентификации диалекта

Йеменский диалект арабского языка – это совокупность разновидностей арабского языка‹ распространённых в Йемене‹ а также на юго-западе Саудовской Аравии‹ в Сомали и Джибути. Йеменский арабский язык считается очень консервативным, так как в нём сохранились многие черты классического арабского, которые не нашли распространения в большей части арабского мира [3, 4, 5, 6, 34, 42, 41, 43, 45, 46, 69, 100].

Йеменский диалект можно разделить на несколько основных диалектных групп, каждая из которых обладает своей лексикой и фонетикой. Наиболее значительными из этих групп являются: севернойеменский (СД) диалект (диалект Саны), южнойеменский (ЮД) диалект (таизско-аденский) и западный (ЗД) тихамейский (Tihamiyya). Количество носителей СД в стране составляет 68,3% от общего населения, количество носителей ЮД - 13,4%, количество носителей ЗД - 12,7% [3, 4, 5, 6, 17, 26, 42 69, 100].

Рассмотрим возможность использования акустических моделей произнесений тестовых слов для оперативной идентификации одного из трех йеменских диалектов в процессе обращения пользователя к автоматической телефонной справочной системе [2, 5, 23, 27, 41, 43, 48, 49, 99]. Рассмотрим случай, когда САРР предназначена для распознавания произнесений названий отдельных цифр.

В таблице 2.1 указаны транскрипции произнесений названий цифр для указанных диалектов. Здесь первая слева транскрипция относится к севернойеменскому диалекту (СД), вторая – к южнойеменскому диалекту (ЮД), третья – к тихамейскому диалекту (западному - ЗД).

Таблица 2.1. Транскрипции произнесений цифр для трех диалектов

Название цифры	**Транскрипция названия цифры по диалектам (СД / ЮД / ЗД)**	
	Фонетический алфавит	**Английский алфавит**
0	[sifr]/ [sifr] / [sifr]	Sifr / Sifr / Sifr
1	[wahid] / [wahid] / [wahid]	Wahid / Wahid / Wahid
2	[ʔiθnajn] / [ʔiθnijn] / [ʔiθnijn]	Ithnajn / Ithnijn / Ithnijn
3	[θalɑːθih] / [θalɑːθah] / [θalɑːθah]	θalaθih/θalaθah/θalaθah
4	[ʔarbʕah] / [ʔarbʕih] / [ʔarbʔih]	Arbaah / Arbaah / Arbaih
5	[xamsih] / [xamsah] / [xamsih]	Khamsih / Khamsah / Khamsih
6	[sittih] / [sittih] / [sittih]	Sittih / Sittih / Sittih
7	[sabʕah] / [sabʕih] / [sabʔih]	Sabaah / Sabaah / Sabaih
8	[θamanjih]/ [θamanjih] / [θamanjih]	θamanjih/ θamanjih / θamanjih
9	[tissʕah] / [tissʕih] / [tissʔih]	Tisaah/Tissaih/Tissaih

В таблице используются знаки транскрипции (ʔ, ʕ), которые обозначают звуки арабской речи, отсутствующие в системе звуков английской речи. Им соответствуют арабские буквы (ع,ء), обозначающие гортанные звуки с твердым приступом [34, 50, 51, 52, 93, 100].

Анализ транскрипций показал, что наиболее сильно различаются по диалектам произнесения цифр: 2, 3, 4, 5, 7, 9. Причем наибольшая степень отличий соответствует цифре 9. Транскрипции названий указанных цифр с использованием международного фонетического алфавита приведены в Таблице 2.2.

Таблица 2.2. Транскрипция наиболее различающихся произнесений цифр

Русское название цифры	**Транскрипция названий цифр по диалектам (СД / ЮД / ЗД)**
2("два")	[ʔiθnajn] / [ʔiθnijn] / [ʔiθnijn]
3("три")	[θalɑːθih] / [θalɑːθah] / [θalɑːθah]
4("четыре")	[ʔarbʕah] / [ʔarbʕih] / [ʔarbʔih]
5("пять")	[xamsih] / [xamsah] / [xamsih]
7("семь")	[sabʕah] / [sabʕih] / [sabʔih]
9("девять")	[tissʕah] / [tissʕih] / [tissʔih]

Анализируя отличия указанных транскрипций по диалектам, можно прийти к выводу о возможности их использовании при построении классификаторов рассматриваемых йеменских диалектов [59 64, 65].

Рассмотрим возможность автоматической идентификации диалекта путем учета различий в произнесениях названий цифр на трех основных диалектах республики Йемен, указанных выше. Целесообразность идентификации диалектов оценим экспериментально путем сравнения результатов автоматического распознавания названий цифр с учетом идентификации, когда для каждого названия цифры на каждом диалекте создается своя акустическая модель (HMM), и без ее учета, когда акустическая модель названия каждой цифры является общей для всех диалектов. Эксперимент проведен с использованием пакета Hidden Markov Model (HMM) Toolbox для системы Matlab. В качестве параметров речевого сигнала использованы 12 МЧКК – MFCC [9, 63, 74, 81, 85, 73, 80].

При создании акустических моделей использовались голоса 18 дикторов - носителей трех диалектов арабского языка (по 25 произнесений от каждого диктора). Тестирование системы осуществлялось с использованием голосов тех же дикторов, но произнесения при тестировании отличались от произнесений, использованных при создании моделей - при обучении системы (другие 25 произнесений).

Блок-схема эксперимента представлена на рисунке 2.1.

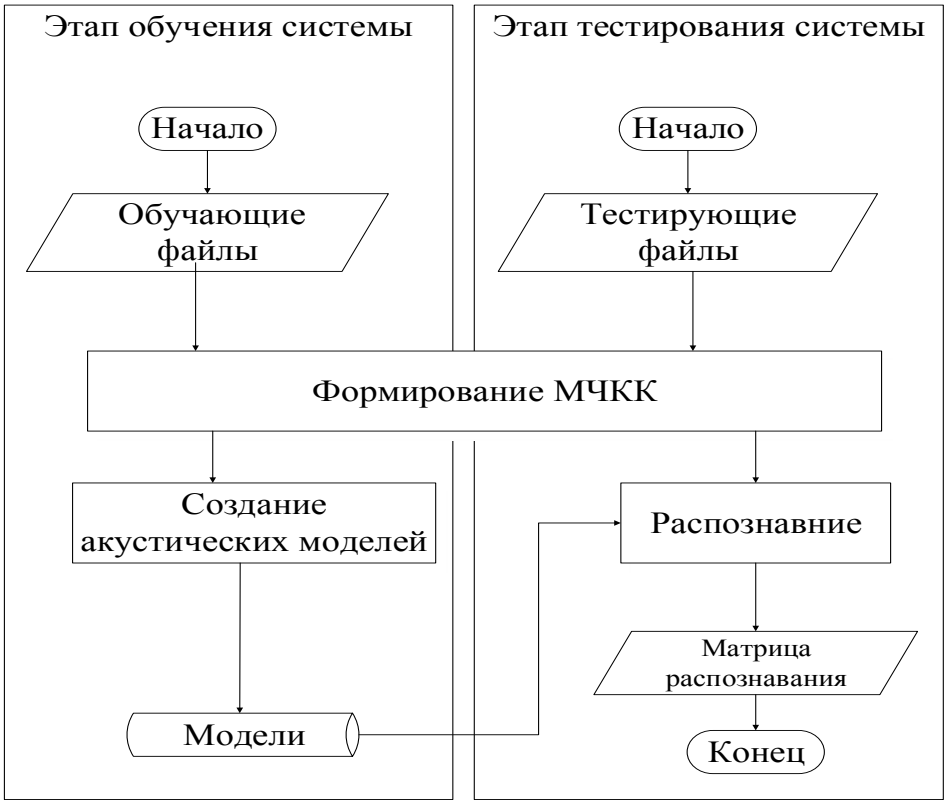


Рисунок 2.1. Блок-схема алгоритма эксперимента

Основные характеристики используемой в эксперименте системы распознавания, следующие:

- Число звуковых файлов, использованных при обучении системы 4500.
- Число звуковых файлов, использованных при тестировании системы 4500.
- Акустические модели произносимых названий используют модели скрытых Марковских процессов (МСМП).
- В качестве параметров речевого сигнала используется 12 MFCC.
- Частота дискретизации речевых сигналов – 16 кГц, количество разрядов квантования 16.

- Размер сегмента речевого сигнала равен 256 отсчетам, величина перекрытия сегментов составляет 128 отсчетов. Количество точек БПФ - 256.

При разработке систем автоматического распознавания речи требуется большой объем звукозаписей (база данных - БД), которые используются как для обучения, так и для тестирования системы. Общепринятых БД на арабском языке не существует. Поэтому для проведения эксперимента были созданы три БД произнесений названий цифр (0 – 9): - отдельная БД для каждого диалекта. Числовые характеристики всех БД одинаковы и представлены в таблице 2.3.

Таблица 2.3. Параметры базы данных

Дикторы		Число повторений		Число звукозаписей	
Обучение системы	Тестирование системы	Обучение	тестирование	Обучение	тестирование
Д1	Д1	25	25	150	150
Д2	Д2	25	25		
Д3	Д3	25	25		
Д4	Д4	25	25		
Д5	Д5	25	25		
Д6	Д6	25	25		
При тестировании системы использованы произнесения диктора Д7					

В таблице обозначены: Д1 – первый диктор, Д2 – второй, Д3 – третий, Д4 – четвёртый, Д5 – пятый, Д6 – шестой, Д7 – седьмой диктор.

При создании каждой модели использованы голоса шести дикторов, которые наиболее сильно отличаются по тембру и манере произнесения названий цифр. Для тестирования использован дополнительно голос седьмого диктора, который не входит в группу голосов дикторов, использованных при построении модели.

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 90,8 %. Если же для каждого диалекта используется своя совокупность моделей для распознаваемых произнесений (случай безошибочной идентификации диалекта), то такая же относительная частота для южного диалекта (ЮД) составляет 98,4%, для северного диалекта (СД) - 97,8% и для западного диалекта (ЗД) - 97,5%. Следовательно, идентификация диалектов позволяет на (7 - 8) % повысить относительную частоту правильного распознавания.

2.2 Идентификация при произнесении одного контрольного слова

Проанализируем 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 название цифры 9. Результатом идентификации является диалект, которому принадлежит акустическая модель, которой с наибольшей вероятностью соответствует произнесение.

Определим вероятность ошибочной идентификации.

$$\begin{aligned} P_{\text{Класс3}} &= P(C)\cdot P(\text{Ю}|C) + P(C)\cdot P(\text{З}|C) + P(\text{Ю})\cdot P(\text{С}|\text{Ю}) + P(\text{Ю}) \\ &\quad \cdot P(\text{З}|\text{Ю}) + P(\text{З})\cdot P(\text{Ю}|\text{З}) + P(\text{З})\cdot P(\text{С}|\text{З}) = \\ &= P(C)\cdot [P(\text{Ю}|C) + P(\text{З}|C)] + P(\text{Ю})\cdot [P(\text{С}|\text{Ю}) + P(\text{З}|\text{Ю})] + P(\text{З}) \\ &\quad \cdot [P(\text{Ю}|\text{З}) + P(\text{С}|\text{З})] \end{aligned} \tag{2.1}$$

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

$$P(C) = P(\text{Ю}) = P(\text{З}) = \frac{1}{3};$$

$$P_{\text{Класс3}} = \frac{1}{3}(P(\text{Ю}|C) + P(\text{З}|C) + P(\text{С}|\text{Ю}) + P(\text{З}|\text{Ю}) + P(\text{Ю}|\text{З}) + P(\text{С}|\text{З})) \quad (2.2)$$

Рассмотрим результаты тестирования идентификатора с использованием названия цифры 9. В таблице 2.4 указаны относительные частоты правильной и ошибочной идентификаций диалектов. Результаты тестирования идентификатора с использованием названий цифр (0 - 8) представлены в приложении 4.

Таблица 2.4. 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

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

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

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

В таблице 2.5. приведены экспериментальные данные по использованию названий других цифр для идентификации. При создании каждой модели использовались голоса шести дикторов, которые наиболее сильно отличаются по тембру и манере произнесения названий цифр. Тестирование алгоритма идентификации проводилось по двум вариантам. При первом варианте для тестирования использовался дополнительно голос седьмого диктора, который не входит в группу голосов дикторов, использованных при построении модели. При втором варианте тестирование проводилось при участии указанных выше шести дикторов. Однако произнесения при тестировании отличались от произнесений, использованных при обучении. Вероятности использования северного, южного и западного диалектов при этом, считались равными.

При определении оценки вероятности ошибки считаем, что полученные экспериментально относительные частоты ошибки классификации являются оценками соответствующих условных вероятностей.

Данные тестирования по всем цифрам представлены во втором столбце Таблицы 2.5. В третьем столбце представлены результаты тестирования по второму варианту. Второй вариант тестирования показал меньшую вероятность ошибки классификации, так как использовались одинаковые голоса дикторов при тестировании и обучении. Такая ситуация мало соответствует реальности. Первый вариант тестирования более реален на практике, так как позволяет избежать ошибочной настройки модели на индивидуальные особенности произнесения названия цифры для данной группы дикторов. Учитываются лишь особенности диалекта.

Таблица 2.5. Оценка вероятности ошибочной классификации на три группы

Название цифр	Оценка вероятности ошибочной классификации	
	Разные дикторы	Одинаковые дикторы
0	0,333	0
1	0,330	0,020
2	0,594	0,015
3	0,330	0,130
4	0,079	0,005
5	0,383	0,002
6	0,119	0,013
7	0,330	0,019
8	0,198	0
9	0,053	0

Из данных таблицы видно, что произнесение названия цифры 9 обеспечивает наименьшую вероятность ошибки идентификации.

2.3 Идентификация при произнесении двух контрольных слов

Рассмотрим теперь возможность использования для идентификации произнесения не одного, а двух названий цифр. При произнесении названий цифр, например, 2 и 3, транскрипция произнесений одинакова для диалектов ЮД и ЗД. Транскрипции диалекта «СД» отличаются от транскрипции диалектов ЮД. и ЗД. Создадим две акустические модели для произнесения названия цифры 2, одна модель «2с» для северного диалекта и вторая «2юз» - общая для южного и западного диалектов. Тогда при произнесении названия этой цифры с различными диалектами возможна классификация диалектов на две группы: «северный» и «южный + западный». Аналогичное рассуждение справедливо для

названий цифр: 5 и 6. При произнесении названий: 5 и 6 возможна классификация диалектов на две группы: «СЗ» и «Ю».

Рассмотрим вероятности ошибок классификации. Пусть имеется две группы диалектов, то есть используются две акустические модели, например, для диалектов: «Ю» и «СЗ». Вероятность ошибки классификации при использовании данных акустических моделей:

$$P_{\text{класс2}} = P(C) \cdot P(\text{ЮЗ}|\text{С}) + P(\text{ЮЗ}) \cdot P(\text{С}|\text{ЮЗ}) = \\ = P(C) \cdot P(\text{ЮЗ}|\text{С}) + (P(\text{Ю}) + P(\text{З})) \cdot P(\text{С}|\text{ЮЗ}) \quad (2.3)$$

Если считать вероятность появления всех диалектов равными, тогда

$$P(C) = \frac{1}{3}\ ;\ P(\text{ЮЗ}) = P(\text{Ю}) + P(\text{З}) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3};$$

$$P_{\text{класс2}} = \frac{1}{3}[P(\text{ЮЗ}|\text{С}) + 2 \cdot P(\text{С}|\text{ЮЗ})]\,, \quad (2.4)$$

где Р(ЮЗ|С)- условная вероятность ошибки классификации, когда диалект классифицируется как ЮЗ, хотя на самом деле он относится к группе «С».

В таблице 2.6. указаны экспериментально полученные в соответствии с последним выражением результаты классификации диалекта на две группы при произнесении названия цифры.

Из таблицы видно, что возможны следующие виды классификации.

- По названию цифр (0, 4, 7, 9) можно классифицировать диалекты на группы: Северный+Южный диалекты и Западный диалект.
- По названию цифр (1, 2, 3) можно классифицировать диалекты на группы: Южный+Западный диалекты и Северный диалект.
- По названию цифр (5, 6, 8) можно классифицировать диалекты на группы: Северный +Западный диалекты и Южный диалект.

Таблица 2.6. Оценка вероятности ошибочной классификации на две группы

Название цифр	Оценка вероятности ошибочной классификации, %			
	Разные дикторы	Одинаковые дикторы	Тип классификации	
0	0,003	0,079	СЮ	З
1	0,026	0	ЮЗ	С
2	0,007	0	ЮЗ	С
3	0,013	0,002	ЮЗ	С
4	0,04	0	СЮ	З
5	0,079	0,013	СЗ	Ю
6	0,066	0	СЗ	Ю
7	0,026	0	СЮ	З
8	0	0	СЗ	Ю
9	0	0	СЮ	З

Для осуществления идентификации необходимо контрольные слова подобрать так, чтобы результаты классификации на две группы после произнесения слов, и при этом дополняли друг друга, а относительные частоты ошибок классификации были минимальными.

Если результаты классификации противоречат друг другу (один из этапов классификации проведен с ошибкой), то фиксируется ошибка идентификации. Данный факт обнаруживается, и выдается сообщение о необходимости повторно произнести контрольные названия цифр. Если оба классификатора сработали с ошибкой, то данная ошибка идентификации не обнаруживается.

Рассмотрим работу идентификатора в случае, когда произносятся названия цифр 2 и 5. Соответствующие экспериментальные данные приведены в таблицах 2.7 и 2.8. Результаты тестирования идентификатора для классификации диалектов на две группы по остальным названиям цифр представлены в приложении 5.

Таблица 2.7 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

Таблица 2.8 Результаты эксперимента по классификации диалектов на две группы (произнесение названия цифры 5).

Результат классификации диалекта	Классифицируемые группы диалектов	
	северный+западный	Южный
	Относительная частота классификации диалекта, %	
Группа "северный+западный диалекты"	90	4
Южный диалект	10	96

Проанализируем работу идентификатора при произнесении названий цифр на северном диалекте. Оценка вероятности ошибочного срабатывания идентификаторов, когда ошибка идентификации не обнаруживается

$$P_2 (юз|с) * P_5 (ю|сз) = 0,02*0,1 = 0,002.$$

Здесь P_2 (юз|с) и P_5 (ю|сз) - вероятности ошибочной классификации при произнесении названий цифр 2 и 5 соответственно. Оценка вероятности правильного срабатывания идентификатора без повторного произнесения названий цифр

$$P_2 (с|с) * P_5 (сз|сз) = 0,98*0,9 = 0,882.$$

Оценка вероятности ошибки идентификации, когда эта ошибка обнаруживается из-за противоречий в результатах классификации (возникает необходимость повторного произнесения названий цифр).

$1 - P_2(\text{с}|\text{с}) * P_5(\text{сз}|\text{сз}) - P_2(\text{юз}|\text{с}) * P_5(\text{ю}|\text{сз}) = P_2(\text{юз}|\text{с}) * P_5(\text{сз}|\text{сз}) + P_2(\text{с}|\text{с}) * P_5(\text{ю}|\text{сз}) =$

$= 1 - 0{,}882 - 0{,}002 = 0{,}116.$

Сравнивая работу идентификатора при произнесении названия одной контрольной цифры, с работой идентификатора, использующего два произнесения названий цифр, можно сделать следующий вывод. Произнесение двух названий обеспечивает намного меньшую вероятность ошибки идентификации, но появляется высокая вероятность повторного произнесения названий контрольных цифр. Необходимость повторного произнесения создает дискомфорт использования САРР.

Рассмотрим теперь совместную работу "двоичного" (произнесение названия цифры 2) и "троичного" (произнесение названия цифры 9) классификаторов в случае идентификации южного диалекта. Вероятность правильного срабатывания обоих классификаторов

$$P_{\text{прав.клсф (2,9)}}(\text{ю}) = P_9(\text{ю}|\text{ю})P_2(\text{юз}|\text{ю}). \qquad (2.5)$$

Тогда вероятность ошибки идентификации

$$P_{\text{ош.клсф (2,9)}}(\text{ю}) = 1 - P_9(\text{ю}|\text{ю})P_2(\text{юз}|\text{ю}) = 1 - P_{\text{прав.клсф (2,9)}}(\text{ю}) \qquad (2.6)$$

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

$$P_{\text{ош.клсф.обнаруж (2,9)}}(\text{ю}) = P_2(\text{юз}|\text{ю})P_9(\text{с}|\text{ю}) + P_2(\text{с}|\text{ю})[P_9(\text{ю}|\text{ю}) + P_9(\text{з}|\text{ю})]. \qquad (2.7)$$

Следовательно, вероятность необнаружения ошибки идентификации

$$P_{\text{ош.клсф.необнаруж (2,9)}}(\text{ю}) = P_{\text{ош.клсф (2,9)}}(\text{ю}) - P_{\text{ош.клсф.обнаруж (2,9)}}(\text{ю}) \qquad (2.8)$$

Аналогично получаем выражения для случая идентификации северного диалекта

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

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

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

Рассмотрим теперь случай западного диалекта

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

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

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

Используя данные таблиц 2.9 и 2.7 (результаты классификации диалектов), получаем значения оценок вероятностей появления ошибки, которую нельзя обнаружить

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

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

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 2 и 3.

$$P_{\text{ош.клсф.обнаруж (2,9)}}(\text{ю}) = 0; P_{\text{ош.клсф.обнаруж (2,9)}}(\text{с}) = 0{,}13; P_{\text{ош.клсф.обнаруж (2,9)}}(\text{з}) = 0{,}04.$$

Наибольшая оценка вероятности повторного произнесения контрольных названий цифр равна 0,13 и соответствует северному диалекту. При равной вероятности использования трех диалектов вероятность повторного произнесения заданных названий цифр, и, соответственно, численное значение оценки вероятности

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

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

2.4 Повышение достоверности распознавания при использовании безошибочной идентификации диалектов

Цель этого подраздела заключается в том, чтобы показать, как идентификация диалектов позволяет повысить точность распознавания голосовых команд. Кроме того, в разделе показано, какую точность идентификации диалектов можно получить при построении идентификатора на основе МСМП.

Сначала решалась задача оценки целесообразности автоматической идентификации диалектов для повышения точности распознавания названий цифр. Затем решалась задача собственно идентификации диалектов. На рисунке 2.2 представлена блок-схема алгоритма проведения эксперимента. При обучении

системы использовалось 4500 звукозаписей и столько же звукозаписей использовалось при тестировании системы.

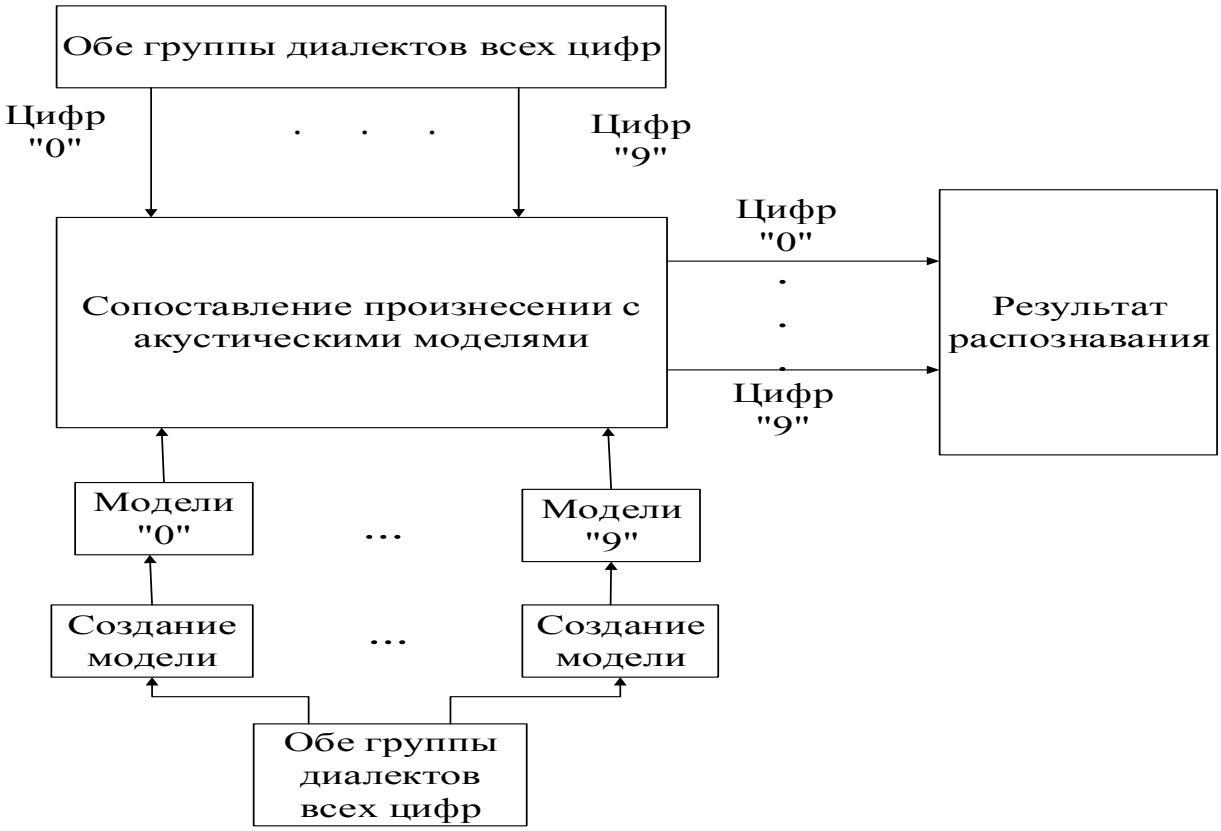


Рисунок 2.2. Алгоритм тестирования системы распознавания в случае отсутствия идентификации

При решении первой задачи сначала для каждого названия цифры создавалась акустическая модель. При создании моделей использовались голоса 18 дикторов носителей трех диалектов арабского языка (по 25 произнесений от каждого диктора). Тестирование системы осуществлялось с использованием голосов тех же дикторов, но произнесения при тестировании отличались от произнесений при обучении (другие 25 произнесений). Результаты тестирования представлены в Таблице 2.9.

Таблица 2.9. Результаты тестирования САРР (акустическая модель каждого названия является общей для всех диалектов)

Название цифры	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0	99,76	0,00	0,00	0,00	2,35	0,47	0,00	10,59	0,00	0,00
1	0,00	99,29	0,47	0,24	2,12	1,41	0,24	0,47	0,94	0,00
2	0,00	0,00	96,71	0,00	0,00	2,12	5,65	0,00	6,12	0,24
3	0,00	0,47	0,00	96,00	0,94	1,18	0,24	0,94	0,00	0,00
4	0,00	0,24	0,00	0,00	92,24	0,00	0,71	0,47	0,24	0,00
5	0,00	0,00	0,00	1,18	0,71	82,59	0,71	4,24	0,47	0,47
6	0,00	0,00	0,24	0,00	0,00	4,71	86,35	2,82	1,41	2,59
7	0,00	0,00	0,24	2,12	0,71	5,65	1,65	75,06	0,00	0,00
8	0,00	0,00	0,24	0,00	0,47	1,41	0,00	0,00	90,82	0,00
9	0,24	0,00	2,12	0,47	0,47	0,47	4,47	5,41	0,00	96,71

Видно, что ошибки присутствуют при распознавании произнесений каждой цифры. Наибольшее количество ошибок соответствует распознаванию названия цифры 7 (100% – 75,06% = 24,94%). Средняя (по диагонали матрицы) оценка точности распознавания системы для трёх диалектов составляет 91,6%.

Затем для каждого диалекта создавалась своя совокупность акустических моделей названий цифр. При создании каждой модели использовались голоса шести дикторов носителей данного диалекта арабского языка (по 25 произнесений от каждого диктора). Тестирование системы осуществлялось с использованием голосов тех же дикторов, но произнесения при тестировании отличались от произнесений при обучении (другие 25 произнесений). На рисунке 2.3 представлена блок-схема алгоритма проведения эксперимента. Результаты тестирования отражены в таблицах: 2.10 – 2.12.

Рисунок 2.3. Алгоритм тестирования системы распознавания при использовании идентификации

Из таблицы 2.10 матрицы распознавания северного диалекта видно, что идентификация диалектов позволяет значительно уменьшить число ошибок распознавания. Самое большое количество ошибок распознавания соответствует распознаванию названия цифры 5 (100% - 92% = 8%), что меньше ранее полученного результата распознавания 24,94%. соответствует распознаванию названия цифры 7. Общая оценка точности распознавания системы для СД составляет 98,1% по диагонали матрицы.

Таблица 2.10. Результаты тестирования системы распознавания произнесений названий цифр (акустическая модель каждого названия создана отдельно для северного диалекта)

Название цифры	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0 ("ноль")	100,00	0,00	0,00	0,00	0,00	0,00	0,00	6,67	0,00	0,00
1("один")	0,00	100,00	2,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2("два")	0,00	0,00	97,3	0,00	0,00	6,00	0,00	0,00	0,00	0,00
3("три")	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00
4("четыре")	0,00	0,00	0,00	0,00	99,33	0,00	0,00	0,00	0,00	0,00
5("пять")	0,00	0,00	0,00	0,00	0,00	92,00	0,00	0,00	0,00	0,00
6("шесть")	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00
7("семь")	0,00	0,00	0,00	0,00	0,67	0,00	0,00	92,67	0,00	0,00
8("восемь")	0,00	0,00	0,00	0,00	0,00	2,00	0,00	0,00	100,00	0,00
9("девять")	0,00	0,00	0,67	0,00	0,00	0,00	0,00	0,67	0,00	100,00

Таблица 2.11. Результаты тестирования системы распознавания произнесений названий цифр (акустическая модель каждого названия создана отдельно для южного диалекта)

Название цифры	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0 ("ноль")	98,67	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,67
1("один")	0,67	100,00	0,00	0,00	2,67	0,00	0,00	0,00	0,00	0,00
2("два")	0,00	0,00	96,67	0,00	0,00	0,00	0,67	0,00	0,00	0,00
3("три")	0,00	0,00	0,00	96,67	0,00	0,00	2,67	0,67	0,00	0,00
4("четыре")	0,00	0,00	0,00	0,00	96,67	0,00	0,00	0,00	0,00	0,00
5("пять")	0,67	0,00	0,00	0,67	0,00	100,00	0,00	0,00	0,00	0,00
6("шесть")	0,00	0,00	0,00	0,00	0,00	0,00	95,33	0,00	0,00	0,67
7("семь")	0,00	0,00	0,00	2,67	0,67	0,00	0,00	99,33	0,00	0,00
8("восемь")	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00
9("девять")	0,00	0,00	3,33	0,00	0,00	0,00	1,33	0,00	0,00	98,67

Из таблицы 2.11 видно, что идентификация диалектов позволяет значительно уменьшить число ошибок распознавания. Самое большое количество ошибок распознавания соответствует распознаванию названия цифры 6. (100% - 95,33% = 4,67%), что меньше ранее полученного результата распознавания 24,94%. (для цифры 7). Общая оценка точности распознавания системы для ЮД составляет 98,2% по диагонали матрицы.

Из таблицы 2.12 матрицы распознавания западного диалекта видно, что идентификация диалектов позволяет значительно уменьшить число ошибок распознавания. Самое большое количество ошибок распознавания соответствует распознаванию названия цифры 7 и 3 (100% - 96% = 4%), что меньше ранее полученного результата распознавания 24,94% (для цифры 7). Общая оценка точности распознавания системы для ЗД составляет 98,3% по диагонали матрицы.

Таблица 2.12. Результаты тестирования системы распознавания произнесений названий цифр (акустическая модель каждого названия создана отдельно для западного диалекта)

Название цифры	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0 ("ноль")	100,00	0,00	0,00	0,00	0,00	0,00	0,00	1,60	0,00	0,00
1("один")	0,00	100,00	0,00	0,80	0,00	0,00	0,00	0,00	0,00	0,00
2("два")	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
3("три")	0,00	0,00	0,00	96,00	0,00	0,00	0,00	0,00	0,00	0,00
4("четыре")	0,00	0,00	0,00	0,80	98,40	0,00	0,00	0,00	0,00	0,00
5("пять")	0,00	0,00	0,00	1,60	0,80	98,40	0,00	0,00	0,80	0,00
6("шесть")	0,00	0,00	0,00	0,00	0,00	0,00	97,60	0,00	0,00	0,00
7("семь")	0,00	0,00	0,00	0,00	0,00	1,60	0,00	96,00	0,80	1,60
8("восемь")	0,00	0,00	0,00	0,80	0,80	0,00	0,00	0,00	98,40	0,00
9("девять")	0,00	0,00	0,00	0,00	0,00	0,00	2,40	2,40	0,00	98,40

Таким образом, идентификация диалектов позволяет значительно снизить относительную частоту ошибки распознавания.

2.5 Вероятность ошибки распознавания голосовых команд при использовании идентификатора диалектов

В данном разделе приведены результаты исследования влияния идентификации диалектов на результаты повышения достоверности системы автоматического распознавания. Высокая степень изменчивости произнесения одних и тех же слов на различных диалектах арабской разговорной речи обусловливает большое количество ошибок при автоматическом распознавании голосовых команд. Данное обстоятельство сдерживает процесс внедрения систем автоматического распознавания речи в телефонии. Поэтому при построении системы автоматического распознавания арабской разговорной речи целесообразно учитывать особенности каждого диалекта.

Идентификация диалекта перед проведением процедуры распознавания позволит использовать акустические модели команд, соответствующие данному диалекту, что повышает точность распознавания. Возникает задача оперативной идентификации диалекта при автоматическом распознавании голосовых команд в телефонии.

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

$$P_{ошГКиднт}(С) = P_{ошГКиднт}(С|С) + P_{ошГКиднт}(Ю|С) + P_{ошГКиднт}(З|С), \qquad (2.9)$$

где $P_{ошГК}(С|С)$ – вероятность ошибки распознавания, когда идентификация диалекта произошла правильно; $P_{ошГК}(Ю|С)$ – вероятность ошибки распознавания, когда идентификация диалекта произошла с ошибкой: вместо

северного определен южный диалект; $Р_{ошГК}(З|С)$ – вероятность ошибки распознавания, когда идентификация диалекта произошла с ошибкой: вместо северного определен западный диалект.

Определим вероятность ошибки распознавания, когда идентификация диалекта произошла правильно

$$Р_{ошГКиднт}(С|С) = Р_{иднт}(С|С) * Р_{ошГК}(С|С), \quad (2.10)$$

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

Определим вероятность

$$Р_{ошГКиднт}(Ю|С) = Р_{иднт}(Ю|С) * Р_{ошГК}(С|Ю), \quad (2.11)$$

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

Определим вероятность

$$Р_{ошГКиднт}(З|С) = Р_{иднт}(З|С) * Р_{ошГК}(С|З), \quad (2.12)$$

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

Следовательно, оценка вероятности точного распознавания при произнесении названия цифры с северным диалектом равна

$$1-Р_{ошГКиднт}(С) \quad (2.13)$$

Рассмотрим случай северного диалекта при использовании одного ключевого слова (цифра 9). Пользуясь ранее полученными данными [3], имеем оценки вероятностей

$$P_{иднт}(С|С) = 0,88 ; P_{иднт}(Ю|С) = 0,08 ; P_{иднт}(З|С) = 0,04.$$

Из последнего эксперимента следует, результаты упомянуты в приложении 4.

$P_{ошГК}(С|С) = 1\text{-}0,98=0,02$; $P_{ошГК}(С|Ю) = 1\text{-}0,434 =0,566$;

$P_{ошГК}(С|З) = 1\text{-}0,344=0,656$. Подставляя числовые данные в выражения (7) – (10), получаем

$$P_{ошГКиднт}(С) = P_{ошГКиднт}(С|С) + P_{ошГКиднт}(Ю|С) + P_{ошГКиднт}(З|С) = 0,88 * 0,02 +$$
$$+0,08 * 0,566 + 0,04 * 0,656 = 0,089.$$

Следовательно, оценка вероятности точного распознавания при произнесении ГК с северным диалектом равна

$$1\text{-}P_{ошГКиднт}(С) = 1\text{-}0,089= 0,911.$$

Случай южного диалекта при использовании ключевого слова (цифра 9). Пользуясь ранее полученными данными [3], имеем оценки вероятностей

$$P_{иднт}(Ю|Ю) = 1 ; P_{иднт}(С|Ю) = 0 ; P_{иднт}(З|Ю) = 0.$$

Из последнего эксперимента следует, результаты упомянуты в приложении 4.

$P_{ошГК}(Ю|Ю) = 1\text{-}0,982=0,018$; $P_{ошГК}(Ю|С) = 1\text{-}0,329 =0,671$;

$P_{ошГК}(Ю|З) = 1\text{-}0,346=0,654$. Подставляя числовые данные в выражения (7) – (10), получаем

$$P_{ошГКиднт}(Ю) = P_{ошГКиднт}(Ю|Ю) + P_{ошГКиднт}(С|Ю) + P_{ошГКиднт}(З|Ю) = 1 * 0,018 + 0 *$$
$$0,671 + 0 * 0,654 = 0,018.$$

Следовательно, оценка вероятности точного распознавания при произнесении ГК с южным диалектом равна

$$1\text{-}P_{ошГКиднт}(Ю) = 1\text{-}0,018= 0,982.$$

Случай западного диалекта при использовании ключевого слова (цифра 9).

Пользуясь ранее полученными данными [3], имеем оценки вероятностей

$$Р_{иднт}(З|З) = 0,96 ; Р_{иднт}(С|З) = 0,04 ; Р_{иднт}(Ю|З) = 0.$$

Из последнего эксперимента следует, результаты упомянуты в приложении 4.

$Р_{ошГК}(З|З) = 1-0,983=0,017$; $Р_{ошГК}(З|С) = 1-0,338 =0,662$;

$Р_{ошГК}(З|Ю) = 1-0,418=0,582$. Подставляя числовые данные в выражения (7) – (10), получаем

$$Р_{ошГКиднт}(З) = Р_{ошГКиднт}(З|З) + Р_{ошГКиднт}(С|З) + Р_{ошГКиднт}(Ю|З) = 0,96 * 0,017 + 0,04 * 0,662 + 0 * 0,582 = 0,0428.$$

Следовательно, оценка вероятности точного распознавания при произнесении ГК с западным диалектом равна

$$1-Р_{ошГКиднт}(З) = 1-0,0428= 0,957.$$

Случай северного диалекта при использовании двух ключевых слов (цифры 2 и 9).

Ранее было получено

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

$$P_{ош.клсф.необнаруж\ (2,9)}(з) = 0$$

Следовательно, оценка вероятности ошибки распознавания при использовании северного диалекта

$\left(1 - P_{ош.клсф.необнаруж\ (2,9)}(с)\right) *$ $Р_{ошГК}(С|С) =(1-0,0024)*(1-0,98)=0,02$.

Так как оценки вероятностей ошибок идентификации западного и южного диалектов равны нулю, то результаты распознавания в данном случае соответствуют работе безошибочного идентификатора.

Рассмотрим влияние идентификации при произнесении названий отдельных цифр. На рисунке 2.4. представлены экспериментальные данные по

достоверности распознавания названий отдельных цифр при отсутствии идентификации и при использовании безошибочной (идеальной) идентификации диалекта.

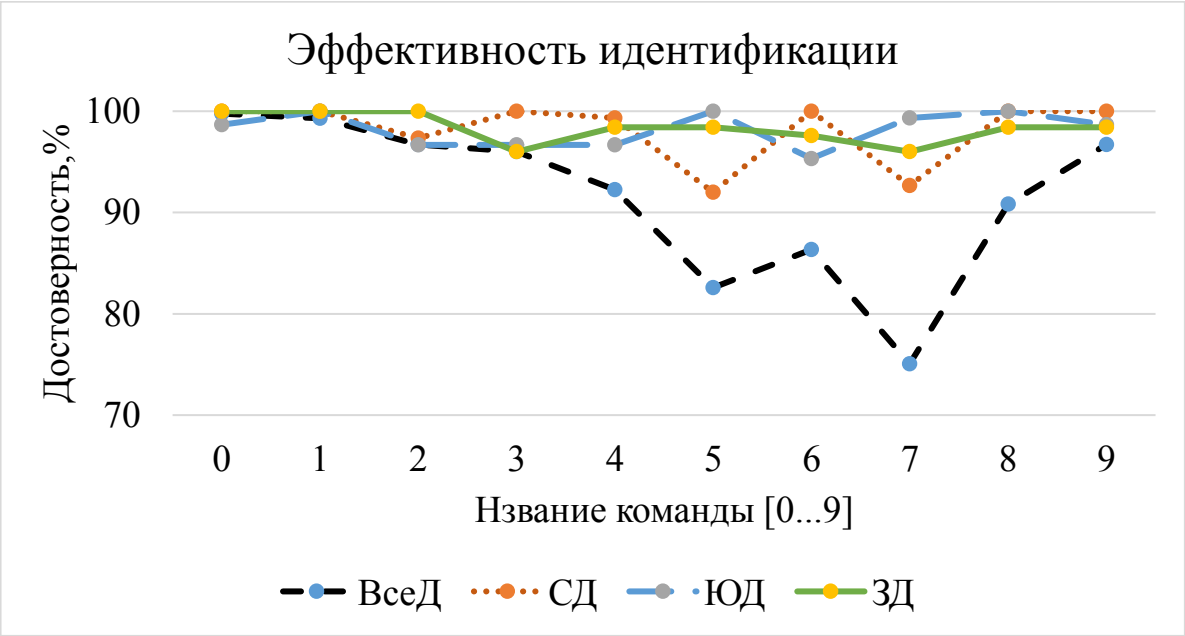


Рисунок 2.4. Достоверность системы распознавания при использовании идентификации

На рисунке обозначены:

- – – Случай, кода идентификация отсутствует;
- – – · Случай, кода используется северный диалект;
- — - Случай, кода используется южный диалект;
- —— Случай, кода используется западный диалект.

2.6 Выводы по разделу 2

1. На основе теоретического анализа и экспериментального исследования показана возможность повышения достоверности распознавания с использованием идентификации диалектов.
2. 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) %.

3. Для оперативной идентификации диалекта при автоматическом распознавании названий цифр целесообразно использовать произнесение одного или нескольких заданных названий цифр. Целесообразно для построения идентификатора диалектов использовать акустические модели таких произнесений названий цифр, которые наиболее сильно отличаются по диалектам
4. В результате анализа причин ошибок идентификации диалекта получены выражения, позволяющие определить вероятность ошибки идентификации.
5. 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%.

ГЛАВА 3 Снижение влияния частотной характеристики канала связи на достоверность распознавания голосовых команд

3.1 Существующие методы нормализации

Нормализация среднего кепстра. Разные микрофоны имеют разные частотные характеристики, и даже один и тот же микрофон имеет разные функции передачи в зависимости от расстояния до микрофона и акустического помещения. В этом подразделе описывается метод, который повышает устойчивость систем распознавания речи к операциям линейной фильтрации [7, 57, 63].

Допустим есть сигнал $x[n]$, и для него вычисляется кепстр в результате чего получается набор из кепстральных векторов $x[x_0, x_1, x_2, x_3, \dots, x_{T-1}]$, их среднее значение $\bar{x}$:

$$\bar{x} = \frac{1}{T}\sum_{T=0}^{T-1} x_t \tag{3.1}$$

Нормализация среднего кепстра (Cepstral Mean Normalization, CMN), состоит из вычитания $\bar{x}$ из каждого вектора x_t

$$\hat{x}_t = x - \bar{x} \tag{3.2}$$

Рассмотрим сигнал, $x[n]$, проходящий через фильтр $h[n]$. На выходе фильтра, получается сигнал $y[n]$, для этого сигнала вычисляются последовательные кепстральные векторы, в результате получится $Y = \{y_0, y_1, \dots, y_{T-1}\}$. В результате дискретного косинусного преобразования (ДКП) определится вектор h:

$$h = C(ln|H(\omega_0)|^2 \dots ln|H(\omega_M)|^2 \dots) \tag{3.3}$$

где C - матрица дискретного косинусного преобразования (Discrete cosine transform, DCT). Тогда на выходе фильтра будет:

$$y_t = x_t + h \tag{3.4}$$

следовательно, среднее значение $\bar{y}_t$:

$$\bar{y} = \frac{1}{T}\sum_{T=0}^{T-1} y_t = \frac{1}{T}\sum_{T=0}^{T-1}(x_t + h) = \bar{x} + h \tag{3.5}$$

и его нормализованный кепстр представляется следующим образом:

$$\hat{y}_t = y_t - \bar{y}_t = \hat{x}_t \tag{3.6}$$

Эта процедура выполняется для каждого высказывания как для обучения, так и для тестирования. Средний вектор $\bar{x}$ передает спектральные характеристики текущего микрофона и акустику помещения. В пределе, когда T $\rightarrow \infty$ для каждого высказывания, следует ожидать, что средние значения для высказываний из одной и той же среды записи будут одинаковыми. Использование нормализации кепстр по среднему значению – Cepstral Mean Normalization CMN для векторов кепстра не изменяет дельта или дельта-дельта кепстр [63].

Проанализируем влияние CMN на короткое высказывание. Предположено, что высказывание содержит одну фонему, скажем / s /. Среднее значение $\bar{x}$ будет очень похоже на кадры в этой фонеме, поскольку / s / довольно стационарно. Таким образом, после нормализации $\hat{x}_t \approx 0$. Аналогичный результат будет иметь место для других фрикативных звуков, что означает, что было бы невозможно различить эти ультракороткие высказывания, и частота ошибок будет очень высокой.

Опытным путем было обнаружено, что нормализация не ухудшает точность распознавания высказываний из одной и той же акустической среды, если они длятся дольше, чем 2–4 секунды. Показано, что использование CMN обеспечивает относительное снижение частоты ошибок на 30% [63].

Следует учесть, что даже для одного и того же микрофона, и акустики помещения расстояние между ртом и микрофоном варьируется для разных дикторов, что вызывает несколько разные передаточные функции. Кроме того, кепстральное среднее характеризует не только передаточную функцию канала, но и среднюю частотную характеристику разных дикторов. Удаляя долгосрочное среднее значение диктора, CMN может выступать в качестве своего рода нормализации диктора [63].

Одним из недостатков CMN является то, что он не различает тишину и голос при вычислении среднего значения высказывания. Дополнение CMN состоит в вычислении различных параметров для шума и речи. То есть разность между вектором среднего значения для речевых кадров в произнесении и вектором среднего значения m_c для речевых кадров в обучающих данных и аналогично для шумовых кадров m_n. Различение речи / шума может быть выполнено путем классификации кадров на речевые кадры и шумовые (паузные) кадры, вычисления среднего кепстра для каждого и вычитания их из среднего значения в обучающих данных. Эта процедура работает хорошо, если классификация речь / пауза точна.

Нормализация кепстра по среднему значению (Cepstral Mean Subtraction - CMN) является широко используемым методом, компенсирующим изменчивость речевого сигнала в кепстральной области. Оно используется в качестве стандартного метода нормализации признаков для большинства систем (Automatic System Recognition, ASR). Основное внимание CMN уделяется

сверточным искажениям, вызванным характеристиками различных каналов связи или записывающих устройств.

Тем не менее, CMN также частично эффективен в снижении эффектов аддитивного шума окружающей среды и изменчивости стиля разговора. С этой точки зрения CMN может быть полезен для нормализации изменчивости, которая появляется в несоответствующих сценариях обучение/ тестирование с нормальной речью и шепотом.

Нормализация по средней дисперсии кепстра (CVN). CVN часто используется в сочетании с CMN, где он способствует устойчивости, масштабируя и ограничивая диапазон отклонений в кепстральных признаках. Этот метод известен как нормализация среднего и дисперсии, и он применяется с помощью следующего уравнения [35, 56]:

$$C_{n,t}^{CVN} = \frac{C_{n,t}^{CMN}}{\sigma_n} \tag{3.7}$$

где $C_{n,t}^{CMN}$ и σ_n Кепстральные коэффициенты после CMN и стандартное отклонение (дисперсия), соответственно. Степень изменчивости PC, уменьшается с использованием нормализация среднего кепстра и дисперсии (Cepstral Mean and Variance Normalization, CMVN) [22, 97, 98].

Нормализация усиления кепстра (CGN). включает CMN и вместо дисперсии, оценивает динамический диапазон выборок в каждом кепстральном измерении непосредственно из максимальных ($C_{n\,max}$) и минимальных ($C_{n\,min}$) значений выборок. Он рассчитывается по следующему уравнению [56]:

$$\mathrm{C}_{n,t}^{CGN} = \frac{c_{n,t}^{CMN}}{(c_{n\,max} - c_{n\,min})} \tag{3.8}$$

Нормализация кепстра в реальном времени (RASTA). CMN требует полного высказывания для вычисления среднего кепстра; таким образом, оно не может

использоваться в системе реального времени, и необходимо использовать приближение.

Можно интерпретировать результаты работы как фильтра верхних частот[40, 63, 66]. с передаточной функцией часто используют:

$$H(z) = 0.1z^4 * \frac{2 + z^{-1} - z^{-3} - 2z^{-4}}{1 - 0.98z^{-1}} \quad (3.9)$$

Низкочастотная фильтрация помогает сгладить некоторые из быстрых спектральных изменений от кадра к кадру. Опытным путем было показано, что фильтр RASTA ведет себя аналогично реализации CMN в реальном времени. Реализация фильтра RASTA, так и CMN в реальном времени. требует, чтобы фильтр был правильно инициализирован. В противном случае первое высказывание может использовать неправильное среднее значение кепстра.[56, 63, 79].

CMN может привести к потере различающей речевой информации, особенно для коротких высказываний. Для этого предложена модификация CMN, чтобы уменьшить эту потерю, преобразовывая каждое тестовое высказывание с шумом в оценку чистого высказывания (средняя оценка данного высказывания, если шум отсутствует) [19, 20, 21, 22, 63].

3.2 Анализ факторов, влияющих на результат нормализации параметров речевого сигнала по среднему значению

Факторы, влияющие на результат нормализации параметров РС по среднему значению [7, 8, 11]***:***

- Неравномерность амплитудно-частотных характеристик звуковых трактов, используемых при обучении системы распознавания речи и при ее тестировании;
- Уровень боковых лепестков оконной функций;

- Отсутствие синхронизации процедур сегментации сигнала на входе и выходе канала связи.

Рассмотрим основные этапы обработки сигнала при формировании МЧКК, которые отображены на схеме, представленной Рисунком 3.1

При определении MFCC используется БПФ, результаты которого, в значительной степени, зависят от вида применяемой оконной функции. Можно предположить, что вид оконной функции влияет на результат определения MFCC. В данной диссертации исследуется влияние вида оконной функции на степень стабилизации значений нормализованных параметров РС при изменении ЧХ канала связи [63]

Звуковые колебания посредством микрофона преобразуются в РС, затем после низкочастотной фильтрации и аналого-цифрового преобразования (АЦП) проводится сегментация РС, каждый сегмент РС взвешивается оконной функцией и осуществляется БПФ – формируется кратковременный спектр сигнала. Для учета особенностей человеческого слуха частотная шкала преобразуется в мел-шкалу согласно выражению $\mathrm{m} = 1125 \ln\left(1 + \frac{\mathrm{f}}{700}\right)$. Далее

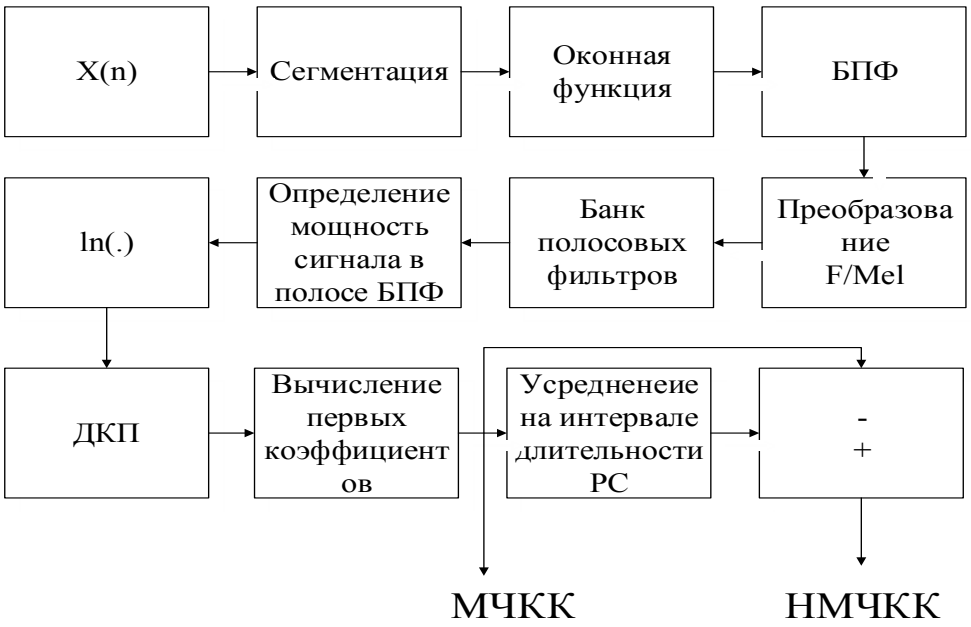


Рисунок 3.1 Схема обработки речевого сигнала при формировании МЧКК

мел-частотный спектр каждого сегмента равномерно разбивается на отдельные полосы набором полосовых фильтров, и определяется мощность сигнала на выходе каждого фильтра. Полученный набор значений мощностей P_{sf} логарифмируется. Затем к результату логарифмирования каждого сегмента применяется ДКП – формируется кепстр РС [63, 81].

Несколько первых коэффициентов ДКП оставляются, остальные коэффициенты удаляются. По полученной временной последовательности наборов (векторов) MFCC определяется среднее значение вектора во времени. Среднее значение вычитается из каждого вектора – последовательность векторов нормализуется по среднему значению. Так как процедура ДКП линейна, то проанализируем нормализацию до проведения ДКП.

Рассмотрим этап создания выборки звуковых сигналов, используемых для обучения системы распознавания – формирования акустических моделей звуков. Пусть $Y_L(p, f)$ – значение мощности сигнала на выходе одного из полосовых фильтров, настроенного на частоту f, для p-го сегмента речевого сигнала.

$$Y_L(p, f) = X_L(p, f)|H_L(f)|^2, \tag{3.10}$$

где $X_L(p, f)$ – значение мощности сигнала на выходе одного из полосовых фильтров, настроенного на частоту f, для p-го сегмента речевого сигнала в идеальном случае, когда амплитудно-частотная характеристика (АЧХ) канала связи – микрофона равномерна в полосе частот РС; $H_L(f)$ – ЧХ канала связи, в данном случае, микрофона, практически используемого при создании обучающей выборки речевых сигналов. После логарифмирования имеем

$$\log\big(Y_L(p, f)\big) = \log(X_L(p, f)) + \log(|H_L(f)|^2), \tag{3.11}$$

После процедуры усреднения результата логарифмирования по всем сегментам (во времени) получаем:

$$\overline{\log\big(Y_L(p, f)\big)} = \overline{\log(X_L(p, f))} + \log(|H_L(f)|^2). \tag{3.12}$$

В результате нормализации величины $\log\big(Y_L(p, f)\big)$ по среднему значению получаем нормализованное значение мощности сигнала на выходе полосового фильтра:

$$Y_{LN}(p,f) = \log\left(Y_L(p,f)\right) - \overline{\log\left(Y_L(p,f)\right)} = \log(X_L(p,f)) - \overline{\log(X_L(p,f))}. \quad (3.13)$$

Видно, что влияние АЧХ микрофона на нормализованное значение мощности сигнала на выходе полосового фильтра устраняется. При обращении пользователя к системе (случай распознавания речи) имеем

$$Y_T(p,f) = X_T(p,f)|H_T(f)|^2, \quad (3.14)$$

где $Y_T(p,f)$ – значение мощности сигнала на выходе одного из полосовых фильтров, настроенного на частоту f, для p-го сегмента речевого сигнала. Повторяя рассуждения, приведенные выше, получаем нормализованные значения мощности сигнала на выходе полосового фильтра.

$$Y_{TN}(p,f) = \log\left(Y_T(p,f)\right) - \log\overline{\left(Y_T(p,f)\right)} = \log(X_T(p,f)) - \log\overline{\left(X_T(p,f)\right)} \quad (3.18)$$

Однако приведенные выше рассуждения не учитывают того факта, что при проведении кратковременного БПФ производится свертка спектра сигнала с ЧХ используемой оконной функции. В результате на результаты спектрального анализа накладывается влияние боковых лепестков ЧХ оконной функции. В частности, если один из компонентов спектра намного больше других компонентов, то наличие боковых лепестков оконной функции может привести к сильному искажению результатов преобразования Фурье для более слабых спектральных компонентов.

Рассмотрим случай, когда спектр сегмента состоит из двух спектральных компонентов. Пусть меньший по уровню компонент находится на уровне бокового лепестка преобразования Фурье сильного компонента. Каждый из компонентов состоит из двух составляющих: постоянной и переменной во

времени. Боковой лепесток также содержит переменную и постоянную во времени составляющие. Постоянная составляющая бокового лепестка суммируется с постоянной составляющей слабого спектрального компонента.

Рассмотрим сначала этап обучения системы. Выделим среди коэффициентов БПФ каждого сегмента речевого сигнала два спектральных компонента: $S_1(p, f_1)$ и $S_2(p, f_2)$, где $S_1(p, f_1)$, $S_1(p, f_2)$ – модули компонентов на частотах f_1 и f_2. Считаем, что фаза компонента на частоте f_1 равна фазе коэффициента БПФ, относящегося к спектральному компоненту на частоте f_2. Компонент $S_1(p, f_1)$ с учетом влияния боковых лепестков оконной функции

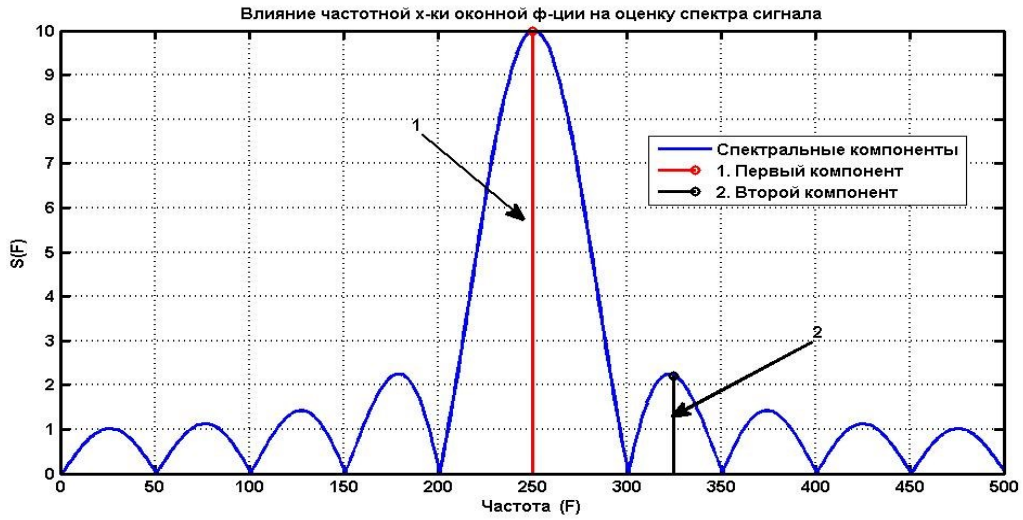


Рисунок 3.2 Влияние вида оконной функции на оценку спектра

$$S_1(p, f_1) = u(p, f_1) \cdot m_1(f_1) + k_{f2}(f_1) \cdot u(p, f_2) \cdot m_1(f_2), \qquad (3.15)$$

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; $k_{f2}(f_1)$ – коэффициент ослабления боковых лепестков оконной функции на

частоте f_1 при применении БПФ к спектральному компоненту на частоте f_2. Преобразуем выражение (3.19) к виду:

$$S_1(p, f_1) = u(p, f_1) \cdot m_1(f_1) \left[1 + \frac{k_{f2}(f_1) u\,(p, f_2) \cdot m_1(f_2)}{u(p, f_1) \cdot m_1(f_1)}\right] \qquad (3.16)$$

Учтем, что второе слагаемое в квадратных скобках намного меньше единицы (это следует из малости боковых лепестков применяемых при БПФ оконных функций). Применяя логарифмирование к выражению (3.20), получаем

$$\log[S_1(p, f_1)] = \log[u(p, f_1)] + \log[m_1(f_1)] + \log\left[\left(1 + \frac{k_{f2}(f_1) u\,(p,f_2) \cdot m_1(f_2)}{u(p,f_1) \cdot m_1(f_1)}\right)\right] \approx \log[u(p, f_1)] + \log[m_1(f_1)] + \frac{k_{f2}(f_1) u\,(p,f_2) \cdot m_1(f_2)}{u(p,f_1) \cdot m_1(f_1)}, \qquad (3.17)$$

Применяя к выражению (3.21) процедуру усреднения по всем сегментам получаем:

$$\overline{\log[S_1(p, f_1)]} \approx \overline{\log[u(p, f_1)]} + \log[m_1(f_1)] + k_{f2}(f_1) \frac{m_1(f_2)}{m_1(f_1)} \cdot \overline{\left(\frac{u(p,f_2)}{u(p,f_1)}\right)}, \qquad (3.18)$$

Вычитая выражение (3.22) из выражения (3.21), получаем результат нормализации параметров сигнала по среднему

$$Y_{LN1}(p, f_1) = \log[S_1(p, f_1)] - \overline{\log[S_1(p, f_1)]} \approx \log[u(p, f_1)] - \overline{\log[u(p, f_1)]} + k_{f2}(f_1) \frac{m_1(f_2)}{m_1(f_1)} \cdot \left[\frac{u(p,f_2)}{u(p,f_1)} - \overline{\left(\frac{u(p,f_2)}{u(p,f_1)}\right)}\right], \qquad (3.19)$$

Аналогично рассуждая для случая распознавания речи, получаем

$$Y_{TN1}(p,f_1) = \log[S_1(f_1)] - \overline{\log[S_1(f_1)]} \approx$$
$$\log[u(p,f_1)] - \overline{\log[u(p,f_1)]} + k_{f2}(f_1)\frac{m_2(f_2)}{m_2(f_1)} \cdot \left[\frac{u(p,f_2)}{u(p,f_1)} - \overline{\left(\frac{u(p,f_2)}{u(p,f_1)}\right)}\right], \quad (3.20)$$

Разность нормированных логарифмов спектров (3.23) и (3.24) для случаев обучения и тестирования

$$delta(p,f_1) = Y_{LN1}(p,f_1) - Y_{TN1}(p,f_1) \approx$$
$$k_{f2}(f_1) \cdot \left[\frac{u(p,f_2)}{u(p,f_1)} - \overline{\left(\frac{u(p,f_2)}{u(p,f_1)}\right)}\right] \cdot \left[\frac{m_1(f_2)}{m_1(f_1)} - \frac{m_2(f_2)}{m_2(f_1)}\right], \quad (3.21)$$

Проанализируем полученное выражение (3.21). Здесь отношения: $\frac{m_1(f_2)}{m_1(f_1)}$ и $\frac{m_2(f_2)}{m_2(f_1)}$ – характеризуют неравномерности АЧХ 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 САР. 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 $\overline{u(p,f_2)}$ разность нормализованных коэффициентов становится большой.

3.3 Экспериментальное исследование факторов, влияющих на нормализацию паромеров речевого сигнала

В работах различных авторов [1, 20, 30, 63] отмечается, что наличие аддитивной помехи в речевом сигнале, а также ограниченность интервала времени усреднения коэффициентов снижают эффект нормализации. Нормализованные параметры одного и того же звукового речевого сигнала на

выходах различных звуковых трактов отличаются друг от друга. В данном разделе исследуется возможность уменьшения указанных отличий.

Основные этапы обработки сигнала при формировании нормализованных МЧКК.

- Каждый сегмент РС взвешивается оконной функцией.
- Взвешенные сегменты подвергаются быстрому преобразованию Фурье (БПФ) – формируется кратковременный спектр сигнала.
- Для учета особенностей человеческого слуха частотная шкала преобразуется в мел-шкалу [63]. Мел-частотный спектр каждого сегмента равномерно разбивается на отдельные полосы набором полосовых фильтров.
- Определяется мощность сигнала на выходе каждого фильтра. Полученный набор значений мощностей сигналов логарифмируется.
- К результату логарифмирования каждого сегмента применяется дискретное косинусное преобразование (ДКП) – формируется кепстр РС.
- Несколько первых коэффициентов ДКП оставляются, остальные коэффициенты удаляются.
- По полученной временной последовательности наборов (векторов) MFCC определяется среднее значение вектора во времени. Среднее значение вычитается из каждого вектора – последовательность векторов нормализуется по среднему значению.
- Так как процедура ДКП линейна, то проанализируем нормализацию до проведения ДКП. Рассмотрим этап создания выборки звуковых сигналов, используемых для обучения системы распознавания – формирования акустических моделей звуков. При проведении кратковременного преобразования Фурье производится свертка спектра сигнала с частотной

характеристикой используемой оконной функции. В результате на результаты спектрального анализа накладывается влияние боковых лепестков частотной характеристики оконной функции. В частности, если один из компонентов спектра намного больше других компонентов, то наличие боковых лепестков оконной функции может привести к сильному искажению результатов преобразования Фурье для более слабых спектральных компонентов.

С целью проверки полученных выше соотношений проведено имитационное моделирование основных процессов формирования МЧКК на этапах создания обучающей и тестирующей выборок данных звука.

Тестовый сигнал пропускается через фильтр, логарифм частотной характеристики которой равен разности логарифмов амплитудно-частотных характеристик (АЧХ) микрофонов, используемых для формирования обучающих и тестовых выборок звуковых файлов [16, 17].

В данном эксперименте использовался нерекурсивный фильтр восьмого порядка. АЧХ и фазо-частотная характеристика (ФЧХ) фильтра представлены на рисунке 3.3. Из анализа ФЧХ следует, что при прохождении сигнала через фильтр он задерживается примерно на четыре периода дискретизации.

Тестовый сигнал, является суммой амплитудно-модулированного гармонического колебания и белого гауссова шума. Отношение сигнал-шум равно 10 дБ. Параметры колебания: частота F=2 кГц, частота модуляции F=1.25 Гц; коэффициент модуляции m=0.3. Частота дискретизации сигнала fs = 8 кГц, длительность колебания T = 12800/fs. Размер сегмента сигнала – 256 отсчетов, величина перекрытия сегментов – 128 отсчетов, число точек быстрого преобразования Фурье- 256.

В качестве меры близости нормализованных параметров сигнала на входе и выходе фильтра рассматривается величина среднеквадратического значения их разности на каждой частоте (для каждого индекса коэффициента БПФ).

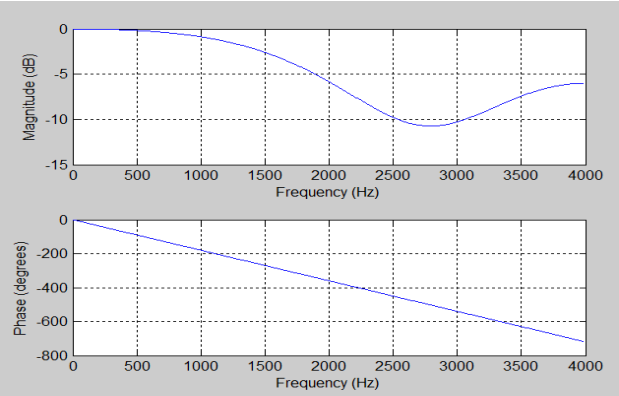


Рисунок 3.3. АЧХ и ФЧХ фильтра

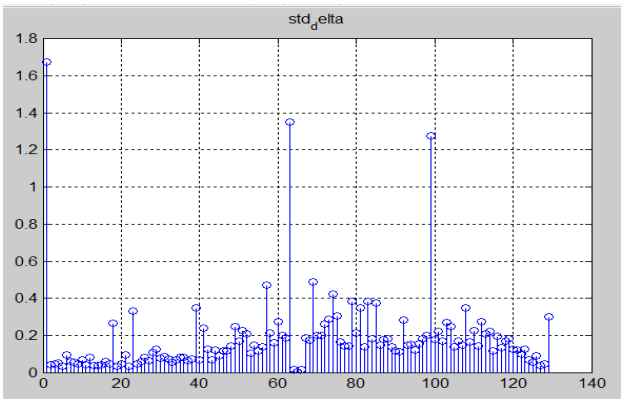


Рисунок 3.4. Зависимость СКЗ от индекса коэффициента БПФ

На рисунке (3.4) представлена зависимость СКЗ от индекса коэффициента БПФ при использовании оконной функции Хэмминга при БПФ. Видно, что для индексов 65, 66 соответствующих частоте сигнала 2000 Гц нормализованные параметры на входе и выходе фильтра отличаются мало. В качестве обобщенной меры близости нормализованных параметров сигнала используется среднее по всем индексам значение СКЗ. Для окна Хэмминга с уровнем подавления боковых лепестков частотной характеристики равным 43 дБ она равна 0,18 дБ. При использовании окна Чебышева с относительным уровнем подавления боковых лепестков более 80 дБ обобщенная мера близости нормализованных параметров сигнала становится меньше, она равна 0,16 дБ.

Проведены две серии экспериментов. В первой серии использовался специально сформированный тестовый сигнал. А во второй серии – РС. В качестве РС использовалась звукозапись последовательности арабских названий цифр от "пять" до "девять".

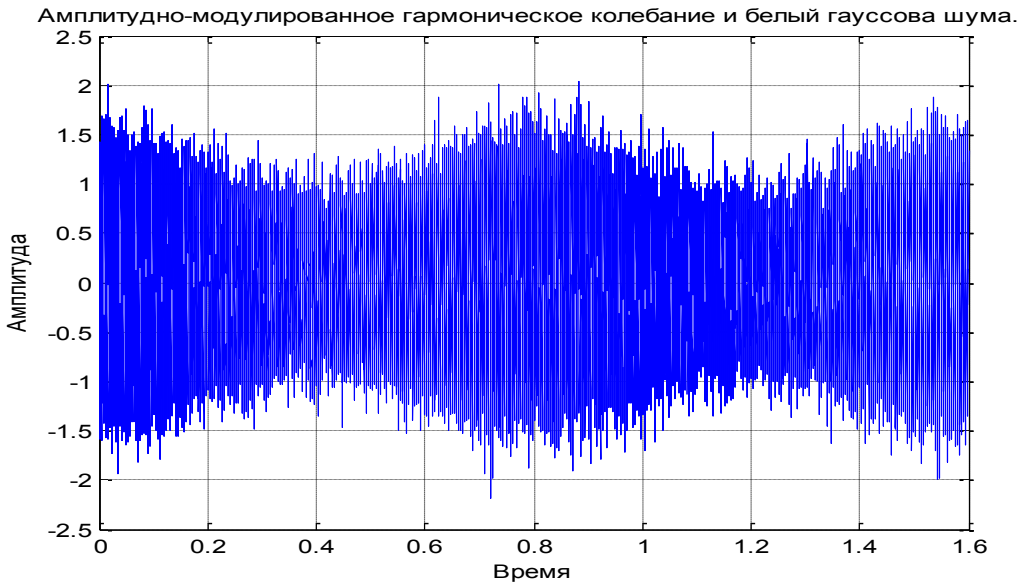


Рисунок 3.5 Сигнал амплитудно-модулированного колебания

Амплитудно-модулированное колебание можно рассматривать как узкополосный РС, спектр которого меняется во времени. Такой выбор тестового сигнала позволяет оценить влияние параметров оконной функции как на вокализованные сегменты сигнала с узким спектром, так и на широкополосные сегменты, к которым, в основном, относятся невокализованные сегменты. При большом отношении сигнал-шум имеет место случай узкополосных сегментов, а при малом – широкополосных.

Тестовый сигнал сначала сегментируется, и для каждого сегмента проводится БПФ. Затем абсолютные значения коэффициентов БПФ возводятся в квадрат и логарифмируются – формируются параметры сигнала для данного эксперимента. Для полученных параметров сигнала проводится процедура их усреднения по всем сегментам (параметры сигнала усредняются во времени). Полученное среднее вычитается из всех параметров сигнала – формируются нормализованные по среднему параметры сигнала. Тестовый сигнал

пропускается также через фильтр. Сигнал на выходе фильтра подвергается обработке, которая полностью повторяет обработку сигнала, описанную выше.

Алгоритмы оценки влияния нормализации на параметры речевого сигнала

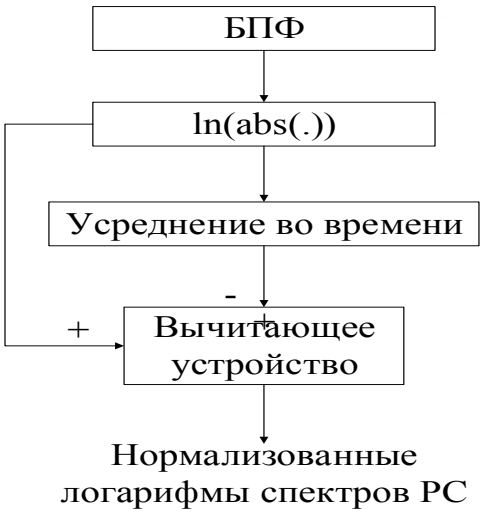


Рисунок 3.6. Обработка сигналов на входе и выходе фильтр-имитатора

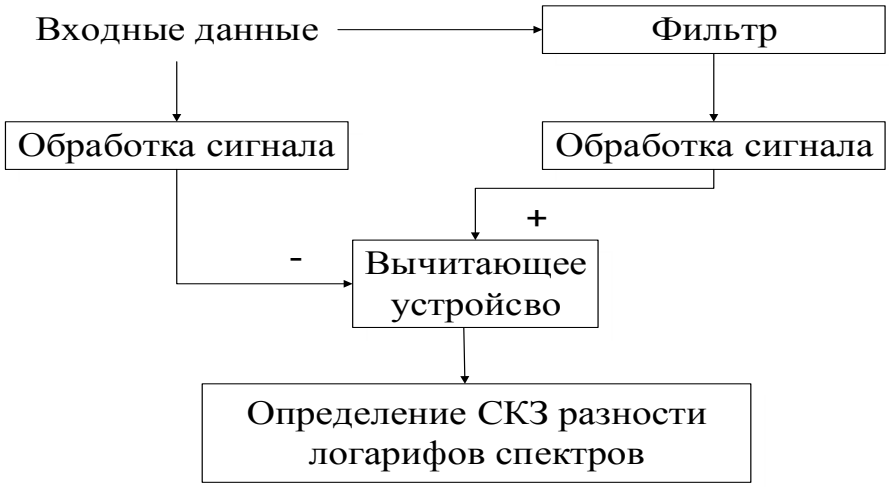


Рисунок 3.7. Алгоритм проведения эксперимента нормализации (LN(p,f) – параметры сигнала)

Нормализованные параметры сигнала на выходе фильтра сравниваются с аналогичными параметрами на входе фильтра – определяется набор разностей

между параметрами. В качестве меры близости нормализованных параметров сигнала на входе и выходе фильтра рассматривается средний по всем сегментам квадрат их разности для каждого значения частоты (для каждого индекса коэффициента БПФ). Для обобщенной оценки близости параметров сигнала на входе и выходе фильтра используется метрика – средний квадрат их разности по всем значениям частоты и по всем сегментам. При использовании речевого сигнала в качестве его параметров рассматривается набор логарифмов спектральных значений мощности, а также набор из 12 MFCC. Результаты экспериментов представлены в Таблице 3.3.

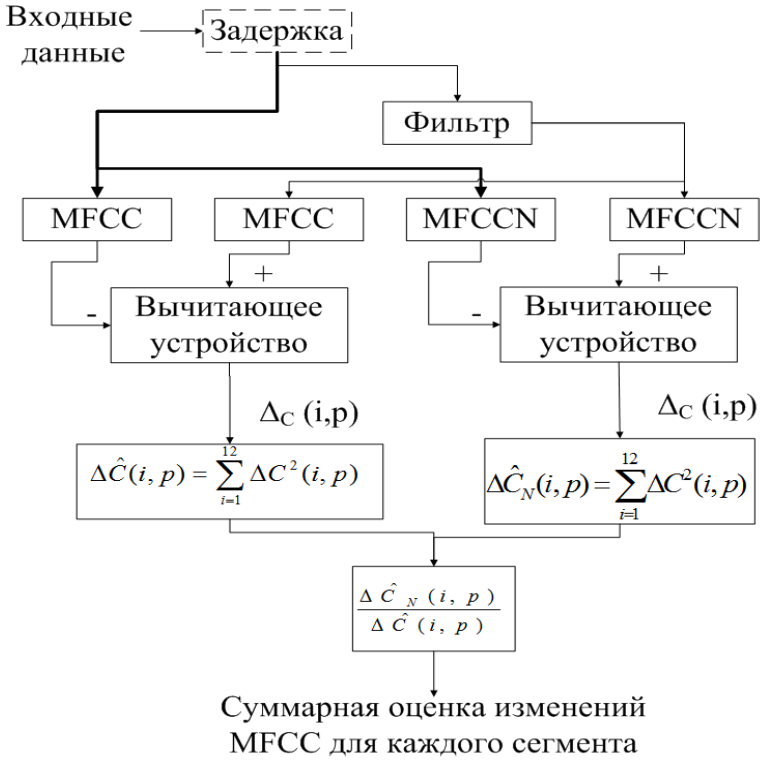


Рисунок 3.8 Алгоритм проведения эксперимента нормализации (MFCC – параметры сигнала)

В таблице (3.3) в первом слева столбце указан вид используемого сигнала: Ш – белый шум, АМ+Ш – сумма амплитудно-модулированного сигнала и белого

шума (при обработке указанных сигналов используется оконная функция Хэмминга), $РС_Х$ – речевой сигнал при использовании оконной функции Хэмминга с уровнем боковых лепестков -43дБ, $РС_ч$. – речевой сигнал при использовании оконной функции Чебышева с меньшим уровнем боковых лепестков равным -80дБ.

Таблица 3.3. Результаты экспериментов

Вид сигнала	**Метрика отличия параметров сигнала (для логарифма спектра мощности)**					
	Фильтр 1			**Фильтр 2**		
	Без нормализации	Нормализация без учета задержки сигнала	Нормализация с учетом задержки сигнала	Без нормализации	Нормализация без учета задержки сигнала	Нормализация с учетом задержки сигнала
Ш	40,2	1,0	0,079	84,3	1,34	0,082
АМ+Ш	40,4	1,28	0.099	84,8	1,68	0,13
РСх	45,5	10,78	3,49	85,6	6,56	2,24
РСч	41,1	2,01	0,28	84,8	2,04	0,12
Метрика отличия параметров сигнала (для MFCC)						
$РС_Х$	0,36	0,013	0,0095	0,46	0,0074	0,0026
$РС_ч$	0,39	0,0017	$1{,}51*10^{-4}$	0,48	0,0025	$6{,}28*10^{-5}$

Во втором столбце указаны значения метрики, характеризующей степень отличий параметров сигналов на входе и выходе фильтра при отсутствии нормализации. В третьем столбце – значения метрики для нормализованных параметров сигнала.

В четвертом столбце – значения метрики для нормализованных параметров сигнала при задержке входного сигнала фильтра на величину группового запаздывания фильтра. Задержка входного сигнала позволяет обеспечить синхронизацию процедур сегментации для входного и выходного сигналов фильтра. Последующие столбцы соответствуют случаю использования второго фильтра.

3.4 Зависимость результатов нормализации параметров речевого сигнала от вида используемой оконной функции

Наиболее часто в качестве параметров речевого сигнала, используемых при распознавании речи, используются мел-частотные кепстральные коэффициенты MFCC [24, 63]. При определении указанных коэффициентов применяется БПФ с последующим преобразованием частотной шкалы в мел-шкалу для учета особенностей слуха человека. Преобразованные коэффициенты БПФ пропускаются через банк полосовых фильтров, логарифмируются, и к ним применяется дискретное косинусное преобразование – формируются мел-частотные кепстральные коэффициенты. В данной работе экспериментально исследовано влияние формы оконной функции, используемой при БПФ, на указанное отличие. Реализован следующий порядок проведения эксперимента.

Звукозаписи произнесений арабских названий цифр пропускались через фильтр, частотная характеристика которого имитирует различие частотных характеристик каналов связи, используемых при обучении и эксплуатации САРР. К сигналам на входе и выходе фильтра применялось БПФ – формировался кратковременный спектр сигнала для каждого сегмента сигнала. Определялись логарифмы абсолютных значений коэффициентов БПФ. Полученные логарифмы выходного сигнала фильтра содержат сумму логарифмов коэффициентов БПФ входного сигнала фильтра и логарифма амплитудно-частотной характеристики фильтра.

- Далее логарифмы коэффициентов БПФ усреднялись во времени, и полученное среднее вычиталось из исходных значений логарифмов – формировались нормализованные значения логарифмов для коэффициентов БПФ. При этом влияние частотной характеристики фильтра на значения логарифмов подавляется. Определялась разность нормализованных значений логарифмов БПФ для входного и выходного

сигналов фильтра. Далее вычислялось среднеквадратическое значение разности (СКЗР) для каждого значения частоты по всем сегментам сигнала. Затем определялось среднее значение СКЗР по всем частотам (СКЗРЧ). Полученное значение СКЗРЧ рассматривается как мера отличий нормализованных параметров речевого сигнала на входе и выходе фильтра. С целью упрощения физической интерпретации результатов анализа в качестве параметров сигнала рассматривается результат логарифмирования набора абсолютных значений коэффициентов БПФ каждого сегмента РС.

Исследование проведено в среде системы Matlab. На основе визуализатора оконных функций wvtool (Window Visualization Tool) разработана программа для исследования влияния формы окна на значения СКЗР и СКЗРЧ [28, 12].

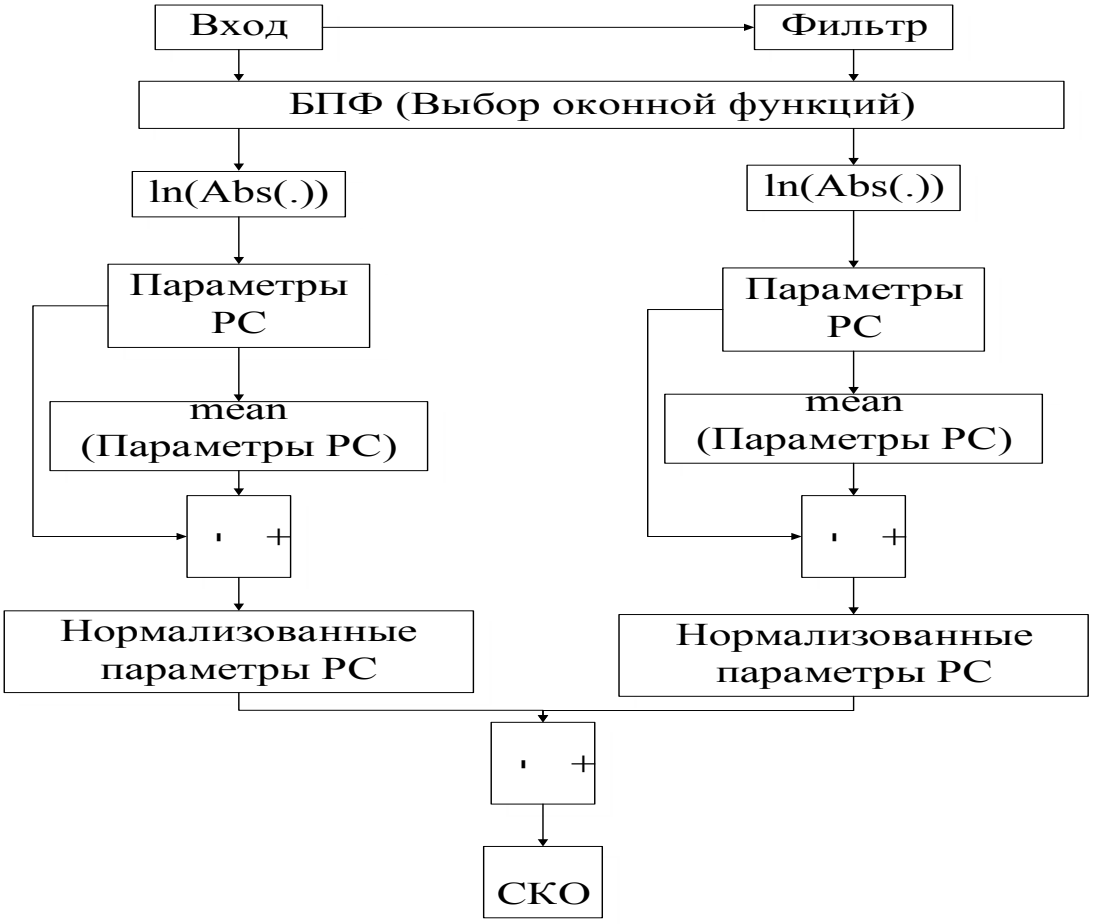


Рисунок 3.9. Алгоритм оценки влияния разных оконных функций на результаты нормализации параметров речевого сигнала

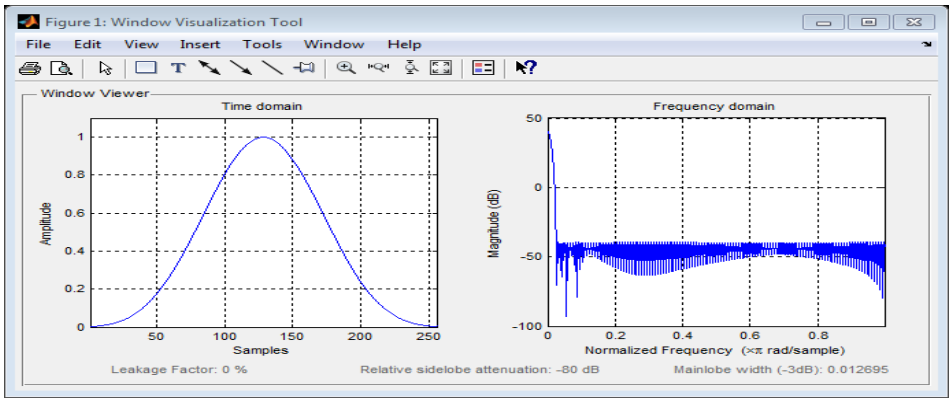


Рисунок 3.10 – Окно Чебышева для уровня боковых лепестков -80 дБ

С помощью созданной программы исследована возможность использования различных оконных функций при вычислении нормализованных

логарифмических значений модулей коэффициентов БПФ. В результате проведения эксперимента определены значения СКЗРЧ для звукозаписей арабских произнесений названий цифр. Параметры звукозаписей: частота дискретизации 16 кГц, число разрядов квантователя 16. Параметры БПФ: размер сегмента речевого сигнала 256 отсчетов, размер перекрытия сегментов 128, число коэффициентов БПФ 256. Для имитации различий частотных характеристик каналов связи, используемых при обучении и эксплуатации САР, применялся нерекурсивный фильтр восьмого порядка, АЧХ которого представлена на рисунке 3.11 Фильтр соответствует разбросу значений АЧХ электретных микрофонов, применяемых в телефонии [16].

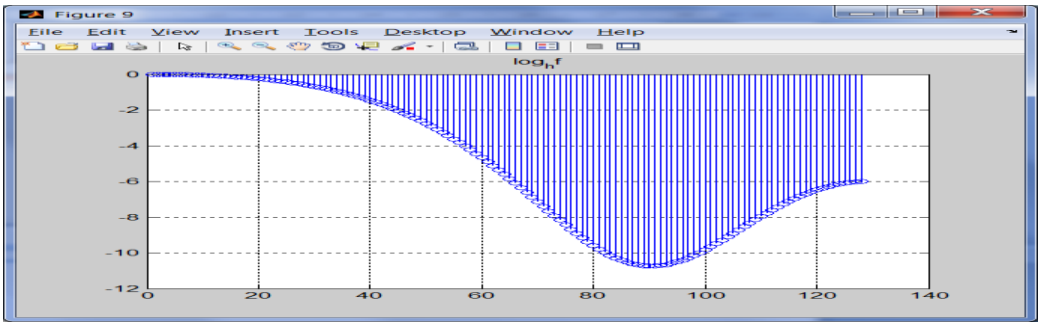


Рисунок 3.11. АЧХ фильтра – имитатора различий частотных характеристик каналов связи

На рисунке максимальное значение индекса коэффициента БПФ соответствует 4 кГц. По оси ординат отложены значения АЧХ в дБ. В таблице (3.4) приведены результаты проведенного эксперимента.

Из данных таблицы (3.4) следует, что малым значениям СКЗРЧ соответствуют оконные функции: Ханна, Блэкмена, Бомена, Тьюки при α=0.5, Парзена. Большим значениям СКЗРЧ соответствуют оконные функции: Чебышева при β=40дБ, Прямоугольная, Кайзера при β=4, Гауссова при α=2.5.

Параметры рассмотренных оконных функций показаны в таблице (3.5) [25, 24, 28].

Таблица 3.4. Результаты эксперимента

Значения СКЗРЧ для случая произнесения арабских названий цифр												
№ п/п	**Оконная функция – "окно"**	**Цифры**										
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**0-9**
1	**Прямоугольное**	3,8	2,48	2,51	2,51	2,18	3,33	3,48	3,62	2,52	2,92	2,85
2	**Треугольное**	0,38	0,88	0,43	0,68	0,81	0,61	0,28	0,47	0,54	0,33	0,45
3	**Бартлетта**	0,51	1,21	0,76	0,91	1,29	0,97	0,41	0,69	0,91	0,43	0,62
4	**Ханна**	0,15	0,13	0,12	0,13	0,12	0,15	0,12	0,12	0,14	0,13	0,16
5	**Хэмминга**	1,93	2,3	2	2,13	2,53	2,15	1,47	2,32	2,09	1,44	1,69
6	**Блэкмена**	0,14	0,13	0,16	0,14	0,15	0,16	0,15	0,14	0,15	0,15	0,18
7	**Блэкмена-Харриса**	0,19	0,2	0,18	0,19	0,22	0,17	0,18	0,18	0,22	0,18	0,21
8	**Наттолла**	0,25	0,4	0,26	0,29	0,29	0,27	0,21	0,24	0,31	0,23	0,26
9	**С плоской вершиной**	0,32	0,43	0,38	0,46	0,45	0,41	0,34	0,36	0,41	0,31	0,37
10	**Бартлетта-Ханна**	0,26	0,58	0,34	0,39	0,42	0,42	0,2	0,33	0,38	0,24	0,31
11	**Бомена**	0,14	0,13	0,14	0,14	0,15	0,16	0,17	0,15	0,16	0,15	0,18
12	**Тьюки при α=0.5**	0,2	0,2	0,19	0,21	0,18	0,18	0,17	0,17	0,19	0,19	0,19
13	**Кайзера при β=4**	1,92	2,33	2,02	2,09	2,52	2,13	1,54	2,28	2,06	1,47	1,69
14	**Кайзера при β=9**	0,28	0,48	0,34	0,33	0,37	0,39	0,26	0,3	0,34	0,22	0,29
15	**Чебышева при β=40дБ**	3,96	2,1	2,03	2,26	1,86	3,49	3,53	3,62	2,25	3,2	3,12
16	**Чебышева при β=80дБ**	0,63	1,43	0,98	1,18	1,41	1,13	0,51	0,84	1,15	0,49	0,78
17	**Гауссово при α=2.5**	1,43	2,08	1,71	1,89	2,37	1,79	1,12	1,78	1,81	1	1,39
18	**Парзена**	0,17	0,17	0,15	0,17	0,17	0,17	0,18	0,16	0,16	0,17	0,19

В таблице 3.5. указаны следующие числовые характеристики оконных функций [28].

- Коэффициент утечки – показывает, какая доля общей мощности окна сосредоточена в боковых лепестках его спектра.
- Уровень максимального из боковых лепестков спектра относительно значения спектра на нулевой частоте.

- Ширина главного лепестка по уровню -3дБ (с учетом отрицательных частот).

Таблица 3.5. Параметры оконных функций

"Окно"	Коэффициент утечки,	Уровень боковых лепестков, дБ	Ширина главного лепестка по уровню -3 дБ
Прямоугольное	9.15	-13.3	0.00683
Треугольное	0.28	-26.5	0.009765
Бартлетта	0.28	-26.5	0.009765
Ханна	0.05	-31.5	0.010742
Хэмминга	0.03	-42.7	0.009765
Блэкмена	0	-58.1	0.012695
Блэкмена- Харриса	0	-92.1	0.014648
Наттолла	0	-97.9	0.014648
С плоской вершиной	96.75	-93	0.023438
Бартлетта-Ханна	0.03	-35.9	0.010742
Бомена	0	-46	0.012695
Тьюки при α=0.5	3.57	-15.1	0.008789
Кайзера при β=4	0.13	-30.1	0.008789
Кайзера при β=9	0.13	-30.1	0.008789
Чебышева при β=40дБ	0.95	-40	0.008789
Чебышева при β=80дБ	0	-80	0.012695
Гауссово при α=2.5	0.01	-43.5	0.010742
Парзена	0	-53.1	0.013672

3.5 Оценки влияния нормализации на результаты достоверности системы распознавания

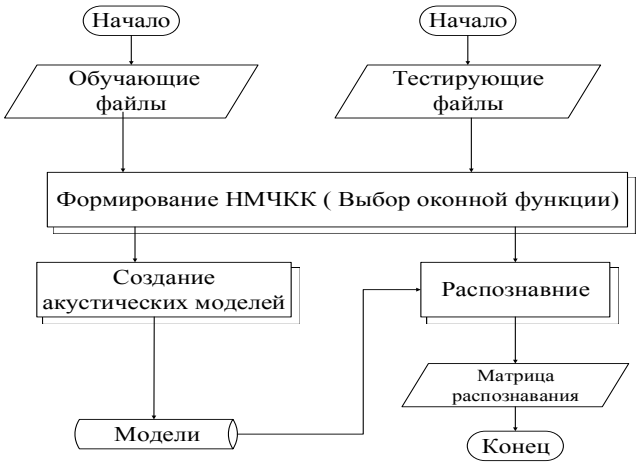


Рисунок 3.12 Алгоритм оценки влияния разных оконных функций на результаты системы распознавания

Оценка достоверности системы распознавания речи при использовании нормализации параметров речевых сигналов показывает, что разность нормализованных логарифмов спектров сигналов и МЧКК на входе и выходе фильтра-имитатора канала связи принимает наименьшее значение при использовании оконных функций Ханна и Хенинга, результаты представлены на рисунке 3.13 и на рисунке 3.14

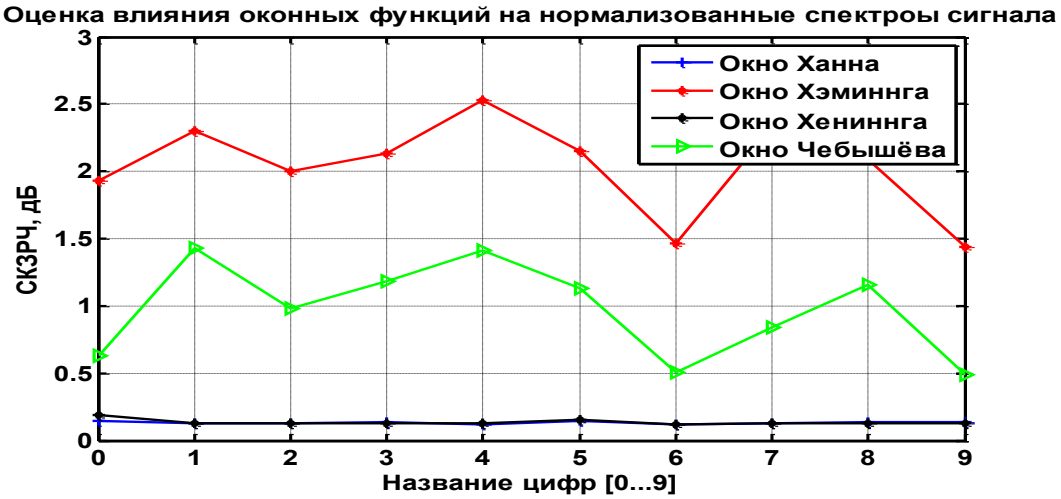


Рисунок 3.13 Влияние вида оконной функции на результаты нормализации логарифма спектра сигнала

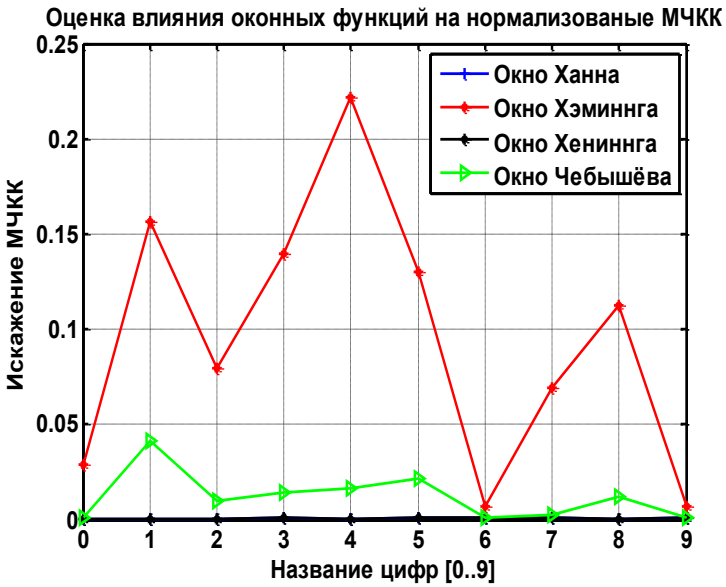


Рисунок 3.14 Влияние вида оконной функции на результаты нормализации МЧКК

Результаты тестирования САРР при использовании нормализованных параметров речевого сигнала - НМЧКК представлены в Таблице 3.6.

Таблица 3.6. Результаты тестирования САРР (акустическая модель каждого названия является общей для всех диалектов)

Название цифры	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0	98,67	0,00	0,00	0,44	0,22	0,00	0,44	5,11	0,00	0,22
1	0,00	100,00	0,00	0,44	0,00	0,67	0,00	0,00	0,00	0,00
2	0,00	0,00	92,89	0,22	0,00	0,00	0,44	0,67	0,22	0,00
3	0,44	0,00	0,00	89,11	2,44	0,22	0,00	1,56	0,00	0,00
4	0,00	0,00	0,67	0,22	95,56	0,22	0,22	0,44	0,22	0,22
5	0,00	0,00	0,00	2,00	0,00	98,67	0,00	1,11	0,00	0,00
6	0,67	0,00	0,89	0,00	0,00	0,00	97,33	0,67	0,00	1,56
7	0,22	0,00	0,00	6,67	1,56	0,00	0,44	89,11	0,00	0,22
8	0,00	0,00	2,44	0,22	0,22	0,22	0,00	0,67	99,56	0,00
9	0,00	0,00	3,11	0,67	0,00	0,00	1,11	0,67	0,00	97,78

Видно, что ошибки присутствуют при распознавании произнесений каждой цифры. Наибольшее количество ошибок соответствует распознаванию названия цифры 7 и 3 (100% – 89,11% = 10,89%). Средняя (по диагонали матрицы) оценка точности распознавания системы для трёх диалектов составляет 95,5%.

Из таблицы 3.7. матрицы распознавания северного диалекта видно, что идентификация диалектов при использовании НЧМКК позволяет значительно уменьшить число ошибок распознавания. Самое большое количество ошибок распознавания соответствует распознаванию названия цифры 7 (100% - 94% = 4%), что меньше ранее полученного результата распознавания 10,89%. соответствует распознаванию названия цифры 7 и 3. Общая оценка точности распознавания системы для СД составляет 99,2% по диагонали матрицы.

Таблица 3.7. Результаты тестирования системы распознавания произнесений названий цифр (акустическая модель каждого названия создана отдельно для северного диалекта)

Название цифры	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0 ("ноль")	100,00	0,00	0,00	0,00	0,00	0,00	0,00	6,00	0,00	0,00
1("один")	0,00	99,33	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2("два")	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
3("три")	0,00	0,67	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00
4("четыре")	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00
5("пять")	0,00	0,00	0,00	0,00	0,00	98,67	0,00	0,00	0,00	0,00
6("шесть")	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00
7("семь")	0,00	0,00	0,00	0,00	0,00	0,00	0,00	94,00	0,00	0,00
8("восемь")	0,00	0,00	0,00	0,00	0,00	1,33	0,00	0,00	100,00	0,00
9("девять")	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00

Таблица 3.8. Результаты тестирования системы распознавания произнесений названий цифр (акустическая модель каждого названия создана отдельно для южного диалекта)

Название цифры	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0 ("ноль")	99,33	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
1("один")	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2("два")	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
3("три")	0,67	0,00	0,00	90,33	0,00	0,00	0,00	0,67	0,00	0,00
4("четыре")	0,00	0,00	0,00	0,00	99,33	0,00	0,67	0,00	0,00	0,00
5("пять")	0,00	0,00	0,00	0,00	0,67	100,00	0,00	0,00	0,00	0,00
6("шесть")	0,00	0,00	0,00	0,00	0,00	0,00	98,67	0,00	0,00	0,67
7("семь")	0,00	0,00	0,00	9,67	0,00	0,00	0,67	99,33	0,00	0,00
8("восемь")	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00
9("девять")	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	99,33

Из таблицы видно, что идентификация диалектов при использовании НЧМКК позволяет значительно уменьшить число ошибок распознавания. Самое большое количество ошибок распознавания соответствует распознаванию названия цифры 3. (100% - 90,33% = 9,67%), что меньше ранее полученного результата распознавания 10,89%. (для цифры 7). Общая оценка точности распознавания системы для ЮД составляет 98,3% по диагонали матрицы.

Таблица 3.9. Результаты тестирования системы распознавания произнесений названий цифр (акустическая модель каждого названия создана отдельно для западного диалекта)

Название цифры	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0 ("ноль")	98,40	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
1("один")	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2("два")	0,00	0,00	98,40	0,00	0,00	0,00	0,00	0,00	0,00	0,00
3("три")	0,00	0,00	0,00	92,80	0,80	0,00	0,00	0,00	0,00	0,00
4("четыре")	0,00	0,00	0,80	1,60	99,20	0,00	0,00	0,00	0,00	0,00
5("пять")	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00
6("шесть")	0,00	0,00	0,00	0,00	0,00	0,00	99,20	0,00	0,00	1,60
7("семь")	1,60	0,00	0,00	5,60	0,00	0,00	0,00	100,00	0,00	0,00
8("восемь")	0,00	0,00	0,80	0,00	0,00	0,00	0,00	0,00	100,00	0,00
9("девять")	0,00	0,00	0,00	0,00	0,00	0,00	0,80	0,00	0,00	98,40

Из матрицы распознавания западного диалекта видно, что идентификация диалектов при использовании НЧМКК позволяет значительно уменьшить число ошибок распознавания. Самое большое количество ошибок распознавания соответствует распознаванию названия цифры 3 (100% - 92,80% = 7,2%), что меньше ранее полученного результата распознавания 10,89% (для цифры 7 и 3). Общая оценка точности распознавания системы для ЗД составляет 98,6% по диагонали матрицы.

Таким образом, идентификация диалектов при использовании НЧМКК позволяет значительно снизить относительную частоту ошибки распознавания.

Рассмотрим теперь зависимость достоверности (точности) распознавания от вида оконной функции. На рисунке 3.14. и рисунок 3.15. приведена зависимость значения достоверности (точности) распознавания, минимального значения по всем названиям цифр от вида оконной функции

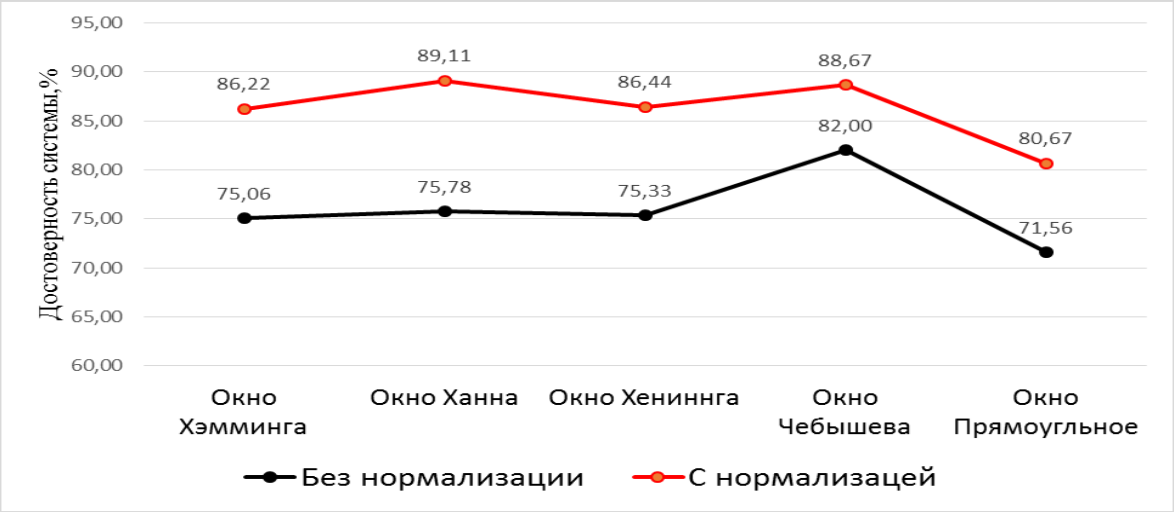


Рисунок 3.15. Достоверность распознавания названий цифр при использовании нормализации параметров речевого сигнала без идентификации

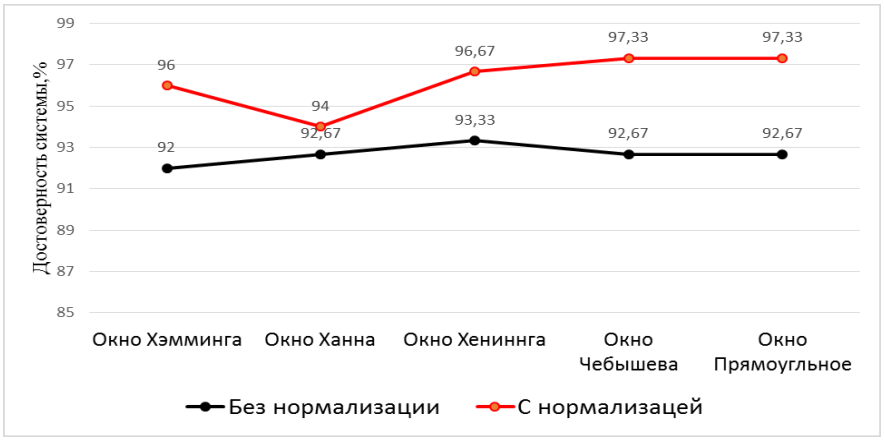


Рисунок 3.16. Достоверность распознавания названий цифр при использовании нормализации параметров речевого сигнала с идентификацией

Видно из графиков на рисунке что точность распознавания увеличивается с использованием нормализации параметров РС.

Таблица 3.9. Результаты зависимости значения достоверности (точности) распознавания, усредненного по всем названиям цифр с и без нормализации при использовании оконных функций Хэмминга и Ханна

Название цифр	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Сред. Точ.**
	Все три диалекта (окно Хэмминга)										
Без_ норм.	99,76	99,29	96,71	96	92,24	82,59	86,35	75,06	90,82	96,71	**91.6**
С_ норм.	97,78	97,33	92	86,22	99,33	98,22	94,89	91,33	99,33	97,56	**95,4**
	Все три диалекта (окно Хана)										
Без_ норм.	99,56	99,56	94,89	92	94,44	85,11	97,11	75,78	88,67	98	**92,5**
С_ норм.	98,67	100	92,89	89,11	95,56	98,67	97,33	89,11	99,56	97,78	**95,9**
	Северный диалект (окно Хэмминга)										
Без_ норм.	100	100	97,33	100	99,33	92	100	92,67	100	100	**98,1**
С_ норм.	100	98,67	100	100	100	97,33	100	96	100	100	**99,2**
	Северный диалект (окно Хана)										
Без_ норм.	99,33	100	98,67	97,33	100	92,67	100	93,33	98,67	100	**99,2**
С_ норм.	100	99,33	100	100	100	98,67	100	94	100	100	**98**
	Южный диалект (окно Хэмминга)										
Без_ норм.	98,67	100	96,67	96,67	96,67	100	95,33	99,33	100	98,67	**98,2**
С_ норм.	97,33	100	100	97,33	100	100	99,33	96	100	98,67	**98,9**
	Южный диалект (окно Хана)										
Без_ норм.	98,67	100	97,33	98,67	98	98,67	96,67	100	99,33	99,33	**98,7**
С_ норм.	99,33	100	100	87,33	99,33	100	98,67	99,33	100	99,33	**98,3**
	Западный диалект (окно Хэмминга)										
Без_ норм.	100	100	100	88,8	99,2	97,6	91,2	97,6	96,8	99,2	**98,3**
С_ норм.	98,4	100	98,4	92	97,6	100	99,2	100	100	98,4	**99,1**
	Западный диалект (окно Хана)										
Без_ норм.	100	100	96,8	98,4	99,2	100	96	99,2	99,2	99,2	**98**
С_ норм.	100	100	99,2	95,2	99,2	100	97,6	96	100	99,2	**98,6**

Видно, что использовании оконной функции Хана улучшает достоверность системы распознавания по сравнению с использованием окна Хэмминга, зависит от типа речевого сигнала.

3.6 Выводы по разделу 3

Видно, что нормализация существенно уменьшает значение метрики - степень отличий параметров сигналов на входе и выходе фильтра. Обеспечение синхронизации существенно снижает значение метрики - повышается эффектность нормализации.

В процессе эксперимента выявлено, что степень отличия нормализованных параметров сигнала на входе и выходе фильтра-имитатора увеличивается в области частот, соответствующей большей неравномерности АЧХ. Данный факт также подтверждает справедливость полученных выше выражений.

Применение оконной функции с малым уровнем боковых лепестков повышает эффективность нормализации. Использование часто используемой при спектральном анализе оконной функции Хэмминга обеспечивает меньшую эффективность нормализации по сравнению с функцией Чебышева. Следует отметить, что при выборе оконной функции важно обеспечить малый уровень боковых лепестков при большой отстройке от главного лепестка оконной функции, так как в речевом сигнале мощность низкочастотных спектральных компонентов обычно больше мощности высокочастотных компонентов.

ГЛАВА 4 Разработка программного обеспечения экспериментального исследования достоверности распознавания

В данном разделе дано описание программного комплекса (ПК) для оценки помехоустойчивости систем распознавания речи в телефонии. При проектировании средств повышения помехоустойчивости систем автоматического распознавания речи (САРР) необходимо оценивать возможности того или иного средства подавления помех различного вида. Для оценки эффективности методов обработки сигналов, поступающих на вход САРР, необходимо иметь соответствующие программные средства [12, 9].

Идентификация диалекта. Наличие диалектов повышает степень изменчивости речи, что обусловливает увеличение числа ошибок распознавания. С использованием ПО можно проводить исследование с диалектами в зависимости от каталога выборок звукозаписей.

Исследование нормализации. Снижение достоверности распознавания во многом обусловлено отличием 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. Комплекс позволяет также оценить возможности нормализации параметров речевого сигнала по их среднему значению во времени.

Исследование средств подавления помех. Автоматическое распознавание речи в телефонии осуществляется в присутствии разнообразных акустических помех, что снижает достоверность распознавания. Разработанный программный комплекс предназначен для оценки эффективность подавления аддитивных помех и с использованием спектрального вычитания (СВ), а также путем использования фильтра Винера ФВ.

Графический интерфейс комплекса значительно сокращает затраты времени на проведение исследований. Программный комплекс разработан для оптимизации параметров РС, параметров шумоподавления и параметров САРР в ходе обучения и тестирования системы с целью повышения достоверности системы распознавания и помехоустойчивости САРР.

4.1 Анализ влияния оконной функции на результат нормализации параметров речевого сигнала

Факторы, влияющие на результат нормализации параметров РС МЧКК и логарифма спектра РС по среднему значению, следующие:

- Вид оконной функции при проведении БПФ;
- Неравномерность амплитудно-частотных характеристик звуковых трактов, используемых при обучении системы распознавания речи и при ее тестировании;
- Уровень боковых лепестков оконной функций;
- Отсутствие синхронизации процедур сегментации сигнала на входе и выходе канала связи.

С целью упрощения физической интерпретации результатов анализа в качестве параметров сигнала рассматривается результат логарифмирования набора абсолютных значений коэффициентов БПФ каждого сегмента РС.

Для оптимизации этих параметров при исследовании системы автоматического распознавании речи разработали программные реализации нескольких алгоритмов для получения оценки. влияния разных факторов на процесс нормализации параметров речевых сигналов.

Разность нормализованных параметров уменьшается с повышением равномерности спектра, уменьшением уровня боковых лепестков оконной функции и отличий ЧХ микрофонов.

Общей целью всех последующих программных обеспечений являются оптимизация САРР арабской речи.

Программа анализа логарифма спектра речевого сигнала

На рисунке 4.1. представлено изображение главного окна программного обеспечения для анализа влияния нормализации логарифмов спектров речи.

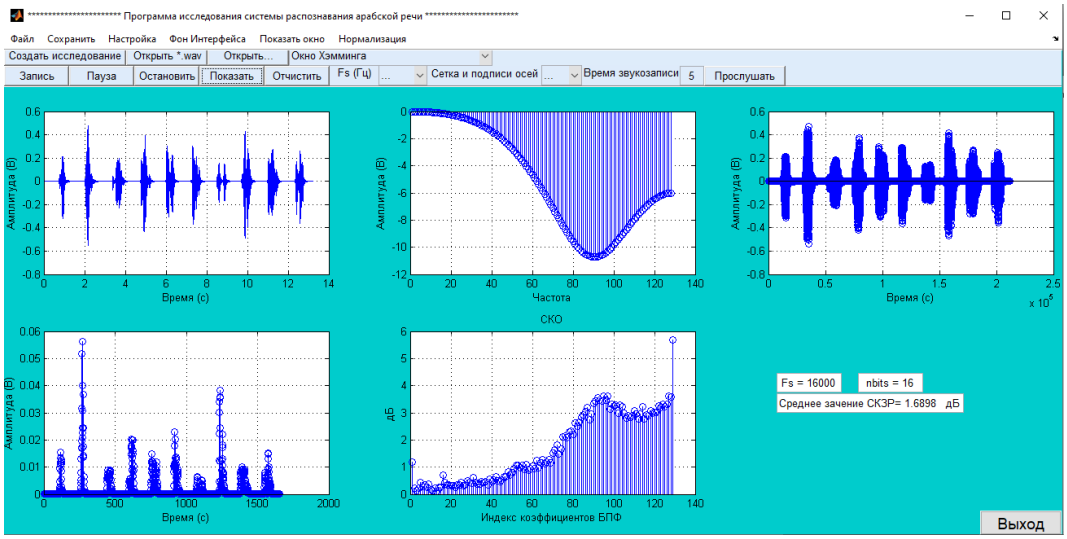


Рисунок 4.1. Программа оценки влияния нормализации логарифмы спектров

На рисунке 4.2 представлено панель управления программы.

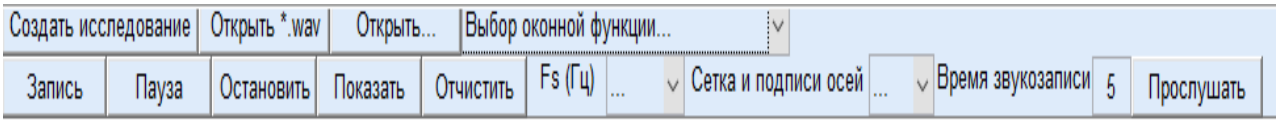


Рисунок 4.2. панель управления программы

С помощью панели управления можно записать речевой сигнал нажатием кнопкой Запись, нажатием кнопкой Пауза можно сделать паузу записи, нажатием кнопкой Остановить можно остановить запись кнопкой Показать показать временную характеристику речевого сигнала на графике**, кнопкой** Отчистить можно непосредственно очистить график. С помощью кнопки Fs (Гц) ... можно выбрать частоту дискретизации при записи сигнала. Кнопкой Прослушать можно прослушать записанный сигнал. Для определения

времени записи служит кнопка Время звукозаписи 5. Для открытия сохраненного речевого сигнала из готового каталога служит кнопка Открыть *.wav . для выхода из программы служит кнопка Выход.

Поскольку программа предназначена для анализа влияния разных оконных функций на результат нормализации логарифмов спектров речевого сигнала, тогда для выбора вида оконной функции служит выпадающее окно, представленое на рисунке 4.3

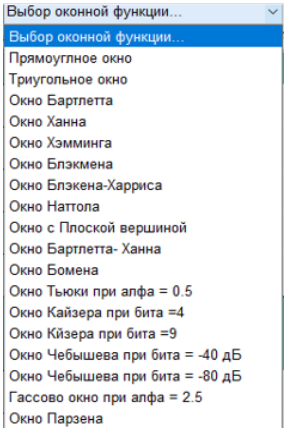


Рисунок 4.3. выпадающее окно выбора оконной функций

Для начала исследования служит выпадающее окно «Нормализация» как показано на рисунке 4.4, Основные этапы проведения исследования.

- открыть готовый речевой сигнал; при открытии можно узнать частоты дискретизации сигнала, число разрядов квантования и форму сигнала, как показано на рисунке 4.1, на графике номер 1 справа в первом ряду.
- открыть фильтр имитатор канала связи, построить его АЧХ и можно анализировать отфильтрованный сигнал, как показано на рисунке 4.1, графики номер 2 и 3 справа в первом ряду.
- Расчет логарифма спектра сигнала на входе и выходе фильтра имитатора канала связи.

- Определение разности (Выходной отфильтрованный задержанный сигнал – входной задержанный сигнал
- Расчёт среднеквадратического отклонения (СКО) разности логарифмов спектров сигнала на входе и выходе канала связи, при этом построится график, показывающий зависимость СКО от индекса коэффициентов БПФ, график 2 во втором ряду справа, и показывается СКО по всем частотам, как показано на рисунке 4.1.

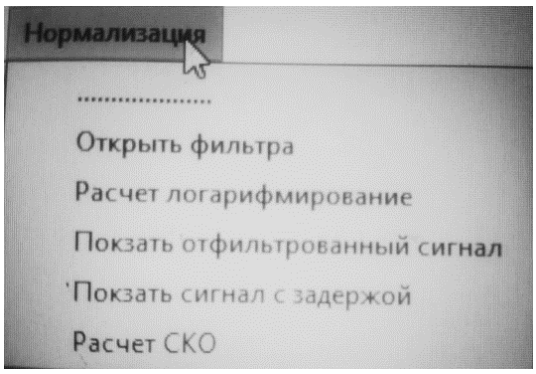
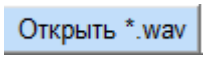
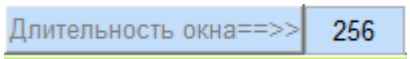


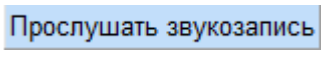
Рисунок 4.4. Выпадающее окно этап анализа

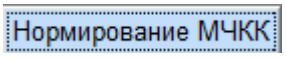
Программа анализа нормализации мел частотных кепстральных коэффициентов.

Цель данной программы заключается в исследование влияния нормализации на МЧКК, а также в исследование процесса синхронизации процесса сегментации сигнала на входе и выходе канала связи. Программа состоит из кнопки 

служит для открытии речевого сигнала. Кнопка 

для того чтобы определить длительность исследуемой оконной функции. Кнопка 

для очистки всех графиков. Кнопка 

служит для прослушивания звукозаписи. Кнопка 

служит для проведения исследования нормализации параметров речевого сигнала МЧКК и процесса сегментации. Помимо

исследования влияния нормализации параметров МЧКК в этой программе можно получить и значение СКО по всем частотам (индексы коэффициентов БПФ).

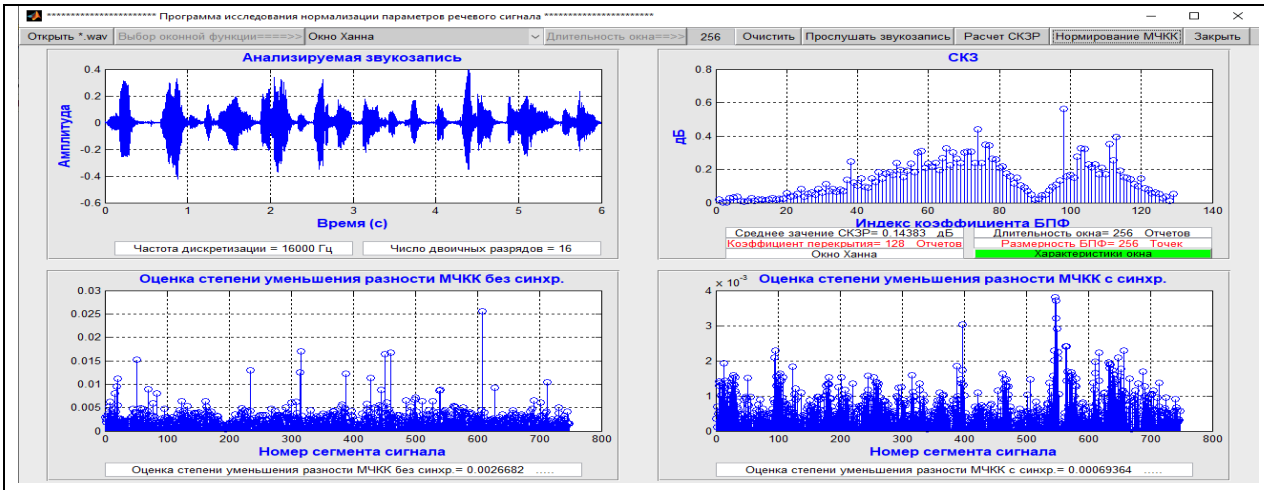


Рисунок 4.5. Программа оценки влияния нормализации на МЧКК и логарифмы спектров

Поскольку программа предназначена для анализа влияния разных оконных функций на результат нормализации параметров речевого сигнала МЧКК, для выбора вида оконной функции служит выпадающее окно, представленное на рисунке 4.6

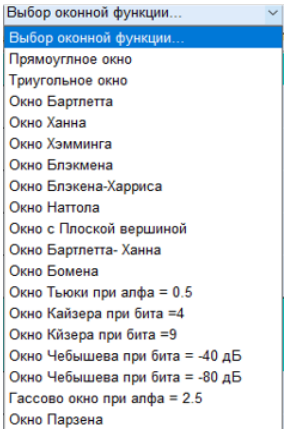


Рисунок 4.6. выпадающее окно выбора оконной функций

В результате использования программы рисунок (4.5) получено следующие результаты:

- Временную характеристику речевого сигнала, длительность исследуемого окна, коэффициент перекрытия и размерность БПФ и тип использованной оконной функции при проведения исследовании.
- Среднее значение среднеквадратического отклонения разности по всем частотам.
- Суммарную оценку степени уменьшения разности параметров МЧКК без синхронизации.
- Суммарную оценку степени уменьшения разности параметров МЧКК с синхронизацией.

С помощью программы можно исследовать разные оконные функции во временной области и в частотной области, как показано на рисунке 4.7.

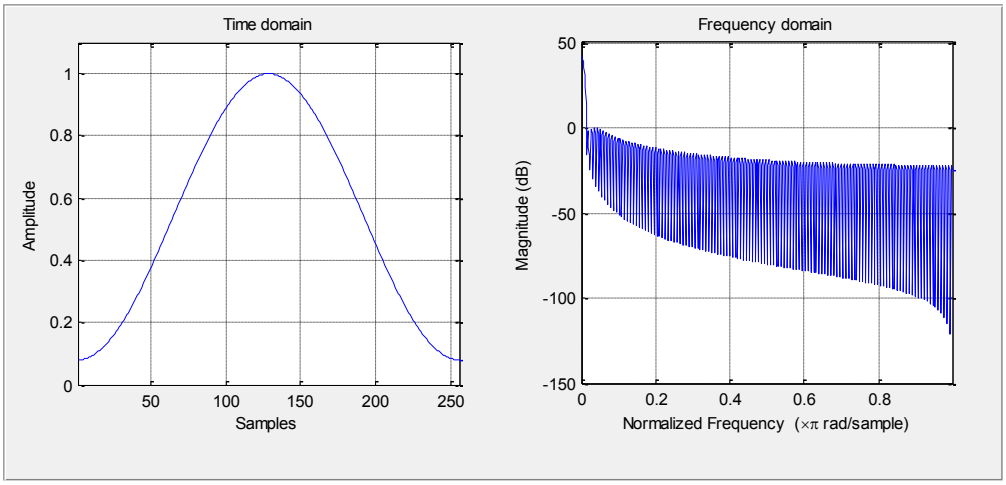


Рисунок 4.7. Временные и частотные характеристики оконной функций Хэмминга

Соответствующие алгоритмы создании программ рассмотрены в главе 3.

4.2 Программный комплекс исследования достоверности системы автоматического распознавания речи

Описание интерфейса программного комплекса

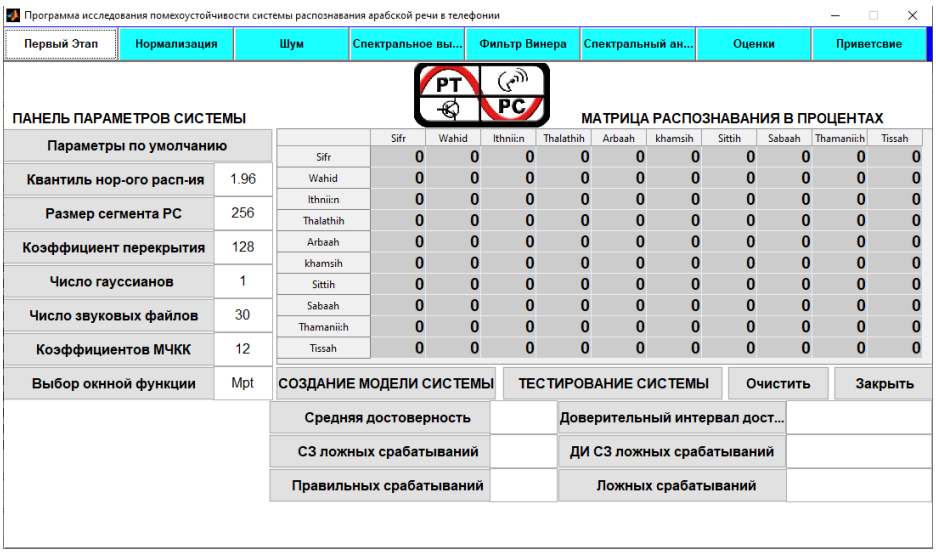


Рисунок 4.8. Изображение главного окна программного комплекса

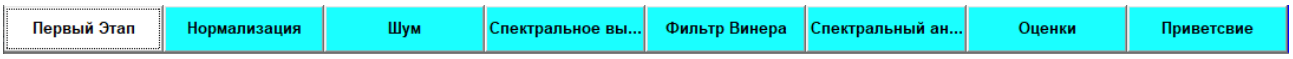


Рисунок 4.9. Главная панель управления программы

На рисунке 4.9 представлено изображение главной панели управления программой. С помощью панели можно проводить исследование в зависимости от выбора подпрограммы.

Рисунок 4.10. Панель управления входных параметров системы распознавания

На рисунке 4.10 отображен рисунок панели параметров системы. Здесь можно задать все параметры исследования. Задать параметр и анализировать его

влияние на достоверность распознавания при обучении и тестировании системы. При выборе подпрограммы для исследования системы в шумовых средах в панели параметров системы введены дополнительно параметры для проведения исследовании с шумами, как показано на рисунке 4.11

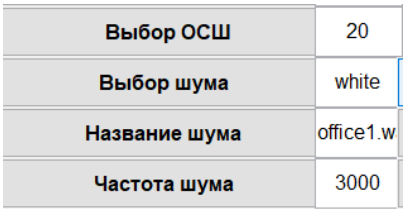


Рисунок 4.11. Панель управления входных параметров системы распознавания

На рисунке 4.11 Показано возможность выбора отношение сигнал-шум (ОШС), в данном случае оно равно 20 дБ. Можно выбрать вид шума для исследования:

- Аддитивный белый гауссовский шум.
- Гармоническая помеха при этом можно исследовать узкополосного шума при разных частотах колебаний.
- Бытовые шумы записаны при различных условиях реальной жизни, например, шума дождя, шум автобуса, шум офиса, можно добавить другие шумы для дальнейшего исследования их влияния на результаты распознавания и исследования эффективности методов шумоподавления. На рисунке 4.12 представлено матрица распознавания в процентах.

МАТРИЦА РАСПОЗНАВАНИЯ В ПРОЦЕНТАХ

	Sifr	Wahid	Ithnii:n	Thalathih	Arbaah	khamsih	Sittih	Sabaah	Thamanii:h	Tissah
Sifr	0	0	0	0	0	0	0	0	0	0
Wahid	0	0	0	0	0	0	0	0	0	0
Ithnii:n	0	0	0	0	0	0	0	0	0	0
Thalathih	0	0	0	0	0	0	0	0	0	0
Arbaah	0	0	0	0	0	0	0	0	0	0
khamsih	0	0	0	0	0	0	0	0	0	0
Sittih	0	0	0	0	0	0	0	0	0	0
Sabaah	0	0	0	0	0	0	0	0	0	0
Thamanii:h	0	0	0	0	0	0	0	0	0	0
Tissah	0	0	0	0	0	0	0	0	0	0

Рисунок 4.12. Матрица результатов распознавания в процентах

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 произнесения. Числа, находящиеся вне диагонали, являются относительными частотами ошибок распознавания. Используя результаты матрицы распознавания можно определить среднюю точность распознавания.

Средняя достоверность		**Доверительный интервал дост...**	
СЗ ложных срабатываний		**ДИ СЗ ложных срабатываний**	
Правильных срабатываний		**Ложных срабатываний**	

Рисунок 4.13. Оценки результаты распознавания

На рисунке 4.13 представлена область программы для отображения результатов работы системы. Отображаются следующие результаты:

- Средняя достоверность.
- Доверительный интервал достоверности распознавания.
- Среднее количество ложных срабатываний.
- Доверительный интервал среднего значения ложных срабатываний.
- Процент правильных срабатываний.
- Процент ложных срабатываний

Для выбора вида исследования (после ввода входных параметров) служит панель, которая показана на рисунке 4.14.

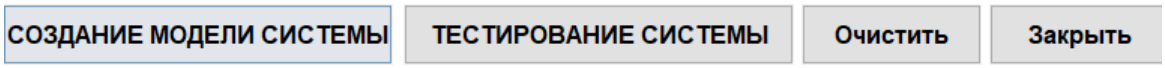
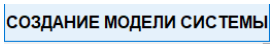
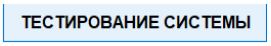


Рисунок 4.14. Панель начала обучения, тестирования системы и кнопки очистки и закрытия программы

Основные кнопки панели представлены на рисунке



служит для создания модели для обучения системы.



служит для тестирования системы.

Кнопки 

для отчистки рабочей среды и закрыти программы соответственно.

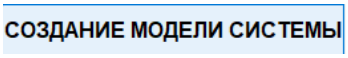
4.3 Экспериментальные исследования САРР помощью программного комплекса

Порядок использования программы начинается заданием следующих входных параметров системы:

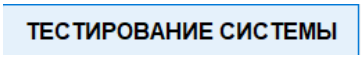
- Квантиль нормального распределения, используется для расчета вероятности доверительного интервала.
- Размер сегмента речевого сигнала.
- Коэффициент перекрытия сегментов речевого сигнала.
- Число гауссианов.
- Число исследуемых звуковых файлов.
- Число коэффициентов МЧКК.
- Выбор вид исследованной оконной функции.

В случае исследования шумов и методов шумоподавления требуются следующие входные параметры:

- Значение ОСШ
- Выбор типа шума
- Выбор частоты гармонического колебаний помехи.
- Тип шума в случае бытовых шумов.

После задания входных параметров следует обучить систему путём нажатия на кнопку 

.

- Для тестирования системы служит кнопка 

- Для очистки результатов исследования служит кнопка 

.
- Для закрытия программы можно использовать кнопку 

или  .

Исследование достоверности распознавания без нормализации и без воздействия помех.

Изображенные закладки позволяют выбрать режим работы комплекса. Закладка "Первый этап" обеспечивает настройку комплекса на параметры САРР. К числу параметров САРР относятся такие, как: размер сегмента РС; вид оконной функции; величина перекрытия сегментов; число гауссианов, которые используются для построения функции плотности распределения вероятностей в зависимости от значения параметров РС; количество МЧКК, которые являются параметрами РС; количество звуковых файлов, которые используются для тестирования САРР.

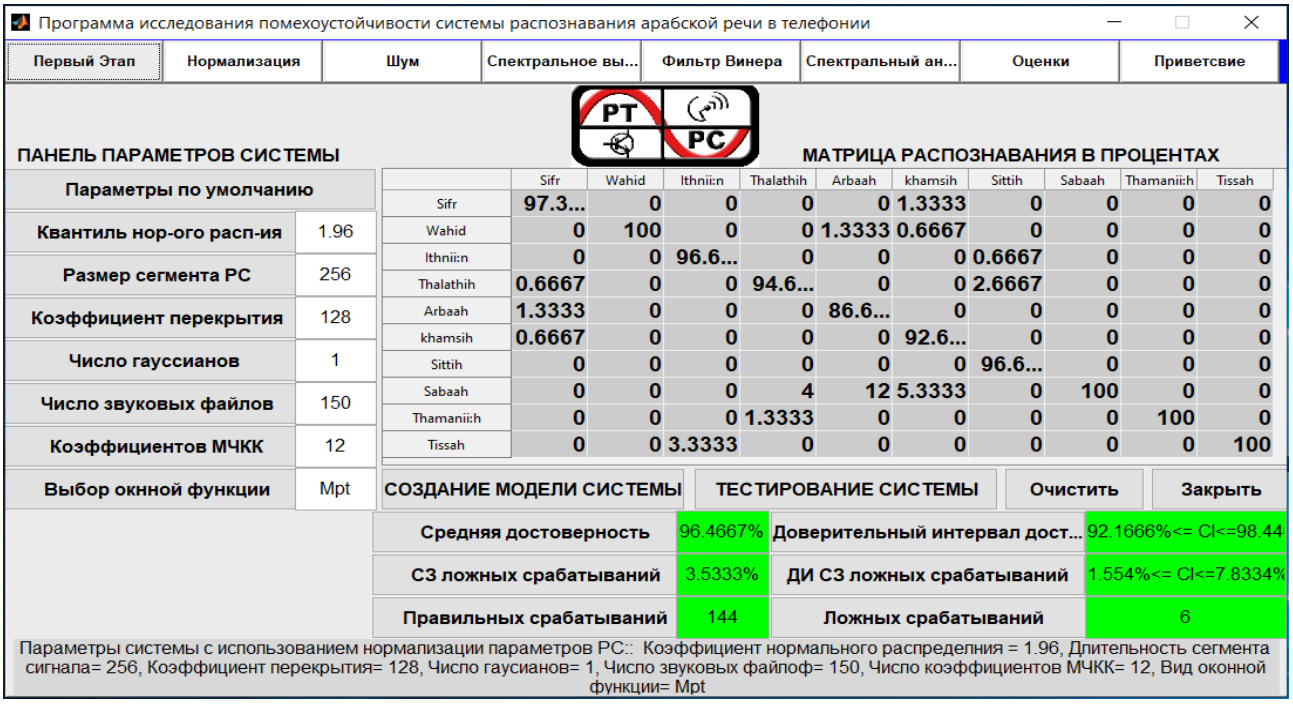


Рисунок 4.15. Закладка первого этапа программного комплекса

При отсутствии помех средняя достоверность распознавания арабских названий цифр составляет более 96 %

Исследование достоверности системы САРР с нормализацией МЧКК параметров речевого сигнала:

Закладка "Нормализация" позволяет определить точность распознавания голосовых команд, когда для построения моделей голосовых команд используются нормализованные МЧКК.

Форма оконной функции, используемая при проведении БПФ влияет на значения нормализованных параметров сигнала. С увеличением неравномерности спектра сигнала и уровням боковых лепестков частотной характеристики оконной функции это влияние увеличивается.

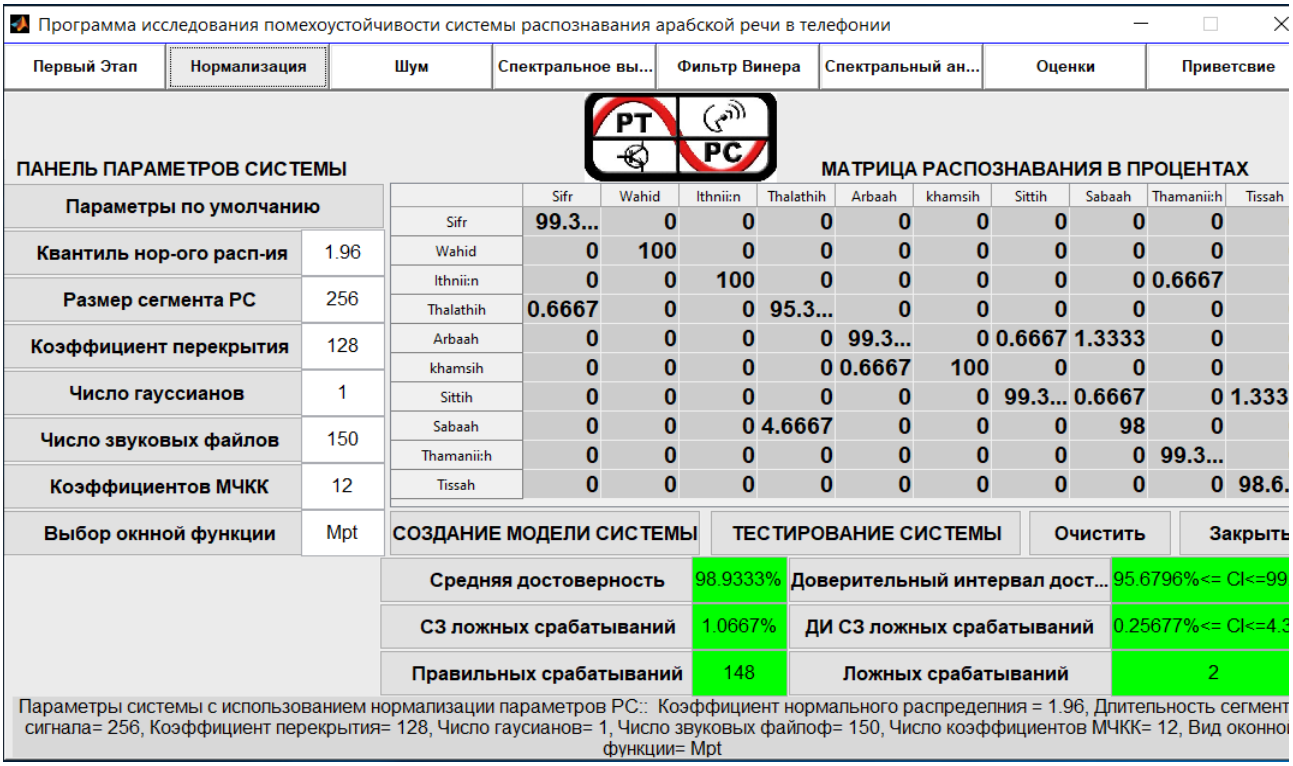


Рисунок 4.16. Составная часть программы, закладка нормализации

При использовании нормализации параметров PC средняя достоверность составила более 98 %

Исследование достоверности системы САРР в шумовых средах

Закладка: "Шум", позволяет протестировать САРР и определить точность распознавания отдельных голосовых команд, когда на сигнал накладывается помеха.

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 PC по сравнению с теми значениями, которые использовались при создании моделей речевых сигналов. [44, 91].

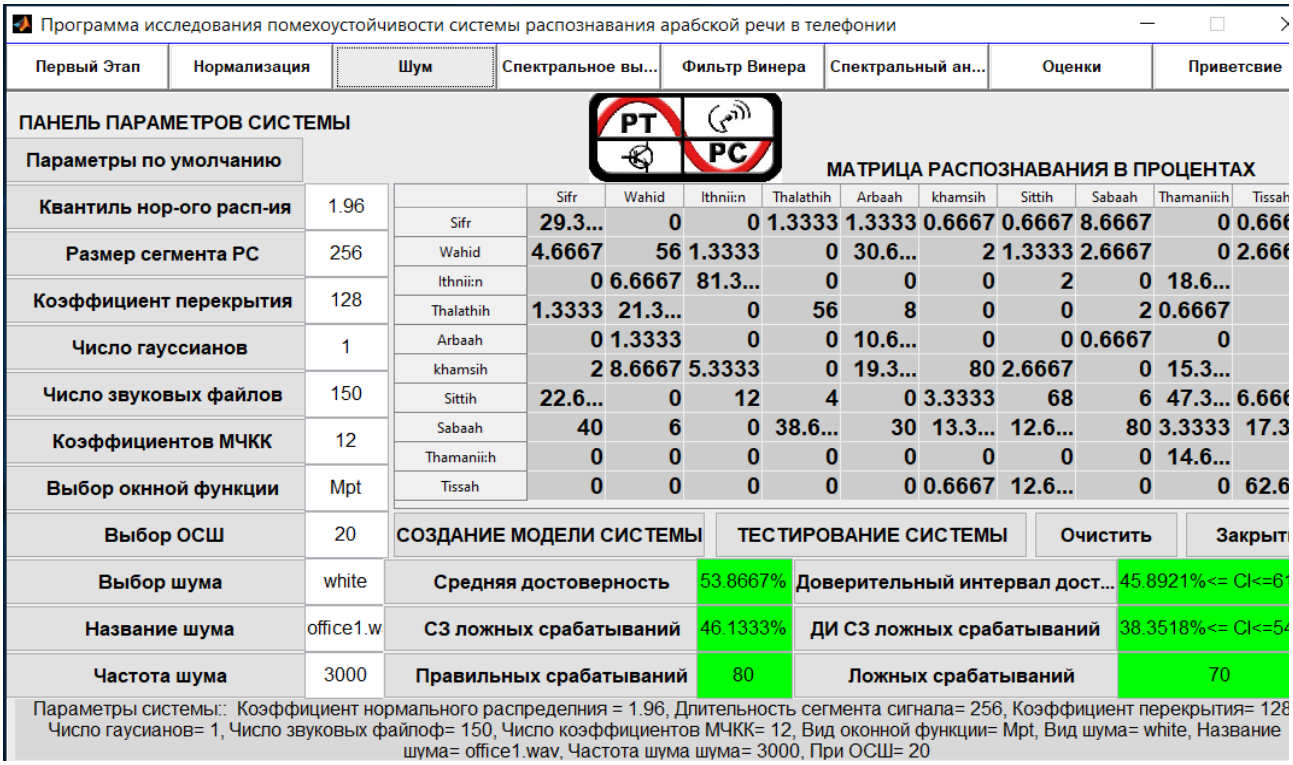


Рисунок 4.17. Составная часть программы, закладка шума

Средняя достоверность распознавания зашумленных произнесений составила 53,8 % при ОСШ 20дБ.

Оценка достоверности использовании спектрального вычитания и фильтра Винера:

Закладки "Спектральное вычитание", «Фильтр Винера» применяются для исследования средств подавления помехи в виде спектрального вычитания и фильтра Винера.

При выборе закладок появляется возможность настройки указанных средств подавления помех по критерию максимума точности распознавания и выбора значения отношения сигнал шум.

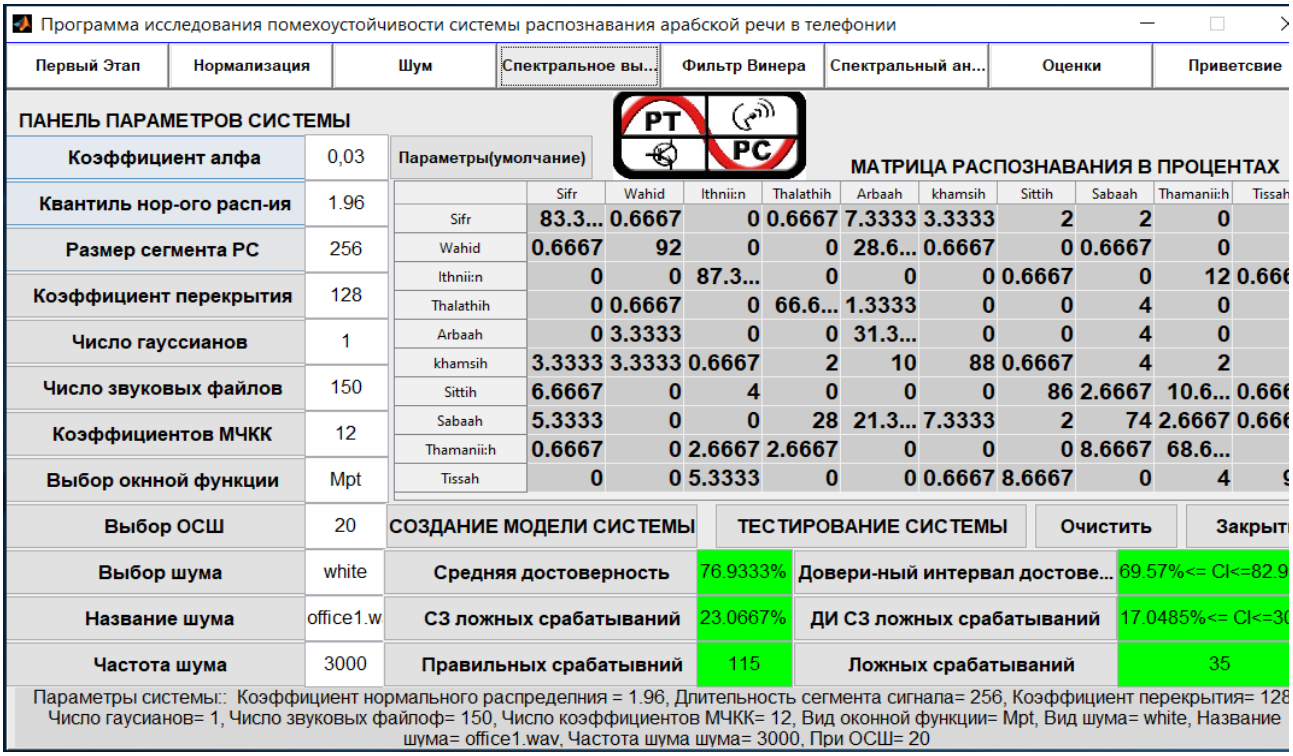


Рисунок 4.18. Составная часть программы, закладка спектральное вычитание

При использовании спектрального вычитания получена средняя достоверность САРР выше 76 % при ОСШ 20дБ

Исследование достоверности системы САРР в шумовых средах с использованием фильтра Винера:

Использование фильтра Винера обеспечивает среднюю достоверность распознавания выше 82 % при ОСШ 20дБ. При использовании спектрального вычитания получена средняя достоверность САРР выше 76 % при ОСШ 20дБ

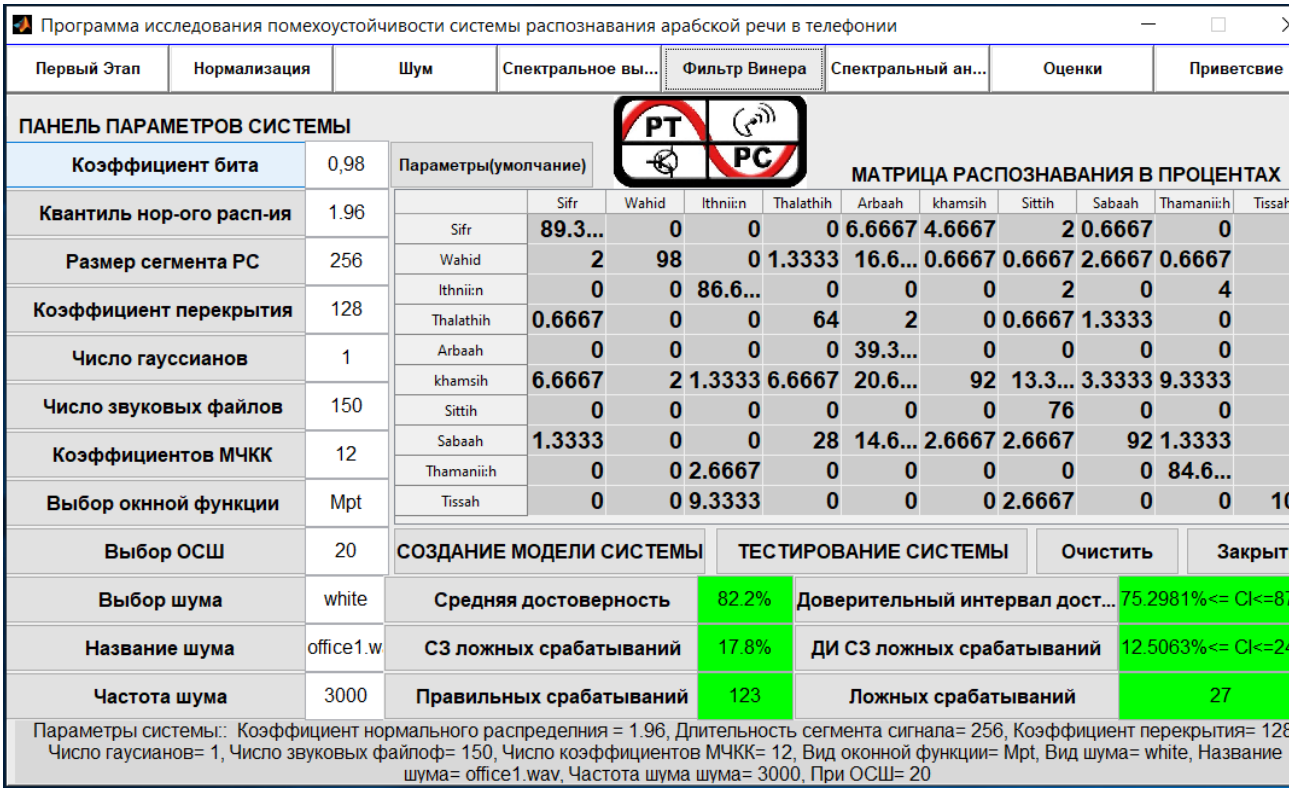


Рисунок 4.19. Составная часть программы, закладка фильтр Винера

.

Результаты работы САРР отражаются таблицей "Матрица распознавания". На рисунке 4.20 отображен вид таблицы для случая, когда распознаются голосовые команды в виде произнесений арабских названий цифр при ОСШ равном 20 дБ. Средства подавления помехи в данном случае не применяются.

	Sifr	Wahid	Ithnii:n	Thalathih	Arbaah	khamsih	Sittih	Sabaah	Thamanii:h	Tissah
Sifr	100	0	0	0	0	0	0	4	0	0
Wahid	0	24	0	0	0	0	0	0	0	0
Ithnii:n	0	0	100	0	0	0	0	0	0	0
Thalathih	0	4	0	100	0	4	0	0	0	0
Arbaah	0	0	0	0	52	0	0	0	0	0
khamsih	0	0	0	0	0	88	0	0	0	0
Sittih	0	72	0	0	44	8	100	0	0	0
Sabaah	0	0	0	0	0	0	0	80	0	0
Thamanii:h	0	0	0	0	0	0	0	0	100	0
Tissah	0	0	0	0	4	0	0	16	0	100

Рисунок 4.20 Результаты распознавания арабских названий цифр при ОСШ =20 дБ

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 голосовой команды. Числа, находящиеся вне диагонали, являются относительными частотами ошибок распознавания. Например, произнесение названия цифры 1 (Wahid) ошибочно распознано как название цифры 6 (Sittih) в 72% случаев.

Помимо точечных оценок вероятностей правильного и ошибочного распознавания возможно получение интервальных оценок. Интервальные оценки вероятности правильного распознавания указываются в полях: "Среднее значение точности" и "Доверительный интервал". Аналогичные поля используются для интервальных оценок вероятностей ошибочного распознавания.

Закладки: «Спектральный анализ", «Оценки" позволяют исследовать частотные и временные характеристики сигнала. Также можно получить оценку точности настройки моделей голосовых команд

Закладка "Приветствие" используется для вывода информации о текущей версии программного комплекса

4.4 Выводы по разделу 4

- С помощью разработанного программного обеспечения можно проводить следующие виды исследования системы распознавания речи
- Оценка достоверности распознавания в условия воздействия реальных помех при применении спектрального вычитания и Фильтра Винера.

- Оценка достоверности распознавания речи при изменении частотных характеристики канала связи, когда исследуются нормализация мел частотных кепстральных коэффициентов по среднему значению.
- Оценка искажений параметров речевых сигналов при воздействии помех и изменении частотной характеристики канала связи.
- 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.

Заключение

Целью диссертационной работы является исследование методов и разработка алгоритмов обработки речевых сигналов для повышения достоверности системы автоматического распознавания арабской речи (республика Йемен) в телефонии.

Применение САРР сдерживается большим количеством ошибок распознавания речи. Одной из причин появления ошибок является отличие значений параметров речевых сигналов, которые были использованы при обучении САР от значений аналогичных параметров, поступающие на вход САР при ее эксплуатации. Отличие параметров вызвано, в частности, влиянием частотной характеристики канала связи и наличием помех в речевых сигналах. Итог проведенных исследований характеризуется следующими результатами:

1. Исследованы возможности включения идентификатора диалектов арабской речи в состав САРР. Разработана методика оценки повышения достоверности распознавания речи при включении идентификатора диалектов арабской речи в состав САРР. Получены выражения оценки вероятности ошибки идентификации. Проведены экспериментальные исследования, позволяющие оценить повышение достоверности распознавания при использовании идентификации.
2. Предложенный алгоритм идентификации диалектов жителей Йемена обеспечивает относительную ошибку идентификации равную 0,24%. что позволяет повысить достоверность распознавания арабских названий цифр, как минимум, на 7%.
3. Проведено экспериментальное исследование возможности применения спектрального вычитания и фильтра Винера для подавления аддитивных помех. Использование указанных средств обеспечивает

повышение достоверности распознавания, если отношение сигнал-щум меньше 35 дБ.

4. Разработаны методика и соответствующий алгоритм оценки эффективности нормализации мел-частотных кепстральных коэффициентов (МЧКК) для снижения влияния частотной характеристики канала связи на достоверность распознавания голосовых команд.
5. Получены выражения для анализа влияния вида оконной функции, используемой при дискретном преобразовании Фурье, на результат нормализации по среднему значению МЧКК.
6. Установлено, что использование оконных функций, частотная характеристика которых имеет малый уровень боковых лепестков, удаленных от главного лепестка, обеспечивает большую степень стабилизации нормализованных параметров речевого сигнала МЧКК. Установлено, что использование оконной функции Ханна наиболее целесообразно при нормализации.
7. В результате проведенных экспериментов установлено, что использование нормализации повышает достоверность распознавания
8. Разработано программное обеспечение, реализующее разработанные алгоритмы исследования САРР, которое позволяет обеспечить разработку САРР с учетом использования идентификатора диалектов, средств подавления помех и нормализации параметров речевого сигнала. Получены свидетельства о государственной регистрации разработанного программного обеспечения.
9. Составлены и обработаны объемные выборки аудиозаписей арабской речи для обучения и тестирования САРР, а также для настройки и тестирования идентификатора диалектов.

Список сокращений и условных обозначений

АЧХ – Амплитудно-частотнная характеристика.

АГБШ – Аддитивный белый гауссовский шум.

БПФ – Быстрое преобразование Фурье.

ВсеД – Все три исследуемые диалекты в группе

САРР – Система автоматического распознавания речи.

СД – Северный диалект.

ДАЯ – Диалектный арабский язык.

РС – Речевой сигнал.

ЧХ – Частотная Характеристики.

МЧКК – Мел-Частотные Кепстральные Коэффициенты.

ЗД – Западный диалект.

ЮД – Южный диалект.

HMM – Hidden Markov Model (Скрытые Марковские модели).

ДКП – DCT – Дискретное Косинусное Преобразование (Discrete cosine transform).

СВ – Спектральное вычитание.

ФВ – Фильтр Винера.

МСМП – Модели скрытых Марковских процессов

СКО – Среднеквадратическое отклонение.

ОСШ – Отношение сигнал-шум.

ASR – Automatic System Recognition.

CMN – Нормализация среднего кепстра (Cepstral Mean Normalization)

CVN – Нормализация средней дисперсии кепстра

CMVN – Cepstral Mean and Variance Normalization

CGN – Нормализация усиления кепстра.

DA – Диалектный арабский язык.

DD – Decision Direct (метод для построения фильтра Винера прямого решения).

EGY – Египетский диалект.

GLF – Диалект страны арабского залива.

GMM - Гауссовский смесь распределения.

IRQ – Иракский диалект

LAV – Левантийский диалект.

MFCC - Mel-Frequency Cepstral Coefficients.

MSA – Modern Standard Arabic (Современный диалект арабского языка).

MFCCN – нормализованные по среднему значению МЧКК – MFCC.

NOR – Североафриканский диалект.

RASTA – Нормализация кепстра в реальном времени.

SNR – Signal Noise Ratio (Отношение сигнал-шум).

TSNR – Tow step noise reduction (двухступенчатое подавление шума).

Yem – Йеменский диалект.

Список использованной литературы

1. Аль-Дайбани А.М. Подавление помех в системе распознавания арабской речи / А.М Аль-Дайбани, Е.К.Левин // Журнал «Проектирование и технологии электронных средств» №4/2016 С.: 14 - 18. —2016.
2. Аль-Дайбани А.М. Анализ возможностей подавления влияния частотной характеристики канала связи на параметры речевого сигнала / А.М Аль-Дайбани, Е.К.Левин // Журнал «Проектирование и технологии электронных средств» №3/2018 С.: 14 - 18. —2018.
3. Аль-Дайбани А.М. идентификация диалектов арабской разговорной речи при автоматическом распознавании голосовых команд вы телефонии / А.М Аль-Дайбани, Е.К.Левин // Журнал «Проектирование и технологии электронных средств» №1/2019 С.: 35 - 40. —2019.
4. Аль-Дайбани А.М. О возможности использования системы распознавания речи при регистрации заявок на получение медицинских услуг в / А.М Аль-Дайбани, Е.К Левин // XII Международной научно-технической конференции «Физика и радиоэлектроника в медицине и экологии – ФРЭМЭ’2016» / С.: 301-303, г. Суздаль, 05 – 07 июля 2016.
5. Аль-Дайбани А.М. Возможности использования автоматизированных систем регистрации заявок на получение медицинских услуг в Йемене / А.М Аль-Дайбани // XII-й Международной научно-технической конференции «Физика и радиоэлектроника в медицине и экологии - ФРЭМЭ2016» / С.: 322-324, г. Суздаль, 05 – 07 июля 2016.
6. Аль-Дайбани А.М., классификация диалектов республики Йемен для повышения точности распознавания речи / А.М Аль-Дайбани // XIII-й Международной научно-технической конференции «Физика и

радиоэлектроника в медицине и экологии - ФРЭМЭ2018», С.: 358-361 г. Суздаль, 03 – 05 июля 2018.

7. Аль-Дайбани А.М. Зависимость результатов нормализации MFCC от вида используемой оконной функции / А.М Аль-Дайбани // XIII-й Международной научно-технической конференции «Физика и радиоэлектроника в медицине и экологии - ФРЭМЭ2018», С.: 361-365 г. Суздаль, 03 – 05 июля 2018.
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-й Международной научно-технической конференции «Физика и радиоэлектроника в медицине и экологии - ФРЭМЭ2018», С.: 365-369 г. Суздаль, 03 – 05 июля 2018.
9. Аль-Дайбани А.М., Программный комплекс для оценки помехоустойчивости систем распознавания речи в телефонии / А.М Аль-Дайбани, Е.К Левин // XIII-й Международной научно-технической конференции «Перспективные технологии в средствах передачи информации - ПТСПИ-2019», С.: 211-212. – г. Суздаль, 03 – 05 июля 2019.
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-й Международной научно-технической конференции «Перспективные технологии в средствах передачи информации - ПТСПИ-2019». – С.: 211-212 г. – Суздаль, 03 – 05 июля 2019.
11. 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.

12. 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.
13. 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.
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. Библиотека звуков [Электронный ресурс] / режим доступа // https://www.liveinternet.ru/users/3629609/post130428638
16. ИНТУИТ (национальный открытый университет). Потребительские свойства телефонных аппаратов / Принципы построения микрофона и телефона. http://www.intuit.ru/studies/courses/1077/211/lecture/5449.
17. Информационные киоски - [Электронный ресурс] / Режим доступа // http://dreamapp.ru/. ата обращения 10.04.2016.

18. Ковалев А. А., Шарбатов Г.Ш. Учебник арабского языка / А. А. Ковалев, Г.Ш. Шарбатов // 3-е изд., перераб. и доп., М.: Восточная лит-ра РАН. Москва. — 753с. — 1998.
19. Левин, Е.К. Компенсация помех при автоматическом распознавании голосовых команд в телефонии [Текст] / Е.К. Левин // Проектирование и технология электронных средств. — 2011. — №3. — С.45-49.
20. Левин, Е.К. Разработка средств исследования и повышения помехоустойчивости систем автоматического распознавания голосовых команд в телефонии: диссертация / Е.К. Левин // Владимир: Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых. — 257с. — 2014.
21. Левин Е. К. Исследование алгоритмов обработки сигналов в системе MATLAB/ Е.К. Левин//метод. указания к лабораторным работам. – Владимир: Изд-во Владим. гос. ун-та, 2011. – 78 с.
22. Меденников И. П. Методы, алгоритмы и программные средства распознавания русской телефонной спонтанной речи / И. П. Меденников // Санкт-Петербург: Санкт-Петербургский государственный университет. — 2016. — 148с.
23. MedVox Интеллектуальная платформа [Электронный ресурс] / Режим доступа: http://www.s2snext.com/ru/resheniya/medvox.html Дата обращения 22.04.2016.
24. Попов В. С. Исследование влияния боковых лепестков спектра окон на погрешности обработки и передачи сигнала / В. С. Попов // МГТУ им. Н. Э. Баумана, кафедра: Информационные системы и телекоммуникации. —2010. —. Режим доступа: http://windowing-matlab.narod.ru/ Дата обращения: 2017.

25. Рабинер, Л.Р. Цифровая обработка речевых сигналов: Пер. с англ. / Л.Р. Рабинер, Р.В. Шафер / ред. М.В. Назаров, Ю.Н. Прохоров. – М.: Радио исвязь, 1981. – 496 с.
26. Романенко П.Н. Разработка системы автоматического распознавания речи для египетского диалекта арабского языка в телефонном канале / А.Н. Романенко // научно-технический вестник информационных технологий, механики и оптики. —Том 16. — № 4. —2016.
27. Регистратура33.рф - [Электронный ресурс] / Режим доступа: http://регистратура33.рф. Дата обращения 20.04.2016.
28. Сергиенко, А. Б. Цифровая обработка сигналов: учебное пособие для вузов по направлению 210300 "Радиотехника" [Text] / А. Б. Сергиенко. – 2-е изд. – СПб: БХВ-Петербург. — 2006 . — 751 с.
29. Сагациян М.В. Разработка и исследование коллективных нейросетевых алгоритмов дикторонезависимого распознавания речевых сигналов / М.В Сагациян // Ярославский государственный университет им. П.Г. Демидова. Владимир. — 134с. — 2015.
30. Тупицин Г.С Предобработка речевых сигналов в системах автоматической идентификации диктора / Г.С Тупицин // Владимирский государственный университет. —Владимир. —137с. — 2015.
31. Топников А.И. Оценка разборчивости и обработка речевых сигналов в задаче шумоподавления / А.И. Топников // Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых. —Владимир. —130с. — 2012.
32. Al-Ayyoub M. Spoken Arabic Dialects Identification: The Case of Egyptian and Jordanian Dialects / M. Al-Ayyoub, M. K. Rihani, N. I. Dalgamoni, N.

A. Abdulla // 2014 5th International Conference on Information and Communication Systems (ICICS). — pp. 1–6— Irbid, Jordan. —2014.

33. Al-Zabibi M. An Acoustic-Phonetic Approach in Automatic Arabic Speech Recognition / M. Al-Zabibi // Ph.D. thesis, Loughborough University of Technology. — pp. 275. — 1990.

34. Arabic Numbers [Электронный ресурс] / режим доступа: https://blogs.transparent.com/arabic/arabic-numbers-1-10-msa-vs-dialects/. дата обращения: октябрь 2017.

35. Adam M. Cepstral variance normalization for audio feature extraction / M. Adam Marek, T. Bocklet // United States: Patent Application Publication. — Nov. 8. — 2018.

36. Ali A. Multi-Dialect Arabic Broadcast Speech Recognition/ A. Ali / Institute for Language, Cognition and Computation School of Informatics // University of Edinburgh. — Pages: 193. — 2018. Режим доступа: https://www.era.lib.ed.ac.uk/handle/1842/31224

37. AbuZeina D. Cross-Word Modeling for Arabic Speech Recognition / D. AbuZeina, M. Elshafei // SpringerBriefs in in Electrical and Computer Engineering. — pages 82. — 2012.

38. Al-Otaibi Y. A. Speech Recognition System of Arabic Digits based on A Telephony Arabic Corpus / Y. A. Alotaibi, M. Alghamdi, F. Alotaiby // Comput. Speech and Language. - Department of Electrical Engineering, King Saud University, Riyadh, Saudi Arabia. — pages 4— 2008.

39. Al-Otaibi Y. A. Comparative Study of ANN and HMM to ArabicDigits Recognition Systems / Y. A. Alotaibi // JKAU— pp: 43:60. — 2008.

40. Acero A. Augmented cepstral normalization for robust speech recognition / Acero A. // IEEE workshop on automatic speech recognition. — pp. 147–148. —1995

41. Akbacak M. Effective Arabic Dialect Classification Using Diverse Phonotactic Models / M. Akbacak, D. Vergyri, A. Stolcke, N. Scheffer, A. Mandal // in:12th Annual Conference of the International Speech Communication Association, Florence, Italy, August 27-31, 2011.— pp. 737–740. — 2011.

42. Abdulghani M. A. The Phonology And Morphology Of Yemeni Tihami Dialect: An Autosegmental Account / M. A Abdulghani [Электронный ресурс]. — July 2010. режим доступа: https://core.ac.uk/display/32600773. дата обращения: октябрь 2017.

43. Alorifl F. S. Automatic identification of arabic dialects using hidden markov models / F. S. Alorifl / Ph.D. thesis // University of Pittsburgh. pp. 132. — 2008.

44. Boll S. F. Suppression of acoustic noise in speech using spectral subtraction / Boll S. F. // IEEE Trans. Acoust., Speech, Signal Process. — Vol. ASSP-27, no. 2. — pp. 113–120. — Apr. 1979.

45. Bougrine H.C.S. Spoken Arabic Algerian Dialect Identification / H.C.S.Bougrine, A. Abdelali // In: IEEE: 2018 2nd International Conference on Natural Language and Speech Processing (ICNLSP). Algiers, Algeria — 2018.

46. Biadsy F. Using Prosody and Phonotactics in Arabic Dialect Identification // F. Biadsy, J. Hirschberg // in: INTERSPEECH 2009,10th Annual Conference of the International Speech Communication Association . — pp. 208–211. Brighton, United Kingdom. — September 6-10. — 2009.

47. Biadsy F. Spoken Arabic Dialect Identification Using Phonotactic Modeling/ Fadi Biadsy, Julia Hirschberg, Nizar Habash/ Published in:Proceeding Semitic '09 Proceedings of the EACL 2009 Workshop on Computational Approaches to Semitic Languages. — Pages 53-61. — 2009.

48. Biadsy F. Dialect and Accent Recognition Using Phonetic-Segmentation Supervectors / F. Biadsy, J. Hirschberg, D. P. W. Ellis,, in: INTERSPEECH 2011,12th Annual Conference of the International Speech Communication Association.— pp. 745–748. — Florence, Italy. — August 27-31. 2011.

49. Boril H. UT-Scope: Towards LVCSR under Lombard effect induced by varying types and levels of noisy background / H. Boril, J.H.L. Hansen // in Proc. IEEE Int. Conf. Acoust. Speech Signal, ICASSP, Prague. — pp. 4472-4475. — 2011

50. Boril H. Unsupervised equalization of Lombard effect for speech recognition in noisy adverse environments / H. Boril, J.H.L. // IEEE Transactions on Audio, Speech, and Language Processing. — vol. 18. — № 6. — pp. 1379–1393. — Aug. 2010.

51. Chen N. F. A linguistically-informative approach to dialect recognition using dialect-discriminating context dependent phonetic models/ N. F. Chen, W. Shen, J. P. Campbell/ in: 2010 IEEE International Conference on Acoustics, Speech and Signal Processing. — pp. 5014–5017. — 2010.

52. Davis S. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences / S. Davis, P. Mermelstein // IEEE transactions on acoustics, speech, and signal processing. — Vol. 28. — Issue 4. — pages 357–366. — 1980.

53. Deng Li Recent advances in deep learning for speech research at Microsoft / Li Deng, Jinyu Li, Jui-Ting Huang, Kaisheng Yao, Dong Yu, Frank Seide, Michael Seltzer, Geoff Zweig, Xiaodong He, Jason Williams, et In // Proc. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. — pp.: 8604 - 8608 — 21 October 2013.

54. Djellab M. Algerian Modern Colloquial Arabic Speech Corpus (AMCASC): regional accents recognition within complex socio-linguistic environments / M. Djellab, A. Amrouche, A. Bouridane, N. Mehallegue // Language Resources and Evaluation. — Vol. 51. — Issue 3. — pages 613–641. — 2017.

55. Donia G. Opinion Mining for Arabic Dialects on Twitter / G. Donia, M. Alfonse, M. E. El-Sayed, A.-B. M.Salem// Egypt Egyptian Computer Science Journal. — Vol. 42. — No.4. — September 2018.

56. Đorđe G. Comparison of Cepstral Normalization Techniques in Whispered Speech Recognition / G. Đorđe, Š. P. Dragana, G. Jovan, M. Branko // Advances in Electrical and Computer Engineering. — Vol. 17. — Number 1. — February 2017.

57. 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 (USBEREIT)/ Publisher: IEEE/ Yekaterinburg. — Russia. pp. — 204-207. — 2019.

58. El Said B. Modern written Arabic: A comprehensive grammar/ B. El Said, M. Carter, A. Gully // Routledge. — 812c. — 2004.

59. Elmahdy M. Modern standard Arabic based multilingual approach for dialectal Arabic speech recognition / M. Elmahdy, R. Gruhn, W. Minker, S.

Abdennadher // In 2009 Eighth International Symposium on Natural Language Processing. — 2009.

60. Eide, E. A parametric approach to vocal tract length normalization / E. Eide H. Gish // Proc. IEEE International Conference on Acoustics, Speech and SignalProcessing (ICASSP).— Vol. 1. — P. 346—348. — 1996.
61. Elaraby M. Deep Models for Arabic Dialect Identification on Benchmarked Data / M. Elaraby M. Abdul-Mageed // Proceedings of the Fifth Workshop on NLP for Similar Languages, Varieties and Dialects. — pages 263–274. — Santa Fe, New Mexico, USA, August 20, 2018.
62. Greenberg C. The 2011 NIST Language Recognition Evaluation / C. Greenberg, A. Martin, M. Przybocki // 13th Annual Conference of the International Speech Communication Association. — September 9-13. — 2012.
63. Huang X. Spoken language processing: A guide to theory, algorithm and system development / X. Huang // Prentice Hall PTR, Englewood Cliffs, NJ. — pages 960. — 2002.
64. Hanani A. Palestinian Arabic Regional Accent Recognition / A. Hanani, H. Basha, Y. Sharaf, S. Taylor // in: 2015 International Conference on Speech Technology and Human-Computer Dialogue (SpeD). — pp. 1–6. —2015.
65. Hahm S. J. Advanced Feature Normalization and Rapid Model Adaptation for Robust In-Vehicle Speech Recognition / S.J. Hahm, H. Boril, A. Pongtep, J.H.L. Hansen // Proc. In: 6th Biennial Workshop on Digital Signal Processing for In-Vehicle Systems. — pp. 14-17. — Seoul, Korea. — 2013.

66. Hermansky H. RASTA processing of Speech / H. Hermansky, N. Morgan // IEEE Transactions on speech and audio processing. — Vol.2. — No. 4. — P. 578 – 589. — 1994.

67. Hagen S. From modern standard Arabic to levantine ASR: Leveraging GALE for dialects / S. Hagen, M. Lidia, F. Biadsy // In Proc. IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). — pp. 266-271. — Waikoloa, HI, USA. — 2011.

68. Habash N. Y. Introduction to Arabic natural language processing. Synthesis Lectures on Human Language Technologies / N. Y. Habash // Synthesis Lectures on Human Language Technologies. —Pages: 1-187— Vol. 3. — Issue 1. — 2010.

69. Janet C. E. W. Aspects of the Phonology and Verb Morphology of three Yemeni Dialects / C. E. W. Janet // The School of Oriental and African Studies. — University of London. —pp.:478. —1989.

70. Katrin K. Novel approaches to Arabic speech recognition: report from the 2002 Johns Hopkins summer workshop / K. Katrin, J. Bilmes, S. Das, N. Duta, M. Egan, G. Ji, F. He, J. Henderson, D. Liu, M. Noamany, P. Schone, R. Schwartz // In Proc. ICASSP. — 2003.

71. Kees V. The Arabic Language / V. Kees // Linguistic Contacts between Arabic and Other Languages. New York: Columbia University Press. — 1997.

72. Kareem D. effective Arabic dialect identification /D. Kareem, H. Sajjad, H. Mubarak // In Proc. 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). — pages 1465–1468. — October 25-29. — Doha, Qatar — 2014.

73. Kathrein A.-K. A Lexical Distance Study of Arabic Dialects/ A.-K. Kathrein M. Saad, S. Chatzikyriakidis, S. Dobnik // The 4th International Conference on Arabic Computational Linguistics (ACLing 2018). — pp. 2-13. — 2018.
74. Kevin M. Hidden Markov Model (HMM) Toolbox for Matlab [Электронный ресурс] / M. Kevin. — 2005. — Режим доступа: https://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html.
75. Lawrence R Rabiner A tutorial on hidden Markov models and selected applications in speech recognition / R. Lawrence Rabiner // Proceedings of the IEEE. — Vol. 77. —Issue 2. — pages 257–286. — 1989.
76. Lamel L. Automatic speech-to-text transcription in Arabic / L. Lamel, A. Messaoudi G. JEAN-LUC, LIMSI-CNRS // ACM Trans Asian Lang Inform Process. —Vol. 8. —Issue 4. — pages 19. — 2009.
77. Li H. Spoken language recognition: From fundamentals to practice/ H. Li, B. Ma, K. Lee // Proceedings of the IEEE. — Vol. 101. — Issue 5. — pages 1136–1159. — 2013.
78. Leena L. Automatic Arabic Dialect Classification Using Deep Learning models / L. Leena, A. Elnagar // The 4th International Conference on Arabic Computational Linguistics (ACLing 2018). — pp. 262–269. — Dubai, United Arab Emirates. — November 17-19. — 2018.
79. Liu F. Efficient Cepstral Normalization for Robust Speech Recognition / F. Liu, R. Stern, X. Huang, A. Acero // Proc. ARPA Human Language Technology Workshop. — 1993. — P. 69–74.
80. Mohamed M. Developing and using a pilot dialectal Arabic treebank / M. Maamouri, A. Bies, T. Buckwalter, M. Diab, N. Habash, O. Rambow, D. Tabessi // In Proc. LREC. — pp.1-6. — 2006.

81. Mel Frequency Cepstral Coefficient (MFCC) tutorial – Practical cryptography [Электронный ресурс]. – Режим доступа: http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/
82. Nour-Eddine L. GMM-Based Maghreb Dialect Identification System / L. Nour-Eddine, A. Adla // Journal of Information Processing Systems. — Vol. 11. — Issue 1. — pages 613–641. — 2017.
83. Nadia H. F. Echo State Networks for Arabic Phoneme Recognition / H. Nadia, T. Allen // World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering. —Vol.7. — No.:7. — 2013.
84. Nadia H. F. Deep Neural Network Acoustic models for Multi-dialect Arabic Speech Recognition / H. F. Nadia // Graduate Department of School of Science and Technology Nottingham Trent University. — pages 154. — England. — 2015.
85. Najafian M. Automatic speech recognition of arabic multi-genre broadcast media / M. Najafian, W.-N. Hsu, A. Ali, J. Glass // In Proc. IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). — Okinawa, Japan — 2017.
86. Nizar H. Issues in Arabic morphological analysis / H. Nizar, S. Abdelhadi, B. Timothy // Arabic computational morphology, Springer. — pages 23–41. — 2007.
87. Nada M. A. S. An acoustic investigation of the rhythm of yemeni arabic and jordanian Arabic / M. A. S. Nada // faculty of languages and linguistics university of malaya kuala lumpur. —pp.:155. —2014.

88. Omar F. Zaidan Arabic dialect identification / F. Zaidan Omar C.-B. Chris C// Computational Linguistics. — Vol. 40. — Issue 1. — pages: 171–202. — 2014.
89. Plapous C. Improved Signal-to-Noise Ratio Estimation for Speech Enhancement / C. Plapous, C. Marro, P. Scalart // Ieee transactions on audio, speech, and language processing. —Pages: 2098 – 2108. — vol. 14. — no. 6. — 2006.
90. Prasad N. V. Improved cepstral mean and variance normalization using Bayesian framework / N. V. Prasad, S. Umesh // 2013 IEEE Workshop on Automatic Speech Recognition and Understanding. Olomouc, Czech Republic. — 8-12 Dec. — 2013.
91. Rao K. S. Language Identification Using Spectral and Prosodic Features/ K. S. Rao, V. R. Reddy, S. Maity // Springer. — 2015.
92. Rong T. Integrating acoustic, prosodic and phonotactic features for spoken language identification / T. Rong, M. Bin, Z. Donglai, L. Haizhou, C. E. // in: 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings. — pages 205-208. —Toulouse, France. — 2006.
93. Sawalha M. Automatically generated, phonemic Arabic-IPA pronunciation tiers for the Boundary AnnotatedQur'an Dataset for Machine Learning (version 2.0) / M. Sawalha, Claire Brierley, Eric Atwell // Computer Information Systems. — LREC 2014 post-conference workshop. Pp. 1-8. — 2014.
94. Scalart, P. Speech enhancement based on a priori signal to noise estimation / P. Scalart, J.V. Filho // 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings. – IEEE, 1996. – Vol. 2. – P. 629–632.

95. Satori H. Introduction to Arabic Speech Recognition Using CMUSphinx System / H. Satori, M. Harti, N. Chenfour. — pages 4. — 2007
96. Versteegh K. The Arabic Language / K. Versteegh // Cambridge University Press. —1997.
97. Vikas J. Modified Cepstral Mean Normalization - Transforming to utterance specific non-zero mean / V. Joshi, N. V. Prasad, S. Umesh // Interspeech 2013. — Lyon, France. — August 25-29, 2013.
98. Vikas J. Modified Mean and Variance Normalization:Transforming to Utterance-Specific Estimates / V. Joshi, N. V. Prasad, S. Umesh // Department of Electrical Engineering, Indian Institute of Technology. — Vol. 35. — Issue 5. — May 2016. — Pages 1593-1609.
99. Yeou M. F0 alignment patterns in arabic dialects / M. Yeou, M. Embarki, S. ALMaqtari, C. Dodane // Proceedings of the 16th International Congress of Phonetic Sciences ICPhS XVI. — Saarbrücken, Germany. — 6-10 August 2007.
100. Yemeni Arabic [Электронный ресурс] / режим доступа: https://en.wikipedia.org/wiki/Yemeni_Arabic дата обращения: октябрь 2017.

Приложение 1. Документы, подтверждающие внедрение основных результатов диссертационной работы

АКТ внедрения ВлГУ

УТВЕРЖДАЮ

Проректор по научной и инновационной работе
Владимирского государственного университета имени
Александра Григорьевича и Николая Григорьевича
Столетовых» (ВлГУ) Федин А. В.

23 09 2019г.

АКТ ВНЕДРЕНИЯ

результатов диссертационной работы
«Исследование методов и разработка алгоритмов обработки сигналов для систем автоматического распознавания телефонной речи в республике Йемен»

Мы, нижеподписавшиеся, заведующий кафедрой радиотехники и радиосистем д.т.н., профессор Никитин О.Р., д.т.н., профессор Полушин П.А, заведующая лабораториями Королева О.В. составили настоящий акт в том, что результаты диссертационной работы Аль-Дайбани Абдулгани Мохаммед Салех внедрены в учебный процесс на кафедре радиотехники и радиосистем Владимирского государственного университета имени Александра Григорьевича и Николая Григорьевича Столетовых (ВлГУ).

Материалы диссертации используются в лекционном курсе и при выполнении лабораторных работ в рамках учебных дисциплин «Устройства приема и обработки сигналов», и «Обработка сигналов», предназначенных, соответственно, для магистрантов и студентов направления «Радиотехника»

Зав. кафедрой радиотехники
и радиосистем д.т.н., профессор Никитин О.Р. 23.09.19
Д.т.н., профессор Полушин П.А. 23.09.19
Заведующая лабораториями Королева О.В. 23.09.19

АКТ внедрения ЦРТ

ЦРТ
Общество с ограниченной ответственностью
«Центр речевых технологий»
Санкт-Петербург, ул. Красуцкого, д. 4
тел. (812) 325-88-48, факс (812) 327-92-97
почтовый адрес: а/я 124, Санкт-Петербург, 196084
e-mail: stc-spb@speechpro.com
http://www.speechpro.ru
ОКПО 20502206, ОГРН 1027810243295
ИНН/КПП 7805093681/783901001

УТВЕРЖДАЮ
Административный директор
«Центр речевых технологий»,

Е.Ю. Марусов

"__" сентября 2019 г.

АКТ

о внедрении результатов диссертационной работы Аль-Дайбани Абдулгани Мохаммед Салеха «Исследование методов и разработка алгоритмов обработки сигналов для систем автоматического распознавания телефонной речи в республике Йемен»

Методика оценки точности работы идентификатора диалектов, предложенная Аль-Дайбани Абдулгани Мохаммед Салехом, была использована при подготовке обучающих данных системы распознавания арабской речи.

Исключительные права на разработанную на ее основе программную систему принадлежат обществу с ограниченной ответственностью «Центр речевых технологий».

Директор научно-
исследовательского департамента
ООО ЦРТ,
кандидат технических наук

К.Е. Левин

Приложение 2. Свидетельства о регистрации программ для ЭВМ

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2019616903

Анализ влияния оконной функции на результат нормализации параметров речевого сигнала

Правообладатель: *Федеральное государственное бюджетной образовательной учреждение высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых» (RU)*

Авторы: *Аль-Дайбани Абдулгани Мохаммед Салех (YE), Левин Евгений Калманович (RU)*

Заявка № 2019615187
Дата поступления 07 мая 2019 г.
Дата государственной регистрации в Реестре программ для ЭВМ *30 мая 2019 г.*

Руководитель Федеральной службы по интеллектуальной собственности Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2019616650

Комплексная программа анализа влияния оконной функции на результаты нормализации параметров речевого сигнала

Правообладатель: *Федеральное государственное бюджетное образовательное учреждение высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых» (RU)*

Авторы: *Аль-Дайбани Абдулгани Мохаммед Салех (YE), Левин Евгений Калманович (RU)*

Заявка № 2019615821
Дата поступления 21 мая 2019 г.
Дата государственной регистрации в Реестре программ для ЭВМ *27 мая 2019 г.*

Руководитель Федеральной службы по интеллектуальной собственности Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2019619052

Оценка эффективности применения спектрального вычитания и фильтра Винера для подавления помех при автоматическом распознавании речи

Правообладатель: *Федеральное государственное бюджетное образовательное учреждение высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых» (RU)*

Авторы: *Аль-Дайбани Абдулгани Мохаммед Салех (YE), Левин Евгений Калманович (RU)*

Заявка № 2019617542
Дата поступления 24 июня 2019 г.
Дата государственной регистрации в Реестре программ для ЭВМ *09 июля 2019 г.*

Руководитель Федеральной службы по интеллектуальной собственности Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2019662082

Оценка эффективности использования нормализованных параметров речевого сигнала в телефонных системах автоматического распознавания речи

Правообладатель: *Федеральное государственное бюджетное образовательное учреждение высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых» (RU)*

Авторы: *Аль-Дайбани Абдулгани Мохаммед Салех (YE), Левин Евгений Калманович (RU)*

Заявка № 2019660724
Дата поступления 30 августа 2019 г.
Дата государственной регистрации в Реестре программ для ЭВМ *16 сентября 2019 г.*

Руководитель Федеральной службы по интеллектуальной собственности Г.П. Ивлиев

Приложение 3. Сертификат участия в конференции IEEE-2019. Диплом за лучший доклад на конференции ФРЭМЭ-2018

Ural Federal University
named after the first President of Russia B.N.Yeltsin
Engineering School of Information Technologies, Telecommunications and Control Systems

USBEREIT

2019 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT)

CERTIFICATE OF PARTICIPATION

This certificate is awarded to

Abdulgani Mohammed Saleh Al-Daibani

For attending the 2019 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT)

Ilya Obabkov,
Director of Engineering School of Information Technologies, Telecommunications and Control Systems

Yekaterinburg, Russia
2019

XIII МЕЖДУНАРОДНАЯ НАУЧНАЯ КОНФЕРЕНЦИЯ «ФИЗИКА И РАДИОЭЛЕКТРОНИКА В МЕДИЦИНЕ И ЭКОЛОГИИ» с научной молодежной школой имени И.Н.Спиридонова ФРЭМЭ'2018

ФРЭМЭ

3 – 5 июля 2018 г.

ДИПЛОМ

ЗА ЛУЧШИЙ ДОКЛАД НАГРАЖДАЕТСЯ

Аль-Дайбани
Абдулгани Мохамед Салех

Доклад рекомендуется к публикации в профильном журнале

Председатель программного комитета, академик РАН, д.ф.-м.н., профессор Ю.В. Гуляев

Владимир – Суздаль, 2018 г.

Приложение 4. Результаты проведения эксперимента классификации диалектов на три группы

Результаты классификации с использованием команды «0»

Таблица П4.1. Одинаковые дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южный диалект, %	Западный диалект, %
Северный диалект	100	0	0
Южный диалект	0	100	
Западный диалект	0	0	100

Таблица П4.2.Разные дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южный диалект, %	Западный диалект, %
Северный диалект	16	0	0
Южный диалект	84	84	100
Западный диалект	0	16	0

Результаты классификации с использованием команды «1»

Таблица П4.3.Одинаковые дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южный диалект, %	Западный диалект, %
Северный диалект	99.33	0	0
Южный диалект	0	94.66	0
Западный диалект	0.66	5.33	100

Таблица П4.4.Разные дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южный диалект, %	Западный диалект, %
Северный диалект	100	0	0
Южный диалект	0	36	36
Западный диалект	0	64	64

Результаты классификации с использованием команды «2»

Таблица П4.5. Одинаковые дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южный диалект, %	Западный диалект, %
Северный диалект	100	0	0
Южный диалект	0	100	4.66
Западный диалект	0	0	95.33

Таблица П4.6. Разные дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южный диалект, %	Западный диалект, %
Северный диалект	100	0	20
Южный диалект	0	20	80
Западный диалект	0	80	0

Результаты классификации с использованием команды «3»

Таблица П4.7. Одинаковые дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южный диалект, %	Западный диалект, %
Северный диалект	97.33	0	0
Южный диалект	1.33	100	3.67
Западный диалект	1.33	0	96.33

Таблица П4.8. Разные дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южный диалект, %	Западный диалект, %
Северный диалект	100	0	0
Южный диалект	0	100	100
Западный диалект	0	0	0

Результаты классификации с использованием команды «4»

Таблица П4.9. Одинаковые дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южный диалект, %	Западный диалект, %
Северный диалект	100	0	0
Южный диалект	0	100	1.66
Западный диалект	0	0	99.33

Таблица П4.10. Разные дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южный диалект, %	Западный диалект, %
Северный диалект	80	4	0
Южный диалект	20	96	0
Западный диалект	0	0	100

Результаты классификации с использованием команды «5»

Таблица П4.11. Одинаковые дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южный диалект, %	Западный диалект, %
Северный диалект	100	0.66	0
Южный диалект	0	99.33	0
Западный диалект	0	0	100

Таблица П4.12. Разные дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южный диалект, %	Западный диалект, %
Северный диалект	92	0	100
Южный диалект	8	92	0
Западный диалект	0	8	0

Результаты классификации с использованием команды «6»

Таблица П4.13. Одинаковые дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южный диалект, %	Западный диалект, %
Северный диалект	100	0	0
Южный диалект	0	100	4
Западный диалект	0	0	96

Таблица П4.14. Разные дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южный диалект, %	Западный диалект, %
Северный диалект	92	0	20
Южный диалект	0	92	0
Западный диалект	8	8	80

Результаты классификации с использованием команды «7»

Таблица П4.15. Одинаковые дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южный диалект, %	Западный диалект, %
Северный диалект	94.33	0	0
Южный диалект	5.66	100	0
Западный диалект	0	0	100

Таблица П4.16. Разные дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южный диалект, %	Западный диалект, %
Северный диалект	0	0	0
Южный диалект	100	100	0
Западный диалект	0	0	100

Результаты классификации с использованием команды «8»

Таблица П4.17. Одинаковые дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южный диалект, %	Западный диалект, %
Северный диалект	100	0	0
Южный диалект	0	100	0
Западный диалект	0	0	100

Таблица П4.18. Разные дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южный диалект, %	Западный диалект, %
Северный диалект	100	0	60
Южный диалект	0	100	0
Западный диалект	0	0	40

Результаты классификации с использованием команды «9» как ключевое слово по всем диалектам (6 дикторов):

Таблица П4.19. Одинаковые дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южный диалект, %	Западный диалект, %
Северный диалект	100	0	0
Южный диалект	0	100	0
Западный диалект	0	0	100

Таблица П4.20. Разные дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южный диалект, %	Западный диалект, %
Северный диалект	88		4
Южный диалект	8	100	0
Западный диалект	4		96

Приложение 5. Результаты проведения эксперимента классификации диалектов на две групп

Результаты классификации с использованием команды «0»

Таблица П5.1. Одинаковые дикторы с разными произнесениями

Название диалекта	Северно-южный диалект, %	Западный диалект, %
Северно-южный диалект	100	24
Западный диалект	0	76

Таблица П5.2. Разные дикторы с разными произнесениями

Название диалекта	Северно-южный диалект, %	Западный диалект, %
Северно-южный диалект	100	100
Западный диалект	0	0

Результаты классификации с использованием команды «1»

Таблица П5.3. Одинаковые дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южно-западный диалект, %
Северный диалект	100	0
Южно-западный диалект	0	100

Таблица П5.4. Разные дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южно-западный диалект, %
Северный диалект	20	0
Южно-западный диалект	80	100

Результаты классификации с использованием команды «2»

Таблица П5.5. Одинаковые дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южно-западный диалект, %
Северный диалект	100	0
Южно-западный диалект	0	100

Таблица П5.6. Разные дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южно-западный диалект, %
Северный диалект	98	0
Южно-западный диалект	2	100

Результаты классификации с использованием команды «3»

Таблица П5.7. Одинаковые дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южно-западный диалект, %
Северный диалект	99.333	0
Южно-западный диалект	0.666	100

Таблица П5.8. Разные дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южно-западный диалект, %
Северный диалект	96	0
Южно-западный диалект	4	100

Результаты классификации с использованием команды «4»

Таблица П5.9. Одинаковые дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южно-западный диалект, %
Северно-южныйДиалект	100	0
Западный диалект	0	100

Таблица П5.10. Разные дикторы с разными произнесениями

Название диалекта	Северный диалект, %	Южно-западный диалект, %
Северно-южныйДиалект	100	12
Западный диалект	0	88

Результаты классификации с использованием команды «5»

Таблица П5.11. Одинаковые дикторы с разными произнесениями

Название диалекта	Северно-западный диалект, %	Южный диалект, %
Северно-западный диалект	99	2
Южный диалект	1	98

Таблица П5.12. Разные дикторы с разными произнесениями

Название диалекта	Северно-западный диалект, %	Южный диалект, %
Северно-западный диалект	90	4
Южный диалект	10	96

Результаты классификации с использованием команды «6»

Таблица П5.13. Одинаковые дикторы с разными произнесениями

Название диалекта	Северно-западный диалект, %	Южный диалект, %
Северно-западный диалект	100	0
Южный диалект	0	100

Таблица П5.14. Разные дикторы с разными произнесениями

Название диалекта	Северно-западный диалект, %	Южный диалект, %
Северно-западный диалект	100	20
Южный диалект	0	80

Результаты классификации с использованием команды «7»

Таблица П5.15. Одинаковые дикторы с разными произнесениями

Название диалекта	Северно-южный диалект, %	Западный диалект, %
Северно-южный диалект	100	0
Западный диалект	0	100

Таблица П5.16. Разные дикторы с разными произнесениями

Название диалекта	Северно-южный диалект, %	Западный диалект, %
Северно-южный диалект	100	8
Западный диалект	0	92

Результаты классификации с использованием команды «8»

Таблица П5.17. Одинаковые дикторы с разными произнесениями

Название диалекта	Северно-западный диалект, %	Южный диалект, %
Северно-западный диалект	100	0
Южный диалект	0	100

Таблица П5.18. Разные дикторы с разными произнесениями

Название диалекта	Северно-западный диалект, %	Южный диалект, %
Северно-западный диалект	100	0
Южный диалект	0	100

Результаты классификации с использованием команды «9»

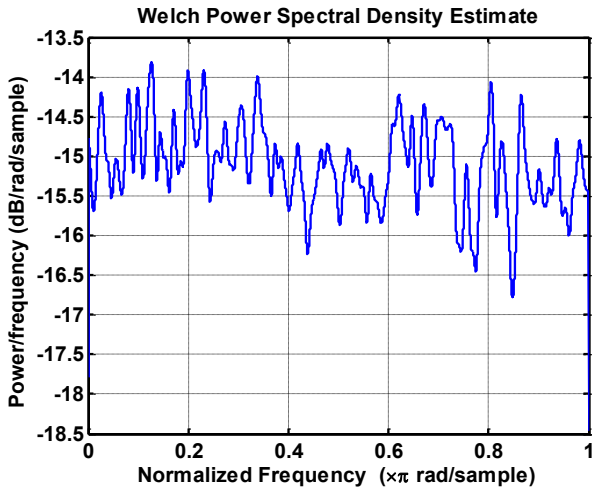
Таблица П5.19. Одинаковые дикторы с разными произнесениями

Название диалекта	Северно-южный диалект, %	Западный диалект, %
Северно-южный диалект	100	0
Западный диалект	0	100

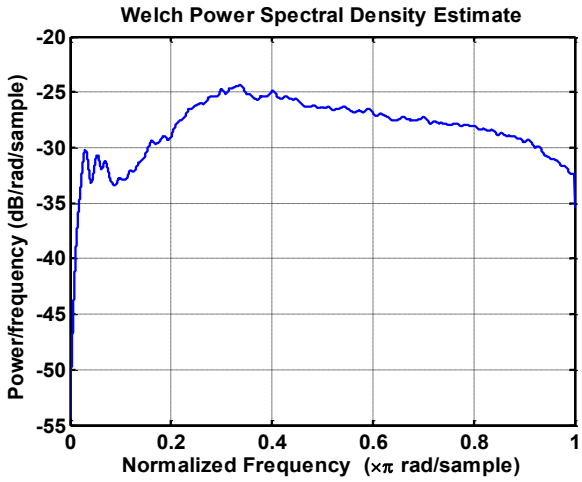
Таблица П5.20. Разные дикторы с разными произнесениями

Название диалекта	Северно-южный диалект, %	Западный диалект, %
Северно-южный диалект	100	0
Западный диалект	0	100

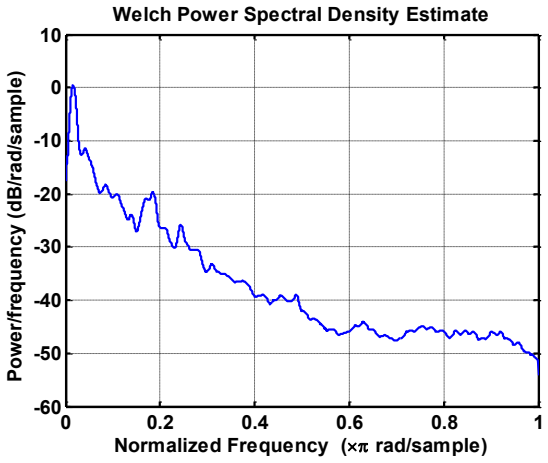
Приложение 6. Характеристики исследуемых шумов и результаты влияния их на достоверность системы распознавания



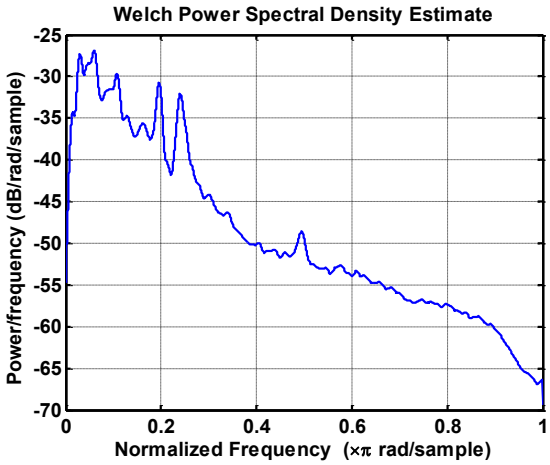
Белый шум



Шум дождя

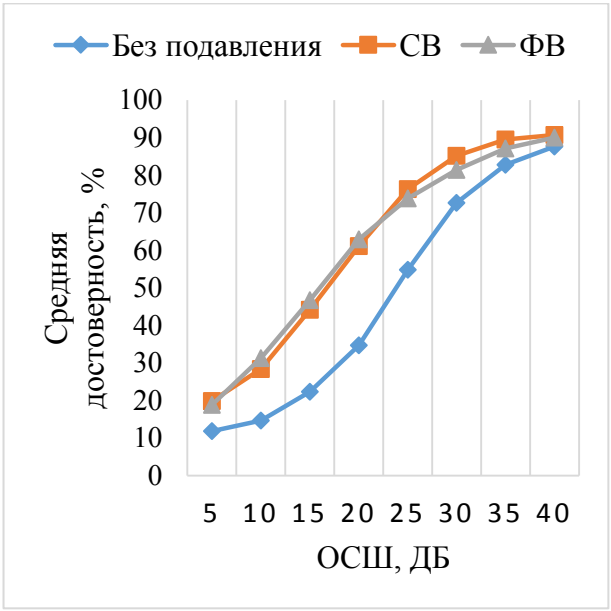


Шум автобуса

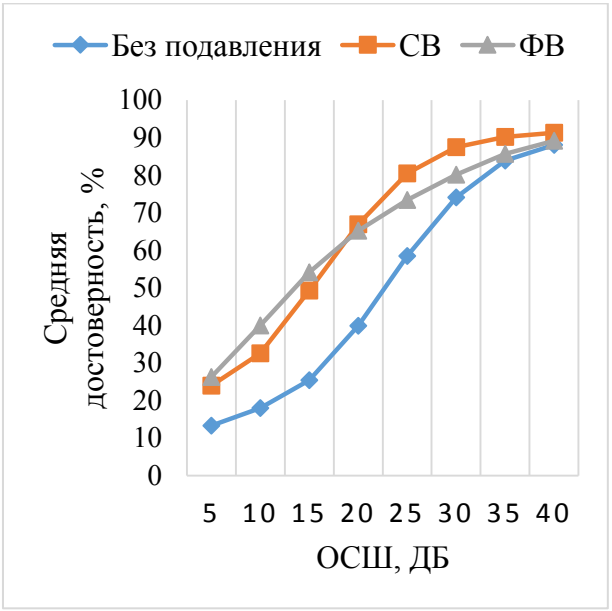


Шум офиса

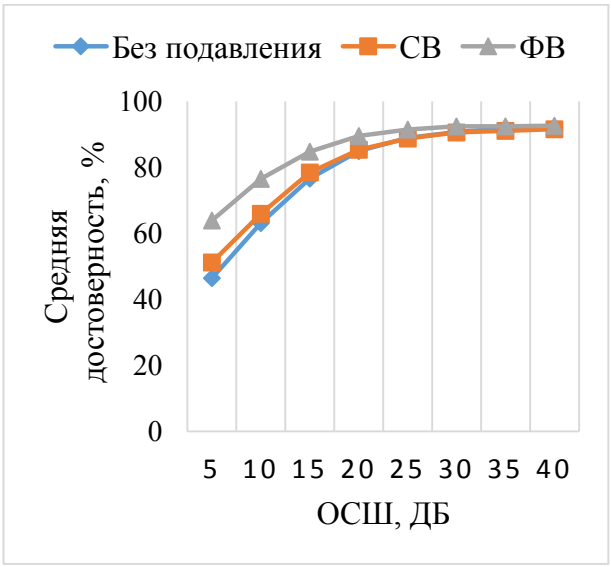
Рисунок П6.1. Спектральная плотность мощности исследуемых шумов



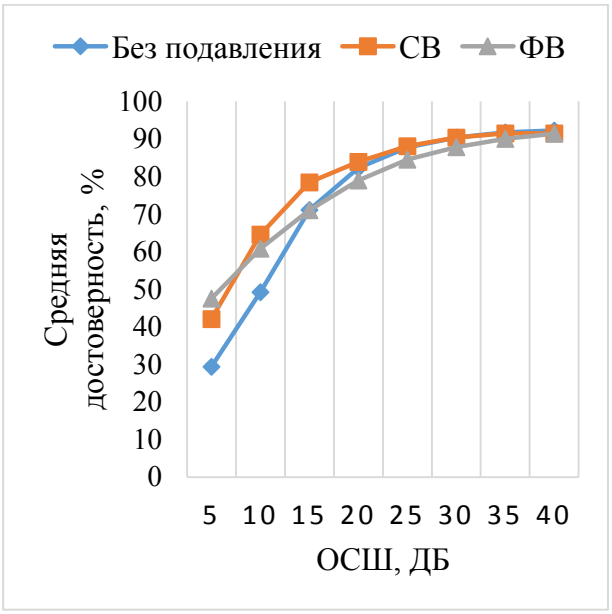
(а)



(б)

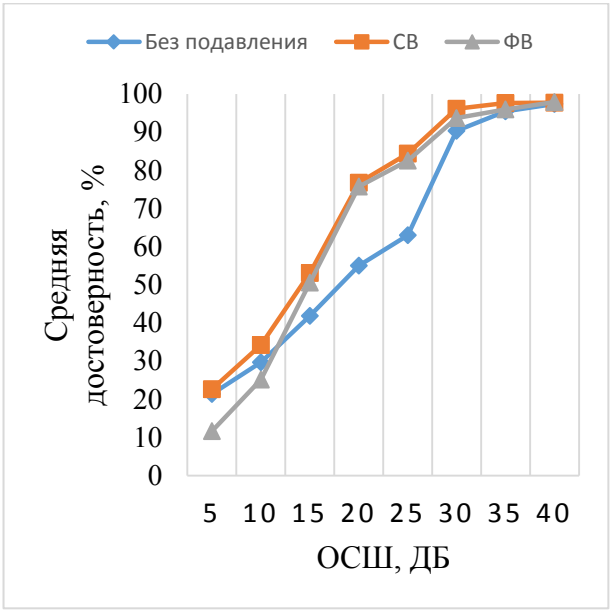


(в)

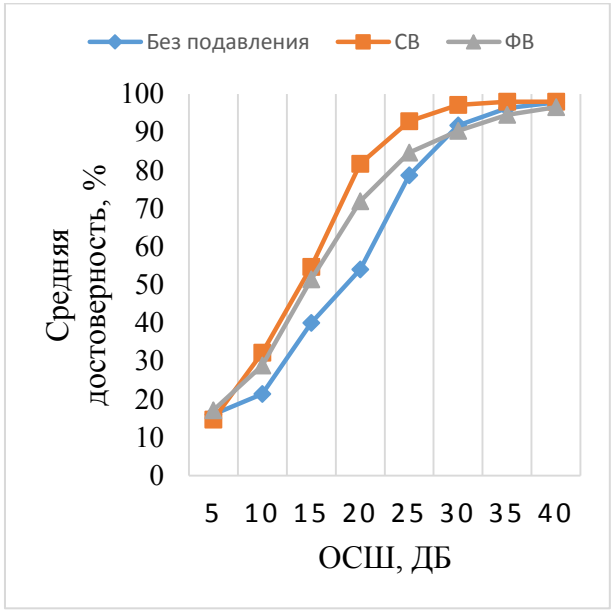


(г)

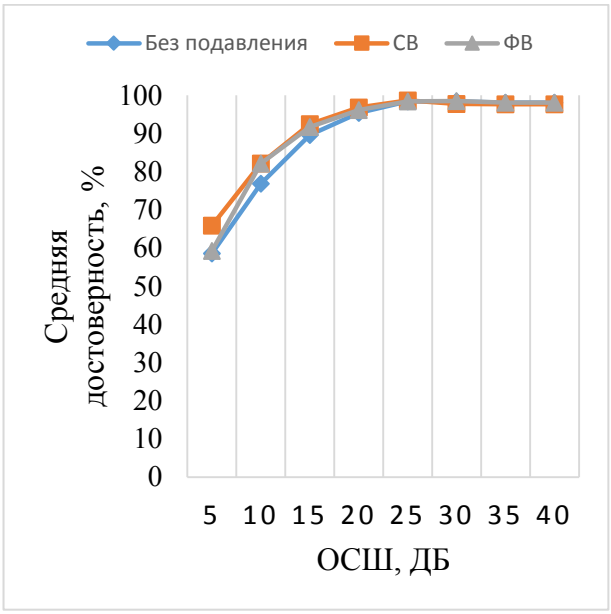
Рисунок П6.2. Средняя достоверность системы для группы диалектов (ВсеД): (а) при присутствии белого шума, (б) при присутствии реального шума (шум дождя), (в) при присутствии реального шума (шум автобуса), (г) при присутствии реального шума (шум офиса)



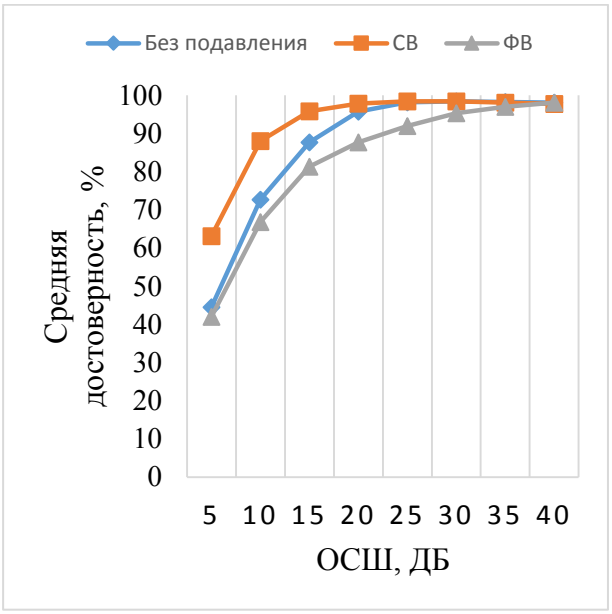
(а)



(б)

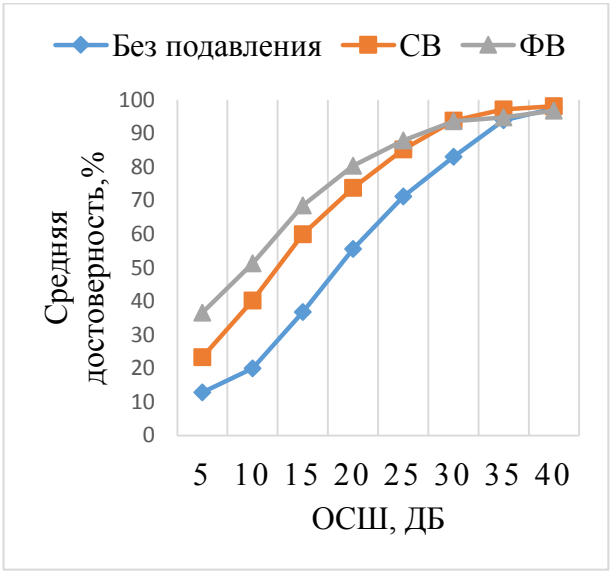


(в)

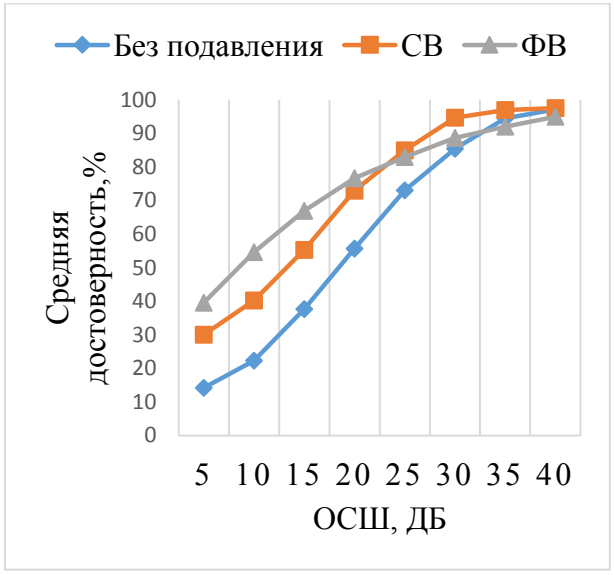


(г)

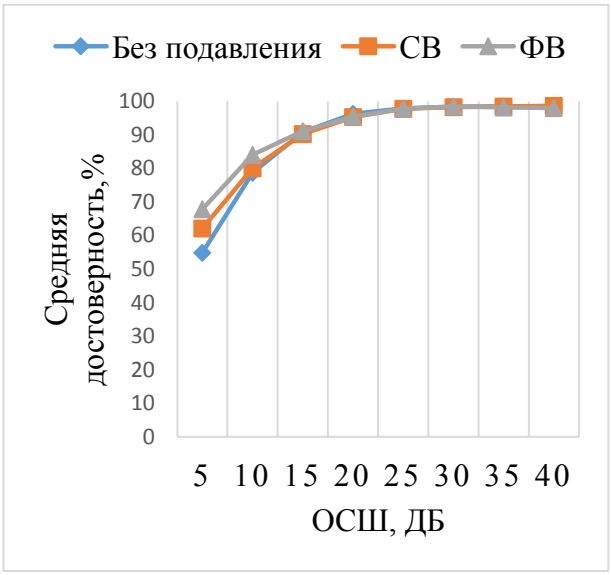
Рисунок П6.3. Средняя достоверность системы для северного диалекта: (а) при присутствии белого шума, (б) при присутствии реального шума (шум дождя), (в) при присутствии реального шума (шум автобуса), (г) при присутствии реального шума (шум офиса)



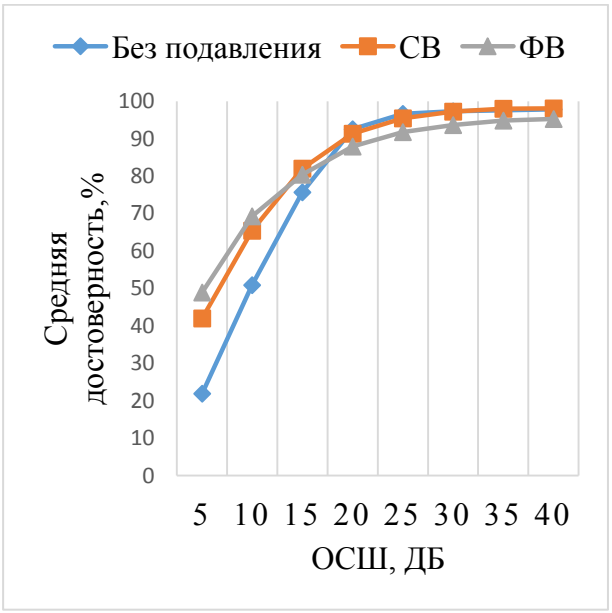
(а)



(б)



(в)



(г)

Рисунок П6.4.Средняя достоверность системы для южного диалекта: (а) при присутствии белого шума, (б) при присутствии реального шума (шум дождя), (в) при присутствии реального шума (шум автобуса), (г) при присутствии реального шума (шум офиса)

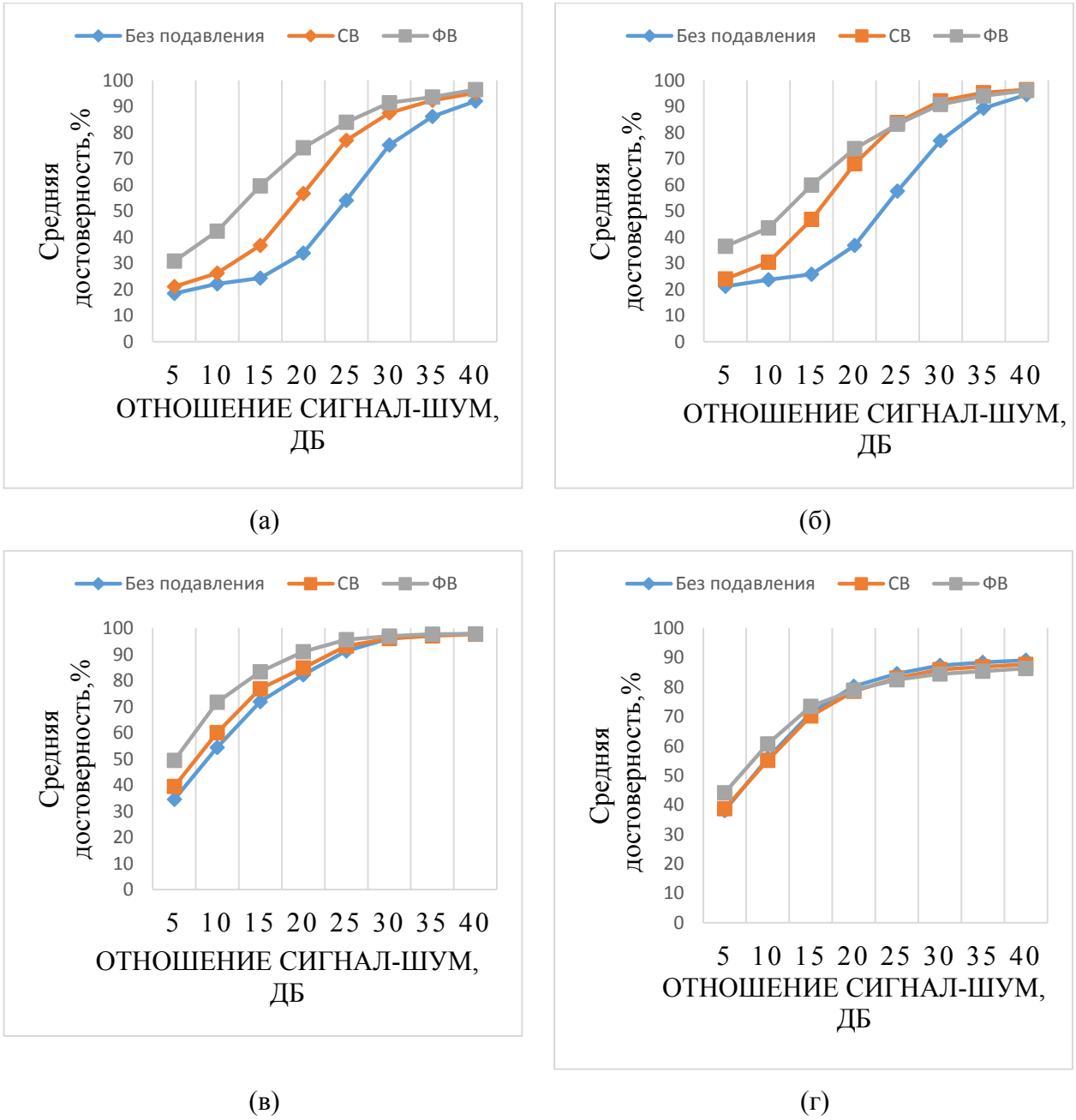


Рисунок П6.5.Средняя достоверность системы для западного диалекта: (а) при присутствии белого шума, (б) при присутствии реального шума (шум дождя), (в) при присутствии реального шума (шум автобуса), (г) при присутствии реального шума (шум офиса)

Приложения 7. Результат тестирования САРР

Таблица П7.1 Результат тестирования идентификации в случае присутствии в САРР северного диалекта

	Матрицы распознавания									
	Обучение СД и тестирование СД									
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0	100,00	0,00	0,00	0,00	0,00	0,00	0,00	6,67	0,00	0,00
1	0,00	100,00	2,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2	0,00	0,00	97,33	0,00	0,00	6,00	0,00	0,00	0,00	0,00
3	0,00	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00
4	0,00	0,00	0,00	0,00	99,33	0,00	0,00	0,00	0,00	0,00
5	0,00	0,00	0,00	0,00	0,00	92,00	0,00	0,00	0,00	0,00
6	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00	0,00	0,00
7	0,00	0,00	0,00	0,00	0,67	0,00	0,00	92,67	0,00	0,00
8	0,00	0,00	0,00	0,00	0,00	2,00	0,00	0,00	100,00	0,00
9	0,00	0,00	0,67	0,00	0,00	0,00	0,00	0,67	0,00	100,00
	Обучение СД и тестирование ЮД									
0	34,67	0,00	0,00	11,33	0,00	0,00	4,67	21,33	0,00	16,67
1	0,00	21,33	0,00	1,33	8,00	0,00	0,00	0,00	0,00	0,00
2	0,00	0,00	58,00	4,00	0,00	0,00	19,33	0,00	1,33	2,00
3	14,00	14,00	2,00	38,00	0,67	12,67	1,33	27,33	16,00	0,00
4	0,67	2,67	10,67	0,67	70,00	0,00	4,00	1,33	0,00	0,00
5	49,33	62,00	19,33	41,33	20,67	87,33	54,00	48,00	82,67	79,33
6	0,00	0,00	10,00	0,00	0,00	0,00	16,67	0,00	0,00	1,33
7	1,33	0,00	0,00	0,67	0,00	0,00	0,00	2,00	0,00	0,00
8	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
9	0,00	0,00	0,00	2,67	0,67	0,00	0,00	0,00	0,00	0,67
	Обучение СД и тестирование ЗД									
0	40,00	0,00	0,00	12,67	0,00	0,00	0,00	11,33	0,00	0,00
1	3,33	29,33	0,00	1,33	0,00	0,00	8,67	0,00	0,67	0,00
2	0,00	0,00	27,33	0,00	0,00	0,00	7,33	2,67	2,00	20,67
3	0,00	6,67	0,00	7,33	14,67	0,00	0,67	0,00	0,00	0,00
4	12,00	0,67	14,67	12,00	53,33	13,33	6,00	0,00	3,33	0,00
5	39,33	63,33	45,33	61,33	29,33	84,67	24,00	85,33	80,67	18,00
6	5,33	0,00	5,33	0,00	0,00	2,00	38,67	0,00	3,33	14,67
7	0,00	0,00	0,00	5,33	2,67	0,00	0,00	0,67	0,00	0,00
8	0,00	0,00	6,00	0,00	0,00	0,00	0,00	0,00	10,00	0,00
9	0,00	0,00	1,33	0,00	0,00	0,00	14,67	0,00	0,00	46,67

Таблица П7.2 Результат тестирования идентификации в случае присутствии в САРР южного диалекта

	Матрицы распознавания									
	Обучение ЮД и тестирование ЮД									
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0	98,67	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,67
1	0,67	100,00	0,00	0,00	2,67	0,00	0,00	0,00	0,00	0,00
2	0,00	0,00	96,67	0,00	0,00	0,00	0,67	0,00	0,00	0,00
3	0,00	0,00	0,00	96,67	0,00	0,00	2,67	0,67	0,00	0,00
4	0,00	0,00	0,00	0,00	96,67	0,00	0,00	0,00	0,00	0,00
5	0,67	0,00	0,00	0,67	0,00	100,00	0,00	0,00	0,00	0,00
6	0,00	0,00	0,00	0,00	0,00	0,00	95,33	0,00	0,00	0,67
7	0,00	0,00	0,00	2,67	0,67	0,00	0,00	99,33	0,00	0,00
8	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00
9	0,00	0,00	3,33	0,00	0,00	0,00	1,33	0,00	0,00	98,67
	Обучение ЮД и тестирование СД									
0	28,00	0,00	15,33	2,00	0,67	0,67	0,00	2,67	8,00	0,00
1	2,00	35,33	2,00	22,67	18,00	0,67	2,00	2,00	0,00	19,33
2	0,00	2,00	54,67	0,00	6,00	6,67	8,00	0,00	18,67	1,33
3	0,00	10,67	0,00	35,33	4,00	0,67	0,00	0,00	0,00	0,00
4	0,00	20,67	0,00	0,00	44,67	2,00	0,00	3,33	16,67	12,67
5	21,33	19,33	8,67	10,00	7,33	72,67	16,00	19,33	7,33	32,67
6	24,00	1,33	2,00	4,00	0,00	0,00	47,33	3,33	0,00	2,67
7	8,67	9,33	2,00	10,67	14,67	9,33	0,00	62,67	1,33	11,33
8	0,00	0,00	1,33	4,00	4,67	0,00	10,67	2,67	33,33	0,00
9	16,00	1,33	14,00	11,33	0,00	7,33	16,00	4,00	14,67	20,00
	Обучение ЮД и тестирование ЗД									
0	52,67	2,00	0,00	15,33	4,00	0,67	3,33	6,67	0,00	0,00
1	3,33	52,00	8,67	2,00	13,33	10,67	0,00	6,00	2,67	0,00
2	1,33	2,00	67,33	0,00	0,00	2,67	16,67	2,00	56,00	26,67
3	8,67	1,33	0,00	26,00	18,00	0,00	0,00	9,33	0,00	0,00
4	0,00	6,67	0,67	6,67	32,00	2,00	0,00	0,00	0,67	0,00
5	0,00	6,67	0,00	22,00	0,67	46,67	0,00	0,00	0,00	0,00
6	4,67	0,00	0,67	0,67	0,00	2,67	54,00	17,33	0,00	10,67
7	16,00	0,67	0,00	19,33	16,00	0,00	0,00	2,00	0,00	0,00
8	0,00	0,67	0,00	2,67	9,33	0,00	0,00	0,67	22,67	0,00
9	13,33	28,00	22,67	5,33	6,67	34,67	26,00	56,00	18,00	62,67

Таблица П7.3 Результат тестирования идентификации в случае присутствии в САРР западного диалекта

	Матрицы распознавания									
	Обучение ЗД и тестирование ЗД									
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0	100,00	0,00	0,00	0,00	0,00	0,00	0,00	1,60	0,00	0,00
1	0,00	100,00	0,00	0,80	0,00	0,00	0,00	0,00	0,00	0,00
2	0,00	0,00	100,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
3	0,00	0,00	0,00	96,00	0,00	0,00	0,00	0,00	0,00	0,00
4	0,00	0,00	0,00	0,80	98,40	0,00	0,00	0,00	0,00	0,00
5	0,00	0,00	0,00	1,60	0,80	98,40	0,00	0,00	0,80	0,00
6	0,00	0,00	0,00	0,00	0,00	0,00	97,60	0,00	0,00	0,00
7	0,00	0,00	0,00	0,00	0,00	1,60	0,00	96,00	0,80	1,60
8	0,00	0,00	0,00	0,80	0,80	0,00	0,00	0,00	98,40	0,00
9	0,00	0,00	0,00	0,00	0,00	0,00	2,40	2,40	0,00	98,40
	Обучение ЗД и тестирование СД									
0	69,60	0,00	0,00	4,00	5,60	0,00	12,80	58,40	2,40	36,80
1	0,00	53,60	0,00	33,60	15,20	0,00	0,00	0,00	0,00	0,00
2	0,80	31,20	32,00	16,80	0,00	14,40	1,60	2,40	40,00	4,80
3	0,00	0,00	0,00	18,40	21,60	0,00	0,80	22,40	0,00	7,20
4	1,60	0,00	0,00	1,60	16,00	0,00	0,00	0,00	0,00	0,00
5	16,80	12,80	32,80	0,80	41,60	74,40	20,00	16,00	0,00	50,40
6	0,00	0,00	3,20	0,00	0,00	0,00	21,60	0,00	0,00	0,00
7	9,60	0,00	0,00	6,40	0,00	1,60	0,00	0,00	0,00	0,00
8	0,00	2,40	12,00	18,40	0,00	5,60	0,00	0,80	57,60	0,00
9	1,60	0,00	20,00	0,00	0,00	4,00	43,20	0,00	0,00	0,80
	Обучение ЗД и тестирование ЗД									
0	44,80	8,80	0,00	9,60	4,00	0,00	0,80	57,60	0,00	17,60
1	0,00	47,20	0,80	19,20	2,40	0,00	0,00	0,00	0,80	0,00
2	0,00	4,80	32,80	0,00	0,80	0,00	0,00	0,00	27,20	3,20
3	2,40	3,20	3,20	26,40	16,00	0,80	12,00	0,00	0,00	0,00
4	18,40	12,00	0,80	34,40	17,60	3,20	12,00	8,80	0,00	0,80
5	17,60	8,80	9,60	4,00	41,60	83,20	24,80	19,20	27,20	26,40
6	7,20	0,00	2,40	0,00	0,00	0,00	33,60	0,00	0,00	6,40
7	2,40	0,00	1,60	0,00	0,00	0,80	0,80	0,80	0,00	28,00
8	2,40	15,20	41,60	6,40	17,60	12,00	4,80	13,60	44,80	2,40
9	4,80	0,00	7,20	0,00	0,00	0,00	11,20	0,00	0,00	15,20