Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»

На правах рукописи

all

Левин Евгений Калманович

РАЗРАБОТКА СРЕДСТВ ИССЛЕДОВАНИЯ И ПОВЫШЕНИЯ ПОМЕХОУСТОЙЧИВОСТИ СИСТЕМ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ ГОЛОСОВЫХ КОМАНД В ТЕЛЕФОНИИ

05.12.13 - Системы, сети и устройства телекоммуникаций

Диссертация

на соискание ученой степени доктора технических наук

Научный консультант – доктор технических наук, профессор Никитин О.Р.

Содержание

ВВЕДЕНИЕ	5
1. МЕТОДЫ ОБЕСПЕЧЕНИЯ ПОМЕХОУСТОЙЧИВОСТИ РАСПОЗНАВАНИЯ	
ГОЛОСОВЫХ КОМАНД	18
1.1. Основные этапы обработки речевого сигнала при распознавании	18
1.1.1. Предварительная обработка сигнала	18
1.1.2. Сопоставление параметров сигнала с акустическими моделями	21
1.2. Подавление помех при предварительной обработке речевого	
сигнала	28
1.3. Подавление помех с использованием акустических моделей	35
1.4. Оценка достоверности распознавания голосовых команд	51
Выводы по разделу 1	55
2. СОКРАЩЕНИЕ ОБЪЕМА ЭКСПЕРИМЕНТОВ ПРИ ОПРЕДЕЛЕНИИ ПАРАМЕТРО)B
ДОСТОВЕРНОСТИ РАСПОЗНАВАНИЯ КОМАНД	57
2.1. Анализ метрики сопоставления параметров речевого сигнала с	
акустическими моделями	57
2.2. Оценка вероятности ложного срабатывания на произнесение	
постороннего слова	67
2.2.1. Метод оценки вероятности	67
2.2.2. Результаты эксперимента	75
2.3. Сравнение результатов распознавания по достоверности	
распознавания	84
Выводы по разделу 2	94
3. ИССЛЕДОВАНИЕ МЕТОДОВ ПОДАВЛЕНИЯ ПОМЕХ НА СТАДИИ	
ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ РЕЧЕВОГО СИГНАЛА	96
3.1. Оценка влияния помех с использованием критерия достоверност	И
распознавания	96
3.2. Оценка влияния помех на параметры огибающей кратковременн	ОГО
спектра сигнала	. 106

3.3. Подавление помех на основе оценки спектра в паузах речевого	
сигнала	110
3.4. Подавление квазипериодических помех	113
Выводы по разделу 3	124
4. ИССЛЕДОВАНИЕ АДАПТИВНОЙ КОМПЕНСАЦИИ РАДИОПОМЕХ	126
4.1. Возможности компенсации широкополосных помех при наличи	ии их
отражений	127
4.2. Адаптивная компенсация помех в двухканальной системе прие	ема 135
4.3. Моделирование адаптивного компенсатора помех	140
4.5. Адаптивная компенсация помех с обратной связью по решению	o 150
4.6. Устройство приема широкополосных сигналов с подавлением	
импульсных помех	157
Выводы по разделу 4	163
5. ПРОГРАММНЫЕ СРЕДСТВА ДЛЯ ИССЛЕДОВАНИЯ ПОМЕХОУСТОЙЧИВОС	ГИ
СИСТЕМ РАСПОЗНАВАНИЯ ГОЛОСОВЫХ КОМАНД	165
5.1. Функциональный состав программного обеспечения, использус	
при разработке САРГК	165
5.2. Исследование алгоритмов предварительной обработки сигнало	в168
5.2.1. Исследование определителя основного тона	169
5.2.2. Исследование алгоритма адаптивной компенсации помех	173
5.2.3. Программный комплекс с использованием среды графического	
программирования	176
5.2. Программные средства для работы с акустическими моделями	179
5.3. Программные средства для оценки достоверности распознаван	ия
команд	182
Выводы по разделу 5	185
6. АППАРАТНОЕ ОБЕСПЕЧЕНИЕ ИССЛЕДОВАНИЙ ПОМЕХОУСТОЙЧИВОСТИ	
СИСТЕМ РАСПОЗНАВАНИЯ ГОЛОСОВЫХ КОМАНД	187

6.1. Система сбора информации об ошибках распознавания при	
организации прямого доступа к абоненту УАТС	188
6.2. Исследование компенсации широкополосных помех	190
Выводы по разделу 6	195
ЗАКЛЮЧЕНИЕ	197
СПИСОК ОСНОВНЫХ СОКРАЩЕНИЙ	201
СПИСОК ЛИТЕРАТУРЫ	203
ПРИЛОЖЕНИЯ	236
П1. Акты внедрения	236
П2. Анализ влияния шума на оценку периода помехи	241
ПЗ. Результаты экспериментов по оценке вероятности ложного	
срабатывания	245
ПЗ.1. Границы доверительных интервалов оценок вероятности по	
результатам тестирования системы (разработка ЦРТ) базами 3 и 4	245
ПЗ.2. Относительная частота ложного срабатывания (FAR)по результа	там
тестирования системы (разработка фирмы Phonexia) базами 1 и 2	247
П4. Листинги т-функций	250
П4.1. Моделирование адаптивного компенсатора помех	
П4.2. Определение вероятности ошибки при сравнении САРГК	252
П4.3. Сравнение по помехоустойчивости результатов распознавания	
команд	253

Введение

Привлекательность речевого «общения» абонента телефонной сети с автоматическими информационными системами, успехи научных исследований автоматического распознавания речи, а также развитие вычислительной техники обусловили в настоящее время развитие систем голосового самообслуживания (СГС). СГС состоит из трех основных составляющих: интерактивного автоответчика, системы автоматического распознавания голосовых команд (САРГК) и устройства синтеза речи [108]. Основные области использования СГС: контактные центры и справочно-информационные службы.

Использование СГС позволяет освободить операторов контактных центров от часто повторяющихся вопросов и тем самым снизить их психологическую нагрузку, обеспечить круглосуточную работу справочной службы и быстрое обновление информации для клиентов.

Необходимость развития СГС обусловило процесс перехода систем автоматического распознавания голосовых команд (САРГК) из научных лабораторий в сферу практической деятельности человека [13, 26, 97, 108, 125, 132, 146, 159, 160].

Особенностью САРГК является работа в условиях сильной изменчивости произнесения команд. Изменчивость команд успешно учитывается формировании акустических моделей отдельных звуков и команд. В качестве моделей обычно используются модели скрытых марковских процессов (МСМП) [153, 159, 183, 184]. МСМП создаются на основе статистической обработки звукозаписей голосовых множества команд, произносимых различными дикторами. В отсутствие помех при дикторонезависимом распознавании команд, когда степень изменчивости произнесения команд наиболее высокая, достигнута достаточно малая относительная частота ошибок распознавания (менее 3%) приемлемая для использования САРГК на практике [185].

Однако наличие акустических, электрических и радиопомех обусловливает большое количество ошибок распознавания, что препятствует широкому применению САРГК [24, 102, 110, 111]. Воздействие помех приводит либо к ложному срабатыванию системы, либо к ошибке в распознавании команды.

Большой вклад в решение проблемы повышения достоверности автоматического распознавания речи внесли следующие ученые: Бондарко Л.В., Винцюк Т.К., Галунов В.И., Гейлс М., Грей А., Маркел Дж.Д., Потапова Р.К., Прохоров Ю.Н., Рабинер Л.Р., Сапожков М.А., Фант Г., Фланаган Дж., Хуанг К, Шафер Р.В., Янг Б. [5, 9, 12, 14, 96, 99, 103, 104, 105, 106, 122, 124, 125, 151, 152, 153, 159, 165]. Работы данных исследователей и их последователей позволили значительно снизить частоту ошибок распознавания. При проведении научных исследований получены приемлемые для практики результаты. Но внедрение САРГК в практическую деятельность человека наталкивается на большие трудности.

Из-за разнообразия помех сложно учесть их влияние в конкретных условиях эксплуатации системы при формировании акустических моделей команд на стадии разработки САРГК. Приходится проводить дополнительную статистическую обработку большого количества звукозаписей с учетом наличия помех при настройке САРГК для работы в конкретных условиях эксплуатации. Проводится большое количество экспериментов для создания звукозаписей и их дальнейшей обработки. Поэтому создание помехоустойчивых САРГК требует больших затрат [26, 97].

Эксперименты проводятся, во-первых, с целью определения параметров алгоритмов обработки речевого сигнала (РС), обеспечивающих требуемую достоверность распознавания команд, а во-вторых, для оценки параметров достоверности. После каждого эксперимента определяются относительные частоты ложного срабатывания системы (False Alarm Rate - FAR) и ложного пропуска команды (False Reject Rate - FRR), которые являются оценками

соответствующих вероятностей и характеризуют достоверность распознавания [194]. Чем выше требования к достоверности распознавания, тем больше объем тестовых выборок звукозаписей и выше трудозатраты.

Особо следует отметить проблему определения FAR. В связи многообразием помех затруднительно создавать достаточно большую тестовую базу для каждого конкретного случая использования САРГК. Обычно создается некоторая большая тестовая звуковая база, на которой тестируются различные системы. Полученное значение FAR оказывается «привязанным» к данной тестовой базе. Так как помехи в конкретных условиях эксплуатации системы, в общем случае, отличаются от тех, которые отражены в тестовой базе, то значение FAR, определенное при тестировании системы, может не соответствовать значению FAR в условиях эксплуатации системы. На этапе внедрения системы приходится проводить сбор дополнительных данных о помехах и особенностях эксплуатации системы для более точной ее настройки.

С расширением сферы использования САРГК растет количество пользователей систем и увеличивается разнообразие помех, воздействующих на системы. Расширение круга пользователей требует повышения достоверности распознавания команд, что наряду с ростом разнообразия помех приводит к увеличению объема экспериментов при разработке САРГК. Соответствующее увеличение трудозатрат еще можно реализовать в рамках научных исследований, но при инженерном проектировании систем такие затраты часто становятся «неподъемными».

проблема Становится актуальной создания средств исследования помехоустойчивости САРГК, обеспечивающих снижение объема экспериментов при выборе параметров алгоритмов обработки РС и оценке достоверности распознавания Снижение объема экспериментов целесообразно команд. обеспечить разработкой соответствующих методов оценки достоверности

распознавания, а также средств оценки степени подавления помех без привлечения мощных вычислительных ресурсов.

Существующие средства разработки, в основном, направлены на создание акустических моделей голосовых команд, реализацию алгоритмов сопоставления параметров речевых сигналов с акустическими моделями и определение параметров речевых сигналов (РС) [174, 181, 205]. Средствам исследования алгоритмов подавления помех уделено мало внимания. Особенно это касается алгоритмов подавления помех на стадии предварительной обработки речевых сигналов до процедуры определения параметров сигналов, используемых при распознавании.

Объектом исследования являются системы автоматического распознавания голосовых команд в телефонии.

Предметом исследования являются средства исследования и повышения помехоустойчивости систем автоматического распознавания голосовых команд в телефонии.

Целью диссертационной работы является создание средств исследования помехоустойчивости САРГК, которые обеспечивают сокращение объема экспериментов при разработке помехоустойчивых САРГК для нужд телефонии, а также создание средств повышения помехоустойчивости САРГК.

Для достижения поставленной цели в диссертационной работе необходимо решить следующие задачи.

- 1. Провести анализ существующих методов обеспечения и исследования помехоустойчивости САРГК.
- 2. Разработать методы исследования помехоустойчивости САРГК, обеспечивающие снижение объема экспериментов при их проектировании.

- 3. Проанализировать возможности и разработать алгоритмы подавления помех на стадии предварительной обработки сигналов.
- 4. Разработать программные и аппаратные средства для исследования алгоритмов подавления помех.

Научная новизна характеризуется следующими результатами диссертационной работы.

- Предложен критерий достоверности распознавания команд, определяемый при тестировании САРГК выборкой звукозаписей.
- Предложен метод сравнения результатов распознавания команд, полученных при разных сеансах тестирования САРГК в процессе их проектирования, по достоверности распознавания.
- Предложен метод оценки вероятности ложного срабатывания системы в зависимости от степени соответствия посторонних произнесений акустической модели ключевого слова команды, и получены результаты экспериментального исследования метода.
- Получены выражения, отражающие зависимость вероятности ошибки сравнения систем по достоверности распознавания от объема тестовой выборки звукозаписей и параметров достоверности сравниваемых систем.
- Получены выражения, отражающие зависимость степени подавления квазипериодической помехи методом неадаптивной компенсации от соотношения уровней помехи, шума и речевого сигнала.
- Получены выражения, отражающие зависимость степени подавления широкополосных радиопомех методом адаптивной компенсации от погрешности реализации требуемых значений задержек в адаптивных трансверсальных фильтрах.
- Предложен алгоритм адаптивной компенсации радиопомех с подавлением влияния сигнала на процесс адаптации путем использования обратной связи по решению.

Теоретическая значимость работы заключается в следующем.

- Проведен анализ метрики сопоставления произнесения с акустическими моделями помехи, «своей» и «чужой» команд;
- оценена вероятность принятия ошибочного решения при сравнении
 САРГК по достоверности распознавания;
- проанализировано влияние аддитивного шума на результат оценки периода помехи при компенсации квазипериодических помех;
- получено выражение оптимального вектора управляющих коэффициентов для процессора Хоуэлса-Эпплбаума, используемого для адаптивной компенсации широкополосных радиопомех с отражениями;
- исследовано влияние обратной связи по решению, используемой для уменьшения влияния сигнала на работу цепи адаптации, на степень подавления помехи с помощью адаптивной компенсации.

Практическая значимость работы заключается в следующем.

- Использование предложенного критерия сравнения результатов распознавания команд, полученных при разных сеансах тестирования САРГК, по достоверности распознавания позволяет уменьшить объем тестовой выборки звукозаписей;
- с целью снижения объема экспериментов, разработана методика выявления помехи и голосовой команды, которым соответствует нижняя граница достоверности распознавания;
- разработана методика оценки вероятности ложного срабатывания САРГК на произнесения слов, не являющихся ключевыми словами команд, которая позволяет снизить объем экспериментов при настройке системы на конкретные условия эксплуатации;

- разработаны программные средства для получения оценки вероятности ложного срабатывания, слабо зависящей от состава тестовой выборки звукозаписей;
- разработан набор функций системы Matlab, для исследования алгоритмов компенсации помех на основе имитационного моделирования;
- разработаны программно-аппаратные средства, обеспечивающие сбор данных о появлении ошибок распознавания команд в процессе эксплуатации САРГК на телефонной линии;
- разработаны структурные схемы комплекса аппаратуры для испытаний устройств компенсации радиопомех.

Внедрение

- 1. Результаты диссертационной работы используются в ООО «Телеком-Сервис» для оценки качества услуг, представляемых в телефонном контакт-центре.
- 2. Предложенный метод сравнения САРГК, использован в ООО «Центр Речевых Технологий» при проектировании системы VoiceDigger автоматического поиска ключевых слов в записях телефонных разговоров.
- 3. Предложенный метод оценки вероятности ложного срабатывания САРГК на произнесения посторонних слов был использован в ООО «НПП Акустика» при разработке системы «Барышня» голосовой навигации.
- 4. Разработанные программные средства, используются на кафедре радиотехники и радиосистем Владимирского государственного университета при проведении лабораторных работ в рамках учебных дисциплин «Компьютерная телефония» и «Цифровая обработка сигналов изображения и звука».

Методы исследования.

При проведении исследований использовались методы математической статистики, матричного анализа, математический аппарат теории случайных функций, адаптивной фильтрации, а также имитационное моделирование устройств обработки сигналов и экспериментальное исследование систем распознавания голосовых команд.

На защиту выносятся

- 1. Метод сравнения результатов распознавания команд по достоверности распознавания, который использует предложенный критерий достоверности, основанный на определении выборочного среднего и выборочной дисперсии.
- 2. Метод оценки вероятности ложного срабатывания с учетом степени несоответствия произнесения постороннего слова акустической модели команды.
- 3. Результаты теоретического анализа и имитационного моделирования компенсации квазипериодических помех на основе оценки периода помехи.
- 4. Результаты теоретического анализа и имитационного моделирования устройств адаптивной компенсации широкополосных радиопомех.
- 5. Результаты теоретического анализа устройств адаптивной компенсации радиопомех с использованием обратной связи по решению для подавления влияния сигнала на процесс адаптации.

Достоверность результатов исследований обусловлена использованием соответствующего математического аппарата, имитационного моделирования и экспериментальной проверкой.

Апробация результатов работы

Основные положения диссертации докладывались на 2...10 МНТК «Перспективные технологии в средствах передачи информации», (1997-2013г.),

Владимир-Суздаль; на 2,4...8,10 МНТК «Физика и радиоэлектроника в медицине и экологии», (1996 - 2012г.), Владимир-Суздаль; на 4...8 межрегиональных НТК «Обработка сигналов в системах двусторонней телефонной связи» (1995...1998), МТУСИ, на 10,11,13 Москва-Пушкинские горы, межрегиональных «Обработка сигналов в системах телефонной связи и вещания», (2000, 2004г.), Москва-Пушкинские горы; МТУСИ, на международной конференции SPECOM'2003. - Moscow State Linguistic University, Moscow, Russia, 2003, Ha МНПК «Фундаментальные проблемы радиоэлектронного приборостроения INTERMATIC-2004», Москва; $(2004\Gamma.),$ на международной конференции SPECOM'2005, – University of Patras, Patras, Greece, 2005.

Публикации

По теме диссертации опубликовано 70 работ, в том числе одна монография, 26 статей, из них 13 статей в изданиях, рекомендованных ВАК, 4 авторских свидетельства на изобретения, 39 тезисов докладов в трудах международных и российских конференций.

Личное участие

Основные теоретические результаты были получены лично автором в период с 1974 по 2014г. Разработка программных и аппаратных средств, а также экспериментальные исследования проводились коллективом сотрудников при активном участии автора в ходе выполнения научно-исследовательских работ, проводимых в качестве исполнителя и научного руководителя на кафедре радиотехники и радиосистем Владимирского государственного университета.

Структура и объем работы.

Диссертация состоит из Введения, шести разделов, Заключения, списка литературы, имеющего 207 наименований отечественных и зарубежных источников, в том числе 70 работ автора, и Приложений. Общий объем диссертации составляет 257 страниц, в том числе 201 страницы основного текста,

33 страницы списка литературы, 50 рисунков, 16 таблиц и 22 страницы приложений.

В первом разделе монографии дана краткая характеристика основных этапов обработки РС при их распознавании, рассмотрены особенности акустических моделей команд, учитывающих сильную изменчивость произнесения. Процесс обработки речевого сигнала (РС) выглядит следующим образом [108, 187]. Сначала проводится предварительная обработка сигнала, во время которой сигнал отделяется от длительных пауз, присутствующих в речи диктора, и осуществляется подавление помех, сопровождающих сигнал. Затем сигнал разделяется на короткие сегменты (интервалы квазистационарности РС), и для каждого сегмента определяется набор (вектор) параметров. Полученная последовательность векторов сопоставляется с акустическими моделями команд, коротких пауз и помех. При сопоставлении определяется наиболее вероятная последовательность моделей, которая и является результатом распознавания [159, 184, 187].

Приводится классификация помех по их источникам и по характеру влияния на результат распознавания ГК, а также сведения о видах параметров РС, используемых при распознавании. Отмечается, что наиболее часто используются мелочастотные кепстральные коэффициенты (mel-frequency cepstral coefficients – MFCC), которые характеризуют огибающую кратковременного спектра сигнала и учитывают особенности человеческого слуха.

Рассмотрены фильтрация и компенсация помех, применяемые на стадии предварительной обработки РС. Наличие пауз в речевом сигнале обусловливает возможность подавления помех методами компенсации. Оценивая параметры помехи в паузе, можно скомпенсировать воздействие помехи на интервале произнесения ключевого слова команды.

Подавление помех на этапе предварительной обработки речевого сигнала обычно осуществляется в спектральной области с помощью спектрального

вычитания и использования фильтра Винера [146, 153, 159]. Следует отметить, что в некоторых случаях воздействия импульсных помех указанные методы не обеспечивают требуемого уровня их подавления.

Рассматриваются существующие подходы к подавлению помех при сопоставлении параметров РС с акустическими моделями команд и помех. Анализируется метрика сопоставления параметров РС с акустической моделью команды — моделью скрытого марковского процесса (МСМП). Рассматривается графическое представление сети акустических моделей — возможных результатов распознавания. В конце раздела проанализированы методы оценки достоверности распознавания команд.

Во втором разделе проведен анализ метрик сопоставления произнесений ключевого и неключевого слов с их акустическими моделями. Модель ключевого слова реализуется последовательным соединением акустических моделей трифонов, а модель неключевого - акустических моделей монофонов. Трифонная модель звука учитывает влияние соседних звуков на данный звук, а монофонная - не учитывает этого влияния. На основе проведенного анализа рассмотрен метод оценки FAR с учетом отличия произнесения постороннего (неключевого) слова от произнесения ключевого слова команды. Рассмотрена методика проведения эксперимента для определения FAR на основе данного метода. Приведены результаты эксперимента.

Предложен критерий сравнения результатов распознавания ключевого слова, полученных при разных сеансах тестирования САРГК в процессе их проектирования, ПО достоверности распознавания. Проведенный двух результатов тестирования САРГК показывает, сравнение достоверности распознавания с использованием данного критерия позволяет сократить объем тестовой звуковой базы при заданной вероятности принятия ошибочного решения по результатам сравнения. Рассмотрена соответствующая методика сравнения результатов распознавания САРГК.

В третьем разделе проанализирована возможность использования рассмотренного в предыдущем разделе критерия достоверности при оценке степени влияния помех на работу САРГК. Приведены результаты эксперимента, которые позволяют определить, какие из помех оказывают наибольшее влияние на достоверность распознавания.

Рассмотрено влияние помех на параметры PC при использовании синтезированных сегментов РС. Проанализированы возможности подавления помехи на основе оценки ее параметров в паузе речевого сигнала. Приведены результаты экспериментального исследования алгоритма подавления помех с использованием фильтра Винера [147]. При проведении эксперимента оценивалось только повышение достоверности распознавания, уменьшение искажений параметров сигнала для всей тестовой выборки звукозаписей.

Проведен анализ возможностей компенсации квазипериодической помехи на основе оценки ее периода в паузах РС. Приведены результаты моделирования процедуры подавления помехи, подтверждающие справедливость проведенного теоретического исследования.

В четвертом разделе рассмотрены возможности адаптивной компенсации радиопомех cпомощью пространственно-разнесенного приема. Проанализировано влияние широкополосных помех, когда на приемные антенны поступает сумма задержанных во времени нескольких копий помехи. Копии формируются за счет отражений помехи от крупных сооружений вблизи станции связи. Получены выражения, связывающие уровень подавления помехи с погрешностью реализации требуемых задержек в адаптивных трансверсальных фильтрах. Справедливость полученных выражений подтверждена моделированием соответствующих процедур обработки сигналов.

Проанализирована возможность подавления влияния сигнала на работу цепей адаптации компенсатора помех путем использования обратной связи по решению.

В пятом разделе рассматривается программное обеспечение, которое требуется при проектировании системы распознавания команд. Дана краткая характеристика всех этапов проектирования: подготовка выборок звуковых файлов для обучения и тестирования системы, создание акустических моделей и акустических моделей \mathbf{c} определением параметров достоверности распознавания, выбор алгоритмов подавления помех с оценкой достоверности распознавания. Приводится классификации программного обеспечения функциональному проектирования ПО назначению организации И ПО пользовательского интерфейса.

Рассмотрены основные программные средства, предназначенные исследования помехоустойчивости САРГК на стадии ее проектирования. В частности, рассмотрена программа, предназначенная ДЛЯ оперативного редактирования сети акустических моделей. Проанализированы возможности комплекса, предназначенного программного ДЛЯ построения программ исследования с помощью графического программирования. Описан сценарий вызова функций системы Matlab для моделирования адаптивного компенсатора радиопомех.

В шестом разделе рассмотрена структура системы сбора информации об ошибках распознавания, которая подключается к учрежденческой автоматической телефонной станции (УАТС) при организации прямого доступа абонента телефонной сети общего пользователя к абоненту УАТС с помощью голосовых команд. Рассмотрено также аппаратное обеспечение оценки степени компенсации радиопомех в составе станции связи.

1. Методы обеспечения помехоустойчивости распознавания голосовых команд

1.1. Основные этапы обработки речевого сигнала при распознавании

1.1.1. Предварительная обработка сигнала

Процесс обработки РС можно разделить на два этапа [108, 159, 184, 205]. Сигнал от микрофона, поступивший по каналу связи на вход системы распознавания, сначала проходит предварительную обработку. Он очищается от помех, с помощью детектора голосовой активности абонента (Voice activity detector – VAD) выделяются только речевые сегменты [129,155, 180, 199], которые присутствуют в речи. Далее определяются параметры речевого сигнала, которые на втором этапе сопоставляются с акустическими моделями (моделями отдельных звуков или произнесений слов), реализованными, как правило, в виде моделей скрытых марковских процессов (МСМП) [136, 159, 205]. Акустические модели создаются на основе множества обучающих произнесений различных слов [159, 186].

Параметры РС определяются на основе модели формирования звуков речи. РС рассматривается как реакция рекурсивного фильтра на широкополосное входное воздействие [96,106]. Изменяющиеся во времени параметры фильтра отражают работу органов артикуляции и определяют вид, тембр и скорость произнесения звука, поэтому они несут основную информацию для процедуры распознавания [159, 184]. Амплитудно-частотная характеристика фильтра определяет огибающую кратковременного спектра сигнала.

Источник возбуждения рекурсивного фильтра отражает работу легких и голосовых связок. Его работу можно описать тремя основными параметрами [184]. Первый параметр - признак «тон-шум» - является классификационным. Если в формировании звука участвуют голосовые связки, то звук является

вокализованным. Признак принимает значение «единица». В противном случае говорят о невокализованных звуках, и признак принимает нулевое значение.

Если звук является вокализованным (например, в случае гласных звуков), то на выходе источника присутствуют короткие импульсы, следующие с частотой $F_{\rm OT}$ основного тона - частотой вибраций голосовых связок (второй параметр). В случае невокализованных звуков (например, шипящих) на выходе присутствует белый шум. По частоте основного тона можно, в частности, определить пол диктора, что в ряде случаев повышает достоверность распознавания. Методы определения $F_{\rm OT}$ и классификации звуков на вокализованные и невокализованные рассматриваются в работах [2, 106, 114, 115, 159]. Третий параметр - коэффициент усиления сигнала источника возбуждения - определяет громкость звука.

Так как речевой сигнал является нестационарным, то все указанные выше параметры определяются на коротких (10...30мс) интервалах квазистационарности сигнала [184]. Параметры модели формирования РС являются основой для определения параметров сигналов, используемых при распознавании речи [42, 96, 99, 106].

Существует несколько методов определения основных информативных параметров РС. Наиболее часто [159, 184, 205] используют анализ спектра с помощью блока полосовых фильтров на основе быстрого преобразования Фурье (БПФ), а также метод линейного предсказания (linear predictive coding – LPC), когда параметры огибающей спектра для анализируемого сегмента сигнала выражаются в виде набора коэффициентов линейного предсказания, которые определяют частотную характеристику рекурсивного фильтра. Набору коэффициентов линейного предсказания равноценен набор коэффициентов отражения, когда рекурсивный фильтр имеет лестничную структуру [106].

Блок фильтров, в частности, используется при определении мелочастотных кепстральных коэффициентов (Mel-Frequency Cepstral Coefficients - MFCC) – параметров РС, широко применяемых при распознавании речи [159, 188, 205].

Спектр, полученный в результате дискретного преобразования Фурье, с помощью банка фильтров разбивается на полосы частот, и в каждой полосе вычисляется среднее значение энергии сигнала. Для учета особенностей слуха человека полоса пропускания фильтров в низкочастотной части спектра устанавливается меньше полосы пропускания фильтров для высокочастотной части спектра. Такое неравномерное разбиение спектра сигнала равноценно преобразованию шкалы частот (в герцах) в мел-шкалу (высота звука оценивается в мелах):

$$Mel(f) = 2595\log_{10}(1 + \frac{f}{700}).$$
 (1.1)

Далее квадраты абсолютных значений коэффициентов спектра умножаются на коэффициенты амплитудно-частотных характеристик фильтров и суммируются на выходе каждого фильтра. Полученные значения спектра логарифмируются и подвергаются дискретному косинусному преобразованию — формируются кепстральные коэффициенты. При формировании параметров МГСС производится дополнительно преобразование частотной шкалы с целью уменьшения степени изменчивости параметров при произнесении одной команды различными дикторами [207].

В работе [147] рассматривается возможность повышения достоверности распознавания путем регулировки значений центральных частот и ширины полос, на которые разбивается спектр РС. В работе [161] предлагается использовать большое количество (порядка 50) коэффициентов дискретного косинусного преобразования для распознавания названий букв английского алфавита.

Малая взаимная корреляция МГСС позволяет использовать диагональные матрицы при работе с моделями скрытых марковских процессов - МСМП, что упрощает вычисления [205]. Кроме того, использование кепстра позволяет уменьшить зависимость параметров РС от изменений частотной характеристики канала связи. Влияние частотной характеристики в кепстральной области проявляется в виде добавления некоторой постоянной составляющей к

параметрам сигнала, которая в дальнейшем относительно просто устраняется [159]. В случае распознавания слитной речи зависимость параметров речи от времени подвергается высокочастотной фильтрации, что позволяет устранить влияние изменений канала передачи речи на результат распознавания (RASTA-метод [157]).

Кроме MFCC, используются PLP-коэффициенты (Perceptual Linear Prediction – PLP) [156, 159, 205]. В отличие от MFCC здесь выходы банка фильтров взвешиваются согласно кривым равной громкости, которые учитывают особенности восприятия человеком громкости звуков на разных частотах. Значения взвешенного энергетического спектра подвергаются преобразованию (компрессии) вида $P(\omega) = [P_0(\omega)]^{0.33}$. Затем находятся коэффициенты линейного предсказания, которые далее преобразуются в кепстральные коэффициенты.

Явного преимущества одной системы параметров над другой не обнаружено. В одних публикациях отдается предпочтение кепстральным коэффициентам, в других — PLP-коэффициентам [172]. Рассмотренные параметры PC обычно дополняются их первыми и вторыми производными, что значительно повышает достоверность распознавания.

1.1.2. Сопоставление параметров сигнала с акустическими моделями

Сопоставление происходит использованием сетевой структуры акустических моделей голосовых команд возможных результатов распознавания. Каждой голосовой команде ставится в соответствие модель. При распознавании параметры произнесенной команды сопоставляются со всеми моделями, которые соответствуют заданному перечню команд. Модель, которой соответствует наибольшая метрика сопоставления с параметрами произнесенной команды, по сути, является результатом распознавания.

Модель команды, как правило, состоит из последовательного соединения моделей звуков. Если модели звуков созданы, то можно быстро получить модель произнесения любого слова — появляется возможность быстрого редактирования голосового меню информационно-справочной системы, где используется САРГК. Следует отметить, что на произнесение звука сильное влияние оказывают соседние звуки (явление коартикуляции) [30], поэтому, как правило, модель звука создается для конкретного случая его окружения двумя соседними звуками. Совокупность центрального и двух соседних звуков называется трифоном [205].

В ряде случаев используют более грубую модель звука без учета влияния соседних звуков. В этом случае говорят о монофонной модели звука [159, 205]. В простых случаях, когда при эксплуатации системы не требуется менять команды, модель команды не разбивается на модели звуков.

Чтобы учесть значительную изменчивость тембра и скорости произнесения звуков голосовой команды, в качестве акустической модели каждого звука используется модель скрытого марковского процесса (МСМП), граф которой представлен на Рисунке 1.1. [137, 184].

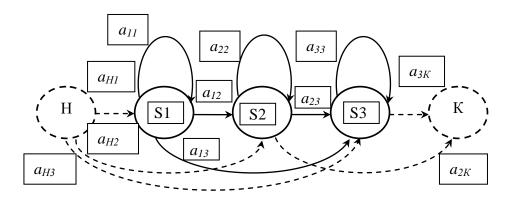


Рисунок.1.1. Граф модели скрытого марковского процесса.

Процесс в каждый момент времени характеризуется одним состоянием из набора $S = \{S_1, S_2, ... S_N\}$ состояний (в данном примере N=3, «Н», «К» - начальное и конечное состояния). Наличие конечного и начального состояний позволяет соединять модели последовательно друг с другом. В каждый из дискретных

моментов времени t = 1, 2, ... происходит либо возврат в исходное состояние, либо переход из одного состояния в другое, находящееся правее данного, согласно набору вероятностей a_{ij} , связанных с данным состоянием. Вероятности a_{ij} перехода из i—го в j—е состояние составляют матрицу A переходных вероятностей. Существует также набор вероятностей возникновения i—го начального состояния $\{a_{ni}\}$. Аналогично имеется набор вероятностей перехода из i—го состояния в конечное состояние $\{a_{ik}\}$.

Каждое j—е состояние (за исключением состояний H и K) характеризуется совместной плотностью $b_j(\vec{X})$ вероятностей значений вектора \vec{X} параметров речевого сигнала. Распределения плотности вероятности $b_j(\vec{X})$ каждого состояния, в общем случае, имеют сложную форму, их часто аппроксимируют композицией гауссовых распределений.

$$b_{j}(\vec{X}) = \sum_{m=1}^{M} c_{jm} N(\vec{X}, \vec{\boldsymbol{\mu}}_{jm}, \Sigma_{jm}), \qquad (1.2)$$

где M — число гауссовых компонентов в распределении вероятностей, c_{jm} - вес m-го компонента смеси, $N(\vec{X}; \vec{\mu}, \Sigma)$ —гауссиан от нескольких переменных с вектором математического ожидания $\vec{\mu}$ и корреляционной матрицей Σ , который определяется как

$$N(\vec{X}, \vec{\mu}_j, \Sigma_j) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_j|}} e^{-0.5(\vec{X} - \vec{\mu}_j)' \Sigma_j^{-1} (\vec{X} - \vec{\mu}_j)}$$
(1.3)

Здесь n — размерность вектора параметров \vec{X} , $|\Sigma_j|$ — определитель корреляционной матрицы Σ_j ., c_{jm} весовой коэффициент гауссиана. Набор параметров $\lambda=(A, B, \vec{\mu}, \Sigma_j, c_{jm})$ полностью характеризует МСМП. Указанные параметры определяются статистическим методом (алгоритм Баума-Уэлча) путём анализа большого количества различных вариаций произнесения каждого звука (слова) [159, 184]. Часто для снижения объема вычислений считают, что

корреляционная матрица параметров является диагональной, и каждый параметр описывается гауссовым законом распределения вероятностей.

Наличие дуг на графе позволяет учесть изменения длительности звука. Возможные изменения тембра звука описываются совокупностью плотностей $b_j(\vec{X})$ вероятностей. Матрица A переходных вероятностей и совокупность плотностей $b_j(\vec{X})$ вероятностей являются параметрами акустической модели и определяются по результатам обработки множества вариантов произнесения данного звука на этапе обучения системы.

Модель позволяет сгенерировать последовательности векторов параметров речевого сигнала для различных вариантов произнесения данного звука, причем длина последовательности векторов параметров (длительность звука) и значения векторов (тембр звука) могут меняться с вероятностью, которая определяется плотностями $b_j(\vec{X})$ вероятностей и матрицей A переходных вероятностей. Наблюдению доступны лишь значения вектора \vec{X} , а сами состояния для наблюдения не доступны.

При распознавании ГК решается следующая задача: необходимо определить такую последовательность состояний МСМП во времени, которая с наибольшей вероятностью соответствует анализируемой последовательности векторов параметров речевого сигнала. Распознавание голосовых команд заключается в определении такой модели из имеющегося перечня моделей, которая с наибольшей вероятностью соответствует наблюдаемой последовательности векторов.

Пусть наблюдается набор из пяти векторов параметров $X = \{\overrightarrow{X_1}, \overrightarrow{X_2}, ..., \overrightarrow{X_5}\}$ для моментов времени: $t_1, t_2, ... t_5$. Процесс смены состояний модели во времени отражается решетчатой диаграммой, представленной на Рисунке 1.2

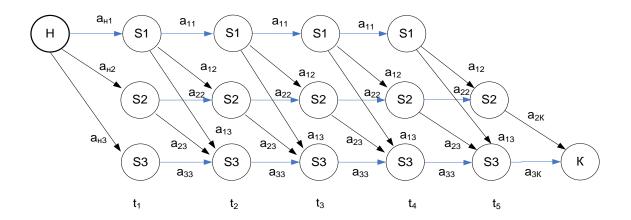


Рисунок 1.2. Решетчатая диаграмма

Нахождение оптимальной последовательности состояний осуществляется с помощью алгоритма Витерби [159]. Эта последовательность состояний выражается в прохождении некоторого пути по решетчатой диаграмме от состояния «Н» до состояния «К».

Пусть для данного случая набора \boldsymbol{X} наблюдений определена оптимальная последовательность состояний $\mathbf{Q} = \{S_1 - S_1 - S_1 - S_2 - S_3\}$. Тогда вероятность появления оптимальной последовательности состояний пропорциональна величине

$$P(X, Q) = a_{H1}b_1(\overrightarrow{X_1})a_{11}b_1(\overrightarrow{X_2})a_{11}b_1(\overrightarrow{X_3})a_{12}b_2(\overrightarrow{X_4})a_{23}b_3(\overrightarrow{X_5})a_{3K}.$$
(1.4)

Обычно в качестве метрики сопоставления произнесения команды с МСМП используется логарифм указанной величины: $M(X/\lambda) = ln[P(X, Q)]$.

В последнее время МСМП была модифицирована — возможные изменения длительности состояний учитываются в модели непосредственно, что позволяет более точно учесть изменчивость произнесения ГК. [159]. Рассмотрим модель с тремя состояниями (Рисунок 1.3)

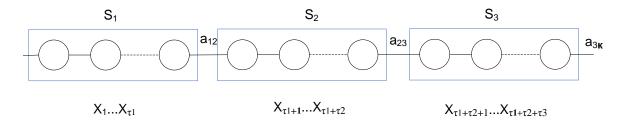


Рисунок 1.3. Граф МСМП с учетом длительности состояний.

Предположим, что после сопоставления параметров некоторого произнесения с МСМП последовательность векторов параметров распределилась между состояниями следующим образом. Первому состоянию S_1 сопоставлено τ_1 векторов, второму состоянию $S_2 - \tau_2$ векторов, третьему $S_3 - \tau_3$ векторов. Метрика сопоставления последовательности \boldsymbol{X} векторов параметров речевого сигнала с данной моделью λ [160]

$$M(X|\lambda) = \sum_{t=1}^{\tau_1} \ln(b_1(\overrightarrow{X_t})) + \sum_{t=\tau_1+1}^{\tau_1+\tau_2} \ln(b_2(\overrightarrow{X_t})) + \sum_{t=\tau_1+\tau_2+1}^{\tau_1+\tau_2+\tau_3} \ln(b_3(\overrightarrow{X_t})) + \sum_{i=1}^{3} \ln(d_i(\tau_i) + \ln(a_{12}) + \ln(a_{23}) + \ln(a_{3K})$$

$$(1.4)$$

Здесь $b_i(\overrightarrow{X_t})$ — плотность вероятности соответствия вектора $\overrightarrow{X_t}$ (t — номер вектора в последовательности векторов) параметров сигнала i-му состоянию модели; $d_i(\tau_i)$. — плотность вероятности соответствия длительности τ_i состоянию S_i ; $a_{i,i+1}$ — вероятность перехода в соседнее состояние. В данной модели непосредственно учтена информация о длительности каждого состояния. Так как метрика состоит из большого числа слагаемых — случайных величин, то она также является случайной величиной.

Так как в одном пункте голосового меню САРГК присутствует несколько ГК, и они могут произноситься с паузами, то для учета произнесения различных команд и наличия пауз используется сеть акустических моделей [205]. Пример ее графического изображения для случая двух ГК представлен на Рисунке 1.4.

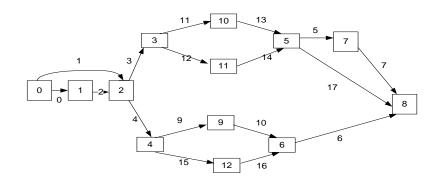


Рисунок 1.4. Сеть акустических моделей.

Данной структуре соответствует перечень узлов и дуг, приведенный в нижеследующей Таблице 1.1.

Таблица 1.1. Состав сети акустических моделей

N=13 L=18	I=7 W=pause2	J=0 S=0 E=1	J=9 S=4 E=9
I=0 W=!START	I=8 W=END	J=1 S=0 E=2	J=10 S=9 E=6
I=1 W=pause1	I=9 W=Помеха1	J=2 S=1 E=2	J=11 S=3 E=10
I=2 W=!NULL	I=10 W=ΓK1	J=3 S=2 E=3	J=12 S=3 E=11
I=3 W=!NULL	I=11 W=ΓK2	J=4 S=2 E=4	J=13 S=10 E=5
I=4 W=!NULL	I=12 W=Помеха2	J=5 S=5 E=7	J=14 S=11 E=5
I=5 W=!NULL		J=6 S=6 E=8	J=15 S=4 E=12
I=6 W=!NULL		J=7 S=7 E=8	J=16 S=12 E=6
			J=17 S=5 E=8

Каждому узлу соответствует акустическая модель. Дуги показывают возможность перехода между узлами. N — число узлов, L — число дуг, I — обозначение узла, W — название модели, соответствующей узлу, J — обозначение дуги, S - обозначает начало дуги, E — обозначает окончание дуги. Первые две колонки - это описания узлов, а последние две - описания дуг (с указанием ее начала и окончания). Узлы типа !NULL являются служебными и используются для объединения дуг.

Предполагается, что командное слово произносится либо с паузой (pause), либо без нее. Вместо произнесения слова возможно появление постороннего

звука - помехи. Слова имеют акустические модели (узлы 10 и 11). Модели помех отражены узлами 11, 12.

В процессе сопоставления некоторого произнесения с сетью моделей определяется та последовательность узлов и дуг, которая соответствует произнесению с наибольшей вероятностью. Так как узлы состоят из МСМП, то оптимальной последовательности узлов и дуг соответствует оптимальная последовательность состояний моделей. Здесь также используется алгоритм Витерби.

Возможны три варианта сопоставления некоторого произнесения с акустическими моделями произнесений ключевых слов ГК, входящими в состав сетевой структуры акустических моделей, которая соответствует некоторому пункту речевого меню. При первом варианте произнесение ключевого слова сопоставляется со «своей» моделью, при втором – с «чужой» моделью, которая относится к другой ГК. При третьем варианте произнесение сопоставляется с моделью произнесения посторонних звуков – помехой.

Из-за изменчивости произнесения ГК и наличия помех возникают ошибки распознавания. Они проявляются либо в виде ложного срабатывания системы на помеху, либо в виде ложного пропуска ГК - трактовки данной команды в качестве помехи или другой команды.

1.2. Подавление помех при предварительной обработке речевого сигнала

Помехи, возникающие в типичном канале телефонной связи, состоящем из двух абонентских телефонных линий и магистрального участка канала связи, по месту возникновения можно подразделить на три группы. Источником помех, вопервых, является сам диктор, который непроизвольно издает посторонние звуки (шум дыхания, покашливание,...) и произносит посторонние слова, не являющиеся ГК. Во-вторых, источником помех является акустическое окружение

диктора (шум вентилятора, звуки автомобиля, щелчки клавиатуры компьютера,...). В-третьих, источниками помех являются средства связи (абонентские телефонные линии, магистральный участок канала связи). Магистральный участок, в общем случае, может включать в себя и радиоканал, в котором присутствуют радиопомехи.

Помехи, создаваемые средствами связи, можно условно разделить на низкочастотные и высокочастотные. Низкочастотные возникают на проводных участках каналов связи, а высокочастотные обусловлены действием радиопомех. Следует отметить, что высокочастотные помехи могут проявить себя в виде низкочастотных на проводных участках канала связи. Так как радиоканал является источником множества различных помех и наиболее сильных искажений сигнала, то влияние его на передачу команд может оказаться весьма существенным. В частности, при цифровой связи и пакетной передаче речи, могут наблюдаться потери пакетов — возникают кратковременные паузы в передаче речи, что ведет к появлению большого числа ошибок распознавания [129].

По отношению к интервалу времени, который занимает ГК, помехи можно подразделить на те, которые попадают в данный интервал, и те, которые не попадают в него. Помехи, находящиеся вне интервала произнесения командного (ключевого) слова, могут вызвать ложное срабатывание системы. Помехи, находящиеся внутри интервала, искажают параметры сигнала, что вызывает ложную трактовку команды либо в качестве помехи, либо в качестве другой команды – имеет место ошибка пропуска команды.

Низкочастотные помехи могут подавляться на стадии предварительной обработки сигнала с помощью фильтрации и компенсации. Если помеха не затрагивает интервал времени присутствия команды, то возможно ее подавление и на этапе сопоставления параметров сигнала с акустическими моделями путем формирования модели помехи.

Радиопомехи при аналоговой связи трансформируются после демодуляции в низкочастотные помехи, а при цифровой связи и пакетной передаче данных — могут привести к потере пакетов данных, поэтому целесообразно их подавление осуществлять до процедуры демодуляции.

Особенностью речевого сигнала является наличие в нем пауз. Во время пауз, которые определяются с помощью детектора голосовой активности, можно с небольшой погрешностью определить параметры помехи. Указанная возможность обусловила широкое распространение метода спектрального вычитания для подавления помех [23, 149, 159, 171, 203]. Оценивается спектральная плотность мощности помехи на интервале паузы с последующим вычитанием ее из спектральной плотности РС, пораженного аддитивной помехой.

Так как оценки спектральной плотности являются случайными величинами, то результат вычитания может принимать отрицательное значение. Чтобы избежать таких ситуаций, малые значения разностей заменяют малой положительной фиксированной величиной, что приводит к появлению помехи типа «музыкальный» шум.

Чтобы сократить вычислительные затраты при оценке уровня шума, предлагается в качестве данной оценки использовать 0,5-квантиль распределения мощности зашумленного сигнала на данной частоте в некоторой окрестности заданного момента времени [148]. Метод может эффективно использоваться для подавления нестационарных помех, у которых скорость изменения спектральной мощности не очень велика.

На основе спектрального вычитания используется фильтр Винера, который минимизирует искажения сигнала, возникающие при его фильтрации [133, 143, 146]. Приближенно фильтр можно реализовать, оценивая кратковременные энергетические спектры в паузе РС, когда помеха проявляет себя наиболее ярко, и на интервале наложения помехи на сигнал. Зная спектральные плотности очищенного и пораженного помехой РС, можно определить частотную

характеристику фильтра, через который следует пропустить сигнал для его фильтрации. В Европейском проекте Aurora, посвященном созданию САРГК, где предварительная обработка РС проводится на терминале пользователя, а собственно распознавание — на удаленном компьютере, указанный метод используется для подавления медленно меняющихся нестационарных помех [140, 146, 158].

С целью повышения степени подавления нестационарных помех в работе [162] предложено параметры фильтра, подавляющего помеху, обновлять, используя не только предыдущие наблюдения, но и последующие. Такой подход позволяет подавить даже импульсивный шум.

В работе [101] рассматривается возможность уменьшения влияния стационарного шума на параметры РС путем усреднения параметров зашумленного сигнала, относящихся к похожим сегментам частично очищенного сигнала. Похожие сегменты ищутся на интервале длительностью примерно 50 мс, окружающем обрабатываемый сегмент. Степень подавления шума больше в случае вокализованных звуков по сравнению с невокализованными.

В работе [144] предлагается спектр входного сигнала, пораженного помехой, разбить на шесть полос. В полосах помехи мало отличаются по форме. Они отличаются, в основном, по мощности. С помощью предварительно обученного искусственного многослойного персептрона влияние помехи в каждой полосе подавляется. Так как форма помехи в каждой полосе слабо зависит от формы широкополосной помехи, сопровождающей РС, то не требуется менять результаты обучения персептрона при изменении формы широкополосной помехи.

В работе [135] исследована возможность усреднения - низкочастотной фильтрации - последовательности векторов МГСС. При этом влияние шума на достоверность распознавания ГК снижается. Следует отметить, что выбор

параметров усреднения проводится экспериментально и зависит от степени различий спектров РС и шума.

В работе [170] рассмотрен метод подавления шума на основе анализа графического отображения спектрограмм зашумленной речи. Результатом анализа является выделение фрагментов изображения, где превалирует шум. Данные фрагменты затем подавляются. Оставшиеся фрагменты используются для распознавания ГК. Следует отметить, что для обеспечения относительной частоты ошибок меньшей 5% эффективность метода незначительна.

В работе [191] рассмотрен метод снижения влияния шума на параметры РС путем расширения полос пропускания фильтров, используемых для мелочастотного преобразования, по сравнению с обычными их значениями. Действие метода проиллюстрировано на примере классификации векторов параметров, представляющих 10 гласных звуков.

Известен подход к подавлению помех на основе вейвлет-преобразования [25]. Суть подхода заключается в определении коэффициентов вейвлет-анализа сигнала, которые относятся к помехе. После обнуления коэффициентов, производится реконструкция сигнала по остальным коэффициентам. Здесь следует отметить сложность подбора типа вейвлета, необходимого для эффективного выделения помехи. Кроме того, эффективность подхода снижается при снижении отношения сигнал-помеха.

В работах [142, 192] описывается метод подавления помех, создаваемых акустическим окружением диктора. Рассматривается возможность использования нескольких каналов записи сигналов с последующей адаптивной компенсацией помех.

В работе [175] рассматриваются различные методы линейной и нелинейной обработки спектральных отсчетов РС, повышающие помехоустойчивость САРГК. В частности, рассматривается метод увеличивающий пики кратковременного спектра РС. Такое увеличение позволяет снизить влияние помех на параметры РС.

Для подавления периодических помех используется метод компенсации помехи, наложенной на сигнал, помехой, которая присутствует на интервале паузы РС и которая задержана во времени на период помехи больший интервала корреляции РС [23]. Следует отметить, что при таком способе подавления помехи отношение сигнал-шум снижается, так как к широкополосному шуму, который присутствует в сигнале, добавляется шум, присутствующий в паузе сигнала.

В ряде работ рассматривается возможность подавления помехи путем использования ее модели. В работе [202] рассматривается возможность подавления нестационарной помехи путем адаптации параметров модели помехи в процессе распознавания названий цифр. Фоновый шум посторонних разговоров (при отношении сигнал-шум 21,5дБ) удалось подавить с помощью адаптации до уровня, обеспечивающего достоверность распознавания 96,69%. Такая достоверность соответствует отношению сигнал-шум 29,5дБ, когда средства подавления шума не используются. Если же в данном случае использовать средства подавления стационарного шума (то есть адаптацию модели шума во времени не проводить), то достоверность распознавания соответствует 95,18%.

В работе [206] предложено использовать искусственные нейросети для преобразования параметров сигнала, прошедшего телефонный канал, в параметры сигнала, записанного с высоким качеством и использованного для создания акустических моделей. В этом случае не требуется длительного переобучения моделей помех, используемых для распознавания речи «телефонного» качества.

Рассматриваются возможности изменения набора PC. параметров используемых для распознавания. В статье [195] для повышения устойчивости распознавания речи К воздействию шума предлагается обычно систем используемый набор параметров РС (МГСС) дополнить значениями частот центроидов частотных поддиапазонов, на которые разбивается диапазон РС. При использовании шести поддиапазонов и при отношении сигнал-шум равном 15 дБ

относительная частота ошибок распознавания снизилась на 20%, а при отношении 30 дБ – на 1,5%.

В работе [189] рассматривается возможность снижения влияния шума на параметры РС, используемые при распознавании. Предлагается вместо процедуры логарифмирования, которая применяется при вычислении МГСС, использовать операцию возведения в дробную степень — извлечения корня степени больше 2. Показано, что возведение в степень 0,08 снижает влияние шума на параметры РС. При этом незначительно увеличивается достоверность распознавания, и сокращаются вычислительные затраты.

В работе [176] рассматривается возможность повышения помехоустойчивости систем распознавания речи с помощью классификации сегментов РС на вокализованные и невокализованные.

В работах [23, 128, 130] приводится обзор методов подавления помех и устранения искажений в РС. Указанные методы предназначены для обработки сигналов с целью их дальнейшего прослушивания человеком. Рассмотрим возможность использования данных методов в САРГК. В частности, методы компенсации помех, описанные в данных работах, используются и при распознавании речи.

Однако, в случае использования фильтров происходит не только подавление помехи, но и изменение параметров сигнала, что может снизить достоверность распознавания. В частности, если помеха узкополосная, то для ее подавления применяются режекторные фильтры. За счет устранения эффекта маскировки слабого РС сильной помехой удается повысить разборчивость РС. Однако при режекции изменяется огибающая спектра РС, следовательно, меняются параметры сигнала, используемые при распознавании. Устраняя искажения спектра за счет действия помехи, метод сам искажает сигнал. Аналогичные рассуждения справедливы и при использовании прочих методов фильтрации. Минимизировать искажения сигнала можно, используя фильтр Винера.

В работе [141] рассматривается возможность классификации неречевых звуков. Показано, что использование динамического программирования и векторного квантования позволяет достичь 70% точности классификации звуков. Результаты классификации можно использовать в дальнейшем для применения соответствующих методов подавления помех.

1.3. Подавление помех с использованием акустических моделей

Воздействие помех на ГК изменяет параметры РС. Это может привести к тому, что произнесенная команда при сопоставлении со «своей» моделью обусловит меньшую метрику, по сравнению с результатом сопоставления с «чужой» моделью. Результатом распознавания является «чужая» модель — возникает ошибка распознавания. Если помеха возникает в паузе между произнесениями ГК, то возможно ложное срабатывание системы.

Наиболее эффективным, но и наиболее трудоемким методом подавления помех является использование акустических моделей, полученных из зашумленной речи [178]. Однако число таких моделей очень велико и обусловлено количеством вариантов проявления помехи. Каждый вариант определяется типом помехи и соответствующим отношением сигнал-помеха.

Существуют методы адаптации параметров МСМП под текущую помеховую обстановку [45, 150, 154, 159, 173]. Однако данные методы требуют достаточно большого объема произнесений команд (более 200 [159]) для настройки моделей, что не всегда возможно.

В работе [163] анализируется возможность исключения из результатов распознавания произнесений, не являющихся ГК. Считается, что при наличии подобного произнесения возможное количество кандидатов на результат распознавания больше по сравнению со случаем произнесения ГК.

В работе [177] рассматривается проблема маркировки произнесений слов, которые не включены в командный словарь системы. Маркировка проводится для различения «словарных» и «несловарных» произнесений. Показано, что путем учета результатов маркировки соседних фрагментов речевого сигнала по отношению к анализируемому фрагменту можно значительно уменьшить число ошибок маркировки. В частности, с этой целью используется модель языка.

работе [126] рассматривается метод выделения «ключевых» произнесений, основанный на трехэтапной процедуре. Сначала с помощью МСМП осуществляется распознавание произнесения на уровне отдельных звуков. Полученная последовательность акустических моделей звуков с помощью модели языка преобразуется в «решетку» возможных последовательностей слогов. Полученные последовательности сопоставляются co слоговыми последовательностями ключевых слов. Подсчитывается мера преобразования строки слогов - результатов распознавания в последовательность слогов ключевого слова. Выбираются те ключевые слова, которым соответствуют небольшие значения меры. Далее для каждого слова полученной совокупности подсчитывается мера «доверительности» на основе данных сопоставления произнесения с акустическими моделями. Выбирается слово с максимальной мерой «доверительности».

В работах [139, 198] рассматривается возможность подавления посторонних звуков и слов при распознавании ключевых слов (словосочетаний) в голосовых командах (ГК), передаваемых по телефону. Эта возможность реализуется построением МСМП для паузы (молчания), шумов передачи и посторонней речи. Данная работа развивает подходы, изложенные в работах [182, 199] и наиболее часто цитируется. Авторы являются сотрудниками AT&T Bell Laboratories (США). Основная решаемая работе, задача, В ЭТО поиск подхода, обеспечивающего подавление большого разнообразия посторонних звуков и слов путем построения относительно небольшого числа соответствующих моделей.

Основные параметры предварительной обработки речевого сигнала (РС) следующие. Сигнал ограничивается полосой 100-3200 Гц, частота дискретизации 6,67 кГц. Сигнал проходит затем через предыскажающий нерекурсивный фильтр верхних частот первого порядка (параметр фильтра а=0,95), разбивается на фреймы длительностью 45 мс со сдвигом равным 15мс. Фреймы взвешиваются окном Хэмминга, и проводится LPC-анализ. Для каждого фрейма определяются 12 кепстральных коэффициентов и 12 их первых производных — всего 24 коэффициента.

Типовые характеристики МСМП: 10 состояний, каждое состояние характеризуется взвешенной суммой девяти гауссовых распределений (с диагональными корреляционными матрицами). В модели используется совокупность экспериментально измеренных плотностей вероятностей логарифма энергии и длительности каждого состояния.

В процессе распознавания применяется технология распознавания непрерывной речи с ограничивающей грамматикой (структурой сети моделей), которая в данном случае рассчитана на произвольное встраивание в поток сторонней речи и фонового шума только одного словарного слова. При распознавании используется сетевой алгоритм поиска с учетом длительностей слов и состояний МСМП [166]. Результатом поиска является совокупность последовательностей слов (ключевых и посторонних, а также пауз), которые являются кандидатами для принятия в дальнейшем окончательного решения по распознаванию ключевого слова.

Особенностью обучения системы является создание нескольких моделей как для каждого ключевого слова, так и для каждого слова посторонней речи. Обучение проводится с помощью сегментирующего алгоритма k-средних [159, 184, 205]. Он разделяет предварительно найденные кластеры из обучающих элементов произнесений. Этот алгоритм использует вероятности, полученные при сопоставлении текущей совокупности моделей с произнесениями, и выделяет из

совокупности обучающих элементов те элементы, которые имеют вероятность меньшую некоторого порога. То есть выделяются все обучающие элементы с низкой вероятностью и создается для них новая модель.

При экспериментах рассматривалось наличие следующих пяти ключевых слов (словосочетаний): collect, calling card, third number, person, operator. Обучающие и тестирующие выборки были получены при разговорах абонентов телефонной сети с оператором. Для эксперимента использовалась выборка произнесений абонентов объемом 7980. Примерно половина выборки использовалась для обучения системы, вторая половина – для тестирования.

1483 произнесения были вручную размечены с выделением словарных слов и сторонней речи для создания отдельных баз данных словарных слов и сторонней речи. База данных сторонней речи использовалась для обучения обобщенных моделей «информационного мусора» (посторонней речи). Число уникальных слов посторонней речи составило 447.

В первом эксперименте было принято, что имеется полностью размеченная база речевых данных, которая состоит из словарных слов и указанных 447 посторонних слов. На основе базы данных были созданы МСМП (10 состояний, смесь из 9 компонентов в каждом состоянии) для каждого из *п* наиболее часто встречающихся посторонних слов и шумов (чмоканье губами, щелчки в телефонной линии, шум дыхания). Кроме того, была создана одна модель для фонового шума (10 состояний, смесь из 9 компонентов в каждом состоянии). Исследовалась зависимость точности распознавания от указанного выше числа *п*.

Из данных эксперимента следует, что точность (более 99%) распознавания изолированных слов не зависит от \boldsymbol{n} , если \boldsymbol{n} больше или равно трем (максимальное значение \boldsymbol{n} =13). Это означает, что, в данном случае, модели для словарных слов и для информационного мусора значительно отличаются. Результаты эксперимента показывают, что при \boldsymbol{n} =1 только 11,6% из всех сегментов сторонней речи используются для создания единственной модели

мусора, и, тем не менее, получена высокая точность распознавания изолированных слов (98,9%).

Когда *п* меняется от 2 до 13, то точность системы растет до *n*=3 (что соответствует 30% использования всех сегментов сторонней речи для построения моделей информационного мусора), и далее она практически не меняется (99% для изолированных слов). Созданием МСМП для наиболее часто встречающихся слов и шумов (с добавкой к МСМП фона) достигнута точность распознавания 94,8% для встроенных в шум и постороннюю речь ключевых слов и 99,1% при отсутствии сторонней речи.

Во втором эксперименте были упрощены требования к разметке речи при создании обучающей базы данных. Выделялись лишь словарные слова, а все остальное, включая фон, относилось к посторонним звукам. Создавалась одна универсальная МСМП для всех посторонних звуков. Кроме того, создавалась отдельная модель для фона. Исследовалась зависимость точности распознавания от количества состояний для МСМП фона.

Из результатов эксперимента видно, что при использовании одной МСМП для посторонних слов и звуков и МСМП фона с одним состоянием была достигнута точность распознавания 98,5% для ключевых слов, произносимых изолированно, и 94% - для этих же слов, окруженных посторонними словами и звуками. При увеличении числа состояний МСМП фона до 10 точность распознавания изолированных слов возросла до 99,2%, а при наличии посторонних слов и звуков – до 94,2%. Указывается, что точность распознавания ключевых слов, которые имели малый объем обучающих произнесений, ниже по сравнению со случаем, когда этот объем был большим.

В следующем эксперименте с помощью алгоритма кластеризации создавалась множественная модель для информационного мусора. Выяснилось, что, если увеличить количество МСМП для посторонних слов и звуков до пяти, то точность распознавания увеличивается до 95,1%.

Для дальнейшего повышения качества системы распознавания была проанализирована возможность отложить принятие решения о результатах распознавания. Получены соотношения между вероятностью ошибки распознавания и вероятностью непринятия решения при распознавании команды (использовалось пять моделей посторонних слов). Эти соотношения получены применением порогового теста для значений вероятностей, вычисляемых в процессе распознавания.

Получены следующие результаты: для достижения 98% вероятности правильного распознавания команд, окруженных посторонней речью, необходимо отказаться от распознавания примерно 10% голосовых команд, поступивших на вход системы распознавания. В случае произнесения изолированных команд точность распознавания достигает 99,7% при отказе от распознавания 4% произнесенных команд.

В следующем эксперименте при создании модели молчания и универсальной модели слов посторонней речи для обучения использовалась неразмеченная база речевых данных. Причем произнесение использовалось для обучения модели посторонних слов даже при наличии в нем словарного слова. Для словарных слов использовался орфографический (не сегментный) уровень разметки в каждом обучающем произнесении.

До начала процесса обучения строится набор МСМП из изолированно произнесенных ключевых слов и одна обобщенная модель информационного мусора из общего фонда данных случайной речи (неразмеченная база данных). Для данного набора моделей начальной загрузки и совокупности обучающих данных, которые содержат словарные слова, используется сегментирующий обучающий алгоритм *k*-средних. При этом обучающие последовательности сегментируются в заданные слова, фон и посторонние слова.

Затем создаются новые модели, и процесс повторяется, пока не сойдется (пока разность между значениями вероятности в соседних итерациях не станет

достаточно малой). При использовании этой полностью автоматизированной обучающей процедуры для создания единственной модели посторонних слов была получена точность распознавания отдельно произносимых заданных слов равная 99,4%, а при произнесении их совместно с посторонними словами – 94,5%.

Приведенные результаты сравнимы с результатами, которые были получены при максимально детальном обучении. Это показывает, что модели фона и посторонних слов могут быть успешно использованы для начальной загрузки без детальной транскрипции обучающих файлов.

В последнем эксперименте была сгенерирована единственная модель для сторонних слов без создания отдельной модели фона. Точность распознавания изолированно произнесенных заданных слов составила 99,1% и 94% при наличии посторонних слов. Это показывает, что может быть создана единственная модель для сторонних слов, которая адекватно характеризует как фон, так и посторонние слова.

По результатам анализа данной работы можно сделать следующие выводы:

- Созданием моделей для наиболее часто встречающихся посторонних слов и звуков можно значительно снизить их влияние на результаты распознавания ключевых (словарных) слов в потоке естественной разговорной речи.
- Используемый в работе алгоритм кластеризации позволяет создать небольшое количество множественных моделей посторонних слов и звуков, что повышает точность распознавания.
- Предложенный подход к автоматическому созданию моделей ключевых слов и «информационного мусора» позволяет использовать неразмеченную обучающую базу данных.
- При распознавании ключевых слов в потоке разговорной речи необходимо использовать технологию непрерывного распознавания с ограничивающей грамматикой.

С целью подстройки модели «информационного мусора» под конкретные требования к достоверности распознавания ключевых слов и подавления посторонних слов и звуков в работе [197] предложено использовать аффинное преобразование логарифмической вероятности модели «информационного мусора»: $f_i(s_i) = \alpha_i s_i + \beta_i$, где s_i - логарифмическая вероятность i-ой непрерывной МСМП информационного мусора; α_i , β_i – параметры аффинного преобразования. Данная модель использовалась для подавления как несловарных слов, так и нестационарных помех в телефонном канале при дикторонезависимом распознавании произнесений названий 10 цифр на испанском языке.

Параметры α_i , β_i подбирались так, чтобы минимизировать функцию стоимости ошибок: $C = L_e E_k$, $+ L_r R_k + L_f F_a$, где E_k относительная частота ошибок распознавания ключевых (словарных) слов; R_k относительная частота подавления ключевых слов; F_{a^-} относительная частота ложной тревоги (постороннее слово распознается как ключевое). L_e , L_r , L_f - весовые коэффициенты, зависящие от вероятности появления ключевых слов и «стоимости» соответствующих ошибок распознавания для конкретной сферы использования системы распознавания. При проведении эксперимента проводилось сравнение предложенного подхода с непосредственным использованием модели информационного мусора. В качестве обучающей и тестирующей баз данных использовались записи с телефонной линии. Для обучения использовалась база, содержащая только цифры, а для распознавания — база с «изолированными» и «встроенными» цифрами. Эти базы использовались для определения относительной частоты ошибок распознавания и относительной частоты подавления ключевых слов.

Отдельно была создана база, содержащая неключевые слова и неречевые звуки: кашель, шум дыхания, смех и т.п. Половина этой базы использовалась для обучения (создания модели информационного мусора), а вторая половина — для определения частоты ошибок типа «ложная тревога».

Приведены результаты определения ошибки распознавания ключевых слов, отдельно сказанных (keyword error), и в окружении «информационного мусора» (embedded error), частоты подавления ключевых слов (keyword rejection, embedded rejection), частоты ложной тревоги (false alarm), а также значения функции стоимости (total cost). Рассмотрены два случая работы системы: с режимом barge-in (возможность пользователя оперативно прервать сообщение САРГК) и без него.

Использование предложенного метода позволяет менять значения характеристик распознавания ключевых слов в окружении посторонних слов и звуков. В частности, в режиме barge-in (обработку сигналов можно прервать произнесением командного слова) достигнуто уменьшение частоты ложной тревоги примерно на 4%. По данной работе можно сделать вывод о целесообразности управления работой модели информационного мусора с помощью введенного аффинного преобразования и коэффициентов функции стоимости.

В работе [134] предлагается метод выделения заданных ключевых слов из произвольной разговорной речи, который не использует модели «информационного мусора». Предложенный метод не требует фиксации начала и конца высказывания, допускает наличие нескольких ключевых слов в высказываниях, а также не требует большого объема вычислений.

Вводится понятие энтропии длительности для некоторого пути в решетчатой диаграмме, найденного по алгоритму Витерби в процессе распознавания:

$$dE = \frac{\sum_{j=1}^{J} \frac{L_j}{L} \log \frac{L_j}{L}}{\log J},$$

где L – длина пути, L_j – количество векторов наблюдения, приходящееся на j - ое состояние модели скрытого марковского процесса (МСМП), J – число состояний МСМП.

Показано, что для ключевых (словарных) слов энтропия меньше по сравнению с неключевыми словами. Если энтропия меньше некоторого порога, то принимается решение о наличии ключевого слова. Введение энтропии в процедуру оценки правильности результата распознавания можно рассматривать как появление дополнительного параметра, используемого для подавления неключевых слов.

В работе [163] рассмотрен подход, при котором для подавления посторонних и плохо произнесенных словарных слов используется тот факт, что неключевое (несловарное) или нечетко произнесенное ключевое (словарное) слово при сопоставлении с моделями словарных слов порождает результаты сопоставления, которые мало отличаются друг от друга. Если же словарное слово произнесено четко, то результат сопоставления его со своей моделью значительно отличается от результатов сопоставления с другими моделями.

Предложенный алгоритм применяет технологию распознавания непрерывной речи с использованием «луча поиска». При такой технологии просматривается не весь, а лишь ограниченный перечень наиболее вероятных гипотез - возможных результатов распознавания. Эксперимент проводился на примере распознавания 10 цифр (ключевых слов) и 895 посторонних слов. Вывод следующий: предложенный метод дает примерно такие же результаты, что и при использовании моделей информационного мусора.

В работе [204] (авторы из AT&T Labs – Research) для обнаружения несловарных слов предложено использовать гибридные модели языка, которые содержат не только слова, но и отдельные фоны (звуки) и слоги. При использовании модели языка учитываются условные вероятности появления какого-либо последующего слова после произнесения данного слова. В работе

решается задача обнаружения несловарных слов в разговорной речи в случае использования больших словарей при распознавании. Обычно из-за ограничений по времени распознавания, а также по объему обучающей базы данных в системе распознавания используются модели не всех возможных слов, а только тех, которые наиболее часто встречаются. Поэтому, если при распознавании встречается несловарное слово, то оно приводит к ошибке распознавания. Так как несловарные слова встречаются редко, то общий поток ошибок увеличивается незначительно, и часто с этим смиряются. Однако в ряде случаев (например, при организации поиска по голосовым командам) возникновение таких ошибок недопустимо.

В работе предлагается в модель языка включать словарные слова как таковые, а несловарные - представлять последовательностями фон (звуков) или слогов. Модель языка строится следующим образом. Словарь для распознавания разделяется на две части. В первой части находятся наиболее часто встречаемые слова, а во второй – остальные слова.

Каждое слово из обучающего набора, принадлежащее второй части словаря, представляется в виде цепочки звуков либо слогов. Далее этот трансформированный обучающий набор используется для получения модели языка. Указанные звуки и слоги (с соответствующим произношением) добавляются в произносительный словарь.

Самым простым из предложенных вариантов обнаружения несловарных слов является определение наилучшей гипотезы, генерируемой моделью языка после анализа некоторого высказывания, которая содержит строку звуков (или слогов), и эта строка не соответствует слову в произносительном словаре.

Были проведены три группы экспериментов с разными количествами словарных и несловарных слов. В каждом из проведенных экспериментов использовался словарь объемом 5000, 10000, 20000 слов, соответственно, который включал в себя N часто используемых слов . Каждое слово из часто используемых

в модели языка представлялось отдельно (0-2000, 0-5000, 0-5000), а остальные (2000-5000, 5000-10000, 5000-2000) слова в модели языка представлялись либо строкой звуков (транскрипционных знаков), либо строкой слогов (третья и четвертая колонки, соответственно). При испытании системы распознавались слова из исходного словаря объемом 5000, 10000 и 20000 слов, соответственно.

Фиксировались относительная частота правильно обнаруженных несловарных слов (OOV detection rate – Det.) и относительная частота ложной тревоги (False alarm rate - FAR), когда словарное слово трактовалось как несловарное. В более сложном случае для обнаружения несловарного слова рассчитывается величина

$$C(OOV \mid L) = \sum_{x \in L} p(x \mid L)C(OOV \mid x, L),$$

где C(OOV|x,L) - количество появлений строк фонем (или слогов) в результате прохождения по пути x структуры L возможных путей распознавания (рассматриваются лишь те пути, которые содержат строки фонем (или слогов), не соответствующие ключевым словам); p(x|L) - вероятность пути. Если величина C(OOV|x,L) больше некоторого порога, то считают, что в высказывании присутствует несловарное слово.

Условия тестирования системы: ключевые слова брались в количестве 2000 как наиболее часто встречающиеся из словаря объемом 5000 слов; слова в интервале 2000 – 5000 считались посторонними; к распознаванию предъявлялись фразы, включающие в себя слова из полного объема словаря 5000.

Представление несловарных слов слогами (hybrid-syls) и фонемами (hybrid-phns) дает выигрыш 10 - 15% в точности распознавания по сравнению с системой, где используется лишь одна модель языка на уровне слов (word).

В работе [169] предлагается использовать модель языка для обнаружения несловарных слов при распознавании непрерывной речи в случае использования больших словарей распознаваемых слов. Особенностью предложенного подхода

является совместное использование моделей языка, созданных на уровне звуков и на уровне слов. Каждая модель отображается сложным графом, в котором вершины (узлы) соответствуют моментам (фреймам) времени, а связи - элементам (token) языка (звукам или словам) совместно с совокупностью числовых характеристик, относящихся к этому элементу. Для связи различных фрагментов графа используются вершины связи (transition).

При распознавании некоторого высказывания на основе моделей языка формируются последовательности звуков. Последовательности звуков, сформированные моделью языка на основе звуков, сравниваются с моделью языка на основе слов. В случае отличия последовательностей делается вывод о наличии несловарного слова.

Благодаря использованию графической модели языка и соответствующих инструментов для работы с ней (Graphical Model Tool Kit - GMTK) [138] затраты времени на обнаружение несловарных слов относительно невелики.

В работе [193] рассматривается задача выделения произнесений ключевых слов, используемых в системах связи, в условиях естественной разговорной речи (на примере языка Таиланда). Распознавание слов здесь производилось на фонемном уровне (марковские модели). Для подавления посторонних слов использовалась схема сетевого поиска результата распознавания, которая включает в себя модели ключевых слов, обобщенные (filler) модели, модели слогов. Последние две группы моделей создаются специально для обеспечения подавления посторонних слов. Для ограничения количества моделей слогов использовались лишь наиболее часто встречаемые в терминах, относящихся к процедуре установления телефонной связи.

Обобщенные модели создавались на уровне фонем, которые не входят в ключевые слова. В состав слогов входили также и те слоги, которые использовались ключевыми словами. Использование обобщенных моделей

увеличивает вероятность определения ключевого слова. Однако, если использовать только их, то увеличивается вероятность ложной тревоги.

Так как модели слогов состоят из таких же фонем, что и ключевые слова, то при использовании таких моделей растет вероятность подавления посторонних слов, сходных по звучанию с ключевыми словами, и уменьшается вероятность ложной тревоги. Поэтому совместно с обобщенными моделями используются модели слогов.

При обучении системы использовалось 195 фраз от 49 дикторов-мужчин и 23 дикторов-женщин. В состав фраз были включены 48 ключевых слов. При тестировании системы использовались 250 фраз, произнесенных 20 мужчинами и тремя женщинами без всяких ограничений. Фразы могли быть без ключевых слов, с одним и более ключевыми словами.

Марковские модели состояли из трех состояний. Параметры сигнала включали в себя 39 коэффициентов (мелочастотные кепстральные коэффициенты, их первые и вторые производные). При распознавании фраз с одним ключевым словом использовалась грамматика, которая предусматривала следующие возможные ситуации произнесения команды: последовательное произнесение ключевого и неключевого слов, произнесение только неключевого слова, отсутствие произнесения каких-либо слов (молчание пользователя).

Результаты распознавания фраз с одним ключевым словом следующие. При использовании одной обобщенной модели, модели молчания и 16 моделей слогов получена точность распознавания ключевых слов 69,63%. Данный результат случаем, (69,16%) co когда сравним использовалась известная распознавания со словарем, ограниченным 48 ключевыми словами. На вход системы поступали фразы, набор слов которых ограничивался всего 62 словами, из которых 48 были ключевыми, заданными а остальные второстепенными, используемыми пользователями при установлении связи.

Полученный результат подтверждает эффективность дополнения набора моделей ключевых слов, наборами обобщенных моделей и моделей слогов.

По результатам проведенного анализа существующих методов подавления помех после определения параметров ГК показано, что задача подавления сторонних звуков и речи может решаться на нескольких уровнях функционирования системы автоматического распознавания речи (САРР).

Самым нижним является уровень определения параметров звуковых сигналов, поступающих на вход САРР. На основе отличия параметров речевых и неречевых сигналов можно подавить влияние неречевых звуков на результаты распознавания. На среднем уровне создаются обобщенные (с различной степенью обобщения) МСМП, соответствующие сторонним звукам и словам. Кроме того, анализируются результаты сопоставления произнесенных голосовых команд (ключевых слов), также посторонних слов **ЗВУКОВ** моделями, соответствующими ключевым словам. Показано, что существуют параметры сопоставления, по которым можно отличить ключевые слова от посторонних слов.

На самом верхнем уровне работы САРГК используются грамматики (сети гипотез результатов распознавания) и модели языка применительно к конкретной решаемой задаче распознавания. Здесь используются модели на уровне слов, слогов и фон, а также комбинации этих моделей. Сравниваются результаты распознавания ключевых слов и посторонних слов и звуков при использовании разнородных моделей языка. Показано, что для ключевых слов, входящих в состав словаря заданных САРГК, результаты распознавания более сопоставимы по сравнению с иными словами и звуками.

На основе проведенного анализа существующих методов подавления сторонних звуков и слов можно рекомендовать следующую сетевую структуру акустических моделей (гипотез), реализующую подавление неречевых звуков и произнесений посторонних слов (Рисунок 1.5).

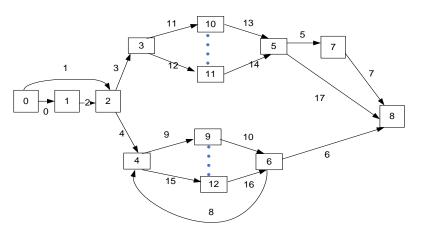


Рисунок 1.5. Сеть моделей, обеспечивающая подавление произнесений неключевых слов.

Данной структуре соответствует перечень узлов и дуг, приведенный в нижеследующей Таблице 1.2.

Таблица1.2. Состав сети моделей, обеспечивающей подавление «посторонних» произнесений

P 0							
N=13 L=18	I=7 W=midsp	J=0 S=0 E=1	J=9 S=4 E=9				
I=0 W=!START	I=8 W=pause2	J=1 S=0 E=2	J=10 S=9 E=6				
I=1 W=pause1	I=9 W=midsp	J=2 S=1 E=2	J=11 S=3 E=10				
I=2 W=!NULL	I=10 W=Ключ. слово1	J=3 S=2 E=3	J=12 S=3 E=11				
I=3 W=!NULL	I=11 W=выдох	J=4 S=2 E=4	J=13 S=10 E=5				
I=4 W=!NULL	I=12 W=k	J=5 S=5 E=7	J=14 S=11 E=5				
I=5 W=!NULL		J=6 S=6 E=8	J=15 S=4 E=12				
I=6 W=!NULL		J=7 S=7 E=8	J=16 S=12 E=6				
		J=8 S=6 E=4	J=17 S=5 E=8				

Здесь каждому узлу соответствует акустическая модель. Дуги показывают возможность перехода между узлами. Первые две колонки - это описания узлов, а последние две - описания дуг.

Предполагается, что каждое слово (ключевое или неключевое) произносится либо с паузой (pause), либо без нее. Вместо произнесения слова возможно появление какой-либо акустической (неречевой) помехи. Все ключевые слова и наиболее часто встречающиеся неречевые помехи (типа шум дыхания) имеют свои акустические модели (например, узел 11). Эти модели подключаются

параллельно узлам 10 и 11. Моделям неключевых слов соответствуют последовательности акустических моделей монофонов (примером служит узел 12). Все необходимые модели монофонов подключаются параллельно узлу 12.

В сетевой модели присутствует обобщенная модель «средней» речи (узел 9 - midsp), которая соответствует звукам, не имеющим своих моделей. Точность ее меньше по сравнению с другими моделями, но она позволяет в какой-то степени учесть произвольный звук. В состав структуры целесообразно также включить «менее обобщенные» модели, каждая из которых представляет группу схожих по артикуляции звуков, наиболее часто встречающихся в «посторонних» произнесениях.

При сопоставлении произнесения с сетевой структурой по алгоритму Витерби определяется сетевой путь, обеспечивающий максимальное значение метрики. Найденный сетевой путь и является результатом распознавания. Следует учесть, что с ростом числа моделей увеличиваются затраты вычислительных ресурсов, используемых при распознавании.

В соответствии с данной рекомендацией проведено экспериментальное исследование эффективности использования сетевой структуры [85]. Использование рекомендованной структуры позволило снизить относительную частоту ложного срабатывания системы на 33,6% по сравнению с использованием лишь одной обобщенной модели «средней» речи.

1.4. Оценка достоверности распознавания голосовых команд

Для пользователя воздействие помех может проявиться в следующей форме: либо при произнесении ГК система ее «не замечает» (имеет место ложный пропуск команды), трактуя ее как помеху, либо произнесенная команда трактуется как другая команда из заданного списка возможных команд. Может возникнуть и другая ситуация: система реагирует на помеху как на ГК – имеет

место ложное срабатывание системы. Рассмотрим существующие подходы для определения параметров достоверности.

В работе [194] рассматривается методика, разработанная в институте стандартов США (National Institute of Standards and Technology - NIST) для оценки качества систем, предназначенных для автоматического поиска информации в архивах звукозаписей речи. Такие системы должны по заданному термину (слову или совокупности слов) найти его звуковое выражение в звукозаписях.

Исходными данными для такой оценки являются относительные частоты P_{Miss} (term, θ) P_{FA} (term, θ) пропуска имеющегося в архивах нужного термина term и ложной тревоги, когда результатом поиска является информация, не соответствующая заданному термину. Обе эти частоты взаимосвязаны порогом θ срабатывания системы, что отражается соответствующим DET (Detect Error Tradeoff) - графиком (равноценен Receiver Operating Characteristic - ROC – графику).

В данной работе указанные частоты определяются следующим образом.

$$P_{Miss}(term, \theta) = 1 - N_{correct}(term, \theta)/N_{true}(term)$$

$$P_{FA}(term, \theta) = N_{spurious}(term, \theta)/N_{NT}(term, \theta)$$

Здесь $N_{correct}$ (term, θ), N_{true} (term, θ), $N_{spuriuos}$ (term, θ), N_{NT} (term, θ) — количества правильных обнаружений термина, истинных присутствий термина в анализируемых архивах звукозаписей, ложных обнаружений термина, благоприятных ситуаций для ложного обнаружения, соответственно. Здесь

$$N_{NT}(term, \theta) = n_{tps} \cdot T_{speec h} - N_{true}(term, \theta),$$

где n_{tps} — число попыток распознавания речи, производимых за одну секунду (далее эта величина принимается равной единице); T_{speech} — объем речевого материала в секундах. Указывается на некоторую произвольность определения N_{NT} (term, θ).

Частоты, соответствующие определенному порогу θ , рассчитываются для каждого термина и затем усредняются на заданной группе терминов.

$$P_{Miss}(\theta) = average_{term} \{P_{Miss}(term, \theta)\}$$

 $P_{FA}(\theta) = average_{term} \{P_{FA}(term, \theta)\}$

Частоты P_{Miss} (term, θ), P_{FA} (term, θ) усредняются лишь по тем терминам, произнесения которых присутствуют в звуковом материале, поэтому частота P_{Miss} (term, θ) всегда существует.

Оценка качества системы обнаружения термина производится применительно к конкретной прикладной задаче путем задания ценности V получения правильного результата работы системы и задания величины C издержек в случае неправильного результата. Используется два варианта определения итоговой ценности Value системы. Первый и второй варианты определяются следующими выражениями, соответственно.

$$\begin{split} Value_o(\theta) &= \frac{\sum_{term} \{V \cdot N_{correct} \ (term, \theta) - C \cdot N_{spirious} \ (term, \theta)\}}{\sum_{term} \{V \cdot N_{true} \ (term)\}} \\ Value_T(\theta) &= 1 - average_{term} \{P_{Miss}(term, \theta) + \beta \cdot P_{FA}(term, \theta)\}. \\ \beta &= \frac{C}{V} \cdot (Pr_{term}^{-1} - 1), \end{split}$$

где Pr_{term} — априорная вероятность появления термина. Рекомендуется $Pr_{term}=10^{-4}$, и C/V=0.1. Если нет пропусков обнаружения термина и нет ложных срабатываний, то система является идеальной, и ее ценность Value=1.

Как отмечается в работе, достоинством первого варианта оценки качества системы является его ориентированность на конкретную область использования системы. Достоинством второго варианта является учет априорной вероятности появления термина. При этом подчеркивается важность правильного распознавания терминов, редко встречающихся в анализируемом звуковом материале.

Тестирующие выборки звуковых файлов сгруппированы по трем языкам (английский, арабский, китайский) и по типу (телефонные разговоры, радионовости, «беседы за круглым столом»). Следует отметить детальность

описания и проработанность тех материалов и документов, в которых фиксируются результаты испытаний систем.

В работе [193] при тестировании систем выделения ключевых слов используются две оценки: "Precision rate (оценка точности)" P=C/D и "Recall rate (оценка правильности)" R=C/A. Здесь C – количество правильных обнаружений ключевых слов, D – общее число срабатываний системы (сумма правильных и неправильных обнаружений ключевых слов), A – общее количество фактически произнесенных ключевых слов.

Чем выше значения указанных оценок, тем выше качество системы. В отличие от рассмотренных выше оценок данные оценки подразумевают тестирование системы изолированно произносимыми словами.

В работе [204] при тестировании систем обнаружения несловарных слов используются две пары оценок. Первая пара: "ООV-detection" (обнаружение несловарных слов) V=C1/A1 и "False Alarm" (ложная тревога) FA=NC/B. Здесь C1 — число произнесений с неключевыми словами, определенных правильно, A1 — общее количество произнесений неключевых слов, NC — количество произнесений неключевых слов, определенных неправильно, В — число высказываний без ключевых слов.

Вторая пара оценок. "Precision (оценка точности)" P=C1/L и "Recall (оценка правильности)" = "OOV-detection" (обнаружение несловарных слов). Здесь L - общее количество высказываний, определенных как имеющие неключевые слова. Чем выше значения указанных оценок, тем выше качество системы. Данные оценки также подразумевают тестирование системы изолированно произносимыми словами.

На основе вышеизложенного можно выделить следующие первичные данные, полученные при тестировании САРГК достаточно представительной по составу дикторов и произносимых слов, характеризующие помехоустойчивость.

 N_{all} – число произнесений всех слов в звукозаписях,

 N_{kw} – число произнесений ключевых слов в звукозаписях,

 $N_{nkw} = N_{all}$ - N_{kw} — число произнесений неключевых слов в звукозаписях,

 N_{det} - число правильных обнаружений ключевых слов в звукозаписях,

 N_{insert} — число ложных срабатываний системы поиска на произнесения неключевых слов (число вставок),

 $N_{replace}$ — число ложных срабатываний системы поиска, когда одно ключевое слово заменяется другим ключевым словом,

 $N_{scip} = N_{kw}$ - N_{det} - число пропусков произнесений ключевых слов в звукозаписях.

На основе первичных данных тестирования определяются параметры помехоустойчивости.

Относительная частота ложного срабатывания $P_{FA} = N_{insert}/N_{nkw}$.

Относительная частота ложного подавления (пропуска) ключевого слова

$$P_{FR} = N_{scip} / N_{kw}$$

Относительная частота правильного распознавания $P_{correct} = N_{det}/N_{kw}$.

Относительная частота замены одного ключевого слова другим ключевым словом из заданного списка ключевых слов $P_{replace} = N_{replace} / N_{kw}$.

Оценка устойчивости САРГК к ложным срабатываниям зависит от состава тестовой выборки звукозаписей. Трудно разделить «вклады» собственно системы распознавания и тестовой выборки в относительную частоту ошибки типа «ложная тревога».

Выводы по разделу 1

1. По месту возникновения помехи, воздействующие на САРГК, можно разделить на следующие группы: помехи, создаваемые диктором, помехи от акустического окружения диктора, помехи от аппаратуры связи. По характеру воздействия на САРГК помехи целесообразно разделить на два класса: помехи, находящиеся вне временного интервала произнесения команды, и

помехи, попадающие в данный временной интервал. Помехи, находящиеся вне интервала ГК, вызывают ложное срабатывание системы, а помехи, попадающие в интервал ГК, обусловливают ошибочный результат ее распознавания.

- 2. Для подавления помех, находящихся вне интервала ГК, в состав сетевой структуры моделей ГК дополнительно вводят модели помех. Модели помех, в частности, включают в себя сетевую структуру моделей монофонов, обеспечивающую подавление произнесений посторонних слов, модели конкретных типов помех, обобщенные модели помех, соответствующие нескольким типам помех, для которых не создавались отдельные модели.
- 3. Для подавления помех, попадающих в интервал ГК, используют модели команд, в которых учтено наличие помехи.
- 4. Для обеспечения высокой достоверности распознавания целесообразно подавление помех с помощью акустических моделей дополнять использованием средств борьбы с помехами на различных стадиях прохождения сигнала по каналу связи.
- 5. Для характеристики помехоустойчивости систем используют относительные частоты ложного срабатывания, ложного пропуска и замены ГК другой командой. Они являются оценками соответствующих вероятностей. Чтобы обеспечить заданную погрешность оценок, необходимо при росте достоверности распознавания увеличивать объем тестовой звуковой базы, что приводит к росту затрат на разработку системы.
- 6. Величина относительной частоты ложного срабатывания зависит от состава тестовой выборки звукозаписей звуковой базы. Из-за отличий звукозаписей тестовой базы от произнесений, распознаваемых при эксплуатации системы, требуется настройка системы на условия эксплуатации.

2. Сокращение объема экспериментов при определении параметров достоверности распознавания команд

При проектировании САРГК приходится сравнивать по достоверности различные варианты построения системы, принимая решение о выборе варианта, обеспечивающего наибольшую достоверность. Так как оценки параметров достоверности являются случайными величинами, то возможно принятие ошибочного решения при сравнении систем, что увеличивает затраты времени на проектирование.

Как указывалось выше, величина относительной частоты FAR ложного срабатывания системы в значительной степени зависит от состава произнесений слов тестовой звуковой базы. Поэтому требуются дополнительные затраты времени на подстройку параметров системы при вводе ее в эксплуатацию.

В данном разделе рассматриваются методы оценки параметров достоверности, обеспечивающие, во-первых, уменьшение вероятности принятия ошибочного решения при сравнении систем, а во-вторых, снижение зависимости FAR от состава произнесений слов тестовой звуковой базы.

2.1. Анализ метрики сопоставления параметров речевого сигнала с акустическими моделями

Пусть после окончания процедуры распознавания произнесению была сопоставлена некоторая модель ГК. Считаем, что часть слитно произнесенных подпоследовательности **ЗВУКОВ** соответствует некоторой непрерывной акустических моделей звуков, входящих в состав модели ГК. Рассмотрим следующий пример. Пусть **ЗВУКИ** модели ГК описываются некоторой последовательностью 14 знаков транскрипции (то есть в составе модели ГК имеется 14 моделей звуков):

abcdefghijklmn.

Пусть звуки произнесения описываются следующей последовательностью такого же количества звуков:

Знак подчеркивания означает «не этот звук». Считаем, что каждый произнесенный звук сопоставлен с соответствующей моделью. Результат сопоставления произнесения с моделью можно отобразить следующей последовательностью символов:

00001111110000

Каждый символ соответствует результату сопоставления произнесенного звука с некоторой моделью звука. Если звук соответствует модели, то имеет место символ единицы. В противном случае присутствует символ нуля. При сопоставлении звука с монофонной моделью возможно лишь два варианта результата сопоставления: соответствует звук модели или не соответствует. Обозначим через $e_{m,j}$ метрику сопоставления для первого результата, а через $d_{m,j}$ — метрику второго варианта. Здесь индексы m, j обозначают соответственно монофонный тип модели и j-ый номер звука в произнесении.

Если звук полностью соответствует трифонной модели с учетом контекстов, то имеет место метрика $e_{tRL,j}$. Индекс RL обозначает, что слева и справа звук окружают звуки, которые соответствуют его трифонной модели. Если звук полностью не соответствует модели, то имеет место метрика $d_{t,j}$. Если модель и звук соответствуют друг другу, но отличаются по одному из контекстов (левому или правому) или по обоим контекстам, то таким ситуациям соответствуют метрики: $e_{tL,j}$, (или $e_{tR,j}$), $e_{t,j}$. Метрика сопоставления всего произнесения с моделью ГК равна сумме метрик сопоставления звуков с соответствующими моделями. Для данного примера метрика

$$M = d_{m,1} + \sum_{j=2}^{4} d_{t,j} + e_{tR,5} + \sum_{j=6}^{9} e_{tRL,j} + e_{tL,10} + \sum_{j=11}^{13} d_{t,j} + d_{m,14}$$

Рассмотрим теперь более общий случай, когда число нулей и единиц в результате сопоставления произнесения с моделью ГК является произвольным. Обозначим через E, L, R количество единиц, количество нулей слева от единиц и количество нулей справа от единиц. Количество случаев полного соответствия звука трифонной модели

$$n_{e,t} = f_{e,t}(E) = \begin{cases} E - 2 npu E \ge 3 \\ E - 2 npu R(L) = 0, L(R) \ne 0 \end{cases}$$

Здесь выражение типа R(L) означает «R или L». Число случаев полного несоответствия звука трифонной модели

$$n_{d,t} = f_{d,t}(L) + f_{d,t}(R), \text{ ede } f_{d,t}(x) = \begin{cases} x - 1 & npu \ x \ge 0 \\ 0 & npu \ x = 0 \end{cases}$$

Суммарное число случаев несоответствия звука трифонной модели только по одному контексту

$$n_{e,t(R+L)} = f_{e,tR}(L) + f_{e,tL}(R), coe f_{e,t}(x) = \begin{cases} 1 & npu \ x > 0 \\ 0 & npu \ x = 0 \end{cases}$$

Число случаев несоответствия звука трифонной модели по двум контекстам

$$n_{e,t} = f_{e,t}(E) = \begin{cases} 1 & npu \ E = 1 \\ 0 & npu \ E > 1 \end{cases}$$

Число случаев несоответствия звука монофонной модели

$$n_{d,m} = f_{d,m}(L) + f_{d,m}(R), \text{ ode } f_{d,m}(x) = \begin{cases} 1 & npu & x > 0 \\ 0 & npu & x = 0 \end{cases}$$

Число случаев соответствия звука монофонной модели

$$n_{e,m} = f_{e,m}(L) + f_{e,m}(R)$$
, so $f_{e,m}(x) = \begin{cases} 1 & npu \ x = 0 \\ 0 & npu \ x > 0 \end{cases}$

Рассмотрим теперь частный случай сопоставления i-ой ГК со «своей» моделью, которая содержит W моделей звуков. В этом случае L=R=0, и метрика сопоставления матрицы $\mathbf{X_i}$ параметров произнесения i-ой ГК, образованной последовательностью векторов параметров, с «трифонным» набором $\Lambda_{t,i}$ моделей звуков данной команды

$$M(\mathbf{X_i}|\Lambda_{t,i}) = e_{m,1} + \sum_{j=2}^{W-1} e_{t,j} + e_{m,W}$$

Рассмотрим теперь результат сопоставления ГК с «чужой» моделью – моделью другой ГК. Считаем, что число звуков в произнесении равно числу моделей звуков в модели команды. Рассмотрим наиболее вероятный случай: E>2, R>1, L>1. Метрика сопоставления матрицы параметров произнесения і-ой ГК с трифонной моделью k-ой ГК, состоящей из набора трифонных и монофонных моделей звуков данной ГК

$$M(\mathbf{X_i}|\Lambda_{t,k}) = d_{m,1} + \sum_{j=2}^{L} d_{t,j} + e_{tR,L+1} + \sum_{j=L+2}^{L+E-1} e_{tRL,j} + e_{tL,L+E} + \sum_{j=L+E+1}^{L+E+R-1} d_{t,j} + d_{m,L+E+R}$$

Рассмотрим теперь случай сопоставления ГК с последовательностью монофонных моделей звуков. Считаем, что каждому из W звуков сопоставлена соответствующая модель. Метрика сопоставления матрицы параметров произнесения i-ой ГК с монофонной моделью i-ой ГК, состоящей из набора только монофонных моделей звуков данной ГК

$$M(\mathbf{X_i}|\Lambda_{m,i}) = \sum_{j=1}^{W} e_{m,j}$$

В общем случае при сопоставлении ГК с монофонными моделями количество сопоставленных по итогам распознавания моделей звуков может отличаться от числа звуков ГК. Это отличие обусловлено небольшой точностью монофонных моделей и изменчивостью произнесения ГК. Следует отметить, что в данном случае метрика окажется больше метрики, полученной согласно формуле, так как метрика формируется по критерию максимального правдоподобия.

Решение о результате распознавания произнесения принимается путем определения модели, обеспечивающей максимальную метрику сопоставления с произнесением. Рассмотрим разности метрик, которые определяются при произнесении ГК. При произнесении ГК происходит сопоставление произнесения с моделями всех ГК из заданного перечня, а также с обобщенными и монофонными моделями.

Рассмотрим разность метрик сопоставления ГК со «своей» трифонной и монофонной моделями, считая, что каждый звук ГК сопоставляется с соответствующей моделью.

$$\Delta M_{t-m,ii} = M(\mathbf{X_i}|\Lambda_{t,i}) - M(\mathbf{X_i}|\Lambda_{m,i}) = e_{m,1} + \sum_{j=2}^{W-1} e_{t,j} + e_{m,W} - \sum_{j=1}^{W} e_{m,j} = \sum_{j=2}^{W-2} (e_{t,j} - e_{m,j}) = \sum_{j=2}^{W-2} \Delta e_{t-m,j}.$$

Здесь $\Delta e_{t-m,j}$ - разность метрик сопоставления матрицы параметров j-го звука с трифонной и монофонной моделями. В процессе тестирования САРГК распознаванию подвергается множество произнесений ГК, поэтому величина $\Delta M_{t-m,ii}$ является случайной величиной равной сумме случайных величин $\Delta e_{t-m,j}$, которые зависят от случайных вариаций длительности и тембра произнесения j-го звука. Можно считать, что ее распределение $p(\Delta M_{t-m,ii})$ плотности вероятностей близко нормальному закону. Если $\Delta M_{t-m,ii} > 0$, то произнесенная ГК распознается верно. Если же $\Delta M_{t-m,ii} < 0$, то ГК трактуется как помеха - имеет место ошибка распознавания.

Математическое ожидание $\mu_{t-m,ii}$ и дисперсия $\sigma^2_{t-m,ii}$ величины $\Delta M_{t-m,ii}$ равны сумме математических ожиданий и дисперсий случайных величин $\Delta e_{t-m,j}$.

$$\mu_{t-m,ii} = \sum_{j=2}^{W-2} \mu_{e,t-m,ij}, \quad \sigma_{t-m,ii}^2 = \sum_{j=2}^{W-2} \sigma_{e,t-m,ij}^2, \qquad (2.1)$$

где $\mu_{e,t-m,ij}$, $\sigma_{e,t-m,ij}^2$, математическое ожидание и дисперсия величины $\Delta e_{t-m,j}$. Учтем, что $\mu_{e,t-m,ij} > 0$, так как модель трифона, в среднем, обеспечивает большую метрику сопоставления со звуком по сравнению с моделью монофона. Следовательно, вероятность того, что ГК будет распознана как постороннее произнесение

$$P_{FR,i} = \int_{-\infty}^{0} p(\Delta M_{t-m,ii}) d(\Delta M_{t-m,ii}) = 0.5 - \Phi(\alpha) , \qquad (2.2)$$

где $\Phi(\alpha) = \frac{1}{\sqrt{2\pi}} \int_0^\alpha e^{-z^2/2} dz$ - табулированная функция Лапласа,

 $\alpha = \mu_{t-m,ii}/\sigma_{t-m,ii}$. С учетом (2.1), видно, что с ростом числа W звуков в ГК растет величина α , что приводит к уменьшению вероятности $P_{FR,i}$ ошибки распознавания.

Рассмотрим разность метрик сопоставления ГК со «своей» и «чужой» трифонными моделями, считая, что число звуков в моделях одинаково, и каждый звук ГК сопоставляется с соответствующей моделью.

$$\Delta M_{t-t,ik} = M(\mathbf{X_i}|\Lambda_{t,i}) - M(\mathbf{X_i}|\Lambda_{t,k}) = e_{m,1} + \sum_{j=2}^{W-1} e_{t,j} + e_{m,W} - d_{m,1} - \sum_{j=2}^{L} d_{t,j} - e_{tR,L+1} - \sum_{j=L+2}^{L+E-1} e_{tRL,j} - e_{tL,L+E} - \sum_{j=L+E+1}^{L+E+R-1} d_{t,j} - d_{m,L+E+R} =$$

$$= (e_{m,1} - d_{m,1}) + \sum_{j=2}^{L} (e_{tRL,j} - d_{t,j}) + (e_{tRL,L+1} - e_{tR,L+1}) + \sum_{j=L+2}^{L+E-1} (e_{tRL,j} - e_{tRL,j}) + (e_{tRL,L+E} - e_{tL,L+E}) + \sum_{j=L+E+1}^{L+E+R-1} (e_{tRL,j} - d_{t,j}) + (e_{m,W} - d_{m,L+E+R})$$

Учитывая одинаковость моделей звуков в интервале позиций от L+2 до L+E-1, получаем

$$\Delta M_{t-t,ik} = \left(e_{m,1} - d_{m,1}\right) + \sum_{j=2}^{L} (e_{tRL,j} - d_{t,j}) + \left(e_{tRL,L+1} - e_{tR,L+1}\right) +$$

$$+(e_{tRL,L+E}-e_{tL,L+E})+\sum_{j=L+E+1}^{L+E+R-1}(e_{tRL,j}-d_{t,j})+(e_{m,L+E+R}-d_{m,L+E+R})$$

В полученном выражении фигурируют три типа разности метрик: $e_m - d_m$, $e_{tRL} - d_t$, $e_{tRL} - d_{tR(L)}$. Первый и третий типы разностей встречаются по два раза, второй тип — L+R-2 раз. В процессе тестирования САРГК эти разности принимают случайные значения. Следовательно, разность $\Delta M_{t-t,ik}$ также является случайной величиной. Так как она состоит из суммы примерно равноценных случайных величин, то целесообразно считать ее гауссовой.

Ошибочный результат распознавания имеет место, когда $\Delta M_{t-t,ik} < 0$. Так как математические ожидания разностей метрик положительны, и самым большим является математическое ожидание разности типа $e_{tRL} - d_t$, то вероятность ошибки распознавания уменьшается, главным образом, за счет увеличения суммы L+R-2=W-E-2.

Рассмотрим теперь ситуацию, когда на вход системы распознавания поступает произнесение постороннего слова. Если метрика сопоставления данного произнесения с совокупностью монофонных моделей его звуков больше метрики сопоставления с моделью k-го ключевого слова, то имеет место правильное распознавание — система относит это произнесение к числу помех. В противном случае имеет место ошибка ложного срабатывания системы. Рассмотрим разность соответствующих метрик.

$$\Delta M_{m-t,ik} = M(\mathbf{X_i}|\Lambda_{m,i}) - M(\mathbf{X_i}|\Lambda_{t,k}) = \sum_{j=1}^{W} e_{m,j} - d_{m,1} - \sum_{j=2}^{L} d_{t,j} - e_{tR,L+1} - \sum_{j=L+2}^{L+E-1} e_{tRL,j} - e_{tL,L+E} - \sum_{j=L+E+1}^{L+E+R-1} d_{t,j} - d_{m,L+E+R} =$$

$$= (e_{m,1} - d_{m,1}) + \sum_{j=2}^{L} (e_{m,j} - d_{t,j}) + (e_{m,L+1} - e_{tR,L+1}) +$$

$$+\sum_{j=L+2}^{L+E-1}(e_{m,j}-e_{tRL,j})+(e_{m,L+E}-e_{tL,L+E})+\sum_{j=L+E+1}^{L+E+R-1}(e_{m,j}-d_{t,j})+\\+(e_{m,L+E+R}-d_{m,L+E+R})$$

В данном выражении присутствуют следующие типы разностей: e_m-d_m , e_m-d_t , $e_m-e_{tR(L)}$, e_m-e_{tRL} . Все эти разности являются случайными величинами. Математические ожидания разностей первого и второго типов имеют положительные значения, четвертого — отрицательное. Математическое ожидание разности третьего типа может быть как отрицательным, так и положительным. Разности первого и третьего типов в полученном выражении встречаются по два раза, второго типа — (L+R-2)=(W-E-2) раз, четвертого типа — E-2 раз.

Видно, что результат распознавания зависит, в основном, от соотношения сумм математических ожиданий разностей второго и четвертого типов. Чем больше величина E-2 и меньше величина W-E-2, тем больше вероятность ложного срабатывания системы. Если рассматривать наиболее вероятную ситуацию, когда сумма математических ожиданий всех разностей, определяющих метрику $\Delta M_{m-t,ik}$, больше нуля, то вероятность ложного срабатывания системы определяется ранее полученным выражением (2.2), где величина α является отношением указанной суммы математических ожиданий к суммарному среднеквадратическому отклонению.

Рассмотрим теперь случай, когда результат сопоставления произнесения с «трифонной» акустической моделью характеризуется наличием не одной группы E, а двух групп E_1 и E_2 :

$$0 \ 0 \dots 0 \ 1 \ 1 \dots 1 \ 0 \ 0 \dots 0 \ 1 \ 1 \dots 1 \ 0 \ 0 \dots 0$$
 (2.3)

Здесь помимо нулевых групп L и R появляется еще одна «серединная» нулевая группа с количеством нулей равным М. В этом случае количество случаев полного соответствия звука трифонной модели

$$n_{e,t} = f_{e,t}(E_1) + f_{e,t}(E_2)$$

Число случаев полного несоответствия звука трифонной модели

$$n_{d,t} = f_{d,t}(L) + f_{d,t}(R) + M$$

Суммарное число случаев несоответствия звука трифонной модели только по одному контексту

$$n_{e,t(R+L)} = f_{e,tR}(L) + f_{e,tL}(R) + 2$$

Число случаев несоответствия звука трифонной модели по двум контекстам

$$n_{e,t} = f_{e,t}(E_1) + f_{e,t}(E_2)$$

Просуммируем число единиц в схеме (2.3) сопоставления произнесения с моделью, а число нулей в группе М разделим на две группы и прибавим их к группам L и R. Тогда величина $n_{e,t}$, полученная для случая с одной группой типа E, должна быть уменьшена на 2, величины: $n_{d,t}$, $n_{e,t}$ не изменяются, величина $n_{e,t(R+L)}$ увеличивается на 2. Можно утверждать, что если преобразовать схему (2.3) с двумя группами типа E, то результаты вычислений, соответствующие одной группе типа E, будут мало отличаться от результатов, соответствующих двум группам типа E. Можно сделать следующий вывод, что схемы вида (2.3) сопоставления произнесения с моделью полностью определяют количество различных вариантов сопоставления произнесений звуков с моделями звуков, и тем самым во многом определяют результат вычисления метрик, а следовательно, результат распознавания.

Рассмотрим теперь случай, когда число звуков постороннего произнесения отличается от числа звуков ГК. Пусть $[\underline{v_1} \ \underline{v_2} \ \underline{v_3} \ v_4 \ v_5 \ \underline{v_6} \ v_7 \ \underline{v_8} \ \underline{v_9} \ v_{10}]$ и $[\underline{n_1} \ \underline{n_2} \ \underline{n_3} \ n_4 \ n_5 \ n_6 \ n_7 \ \underline{n_8} \ n_9 \ \underline{n_{10}} \ \underline{n_{11}} \ n_{12}]$ — транскрипции ГК и постороннего произнесения, соответственно. Совпадающие знаки транскрипции подчеркнуты. Каждому знаку

транскрипции ГК соответствует модель. Все модели, кроме крайних, являются трифонными, а крайние — монофонными. Если модели звуков достаточно точные, то звук $\underline{\mathbf{n}}_1$ сопоставляется с монофонной моделью звука $\underline{\mathbf{v}}_1$. Аналогично звук $\underline{\mathbf{n}}_2$ сопоставляется со «своей» трифонной моделью звука $\underline{\mathbf{v}}_2$. Звуки: \mathbf{n}_4 \mathbf{n}_5 \mathbf{n}_6 \mathbf{n}_7 сопоставляются с «чужими» трифонными моделями. Звук $\underline{\mathbf{n}}_8$ сопоставляется со «своей» трифонной моделью звука $\underline{\mathbf{v}}_6$. Однако его окружение с обеих сторон не соответствует трифонной модели звука $\underline{\mathbf{v}}_6$. Рассуждая аналогично, для остальных звуков постороннего произнесения, получим суммарную метрику сопоставления произнесения данного і-го постороннего слова с моделью данной j-ой ГК

$$M_{i\Gamma K j} = e_{m1} + \ e_{tRL2} + \ e_{tL3} + \ d_{t4} + \ d_{t5} + \ d_{t6} + \ d_{t7} + \ e_{t8} + \ d_{t9} + \ e_{tR10} + \ e_{tL11} + \ d_{m12}.$$

Метрика сопоставления данного произнесения с последовательностью монофонных моделей соответствующих звуков M_{im} = e_{m1} + e_{m2} +...+ e_{m12} . Разность метрик: $\Delta M_{im\text{-}\Gamma Kj}$ = M_{im} - $M_{i\Gamma Kj}$ = $(e_{m1}$ - $e_{m1})$ + $(e_{m2}$ - $e_{tRL2})$ + $(e_{m3}$ - $e_{tL3})$ + $(e_{m4}$ - $d_{t4})$ + $(e_{m5}$ - $d_{t5})$ + $(e_{m6}$ - $d_{t6})$ + $(e_{m7}$ - $d_{t7})$ + $(e_{m8}$ - $e_{t8})$ + $(e_{m9}$ - $d_{t9})$ + $(e_{m10}$ - $e_{tR10})$ + $(e_{m11}$ - $e_{tL11})$ + $(e_{m12}$ - $d_{m12})$.

Так как величина метрики сопоставления зависит от степени соответствия произнесения звука модели, то математическое ожидание величины $\Delta M_{\text{im-}\Gamma Kj}$ состоит из суммы отрицательных и положительных слагаемых. К положительным слагаемым относятся математические ожидания, соответствующие следующим типам разностей метрик: $(e_m - d_t)$, $(e_m - d_m)$. Наибольшей является разность $(e_m - d_t)$. Отрицательные слагаемые: $(e_m - e_{tRL})$, $(e_m - e_{tL})$, $(e_m - e_{tR})$, (

Чем больше метрик типа d и меньше метрик типа e, тем больше отношение α в выражении (2.2) и, следовательно, меньше вероятность ложного срабатывания. Соотношение чисел метрик различных типов характеризует вероятность ложного срабатывания. Если предположить что метрики типов e_{tR} , e_{tL}

примерно равны, то их можно объединить, тогда для рассмотренного примера произнесение постороннего слова относительно данной ГК можно охарактеризовать следующей комбинацией типов метрик (Таблица 2.1).

Таблица 2.1. Пример комбинации типов разностей метрик

e_{tRL}	$e_{t(R)L}$	e_t	$e_{\rm m}$	$d_{\rm m}$	d_t
1	3	1	1	1	5

Набор чисел во второй строке таблицы указывает количество разностей соответствующего типа, и его можно трактовать как код метрики, который определяет вероятность ложного срабатывания.

Проведенный анализ метрики сопоставления произнесения с моделями команд и помех показывает, что вероятности ложного срабатывания системы и распознавания произнесенной данной ГК в виде другой команды во многом определяются числом звуков произнесения, которые соответствуют трифонным моделям звуков, входящим в состав какой-либо модели ГК.

2.2. Оценка вероятности ложного срабатывания на произнесение постороннего слова

2.2.1. Метод оценки вероятности

Материалы данного подраздела отражены в публикациях [70, 71, 85, 107]. Так как при произнесении команды непроизвольно произносятся и посторонние слова, то возможно ложное срабатывание системы автоматического распознавания голосовых команд (САРГК) на данные произнесения. Решение о результате распознавания принимается с учетом сопоставления произнесений слов не только с моделями команд, но и с моделями посторонних слов – моделями помех. Если метрика сопоставления произнесения с моделью ГК (ключевого слова) больше метрики сопоставления с моделью помехи, то принимается решение «произнесено ключевое слово».

Однако использование моделей помех может привести к тому, что произнесенная команда трактуется как постороннее слово, и система на нее не реагирует – имеет место ложное подавление (пропуск) команды. Совокупность вероятностей ложного срабатывания (ложной тревоги) и ложного подавления характеризует качество САРГК.

Указанные показатели качества определяют при тестировании системы на достаточно представительной по составу слов и дикторов звуковой базе. При заданной относительной частоте подавления (пропуска) ключевого слова (false reject rate - FRR) фиксируется относительная частота ложного срабатывания (ложной тревоги (false alarm rate - FAR)) системы на произнесение неключевого слова. Частоты FRR и FAR являются точечными оценками вероятностей ложного срабатывания системы и ложного пропуска ключевого слова.

Параметр FAR обычно определяется как отношение числа ложных тревог к числу произнесений неключевых слов. В условиях слитной речи, когда ложное срабатывание может возникнуть на стыке произнесений слов, под количеством произнесений неключевых слов обычно понимают величину равную отношению длительности тестирующей звуковой базы за вычетом общей длительности произнесений ключевых слов к средней длительности слова в базе [194].

Величина FAR имеет сильную зависимость от количества фрагментов звуковой базы, которые созвучны ключевому слову, но к нему не относятся. Поэтому величину FAR используют лишь для сравнения систем на заданной звуковой базе. Нет возможности сравнивать системы, если их тестирование проводилось на разных базах. Следует отметить высокую стоимость тестовых баз, что ограничивает число баз у разработчика систем.

Ряд вопросов возникает и при анализе существующего способа подсчета числа ложных тревог при тестировании системы. Например, при ключевом слове «вопрос» система сработала на произнесение слова «вопросами», которое содержит все звуки ключевого слова. Считать ли ложной тревогой срабатывание

системы на произнесение данной словоформы ключевого слова? Необходимо оговорить условия, при которых результат распознавания следует считать ложной тревогой.

Второй вопрос возникает при подсчете числа ложных тревог в условиях сильной «созвучности» неключевого слова с ключевым. Например, результатом распознавания является слово «опрос» при наличии ключевого слова «вопрос». Считать данный случай ложной тревоги равноценным ошибочному срабатыванию системы, например, на менее созвучное слово «Россия», наверное, неправильно.

В данном разделе предлагается и анализируется метод определения числа фрагментов тестовой звуковой базы, которые могут вызвать ложное срабатывание системы и поэтому их можно трактовать как произнесения неключевых слов при оценке вероятности ложного срабатывания системы [55, 70, 71]. Использование метода позволяет снизить зависимость оценки вероятности ложного срабатывания от состава тестовой звуковой базы. Кроме того, анализируется зависимость вероятности ложного срабатывания системы от степени отличия произнесения неключевого слова от акустической модели ГК. Данная зависимость позволяет более точно оценить устойчивость системы к воздействию произнесений посторонних слов.

Анализ ложных срабатываний, проведенный по результатам экспериментов, показал наличие в произнесениях, которые вызвали ложное срабатывание, усеченной и прореженной последовательности звуков ключевого слова, в которую могут быть вставлены звуки неключевого слова. Предлагается для каждого ключевого слова выделять различные структуры помех - фрагментов базы, которые вызвали ложное срабатывание системы, и для каждой структуры определять количество фрагментов базы, которые могли бы вызвать ложное срабатывание системы.

На основе изложенного в предыдущем подразделе материала целесообразен следующий порядок оценки вероятности ложного срабатывания.

Для заданной ГК при тестировании определяются те фрагменты слитно произносимых слов звуковой базы, которые вызвали ложное срабатывание САРГК. Для каждого фрагмента определяется структура произнесений, которые могут вызвать ложное срабатывание. В структуре указываются те звуки произнесения (транскрипционные знаки), которые совпали со знаками транскрипции ГК. Кроме того, помечаются те знаки произнесения, которые отличаются от соответствующих знаков транскрипции ГК.

Пусть $[\underline{v_1} \ \underline{v_2} \ \underline{v_3} \ v_4 \ v_5 \ \underline{v_6} \ v_7 \ \underline{v_8} \ \underline{v_9} \ v_{10}]$ и $[\underline{n_1} \ \underline{n_2} \ \underline{n_3} \ n_4 \ n_5 \ n_6 \ n_7 \ \underline{n_8} \ n_9 \ \underline{n_{10}} \ \underline{n_{11}} \ n_{12}]$ – транскрипции ГК и постороннего произнесения, соответственно. Совпадающие знаки транскрипции подчеркнуты. Для рассматриваемого примера структура произнесения, которое вызвало ложное срабатывание, выглядит следующим образом: $[\underline{v_1} \ \underline{v_2} \ \underline{v_3} \ ***** \ \underline{v_6} \ * \ \underline{v_8} \ \underline{v_9} \ *]$. Здесь знаком * помечены несовпадающие знаки. Если знак * заменить нулем, то структура примет вид $[\underline{v_1} \ \underline{v_2} \ \underline{v_3} \ 0000 \ \underline{v_6} \ 0 \ \underline{v_8} \ \underline{v_9} \ 0]$. Далее определяются в транскрипции звуковой базы все фрагменты, которые соответствуют данной структуре. Определяется величина FAR как отношение числа фактических ложных срабатываний к числу фрагментов, которые соответствуют данной структуре. Для каждой структуры определяется своя величина FAR. Аналогично определяется FAR для каждой ГК из заданного их перечня.

Так как каждому коду метрики, пример которого приведен в Таблице 2.1, соответствует несколько «помеховых» структур, то для получения обобщенной оценки целесообразно определять величину FAR и для кода метрики. В этом случае набор значений FAR для соответствующих кодов метрик характеризует устойчивость САРГК к воздействию посторонних произнесений относительно данной ГК. Аналогичную обобщенную оценку можно использовать и для всего набора заданных команд.

Рассмотрим следующий пример из результатов тестирования системы. Транскрипция [r a1 s' i0 j a4] ключевого слова «Россия», состоит из шести знаков.

«Созвучный» фрагмент звуковой базы «профессор», который вызвал ложное срабатывание, имеет транскрипцию [p r a1 f' e0 s a4 r], содержит три знака: r, a1, a4, - входящих в состав транскрипции ключевого слова. Здесь цифры: 1,0,4 - стоящие после транскрипционных знаков, означают соответственно, произнесение звука в предударном, в ударном и заударном слогах. Знак апострофа обозначает смягченное произнесение звука [30].

Если имеется достаточно большое количество произнесений слова «профессор», то можно определить оценку вероятности ложного срабатывания для пары: (ключевое слово – «Россия») – (помеха – «профессор») с приемлемой для ее практического использования погрешностью. Данная оценка для ключевого слова «Россия» не изменится, если в другой тестовой базе сохранится тот же диктор и то же слово-помеха «профессор».

Однако практическая ценность такой оценки мала, так как в иной тестовой базе может отсутствовать слово «профессор» или же число его произнесений окажется недостаточным для получения оценки с достаточно малой погрешностью.

В предыдущем подразделе показано, что вероятность ложного срабатывания во многом определяется количеством и размерами слитных подпоследовательностей транскрипционных знаков, которые присутствуют в модели ГК и в произнесении постороннего слова. Поэтому целесообразно определять вероятность ложной тревоги в зависимости от данных параметров, которые можно отразить структурой помехи.

В данном примере помеха имеет вариант «0 r a1 000 a4 0» структуры, которую можно представить кодовой комбинацией «0200010». Здесь два звука помехи, входящих в состав звуков ключевого слова, образуют слитную подпоследовательность звуков, которая имеется в ключевом слове, а один «созвучный» звук стоит отдельно. Между парой звуков и одним звуком стоят три транскрипционных знака, которые отсутствуют в транскрипции ключевого слова.

«Созвучные» звуки помехи окружены звуками, отсутствующими в ключевом слове. Данной кодовой комбинации соответствует девять фрагментов звуковой базы, и четыре из них вызвали ложное срабатывание системы.

Следует отметить, что в условиях слитной речи (при тестировании читается текст) появление пауз — явление относительно редкое, поэтому при обработке результатов тестирования наличие пауз не учитывается.

В предыдущем разделе показано, что за счет взаимного влияния соседних звуков на их произнесение (явление коартикуляции) значимость различных транскрипционных знаков слова-помехи, совпадающих со знаками ключевого слова, для созвучности разная. Если знаки ключевого и неключевого слов имеют одинаковое окружение с обеих сторон, то соответствующий звук неключевого слова с относительно высокой точностью соответствует акустической модели трифона (модели скрытого марковского процесса) ключевого слова. Если же окружение совпадает только с одной стороны, то точность соответствия звука модели снижается. Если же окружение звука неключевого слова полностью отличается от окружения соответствующего звука ключевого слова, то точность соответствия звука модели является самой низкой.

Введем числовые параметры, характеризующие степень отличия в произнесениях ключевого и неключевого слов:

- w число транскрипционных знаков ключевого слова;
- а число одинаковых транскрипционных знаков ключевого и неключевого слов;
- е число одинаковых транскрипционных знаков ключевого и неключевого слов, имеющих слева и справа одинаковое окружение;
- s число одинаковых знаков ключевого и неключевого слов, имеющих одинаковое окружение только с одной стороны;
- d число одинаковых знаков ключевого и неключевого слов, имеющих разное окружение с каждой стороны;

і – число ложных тревог (срабатываний);

g — число фрагментов звуковой базы, «созвучных» ключевому слову и соответствующих данной структуре помехи.

Совокупность параметров {w,e,s,d} определяет степень «созвучности» ключевого и неключевого слов, поэтому от нее зависит обобщенная оценка вероятности ложной тревоги для группы ключевых слов с числом w транскрипционных знаков. Указанная совокупность параметров отличается от параметров кода метрики, представленных в предыдущем разделе, тем, что часть параметров метрики просуммирована и является параметром из рассматриваемой совокупности. Такое суммирование позволяет снизить объем тестирующей выборки произнесений с сохранением величины погрешности полученных оценок достоверности.

Если используется не одно ключевое слово, а целый список, то число ложных срабатываний и количество фрагментов базы «созвучных» списку слов определяется соответствующими числами для каждого ключевого слова в отдельности. Если ложные срабатывания разных ключевых слов вызваны различными структурами помех, то для списка этих слов ложные срабатывания и «созвучные» фрагменты базы суммируются. Если же структуры помех одинаковы для разных слов, то следует суммировать лишь случаи различающихся помех, которые соответствуют одной структуре помехи.

На основе проведенного анализа влияния «созвучности» на число ложных срабатываний САРГК можно рекомендовать следующую методику оценки вероятности ложного срабатывания системы с учетом «созвучности» произнесений ключевых и неключевых слов.

1. Для заданного ключевого слова составляются все возможные варианты транскрипций для данного состава дикторов.

- 2. Проводится процедура распознавания заданного ключевого слова с использованием тестовой звуковой базы, и фиксируются случаи ложных срабатываний.
- 3. Для каждого случая ложного срабатывания фиксируются число транскрипционных знаков, совпадающих со знаками ключевого слова, и вариант структуры помехи фрагмента звуковой базы, который вызвал ложное срабатывание системы.
- 4. Для каждой пары «ключевое слово вариант структуры помехи» определяется число соответствующих фрагментов базы. Они рассматриваются как помехи, которые могут вызвать ложное срабатывание системы.
- 5. Определяются оценки вероятностей ложного срабатывания для каждой пары.
- 6. Если объем базы недостаточно велик для получения оценок с требуемой погрешностью, то пункты 5, 6 выполняются для пары «ключевое слово тип структуры помехи».
- 7. Для получения обобщенных оценок для нескольких ключевых слов числа ложных срабатываний и созвучных ключевым словам фрагментов звуковой базы группируются по значениям набора параметров {w,e,s,d}.

Для оценки вероятности ложного срабатывания на новом звуковом материале проводится анализ транскрипции звукового материала, на котором планируется использовать систему распознавания и оценивается вероятность появления различных структур помехи. С использованием оценок, полученных при тестировании, определяется оценка вероятности ложного срабатывания на данном звуковом материале. Используемые тестовые базы должны иметь следующие компоненты.

- Звуковой материал, размеченный во времени до уровня отдельного звука. Тематика материала должна соответствовать или перекрывать тематику того материала, на котором планируется использование САРГК.
- Текст базы с учетом оговорок дикторов.
- Транскрипцию звукового материала с учетом пауз, оговорок, звуков хезитации.

Подготовка списка ключевых слов для тестирования системы (если список специально не задан).

- Задаются минимальным и максимальным значениями количества транскрипционных знаков в ключевых словах.
- Все слова базы группируются по числу транскрипционных знаков в пределах, указанных в предыдущем пункте.
- Для каждой группы составляется список ключевых слов, обеспечивающих максимальный охват типов звуков.
- При тестировании рассматривается по очереди каждое ключевое слово (случай однословного словаря).
- Проводится формирование списков ключевых слов с учетом их созвучности для последующего тестирования системы со списками ключевых слов и определения числа замен.

2.2.2. Результаты эксперимента

На звуковой базе (далее – база 1), представленной ООО «Центр Речевых Технологий» (г. Санкт-Петербург) – далее ЦРТ [108], протестирована система KWS распознавания произнесений ключевых слов. Система разработана чешской фирмой Phonexia [179]. Характеристика тестовой звуковой базы: 90 файлов звукозаписей фраз журналистов, ведущих радиопередачи, (четыре диктора: один

мужчина и три женщины); количество произнесенных слов: 2645, длительность базы 1211с.

Условия тестирования: порог обнаружения ключевого слова установлен минимально возможным (нулевым) для обеспечения максимального числа ложных тревог и минимального количества пропусков ключевого слова; фиксировались все фрагменты звуковой базы, которые вызвали срабатывание системы. За результат правильного распознавания принималось срабатывание фрагмент звуковой базы, системы на последовательность транскрипционных которого совпадала знаков c последовательностью транскрипционных знаков ключевого слова.

Аналогичные эксперименты проводились и со второй звуковой базой (далее база 2) [73]. Характеристики базы: 60 файлов звукозаписей трех текстов, прочитанных двадцатью дикторами (10 мужчин и 10 женщин). Количество произнесенных слов: 2098, длительность базы 1226с. Число ключевых слов – 7. Состав ключевых слов представлен таблицей 2.2.

Таблица 2.2. Состав ключевых слов для первого эксперимента.

База 1

свобода, Россия, человек, время, город, вопрос, программа, президент, радио, государство, газета, корреспондент, история, сторона, решение, Москва, ребенок, палата, страна, власть

База 2

автомобилистов, автомобилисту, автомобилисты, парковки, парковок, водителей, водителю

Кроме того, проводилось тестирование системы, разработанной ЦРТ, на двух базах (базы 3 и 4), представленных ЦРТ, с использованием одинакового состава следующих ключевых слов [107].

Таблица 2.3. Состав ключевых слов для второго эксперимента.

была, вечером, говорила, даже, действительно, женщина, кажется, комнату, меня, надо, ничего, обыкновенно, очень, подумал, понимаешь, последнее, сегодня, сказал, тебе, телефону, хорошо

Базы 3,4 — это результат прочтения четырьмя дикторами (две женщины, двое мужчин) текстов двух рассказов. Соответствующие звуковые файлы (частота дискретизации - 8кГц, глубина квантования - 16 двоичных разрядов) — это результат прохождения микрофонных сигналов через фильтр, имитирующий частотную характеристику стандартного телефонного канала. База 3 длительностью 8662с содержит 19068 слов, а база 4 имеет длительность 7668с и содержит 16776 слов.

Рассмотрим сначала процесс обработки результатов тестирования САРГК на базах 3 и 4. Сам процесс тестирования проводился работниками ЦРТ. На основе результатов тестирования для каждого ключевого слова составлялась таблица с результатами первичной обработки результатов тестирования [107]. Результаты первичной обработки представлены Таблицей 2.4 (фрагмент большой таблицы).

Таблица. 2.4. Результаты первичной обработки данных тестирования САРГК

слово	транскрипция	w	a	Структ-	e	S	d	совп.	транскрип.	файл помехи
				ypa				знаки	помехи	
сегодня	s' i1 v o0 d' n' a4	7	4	040	2	2	0	* s' i1 v o0	f s' i1 v o0 p	mtb0032.seg_B2
								* s' i1 v o0	f s' i1 v o0 n	mtb0480.seg_B2
		7	3	010020	0	2	1	* s' * * v o0 *	e0 s' t' s v o0 j	mtb0091.seg_B2
								* s' * * v o0 *	a0 s' e0 1 v o0 z	mtb0325.seg_B2
								* s' * * o0 d' *	f s' e0 h o0 d' i4	mtb0323.seg_B2
								* s' * * d' n' *	o0 s' p a1 d' n' a0	mtb0390.seg_B2
								* i1 * * o0 d' *	a2 i1 s h o0 d' i4	mtb0080.seg_B2

Затем результаты первичной обработки суммировались, и формировалась таблица следующего вида (Таблица 2.5) (фрагмент большой таблицы)

Таблица 2.5. Сумма данных первичной обработки результатов эксперимента

	База 4										
Слово	Транскрипция	W	Структура помехи	e	S	d	i	g			
была	b y1 1 a0	4	01010	0	0	2	4	58			
даже	d a0 zh y4	4	010	0	0	1	10	1432			
даже	d a0 zh y4	4	01010	0	0	2	2	59			
даже	d a0 zh y4	4	010010	0	0	2	1	60			
даже	d a0 zh y4	4	020	0	2	0	6	118			
даже	d a0 zh y4	4	030	1	2	0	2	12			
очень	o0 ch i4 n'	4	010	0	0	1	3	1934			
была	b y1 1 a0	4	020	0	2	0	2	21			
меня	m' i1 n' a0	4	010010	0	0	2	1	116			
была	b y1 1 a0	4	010010	0	0	2	6	52			
была	b y1 1 a0	4	01020	0	2	1	1	1			
была	b y1 1 a0	4	02010	0	2	1	6	7			

Далее проводилось суммирование чисел ложных срабатываний (i, g) и фрагментов базы, созвучных каждому ключевому слову, по категориям: «слово», «структура помехи». Результаты суммирования по категориям представлены в виде Таблиц 2.6 и 2.7 (фрагменты больших таблиц).

Таблица 2.6. Зависимость числа ложных срабатываний от структуры помехи

База 4 (і - число ложных срабатываний)										
Структура помехи	была	вечером	говорила	даже	женщина	кажется				
010	54	2	5	42	1	0				
010000010	0	0	1	0	0	0				
010000030	0	0	0	0	0	0				
01000010	0	0	1	0	0	0				
0100010	1	0	0	0	0	0				
010010	9	0	0	3	0	0				
010010010	0	0	0	0	0	0				
0100110	0	0	0	0	0	0				
010020	0	0	0	0	0	0				
010050	0	0	3	0	0	0				
01010	17	0	0	15	1	0				
0101000010	0	0	1	0	0	0				
01010010010	0	0	2	0	0	0				
0101010	0	0	0	0	0	0				

Таблица.2.7. Зависимость числа возможных ложных срабатываний от структуры помехи

База 4 (g - числ	База 4 (g - число фрагментов базы «созвучных» ключевому слову)										
Структура помехи	была	вечером	говорила	даже	женщина	кажется					
010	6437	2656	8442	5719	2717	0					
010000010	0	0	192	0	0	0					
010000030	0	0	0	0	0	0					
01000010	0	0	226	0	0	0					
0100010	44	0	0	0	0	0					
010010	155	0	0	117	0	0					
010010010	0	0	0	0	0	0					
0100110	0	0	0	0	0	0					
010020	0	0	0	0	0	0					
010050	0	0	3	0	0	0					
01010	227	0	0	234	256	0					
0101000010	0	0	26	0	0	0					
01010010010	0	0	2	0	0	0					
0101010	0	0	0	0	0	0					

Для каждой пары значений (i,g) вычислялись нижняя p_1 и верхняя p_2 границы доверительного интервала оценки вероятности ложной тревоги [15]:

$$p_1 = \frac{g}{t^2 + g} \left[w + \frac{t^2}{2g} - t \sqrt{\frac{w(1 - w)}{g} + \left(\frac{t}{2g}\right)^2} \right]$$
 (2.4)

$$p_2 = \frac{g}{t^2 + g} \left[w + \frac{t^2}{2g} + t \sqrt{\frac{w(1 - w)}{g} + \left(\frac{t}{2g}\right)^2} \right]$$
 (2.5)

Здесь w=i/g; t=1,645, что соответствует доверительной вероятности равной 0,9.

На Рисунке 2.1 представлено графическое отображение границ доверительных интервалов для баз 3, 4. Условные обозначения на графике: о — База 3, * — База 4. По вертикальной оси отложены значения десятичных логарифмов границ интервалов, а по горизонтальной - номера пар «ключевое слово — структура помехи». Данные для рисунка представлены в Приложении таблицами ПЗ.1.1 и ПЗ.1.2.

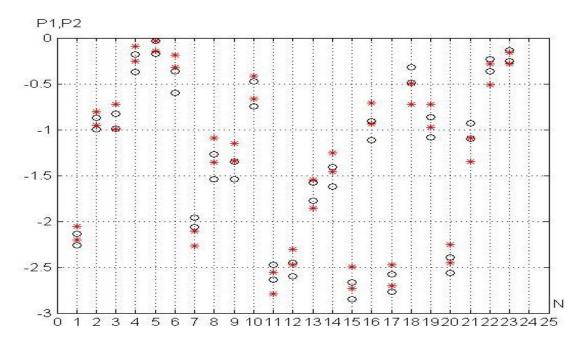


Рисунок 2.1. Доверительные интервалы оценки вероятности ложного срабатывания.

Можно сделать следующие выводы. Во-первых, из таблиц следует, что с увеличением у слова-помехи числа транскрипционных знаков одинаковых со знаками ключевого слова, растет вероятность ложного срабатывания. Во-вторых, из рисунка следует, что, несмотря на изменение состава неключевых слов базы для одинаковых пар «ключевое слово-тип структуры помехи» интервальные оценки перекрываются или очень близки.

Рассмотрим возможность обобщенной оценки вероятности ложной тревоги для заданного перечня ключевых слов и фиксированного набора параметров {w, e, s, d}. На Рисунке 2.2 представлено графическое отображение интервальных оценок для различных значений набора параметров {w,e,s,d}. Данные для рисунка представлены Таблицами ПЗ.1.3, ПЗ.1.4 Приложения.

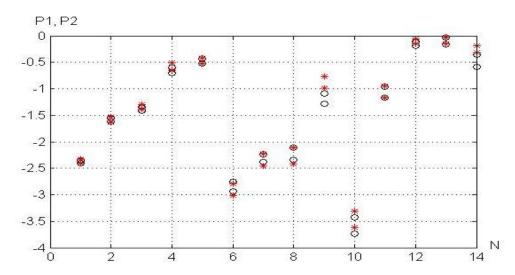


Рисунок 2.2. Доверительные интервалы оценки вероятности ложного срабатывания для списка ключевых слов

Видно, что интервальные оценки для разных баз перекрываются. Так как База 3 больше Базы 4 на 2292 слова, то интервалы оценок вероятности для нее уже. Рост объема базы приводит к увеличению числа оценок с узкими интервалами (соответствует количеству ложных срабатываний для каждой структуры помехи большему 10) примерно в 2 раза.

Рассмотрим результаты тестирования второй системы (разработка фирмы Phonexia) распознавания команд на базах 1 и 2. На основе Таблиц ПЗ.2.1 и ПЗ.2.2 Приложения построена диаграмма, представленная Рисунком.2.3, которая показывает зависимость относительной частоты ложного срабатывания (FAR) от параметров, характеризующих структуру помехи. Так как указанным данным, в основном, соответствует величина d/w = 0, то показана зависимость FAR лишь от двух параметров: e/w u s/w.

Из Рисунка 2.3 следует, что зависимости FAR от параметров, характеризующих структуру помехи, для разных баз похожи. Чем больше значения e/w, s/w, тем больше FAR. Причем определяющую роль играет величина e/w – доля звуков помехи, которые одинаково произносятся со звуками ключевого слова с учетом коартикуляции.

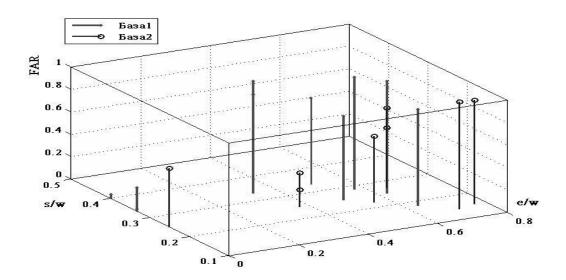


Рисунок.2.3. Зависимость относительной частоты ложного срабатывания от параметров структуры помехи.

Определим оценку вероятности интервальную ложной тревоги зависимости от величины q=a/w произнесений неключевого слова и голосовой команды. Так как отдельным значениям q соответствуют малые количества ложных тревог и «созвучных» фрагментов, что снижает точность оценки, то для повышения точности сгруппируем значения «созвучности», соответствуют достаточно большие количества ложных срабатываний системы. Исходные данные для расчета представлены результатами тестирования на базах 1 и 2 в Таблице ПЗ.2.3 Приложения.

При составлении Таблицы ПЗ.2.3 для базы 2 было учтено следующее обстоятельство. В неключевых словах: «парковки», «парковаться» - имеются группы звуков (р a1 r k o0, р a2 r k a1 v a0 c), совпадающих со звуками ключевого слова «парковок» [р a1 r k o0 v a4 k]. Количество совпадающих звуков для слова «парковаться» меньше по сравнению со словом «парковки», но величина FAR для него больше. Неоднократное прослушивание произнесений слов: «парковок», «парковки», «парковаться» - показало, что «предпредударный» звук а2 в слове «парковаться» нельзя отличить от предударного звука а1 слов

«парковки», «парковок». Было принято решение считать звук а2 в слове «парковаться» совпадающим со звуком а1 ключевого слова «парковок».

На основе полученных данных с доверительной вероятностью 0,95 определены интервальные оценки вероятности ложной тревоги для каждого диапазона значений «созвучности» согласно формулам (2.4), (2.5). В Таблице 2.8 представлены результаты вычислений

Таблица 2. 8. И	Інтервальные оценки Г.	'AR по диапазонам	значений «созвучности».

Диапазон значений q	\mathbf{p}_1	$\mathbf{p_2}$	$a=0,5(p_1+p_2)$					
База 1								
0.1250.375	0,0039	0,0073	0,0056					
0.40.571	0,07	0,12	0,096					
0.60.8	0,43	0,61	0,52					
0.8181	0,75	0,93	0,84					
База 2								
0.6250.778	0,26	0,40	0.33					
0.8570.923	0,78	0,88	0.83					

Здесь p_1 , p_2 - границы доверительного интервала, где может находиться истинное значение вероятности; t=1,96 – квантиль нормального распределения вероятностей, соответствующий доверительной вероятности 0,95; g – число фрагментов тестовой базы созвучных ключевому слову; w=i/g – относительная частота ложной тревоги; i – число ложных тревог.

Анализируя таблицу 2.8, можно сделать следующий вывод. Несмотря на то, что поменялись составы дикторов, произнесенных слов вообще и ключевых слов в частности, интервальные оценки вероятностей ложной тревоги очень близки. Наблюдается рост вероятности ложной тревоги с увеличением степени «созвучности» помех ключевым словам. Следовательно, можно говорить о том, что полученные оценки во многом определяются собственно системой распознавания, то есть достаточно объективно характеризуют качество системы.

При сохранении состава дикторов либо при обеспечении их достаточного разнообразия оценки вероятности ложной тревоги, полученные с учетом «созвучности» на тестовой звуковой базе, можно использовать при оценке

вероятности ложной тревоги на звуковом материале с иным составом произносимых слов. При расчете вероятности p_{FAreal} ложной тревоги для заданного ключевого слова применительно к конкретному звуковому материалу следует воспользоваться формулой

$$p_{FA_real} = \sum_{i=1}^{N} p_{FA,i} p_i$$
 (2.6)

Здесь $p_{FA,i}$ — вероятность ложного срабатывания для i —ой помеховой структуры; N — общее число помеховых структур, учитываемых при расчете; p_i — вероятность появления i—ой помеховой структуры в звуковом материале.

Следует отметить, что в сети моделей – возможных результатов распознавания следует указывать все возможные произнесения словоформ ключевого слова. В этом случае можно сократить число ложных срабатываний системы, так как словоформы имеют высокую степень «созвучности» с ключевым словом.

На основе вышеизложенного можно считать, что интервальная оценка вероятности ложной тревоги для заданной пары «ключевое слово – структура помехи» объективно характеризует устойчивость САРГК к произнесениям неключевых слов. Рассмотренный метод оценки вероятности ложной тревоги снижает зависимость оценки от состава неключевых слов базы, что позволяет прогнозировать оценку вероятности ложной тревоги для данного ключевого слова при смене состава слов звукового материала.

2.3. Сравнение результатов распознавания по достоверности распознавания

При разработке САРГКГК возникает задача сравнения различных вариантов реализации системы и выбора варианта, обеспечивающего наибольшую достоверность распознавания. Оценки вероятностей ошибок распознавания (относительные частоты) определяются по результатам тестирования САРГК на

большой звуковой базе [4, 38, 166, 194]. Так как сравнение вариантов системы происходит по оценкам, которые являются случайными величинами, то возможно принятие ошибочного решения по результатам сравнения. Чем выше достоверность распознавания, тем больше должна быть тестовая база, чтобы вероятность ошибки сравнения сохранялась на заданном уровне. Однако увеличение объема базы обусловливает рост ее стоимости и трудоемкости тестирования.

В данном разделе рассматривается метод сравнения САРГК, обеспечивающий снижение объема тестовой выборки голосовых команд по сравнению с существующим методом [69, 79, 82]. С помощью рассматриваемого ниже подхода определяются оценки вероятностей замены одной команды другой из списка команд, ложного срабатывания при воздействии помех, ложного пропуска команды. В дальнейшем все указанные вероятности называются вероятностью ошибки распознавания.

Суть метода заключается в следующем. При оценке вероятности ошибки распознавания i-ой команды в процессе тестирования САРГК, которая предназначена для дикторонезависимого распознавания нескольких команд, определяется метрика M_{ii} сопоставления текущего произнесения команды со «своей» моделью. При этом в сети моделей для сопоставления доступна только эта модель. Затем определяется наибольшая метрика M_{ij} сопоставления данного произнесения с j-ой «чужой» моделью. При этом в сети моделей доступны все модели, кроме i-ой. Далее определяется разность метрик $\Delta M_i = M_{ii}$ - M_{ij} . Если ΔM_i > 0, то результат распознавания считается правильным. В противном случае имеет место ошибка распознавания.

Величина разности ΔM_i определяется изменчивостью произнесения i-ой команды. Изменчивость, в свою очередь, определяется множеством факторов: временной и тембральной изменчивостью произнесения, отличиями голосов разных дикторов, воздействием помех,... Поэтому распределение вероятностей

случайной величины ΔM_i можно считать нормальным. Данное утверждение подтверждается проверкой результатов экспериментов по критерию «хи-квадрат» [15].

Если определить оценки $\hat{\mu}$ математического ожидания и среднеквадратического отклонения $\hat{\sigma}$ для данного распределения, то, используя отношение $\hat{\alpha} = \hat{\mu}/\hat{\sigma}$ при сравнении различных САРГК, можно сократить требуемое число произнесений этой команды в тестирующей звуковой базе по сравнению с известным методом. При этом вероятность принятия ошибочного решения по результатам сравнения систем не уменьшается.

Исходя из данного выше определения случайной разности ΔM_i , вероятность $P_{ou.\Gamma K}$ ошибки распознавания команды

$$P_{out.\Gamma K} = \int_{-\infty}^{0} p(\Delta M_i) d(\Delta M_i) = 0.5 - \Phi(\alpha), \qquad (2.7)$$

где $\Phi(\alpha) = \frac{1}{\sqrt{2\pi}} \int_0^\alpha e^{-z^2/2} dz$ - табулированная функция Лапласа; $\alpha = \mu/\sigma$; μ , σ - математическое ожидание и среднеквадратическое отклонение нормального распределения, соответственно; $p(\Delta M_i)$ - плотность нормального распределения вероятностей случайной величины ΔM_i [15, 37]. Из приведенной формулы следует однозначное соответствие вероятности ошибки распознавания величине α : с увеличением α снижается вероятность ошибки распознавания.

Пусть САРГК2 имеет меньшую вероятность ошибки распознавания по сравнению с САРГК1, тогда $\Delta\alpha=\alpha_2$ - $\alpha_1>0$. На практике можно использовать лишь статистические оценки $\hat{\alpha}=\hat{\mu}/\hat{\sigma}$. Если $\Delta\widehat{\alpha}=\widehat{\alpha_2}-\widehat{\alpha_1}<0$, то возникает ошибка сравнения САРГК. Вероятность $P_{ou.\ cpas.}$ ошибки

$$P_{ou.cpas.} = \int_{-\infty}^{0} p(\Delta \widehat{\alpha}) d(\Delta \widehat{\alpha}), \qquad (2.8)$$

где $p(\Delta \widehat{\alpha})$ – плотность распределения вероятностей случайной величины $\Delta \widehat{\alpha}$.

Проанализируем сначала, насколько достоверным является сравнение САРГК по относительной частоте \bar{p} возникновения ошибки распознавания. Рассмотрим случайную величину $\Delta \bar{p} = \bar{p}_1 - \bar{p}_2$, где \bar{p}_1 , \bar{p}_2 - относительные частоты возникновения ошибки распознавания для первой САРГК1 и второй САРГК2, соответственно. Для определенности считаем, что вторая САРГК2 обеспечивает более достоверное распознавание ГК, то есть вероятность ошибки для второй САРГК меньше, чем для первой: $p_2 < p_1$. Если при сравнении САРГК случайная величина $\Delta \bar{p}$ окажется положительной, то будет принято верное решение: САРГК2 обеспечивает большую достоверность по сравнению с САРГК1. В противном случае принимается ошибочное решение по результатам сравнения САРГК. Определим вероятность того, что $\Delta \bar{p}$ примет отрицательное значение при соблюдении условия: $p_2 < p_1$.

Распределение вероятностей относительной частоты p возникновения события при проведении n испытаний подчиняется биномиальному закону[15]. Однако при больших n приблизительно описывается нормальным законом с математическим ожиданием равным вероятности p возникновения события и среднеквадратическим отклонением

$$\sigma_p = \sqrt{p(1-p)/n}$$

В этом случае случайная величина Δp также является нормально распределенной с математическим ожиданием и среднеквадратическим отклонением, соответственно

$$\Delta p = p_1 - p_2, \, \sigma_{\Delta p} = \sqrt{\frac{p_1(1-p_1) + p_2(1-p_2)}{n}}$$
 (2.9)

Вероятность принятия ошибочного решения при сравнении САРГК по относительной частоте возникновения ошибки

$$p_{p.\text{OIII}} = P(-1 < \Delta \bar{p} < 0) = \Phi\left(\frac{0 - \Delta p}{\sigma_{\Delta p}}\right) - \Phi\left(\frac{-1 - \Delta p}{\sigma_{\Delta p}}\right) = 0.5 - \Phi\left(\frac{\Delta p}{\sigma_{\Delta p}}\right). \quad (2.10)$$

Если задана вероятность принятия ошибочного решения, то, пользуясь таблицей значений функции Лапласа, из выражения (2.10) можно определить отношение $\frac{\Delta p}{\sigma_{\Delta p}}$. Затем, используя выражение (2.9) и задаваясь ожидаемыми значениями вероятностей p_1 и p_2 , можно определить требуемый объем тестирующей выборки

$$n_p = \left(\frac{\Delta p}{\sigma_{\Delta p}}\right)^2 \frac{p_1(1 - p_1) + p_2(1 - p_2)}{\Delta p^2}$$
 (2.11)

Оценим теперь вероятность принятия ошибочного решения при использовании предлагаемого подхода

$$P_{out.cpas.} = \int_{-\infty}^{0} p(\Delta \widehat{\alpha}) d(\Delta \widehat{\alpha}) = P(-\infty < \widehat{\alpha}_{2} - \widehat{\alpha}_{1} < 0) =$$

$$= P\left(-\infty < \frac{\widehat{\mu}_{2}}{\widehat{\sigma}_{2}} - \frac{\widehat{\mu}_{1}}{\widehat{\sigma}_{1}} < 0\right) = P(-\infty < \widehat{\mu}_{2} \cdot \widehat{\sigma}_{1} - \widehat{\mu}_{1} \cdot \widehat{\sigma}_{2} < 0).$$

Случайные величины $\hat{\mu}$ являются нормально распределенными [15] с математическим ожиданием μ и среднеквадратическим отклонением $\sigma_{\mu} = \sigma/\sqrt{n}$ Случайные величины $\hat{\sigma}$ распределены по закону «Распределение модуля [27] с математическим ожиданием многомерного вектора» среднеквадратическим отклонением $\sigma_{\sigma} \approx \sigma / \sqrt{2(n-1)}$. Учитывая, что случайные величины $\hat{\mu}$ и $\hat{\sigma}$ взаимно независимы, считаем случайную величину $\hat{\beta} = \widehat{\mu_2} \cdot \widehat{\sigma_1}$ распределенной нормально (проверено моделированием) математическим ожиданием среднеквадратическим отклонением, соответственно [22]:

$$\beta = \mu_{2} \cdot \sigma_{1} - \mu_{1} \cdot \sigma_{2},$$

$$\sigma_{\beta} = \sqrt{(\hat{\beta} - \beta)^{2}} = \sqrt{(\sigma_{\mu 2}^{2} \sigma_{1}^{2} + \sigma_{\sigma 1}^{2} \mu_{2}^{2} + \sigma_{\mu 2}^{2} \sigma_{\sigma 1}^{2} + \sigma_{\mu 1}^{2} \sigma_{2}^{2} + \sigma_{\sigma 2}^{2} \mu_{1}^{2} + \sigma_{\mu 1}^{2} \sigma_{\sigma 2}^{2})} \approx$$

$$\approx \frac{\sigma_{1} \sigma_{2}}{\sqrt{n/2}} \sqrt{1 + 0.25(\alpha_{1}^{2} + \alpha_{2}^{2})}, \alpha_{1} = \mu_{1} / \sigma_{1}, \alpha_{2} = \mu_{2} / \sigma_{2}$$
(2.12)

Вероятность принятия ошибочного решения при сравнении САРГК

$$P_{oui.cpagh} = P(-\infty < \hat{\beta} < 0) = 0.5 - \Phi(\alpha_{\beta}), \alpha_{\beta} = \beta / \sigma_{\beta} \approx \sqrt{n} \frac{\alpha_{2} - \alpha_{1}}{\sqrt{2 + 0.5(\alpha_{1}^{2} + \alpha_{2}^{2})}}. \tag{2.13}$$

При заданной вероятности P_{oui} сравн ошибочного решения и ожидаемых значениях вероятностей p_1 , p_2 ошибок распознавания для САРГК1 и САРГК2 можно, пользуясь выражениями (2.12), (2.13), определить требуемый объем тестирующей выборки

$$n_{\alpha} \approx \alpha_{\beta}^{2} \frac{2 + 0.5(\alpha_{1}^{2} + \alpha_{2}^{2})}{(\alpha_{2} - \alpha_{1})^{2}}$$
 (2.14)

Оценим необходимый объем тестирующей выборки при сравнении САРГК по относительной частоте и по предложенному методу на числовом примере. Пусть вероятность $P_{out\ cpaвн}$ принятия ошибочного решения при сравнении САРГК задана равной 10%. Пусть вероятность ошибки p_1 распознавания для САРГК1 равна 3%, а для САРГК2 — вероятность ошибки p_2 =2%. Из формул (2.10), (2.13) следует

$$\Phi(\Delta p/\sigma_{\Delta p}) = \Phi(\beta/\sigma_{\beta}) = 0.5 - P_{oui\ cpagh} = 0.4.$$

Пользуясь таблицей значений функции Лапласа [15, 37], находим $\Delta p/\sigma_{\Delta p}=\beta/\sigma_{\beta}=1,28.$ Из выражений (2.11) и (2.14) находим требуемое число испытаний при сравнении САРГК по относительной частоте и по параметрам распределения случайной величины ΔM_i , соответственно: n_p =798, n_{α} =298.

Полученные значения были проверены с помощью моделирования в среде Matlab. Были сгенерированы две независимые выборки значений случайной нормально распределенной величины ΔM_i объемом n — моделировались результаты тестирования САРГК1 и САРГК2. Каждой выборке соответствовала своя пара заданных значений математического ожидания μ и среднеквадратического отклонения σ . Значение $\alpha = \mu/\sigma$ для второй выборки было

больше, чем для первой, что задавало большую вероятность ошибки для САРГК1 ($P_1 > P_2$).

Для каждой выборки подсчитывалось количество m событий, когда $\Delta M_i < 0$ (появление ошибки распознавания). Величина относительной частоты $\stackrel{-}{p} = m/n$ появления события фиксировалась. Кроме того, для каждой выборки фиксировалась величина $\hat{\alpha} = \hat{\mu}/\hat{\sigma}$ и определялись разности $\overline{\Delta p} = \overline{p_1} - \overline{p_2}$, $\Delta \widehat{\alpha} = \widehat{\alpha}_2 - \widehat{\alpha}_1$. Здесь $\hat{\mu}$, $\hat{\sigma}$ - оценки математического ожидания и среднеквадратического отклонения.

Далее описанный эксперимент повторялся M раз. При повторении подсчитывалось количество E_p событий, когда $\overline{\Delta p} < 0$ (число неправильных решений при сравнении САРГК по относительной частоте появления ошибки распознавания). Подсчитывалось также количество E_α событий, когда $\Delta \widehat{\alpha} < 0$ (число неправильных решений при сравнении САРГК по величине $\widehat{\alpha}$).

Далее подсчитывались значения оценок вероятностей принятия ошибочных решений по \bar{p} и $\hat{\alpha}$: $\bar{p}_{p.\text{ош}} = E_p/M$, $\bar{p}_{\alpha.\text{ош}} = E_\alpha/M$

При проведении моделирования изменялся объем n выборки так, чтобы полученные значения $\bar{p}_{p.\text{ош}}$, $\bar{p}_{\alpha.\text{ош}}$ сравнялись с заданной величиной 10% вероятности принятия ошибочного решения.

Из заданных значений p_1 =0,03, p_2 =0,02 и выражения (1) следуют значения α_1 =1,88 и α_2 =2,06. Исходя из данных значений α , были установлены следующие значения параметров распределений: μ_1 =1,88, μ_2 =2,06, σ_1 = σ_2 =1.

При $M=10^6$ получены следующие результаты моделирования: $n_{p.мод.}=640$, n_{α} $n_{\rho.мод.}=640$, $n_{\alpha}=640$, n_{α

Из данного примера видно, что сравнение САРГК по величине α позволяет сократить объем тестирующей выборки в 2,15 раза.

Рассмотренный подход к сравнению САРГК на основе использования оценки $\hat{\alpha}$ был проверен при проектировании САРГК (распознавание названий цифр 0...9) в ЦРТ. Одна из САРГК при обучении прошла две итерации алгоритма Баума-Уэлча [159], а вторая — четыре, что обеспечивает более высокую достоверность распознавания. Системы тестировались выборкой с объемом n=560. В процессе тестировании ошибки распознавания не были обнаружены.

При обработке результатов измерения значений случайной величины ΔM_i были получены следующие значения оценок математического ожидания и среднеквадратического отклонения:

$$\widehat{\mu_1} = 397.9; \ \widehat{\mu_2} = 419.6; \cdot \ \widehat{\sigma_1} = 157.7; \ \widehat{\sigma_2} = 161.9.$$

Отсюда следует $\widehat{\alpha_1}=2,52$; $\widehat{\alpha_2}=2,59$. Видно, что $\widehat{\alpha_1}<\widehat{\alpha_2}$ - подтверждается факт, что вторая САРГК имеет меньшую вероятность ошибки. Если в выражении (2.1) вместо значений величины α подставить ее оценки, то получим оценки вероятностей ошибок: $\widehat{P}_{1\ ou.\Gamma K}=0,0059$; $\widehat{P}_{2\ ou.\Gamma K}=0,0048$. Такая подстановка в данном случае оправдана, так как благодаря большому объему выборки относительная погрешность оценок: $\widehat{\mu}$, $\widehat{\sigma}$, $\widehat{\alpha}$ - невелика. Можно показать, что она не превышает 5%.

Аналогично из выражения (2.13) следует значение вероятности ошибки данного заключения: $\hat{P}_{ou.cpaвh} = 0,29$. Для снижения вероятности ошибочного решения до величины $\hat{P}_{ou.cpaвh} = 0,1$ следует увеличить объем тестирующей выборки до 3024 (следует из выражений (2.13) и (2.14).

Использование рассмотренного подхода к сравнению САРГК в данном случае позволило не только оценить вероятности ошибок распознавания, но и оценить достоверность заключения по результатам сравнения САРГК, а также

определить необходимый объем тестирующей выборки для реализации требуемой достоверности заключения.

Был проведен еще один эксперимент. Тестировались две системы, предназначенные для распознавания голосовых команд - названий цифр от 0 до 9. В первой системе использовалось семь гауссовых кривых для описания законов распределения в состояниях моделей скрытых марковских процессов, а во второй – 10, что обеспечивает более высокую точность акустических моделей звуков. Объем тестирующей выборки для каждой команды – 56. Результаты тестирования представлены в Таблице 2.9.

Таблица 2.9. Резу	упьтаты сравне	ния лвух neavi	ьтатов тести	пования
1 иолица 2.7. 1 сэ	ymbrarbi cpabiic	лии двух резул	ibiaiob iccin	pobalini

Команда	μ_1	σ_1	μ_2	σ_2	α_1	α_2	p ₁ %	p ₂ %	Рош.реш
ноль	73.260	19.354	82.285	19.642	3.78	4.18	0.008	0.001	0.24
один	104.410	17.999	110.123	19.176	5.80	5.74	0.000	0.000	0.38
два	136.440	40.878	139.641	41.711	3.34	3.35	0.042	0.041	0.49
три	39.600	18.012	46.233	18.280	2.20	2.53	1.395	0.572	0.18
четыре	108.538	28.938	121.115	29.857	3.75	4.06	0.009	0.002	0.32
АТКП	98.421	22.209	103.572	25.805	4.43	4.01	0.000	0.003	0.24
шесть	104.882	24.362	108.877	26.021	4.30	4.18	0.001	0.001	0.42
семь	59.892	25.951	69.638	23.210	2.31	3.00	1.050	0.135	0.04
восемь	123.332	31.919	127.009	35.505	3.86	3.58	0.006	0.017	0.48
девять	104.419	35.021	106.646	35.418	2.98	3.01	0.14	0.13	0.47

Условные обозначения: μ_1 , μ_2 , σ_1 , σ_2 — средние значения и оценки среднеквадратического отклонения распределения случайной величины ΔM_i для первой и второй систем соответственно; p_1 , p_2 , $P_{\text{ош.реш}}$ — оценки вероятностей ошибок для первой и второй систем соответственно, вероятность ошибочного решения при сравнении систем по достоверности распознавания; $\alpha_i = \mu_i / \sigma_i$ - критерий сравнения САРГК.

В целом расчетные по результатам тестирования вероятности ошибок для второй САРГК меньше. Видно, что для команд: «»один», «два», «шесть», «восемь» - вероятности ошибочных решений по результатам сравнения систем велики, что говорит о необходимости увеличить объем тестирующей выборки для

этих команд для получения более достоверного вывода по сравнению САРГК. Следует отметить, что ошибки распознавания при тестировании отсутствовали, что не позволяет сравнить системы по относительной частоте ошибок распознавания.

Проанализируем зависимость вероятности принятия ошибочного решения по результатам сравнения САРГК от объема выборки. Путем описанной выше процедуры моделирования построены графики (Рисунок 2.4) зависимостей вероятностей ошибочного решения при использовании известного и предложенного методов (на графике – P_{err}), (n – объем тестирующей выборки) для различных значений вероятностей P1, P2 ошибок распознавания голосовых команд для САРГК1 и САРГК2, соответственно.

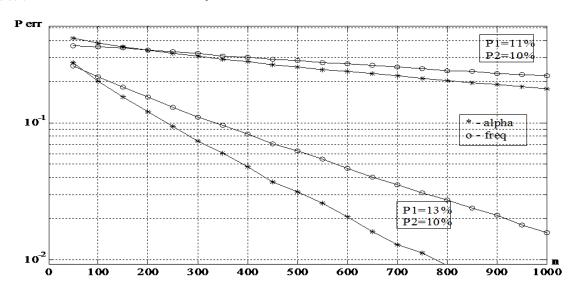


Рисунок 2.4. Зависимость вероятности ошибочного решения при сравнении САРГК от объема тестовой выборки звукозаписей

Здесь знаком «*» помечены данные, полученные при использовании параметра α , знак «о - freq» относится к данным, полученным на основе определения относительной частоты ошибки распознавания. Графики являются достаточно гладкими, что говорит о малости погрешностей статистических оценок $\widehat{P}_{w_{\text{ош}}\text{-срав.}}(n)$ и $\widehat{P}_{\alpha_{\text{ош}}\text{-срав.}}(n)$. Небольшая величина погрешностей обусловлена достаточно большой величиной $M=10^5$ числа сравнений систем по вероятности ошибки распознавания команд.

Анализ графиков показывает, что с увеличением объема тестирующей выборки уменьшается вероятность ошибки сравнения САРГК и растет выигрыш (уменьшение объема выборки) при применении предложенного метода сравнения САРГК. Выигрыш также увеличивается при уменьшении степени отличия сравниваемых САРГК по вероятности ошибки распознавания команд.

Проведенное исследование показывает, что выигрыш (сокращение объема тестирующей выборки) от применения метода сравнения САРГК с использованием вероятностных характеристик сопоставления произнесений команды с акустическими моделями звуков растет с повышением требований к точности сравнения. На основании вышеизложенного можно рекомендовать следующую методику сравнения САРГК.

- -Задаются допустимой вероятностью ошибочного решения при сравнении САРГК и определяют из выражения (2.13) величину $\alpha_{\scriptscriptstyle B}$.
- -Задаются ожидаемыми значениями вероятностей ошибок распознавания и из выражения (2.7) определяют значения α_1 и α_2 .
 - -С помощью выражения (2.14) определяют объем тестирующей выборки.
 - -Проводят тестирование САРГК и определяют значения $\alpha_1; \alpha_2$.
- -По полученным значениям $\overline{\alpha}_1; \overline{\alpha}_2$ уточняют объем выборки и проводят повторное тестирование.
 - -Формулируют заключение по результатам сравнения САРГК.

Выводы по разделу 2

1. Предложен метод, обеспечивающий снижение зависимости оценки вероятности ложного срабатывания от состава тестовой выборки звукозаписей (звуковой базе). Метод использует следующие положения.

- Число фрагментов звуковой базы, которые могут вызвать ложное срабатывание, определяется с учетом отличия последовательности звуков произнесения слова от последовательности звуков ключевого слова ГК.
- Различные типы отличий фрагментов тестовой звуковой базы от последовательности звуков ГК характеризуются соответствующими цифровыми кодовыми комбинациями. Для каждой кодовой комбинации определяется оценка вероятности ложного срабатывания.
- Степень отличий заданного перечня ГК от «некомандных» произнесений характеризуется набором параметров. Для каждой совокупности значений параметров определяется оценка вероятности ложного срабатывания.
- 2. Предложенный метод позволяет снизить объем экспериментов при определении оценки вероятности ложного срабатывания для звукового материала, отличающегося от состава тестовой выборки звукозаписей.
- 3. Рассмотрен метод сравнения результатов распознавания, полученных в процессе тестирования САРГК, по достоверности распознавания. Метод использует предложенный критерий достоверности распознавания ГК, который определяется путем вычисления среднего и среднеквадратического отклонения для выборки значений разностей метрик сопоставления ключевого слова со «своей» и «чужой» акустическими моделями. Получены выражения, которые позволяют определить объем тестовой звуковой базы в зависимости от заданной вероятности ошибки сравнения
- 4. Предложенный метод обеспечивает снижение объема тестовой звуковой базы по сравнению с существующим методом. В частности, если вероятности ложного пропуска команды сравниваемых результатов распознавания равны 10% и 13%, то при заданной вероятности ошибки сравнения равной 5% использование критерия позволяет сократить объем тестовой выборки звукозаписей голосовой команды с 580 до 400 звукозаписей (на 31%).

3. Исследование методов подавления помех на стадии предварительной обработки речевого сигнала

В данном разделе рассматриваются помехи, присутствующие на интервале длительности ключевого слова команды. Воздействие помехи меняет параметры речевого сигнала, следовательно, меняются и метрики сопоставления параметров с акустическими моделями, входящими в состав сети моделей — возможных результатов распознавания. Достоверность распознавания меняется.

Для подавления помех используются методы фильтрации и компенсации. В общем случае, наряду с подавлением помехи указанные методы меняют параметры РС. Поэтому для оценки целесообразности использования какого-либо метода следует оценивать степень снижения достоверности распознавания, а также изменения параметров РС до и после его применения.

Помехи, проявляющие себя на стадии предварительной обработки РС, по форме можно подразделить на шумовые, импульсные, узкополосные. Особо следует обратить внимание на периодические помехи. Наличие периодичности упрощает подавление помехи.

Анализ влияния помехи и метода ее подавления целесообразно рассматривать как на уровне отдельных фрагментов сигнала, так и на уровне тестовых выборок голосовых команд, когда оцениваются усредненные характеристики изменений параметров РС, а также снижение достоверности распознавания.

3.1. Оценка влияния помех с использованием критерия достоверности распознавания

Логично оценивать степень влияния помех на достоверность распознавания величиной b изменения вероятности ошибки распознавания при воздействии помехи относительно вероятности ошибки, когда помеха отсутствует.

Воздействие помехи меняет параметры речевого сигнала. При этом метрика сопоставления произнесения со «своей» моделью меняется. Одновременно меняются метрики его сопоставления с остальными моделями, присутствующими в сети гипотез. Наиболее часто воздействие помехи уменьшает степень соответствия произнесения со «своей» моделью — соответствующая метрика уменьшается. Растет вероятность ложного пропуска команды. Однако может возникнуть ситуация, когда метрика сопоставления произнесения с ближайшей «чужой» моделью станет меньше по сравнению с метрикой сопоставления со «своей» моделью. При этом воздействие помехи приводит к уменьшению вероятности ошибки распознавания. Поэтому целесообразно учитывать абсолютную величину изменения вероятности ошибки при оценке влияния помех на работу системы.

$$b = \frac{|p_{\text{ош.п}} - p_{\text{ош.0}}|}{p_{\text{ош.0}}}.$$
(3.1)

Здесь $p_{\text{ош.п}}$, $p_{\text{ош.0}}$ - вероятности ошибки распознавания при наличии и соответственно. В тестирования САРГК помех, результате определяются оценки вероятностей ошибки. Поэтому на практике в выражении (3.1)онжом использовать $\hat{p}_{ ext{out.II}},\hat{p}_{ ext{out.0}}$ вероятностей ЛИШЬ оценки соответствующие относительные частоты ошибок, и вместо величины b можно определить лишь ее оценку

$$\hat{b} = rac{|\hat{p}_{ ext{oul.II}} - \hat{p}_{ ext{oul.0}}|}{\hat{p}_{ ext{oul.0}}}.$$

Во втором разделе показано, что выборка значений разности метрик $\Delta M_{ij} = M_{ii}$ - M_{ij} сопоставления произнесения со «своей» и наиболее близкой «чужой» моделями несет информацию о достоверности распознавания. Параметр $\widehat{\propto}$ выборки можно использовать в качестве критерия сравнения систем по достоверности распознавания. Использование критерия $\widehat{\propto}$ при сравнении САРГК по достоверности обеспечивает меньшую вероятность ошибки сравнения

относительно случая, когда используются относительные частоты ошибок распознавания.

Проанализируем возможность определения степени влияния помех на достоверность распознавания с помощью оценки €. Из выражения (2.7) следует

$$\hat{b} = \frac{|\Phi(\hat{\alpha}_0) - \Phi(\hat{\alpha}_\Pi)|}{0.5 - \Phi(\hat{\alpha}_0)}.$$
(3.2)

Здесь $\hat{\alpha}_0$, $\hat{\alpha}_{\scriptscriptstyle \Pi}$ - значения критерия $\widehat{\propto}$ для случаев отсутствия и наличия помехи.

Так как функция $\Phi(\alpha_n)$ Лапласа является монотонной, то разность, стоящая в числителе выражения, монотонно зависит от разности $\hat{\alpha}_0 - \hat{\alpha}_n$. Следовательно, использование критерия $\hat{\alpha}$ при вычислении \hat{b} обеспечивает более точный учет влияния помех на достоверность распознавания по сравнению с использованием относительных частот. Оценим влияние помех, возникающих в телефонном канале, на достоверность распознавания ГК.

Помехи и искажения в телефонном канале (ТК), образованном двумя каналами тональной частоты и магистральным каналом, регламентируются государственными стандартами [16, 17, 109] и эксплуатационными нормами на электрические параметры коммутируемых каналов телефонных сетей общего пользования [131]. В телефонной линии (ТЛ) можно выделить три основных типа помех: шум в полосе частот ТК; гармоническая помеха; импульсная помеха.

Типовая амплитудно-частотная характеристика (AЧX) всего канала связи имеет вид, приведенный на Рисунке 3.1. Максимально допустимые искажения AЧX нормируются (Рисунок 3.2).

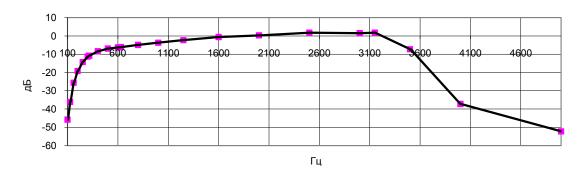


Рисунок 3.1. Типовая амплитудно-частотная характеристика телефонного канала связи

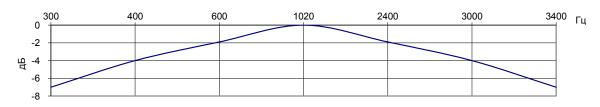


Рисунок 3.2. Максимально допустимые искажения амплитудно-частотной характеристики телефонного канала связи

Проведено экспериментальное исследование влияния указанных типов помех и искажений на достоверность распознавания двух команд: «ноль» и «один» [43, 72, 86, 87]. Соответствующие транскрипции команд: [n o0 l'] и [a1 d i0 n]. Здесь цифры 0 и 1 соответствуют ударному и предударному слогам. На Рисунках 3.3 и 3.4 приведены соответствующие «осциллограммы» для мужского голоса. По оси абсцисс отложены значения времени в секундах, а по оси ординат — нормированные значения амплитуды. Следует отметить, что сигнал команды «один» в отличие от команды «ноль» содержит участок с малым значением амплитуды (соответствует звуку «d»).

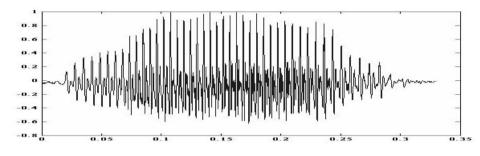


Рисунок 3.3. «Осциллограмма» речевого сигнала «ноль»

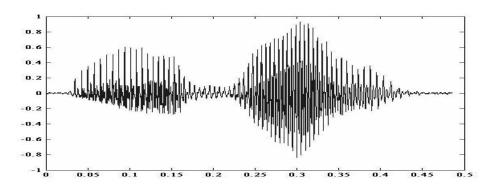


Рисунок 3.4. «Осциллограмма» речевого сигнала «один»

На Рисунке 3.5 приведены графики огибающих кратковременного спектра для звуков: «о0», «i0», «п» «l'». Значения огибающих спектра получены после определения коэффициентов линейного предсказания на основе рекурсивной модели РС [112]. По оси абсцисс отложены номера спектральных отсчетов при частоте дискретизации 8 кГц, по оси ординат отложены значения отсчетов в дБ.

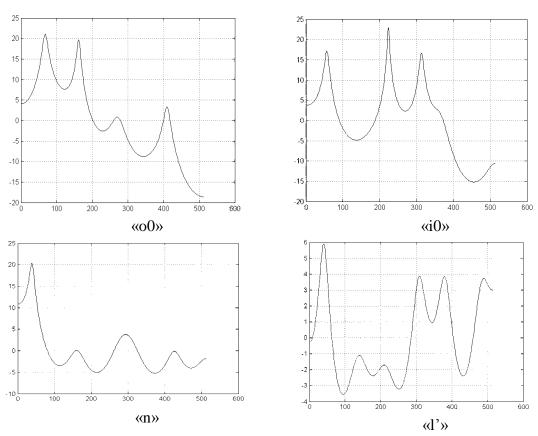


Рисунок 3.5. Огибающие кратковременных амплитудных спектров звуков.

Следует отметить значительную неравномерность спектров ярко Воздействие выраженную формантную структуру указанных звуков. сосредоточенных по спектру помех вызывает появление дополнительного «всплеска» огибающей спектра сигнала, что существенно влияет на результаты распознавания, так как именно параметры огибающей спектра несут основную информацию для САРГК. На Рисунке 3.6 показано влияние помехи с частотой 400 Гц на огибающую спектра сигнала, соответствующего звуку «п» команды «Один». По оси абсцисс отложены значения частоты в Гц, по оси ординат значения амплитудного спектра в дБ.

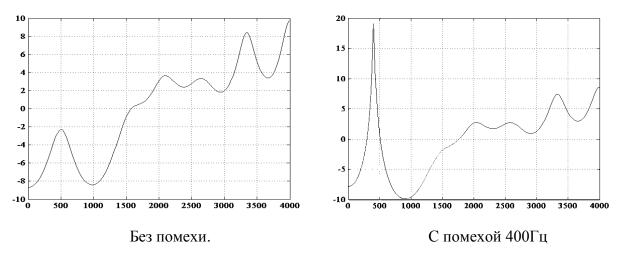


Рисунок 3.6. Влияние гармонической помехи на огибающую спектра звука «n»

Сетевая структура моделей, использованная в данном эксперименте, состояла из двух параллельных цепочек моделей. В первой цепочке находились последовательно соединенные акустические модели паузы и команды «ноль», а во второй — акустические модели паузы и команды «один». Все модели, используемые в данной серии экспериментов, за исключением модели паузы, имеют простую лево-правую структуру, то есть без возврата в предыдущее состояние. В модель паузы дополнительно включена ненулевая вероятность перехода в предыдущее состояние.

Для создания акустических моделей - обучения системы использовались отдельные произнесения команд (550 произнесений на каждую команду). При

создании обучающей выборки использовались голоса 11 дикторов (четыре женщины, семь мужчин).

В процессе обучения были сформированы модели звуков, входящих в состав команд в соответствии с их транскрипциями. Команда «ноль» была разбита на звуки «n», «o0», «l'», а команда «один» - на звуки «a1», «d», «i0», «n». Поскольку звук «н» входит в состав обеих команд, то всего было создано шесть моделей звуков. Каждая модель звука состояла из трех состояний.

В качестве системы параметров речевого сигнала использовались 12 кепстральных коэффициентов, полученных на основе метода линейного предсказания, логарифм энергии, а также первые две производные указанных параметров (всего 39 параметров) [205]. Размер окна анализа 25 мс, смещение кадров 10 мс, тип окна – окно Хэмминга. Для повышения устойчивости системы распознавания к частотным искажениям использовались параметры с нулевым математическим ожиданием. Как было показано выше, при использовании кепстральных коэффициентов влияние искажений АЧХ телефонного канала на параметры РС ОНЖОМ уменьшить, вычитая из вычисленных временных траекторий параметров их средние значения на интервале длительности ГК.

На вход САРГК подавались звукозаписи произнесений команд. К звукозаписям добавлялись имитации помех, характерных для телефонной линии. Были выделены основные виды помех в телефонном канале, а также их предельно допустимые величины:

- шум с равномерной спектральной плотностью в полосе 300-3400 Гц минимальное отношение сигнал-шум 25 дБ;
- последовательность импульсов (помехи от источников питания), частота следования импульсов 100 Гц, минимальное отношение сигнал-шум 25 дБ;
- гармонические колебания (помехи от источников питания, вызывных сигналов). Частоты колебаний соответствовали границам частотного диапазона ТК и его середине (400, 1000, 2600 Гц). Минимальное отношение сигнал-шум для

каждого колебания составило 32 дБ. Исследование достоверности распознавания ГК производилось отдельно для каждого колебания.

Анализировалось также влияние искажений амплитудно-частотной характеристики телефонного канала на достоверность распознавания (Рисунки 3.1 и 3.2).

При распознавании команд на вход системы подавалась тестирующая выборка, состоящая из 550 произнесений от 11 дикторов (те же дикторы, что и в обучающей выборке), Сначала произносилась одна команда и фиксировалась метрика ее сопоставления со «своей» моделью. Затем доступ к «своей» модели блокировался, и определялась метрика сопоставления с «чужой» моделью. Определялась разность метрик. Такие же эксперименты проводились и для другой команды. Определялись разности ΔM_{ij} метрик, и строились гистограммы распределения их значений.

Всего было проведено семь экспериментов: для каждого вида помехи и для речевого сигнала при отсутствии помех. На Рисунках 3.7 и 3.8 приведены результаты эксперимента в виде графиков, аппроксимирующих соответствующие гистограммы. По оси абсцисс отложены значения разностей ΔM_{ij} метрик.

Положительные значения разности соответствуют правильным решениям распознавания, отрицательные — ошибкам распознавания. Жирной вертикальной чертой разделены границы областей правильного и ошибочного распознавания. По оси ординат отложен процент от общего числа произнесений, попадающий в определенную область разностей метрик.

Линия «300-3400» соответствует записям без помех; линия «w_noise» - записям с белым шумом; линия «pulse» - записям с импульсными помехами; линии: «sin_left», «sin_middle» и «sin_right» - соответствуют записям с гармоническими помехами на частотах: 400, 1000 и 2600 Гц, соответственно; линия «fading» соответствует записям с частотными искажениями в телефонном канале согласно Рисункам 3.1 и 3.2.

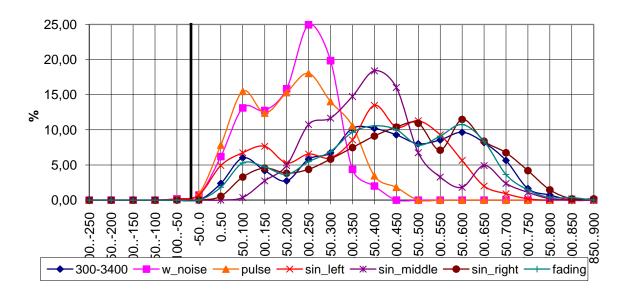


Рисунок 3.7. Аппроксимация гистограмм значений разности метрик сопоставления произнесений слова «ноль» со «своей» и «чужой» моделями при воздействии помех.

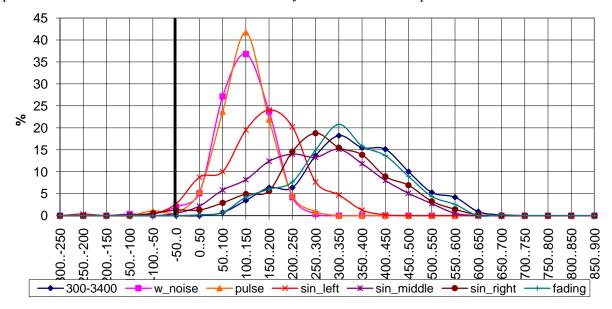


Рисунок 3.8. Аппроксимация гистограмм значений разности метрик сопоставления произнесений слова «один » со «своей» и «чужой» моделями при воздействии помех.

В Таблице 3.1 приведены значения выборочного среднего $\hat{\mu}$, выборочного среднеквадратического отклонения $\hat{\sigma}$, а также величин: $\hat{\alpha}$, \hat{b} , $\hat{p}_{\text{ош}}$.

Таблица 3.1. Значения критерия достоверности и относительной частоты ошибки

распознавания при воздействии помех

Помеха	300-3400	w_noise	pulse	sin_left	sin_middle	sin_right	fading				
	«Ноль»										
$\hat{\mu}$	403,05	187,20	189,66	334,93	369,35	441,21	401,78				
$\hat{\sigma}$	183,51	87,60	103,13	172,06	131,53	180,27	178,47				
â	2,20	2,14	1,84	1,95	2,81	2,45	2,25				
$\Phi(\widehat{\alpha})$	0,4861	0,4838	0,4671	0,4744	0,4975	0,4931	0,4878				
\widehat{b}	0	0,1636	1,3652	0,8404	0,8218	0,4863	0,1208				
$\hat{p}_{ ext{out}}$ (%)	0,00	0,91	1,09	0,91	0,00	0,00	0,00				
			«Од	(ин»							
$\hat{\mu}$	355,78	118,26	121,27	161,61	273,91	301,96	340,85				
$\hat{\sigma}$	115,03	55,32	54,15	91,13	124,57	123,37	108,95				
â	3,09	2,14	2,24	1,77	2,20	2,45	3,13				
$\Phi(\widehat{\alpha})$	0,499	0,4838	0,4875	0,4616	0,4861	0,4929	0,4991				
\hat{b}	0	15,1647	11,5357	37,3336	12,8926	6,1372	0,1267				
$\hat{p}_{\text{ош}}$ (%)	0,00	2,73	1,82	3,64	0,73	1,82	0,00				

Из таблицы следует, что уменьшение числа ошибок распознавания сопровождается уменьшением отношения \hat{b} . Использование данного отношения становится особенно полезным, если число ошибок невелико или их вообще нет (при проведении тестирования на звуковой базе относительно небольшого объема), что не позволяет судить о значении вероятности ошибки с достаточной точностью.

Из таблицы также следует, что на результаты распознавания команды «Ноль» наибольшее влияние оказывает импульсная помеха. Следует отметить, что для данной команды воздействие среднечастотной гармонической помехи увеличивает достоверность распознавания относительно случая отсутствия помех. Использование относительной частоты $\hat{p}_{\text{ош}}$ ошибки распознавания в качестве оценки степени влияния помех на работу САРГК не позволяет обнаружить данный факт.

На результаты распознавания команды «Один» наибольшее влияние оказывает низкочастотная гармоническая помеха. В целом влияние помех на

распознавание данной команды намного сильнее по сравнению с ГК «Ноль», что можно объяснить наличием маломощного звука «d» в команде «Один».

Для обеих команд искажения AЧХ телефонной линии проявили себя одинаково слабо, что обусловлено выбранной системой параметров речевого сигнала, которая учитывает эти искажения.

3.2. Оценка влияния помех на параметры огибающей кратковременного спектра сигнала

Оценка влияния помех на достоверность распознавания осуществляется путем тестирования на достаточно объемной звуковой базе. Процедура тестирования и обработки его результатов требуют больших затрат времени. Поэтому на промежуточных этапах проектирования целесообразно оценить степень влияния на распознавания путем оценки изменений параметров огибающей спектра речевого сигнала (РС), которые являются исходными данными для алгоритма распознавания.

Наиболее целесообразно такой анализ проводить с использованием синтезированных звуков, которые широко используются при исследовании алгоритмов эффективного кодировании (сжатия) речевых сигналов [106, 114, 115]. Суть данного подхода заключается в следующем. Речевой сигнал, соответствующий ГК, разбивается на отдельные сегменты, длительность которых равна интервалу стационарности сигнала. Каждому звуку соответствует несколько сегментов. Для каждого сегмента по методу линейного предсказания представленная определяется огибающая спектра, либо коэффициентами линейного предсказания либо коэффициентами частной корреляции (КЧК), и сигнал остатка предсказания [96, 106].

Количество вариантов произнесения звука можно представить как произведение числа вариантов огибающей спектра на количество вариантов остатка предсказания. Информация для САРГК сосредоточена в огибающей

спектра, поэтому разнообразие вариантов произнесения звука можно свести к разнообразию вариантов огибающей спектра. Сигнал остатка предсказания можно упрощенно моделировать либо последовательностью коротких импульсов (для вокализованных звуков), либо белым шумом (для невокализованных звуков) [106]. Сигнал синтетического звука образуется прохождением модели сигнала остатка предсказания через синтезирующий фильтр, частотная характеристика которого определяется параметрами огибающей спектра, найденными ранее в процессе анализа сегмента речевого сигнала.

Схема использования синтезированных звуков при анализе влияния помех на параметры огибающей спектра, представлена на Рисунке.3.9.

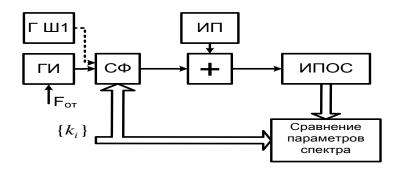


Рисунок 3.9.

Генератор белого шума (ГШ) и генератор импульсов (ГИ), используемые, соответственно, при синтезе невокализованных и вокализованных звуков, формируют сигнал возбуждения синтезирующего фильтра (СФ), частотная характеристика которого определяется набором $\{k_i\}$ коэффициентов частной корреляции (либо линейного предсказания). Параметр $F_{\rm or}$ (частота основного тона) задает частоту следования импульсов. На выходе СФ формируется сегмент сигнала соответствующего звука, к которому имитатор помех (ИП) добавляет помеху заданного вида и уровня. Измеритель параметров огибающей спектра (ИПОС) требуется для определения параметров огибающей спектра, которые используются в алгоритме распознавания ГК. Сравнение параметров спектра

позволяет определить степень отклонения параметров при воздействии помехи от их значений при отсутствии помехи.

В соответствии с данной схемой был проведен эксперимент по исследованию влияния шума и гармонических помех на значения 10 коэффициентов частной корреляции, соответствующих звуку «а0» (гласная «а» в ударном слоге) русской речи [58, 59].

Размер сегмента речевого сигнала был равен 200 отсчетам РС, что при частоте дискретизации 8кГц соответствовало длительности сегмента 25мс. Период основного тона данного РС равен 6,4 мс (51 отсчет) – частота основного тона 156 Гц. Рассматривалось отдельно воздействие шума и гармонической помехи, частота которой принимала значения 2600, 1000 и 400 Гц. Фиксировались измеренные (использовалось временное окно Хэмминга) значения КЧК при отсутствии помех, а также при изменении отношения сигнал-помеха в пределах от 15 до 45 дБ. Для синтеза сегмента РС, а также для его анализа использовались синтезирующий и анализирующий фильтры лестничной структуры.

На Рисунке 3.10 (а, б) приведены, соответственно, графики зависимостей оценок математических ожиданий (МО) и среднеквадратических отклонений (СКО) для каждого коэффициента частной корреляции (усреднение проводилось по 1000 реализациям) при воздействии шума с равномерной спектральной плотностью.

Видно, что зависимости средних значений, в общем случае, немонотонны, аппроксимацию аналитическими выражениями. что затрудняет ИХ уменьшением отношения сигнал-шум увеличивается СКО и величина отклонения МО для каждого КЧК. Справа от графика СКО для сравнения приведены СКО при естественной изменчивости натурального значения онжктопп произнесенного звука «а». Как видно, естественная изменчивость параметров звуков сравнима с СКО и отклонениями МО при отношениях сигнал-шум меньших 35 - 40 дБ.

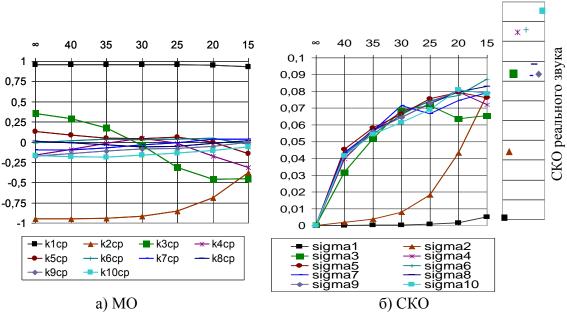


Рисунок 3.10. Графики зависимостей оценок математических ожиданий и среднеквадратических отклонений для каждого коэффициента частной корреляции от отношения сигнал-шум

На Рисунке 3.11 (а, б, в) приведены графики зависимостей КЧК от отношения сигнал-помеха для случая гармонической помехи на частотах 400, 1000 и 2600 Гц, соответственно.

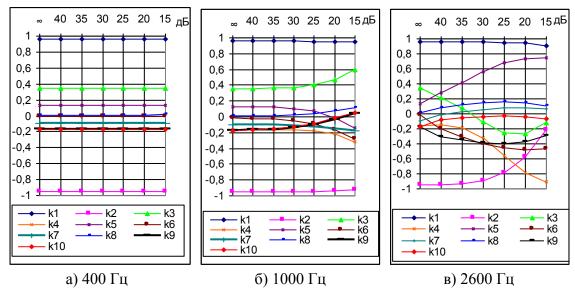


Рисунок 3.11. Графики зависимостей оценок математических ожиданий и среднеквадратических отклонений для каждого коэффициента частной корреляции от отношения сигнал-помеха при воздействии гармонической помехи.

По результатам эксперимента можно сделать следующие выводы. Погрешность растет немонотонно с уменьшением отношения сигнал-шум. Погрешность в определении КЧК с большими индексами больше погрешностей КЧК с малыми индексами. С уменьшением частоты помехи снижается степень ее влияния на КЧК, что определяется формой огибающей кратковременного спектра синтезированного звука (Рисунок 3.12). Видно, что в области низких частот мощная форманта на 500 Гц. Поэтому наличие присутствует частоте гармонической помехи с частотой 400 Гц относительно слабо сказывается на изменении КЧК.

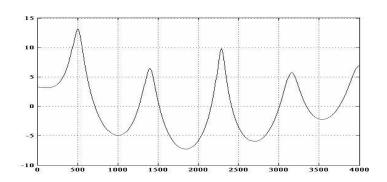


Рисунок 3.12. Огибающая кратковременного амплитудного спектра синтезированного звука Главный вывод: на параметры данного звука наибольшее влияние оказывают высокочастотные помехи.

3.3. Подавление помех на основе оценки спектра в паузах речевого сигнала

Если помеха является стационарной на интервале, длительность которого равна сумме длительностей интервалов ГК и предшествующей паузы, то, оценив параметры помехи на интервале паузы при отсутствии сигнала, можно использовать полученные оценки для подавления помех на интервале действия ГК. Однако, если длительность паузы невелика, то погрешность оценки значительна, что сказывается на степени подавления помехи.

В стандарте Европейского института стандартов в области телекоммуникаций ETSI ES 202 212 V1.1.2 (2007-01) [146] для распределенных систем автоматического распознавания речи описывается метод подавления помех на основе реализации, близкой к фильтру Винера. АЧХ физически нереализуемого фильтра Винера описывается выражением [40]

$$H(\omega) = \frac{G(\omega)}{G(\omega) + N(\omega)} = \frac{\eta(\omega)}{1 + \eta(\omega)},$$

где $G(\omega)$, $N(\omega)$, $\eta(\omega) = G(\omega)/N(\omega)$ - спектры мощностей сигнала, помехи и отношение сигнал-помеха, соответственно. В указанном методе используются соответствующие оценки спектров, полученные с помощью быстрого преобразования Фурье.

В исследуемом алгоритме используется двухэтапная фильтрация помех. На первом этапе с помощью детектора голосовой активности абонента выделяются интервалы времени, где присутствует только помеха. На этих интервалах производится оценка спектра помехи с помощью быстрого преобразования Фурье. Оценка спектра осуществляется по кадрам: каждые 10 мс с длительностью кадра 25 мс при частоте дискретизации 8 кГц. Далее вычисляются квадраты модулей коэффициентов БПФ, которые выступают в роли оценок спектров мощности помехи и сигнала совместно с помехой, и вся последующая обработка сигнала происходит в частотной области. Оценка спектра сигнала без помехи осуществляется согласно следующему выражению.

$$X_{\text{den}}(\omega, t) = \beta X_{\text{den}}(\omega, t - 1) + (1 - \beta)[X(\omega, t) - X_{\text{noise}}(\omega, t)]$$
(3.3)

Здесь $X_{\rm den}(\omega,t)$ - модуль коэффициента БПФ сигнала, очищенного от помехи, на частоте ω , определенный на фрейме в момент времени t; $X_{\rm den}(\omega,t-1)$ - аналогичный коэффициент для предыдущего момента времени; $X_{\rm noise}(\omega,t)$ - аналогичный коэффициент для помехи, определенный на интервале паузы; β - коэффициент близкий единице. На этом же этапе для каждого фрейма (кадра) речевого сигнала формируется частотная характеристика фильтра

$$H_1(\omega) = \frac{\sqrt{\hat{\eta}(\omega)}}{1 + \sqrt{\hat{\eta}(\omega)}},$$

где $\hat{\eta}(\omega) = (X_{\text{den}}(\omega, t)/X_{\text{noise}}(\omega, t))^2$ — оценка отношения сигнал-помеха, полученная в виде отношения соответствующих оценок спектров мощности.

Затем речевой сигнал пропускается через этот фильтр — осуществляется предварительная фильтрация сигнала. На втором этапе обрабатывается сигнал после предварительной фильтрации: вновь оценивается спектр помехи, формируется частотная характеристика фильтра, и осуществляется повторная фильтрация сигнала.

Для оценки эффективности указанного метода подавления помех, был проведен следующий эксперимент [90, 93, 167]. На речевые сигналы, которые соответствовали произнесениям ГК, накладывалась помеха, представляющая собой аддитивную смесь «розового» шума и периодической импульсной последовательности (запись с телефонной линии). Спектр помехи представлен на Рисунке 3.13.

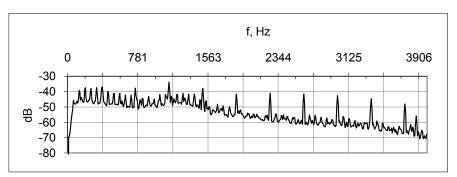


Рисунок 3.13. Спектр помехи (запись с телефонной линии)

В качестве голосовых команд использовались названия всех цифр, а также произнесения слов: «да», «нет» и «ошибка». Отношение сигнал-помеха при этом составило 25 дБ.

Полученная смесь сигнала и помехи подвергались фильтрации. Затем для выборки голосовых команд без помехи, выборки с шумом и выборки, полученной на выходе фильтра, вычислялись кепстральные коэффициенты, а также их первые и вторые производные. Данные наборы параметров сравнивались между собой:

вычислялось две дисперсии: разности параметров сигнала при отсутствии помехи и сигнала с помехой, а также разности параметров сигнала при отсутствии помехи и сигнала с отфильтрованной помехой. За счет фильтрации дисперсия разности снизилась на 39%.

Затем проводилось контрольное распознавание голосовых команд с использованием фильтрации. В эксперименте использовалась тестирующая выборка, состоящая из 3900 произнесений указанных выше команд, записанных шестью дикторами (1 женщина, 5 мужчин). Каждый диктор по 50 раз произнес каждую команду. Данная выборка была записана в лабораторных условиях. При этом отношение сигнал – помеха составило более 45 дБ.

При создании моделей ГК в качестве обучающей использовалась указанная выше тестирующая выборка ГК. Результаты распознавания: для тестовой выборки без помехи частота ошибок распознавания составила 0,18%; для выборки с помехой - 39%; для выборки с помехой после фильтрации – 6,3%.

Наиболее эффективно рассмотренный алгоритм фильтрации работает при достаточно медленных изменениях параметров помехи. Только в этом случае можно оценить спектр помехи с малой погрешностью

Следует отметить, что при оценке эффективности рассмотренного метода подавления помех использовалась статистическая оценка степени искажений параметров РС для набора ГК. Существенное снижение степени искажений параметров РС обусловило значительный рост достоверности распознавания. Использование алгоритмов, похожих на данный алгоритм, нашло отражение во многих публикациях [201].

3.4. Подавление квазипериодических помех

Недостатком рассмотренного выше метода подавления помехи является искажение спектра сигнала, обусловленное прохождением сигнала через фильтр.

Кроме того, рассмотренный выше алгоритм имеет ограниченные возможности подавления импульсных помех. Рассмотрим случай импульсной квазипериодической помехи и возможность ее подавления, когда указанное подавление не сопровождается значительными искажениями сигнального спектра. Считаем, что помеха $\xi(t)$ является суммой стационарного шума n(t) и медленно меняющейся во времени мощной периодической составляющей u(t).

$$\xi(t) = u(t) + n(t).$$
 (3.4)

Периодическую составляющую можно подавить, вычитая копию фрагмента помехи в паузе, смещенную во времени на интервал кратный периоду помехи, из сигнала, пораженного помехой [39, 49, 63]. В работах [39, 49] рассмотрены возможности адаптивной и неадаптивной компенсации помехи без учета погрешности определения периода помехи. Следует отметить, что из-за малой длительности команд сложно получить точные значения коэффициентов весового суммирования при адаптивной компенсации. В данном разделе рассматривается влияние указанной погрешности на результаты неадаптивной компенсации. Более детально алгоритм подавления помехи выглядит следующим образом [63].

Детектор голосовой активности [54,159] определяет интервал паузы, где нет сигнала, и интервал ГК. На интервале паузы определяется период T_p детерминированной составляющей u(t). Затем из паузы формируется фрагмент, длительность которого равна целому числу периодов детерминированной составляющей помехи. Данный фрагмент задерживается на целое число L периодов (сдвигается по оси времени), умножается на некоторый коэффициент W и вычитается из ГК, пораженной помехой — происходит взаимная компенсация помехи, присутствующей на интервале действия ГК, и фрагмента помехи в паузе. Затем фрагмент вновь сдвигается по оси времени и вновь вычитается из сигнала, пораженного помехой. Процессы сдвига и вычитания продолжаются до тех пор, пока сигнал на протяжении всей своей длительности не будет очищен от помехи.

Проанализируем эффективность данного подхода, используя теорию адаптивной компенсации помех [98, 118]. Из теории следует, что, если мгновенные мощности копии «паузной» помехи и помехи, присутствующей на интервале ГК, одинаковы, то для максимального подавления помехи (по критерию минимума среднего квадрата ошибки) необходимо «паузную» помеху, задержанную во времени, умножить на коэффициент

$$W_0 = \frac{\overline{\xi(t)\xi(t - LT_p)}}{\overline{\xi(t - LT_p)^2}} = \rho_{\xi}(LT_p).$$
(3.5)

Здесь $\xi(t)$, $\xi(t-LT_p)$ - реализации помехи в моменты времени t и $(t-LT_p)$, где L – целое число, T_p – период детерминированной составляющей помехи, LT_p - величина задержки паузной помехи. Черта сверху означает процедуру усреднения по ансамблю реализаций. Так как $\overline{\xi^2(t-LT_p)}=\overline{\xi^2(t)}$, то $W_0=\rho_\xi(LT_p)$, где $\rho_\xi(LT_p)$ - нормированная автокорреляционная функция $\rho_\xi(\tau)$ помехи при $\tau=LT_p$.

Учитывая, что $\overline{\xi(t)\xi(t-LT_p)}=u^2(t)+\overline{n(t)n(t-LT_p)},$ а $\overline{\xi(t)^2}=u^2(t)+\overline{n^2(t)}$ и считая, что процессы n(t) и $n(t-LT_p)$ не коррелированы, получаем

$$W_0 = \frac{u^2(t)}{u^2(t) + \overline{n^2(t)}}. (3.6)$$

Усредняя W_0 на периоде T_p , получаем

$$W_{01} = \frac{P_u}{P_u + P_n} = \frac{q}{1+q} , \qquad (3.7)$$

где $q = P_u/P_n$. Здесь $P_u = \langle u^2(t) \rangle$, P_n , - средние мощности периодической составляющей и шума, соответственно.

Определим теперь остаточную мощность помехи после окончания процесса компенсации.

$$\overline{\varepsilon^{2}(t)} = \overline{[s(t) + \xi(t) - W_{01}\xi(t - LT_{p})]^{2}} = \overline{s^{2}(t)} +$$
(3.8)

$$+\overline{\xi(t)^2}(1-W_0^2) = \overline{s^2(t)} + [u^2(t) + \overline{n^2(t)}](1-W_0^2)$$

Здесь $\varepsilon(t) = s(t) + \xi(t) - W_{01}\xi(t-LT_p)$ — напряжение на выходе вычитающего устройства; s(t) - полезный сигнал. Последнее слагаемое в полученном выражении соответствует мощности помехи после компенсации. Усредняя данное слагаемое на периоде T_p , получаем среднюю мощность помехи после компенсации

$$P_{\xi \text{ oct}} = (P_u + P_n) \left[1 - \left(\frac{P_u}{P_u + P_n} \right)^2 \right] = P_n \left(1 + \frac{q}{1 + q} \right)$$
 (3.9)

Из полученного выражения следует, что при увеличении q от нуля до бесконечности остаточная мощность помехи меняется от P_n до $2P_n$. Мощность остаточной после компенсации помехи практически определяется лишь уровнем шума. Если периодическая оставляющая помехи достаточно мощная, то величина W_{01} близка единице.

Так как при определении значения $\rho_{\xi}(LT_p)$ автокорреляционной функции помехи появляются погрешности, которые могут быть значительными при малой длительности паузы, то при большой мощности периодического компонента (q>>1) целесообразно вместо оптимального коэффициента W_{01} использовать W=1. Определим мощность $P_{\xi \text{ ост}}$ остаточной помехи в этом случае. Так как $\varepsilon(t) = s(t) + \xi(t) - \xi(t - LT_p)$, то суммарная мощность сигнала и помехи на выходе компенсатора

$$\overline{\varepsilon^2(t)} = \overline{[s(t) + \xi(t) - \xi(t - LT_p)]^2} =$$

$$= \overline{s^2(t)} + \overline{2\xi^2(t)} - \overline{2\xi(t)\xi(t - LT_p)}$$
(3.10)

Остаточная мощность помехи

$$P_{\xi \text{ oct } 1} = \langle \overline{2\xi^2(t)} - \overline{2\xi(t)\xi(t - LT_p)} \rangle = 2\overline{n^2(t)} = 2P_n$$
 (3.11)

Сравним мощности остаточной помехи при $W=W_{opt}$ и при W=1.

$$\lambda = \frac{P_{\xi \text{ oct } 1}}{P_{\xi \text{ oct}}} = 1 + \frac{1}{1 + 2q}$$
 (3.12)

При q>2 величина λ <1,2. То есть проигрыш в подавлении помехи при использовании W=1 вместо W= W_{opt} очень мал и составляет менее 0,8дБ. Следовательно, целесообразно использовать W=1 и не определять значение $\rho_{\xi}(LT_p)$ автокорреляционной функции помехи.

Определим минимальное значение q, при котором компенсация обеспечивает достаточно глубокое подавление помех при W=1. Уровень подавления помехи

$$\gamma = \frac{P_{\xi}}{P_{\xi \text{ oct } 1}} = \frac{P_u + P_n}{2P_n} = 0.5(1+q)$$
(3.13)

Если задаться величиной $\gamma > 2$, то величина q должна быть больше трех.

Рассмотрим теперь влияние погрешности в определении периода T_p на степень подавления помехи. Определим период как минимальное значение аргумента автокорреляционной функции помехи не равное нулю, при котором функция принимает максимальное значение. Имеют место два источника погрешностей при определении аргумента: дискретизация аргумента функции и влияние шума на результат определения максимума функции.

Шаг дискретизации равен периоду T_d дискретизации речевого сигнала. Максимальная погрешность определения аргумента равна половине периода дискретизации. Если помеха содержит гармоническое колебание с частотой f, то нормированная автокорреляционная функция колебания

$$\rho_{u}(\tau) = \cos(2\pi f \tau). \tag{3.14}$$

Первый максимум функции имеет место при $\tau = \tau_0 = 1/f$. Если определяется аргумент не первого, а m–го максимума, то в случае максимальной погрешности определения аргумента за счет дискретизации значение автокорреляционной функции

$$\rho_u(m\tau_0 + 0.5T_d) = \cos(2\pi m f \tau_0 + \pi f T_d) = \cos(\pi f T_d) = \cos(\pi f / f_d). \quad (3.15)$$

Здесь f_d =1/ T_d — частота дискретизации. Погрешность определения периода в этом случае не превышает величины $0.5T_d/m$. Погрешность, возникающая при

определении времени задержки копии паузной помехи, не превышает $0.5T_d$, если $L \leq m$. В противном случае погрешность определения задержки возрастает. Если указанное условие выполняется, то при W=1 остаточная после компенсации мощность помехи

$$P_{\xi \text{ oct } 1} = \langle \overline{2\xi^{2}(t)} - \overline{2\xi(t)\xi(t - LT_{p} - 0.5T_{d})} \rangle = 2P_{u} + 2P_{n} -$$

$$-2P_{u}\rho_{u}(LT_{p} + 0.5T_{d}) = 2P_{n} + 2P_{u}[1 - \cos(\pi f/f_{d})].$$
(3.16)

Величина подавления помехи

$$\gamma = \frac{P_{\xi}}{P_{\xi \text{ oct } 1}} = \frac{P_u + P_n}{2P_n + 2P_u[1 - \cos(\pi f/f_d)]} = \frac{0.5(1+q)}{1 + q[1 - \cos(\pi f/f_d)]},$$
 (3.17)

Из выражения (3.15) следует, что для уменьшения погрешности, возникающей за счет дискретизации, целесообразно частоту дискретизации выбирать достаточно большой. В частности, при $f/f_d \le 0.25$ и q=4 величина $\gamma \ge 1.15$ (0.6дБ). Если задаться величиной γ подавления помехи, то можно определить максимально допустимое отношение f/f_d .

Оценим теперь влияние шума на погрешность определения периода детерминированной составляющей помехи. Считаем, что период детерминированной составляющей помехи равен целому числу периодов дискретизации, и погрешность за счет дискретизации отсутствует. Определим влияния шума оценка $\hat{\rho}_u(m\tau_0 + T_d)$ за счет вероятность автокорреляционной функции станет больше оценки $\hat{
ho}_u(m au_0)$. В этом случае возникает погрешность определения периода равная периоду дискретизации.

Если период T_p определяется по аргументам максимумов автокорреляционной функции помехи, то необходимо проанализировать значение случайной величины $\Delta \hat{R}(1) = \hat{R}(LB) - \hat{R}(LB+1)$, где $\hat{R}(LB)$ и $\hat{R}(LB+1)$ оценки (ковариационный метод) [106,112] автокорреляционной функции

$$\hat{R}(LB) = \frac{1}{N - LB} \sum_{i=LB}^{N-1} \xi(i) \ \xi(i - LB)$$

$$\widehat{R}(LB+1) = \frac{1}{N-LB-1} \sum_{i=LB+1}^{N-1} \xi(i) \ \xi(i-LB-1)$$
(3.18)

Здесь B — период детерминированной составляющей помехи, выраженный в виде числа периодов T_d дискретизации; L - порядковый номер рассматриваемого максимума автокорреляционной функции; N — длительность паузы, выраженная через число периодов T_d дискретизации. Величина $\xi(i)$ определяется выражением (3.4).

Если случайная величина $\Delta \hat{R}(1) > 0$, то при значении дискретного времени равном LB оценка автокорреляционной функции периодической составляющей помехи принимает максимальное значение, и величина периода детерминированной составляющей помехи определяется без погрешности. Если же $\Delta \hat{R}(1) < 0$, то оценка автокорреляционной функции принимает максимальное значение при значении дискретного времени равном LB+1. В этом случае величина периода детерминированной составляющей помехи определяется с погрешностью равной периоду дискретизации.

Если шум гауссов, то вероятность $p_{\text{ош}}$ того, что $\Delta \hat{R}(1) < 0$, определяется выражением [15]

$$p_{\text{ош}} = 0.5 - \Phi(\alpha),$$

$$\Phi(\alpha) = \frac{1}{\sqrt{2\pi}} \int_{0}^{\alpha} e^{-z^{2}/2} dz , \alpha = \frac{\mu_{\Delta}}{\sqrt{D_{\Delta}}}$$
(3.19)

Здесь $\Phi(\alpha)$ - табулированная функция Лапласа.

Математическое ожидание μ_{Δ} случайной величины $\Delta \hat{R}(1)$ равно разности математических ожиданий случайных величин: $\hat{R}(LB)$, $\hat{R}(LB+1)$

$$\mu_{\Delta} = \mu(LB) - \mu(LB + 1) =$$

$$= \frac{1}{N - LB} \sum_{i=0}^{N-1-LB} u^{2}(i) - \frac{1}{N - LB - 1} \sum_{i=0}^{N-2-LB} u(i)u(i+1) \approx$$

$$\approx R_{u}(0) - R_{u}(1)$$
(3.20)

Приближенное равенство тем точнее выполняется, чем в большей степени соблюдается условие N-LB>>LB, и чем больше L. Величина дисперсии D_{Δ} случайной величины $\Delta \hat{R}(1)$ определяется выражением

$$D_{\Delta} = \sigma_n^2 \sum_{i=0}^{N-LB-2} \left[\left(\frac{u(i+1)}{N-LB} - \frac{u(i)}{N-LB-1} \right)^2 + \left(\frac{u(i)}{N-LB} - \frac{u(i+1)}{N-LB-1} \right) \right]^2 +$$

$$+ \sigma_n^2 \frac{u^2(N-1) + u^2(0)}{(N-LB)^2} + 2\sigma_n^4 \frac{(N-LB)^2 - (N-LB) + 1}{(N-LB)^2 * (N-LB-1)}$$
(3.21)

Выражения для μ_{Δ} и D_{Δ} получены согласно выкладкам, приведенным в Приложении. Полученные выражения были проверены с помощью моделирования. Помеха была представлена суммой гауссова белого шума со среднеквадратическим отклонением σ =8 и гармонического колебания с частотой f_0 =400 Γ ц и амплитудой A=10*2 0,5 . Величина q= $A^2/(2*\sigma^2)$ =100/64=1,5625, что составляет 3,9дБ. Частота дискретизации f_d =8к Γ ц. Длительность паузы N=30 отсчетов сигнала, период B= f_d/f_0 =8000/400=20 отсчетов сигнала, номер максимума автокорреляционной функции L=1.

Было сформировано 10^5 реализаций шума, и для каждой реализации вычислялись значения $\mu(LB)$ и $\sigma(LB)$. Расчетные значения составили 100 и 41,1, соответственно, а результаты моделирования: 100,0945 и 41,0005. Для значений $\mu(LB+1)$ и $\sigma(LB+1)$ имеем результаты вычислений: 84,54 и 41,65 и результаты моделирования: 84,28 и 41,68. Результаты моделирования подтверждают правильность формул для $\mu(LB)$, $\sigma(LB)$, $\mu(LB+1)$ и $\sigma(LB+1)$.

Вычисленная величина среднеквадратического отклонения разности оценок значений корреляционной функции $\sqrt{D_{\Delta}}=34,98$. Вычисленное значение

$$\alpha = \frac{\mu_{\Delta}}{\sqrt{D_{\Delta}}} = \frac{100 - 84,54}{34,98} = 0,44$$

Соответствующее значение вероятности появления погрешности равной периоду дискретизации

$$p_{\text{погреш.}} = 0.5 - \Phi(\alpha) = 0.5 - 0.17 = 0.33$$

По результатам моделирования $p_{\text{погреш.мод.}} = 0.32$. Следовательно, моделирование подтверждает справедливость выражений (3.27) и (3.28).

Проанализированный алгоритм компенсации помех был использован для подавления импульсной помехи (сигнал «занято» в телефонной линии). Форма помехи: «пачки» длительностью 0,25с искаженных импульсов, следующих с частотой 0,3Гц. Частота следования импульсов внутри пачки - 500Гц. Кроме того, на речевой сигнал накладывались шум и относительно импульсов слабая гармоническая помеха с частотой 50Гц. Частота дискретизации составила 10кГц. Глубина квантования — 16 бит. На Рисунке 3.14 представлена «осциллограмма» сигнала, пораженного помехой.

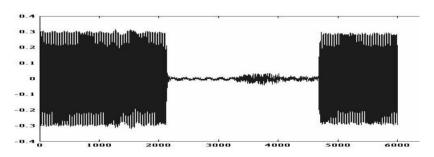


Рисунок 3.14. «Осциллограмма» сигнала, пораженного помехой

На Рисунке 3.15 представлена «осциллограмма» фрагмента помехи. По осям абсцисс на обоих рисунках отложены номера отсчетов сигналов.

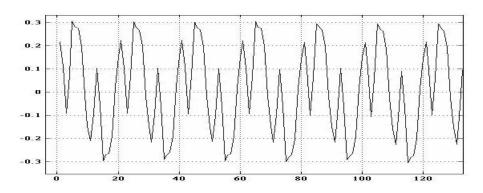


Рисунок 3.15. «Осциллограмма» фрагмента помехи

Длительность фрагмента паузы, используемого для компенсации, была установлена равной периоду следования импульсов — 2мс. С целью исключения процедуры компенсации на интервале отсутствия импульсов контролировалось изменение энергии сигнала, пораженного помехой, при проведении каждого этапа компенсации. Если энергия увеличивалась, то процедура не проводилась.

На Рисунке 3.16 представлена «осциллограмма» сигнала после компенсации.

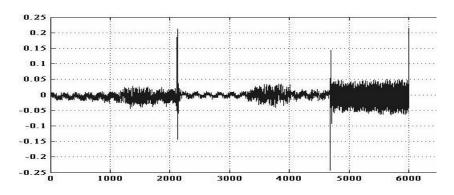


Рисунок 3.16. «Осциллограмма» сигнала после компенсации

Подавление помехи на интервале длительности первой «пачки» импульсов составило 27,3дБ, а на интервале второй - 17,3дБ. Вторая «пачка» подавлена в меньшей степени за счет нестабильности временного сдвига последовательности импульсов в пачке. Следует отметить, что длительность остаточных после компенсации импульсных помех в моменты времени, соответствующие началу и окончанию пачки, растет с увеличением длительности фрагмента помехи, используемого для компенсации.

 \mathbf{C} целью подавления гармонической помехи проведена процедура фрагментом 20мс. повторной компенсации длительностью паузы «Осциллограмма» сигнала после повторной компенсации представлена на Рисунке 3.17.

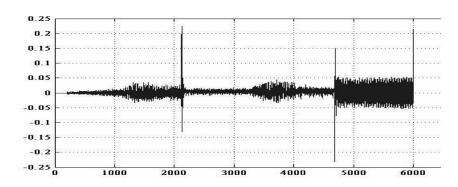


Рисунок 3.17. «Осциллограмма» сигнала после повторной компенсации

Величина подавления гармонической помехи на интервале длительности (0...0,1c) первой «пачки» импульсов составила 8,9дБ. Суммарная величина подавления помех на указанном интервале действия помехи равна 27,3+8,9=36,2 дБ. Подавление помехи на интервале длительности второй «пачки» импульсов составило 0,6дБ. Суммарная величина подавления помех на указанном интервале действия помехи равна 17,3+0,6=17,9 дБ. Следует отметить, что уровень шума после компенсации становится выше.

Для увеличения степени подавления помех в последующих пачках, необходимо учесть изменение состава текущей пачки по отношению к соседним пачкам. Из текущей пачки выделяется фрагмент помехи длительностью равной периоду. Он используется для компенсации. Соответствующий алгоритм подавления помехи выглядит следующим образом.

- 1. С помощью пороговой функции определяются номера отсчетов сигнала, пораженного помехой, которые соответствуют наличию импульсов помехи.
- 2. С помощью корреляционной функции определяется период Т колебаний, «заполняющих» импульс помехи.

- 3. По массивам данных свободных от помехи определяется интервал τ_s корреляции речевого сигнала. Формируется копия массива данных сигнала, пораженного помехой, задержанная во времени на величину $\tau = kT \ge \tau_s$.
- 4. Копия вычитается из массива данных сигнала, пораженного помехой.
- 5. Анализируется полученная разность. На интервалах времени, соответствующих присутствию импульсов помехи, находятся сегменты сигнала длительностью T, имеющие минимальную энергию на данном интервале. Фиксируются номера отсчетов, соответствующие границам сегмента.
- 6. Из массива данных сигнала, пораженного помехой, формируются сегменты на основе зафиксированных номеров отсчетов сигнала. Для каждого импульса помехи формируется свой сегмент.
- 7. Путем перемещения копии полученных сегментов в пределах соответствующих импульсов помехи выполняется компенсация помехи.
- 8. Для устранения остаточных после компенсации коротких импульсов помехи используется экстраполяция соседних с импульсами фрагментов сигнала на интервале длительности короткого импульса.

Выводы по разделу 3

- 1. Предложена методика оценки влияния помех на достоверность распознавания голосовых команд, использующая критерий достоверности распознавания.
- 2. Методика обеспечивает повышение чувствительности оценки по сравнению с известным методом. В частности, при распознавании команды «ноль» в случае воздействия гармонической помехи частотой 1000 Гц относительное изменение параметра b, монотонно связанного с относительным изменением вероятности ошибки, равно 0.8218. При этом относительная частота ошибки распознавания не изменилась, оставаясь равной нулю.

- 3. Проведенное экспериментальное исследование влияния помех на параметры огибающей кратковременного спектра сигнала показывает, что влияние узкополосной помехи снижается при совпадении ее частоты с частотой форманты. Целесообразно анализ влияния узкополосных помех проводить с учетом формантной структуры синтезированных сегментов речевого сигнала.
- 4. Показано, что параметры огибающей кратковременного спектра немонотонно зависят от отношения сигнал-помеха, что затрудняет аппроксимацию указанной зависимости. Требуется проводить экспериментальное исследование системы для различных значений отношения сигнал-помеха.
- 5. Проведенные эксперименты показали, что метод спектрального вычитания обеспечивает значительное снижение влияния помех на параметры огибающей кратковременного спектра сигнала И на достоверность распознавания. Целесообразно тестирование системы с определением параметров достоверности распознавания предварять менее затратной оценкой степени изменения параметров РС. При исследовании использована статистическая оценка степени искажений параметров РС для набора голосовых команд. Отмечено снижение степени подавления помех при увеличении скорости изменения их параметров.
- 6. Предложен алгоритм компенсации помех, представляющих собой аддитивную смесь шума и периодических компонентов с медленно меняющейся частотой.
- 7. Получены выражения, отражающие влияние частоты дискретизации и отношений «сигнал-шум», «периодический компонент-шум» на погрешность оценки периода помехи. Достоверность выражений подтверждена моделированием в системе Matlab, а эффективность компенсации результатами обработки записи сигнала с телефонной линии. Импульсная помеха была подавлена на 27,3 дБ.
- 8. Полученные выражения позволяют рассчитать минимально допустимую частоту дискретизации сигналов, а также минимально допустимые значения отношений «сигнал-шум», «периодический компонент-шум».

4. Исследование адаптивной компенсации радиопомех

Воздействие радиопомех в аналоговых системах связи проявляется в виде электрических помех на этапе предварительной обработки РС. В цифровых системах связи при пакетной передаче данных речи возможно пропадание пакетов, что проявляется в виде дополнительных пауз в речи. Наличие пауз значительно снижает достоверность распознавания команд.

Проблема обеспечения помехоустойчивости наиболее остро стоит при передаче информации по каналам связи с селективно-частотными замираниями сигнала. Такие каналы характерны, в частности, для мобильной и тропосферной радиосвязи.

Для борьбы с помехами в каналах с замираниями сигнала применяют разнесенный прием, помехоустойчивое кодирование и интерливинг [1, 11, 19, 20, 28, 31, 113]. Повышение помехоустойчивости требует увеличения мощности помехоустойчивого кода, что с учетом ограниченной пропускной способности канала связи приводит к снижению потока данных, предназначенных конечному пользователю системы связи.

Существует [137, 146] вариант решения данной проблемы путем организация передачи данных, предназначенных только для САРГК. В этом случае можно увеличить мощность помехоустойчивого кода без большого ущерба для передачи других потоков данных в канале связи, так как объем параметров РС для САРГК меньше по сравнению с объемом данных речи, предназначенных для прослушивания человеком. Однако такой подход требует разработки специальных программно-аппаратных средств для пользователя.

Значительного подавления помех можно достичь с помощью адаптивной компенсации помех (АКП) путем организации пространственно-разнесенного приема [21, 32, 98]. Однако при реализации АКП возникают следующие проблемы. Во-первых, из-за наличия сооружений вблизи приемных антенн,

которые могут выступать в роли отражателей помех, возможно снижение коррелированности помеховых колебаний на приемных антеннах, что снижает степень их подавления. Во-вторых, влияние сигнальных компонентов на работу системы адаптивной компенсации может привести к их ослаблению на входе демодулятора.

В данном разделе рассмотрены возможности повышения степени компенсации широкополосных помех за счет использования адаптивных фильтров в трактах приема сигнала. Кроме того, рассмотрены меры по подавлению влияния полезного сигнала на работу алгоритма адаптивной компенсации помех.

4.1. Возможности компенсации широкополосных помех при наличии их отражений

Эффективным средством подавления помех является ИХ взаимная компенсация на основе использование нескольких приемных антенн. Часто рядом расположены сооружения, антеннами крупные поэтому непосредственным воздействием излучения от источников помех, на антенны поступают и большие по уровню отражения помех от этих сооружений. Отражения, имея различные временные запаздывания, уровни и сдвиги фаз, суммируются с помехами, поступившими непосредственно от источников излучения, и вызывают уменьшение степени корреляции между помеховыми компонентами в каналах приема. Поэтому эффективность использования весового суммирования (ВС) колебаний с выходов приемных антенн для подавления помех снижается. Это снижение тем значительнее, чем шире спектр помехи [98].

Помещением трансверсальных фильтров (ТФ) в каналы приема обеспечивается подавление помех [98]. Но для глубокого подавления помех требуется большое число отводов линии задержки (ЛЗ), что вызывает значительное усложнение системы адаптивного ВС.

В данном разделе рассматривается возможность использования ТФ с неравномерным расположением небольшого числа отводов в ЛЗ, что позволяет снизить сложность последующего адаптивного ВС [52, 65, 83, 121]. Такая возможность появляется, если известны значения задержек отражений. Данные фильтры моделируют тракты распространения помехи от источника до приемных антенн. На Рисунке 4.1 представлена структурная схема двухвходового компенсатора помех (КП) совместно с фильтрами, моделирующими тракты распространения помех.

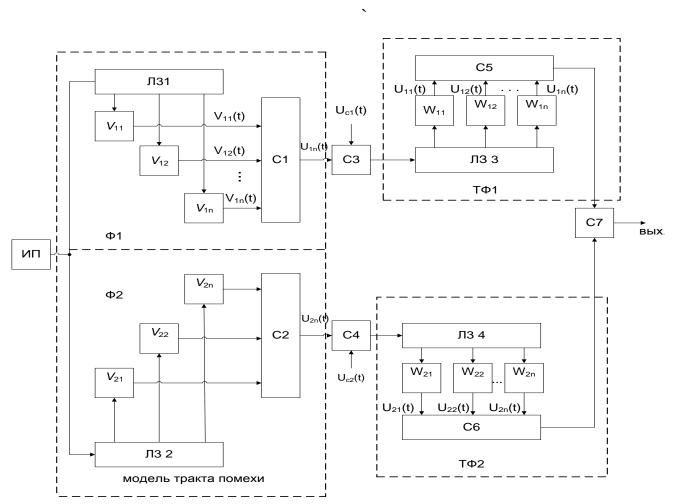


Рисунок 4.1. Структурная схема двухвходового компенсатора помех.

Здесь Ф1 и Ф2 - моделирующие ТФ для первой и второй антенн приемника, соответственно, ТФ1 и ТФ2 - трансверсальные фильтры, входящие в состав КП, ИП - источник помехи, Л31-Л34 - многоотводные линии задержки, С1 ... С7 –

сумматоры, V_{1i}, V_{2i}, W_{1i} и $W_{2i},$ где $i \in [1, \widetilde{M}],$ - коэффициенты весового суммирования ТФ, $U_{c1}(t), U_{c2}(t), U_{n1}(t), U_{n2}(t)$ - напряжения сигналов и помех на первой и второй антеннах, соответственно. Длины ЛЗ 3, ЛЗ 4 определяются максимальным запаздыванием помехи, отраженной от переизлучателя или принятой непосредственно от источника, относительно самого раннего момента времени t_0 приема помехи одной из антенн. Расстояние между используемыми отводами ЛЗ определяется запаздыванием отражений относительно Коэффициенты: w_{1i} , w_{2i} весового суммирования напряжений с выводов ЛЗ при погрешности ($\Delta \tau = 0$) реализации нулевой требуемых задержек соответствующим комплексным коэффициентам v_{2i}, v_{1i} передачи помехового напряжения по трактам распространения "источник излучения - отражатель антенна приемника".

Реализация требуемой величины относительного временного запаздывания возможна лишь с некоторой погрешностью, обусловленной неточностями измерения запаздываний и изготовления ЛЗ. Определим минимально возможное в этом случае отношение q_n помеха-сигнал на выходе КП при оптимальных значениях коэффициентов весового суммирования (КВС) в ТФ. Ситуация, наиболее трудная для подавления помех, характеризуется следующими положениями:

- помеха полностью перекрывает по спектру полезный сигнал;
- -отражения помехи поступают на антенны с различных направлений от достаточно далеко разнесенных отражателей, что приводит к некоррелированности помеховых компонентов на соседних отводах ЛЗ;
- абсолютные величины погрешностей реализации относительных запаздываний в отводах ЛЗ максимальны, причем знак погрешности в ТФ1 противоположен знаку погрешности во втором ТФ.

Считаем, что интервал корреляции широкополосного сигнала, который при

разработке стараются сделать минимальным, меньше расстояния между соседними отводами ЛЗ в составе КП (случай "больших задержек"). Поэтому сигнальные компоненты на выходах ЛЗ не коррелированы. Пространственное разнесение антенн обусловливает слабую корреляцию селективно-частотных замираний сигналов в каналах приема. В работах [48, 88] показано, что при достаточно широкой полосе принимаемых сигналов и наличии селективно-частотных замираний корреляционную матрицу сигналов на входах компенсатора можно считать диагональной. На основании вышеизложенного корреляционная матрица сигналов

$$R_{c} = P_{co} \begin{bmatrix} E & 0 \\ 0 & E \end{bmatrix} \tag{4.1}$$

где P_{cO} - мощность сигнала на входе ЛЗ; E, 0 - соответственно единичная и нулевая матрицы размерности М \times М. При указанных условиях корреляционная матрица помех

$$R_{\Pi} = P_{\Pi 0} \begin{bmatrix} \vec{F}_{1}^{+} \vec{F}_{1} E & \rho_{\Delta} \vec{F}_{2} \vec{F}_{1}^{+} \\ \rho_{\Delta} \vec{F}_{1} \vec{F}_{2}^{+} & \vec{F}_{2}^{+} \vec{F}_{2} E \end{bmatrix}, \tag{4.2}$$

где $P_{\text{по}} = \overline{U}_{\text{m}}^2/2R_{\text{вх}}$ - мощность максимального по уровню переотражения на входе ЛЗ; $R_{\text{вх}}$ - входное сопротивление ЛЗ; $U_m(t)$ - модуль комплексной огибающей максимального по уровню отражения (или прямого прохождения помехи) на входе ЛЗ; $\vec{F}_1 = \Phi_2 \vec{\theta}_1$; $\vec{F}_2 = \Phi_1 \vec{\theta}_2$; $\Phi_1 = diag[\exp(j\varphi_{ii}^{(1)})]$ - диагональная матрица фазовых сдвигов на выводах ЛЗЗ, обусловленных погрешностями реализации запаздываний отражений, взятых с противоположными знаками, для ТФ1; $\Phi_2 = diag\left[\exp\left(j\varphi_{ii}^{(2)}\right)\right]$ - матрица фазовых сдвигов для ТФ2; $\vec{\theta}_1^T = [V_{11}, V_{12}, \cdots, V_{1M}]$ - вектор коэффициентов ослаблений отражений для первого тракта распространения помехи; $\vec{\theta}_2^T = [V_{21}, V_{22}, \cdots, V_{2M}]$ - вектор для второго тракта; $\rho_{\Delta} = |\rho(\Delta \tau)|$ - модуль нормированной корреляционной функции

помехового излучения при временном сдвиге равном удвоенной абсолютной величине максимальной погрешности реализации относительных запаздываний.

Поставим целью весового суммирования (BC) получение минимально возможного отношения $q_{\rm n}$ помеха-сигнал на выходе сумматора. В общем случае многоканального BC [98] отношение

$$q_{\Pi} = \frac{P_{\Pi\Sigma}}{P_{c\Sigma}} = \frac{\overrightarrow{W}^{+} R_{\Pi} \overrightarrow{W}}{\overrightarrow{W}^{+} R_{c} \overrightarrow{W}}, \tag{4.3}$$

где $P_{\Pi\Sigma}$, $P_{c\Sigma}$, — мощности сигнала и помехи на выходе сумматора; $\overrightarrow{W}^{\mathrm{T}} = [w_1, w_2, ..., w_{\mathrm{M}}]$ — вектор коэффициентов BC (КВС); $R_{\mathrm{C}} = \overline{X_{\mathrm{C}}^*(t)} \, \overline{X_{\mathrm{C}}^{\mathrm{T}}(t)}, R_{\mathrm{\Pi}} = \overline{X_{\mathrm{\Pi}}^*(t)} \, \overline{X_{\mathrm{\Pi}}^{\mathrm{T}}(t)}$ - корреляционные матрицы комплексных огибающих сигнала и помехи; * , $^+$, $^{\mathrm{T}}$ - знаки комплексного, эрмитового сопряжения и транспонирования матриц; $X_{\mathrm{C}}^{\mathrm{T}}(t) = [x_{\mathrm{C1}}(t), x_{\mathrm{C2}}(t), ..., x_{\mathrm{CM}}(t)], X_{\mathrm{\Pi}}^{\mathrm{T}}(t) = [x_{\mathrm{\Pi1}}(t), ..., x_{\mathrm{\PiM}}(t)]$ - векторы комплексных огибающих сигнала и помехи на входах весового сумматора (ВС); М - число входов ВС; $\overline{(\cdot)}$ - обозначение операции усреднения [98].

Минимизация величины $q_{\rm n}$, выражаемой через обобщенное отношение (4.3) Релея, приводит к обобщенной проблеме собственных чисел, которая формулируется в виде матричного уравнения [10, 37]:

$$R_{\Pi} \overrightarrow{W}_{oi} = \lambda_i R_c \overrightarrow{W}_{oi}, \tag{4.4}$$

где λ_i и \overrightarrow{W}_{oi} — i-ое собственное число и соответствующий ему собственный вектор. Минимальное собственное число $\lambda_1 = q_{\Pi.\text{мин}}$, где $q_{\Pi.\text{мин}}$ - минимальное значение q_{Π} при изменении вектора \overrightarrow{W} . Учитывая, что $\overrightarrow{W^T} = [\overrightarrow{W_1^T}, \overrightarrow{W_2^T}]$ где $\overrightarrow{W_1}$ - вектор КВС для ТФ1, а $\overrightarrow{W_2}$ - вектор КВС для ТФ2, подставляем (4.1),(4.2) в (4.3) (4.4) и получаем систему уравнений относительно собственного числа λ и соответствующих собственных векторов $\overrightarrow{W_{0_1}}$ и $\overrightarrow{W_{0_2}}$:

$$\begin{cases} P_{\pi o} \left(\overrightarrow{F_1^+} \overrightarrow{F_1} \overrightarrow{W}_{01} + \rho_{\Delta} \overrightarrow{F_2} \overrightarrow{F_1^+} \overrightarrow{W}_{02} \right) = \lambda P_{co} \overrightarrow{W}_{01} \\ P_{\pi o} \left(\rho_{\Delta} \overrightarrow{F_1} \overrightarrow{F_2^+} \overrightarrow{W}_{01} + \overrightarrow{F_2^+} \overrightarrow{F_2} \overrightarrow{W}_{02} \right) = \lambda P_{co} \overrightarrow{W}_{02}. \end{cases}$$
(4.5)

Учитывая, что $\vec{F}_1^+ \vec{W}_{02}$ и $\vec{F}_2^+ \vec{W}_{01}$ – некоторые числа, имеем

$$\vec{W}_{01} = \alpha \vec{F}_2, \vec{W}_{02} = \beta \vec{F}_1, \tag{4.6}$$

где α , β - комплексные числа. Подставляя (4.6) в (4.5), получаем систему скалярных уравнений

$$\begin{cases} P_{\text{по}} \hat{P}_{1}(\eta + \rho_{\Delta}) = \lambda \eta P_{\text{co}} \\ P_{\text{по}} \hat{P}_{2}(\eta \rho_{\Delta} + 1) = \lambda P_{\text{co}} \end{cases}$$
(4.7)

где $\hat{P}_1 = \vec{F}_1^+ \vec{F}_1;$ $\hat{P}_2 = \vec{F}_2^+ \vec{F}_2;$ $\eta = \alpha/\beta$. Решая систему уравнений (4.7) относительно минимального собственного числа λ_1 , получаем

$$\eta_1 = \left(2\rho_{\Delta}\hat{P}_2\right)^{-1} \left[\hat{P}_1 - \hat{P}_2 - \left(\hat{P}_1 + \hat{P}_2\right)\sqrt{1 - 4\hat{P}_1\hat{P}_2(1 - \rho_{\Delta}^2)/\left(\hat{P}_1 + \hat{P}_2\right)^2}\right], \tag{4.8}$$

$$\lambda_1 = P_{\pi\sigma}\hat{P}_2P_{\sigma\sigma}^{-1}(1 + \eta_1\rho_{\Delta})$$
 При малых погрешностях реализации относительных запаздываний

При малых погрешностях реализации относительных запаздываний $1 - \rho_{\Delta}^2 \ll 1$. Поэтому $4\hat{P}_1\hat{P}_2(1 - \rho_{\Delta}^2)/(\hat{P}_1 + \hat{P}_2)^2 \ll 1$, и

$$\eta_1 \approx \rho_{\Delta}^{-1} [\hat{P}_1 (\hat{P}_1 + \hat{P}_2)^{-1} (1 - \rho_{\Delta}^2) - 1],
\lambda_1 \approx (1 - \rho_{\Delta}^2) (P_{co} \hat{P}_1^{-1} + P_{co} \hat{P}_2^{-1})^{-1} P_{\Pi o}$$
(4.9)

Рассмотрим теперь случай "малых задержек", когда величины относительных запаздываний отражений намного меньше интервала корреляции сигнала. При этом сигнальные компоненты на выводах ЛЗ сильно коррелированы, и корреляционная матрица сигналов

$$R_c = P_{\rm co} \begin{bmatrix} \vec{S}_1 \vec{S}_1^+ & 0\\ 0 & \vec{S}_2 \vec{S}_2^+ \end{bmatrix}, \tag{4.10}$$

где $\overrightarrow{S_1}^*$ и $\overrightarrow{S_1}^*$ – сигнальные фазовые векторы на выводах Л31 и Л32, соответственно; $P_{\text{со}}$ – мощность полезного сигнала на входах Л31 и Л32. Найдем минимальное возможное значение отношения q_n и соответствующие векторы КВС для данного случая. Используя (4.3) и (4.4), с учетом (4.12) получаем

$$R_{\Pi} \overrightarrow{W}_0 = k \begin{bmatrix} \beta \vec{S}_1 \\ \vec{S}_2 \end{bmatrix}, \tag{4.11}$$

где k — произвольная постоянная, а

$$\beta_0 = \frac{\vec{S}_1^+ \, \vec{W}_{01}}{\vec{S}_2^+ \vec{W}_{02}}.\tag{4.12}$$

Так как отношение сигнал-помеха не меняется при умножении вектора \overrightarrow{W}_0 на любое число, то, в частности, k=1.

С целью определения граничных значений $q_{\text{мин}}$ рассмотрим случай, когда матрица $R_{\text{п}}$ описывается выражением (4.2). Тогда выражение (4.11) при k=1 принимает вид

$$\begin{cases}
q_0^{\text{BX}}(\widehat{P}_1 \overrightarrow{W}_{01} + \rho_\Delta \widetilde{\theta}_2 \vec{F}_2) = \beta_0 \vec{S}_1 \\
q_0^{\text{BX}}(\rho_\Delta \widetilde{\theta}_1 \vec{F}_1 + \widehat{P}_2 \overrightarrow{W}_{02}) = \vec{S}_2
\end{cases}$$
(4.13)

$$q_0^{\text{BX}} = P_{\text{no}}/P_{\text{co}}, \tilde{\theta}_1 = \overrightarrow{F_2^+} \overrightarrow{W_{01}}, \ \tilde{\theta}_2 = \overrightarrow{F_1^+} \overrightarrow{W_{02}}. \tag{4.14}$$

Решая систему уравнений (4.13) и учитывая (4.14) и (4.12), получаем

$$\vec{W}_{01} = \frac{1}{q_0^{\text{BX}} \hat{P}_1} \left[\beta_0 \vec{S}_1 + \frac{\beta_0 \rho_\Delta^2 \alpha_1 - \rho_\Delta \alpha_2}{\hat{P}_2 (1 - \rho_\Delta^2)} \vec{F}_2 \right]
\vec{W}_{02} = \frac{1}{q_0^{\text{BX}} \hat{P}_2} \left[\vec{S}_2 + \frac{\rho_\Delta^2 \alpha_2 - \beta_0 \rho_\Delta \alpha_1}{\hat{P}_1 (1 - \rho_\Delta^2)} \vec{F}_1 \right]$$
(4.15)

где $\alpha_1 = \vec{S}_1^+ \vec{F}_2$; $\alpha_2 = \vec{S}_2^+ \vec{F}_1$;

$$\beta_{0} = (2\rho_{\Delta}\alpha_{1}^{*}\alpha_{2})^{-1} \left[\hat{\mathbf{d}} - \sqrt{\hat{\mathbf{d}}^{2} + 4\rho_{\Delta}^{2}|\alpha_{1}\alpha_{2}|^{2}} \right];$$

$$\hat{\mathbf{d}} = \rho_{\Delta}^{2}(|\alpha_{2}|^{2} - |\alpha_{1}|^{2}) + (\hat{P}_{1}\hat{P}_{c2} - \hat{P}_{2}\hat{P}_{c1})(1 - \rho_{\Delta}^{2});$$

$$\hat{P}_{c1} = \vec{S}_{1}^{+}\vec{S}_{1}; \ \hat{P}_{c2} = \vec{S}_{2}^{+}\vec{S}_{2}.$$

$$(4.16)$$

Подставляя (4.15) в (4.4) и учитывая (4.16), получаем

$$q_{\text{MUH}} = \frac{q_0^{\text{BX}} \hat{P}_1 \hat{P}_2 (1 - \rho_{\Delta}^2)}{\hat{P}_1 \hat{P}_{c2} (1 - \rho_{\Delta}^2) + \rho_{\Delta}^2 |\alpha_2|^2 - 0.5 \left(\hat{\mathbf{d}} - \sqrt{\hat{\mathbf{d}}^2 + 4\rho_{\Delta}^2 |\alpha_1 \alpha_2|^2}\right)}. \tag{4.17}$$

Если величина $1-\rho_{\Delta}^2\ll 1$, то

$$\beta_0 = -\alpha_1/\alpha_2, \ q_{\text{MMH}} \approx \frac{q_0^{\text{BX}} \hat{P}_1 \hat{P}_2 (1 - \rho_{\Delta}^2)}{|\alpha_1|^2 + |\alpha_2|^2}$$
(4.18)

Из выражения (4.18) следует, что в данном случае степень подавления помех определяется величиной $1-\rho_{\Delta}^2$, как и в случае некоррелированных сигнальных компонентов на выводах ЛЗ. Видно, что чем более "параллельны" векторы \vec{S}_1 и \vec{F}_2 , а также \vec{S}_2 и \vec{F}_1 , тем больше величины $|\alpha_1|$ и $|\alpha_2|$, тем большая мощность сигнала обеспечивается на выходе КП.

Полученные в результате анализа выражения (4.15) и (4.8) соответствуют случаю, когда интервал корреляции τ_{Π} удовлетворяет условию $\tau_{\Pi} \leq \Delta \tau$, где $\Delta \tau$ - удвоенная максимальная погрешность реализации требуемых задержек. Если условие не выполняется, то имеет место сужение спектра помехи. Чем уже спектр помехи, тем больше степень коррелированности помеховых компонентов на отводах ЛЗ, относящихся к разным входам КП. Следовательно, растет степень подавления помех. Поэтому полученные выражения относятся к верхним границам отношения $q_{\text{мин}}$ помеха-сигнал на выходе КП.

При прохождении сигнала через ТФ он искажается, распадаясь на ряд задержанных во времени копий разного уровня и с различным фазовым сдвигом. Совокупность радиоканала и ТФ можно рассматривать как некоторый фильтр с импульсной характеристикой равной свертке импульсных характеристик ΤФ. тропосферного канала Поэтому результирующая импульсная характеристика имеет длительность т_с равную сумме длительностей импульсных характеристик канала и ТФ: $\tau_{\scriptscriptstyle T}$, $\tau_{\scriptscriptstyle T\varphi}$. При суммировании сигнальных компонентов, прошедших через ТФ1 и ТФ2, импульсная характеристика КП совместно с каналом имеет максимальную длительность $\tau_{\text{ким}}$ равную наибольшей из длительностей τ_{c1} и τ_{c2} . Следовательно, $\tau_{\text{кпм}} = \tau_{\text{т}} + \tau_{\text{лз}}$, где $\tau_{\text{лз}}$ - максимальная задержка ЛЗ, входящих в состав ТФ. Из-за дополнительного "растягивания" сигнала во времени может возникнуть явление межсимвольной интерференции

[68], приводящее к ошибочному приему. Для исключения этого явления необходимо соблюдать условие: $T > \tau_{K\Pi M} + \tau_{M}$ где длительность τ_{u} информационного импульса на выходе согласованного фильтра приемника, Т – период следования импульсов, либо принимать меры устранению ПО межсимвольной интерференции [36].

4.2. Адаптивная компенсация помех в двухканальной системе приема

Рассмотрим возможность подавления многолучевых помех с помощью процессора Хоуэлса-Эпплбаума [98] на примере системы приема цифровых широкополосных сигналов [8] с двукратным пространственным разнесением антенн. В процессе анализа считаем мощность помехи намного большей мощности полезного сигнала, что позволяет пренебречь учетом его мешающего влияния на настройку коэффициентов весового суммирования. Структурная схема соответствующего адаптивного компенсатора помех представлена на Рисунке 4.2.

Напряжения с антенн $U_1(t)$ и $U_2(t)$ поступают на ЛЗІ и ЛЗ2 – линии задержки. С выводов ЛЗ1 и ЛЗ2 помеховые и сигнальные компоненты проходят через формирователи квадратурных составляющих ФКС1 и ФКС2, каждый из которых состоит из М фазовращателей на 90°. Затем напряжения поступают на формирователи весовых коэффициентов суммирования ФВК, выходы которых подключены к многовходовому сумматору С. Выход сумматора через усилитель У подключен ко всем ФВК.

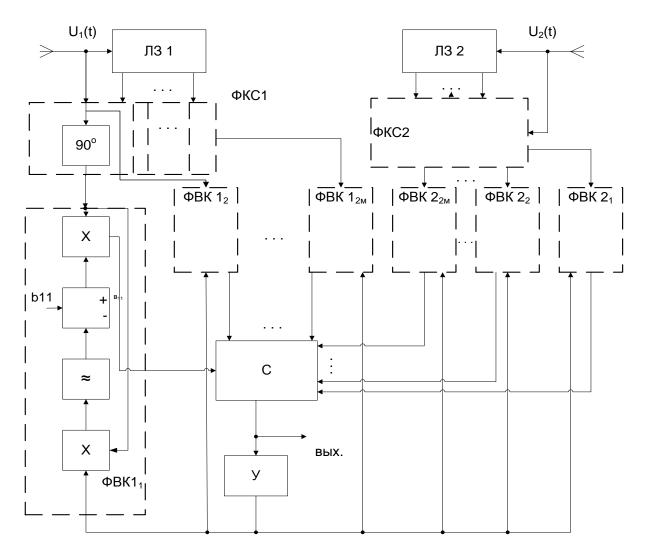


Рисунок 4.2. Структурная схема адаптивного компенсатора помех

Каждый ФВК состоит из двух перемножителей, вычитающего устройства, на суммирующий вход которого поступает некоторое постоянное напряжение b_{ij} и фильтра, обеспечивающего усреднение результата перемножения. Установившееся значение вектора КВС [98]

$$\overrightarrow{W}_{\infty} = (E + \mu R_{\Pi})^{-1} \overrightarrow{B}, \tag{4.19}$$

где E - единичная матрица размерности $2M\times 2M$; $R_{\Pi}=\overrightarrow{X_{\Pi}^*(t)}\overrightarrow{X_{\Pi}^T(t)}$ - корреляционная матрица помех; $\overrightarrow{X_{\Pi}(t)}$; \overrightarrow{W} ; \overrightarrow{B} – векторы помеховых компонентов на выходах ЛЗ, коэффициентов весового суммирования и управления

размерности 2M соответственно; μ - коэффициент усиления в петле обратной связи. Здесь под векторами понимаются суммы векторов соответствующих квадратурных величин.

Определим оптимальное значение вектора управления \vec{B} , обеспечивающее наиболее эффективное подавление помех. Рассмотрим сначала случай "больших задержек", когда корреляционная матрица R_{Π} помех и сигнальная матрица $R_{\rm c}$ описываются выражениями (4.2) и (4.1). Подставляя (4.2) в формулу (4.19), получаем

$$\begin{cases}
\widehat{\mu} \left(\widehat{P}_{1} \overrightarrow{W}_{1} + \rho_{\Delta} \vec{F}_{2} \vec{F}_{1}^{+} \overrightarrow{W}_{2} \right) + \overrightarrow{W}_{1} = \overrightarrow{B}_{1} \\
\widehat{\mu} \left(\rho_{\Delta} \vec{F}_{1} \vec{F}_{2}^{+} \overrightarrow{W}_{1} + \widehat{P}_{2} \overrightarrow{W}_{2} \right) + \overrightarrow{W}_{2} = \overrightarrow{B}_{2}
\end{cases} (4.20)$$

где $\hat{\mu} = \mu P_{\text{по}}$; \vec{B}_1 , \vec{B}_2 , \vec{W}_1 , \vec{W}_2 – векторы управляющих и весовых коэффициентов для первого и второго трансверсальных фильтров. Учитывая, что

$$\vec{F}_1^+ \vec{W}_2 = \theta_2, \ \vec{F}_2^+ \vec{W}_1 = \theta_1 -$$
 (4.21)

некоторые числа, получаем

$$\begin{cases}
\vec{W}_{1} = (1 + \hat{\mu}\hat{P}_{1})^{-1}(\vec{B}_{1} - \hat{\mu}\rho_{\Delta}\theta_{2}\vec{F}_{2}) \\
\vec{W}_{2} = (1 + \hat{\mu}\hat{P}_{2})^{-1}(\vec{B}_{2} - \hat{\mu}\rho_{\Delta}\theta_{1}\vec{F}_{1})
\end{cases} (4.22)$$

где $\hat{P}_1 = \vec{F}_1^+ \vec{F}_1$ и $\hat{P}_2 = \vec{F}_2^+ \vec{F}_2$ — нормированные по $P_{\text{по}}$ мощности помех в первом и втором каналах, соответственно. Подставляя выражения (4.22) в формулы (4.21), получаем

$$\theta_{1} = \frac{\gamma_{1} + \widehat{\mu} \widehat{P}_{2} (\gamma_{1} - \rho_{\Delta} \gamma_{2})}{1 + \widehat{\mu} (\widehat{P}_{1} + \widehat{P}_{2}) + \widehat{\mu}^{2} \widehat{P}_{1} \widehat{P}_{2} (1 - \rho_{\Delta}^{2})};$$

$$\theta_{2} = \frac{\gamma_{2} + \widehat{\mu} \widehat{P}_{1} (\gamma_{2} - \rho_{\Delta} \gamma_{1})}{1 + \widehat{\mu} (\widehat{P}_{1} + \widehat{P}_{2}) + \widehat{\mu}^{2} \widehat{P}_{1} \widehat{P}_{2} (1 - \rho_{\Delta}^{2})};$$
(4.23)

где $\gamma_2 = \vec{F}_1^+ \vec{B}_2$, $\gamma_1 = \vec{F}_2^+ \vec{B}_1$. Из выражения (4.6) следует, что для максимального приближения векторов \vec{W}_1 , \vec{W}_2 к оптимальным значениям \vec{W}_{01} и \vec{W}_{02} целесообразно установить $\vec{B}_1 = k_1 \vec{F}_2$; $\vec{B}_2 = k_2 \vec{F}_1$. Здесь k_1 , k_2 — некоторые

коэффициенты. В этом случае $\gamma_1=\vec{F}_2^+\vec{B}_1=k_1\vec{F}_2^+\vec{F}_2=k_1\hat{P}_2$, $\gamma_2=\vec{F}_1^+\vec{B}_2=k_2\vec{F}_1^+\vec{F}_1=k_2\hat{P}_1$. Подставляя полученные выражения в формулы (4.23), получаем

$$\theta_{1} = \frac{k_{1}\hat{P}_{2} + \hat{\mu}\hat{P}_{2}(k_{1}\hat{P}_{2} - k_{2}\rho_{\Delta}\hat{P}_{1})}{1 + \hat{\mu}(\hat{P}_{1} + \hat{P}_{2}) + \hat{\mu}^{2}\hat{P}_{1}\hat{P}_{2}(1 - \rho_{\Delta}^{2})};$$

$$\theta_{2} = \frac{k_{2}\hat{P}_{1} + \hat{\mu}\hat{P}_{1}(k_{2}\hat{P}_{1} - k_{1}\rho_{\Delta}\hat{P}_{2})}{1 + \hat{\mu}(\hat{P}_{1} + \hat{P}_{2}) + \hat{\mu}^{2}\hat{P}_{1}\hat{P}_{2}(1 - \rho_{\Delta}^{2})}.$$
(4.24)

Выражения (4.22) приобретают вид

$$\begin{cases}
\overrightarrow{W}_{1} = (1 + \hat{\mu}\hat{P}_{1})^{-1} (k_{1} - \hat{\mu}\rho_{\Delta} \frac{k_{2}\hat{P}_{1} + \hat{\mu}\hat{P}_{1}(k_{2}\hat{P}_{1} - k_{1}\rho_{\Delta}\hat{P}_{2})}{1 + \hat{\mu}(\hat{P}_{1} + \hat{P}_{2}) + \hat{\mu}^{2}\hat{P}_{1}\hat{P}_{2}(1 - \rho_{\Delta}^{2})})\overrightarrow{F}_{2} \\
\overrightarrow{W}_{2} = (1 + \hat{\mu}\hat{P}_{2})^{-1} (k_{2} - \hat{\mu}\rho_{\Delta} \frac{k_{1}\hat{P}_{2} + \hat{\mu}\hat{P}_{2}(k_{1}\hat{P}_{2} - k_{2}\rho_{\Delta}\hat{P}_{1})}{1 + \hat{\mu}(\hat{P}_{1} + \hat{P}_{2}) + \hat{\mu}^{2}\hat{P}_{1}\hat{P}_{2}(1 - \rho_{\Delta}^{2})})\overrightarrow{F}_{1},
\end{cases} (4.25)$$

Рассмотрим отношение

$$\beta_{2} = \frac{\left(1 + \hat{\mu}\hat{P}_{2}\right)^{-1} \left(k_{2} - \hat{\mu}\rho_{\Delta}\frac{k_{1}\hat{P}_{2} + \hat{\mu}\hat{P}_{2}\left(k_{1}\hat{P}_{2} - k_{2}\rho_{\Delta}\hat{P}_{1}\right)}{1 + \hat{\mu}(\hat{P}_{1} + \hat{P}_{2}) + \hat{\mu}^{2}\hat{P}_{1}\hat{P}_{2}\left(1 - \rho_{\Delta}^{2}\right)}\right)}{\left(1 + \hat{\mu}\hat{P}_{1}\right)^{-1} \left(k_{1} - \hat{\mu}\rho_{\Delta}\frac{k_{2}\hat{P}_{1} + \hat{\mu}\hat{P}_{1}\left(k_{2}\hat{P}_{1} - k_{1}\rho_{\Delta}\hat{P}_{2}\right)}{1 + \hat{\mu}(\hat{P}_{1} + \hat{P}_{2}) + \hat{\mu}^{2}\hat{P}_{1}\hat{P}_{2}\left(1 - \rho_{\Delta}^{2}\right)}\right)} = \frac{\left(1 + \hat{\mu}\hat{P}_{1}\right)\left(\alpha_{2} - \hat{\mu}\rho_{\Delta}\frac{\hat{P}_{2} + \hat{\mu}\hat{P}_{2}\left(\hat{P}_{2} - \alpha_{2}\rho_{\Delta}\hat{P}_{1}\right)}{1 + \hat{\mu}(\hat{P}_{1} + \hat{P}_{2}) + \hat{\mu}^{2}\hat{P}_{1}\hat{P}_{2}\left(1 - \rho_{\Delta}^{2}\right)}\right)}{\left(1 + \hat{\mu}\hat{P}_{2}\right)\left(1 - \hat{\mu}\rho_{\Delta}\frac{\alpha_{2}\hat{P}_{1} + \hat{\mu}\hat{P}_{1}\left(\alpha_{2}\hat{P}_{1} - \rho_{\Delta}\hat{P}_{2}\right)}{1 + \hat{\mu}(\hat{P}_{1} + \hat{P}_{2}) + \hat{\mu}^{2}\hat{P}_{1}\hat{P}_{2}\left(1 - \rho_{\Delta}^{2}\right)}\right)}$$

$$(4.26)$$

Видно, что β_2 (а следовательно, и оптимальное значение вектора управляющих коэффициентов) определяется не конкретными значениями k_2, k_1 , а лишь их отношением $\alpha_2 = k_2/k_1$. С целью определения данного отношения рассмотрим сначала частный случай, когда отражения отсутствуют. Для данной ситуации имеем $F_1 = 1$, $F_2 = k$, где k — некоторая постоянная величина. Корреляционная матрица помех принимает вид

$$R_{\Pi} = P_{\Pi 0} \begin{bmatrix} E & \rho_{\Delta} k \\ \rho_{\Delta} k & k^2 E \end{bmatrix}, \tag{4.27}$$

Система уравнений (4.20) принимает вид

$$\begin{cases}
\widehat{\mu} (w_1 + \rho_{\Delta} k w_2) + w_1 = b_1 \\
\widehat{\mu} (\rho_{\Delta} k w_1 + k^2 w_2) + w_2 = b_2
\end{cases}$$
(4.28)

Здесь все векторные величины становятся скалярными. Решая систему уравнений, получаем

$$w_{1} = \frac{b_{1} + \widehat{\mu} k(kb_{1} - \rho_{\Delta}b_{2})}{(1 + \widehat{\mu})(1 + \widehat{\mu} k^{2}) - (\rho_{\Delta}\widehat{\mu} k)^{2}};$$

$$w_{2} = \frac{b_{2} + \widehat{\mu} (b_{2} - \rho_{\Delta}b_{1}k)}{(1 + \widehat{\mu})(1 + \widehat{\mu} k^{2}) - (\rho_{\Delta}\widehat{\mu} k)^{2}}.$$
(4.29)

Рассмотрим отношение

$$\beta = \frac{w_2}{w_1} = \frac{b_2 + \widehat{\mu} (b_2 - \rho_\Delta b_1 k)}{b_1 + \widehat{\mu} k (k b_1 - \rho_\Delta b_2)} \approx \frac{\alpha - \rho_\Delta k}{k^2 - \alpha \rho_\Delta k}$$
(4.30)

Здесь принято, что $\propto = \frac{b_2}{b_1}$, и величина $\widehat{\mu}$ достаточно велика. Рассмотрим отношение помеха-сигнал на выходе компенсатора

$$q_{\text{вых}} = q_0 \frac{1 + (\beta k)^2 + 2\rho_{\Delta}\beta k}{1 + \beta^2}.$$
 (4.31)

Здесь $q_0 = P_{\pi 0}/P_{c0}$ и учтено, что сигнальные компоненты на входах компенсатора не коррелированы. Подставляя (4.30) в выражение (4.31) и дифференцируя полученное выражение по \propto , получаем оптимальные значения $\propto_{01}, \propto_{02}$, обеспечивающие соответственно максимум и минимум $q_{\text{вых}}$.

$$\alpha_{01(02)} = \frac{k^2 - 1 \mp (k^2 + 1) \sqrt{1 - 4 \frac{1 - \rho_{\Delta}^2}{(k^{-1} + k)^2}}}{2k\rho_{\Delta}}.$$
(4.32)

Считая $1 - \rho_{\Delta}^2 \ll 1$, получаем

$$\alpha_{01} \approx \frac{k}{\rho_{\Delta}} \cdot \left(1 - \frac{1 - \rho_{\Delta}^2}{k^2 + 1}\right); \quad \alpha_{02} \approx \frac{-1}{k\rho_{\Delta}} \cdot \left(1 - \frac{1 - \rho_{\Delta}^2}{k^{-2} + 1}\right). \tag{4.33}$$

На основе проведенных рассуждений считаем, что для общего случая

наличия отражений при $1-\rho_{\Delta}^2\ll 1$ величина $\alpha_2\approx -1/\rho_{\Delta}$.

4.3. Моделирование адаптивного компенсатора помех

С целью проверки полученных выражений промоделируем работу адаптивного компенсатора помех и их отражений. Рассмотрим сначала простейший вариант помехи, когда отражения отсутствуют. В этом случае вместо многоотводных линий задержки (ЛЗ) используются ЛЗ без отводов. Считаем, что в каналах приема установлены ЛЗ, обеспечивающие совмещение помех на входах компенсатора во времени с некоторой малой погрешностью. Считаем, что сигнальные компоненты на входах компенсатора не коррелированы и настолько малы по сравнению с помехой, что не оказывают влияние на процесс адаптации.

Основная цель моделирования — поиск оптимального вектора \vec{B} . Для уяснения «физического» смысла полученных выше выражений считаем, что помехи с выходов линий задержки синфазны. Параметры моделирования данной структуры АКП следующие. Помехой является последовательность гауссовых радиоимпульсов с центральной частотой спектра 70 МГц, шириной спектра 2 МГц (на уровне 6дБ). Длительность импульса 0,4мкс (на уровне 6дБ), период следования 4мкс, длительность интервала анализа 100мкс. Максимальная амплитуда радиоимпульса на выходе источника помехи равна единице. На Рисунке.4.3 представлена «осциллограмма» импульса

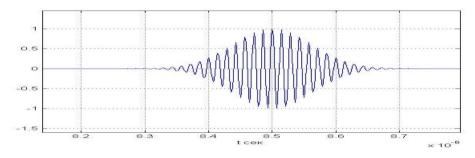


Рисунок 4.3. «Осциллограмма» импульса помехи

Коэффициенты передачи помехи от источника до выходов трактов

промежуточной частоты приемников — входов АКП принимались равными v_1 =2, v_2 =4. Следовательно, k= v_2 / v_1 =2. Задержки в трактах распространения помехи от источника до первого и второго входов компенсатора, соответственно $\tau 1$ =5мкс, $\tau 2$ = $\tau 1$ - $\Delta \tau$, где $\Delta \tau$ =4/70 мкс — погрешность совмещения импульсов помехи во времени, обусловленная дискретностью реализации требуемой задержки с помощью линий задержки. Для определения коэффициента весового суммирования использовалось выражение (4.19) при μ = 10^6 . Корреляционная матрица помех в этом случае

$$R_n = \begin{bmatrix} 0.158 & 0.3088 \\ 0.3088 & 0.6321 \end{bmatrix} = 0.158 \cdot \begin{bmatrix} 1 & 1.9539 \\ 1.9539 & 4 \end{bmatrix}$$

Из матрицы следует, что коэффициент взаимной корреляции $\rho_{\Delta}=0,9769.$ Используя выражения (4.30), (4.31) получаем график зависимости $q_{\text{вых}}(\alpha)$ (Рисунок 4.4).

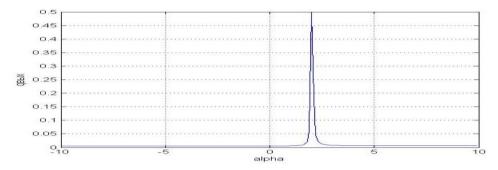


Рисунок 4.4. Зависимость отношения помеха-сигнал на выходе компенсатора от отношения α коэффициентов управления при $\rho_{\Delta}=0,9769$

Из графика следует, что при изменении α от -10 до 1 величина $q_{\rm вых.}$ достаточно мала и меняется от 5,92*10⁻³ до 6,18*10⁻³ . При изменении α от 3 до 10 величина $q_{\rm вых.}$ также принимает малые значения и меняется от 8,07*10⁻³ до 6,11*10⁻³. Отношение помеха-сигнал принимает максимальное значение $q_{\rm вых.макс}(\alpha=2,028)=0,784$. Отношение помеха-сигнал принимает минимальное значение $q_{\rm вых.макс}(\alpha=-0,5)=5,8*10^{-3}$.

Сопоставим результаты моделирования с результатами анализа. Из выражений (4.33) следует $\alpha_{01}=2,028;$ $\alpha_{02}=-0,493.$ То есть расхождение

результатов расчета и моделирования для α_{01} равно нулю, а для α_{02} составляет 0,014. Данное расхождение обусловлено дискретностью задания данных для моделирования. Если значение $\alpha = -0.5$ подставить в выражения (4.30) и (4.31), полученный результат полностью совпадает co значением $q_{\scriptscriptstyle \mathrm{BHX.MИH}}$. Аналогичная ситуация наблюдается Следовательно, И ДЛЯ $q_{\scriptscriptstyle \mathrm{BЫX.MAKC}}.$ полученных выражений справедливость при анализе моделированием подтверждается.

Из проведенного анализа следует, что значение α можно менять в значительных пределах при этом отношение помеха-сигнал изменяется мало и является небольшой величиной. Однако величина α не должна находиться в окрестности α_{01} , которое в терминологии адаптивных антенных решеток соответствует случаю совпадения направлений прихода сигнала и помехи. Наиболее целесообразно величину α устанавливать равной α_{02} , которая определяется корреляционной матрицей помех.

Если величина $\rho_{\Delta} \ll 1$, то компенсация нецелесообразна, что подтверждается графиком зависимости $q_{\text{вых}}(\alpha)$ при $\rho_{\Delta}=0.56$, изображенным на Рисунке 4.5.

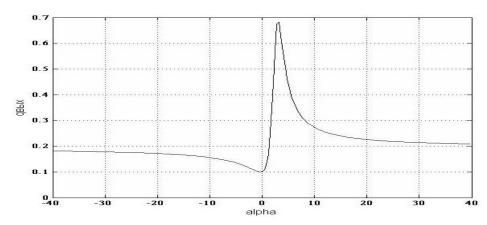


Рисунок 4.5. Зависимость отношения помеха-сигнал на выходе компенсатора от отношения α коэффициентов управления при $\rho_{\Delta}=0.56$

Из графика следует, что $q_{\text{вых. мин}} = 0.1$. При отсутствии компенсации отношение помеха-сигнал в первом канале приема равно 0.158 (считая, что

мощность сигнала равна единице). То есть компенсация в данном случае уменьшает отношение помеха-сигнал всего лишь в 1,58 раза.

Рассмотрим теперь результаты моделирования при наличии прямого прохождения помехи на вход приемника и наличия одного ее отражения. В качестве помехи выступает периодическая последовательность гауссовых радиоимпульсов с шириной спектра на уровне -6дБ равной 6МГц и периодом Т=8мкс. На каждый из двух входов приемника поступают по две копии помехи. Параметры задержки и коэффициентов передачи указаны в Таблице 4.1

Таблица 4.1. Параметры моделирования помехи

Выходы линейных трактов	τ_1	τ_2	$\mathbf{v_1}$	\mathbf{v}_2
приемника	(мкс)	(мкс)		
Выход 1	0,5	1,5	1	2
Выход 2	1	3	3	4

Здесь τ_1 , τ_2 - задержки копий помехи; v_1 , v_2 - коэффициенты передачи от источника помехи до выхода линейного тракта. Задержки τ_1 , τ_2 выбирались так, чтобы реализовать случай «больших» задержек. Необходимо исключить нежелательную взаимную коррелированность копий помехи на выходах линий задержки между первым и вторым адаптивными фильтрами. Взаимная коррелированность должна присутствовать только для тех копий помехи первого и второго фильтров, которые должны быть совмещены во времени с помощью линий задержки. На Рисунке 4.6 показана «осциллограмма» одного периода последовательности импульсов помехи.

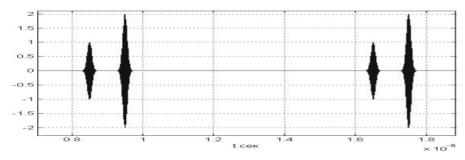


Рисунок 4.6. «Осциллограмма» периода последовательности импульсов помехи с одним отражением

В данном случае корреляционная матрица помех на отводах линий задержки в компенсаторе помех имеет следующий вид

$$\widehat{R}_{n} = P_{n0} \begin{bmatrix} (v_{1}^{2} + v_{2}^{2}) & 0 & v_{1}v_{3}\rho & v_{2}v_{3}\rho \\ 0 & (v_{1}^{2} + v_{2}^{2}) & v_{1}v_{4}\rho & v_{2}v_{4}\rho \\ v_{1}v_{3}\rho & v_{1}v_{4}\rho & (v_{3}^{2} + v_{4}^{2}) & 0 \\ v_{2}v_{3}\rho & v_{2}v_{4}\rho & 0 & (v_{3}^{2} + v_{4}^{2}) \end{bmatrix}$$
(4.34)

Здесь коэффициенты v_3 , v_4 соответствуют коэффициентам передачи переотражений для выхода второго линейного тракта приемника. В данном случае вектор $\mathbf{F_1}=[v_1\ v_2\]=[1\ 2]$, а вектор $\mathbf{F_2}=[v_3\ v_4\]=[3\ 4]$.

При проведении первого эксперимента определялась зависимость отношения сигнал шум от отношения α_2 . Вектор управляющих коэффициентов $B=10^4*[\alpha_2(3\ 4)\ 1\ 2]$. Погрешность реализации требуемых задержек в одном канале приема $\Delta \tau=1/70$ мкс. График зависимости отношения сигнал-помеха от отношения α_2 представлены на Рисунках 4.7, 4.8. Видно, что для $\alpha_2=0,2$ отношение сигнал-помеха принимает минимальное значение, что соответствует максимальному сближению «направлений» прихода сигнала и помехи. На Рисунке 4.8 видно, что имеется слабо выраженный максимум отношения сигнал-помеха.

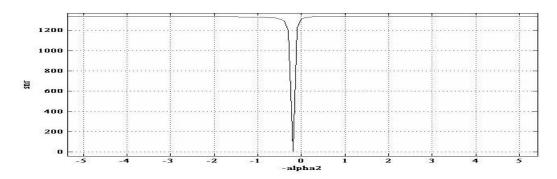


Рисунок 4.7. Зависимость отношения сигнал-помеха на выходе компенсатора от отношения α коэффициентов управления при наличии одного отражения помехи

Затем с целью выработки алгоритма поиска оптимального вектора **В** управления определялись результаты компенсации при реализации требуемых задержек в адаптивных фильтрах компенсатора без погрешностей и с

погрешностью $\Delta \tau = 1/70$ мкс. Сначала использовался вектор управляющих коэффициентов ${\bf B} = 10^4 * [~3~4~-1~-2]$, что соответствует значению $\alpha_2 = -1$.

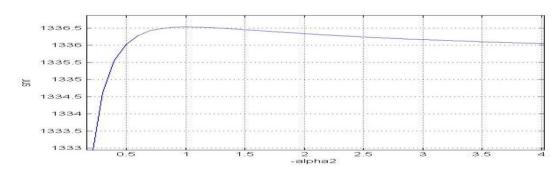


Рисунок 4.8. Зона максимума зависимости отношения сигнал-помеха на выходе компенсатора от отношения α коэффициентов управления при наличии одного отражения помехи

Если погрешность $\Delta \tau = 0$, то оценка корреляционной матрицы помех на выходах линий задержки имеет следующий вид

$$\widehat{\mathbf{R}}_{n1} = \begin{bmatrix} 0.0345 & 0 & 0.0207 & 0.0414 \\ 0 & 0.0345 & 0.0276 & 0.0553 \\ 0.0207 & 0.0276 & 0.1727 & 0 \\ 0.0414 & 0.0553 & 0 & 0.1727 \end{bmatrix}$$

Если $\Delta \tau = 1/70$ мкс, то в этом случае корреляционная матрица помех

$$\widehat{\mathbf{R}}_{n2} = \begin{bmatrix} 0.0345 & 0 & 0.0205 & 0.0409 \\ 0 & 0.0345 & 0.0273 & 0.0545 \\ 0.0205 & 0.0273 & 0.1727 & 0 \\ 0.0409 & 0.0545 & 0 & 0.1727 \end{bmatrix}$$

Так как матрица \widehat{R}_{n1} получена при нулевой погрешности реализации требуемых задержек, то коэффициент корреляции в выражении (4.35) равен единице. Для получения информации о коэффициенте корреляции ρ_{Δ} при $\Delta \tau$ =(1/70) мкс разделим поэлементно матрицу \widehat{R}_{n2} на матрицу \widehat{R}_{n1} .

$$\widehat{\mathbf{R}}_{n2}.\,/\,\widehat{\mathbf{R}}_{n1} = \begin{bmatrix} 1 & 1 & 0.987 & 0.987 \\ 1 & 1 & 0.987 & 0.987 \\ 0.987 & 0.987 & 1 & 1 \\ 0.987 & 0.987 & 1 & 1 \end{bmatrix}$$

Отсюда следует, что коэффициент взаимной корреляции между копиями помехи на выводах линий задержки ρ_{Δ} =0.987. Величина $P_{n0}=\frac{0.0345}{1^2+2^2}=0,0069$. Отношение сигнал-помеха на выходе компенсатора SNR=1336,5. Вектор коэффициентов весового суммирования W=[40.1;53.5;-13.25;-26.5]. Считая $P_{co}=1;\ P_{no}=0,0069;\ \hat{P}_1=5;\ \hat{P}_2=25$ и используя выражения (4.9), получаем $\eta_1\approx 0.987^{-1}[5(5+25)^{-1}(1-0.987^2)-1]\approx -1,$ $\lambda_1\approx (1-0.987^2)(1*5^{-1}+1*25^{-1})^{-1}*0.0069=7.426*10^{-4}$

Следовательно, расчетное отношение сигнал-помеха SNR_t= $1/\lambda_1=1346.5\approx$ SNR = 1336.5 , что свидетельствует о подтверждении результатов теоретического анализа моделированием.

Если проанализировать фрагмент корреляционной матрицы \hat{R}_{n1} с первой по вторую строку и с третьего по четвертый столбец, то видно, что строки этого фрагмента отражают соотношение между уровнями копий помехи, приходящими на первую антенну, а столбцы соответствуют копиям помехи на второй антенне. Поэтому можно сформировать вектор управляющих коэффициентов, пользуясь следующим правилом: меняем знак строки указанного фрагмента матрицы на противоположный, и полученный набор коэффициентов используем для управления адаптивного фильтра, подключенного к выходу линейного тракта второго приемника, а столбец фрагмента (без инвертирования) используем для управления адаптивного фильтра, подключенного к выходу линейного тракта первого приемника.

В Таблице 4.2 приведены результаты моделирования и расчета согласно выражениям (4.8) для различных значений погрешности $\Delta \tau$ реализации задержки

Таблица 4.2. Результаты моделирования и расчета

$\Delta \tau$ (MKC)	$\alpha_{2_{3\kappa cn}}$	$\alpha_{2_pac4} = -\rho_{\Delta}^{-1}$	SNR _{эксп}	SNR _{pac4}
1/70	-1,01	-1,013	1336 (31дБ)	1342
2/70	-1,04	-1,054	344 (25дБ)	344
4/70	-1,16	-1,23	96 (19,8дБ)	96

Здесь $\alpha_{2__{9KCR}}$ - это значение α_2 , обеспечивающее максимум отношения SNR $_{9KCR}$ сигнал-помеха при моделировании. Максимум отношения SNR $_{pac4}$ получен согласно выражениям (4.8). Отношения SNR для α_2 расу практически не отличаются от SNR $_{9KCR}$.

Если уменьшать ширину спектра помехи, то степень подавления помехи увеличивается. В частности, при сужении ширины спектра помехи до 0.5МГц отношение сигнал-помеха SNR=156980 (52дБ). При значениях вектора управляющих коэффициентов B1=[7787, 11163,-10496, -7787] и B2=[30000,40000, -10000,-20000] уровень подавления помехи не меняется, несмотря на значительные изменения вектора. Должно выполняться лишь главное условие: вектор \boldsymbol{B} не должен быть параллельным вектору коэффициентов передачи \boldsymbol{V} . В противном случае отношение сигнал-помеха снижается до значения SNR= 46142 (46.6дБ). В случае расширения полосы помехи до 6 МГц отношение сигнал-помеха снижается до значения SNR= 9.26 (9.6дБ).

Рассмотрим теперь результаты моделирования для случая «малых» задержек, когда сигнальные компоненты на отводах линии задержки сильно коррелированы. Попрежнему считаем, что сигнальные компоненты на выходах линейных трактов приемников не коррелированы. Общий порядок моделирования следующий: для ранее рассмотренного случая компенсации помехи с одним отражением определяются коэффициенты весового суммирования в виде вектора W. Далее на входы весового сумматора подается двухпозиционный фазоманипулированный сигнал. Сигнал модулирован псевдослучайной последовательностью бинарных символов. Для устранения коррелированности сигнальных компонентов на входах сумматора сигнал на одном входе задержан относительно сигнала на другом входе на величину периода T_d следования информационных символов.

Величина T_d =15мкс - выбрана достаточно большой по сравнению с расстоянием между отводами линии задержки, чтобы обеспечить коррелированность сигнальных компонентов на них. Частота f0 несущей и частота Fs дискретизации выбраны такими же, что и при моделировании помехи. Количество N информационных посылок равно

100.

Учитывая, что вектор $F1=[1\ 2]$, а вектор $F2=[3\ 4]$, определим согласно формуле (4.18) величину

$$\beta_0 = -\alpha_1/\alpha_2 = -\frac{\vec{S}_1^+ \vec{F}_2}{\vec{S}_2^+ \vec{F}_1} = -7/3;$$

Отсюда следует, что в данном случае оптимальный вектор управления B=[-7/3 -7/3 1 1]. Было проведено 14 экспериментов с различными псевдослучайными последовательностями модулирующих символов, которые порождали различные оценки корреляционной матрицы сигналов на выводах линий задержек. Наибольшее значение сигнал-помеха составило 1231 (31дБ). Соответствующая оценка корреляционной матрицы сигналов

$$\widehat{\mathbf{R}}_s = \begin{bmatrix} 0.5 & 0.4258 & -0.0373 & -0.0527 \\ 0.4258 & 0.5 & 0.037 & 0.0005 \\ -0.0373 & 0.037 & 0.5 & 0.4636 \\ -0.0527 & 0.0005 & 0.4636 & 0.5 \end{bmatrix}$$

Видно, что сигнальные компоненты на отводах разных линиях задержки слабо коррелированы, а на отводах одной линии задержки коррелированы сильно. Учитывая, что

$$lpha_1=7;\ lpha_2=3;\
ho_\Delta=0.987;
m P_{s0}=0.5;\ \widehat{R}_n(1,1)=P_{n0}(v_{11}^2+v_{12}^2)=$$
 $P_{n0}(1+4)=0.0345;\ P_{n0}=\frac{0.0345}{5}=0,0069;\ \widehat{P}_1=(v_{11}^2+v_{12}^2)=1+4=5; \widehat{P}_2=(v_{21}^2+v_{22}^2)=9+16=25,\ \ и\ \ \$ используя выражение (4.20), получаем $eta_0=-lpha_1/lpha_2=-7/3.$

$$q_{\text{\tiny MUH}} \approx \frac{q_0^{\text{\tiny BX}} \widehat{P}_1 \widehat{P}_2 (1 - \rho_{\Delta}^2)}{|\alpha_1|^2 + |\alpha_2|^2} = \frac{(0.0069/0.5) * 5 * 25 * (1 - 0.987^2)}{49 + 9} = 7.68 * 10^{-4}$$

Отношение сигнал-помеха SNR= $1/q_{\text{мин}}=1301$. Расхождение результатов моделирования и теоретического анализа незначительно и составляет 5,7%, что подтверждает справедливость полученных выше соотношений. Расхождение можно объяснить тем, что вид оценки \hat{R}_s несколько отличается от структуры матрицы (4.10), которая была использована при анализе.

Были проведены эксперименты на компьютере, которые ставили целью определить зависимость отношения сигнал-помеха от вида вектора управляющих коэффициентов. Результаты представлены в Таблице 4.3.

Таблица 4.3. Зависимость отношения сигнал-помеха от вектора управляющих коэффициентов (результаты моделирования)

N	Вектор	Вектор	$SNR = \frac{W'W}{W}$	$SNR - \frac{W'R_sW}{}$
ЭКСП.	V	В	$\frac{3NR}{W'R_nW}$	$SNR = \frac{W'R_nW}{W'R_nW}$
1	[1234]	[3 4 -1 -2]	1336	1231
2	[1234]	[-7/3 -7/3 1 1]	1336	1231
3	[1-234]	[3 4 -1 2]	1336	1013
4	[1-234]	[7 7 1 1]	1335	1014

В таблице представлены значения отношений сигнал-помеха при различных значениях вектора **V** коэффициентов передачи отражений помехи от источника помехи до входа АКП. Во втором столбце таблицы представлены оптимальные значения вектора **B** управляющих коэффициентов для случаев «больших» (эксперименты 1 и 3) и «малых» задержек (эксперименты 2 и 4). В четвертом и пятом столбцах таблицы представлены значения сигнал-помеха для случаев «больших» и «малых» задержек, соответственно.

Видно, что изменение вектора управляющих коэффициентов от оптимального значения при «больших» задержках до оптимального при «малых» задержках слабо сказывается на степени подавления помех. Существенное влияние на подавление помех при «малых» задержках оказывает вектор \mathbf{V} .

Определение оптимального вектора B для случаев «больших» и «малых» задержек осуществляется путем расчета векторов F_1 , F_2 (показано выше) на основе анализа матрицы помех R_n и соотнесения расстояний между соседними отводами линий задержки с интервалом корреляции сигнала.

На практике часто встречаются мощные радиоимпульсные помехи [3, 7]. На основе проведенного анализа, а также работы [84] предложены разработанные с участием автора устройства приема с компенсацией радиоимпульсных помех [120, 121]. Устройство [120] предназначено для приема аналоговых сигналов, а устройство

[121] – для широкополосных цифровых [8].

При отсутствии помех в устройстве происходит обычный прием разнесенных сигналов с линейной схемой сложения [1, 20, 100]. Когда появляется помеха, то срабатывают обнаружители помех, управляющие коммутаторами сигнала, которые на входы линейного сумматора подают один и тот же полезный сигнал, очищенный от помех. Тем самым разнесенный прием лишь на короткое время действия импульсной помехи превращается в прием без разнесения.

4.5. Адаптивная компенсация помех с обратной связью по решению

Использование адаптивных компенсаторов помех (АКП) наталкивается на опасность ошибочного подавления полезного сигнала за счет влияния его на работу цепей адаптации [118]. Встает задача разработки алгоритмов адаптивной компенсации помех, обеспечивающих нечувствительность цепей адаптации к сигналу. Известные методы защиты АКП от ошибочной настройки становятся очень сложными либо малоэффективными в случае многолучевых каналов, когда спектр принимаемого сигнала характеризуется значительной априорной неопределенностью.

Ниже рассматривается возможность уменьшения влияния полезного фазоманипулированного сигнала, прошедшего канал с достаточно медленно меняющейся частотной характеристикой, на настройку АКП при подавлении внешних помех с широкой полосой частот [50,51]. Суть предлагаемого алгоритма настройки анализируется на примере простейшего двухэлементного АКП, структура которого изображена на Рисунке 4.9.

Схема управления (СУ) определяет коэффициент W весового суммирования сигнала с комплексной огибающей U_1 , поступающего от антенны A_1 , с сигналом комплексной амплитуды U_0 от антенны A_0 . Перемножитель (П) осуществляет операцию комплексного перемножения $U_1 * W$. Инвертор (Ин) и фильтр нижних

частот (ФНЧ) служат для установки требуемого знака и реализации необходимого усреднения корреляционной обратной связи в АКП. К выходу сумматора (∑) подключается линейный тракт (ЛТ) приемника с демодулятором (Дм) на выходе. Структура АКП и дальнейший ее анализ приведены для комплексных огибающих сигнала и помехи.

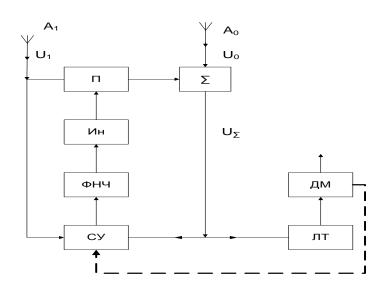


Рисунок 4.9. Структура двухвходового компенсатора помех с компенсацией влияния сигнала на цепи адаптации.

При реализации часто используемого алгоритма наименьших квадратов СУ содержит только перемножитель комплексных огибающих, выход которого подключен к указанному на схеме ФНЧ [98]. Значение комплексного коэффициента W весового суммирования напряжений U_1 и U_0 определяется выражением:

$$W = -\mu \bar{y}, y = U_1^* U_{\Sigma} = U_1^* U_0 + W U_1^* U_1$$
 (4.37)

где: μ - коэффициент передачи петли корреляционной обратной связи, * - знак комплексного сопряжения, а черта сверху означает операцию усреднения во времени. Пусть

$$U_i = U_{ic} + U_{in} + U_{i \text{ m}}, (4.38)$$

где индекс i принимает значения либо 1, либо 0; U_{ic} , U_{in} , $U_{i\, \text{ш}}$ — комплексные огибающие сигнала, помехи и белого гауссова шума на соответствующей

антенне. В дальнейшем считаем, что комплексные огибающие шума на элементах АКП не коррелированы. Сигнал и помеха также взаимно не коррелированы. Из выражений (4.37), (4.38) - следует:

$$y = (U_{1c} + U_{1\pi} + U_{1\underline{\mathbf{u}}})^* (U_{\Sigma c} + U_{\Sigma \pi} + U_{\Sigma \underline{\mathbf{u}}}) =$$

$$= U_{1c}^* U_{\Sigma c} + U_{1c}^* (U_{\Sigma \pi} + U_{\Sigma \underline{\mathbf{u}}}) + (U_{1\pi} + U_{1\underline{\mathbf{u}}})^* (U_{\Sigma c} + U_{\Sigma \pi} + U_{\Sigma \underline{\mathbf{u}}})$$

$$(4.39)$$

Из полученного выражения следует, что при последующем усреднении величины у первое слагаемое в выражении отражает нежелательное влияние полезного сигнала на настройку коэффициента весового суммирования W, второе слагаемое при усреднении превращается в нуль, третье слагаемое - обусловливает настройку АКП на подавление помехи. Рассмотрим выражение:

$$y_{\tau} = U_1^* U_{\Sigma \tau} = (U_{1c} + U_{1\pi} + U_{1m})^* (U_{\Sigma c\tau} + U_{\Sigma \pi \tau} + U_{\Sigma m\tau})$$
(4.40)

где: $U_{\Sigma c \tau}$ - комплексная огибающая U_{Σ} на выходе сумматора, задержанная на $\tau = kT$, где k - целое, T - период следования информационных посылок фазоманипулированного сигнала. Пусть значение k выбрано так, что

$$\tau_k > \tau > \tau_{\Pi} \tag{4.41}$$

где τ_k - интервал времени, на котором частотная характеристика радиоканала остается неизменной, $\tau_{\rm n}$ - интервал корреляции помехи. По аналогии с выражением (4.39) влияние сигнала на настройку АКП и в данном случае обусловлено аналогичным слагаемым в выражении (4.40) $a_{\tau} = U_{1c}^* U_{\Sigma c \tau}$.

Учитывая условия выбора величины τ , считая, что источники помех перемещаются достаточно медленно, то есть $W(t)=W(t-\tau)$, и учитывая, что полезный сигнал является фазоманипулированным, имеем

$$U_{\Sigma c\tau} = N U_{\Sigma c}, \tag{4.42}$$

где N=1, если информационные посылки, разделенные интервалом времени τ , одинаковы, и N=-1 в противном случае. Если N известно, то можно подавить влияние сигналов на настройку АКП путем алгебраического сложения величин y и y_{τ} . Знак суммирования при этом выбирается так, чтобы произошла взаимная компенсация слагаемых $U_{1c}^*U_{\Sigma c}$ и $U_{1c}^*U_{\Sigma c\tau}=NU_{1c}^*U_{\Sigma c}$ из выражений (4.39) и (4.42).

Оценку \widehat{N} можно получить с помощью демодулятора сигналов с относительной фазовой манипуляцией [3], дополняя его схемой сравнения информационных символов, разделенных интервалом $\tau = kT$. Используя оценку \widehat{N} , то есть, организуя обратную связь по решению (ОСР), определим значение весового коэффициента W с помощью выражения:

$$W = -\mu \overline{\left[y - \widehat{N}y_{\tau}\right]}.\tag{4.43}$$

Подставляя в формулу (4.43) выражения (4.39), (4.40), после простых преобразований получаем

$$W = -\mu \left[\overline{U_{1c}^* U_{\Sigma c} (1 - N\widehat{N})} + \overline{(U_{1\pi} + U_{1\pi})^* (U_{\Sigma\pi} + U_{\Sigma\pi})} \right] =$$

$$= -\mu \left[\overline{U_{1c}^* (W U_{1c} + U_{oc})} \left(\overline{1 - N\widehat{N}} \right) + \overline{U_{1\pi}^* (W U_{1\pi} + U_{o\pi})} + \overline{+U_{1\pi}^* (W U_{1\pi} + U_{o\pi})} \right]$$
(4.44)

Учитывая, что $\overline{N}\widehat{N}=1-2\hat{p}$, где \hat{p} - вероятность ошибочного определения N, получаем

$$W = \frac{-\mu \left(2\hat{p}\overline{U_{1c}^*U_{oc}} + \overline{U_{1\pi}^*U_{o\pi}}\right)}{1 + \mu \left(2\hat{p}\overline{|U_{1c}|^2} + \overline{|U_{1\pi}|^2} + \overline{|U_{1\pi}|^2}\right)} = \frac{-\mu \left(2\hat{p}\rho_c\sqrt{\overline{|U_{1c}|^2}\overline{|U_{oc}|^2}} + \rho_{\pi}\sqrt{\overline{|U_{1\pi}|^2}\overline{|U_{o\pi}|^2}}\right)}{1 + \mu \left(2\hat{p}\overline{|U_{1c}|^2} + \overline{|U_{1\pi}|^2} + \overline{|U_{1\pi}|^2}\right)},$$

$$(4.45)$$

где $\rho_{c(\pi)} = U_{1c(\pi)}^* U_{oc(\pi)} / \sqrt{\left|U_{1c(\pi)}\right|^2} \left|\overline{U_{oc(\pi)}}\right|^2$ - коэффициент корреляции комплексных огибающих на входах АКП. Из полученного выражения видно, что при достаточно малой величине \hat{p} настройка W слабо зависит от сигнала. Так как приемные антенны находятся близко друг от друга, то принимаем

$$\overline{|U_{1c}|^2} = \overline{|U_{oc}|^2} = \overline{|U_c|^2}, \overline{|U_{1\pi}|^2} = \overline{|U_{o\pi}|^2} = \overline{|U_{\pi}|^2},
\overline{|U_{1m}|^2} = \overline{|U_{om}|^2} = \overline{|U_{m}|^2}, |\rho_c| = |\rho_{\pi}| = 1.$$
(4.46)

С учетом равенств (4.46) выражение (4.45) приводится к виду

$$W = \frac{2\hat{p}|\overline{U_c}|^2 \exp(j\varphi_c) + |\overline{U_{\Pi}}|^2 \exp(j\varphi_{\Pi})}{\mu^{-1} + 2\hat{p}|\overline{U_c}|^2 + |\overline{U_{\Pi}}|^2 + |\overline{U_{\Pi}}|^2},$$
(4.47)

где φ_c , φ_{Π} - разности аргументов комплексных огибающих сигналов и помехи на входах АКП

$$q_{\Sigma} = (p_{\Sigma_{III}} + p_{\Sigma_{II}})/p_{\Sigma_{C}} = (\overline{|U_{\Sigma_{III}}|^{2}} + \overline{|U_{\Sigma_{II}}|^{2}})/\overline{|U_{\Sigma_{C}}|^{2}} = \frac{\overline{|U_{\Pi}|^{2}} * \{1 + |W|^{2} + 2Re[W^{*} \exp(j\varphi_{\Pi})]\} + \overline{|U_{III}|^{2}}(1 + |W|^{2})}{|U_{C}|^{2} \{1 + |W|^{2} + 2Re[W^{*} \exp(j\varphi_{C})]\}},$$

$$(4.48)$$

где $p_{\Sigma \text{ш}},\, p_{\Sigma \text{п}},\, p_{\Sigma \text{c}}$ - мощности шума, помехи, сигнала на выходе АКП.

С целью упрощения выражения для q_{Σ} , конкретизируем условия работы АКП. Без потери общности анализа считаем $\varphi_c = 0$. В зависимости от направления прихода помехи после окончания процесса адаптации уровень сигнала на выходе АКП изменяется, так как меняется аргумент комплексного коэффициента W весового суммирования. Чем ближе значение φ_{Π} к нулю, тем АКП. меньше сигнал после настройки поэтому эффективность рассматриваемого алгоритма ниже. Ограничимся ситуацией, когда $\phi_{\scriptscriptstyle \Pi} \ge 90^\circ$. При этом наиболее "неприятным" является случай наименьшего сигнала $U_{\Sigma {
m c}},$ когда $\varphi_{\rm n} = 90^{\circ}$. Его в дальнейшем и будем рассматривать. При последующем анализе будем также считать, что коэффициент усиления μ достаточно велик, и $\mu^{-1} \ll$ $\overline{|U_{\rm III}|^2}$, что позволяет упростить выражение (4.47). С учетом сделанных допущений выражения (4.47), (4.48) приобретают вид:

$$W \approx -\frac{2\hat{p}|\overline{U_c}|^2 + j|\overline{U_{\Pi}}|^2}{2\hat{p}|\overline{U_c}|^2 + |\overline{U_{\Pi}}|^2 + |\overline{U_{\Pi}}|^2} = \frac{-2\hat{p} - jq_{\Pi}}{2\hat{p} + q_{\Pi} + q_{\Pi}},$$
(4.49)

$$q_{\Sigma} = \frac{q_{\Pi}(1 + |W|^2 + 2Re(jW^*)) + q_{\Pi}(1 + |W|^2)}{1 + |W|^2 + 2ReW^*},$$
(4.50)

где $q_{\Pi} = \overline{|U_{\Pi}|^2}/\overline{|U_{\rm c}|^2}$; $q_{\rm m} = \overline{|U_{\rm m}|^2}/\overline{|U_{\rm c}|^2}$. Подставляя выражение (4.49) в (4.50),

получаем

$$q_{\Sigma} = f_{2}(\hat{p}) =$$

$$= \frac{q_{\Pi}[4\hat{p}^{2} + (q_{\Pi} + 2\hat{p})^{2}] + q_{\Pi}[(2\hat{p} + q_{\Pi} + q_{\Pi})^{2} + 4\hat{p}^{2} + q_{\Pi}^{2}]}{(q_{\Pi} + q_{\Pi})^{2} + (q_{\Pi})^{2}}$$

$$(4.51)$$

Здесь первое слагаемое числителя характеризует уровень помехи на выходе АКП, а второе - шума. Для нахождения q_{Σ} необходимо определить зависимость вероятности ошибки \hat{p} от уровней сигнала, помехи и шума на входе демодулятора. Так как решение о значении N в выражении (4.43) принимают в результате логической обработки двух информационных посылок, задержанных относительно друг друга на величину $\tau = kT$ то

$$\hat{p} = f_1(p_{\text{MM}}) = 2p_{\text{MM}}(1 - p_{\text{MM}}), \tag{4.52}$$

где $p_{\rm дм}$ - вероятность ошибки демодулятора при приеме одного информационного символа фазоманипулированного сигнала.

Для упрощения выкладок ограничимся случаем, когда и помеха, и шум являются гауссовыми процессами с равномерной спектральной плотностью в полосе пропускания. В этом случае вероятность ошибки зависит лишь от отношения q_{Σ} на выходе АКП (считаем, что линейный тракт приемника не меняет существенно величину q_{Σ}). Подставляя выражение (4.52) формулу (4.51) и используя известную [36] зависимость $p_{\text{дм}}(q_{\Sigma})$ получаем уравнение

$$q_{\Sigma} = f_2(\hat{p}) = f_2[f_1(p_{\text{дм}})] = f_2\{f_1[p_{\text{дм}}(q_{\Sigma})]\},$$

решением которого является искомая величина q_{Σ} . Однако ввиду трансцендентности выражения $p_{\phi_{\rm M}}(q_{\Sigma})$ такой путь нахождения q_{Σ} достаточно сложен. Поэтому определим q_{Σ} путем численного решения системы двух уравнений

$$\begin{cases} \hat{p} = f_1(p_{\text{MM}}(h_{\Sigma})) = 2p_{\text{MM}}(1 - p_{\text{MM}}), \\ h_{\Sigma} = [q_{\Sigma}(\hat{p})]^{-1} = \{ [q_{\Sigma}(p_{\phi_{M}})] \}^{-1} \end{cases}$$
(4.53)

где: $h_{\Sigma} = q_{\Sigma}^{-1} =$ - отношение сигнал-шум, $q_{\Sigma}(\hat{p})$ определяется выражением (4.52),

а зависимость $p_{\rm дм}(h_{\Sigma})$ для сигнала с относительной фазовой манипуляцией заимствована из [36, 123].

$$p_{\text{дм}} = p_{\text{офм}} = 2p_{\text{фм}} (1 - p_{\text{фм}}),$$

$$p_{\text{фм}} = 1 - F(\sqrt{2h_{\Sigma}}), \qquad F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp[E - 0.5t^{2}) dt$$

$$h_{\Sigma} = p_{c}/(p_{\text{III}} + p_{\pi}) = E/N_{0}$$
(4.54)

Здесь F(x) - табулированная функция Лапласа; E - энергия одного информационного символа, $N_{\rm o}$ - односторонняя спектральная плотность мощности белого шума, образованного помехой и собственным шумом приемника. При решении предполагается, что в первый момент времени демодулятор работает под воздействием шума и некомпенсированной помехи, определяется вероятность $p_{\rm дм}$ ошибки демодулятора. Вероятность ошибки затем используется для вычисления h_{Σ} . Затем найденное значение h_{Σ} используется для вычисления второго значения $p_{\rm дм}$, которое используется для вычисления второго значения h_{Σ} . Описанный процесс нахождения h_{Σ} итеративно повторяется несколько раз. В результате находится значение h_{Σ} , соответствующее решению системы указанных уравнений.

На Рисунке 4.10 приведены графики зависимостей $h_{\Sigma} = h_{\rm out} (h_{\rm in} = (q_{\rm in} + q_{\rm in})^{-1})$. Графики построены для разных значений $q_{\rm in}$. Величина $q_{\rm in}$ изменяется в пределах: 0...4. Соответствие графиков значениям $q_{\rm in}$ отражено Таблицей 4.3. Здесь в третьей и четвертых колонках указаны соответствующие значения сигнал-шум и вероятность ошибки демодулятора ОФМ-сигналов при наличии только шума (внешняя помеха отсутствует).

Таблица 4.3. Вероятность ошибки демодулятора ОФМ – сигнала в зависимости от отношения сигнал-шум

№ графика	q_{uu}	h _{ш дБ}	$p_{o\phi_{\mathcal{M}}}$
1	0,0839	10,76	10^{-6}
2	0,1028	9,88	10^{-5}
3	0,1327	8,77	10^{-4}

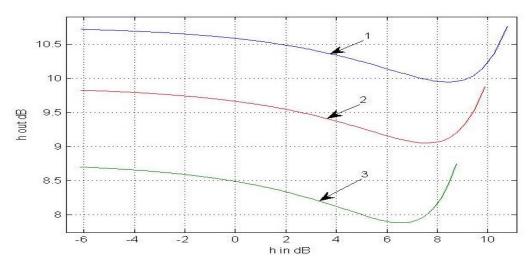


Рисунок 4.10. Графики зависимости отношения сигнал-шум на выходе компенсатора от отношения сигнал-шум на его входе

Из графиков следует, что при значительных изменениях величины q_{Π} отношение h_{Σ} «сигнал/(помеха+шум)» меняется незначительно и имеет большое значение, которое обеспечивает малую вероятность ошибки демодулятора. С уменьшением уровня шума относительно сигнала возрастает отношение h_{Σ} , что обеспечивает меньшую вероятность ошибки. Минимальные значения h_{Σ} , наблюдаемые на графиках, соответствуют случаю, когда внешняя помеха становится соизмеримой с уровнем шума.

4.6. Устройство приема широкополосных сигналов с подавлением импульсных помех

Радиоимпульсные помехи являются очень мощными и часто встречаются на практике. В описании к авторскому свидетельству [I20] была предложена схема разработанного с участием автора устройства приема аналоговых сигналов с двукратным разнесением, в котором осуществляется компенсация радиоимпульсных помех. При отсутствии помех в устройстве происходит обычный прием разнесенных сигналов с линейной схемой сложения [17, 46, 93].

Когда появляется помеха, то срабатывают обнаружители помех, управляющие коммутаторами сигнала, которые на входы линейного сумматора подают один и тот же полезный сигнал, очищенный от помех. Тем самым разнесенный прием на время действия помехи превращается в прием без разнесения. После прекращения действия помехи вновь восстанавливается режим пространственно-разнесенного приема.

При пространственно-разнесенном приеме широкополосных цифровых сигналов частотно-селективными замираниями ИЗ входных сигналов опорный, необходимый для демодуляции формируется который сигнал, очищается от шумов путем узкополосной фильтрации [6, 8, 9]. Если при компенсации радиоимпульсных помех осуществлять коммутацию входных сигналов, как предлагается в [120], то возникает длительный переходный процесс изменения опорного сигнала, который по длительности равен нескольким тактам следования информационных посылок. Поэтому для эффективного применения устройства, предложенного в [120], необходимо, чтобы параметры помехи удовлетворяли следующим условиям.

$$\tau_{\rm M} \gg T_{\rm on} + T_{\rm oc}, \tau_3 \gg T_{\rm M} \gg \tau_{\rm M} + T_{\rm on} + T_{\rm o2};$$
 (4.55)

где $\tau_{\rm u}$, $T_{\rm u}$ - длительность и период следования помеховых импульсов; $T_{\rm on}$, $T_{\rm oc}$ - времена срабатывания обнаружителя помех и формирователя опорного сигнала демодулятора; τ_3 - постоянная времени разряда запоминающих конденсаторов УВХ, входящих в состав адаптивных трансверсальных фильтров. Из (4.36) следует, что применение устройства [120] целесообразно либо при достаточно больших значениях $\tau_{\rm u}$ и $T_{\rm u}$, которые превышают длительность нескольких информационных посылок, либо, когда допустимо использовать прием без разнесения.

На практике часто требуется обеспечить подавление более коротких импульсов. Рассмотрим изображенную на рис. 4.11. схему предложенного, при

участии автора [121], устройства приема двоичных широкополосных сигналов, подавляющего радиоимпульсные помехи. Проанализируем сначала работу устройства при отсутствии помех. С выходов блоков усиления и преобразования частоты (БУП) сигналы непосредственно и через линию задержки ЛЗ поступают на входы каналов демодуляции. Так как на выходах обнаружителей помех (ОП) импульсы отсутствуют, то размыкающие ключи РК1, РК2 находятся в замкнутом состоянии, а переключающие ключи ПК подключают выходы демодуляторов ДМ к сумматору С2. При этом на выходах согласованных фильтров СФ1,2 возникают серии узких импульсов, соответствующих "сжатым" во времени широкополосным сигналам, прошедшим многолучевой канал. Каждый согласованный фильтр выделяет лишь соответствующий двоичный символ.

С выходов согласованных фильтров сигналы поступают на входы демодуляторов ДМ на входы сумматора Σ . Очищенные от шумов с помощью гребенчатого фильтра накопления ГФН, построенного на основе рециркулятора, импульсы поступают на опорные входы d демодуляторов. Каждый демодулятор содержит два коррелятора Кор1 и Кор2, на входы которых поступают импульсы с выходов согласованных фильтров СФ1 и СФ2.

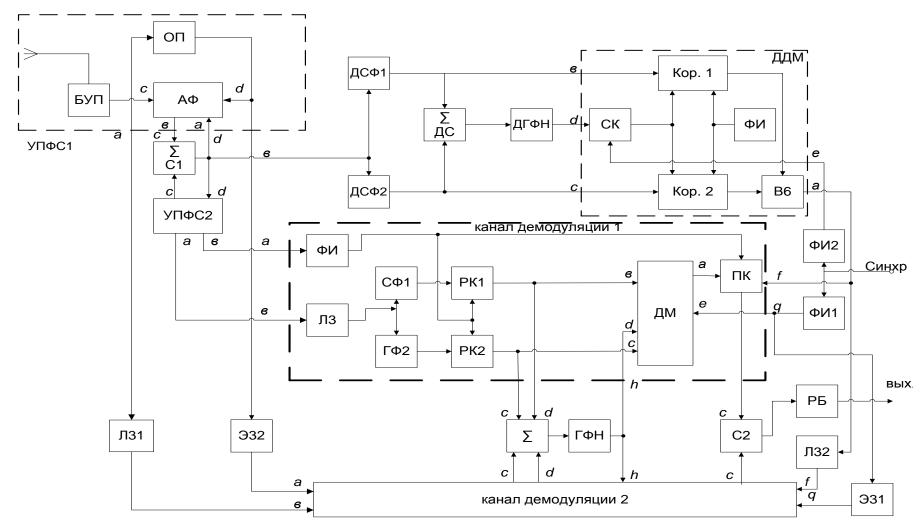


Рис. 4.11. Компенсатор широкополосных помех

На другие входы корреляторов через стробирующий ключ СК поступает опорный сигнал (вход d). Ключ СК управляется импульсами синхронизации, длительность которых определяется формирователем импульсов ФИ2 и равна длительности интегрирования сигналов в корреляторах, которая, в свою очередь, определяется длительностью сигнала на выходе согласованных фильтров. После окончания интегрирования импульс синхронизации своим задним фронтом запускает формирователь коротких импульсов ФИ, который сбрасывает интеграторы корреляторов Кор 1, 2 в исходное состояние.

С выходов корреляторов сигналы поступают на вычитающий блок ВБ, с выхода которого сигналы, суммируясь в сумматоре С2, поступают на решающий блок РБ. Задержка импульсов управления стробирующим ключом СК, осуществляемая первым элементом задержки ЭЗ1, равна задержке сигналов, осуществляемой первой линией ЛЗ1 задержки, и обеспечивает временное разнесение принимаемых сигналов.

При появлении помехи на антеннах на выходах обнаружителей помехи возникают импульсы, которые поступают на стробирующие входы d адаптивных трансверсальных фильтров $A\Phi$, обеспечивая их функционирование, и запускают формирователи Φ И импульсов каналов демодуляции. На выходах этих формирователей появляются импульсы с длительностью равной длительности переходных процессов в согласованных фильтрах каналов демодуляции при воздействии на них радиоимпульсных помех. На это время размыкающие ключи PK1 и PK2 отключают согласованные фильтры от входов демодуляторов и сумматора Σ . Одновременно на такое же время переключающие ключи каналов демодуляции подключают к входам второго двухвходового сумматора C2 выход дополнительного демодулятора ДДМ непосредственно и через вторую линию задержки $\Pi32$.

На входы b и c дополнительного демодулятора ДДМ информационные сигналы поступают с подавленной помехой, поэтому она не влияет на работу

решающего блока. После окончания мешающего влияния помехи на согласованные фильтры все ключи каналов демодуляции возвращаются в исходное состояние, обеспечивая устройству прежни режим работы. Линия задержки ЛЗ, стоящая в канале демодуляции, обеспечивает задержку сигнала на время, необходимое для срабатывания обнаружителей помехи ОП и ключей канала демодуляции. Тем самым исключается прохождение помеховых импульсов на выходы каналов демодуляции.

Принятые антеннами помехи после усиления и преобразования частоты в блоках БУП поступают на адаптивные фильтры АФ. Эти фильтры корректируют форму, уровень и фазу радиоимпульсной помехи так, что на выходах адаптивных фильтров помехи имеют одинаковую форму, уровень, но сдвинуты по фазе на 180°. Складываясь далее на первом двухвходовом сумматоре С1, они взаимно компенсируются.

Очищенный от помехи сигнал поступает на дополнительные согласованные демодулятора. фильтры ДСФ1,2 далее на входы дополнительного Дополнительный двухвходовый сумматор ДС и дополнительный гребенчатый фильтр ДГФН накопления формируют очищенный от шумов опорный сигнал для дополнительного демодулятора (вход d). Вторая линия задержки Л32 имеет такое же время задержки, что и первая линия задержки Л31, и служит для согласования во времени работы дополнительного демодулятора и демодулятора первого ПК демодуляции. Переключающие ключи подключают выход вычитающего блока ВБ к входу решающего блока РБ.

Коммутация каналов приема происходит по выходам вычитателей, где отсутствуют высокоизбирательные цепи. Поэтому длительность переходных процессов мала. Размыкающие ключи РК1 и РК2 на время действия помехи и переходных процессов в согласованных, фильтрах: СФ1, 2 отключают от них узкополосный тракт формирования опорного сигнала. Тем самым длительность переходных процессов при отключении КП определяется лишь величиной $T_{C\Phi}$. При широкой полосе согласованных фильтров и коротких импульсах помехи

процессы в тракте формирования опорного сигнала изменяются незначительно. Поэтому при обратном переключении в режим приема без помех тракт опорного сигнала будет готов к приему сигналов.

Эффективное подавление помех как следует из (4.55), обеспечивается данным устройством при соблюдении следующих условий:

$$T_{\rm oc} \gg \tau_{\rm u} \gg T_{\rm on}, \tau_3 \gg T_{\rm u} \gg \tau_{\rm u} + T_{\rm on}.$$
 (4.56)

В этом случае устройство подавляет помехи на протяжении почти всей длительности импульса и обеспечивает разнесенный прием в течение практически полного периода следования импульсов. Если же длительность помеховых импульсов очень велика, и $\tau_{\rm u} \geq T_{\rm oc}$, то опорный сигнал на выходе гребенчатого фильтра за время действия помехи значительно уменьшается. Для его восстановления после прекращения действия помехи требуется дополнительное время, в течение которого устройство должно оставаться в режиме приема без разнесения. В данном случае обнаружитель помех после окончания импульса вырабатывает соответствующий сигнал управления с некоторой задержкой $\tau_{\rm on} \approx T_{\rm on}$.

Рассмотренное устройство позволяет эффективно подавлять многолучевые импульсные помехи при изменении длительности и периода следования импульсов в широких пределах. Оно обеспечивает разнесенный прием в течение почти всего периода следования импульсов. Прием становится не разнесенным лишь на время действия помехового импульса. Поэтому надежность связи при действии помех, характеризующихся большой скважностью, уменьшается незначительно.

Выводы по разделу 4

1. Получены выражения, отражающие влияние погрешности реализации трансверсальных фильтров в составе компенсатора широкополосных помех, на степень подавления помех.

- 2. По результатам анализа адаптивного весового сумматора на основе процессора Хоуэлса-Эпплбаума определена зависимость оптимального вектора управляющих коэффициентов процессора от корреляционных матриц помехи и сигнала на отводах линий задержки. Моделирование адаптивного компенсатора помех подтвердило справедливость полученных при анализе выражений.
- 3. Предложен алгоритм подавления влияния сигнала на работу цепей адаптации компенсатора помех. Алгоритм использует обратную связь по решению.
- 4. В результате анализа предложенного алгоритма получены выражения, отражающие зависимость отношения сигнал-помеха на выходе компенсатора от отношения сигнал-помеха на его входе.
- 5. Использование алгоритма позволяет подавить помеху даже в случае, когда отношение сигнал-помеха на входе компенсатора больше единицы. В частности, при отношении сигнал-помеха на входе равном 4 дБ отношение сигнал-помеха на выходе равно 8,2 дБ.
 - 5. Предложены устройства компенсации импульсных помех.

5. Программные средства для исследования помехоустойчивости систем распознавания голосовых команд

5.1. Функциональный состав программного обеспечения, используемого при разработке САРГК

Рассмотрим основные этапы разработки. Сначала из имеющегося звукового материала формируются две выборки произнесений различных слов, а также помех, возникающих вне интервала произнесений слов [38]. Одна выборка предназначена для создания моделей отдельных звуков речи и помех (для «обучения» системы), а вторая – для тестирования системы с целью определения параметров достоверности распознавания. Обучающая выборка снабжается метками времени, обозначающими границы слов, пауз, звуков. На данном этапе используются программные средства, обеспечивающие прослушивание звукового материала, наблюдение за формой и спектром сигналов, а также простановку меток времени.

На втором этапе определяются параметры PC, выбирается алгоритм фиксации голосовой активности диктора (Voice Activity Detector – VAD) и алгоритмы подавления помех на стадии предварительной обработки сигнала. На данном этапе используются программные средства, обеспечивающие моделирование источников помех, моделирование (синтез) отдельных сегментов PC, оценку целесообразности использования указанных выше алгоритмов, параметров PC. Анализируются форма и спектр помех. Оценивается степень изменения параметров PC при воздействии помех и использовании алгоритмов их подавления. Степень изменения параметров PC оценивается как для отдельных сегментов сигнала, так и для всей тестовой выборки произнесений слов.

На третьем этапе создаются акустические модели помех, пауз и звуков, а также обобщенные акустические модели помех и звуков, относящиеся к нескольким звукам речи и помехам. Создается сеть акустических моделей в

соответствии с возможными результатами распознавания, исследуются алгоритмы сопоставления параметров РС с акустическими моделями.

На четвертом этапе система тестируется с целью определения параметров достоверности распознавания — оценок вероятностей ложного срабатывания системы и ложного пропуска команды. По результатам тестирования уточняются параметры алгоритмов подавления помех состав обобщенных моделей, а также параметры акустических моделей.

При разработке САРГК используется программные средства [6, 35, 44, 57, 164, 174, 181, 205]. По функциональному составу и общности используемых алгоритмов программные средства, применяемые при разработке САРГК, можно разделить на несколько основных категорий. К первой категории относятся программы, предназначенные для анализа сигналов и помех (наблюдение за формой, спектром, прослушивание), а также для простановки меток времени, которые указывают границы отдельных звуков и произнесений слов. Данные задачи решаются во многих областях науки и техники, поэтому здесь целесообразно использовать универсальные программные средства. В частности, существует несколько звуковых редакторов, которые могут использоваться для решения указанных задач [23]. Простановка меток времени, которые определяют границы слов и звуков, осуществляется с помощью специализированного программного обеспечения [47, 54, 80]. Примером такого ПО являются разработанные Центром Речевых Технологий программные средства, (г.С.Петербург) [108].

К отдельной категории относятся программы, которые обеспечивают исследование алгоритмов предварительной обработки сигналов. С помощью указанных программ проводится моделирование источников помех и синтез отдельных сегментов РС, исследуются алгоритмы выделения пауз, формирования параметров сигнала, подавления помех.

К третьей категории относятся программы, предназначенные для работы с акустическими моделями. С помощью программ создаются модели, сеть моделей, а также исследуются алгоритмы сопоставления параметров РС с моделями.

К четвертой категории относятся программы, обеспечивающие оценку достоверности распознавания. Они обеспечивают формирование транскрипции текста, слова которого произносят дикторы при тестировании системы, фиксацию фактов ложного срабатывания системы и пропуска команды, замены одной команды на другую, с их помощью проводят статистическую обработку результатов тестирования.

По способу организации пользовательского интерфейса программные средства можно подразделить на следующие категории. К первой категории следует отнести ПО, которое использует текстовый язык сценариев для последовательного вызова требуемых функций. Такой способ организации интерфейса позволяет быстро менять программу исследований. Примером такого ПО является система Matlab. Она имеет богатую библиотеку функций, связанных с обработкой сигналов, а также функций математической статистики [18, 174]. Поэтому целесообразно ее использовать в качестве программной среды для решения соответствующих задач.

Однако следует учесть, что реализуемые в системе Matlab сценарии вызова различных функций могут быть очень длинными и трудными для уяснения логики обработки сигналов. Данное обстоятельство затрудняет процесс изменения программы исследования. Кроме того, из-за отсутствия процедуры компиляции программного кода, программы в среде Matlab выполняются медленнее по сравнению с программами, созданными с помощью универсальных языков программирования.

Если задачи исследования в процессе его проведения меняются мало, и требуются большие вычислительные ресурсы, то следует использовать указанные выше языки программирования и хорошо развитый интерфейс пользователя.

Если необходимо часто менять порядок исследования, которое требует решения ряда типичных задач, то целесообразно использование метода графического программирования, когда имеется несколько программных модулей, которые относительно быстро связываются пользователем в требуемую структуру, отображаемую графически.

Существует несколько программных пакетов, предназначенных разработки систем распознавания речи [164, 181, 205]. Из них наиболее распространена программная среда Hidden Markov Models Tool Kit - HTK (разработана в английском университете г. Кембриджа) [205]. Кроме того, разработан набор соответствующих функций для работы в среде Matlab [174]. Все указанные программные средства сориентированы, в основном, на создание систем, использующих акустические модели для подавления помех. Алгоритмам подавления помех на стадии предварительной обработки сигналов уделено мало внимания. Кроме τογο, данные программные средства рассчитаны использование в научных лабораториях, когда круг решаемых задач достаточно широкий. Использование их при проектировании систем, предназначенных для конкретных условий эксплуатации, требует больших затрат времени.

Ниже рассмотрены особенности программных средств, которые, в основном, предназначены для обеспечения помехоустойчивости САРГК на стадии предварительной обработки сигналов. Кроме того, решается задача сокращения объема экспериментов при настройке системы на конкретные условия эксплуатации.

5.2. Исследование алгоритмов предварительной обработки сигналов

Автор принял участие в разработке ряда программных средств для исследования алгоритмов формирования параметров РС. Разработки проводились совместно с П.С.Рагузиным и К.Е.Левиным. Исследованы алгоритмы определения коэффициентов частной корреляции на основе использования лестничных фильтров, а также определения частоты основного тона на основе использования

функции средней разности (Average Magnitude Difference Function – AMDF) [6, 35, 44, 46, 47, 53, 56, 60, 61, 62, 64, 67, 68, 75, 76, 78, 116, 117]. Созданное ПО позволяет выбрать оптимальный порог классификации сегментов сигнала на вокализованные и невокализованные, оценить погрешность моделирования сегмента РС в зависимости от порядка предсказания, а также зафиксировать траектории изменения параметров РС во времени и признаки траекторий – характерные их отличия для различных команд.

Разработана программа для исследования алгоритма удаления пауз из речевого сигнала [47, 80]. Исследовался алгоритм, основанный на подсчете числа переходов сигнала через ноль. Удалось сократить число ошибок фиксации границ паузы путем модернизации известного алгоритма. Для контроля правильности работы алгоритма обеспечен просмотр «осциллограмм» голосовых команд после удаления пауз.

Ниже рассмотрены возможности некоторых разработанных программных средств.

5.2.1. Исследование определителя основного тона

Определение частоты F_{OT} (или периода T_{OT}) основного тона речевого сигнала в ряде случаев позволяет повысить достоверность распознавания за счет выбора совокупности моделей команд по результатам предварительной классификации голосов дикторов на женские и мужские. Классификация основана на том факте, что частота основного тона для женских голосов больше по сравнению с мужскими голосами [106]. Определение F_{OT} осуществляется для вокализованных сегментов речевого сигнала после классификации сегментов речевого сигнала на вокализованные и невокализованные.

Определение периода T_{OT} основного тона и классификация сегментов сигнала могут осуществляться различными методами [114, 115]. В частности, используется кратковременная функция среднего значения разности AMDF (Average Magnitude Difference Function). Метод основан на том, что для строго

периодической функции X(n) с периодом P последовательность d(n) = X(n) - X(n+k) равна нулю при $k = 0, \pm P, \pm 2P, \ldots$ Для сегментов вокализованного речевого сигнала естественно ожидать, что последовательность d(n) будет близка k нулю при k, кратном периоду основного тона. Кратковременное среднее значение величины d(n) как функция k становится минимальным, если значение k близко k периоду основного тона. Кратковременная функция среднего значения разности (AMDF) определяется как

$$y(k) = \sum_{n=0}^{N-1} |X(n) - X(n+k)|,$$

где X(n) — значение входного сигнала в момент времени $nT_{_{\rm J}}$; $T_{_{\rm J}}$ — период дискретизации; N — число выборок в сегменте сигнала. В общем случае X(n) — сумма периодического и случайного компонентов, поэтому данная функция является случайной. Очевидно, что если последовательность X(n) близка к периодической функции на интервале, выделенном с помощью временного окна, то y(k) принимает минимальные значения при $k=P,2P\ldots$

Период T_{ot} основного тона определяется расстоянием между двумя минимумами функции. Если отношение минимума функции к максимуму больше заданного порогового значения $r_{k\cdot}$, то принимается решение о невокализованности звука. Если отношение меньше порога, то звук считается вокализованным.

В процессе исследования определяется оптимальное значение порога классификации, обеспечивающее минимум ошибок классификации. Для оптимального порога определяются значения ошибок классификации, и определяется период (частота) основного тона. Воздействие помех изменяет значение оптимального порога.

Исследования базируются на следующих процедурах обработки сигнала. Для заданного звукового файла определяются коэффициенты частной корреляции (КЧК) всех сегментов соответствующего речевого сигнала. Найденные коэффициенты используются для формирования частотной характеристики нерекурсивного инверсного фильтра. Его АЧХ является обратной по отношению к огибающей спектра соответствующего сегмента речевого сигнала. Речевой сигнал проходит через инверсный фильтр — формируется сигнал остатка предсказания.

По «осциллограмме» сигнала остатка предсказания исследователь «вручную» классифицирует сегменты на вокализованные и невокализованные. Результат классификации соответствует работе идеального классификатора. Для вокализованных сегментов вручную определяется частота основного тона.

Далее для исследуемого алгоритма классификации устанавливается оптимальный порог классификации путем сравнения результата классификации с результатом идеальной классификации. Меняя порог, добиваются минимального числа ошибок. После установки оптимального порога определяется период основного тона на вокализованных сегментах.

Затем с помощью медианной фильтрации устраняются одиночные случайные выбросы-сбои в результате измерения. Полученный после медианной фильтрации результат принимается за результат определения периода основного тона для исследуемого алгоритма классификации.

Далее сравниваются результаты идеального и «реального» измерения периода. По результатам сравнения формируется таблица, где указывается число ошибок измерения различного типа. Проанализируем указанные выше процедуры обработки сигнала более детально.

Речевой сигнал сначала проходит через предыскажающий фильтр с системной функцией $h(z)=1-0.95z^{-1}$, который обеспечивает более точную последующую оценку спектра, затем сигнал разбивается на ряд перекрывающихся во времени сегментов.

Перекрытие сегментов во времени обеспечивает более точное отслеживание изменения параметров сигнала во времени. Далее каждый сегмент взвешивается окном Хэмминга. Затем определяется корреляционная функция для каждого сегмента, и с помощью алгоритма Левинсона-Дарбина [106] определяются КЧК.

Найденные коэффициенты используются для построения инверсного фильтра. На выходе фильтра формируется сигнал остатка предсказания (Рисунок 5.1).

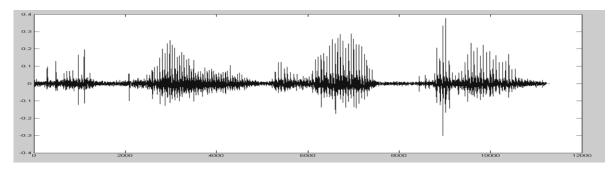


Рисунок. 5.1. Сигнал остатка предсказания

Для каждого сегмента сигнала остатка предсказания определяется функция среднего значения разности. Типичный график функции для вокализованного сегмента представлен на Рисунке 5.2.

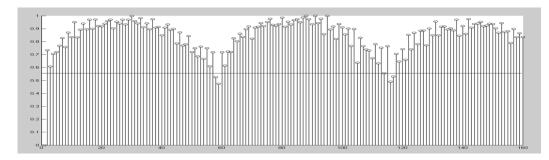


Рисунок. 5.2. Функция среднего значения разности

Прямая линия соответствует пороговому значению классификации. Эта величина задаётся исследователем. Для вокализованных звуков она должна превышать минимальное значение AMDF. Выбранный порог одинаков для всех сегментов. Для некоторых невокализованных сегментов порог превышает минимальное значение AMDF, а для некоторых вокализованных порог меньше минимального значения AMDF. В первом случае сегмент классифицируется как вокализованный, а во втором случае — невокализованный, возникают ошибки классификации. Оптимизация порога минимизирует число ошибок классификации.

Для определения периода основного тона определяется расстояние между двумя соседними минимумами AMDF. Ошибки в определении периода основного тона бывают трех типов: когда результат измерения в несколько раз больше,

когда результат измерения в несколько раз меньше, когда ошибка измерения не относится к типам ошибок, указанных выше [106, 196].

На Рисунке 5.3 представлены результаты определения периода основного тона до медианной фильтрации (верхний график) и после (нижний график).

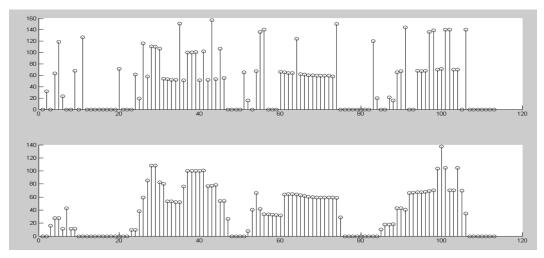


Рисунок.5.3. Траектории значения периода основного тона до и после медианной фильтрации Видно, что на втором графике количество «кратных» ошибок уменьшилось.

5.2.2. Исследование алгоритма адаптивной компенсации помех

При решении вопроса о целесообразности применения адаптивной компенсации широкополосных помех, возникающих в радиоканале, необходимо оценить, насколько повышается отношение сигнал-помеха при использовании компенсации. Основными исходными данными для получения оценки являются параметры тракта распространения помехи: значение частоты несущей, форма помехи, величины задержек отражений помехи, коэффициенты отражений, погрешность реализации требуемых задержек в адаптивных фильтрах компенсатора, а также длительность информационной посылки сигнала.

В результате работы программы на выходе компенсатора определяются: отношение сигнал-помеха, массивы данных сигнала и помехи, корреляционные матрицы помехи и сигнала на выходах многоотводных линий задержек, входящих в состав адаптивных фильтров компенсатора [52]. Ниже приведен фрагмент сценария вызова соответствующих функций системы Matlab для случая

воздействия радиоимпульсной помехи на прием фазоманипулированного сигнала. Согласно данному сценарию проведено исследование адаптивного компенсатора радиопомех, результаты которого представлены в разделе 4. Текст m-функций приведен в Приложении.

Сценарий вызова функций для исследования адаптивной компенсации помех

<u>% Моделирование помех на входах компенсатора в тракте промежуточной частоты приемного устройства</u>

f0=70e6; % частота несущей 70 МГц

Fs=700e6; % частота дискретизации 700МГц

bwr=-6; % уровень (дБ) задания ширины спектра радиоимпульсной помехи % (гауссов импульс)

bw=6/70; % относительная ширина спектра помехи

Т=8е-6; % период следования импульсов помехи 8 мкс

tau1=[0.5e-6,1.5e-6]; tau2=[1e-6,3e-6]; % задержки (в секундах) для случая

% двулучевого приема помехи на входах двухвходового компенсатора помех

 $v1=[1\ 2];\ v2=[3\ 4];\%$ коэффициенты передачи помехи по каждому лучу от

% источника до входа компенсатора помех

delta_tau=1/7e-7; % погрешность реализации требуемой задержки в % компенсаторе помех

Td=15e-6; % длительность информационной посылки

t_end=100e-6; % длительность интервала наблюдения

[N1,N2]=pomexa_sim12(t_end,Fs,f0,T,bw,bwr,tau1,tau2,v1,v2); % результат % (N1,N2) прохождения помехи от источника до входов компенсатора

% Помеха на выходе компенсатора помех

 $[snr2,W2,Nout2,inputN2,R2]=compens10(1e4*[v2,-v1]',N1,N2,tau2+delta_tau,tau1,Fs,t_end,T);$

% пропускаем помеху через компенсатор и определяем отношение %snr2 сигнал-помеха на его выходе для случая больших задержек, % W2 - вектор весовых коэффициентов,

, so the second to spanning s,

% Nout2 - напряжение помехи на выходе компенсатора,

% inputN2 и R2 - вектор напряжений помехи и корреляционная матрица % помехи на выводах линий задержки компенсатора; вектор управления % оптимален

<u>% Моделирование сигналов на входах компенсатора помех</u> [s1,s2]=sig_psk2(Fs,f0,Td);

% формируем две не коррелированных копии фазоманипулированного %сигнала: одна копия задержана относительно другой на длительность %Td информационной посылки

% Сигнал на выходе компенсатора помех

[s_out2, inputS2, Rs2, Ps_out2]=sig_out4(s1,s2,Fs,tau2+delta_tau,tau1,W2); % пропускаем сигнал через компенсатор и определяем сигналы %inputS2 на % выходах двух адаптивных фильтров компенсатора, %корреляционную матрицу % Rs2 сигнала на выходах линий задержки и мощность %Ps_out2 сигнала %s_out2 на выходе компенсатора

<u>% Отношение сигнал-помеха на выходе компенсатора помех</u> Pn_out2=W2'*R2*W2; % мощность помехи на выходе компенсатора SNR2=Ps_out2/Pn_out2; % отношение сигнал-помеха на выходе компенсатора

С помощью рассмотренного сценария можно определить максимально допустимую величину погрешности реализации задержек в адаптивных фильтрах компенсатора для получения требуемой степени подавления помехи.

5.2.3. Программный комплекс с использованием среды графического программирования

При проведении исследований часто приходится применять ограниченный набор типовых функций с изменяемыми параметрами, которые используются в различных сценариях. В этом случае затраты времени на создание программного обеспечения исследований можно значительно сократить, используя среду графического программирования. С этой целью в среде C++ Builder был разработан соответствующий программный комплекс [57, 168].

Комплекс содержит набор динамических библиотек - функциональных блоков, которые моделируют источники сигналов и помех, средства обработки сигналов и оценки результатов экспериментов. При проведении исследований организуются соответствующие связи между функциональными блоками. Блоки и связи между ними отображаются графически. Главное окно программы изображено на Рисунке 5.4.

Здесь отображена схема организации эксперимента, целью которого является сравнение параметров речевого сигнала до и после воздействия помехи. В блок «Файло» загружается совокупность звуковых файлов, содержащих множество произнесений ключевого слова команды. «Детектор активности абонента» определяет границы между информативной частью речевого сигнала и паузами. «Усилители-аттенюаторы» преобразуют значения границ, выраженные в номерах отсчетов сигнала, в номера его сегментов — интервалов квазистационарности сигнала, которые формируются при его анализе.

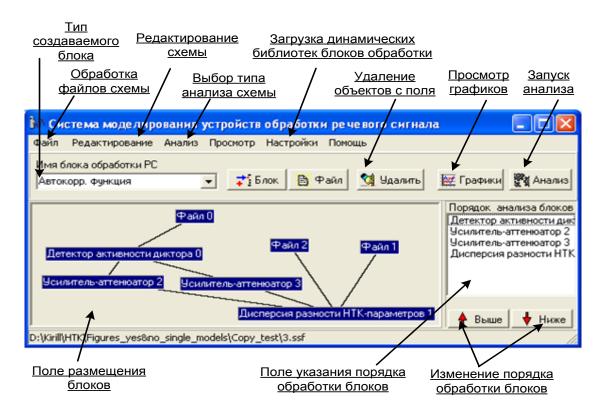


Рисунок 5.4. Главное окно программы

Блоки: «Файл1» и «Файл2» - содержат массивы параметров сигнала без помехи и с помехой. Блок «Дисперсия разности НТК-парметров1» определяет дисперсию разности между массивами параметров информативной части сигнала. В правой части окна указана последовательность «прохождения» функциональных блоков при проведении эксперимента.

На Рисунке 5.5 изображена схема определения зависимости параметров огибающей спектра речевого сигнала от уровня помехи типа «белый шум» для одного сегмента.

Совокупность блоков «Генератор импульсов» и «Синтезирующий фильтр» обеспечивает формирование вокализованного сегмента речевого сигнала. Параметры огибающей спектра сегмента поступают из блока «Файл0». В сумматоре «Сумматор3» к синтезированному сегменту сигнала с помощью блока «Генератор шума 2» добавляется помеха — шум. Анализирующий фильтр «Ан.фильтр» определяет параметры огибающей кратковременного спектра сигнала — коэффициенты частной корреляции (КЧК). Для графического представления результатов определения КЧК используется блок «Извлечение

КЧК5», в котором последовательность векторов КЧК преобразуется в траектории отдельных КЧК.

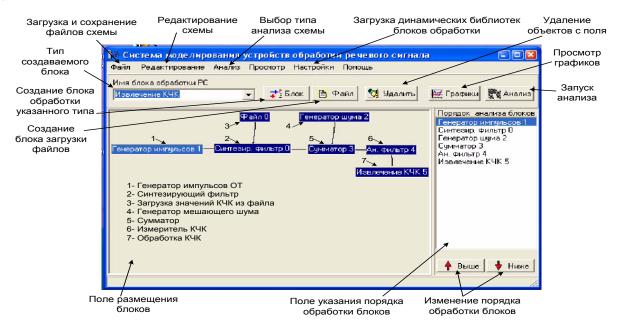


Рисунок 5.5.Определение зависимости параметров огибающей спектра речевого сигнала от уровня помехи

Для каждого КЧК генерируется множество реализаций шума. Результаты определения КЧК усредняются. Усреднение результатов проводится с использованием блока итерационного анализа. Данная процедура повторяется с измененным значением мощности генератора шума — определяется зависимость КЧК от мощности шума. Окно настройки блока итерационного анализа приведено на Рисунке 5.6.

В левой части окна приведены параметры, характеризующие пределы изменения мощности (в данном случае - среднеквадратического отклонения) генератора шума.

Помимо приведенных примеров, программный комплекс используется для определения частоты основного тона, графического отображения формы сигнала и различных зависимостей.

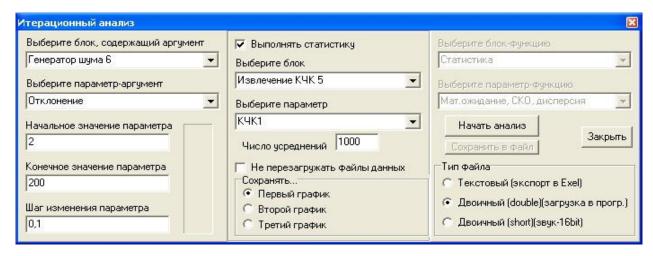


Рисунок 5.6. Окно настройки блока итерационного анализа

5.2. Программные средства для работы с акустическими моделями

Существует несколько пакетов программ, с помощью которых создаются акустические модели и реализуются алгоритмы распознавания команд [164, 181, 205]. Наиболее часто используется пакет Hidden Markov Model Toolkit (HTK) [205]. На основе пакета НТК была разработана методика формирования акустических моделей трифонов, состоящая из нескольких этапов [92]. Точность созданных моделей была проверена путем тестирования САРГК, рассчитанной на распознавание русских названий цифр и произнесений слов: «да», «нет», «ошибка». Результаты тестирования САРГК показали, что относительная частота ошибок распознавания составила 94% при общем числе произнесений равном 2600.

В работах [66, 77] рассмотрен вопрос сокращения вычислительных затрат при проверке правильности созданных акустических моделей и алгоритма сопоставления параметров РС с моделями. Предложен алгоритм формирования относительно небольшого числа последовательностей векторов параметров РС в соответствии с созданными моделями. Разработана программа, которая реализует процесс распознавания сформированных последовательностей векторов с помощью моделей. По результатам распознавания оцениваются параметры моделей. Найденные оценки сравниваются с известными значениями параметров.

По результатам сравнения делается вывод о соответствии оценок параметрам моделей.

Как указывалось в разделе 1, для подавления помех, которые находятся вне интервала произнесения команды, в состав сети моделей вводятся модели помех. Модели помех могут быть представлены либо отдельным узлом в графе сети гипотез (например, шум дыхания), либо некоторой структурой, состоящей из узлов и дуг. В частности, такой структурой моделируется непроизвольное произнесение неключевого слова. В последнем случае узлы представляют собой монофонов, объединяются которые дугами В соответствующую целью снижения вычислительных структуру. C затрат при проведении распознавания, кроме указанных выше видов моделей помех, используются обобщенные модели помех, которые содержат суммарную информацию о нескольких видах помех и звуках речи.

При формировании сети моделей («грамматики») требуется часто менять вид и состав моделей помех, а также параметры моделей, определяющих их значимость при распознавании. Для снижения затрат времени на редактирование разработана программа-редактор, сети была которая предназначена ДЛЯ оперативного изменения состава моделей команд и помех, а также их параметров. обладает графическим Программа развитым интерфейсом Рассмотрим особенности программы на примере редактирования сети моделей, граф которой изображен на Рисунке 5.7.

Данной структуре соответствует перечень узлов и дуг, приведенный в Таблице 5.1. В двух колонках справа перечислены дуги с указанием номеров узлов начала и окончания дуги. Узлы 10...59 содержат модели произнесений ключевых слов, а модели монофонов, из которых состоят непроизвольные произнесения неключевых слов, а также обобщенные модели монофонов и помех находятся в узлах 60...88.

В процессе формирования сети моделей определяется состав моделей помех, а также значения их параметров (em_add, wd_add), которые обеспечивают

максимальную достоверность распознавания произнесений заданного состава ключевых слов. Модели помех midsp являются обобщенными в наибольшей степени и соответствуют совокупности помех, для которых не созданы модели, соответствующие узлам 60...88. При редактировании сети моделей оперативно меняются состав и параметры моделей, расположенных между нулевыми узлами 3, 5 и нулевыми узлами 4, 6. На Рисунке 5.8 приведено окно программы.

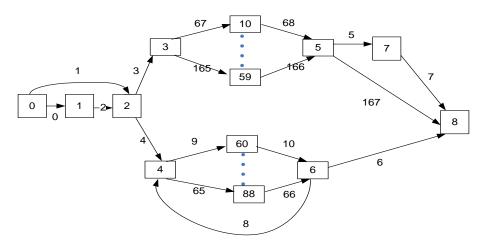


Рисунок 5.7. Граф редактируемой сети моделей

Таблица 5.1. Состав редактируемой сети моделей

N=13 L=18		I=9 W=midsp		J=0	S=0	E=1	J=9 S=4 E=60
I=0	W=!START	I=10	W=Алдушин	J=1	S=0	E=2	J=10 S=60 E=6
I=1	W=pause1			J=2	S=1	E=2	
I=2	W=!NULL			J=3	S=2	E=3	
I=3	W=!NULL	I=59	W=Ильюшин	J=4	S=2	E=4	J=165 S=3 E=59
I=4	W=!NULL	I=60	W=00	J=5	S=5	E=7	J=166 S=59 E=5
I=5	W=!NULL			J=6	S=6	E=8	J=167 S=5 E=8
I=6	W=!NULL			J=7	S=7	E=8	
I=7	W=midsp	I=88	W=k	J=8	S=6	E=4	
I=8	W=pause2						

Здесь исходными для работы программы являются списки - словари ключевых слов, а также списки звуков помех и сочетаний звуков, слогов, которые входят в состав сети гипотез. В окне программы необходимо указать пути, где находятся соответствующие файлы. В нижней части окна частично показан текст, соответствующий сформированной сети моделей.

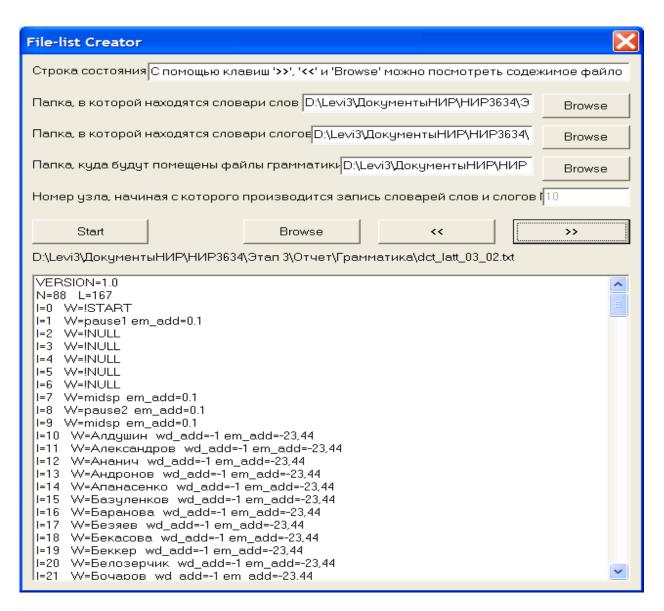


Рисунок 5.8. Окно программы для редактирования сети моделей

5.3. Программные средства для оценки достоверности распознавания команд

Все исследования, о которых говорилось во втором разделе, проводились с использованием специально разработанных программ. В частности, создана программа для определения необходимого количества испытаний при сравнении систем по достоверности распознавания с учетом заданной ошибки принятия неправильного решения по результатам сравнения.

Разработана программа для оценки вероятности ложного срабатывания с учетом степени несоответствия произнесения постороннего слова акустической модели команды

При создании системы распознавания команд необходимо определить относительную частоту ошибок распознавания при заданном уровне помехи и сравнить ее с максимально допустимым значением. Если найденное значение частоты ошибок превышает допустимое значение, то следует изменить состав команд для данного пункта меню. Из заданного перечня команд пункта голосового меню следует определить ту команду, которая обусловливает наибольшую частоту ошибок, и исключить ее данного перечня команд.

На основе критерия достоверности распознавания, рассмотренного в разделе 2, в системе Matlab с применением пакета HMM Toolbox [174] разработан набор функций и соответствующий сценарий их вызова для решения данной задачи. При последовательном вызове функций на вход системы распознавания подается тестовая выборка произнесений команд, которые относятся к данному пункту голосового меню. К произнесениям добавляется имитация помехи заданного вида и уровня. Для каждой команды фиксируется значение критерия сравнения САРГК по достоверности распознавания с целью выявления команды с Для минимальной достоверностью распознавания. выявленной команды определяется относительная вероятность ошибки, которая сравнивается с максимально допустимым значением.

На Рисунке 5.9 приведены графики зависимости критерия достоверности для десяти команд (названия цифр) от отношения сигнал-шум (дБ) при воздействии помехи типа «белый шум».

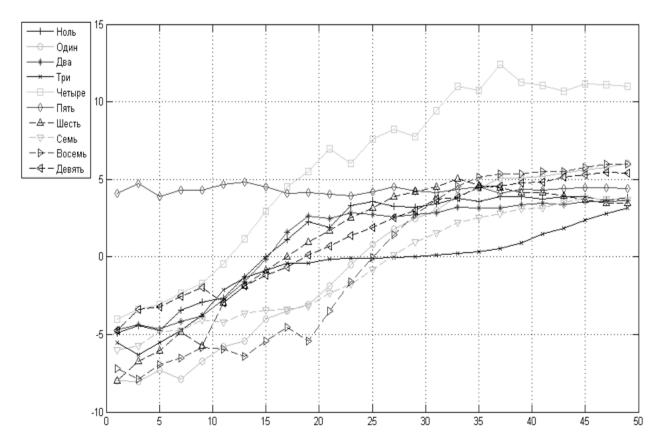


Рисунок 5.9. Графики зависимости критерия достоверности от отношения сигнал-шум Для построения графиков используется функция drawAlphaToSNRGraph. Листинг функции приведен в Приложении ПЗ.2.

Из графиков следует, что в диапазоне (28...48 дБ) ожидаемых значений отношения сигнал-помеха наименьшую достоверность распознавания обусловливает команда «три». Соответствующая матрица результатов распознавания (confusion matrix) всех команд представлена на Рисунке 5.10.

[]	'0'	'1'	'2'	'3'	'4'	'5'	'6'	'7'	'8'	'9'
'Ноль'	100	0	0	0	0	0	0	0	0	0
'Один'	0	100	0	0	0	0	0	0	0	0
'Два'	0	0	100	0	0	0	0	0	0	0
'Три'	0	0	0	94	0	0	0	0	0	0
'Четыре'	0	0	0	2	100	0	0	0	0	0
'Пять'	0	0	0	0	0	100	0	0	0	0
'Шесть'	0	0	0	0	0	0	100	0	0	0
'Семь'	0	0	0	0	0	0	0	100	0	0
'Восемь'	0	0	0	0	0	0	0	0	100	0
'Девять'	0	0	0	4	0	0	0	0	0	100

Рисунок 5.10. матрица результатов распознавания (confusion matrix)

Злесь верхней строке представлены цифры, названия которых произносились. В левом столбце представлены названия моделей команд – результатов распознавания. На пересечениях строк и столбцов указаны значения (%)относительных частот соответствия произнесения Из И модели. представленных результатов распознавания следует, что команда «три» в двух процентах случаев была распознана как команда «четыре», а в четырех процентах случаев – как «девять». Следовательно, для данной команды относительная частота ошибок распознавания является максимальной по сравнению с другими командами и составляет 6%, что превышает предельно допустимое значение 5% при заданном отношении сигнал-помеха равном 41 дБ. Целесообразно для повышения помехоустойчивости данный пункт меню разбить на два подпункта «названия цифр от нуля до четырех» и «названия цифр от пяти до девяти».

Выводы по разделу 5

- 1. Программное обеспечение разработки САРГК по функциональному составу предложено подразделить на следующие категории:
 - анализ речевых сигналов и помех, а также определение границ отдельных звуков и произнесений слов в звукозаписях;
 - исследование алгоритмов предварительной обработки сигналов: определение параметров сигналов, подавление помех, выделение пауз, моделирование источников помех и сегментов речевых сигналов;
 - создание акустических моделей, сети моделей, а также сопоставление параметров сигналов с моделями;
 - оценка параметров достоверности распознавания.
 - 2. По способу организации интерфейса пользователя программные средства целесообразно подразделить на следующие категории.

- Использование текстового сценария вызова функций из заранее созданного набора функций, что позволяет быстро менять вид исследований. В частности, такой интерфейс целесообразен при исследовании различных алгоритмов подавления помех, совмещенных во времени с голосовыми командами.
- Использование графической среды программирования. Целесообразно при проведении исследований с меньшей степенью разнообразия по сравнению с применением текстовых сценариев.
- Использование развитого графического интерфейса универсальных языков программирования – целесообразно при исследований большим проведении однотипных объемом вычислений. Скорость проведения исследований при ЭТОМ повышается. Такой интерфейс, в частности, целесообразен при анализе сигналов при формировании тестовых выборок звукозаписей.
- 3. Разработанное программное обеспечение исследований помехоустойчивости САРГК позволяет решать следующие основные задачи.
 - Исследование алгоритмов подавления помех с целью определения их оптимальных параметров.
 - Определение степени изменения параметров речевого сигнала при воздействии помех и использовании алгоритмов их подавления.
 - Оценка достоверности распознавания ГК при воздействии помех и использовании алгоритмов их подавления.

6. Аппаратное обеспечение исследований помехоустойчивости систем распознавания голосовых команд

Так как при разработке САРГК не всегда имеется полная информация о видах помех, воздействующих на систему в конкретных условиях ее эксплуатации, то целесообразно осуществить сбор дополнительных данных о возможных ошибках распознавания еще на стадии разработки. Необходимо иметь систему сбора информации, в которой фиксируются ошибки распознавания и условия их возникновения. Данная система подключается к уже используемой системе обслуживания телефонных звонков абонентов, которую планируется модернизировать путем использования голосового управления.

Необходимо так построить систему, чтобы сбор необходимой информации происходил с минимальными потерями времени для пользователей системы. Кроме того, необходимо учесть возможность оперативного изменения состава пунктов голосового меню, а также состава команд для каждого пункта меню.

В данном разделе рассматривается реализация системы сбора информации на примере организации голосового управления прямым доступом абонента телефонной сети общего пользования (ТСОП) к абоненту учрежденческой телефонной сети (УАТС) [91, 94]. Для подключения САРГК к телефонной линии требуются специальные программно-аппаратные средства.

При оценке эффективности адаптивного компенсатора радиопомех (АКП) на последнем этапе его разработки необходимо определить соответствующее повышение помехоустойчивости станции связи. Целесообразно это исследование провести в рамках кольцевой проверки работоспособности станции, когда выход передатчика после ослабления подается на вход приемника. Для подключения компенсатора помех к станции связи требуются специальные аппаратные средства. В разделе рассматривается вариант подключения АКП к станции дальней тропосферной связи.

6.1. Система сбора информации об ошибках распознавания при организации прямого доступа к абоненту УАТС

Прямой доступ абонента ТСОП к абоненту УАТС осуществляется обычно тональным набором дополнительного номера после установления соединения с УАТС [91]. Такая процедура достаточно длительная, и, кроме того, невозможна, ТСОП если телефонный аппарат абонента имеет возможность импульсного набора номера. Подача голосовой команды, например, в форме произнесения фамилии работника учреждения значительно упрощает для процесс установления требуемого телефонного пользователя соединения. Рассмотрим порядок «общения» пользователя с такой системой.

«Общение» начинается с момента телефонного вызова, поступившего на вход системы. Пользователь прослушивает короткое приветствие, в котором сообщается название организации, куда он обратился. Если пользователь знает правила «общения» с системой, то он может подать команду, не дожидаясь окончания приветствия. Если же абонент телефонной сети звонит в организацию первый раз после установки САРГК и не знает, что система может реагировать на ГК, то произнесенные им слова не соответствуют команде, и САРГК воспринимает произнесение, как постороннее - происходит отказ от распознавания. Данный отказ фиксируется, и пользователь прослушивает правила правильного произнесения команды.

После произнесения команды пользователю сообщаются результаты распознавания с просьбой подтвердить правильность или отметить факт ошибки распознавания. Если ГК распознана правильно, она выполняется. Если же имеет место ошибка, то к общению с пользователем подключается оператор. Весь диалог пользователя с САРГК, когда произошла ошибка распознавания, фиксируется для дальнейшего анализа. Оператор подключается также и в случае отказа пользователя от «общения» с САРГК.

Рассмотрим схему обработки телефонного вызова при организации прямого доступа с использованием САРГК (Рисунок 6.1).

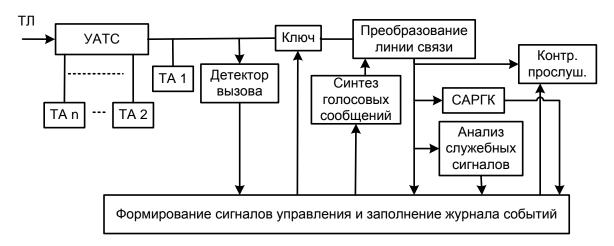


Рисунок 6.1. Схема организации прямого доступа с использованием САРГК

При поступлении вызова от абонента ТСОП срабатывает детектор вызова, что обеспечивает срабатывание ключевой схемы (Ключ) – происходит занятие телефонной линии и начинается «диалог» системы с абонентом. Блок «Преобразование линии связи» преобразует двухпроводную линию связи в четырехпроводную, появляется возможность использования синтеза голосовых сообщений (Синтез голосовых сообщений), также распознавания голосовых команд. Этот же блок обеспечивает гальваническую развязку телефонной линии и персонального компьютера, а также подавление местного эффекта [17, 91]. По результатам распознавания формируются сигналы управления блоками, указанными на схеме.

Если ошибки распознавания нет, то с помощью блока «Ключ» осуществляется набор номера вызываемого абонента УАТС. Далее САРГК в обслуживании телефонного вызова не используется. Если же произошла ошибка распознавания, то весь предыдущий «диалог» системы с абонентом записывается с указанием даты и времени события. Одновременно абоненту предлагается повторить команду. Если он отказывается от дальнейшего взаимодействия с САРГК, то фиксируется либо повторная ошибка, либо факт превышения длительности паузы заданного порога. В этом случае САРГК отключается, и блок

«Ключ» набирает номер телефонного аппарата (ТА) оператора, который обеспечивает обработку вызова.

Блок «Анализ служебных сигналов» предназначен для обработки служебных сигналов типа «Отбой» в телефонной линии, когда, например, абонент ТСОП внезапно положит трубку. Блок «Контр. прослуш.» обеспечивает контрольное прослушивание «диалога» системы с абонентом для уточнения причины возникновения ошибки распознавания либо для ускорения процесса обслуживания телефонного вызова.

Рассмотрим один из вариантов реализации схемы, изображенной на Рисунке 6.1. Блоки: «Ключ», «Детектор вызова» и «Преобразование линии связи» реализуются аппаратно в виде блока сопряжения персонального компьютера одного из работников учреждения с телефонной линией. Управление блоком осуществляется через коммуникационный порт компьютера. Звуковая карта компьютера обеспечивает посылку голосовых сообщений в линию и прием голосовых команд и служебных сигналов.

Все остальные блоки схемы реализуются программно. Соответствующее ПО устанавливается на компьютере работника учреждения. Телефонный аппарат работника обозначен на схеме как ТА 1. Указанный работник выполняет функции оператора системы. Если же работник отсутствует на рабочем месте, то функции оператора выполняет другой работник.

Рассмотренный вариант реализации схемы является достаточно экономичным: для автоматизированного обслуживания телефонного вызова не требуется выделения дополнительного порта УАТС и дополнительного компьютера.

6.2. Исследование компенсации широкополосных помех

При разработке средств адаптивной компенсации широкополосных помех, рассмотренных в разделе 4, возникает задача определения степени

подавления помех. Для решения данной задачи целесообразно использовать устройство компенсационной обработки сигналов (УКОС), структурная схема которого приведена на Рисунке 6.2.

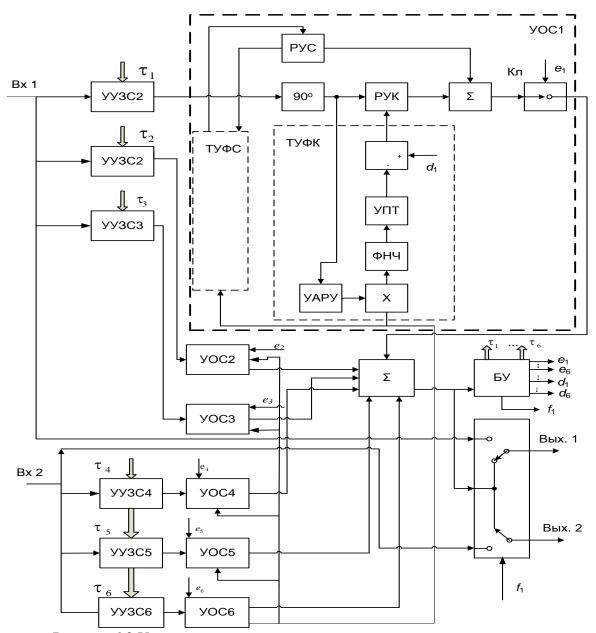


Рисунок 6.2. Устройство адаптивной компенсации широкополосных помех

Сигналы с входов устройства поступают на управляемые устройства задержки сигналов УУЗС. Их задержка изменяется в соответствии с цифровым кодом, поступающим от блока управления БУ. Выходы УУЗС подключены к устройствам обработки сигналов УОС, которые являются амплитуднофазовыми регуляторами (АФР), каждый из которых состоит из двух квадратурных каналов регулирования уровня входного сигнала и

фазовращателя на 90°.

Синфазный канал содержит регулируемый усилитель (РУС) и тракт усиления и фильтрации (ТУФС) сигнала. Квадратурный канал также содержит регулируемый усилитель РУК и тракт усиления и фильтрации ТУФК. Каждый тракт усиления и фильтрации содержит перемножитель, вход которого подключен к выходу усилителя с автоматической регулировкой усиления (УАРУ), фильтр нижних частот (ФНЧ), усилитель постоянного тока (УПТ) и вычитающее устройство.

Усилитель УАРУ обеспечивает постоянный уровень сигнала на входе перемножителя, что ведет к сужению требуемого динамического диапазона последнего. Ключ Кл. на выходе АФР служит для подключения и отключения блоков АФР от сумматора. Блок управления обеспечивает установку требуемых величин задержек τ_1 ... τ_6 , задание вектора управляющих коэффициентов d_1 ... d_6 , включение и выключение блоков АФР (сигналы управления ключами e_1 ... e_6), а также включение и выключение режима компенсации помех (сигнал управления ключами f_1).

Степень подавления помех и повышения помехоустойчивости приема целесообразно оценивать, используя кольцевой проверки средства работоспособности станции связи. Соответствующая схема приведена на Рисунке 6.3. Генератор псевдослучайного кода ПСК генерирует последовательность нулей информационное сообщение. единиц, имитирующих цифровом преобразующем устройстве ЦПУ происходит кодирование информации, которая затем поступает на модулятор М. Далее в передающем устройстве П сигнал переносится на рабочую частоту и, пройдя через высокочастотный регулируемый аттенюатор А, имитирующий затухание сигнала при его распространении, попадает на малошумящие усилители МШУ приемного устройства станции. После преобразователя частоты ПЧ сигнал поступает на входы предварительных усилителей сигнала промежуточной частоты ПУПЧ. Выходные напряжения

ПУПЧ суммируются с колебаниями имитатора помех ИП и поступают на входы УКОС.

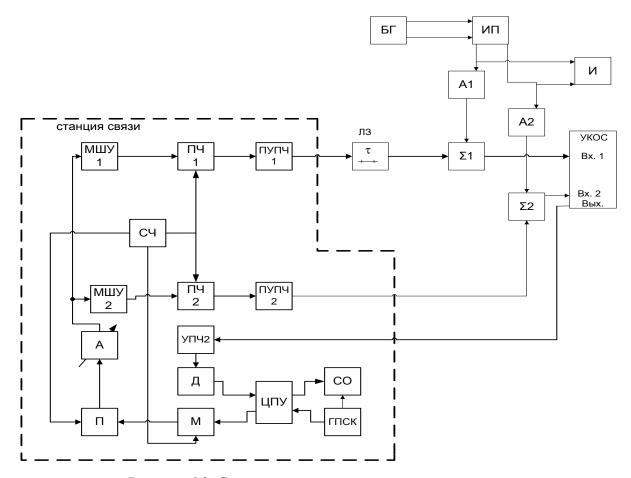


Рисунок 6.3. Схема кольцевой проверки станции связи

В УКОС происходит подавление помехи, и очищенный сигнал через усилитель промежуточной частоты УПЧ попадает на вход демодулятора Д. Далее, пройдя через ЦПУ, сигнал превращается в последовательность нулей и единиц, которая поступает на счетчик ошибок (СО). Счетчик фиксирует количество несовпадений символов принятой и исходной последовательности 'нулей и единиц за определенный интервал времени, позволяя оценить вероятность ошибки приема информации.

Синтезатор частоты СЧ обеспечивает необходимый для работы станции набор гармонических сигналов с требуемыми значениями частот. Все перечисленные блоки за исключением УКОС и имитатора помех ИП входят в состав станции связи. На схеме изображен также блок БГ генераторов помеховых колебаний, предназначенный для формирования колебаний различной формы,

имитирующих помехи. Кроме того, на схеме показаны фиксированные аттенюаторы A1, A2, служащие для снижения мощности сигналов помех, линия задержки ЛЗ необходимая для реализации фазовых различий полезных сигналов, принимаемых пространственно-разнесенными антеннами. Измеритель И, подключенный к выходам ИП, необходим для контроля уровня помех с выходов ИП. ИП предназначен для моделирования двух многолучевых каналов распространения помехи.

Степень повышения помехоустойчивости станции связи определяется по следующей методике. Сначала с помощью регулируемого аттенюатора А устанавливаются такие уровни сигналов станции, при которых относительная частота ошибки приема, определяемая счетчиком ошибок, равняется величине $p < p_{\text{пор}}$.

Затем, оставляя компенсатор помех УКОС включенным в сквозном режиме, увеличивают уровень помехи до тех пор, пока относительная частота появления ошибок не станет равной величине $p_{\text{пор}}$. Далее УКОС переводится в режим компенсации помех, и уровень помехи увеличивается до такого уровня, пока вновь относительная частота появления ошибок не станет равной величине $p_{\text{пор}}$. Показания N_1 и N_2 регулятора уровня помехи в первом и во втором случаях, соответственно, фиксируются. Отношение уровней помехи $d=N_2/N_1$ принимается за величину повышения помехоустойчивости станции. Величина d показывает, во сколько раз допустимо увеличение уровня помехи при включении компенсатора помех в состав станции связи.

При автоматизации исследований адаптивной компенсации радиопомех необходимо менять частоту несущих колебаний помехи по заданной программе. Использование для этой цели синтезаторов частоты приводит к значительному росту стоимости аппаратного обеспечения исследований.

На предприятиях, изготавливающих аппаратуру связи, широко применяются электронно-счетные частотомеры (ЭСЧ). Объединяя ЭСЧ с генератором, управляемым напряжением, в систему цифровой частотной

автоподстройки (ЦЧАП), можно реализовать функцию программного управления частотой помехи с меньшими затратами [29, 33, 34, 119, 127]. Погрешность установки частоты проанализирована в работе [95].

Одной из проблем реализации ЦЧАП является необходимость обеспечения достаточно быстрой перестройки частоты при сохранении высокой точности установки ее значения. С целью решения данной проблемы предложено менять точность измерения частоты в течение длительности переходного процесса установки нового значения частоты [41, 81, 89, 119].

С целью сокращения вычислительных затрат при определении разности между требуемым и текущим значениями частоты предложено использовать не все разряды многоразрядных чисел, представляющих значения частоты с высокой точностью, а лишь группу разрядов [127]. На первых этапах переходного процесса установки нового значения частоты используются лишь старшие разряды частотомера, так как не требуется определять частоту с высокой точностью. По мере уменьшения разности между текущим и требуемым значениями частоты точность измерения повышается, и происходит сдвиг группы разрядов числа, представляющего значение частоты, участвующих в вычислении разности, в сторону младших разрядов.

После окончания переходного процесса стабилизация значения частоты происходит с участием лишь небольшого количества младших разрядов результата измерения частоты. Для того, чтобы вычисленное по группе разрядов значение разности соответствовало ее истинному значению, в структуру ЦЧАП вводится дополнительно корректор разности.

Применение особого алгоритма вычисления разности позволяет ускорить процесс установки частоты и уменьшить объем вычислений.

Выводы по разделу 6

1. Для окончательной настройки САРГК разработана система сбора

информации об ошибках распознавания голосовых команд при эксплуатации САРГК в составе системы, которая обеспечивает прямой доступ клиента к абоненту учрежденческой АТС. Разработанная система предназначена для работы совместно с существующей системой обработки телефонных вызовов, и она оказывает незначительное влияние на возможное снижение скорости обслуживания телефонных вызовов.

- 2. Оценку помехоустойчивости связной радиостанции при компенсации широкополосных помех целесообразно проводить в режиме кольцевой проверки станции, используя отдельное устройство компенсации помех.
- 3. Для сокращения затрат на проведение исследований помехоустойчивости и снижения их стоимости целесообразно изменять частоту имитатора помехи автоматически, используя предложенные устройства цифровой частотной автоподстройки на базе электронно-счетного частотомера.

Заключение

- 1. Из-за высокой изменчивости произнесения ГК требуется большое количество экспериментов для определения оптимальных параметров алгоритмов обработки РС в условиях воздействия помех, а также для определения параметров достоверности распознавания ГК. Помехи создаются самим диктором (непроизвольное произнесение посторонних слов и звуков), его акустическим окружением и каналом связи. Параметрами достоверности распознавания являются относительные частоты ложного пропуска ГК и ложного срабатывания САРГК.
- 2. С расширением сферы использования САРГК увеличивается разнообразие помех и расширяется круг пользователей САРГК. Расширение круга пользователей требует повышения достоверности распознавания команд, что наряду с ростом разнообразия помех приводит к увеличению объема экспериментов при разработке САРГК.
- 3. Для обеспечения требуемой достоверности распознавания ГК в условиях действия помех подавление помех осуществляется на всех стадиях обработки РС в процессе распознавания ГК. На первой стадии помехи подавляются сразу же после отделения ГК от пауз, ее окружающих. Затем помехи подавляются на стадии определения набора (вектора) параметров РС соответствующей ГК. Подавление помех осуществляется и на последней стадии стадии собственно распознавания, когда последовательность векторов параметров РС сопоставляется с акустическими моделями команд с целью определения той модели, которая с наибольшей вероятностью соответствует данной последовательности.
- 4. Предложен метод определения относительной частоты ложного срабатывания с учетом отличия произнесений посторонних слов от произнесения ключевого слова ГК. Метод позволяет снизить объем экспериментов при определении оценки вероятности ложного срабатывания для звукового материала,

отличающегося от тестовой выборки звукозаписей. В частности, снижается объем экспериментов при настройке САРГК на конкретные условия ее эксплуатации.

- 5. Предложен метод сравнения результатов распознавания, полученных в ходе тестирования САРГК, по достоверности распознавания. В основе метода лежит критерий достоверности распознавания, который использует параметры статистического распределения значений разности метрик, полученных при сопоставлении произнесения со «своей» и ближайшей «чужой» акустическим моделями.
- 6. Использование методов позволяет сократить объем экспериментов при тестировании САРГК. В частности, если вероятности ложного пропуска команды сравниваемых результатов распознавания равны 10% и 13%, то при заданной вероятности ошибки сравнения равной 5% использование критерия позволяет сократить объем тестовой выборки звукозаписей голосовой команды с 580 до 400 звукозаписей (на 31%).
- 7. Использование синтезированных сегментов речевого сигнала позволяет исследовать влияние помех на параметры огибающей кратковременного спектра РС. В частности, показано, что влияние узкополосной помехи снижается при совпадении ее частоты с частотой форманты.
- 8. Показано, что значения параметров огибающей кратковременного спектра РС немонотонно зависят от отношения сигнал-помеха. Необходимо проводить экспериментальное исследование САРГК для различных значений отношения сигнал-помеха.
- 9. На примере экспериментального исследования метода подавления помех, основанного на вычитании спектра помехи в паузе РС из спектра сигнала, пораженного помехой, установлено следующее. При исследовании методов подавления помех целесообразно для снижения объема экспериментов определять параметры достоверности распознавания после определения степени изменения параметров РС. Контроль за изменением параметров РС позволяет

оценить целесообразность использования метода без проведения объемных экспериментов.

- 10. Предложен алгоритм подавления помехи, совмещенной во времени с произнесением ключевого слова ГК, путем взаимной компенсации с задержанной копией попадающего фрагмента помехи, В интервал паузы. Алгоритм обеспечивает подавление помех, представляющих собой сумму шума и периодических компонентов с медленно меняющейся частотой. Результаты эксперимента показывают, использование алгоритма, частности, что обеспечивает подавление импульсных помех на 27дБ.
- 11. Получены выражения, отражающие влияние частоты дискретизации и отношений «сигнал-шум», «периодический компонент-шум» на погрешность оценки периода помехи. Они позволяют определить минимально допустимую частоту дискретизации сигнала, а также минимально допустимые отношения «сигнал-шум», «периодический компонент-шум», обеспечивающие подавление помех.
- 12. Наличие пространственно-разнесенного приема обусловливает возможность подавления широкополосных радиопомех путем установки в тракты приема адаптивных трансверсальных фильтров. Получены выражения, отражающие зависимость уровня помехи на выходе компенсатора помех от погрешности реализации требуемых задержек в трансверсальных фильтрах. Справедливость выражений подтверждена моделированием адаптивного компенсатора помех.
- 13. Для повышения степени компенсации помех предложен и проанализирован алгоритм подавления влияния сигнала на работу цепей адаптации компенсатора помех. Алгоритм использует обратную связь по решению. Использование алгоритма позволяет подавить помеху даже в случае, когда отношение сигнал-помеха на входе компенсатора больше единицы. В частности, при отношении сигнал-помеха на входе равном 4 дБ отношение сигнал-помеха на выходе равно 8,2 дБ.

- 14. Предложены устройства адаптивной компенсации импульсных радиопомех, защищенные авторскими свидетельствами на изобретения.
- 15. Предложена классификация программных средств по функциональным возможностям необходимых при разработке помехоустойчивых САРГК. Разработка программных средств в соответствии с классификацией позволяет сократить их объем за счет использования одного и того же средства на разных стадиях разработки САРГК.
- 16. Разработаны программные средства для исследования помехоустойчивости САРГК, обеспечивающие, в частности, моделирование процессов компенсации помех, влияние помех на параметры РС, оценку параметров достоверности распознавание, редактирование сети акустических моделей, расчет объема тестовой выборки звукозаписей.
- 17. Разработана система сбора информации об ошибках распознавания голосовых команд при организации прямого доступа абонента телефонной сети общего пользования к абоненту учрежденческой АТС. Влияние ошибок распознавания на скорость обслуживания телефонных вызовов минимизировано.
- 18. Разработана структура аппаратных средств для оценки степени подавления широкополосных радиопомех при использовании адаптивной компенсации помех в составе станции связи.

Список основных сокращений

АКП – адаптивный компенсатор помех

АКФ – автокорреляционная функция

АО – аппаратное обеспечение

АТС – автоматическая телефонная станция

АТФ – адаптивный трансверсальный фильтр

АФ – анализирующий фильтр

АЧХ – амплитудно-частотная характеристика

БПФ – быстрое преобразование Фурье

ВОТ - выделитель основного тона

ВС – весовое суммирование

ГИ – генератор импульсов

ГК – голосовая команда

ГШ – генератор шума

ДПФ – дискретное преобразование Фурье

КЛП - коэффициенты линейного предсказания

КЧК - коэффициенты частной корреляции

ЛЗ – линия задержки

МСМП – модель скрытого марковского процесса

ОСР – обратная связь по решению

ПО – программное обеспечение

РС – речевой сигнал

СГС – система голосового самообслуживания

САРГК – система автоматического распознавания голосовых команд

СКО - среднеквадратическое отклонение

ТА – телефонный аппарат

ТК – телефонный канал

ТЛ – телефонная линия

ТСОП - телефонная сеть общего пользования

ТФ – трансверсальный фильтр

УАТС – учрежденческая автоматическая телефонная станция

ЧОТ - частота основного тона

AMDF – average magnitude difference function - функция средней разности

FAR – False alarm rate – относительная частота ложного срабатывания

FRR – False reject rate – относительная частота ложного пропуска

LPC – linear predictive coding - кодирование с линейным предсказанием

MFCC – mel-frequency cepstral coefficients – мелочастотные кепстральные коэффициенты

PLP – perceptual linear prediction - линейное предсказание с учетом модели человеческого слухового восприятия

VAD - Voice activity detector -детектор голосовой активности абонента

Список литературы

- 1. Андронов, И.С. Передача дискретных сообщений по параллельным каналам [Текст] / И.С. Андронов, Л.М. Финк. М.: Сов. радио, 1971. 403 с.
- Архипов И.О. Адаптивный алгоритм принятия решения «ТОН-НЕ ТОН», синхронный с основным тоном [Электронный ресурс] / И. А. Архипов, В. Б. Гитлин, Д. А. Лузин // Речевые технологии. 2009. №1. С.80-93. Режим доступа: http://speechtechnology.ru/files/1-2009.pdf Дата обращения: 11.01.2014.
- Берноскуни, Ю.В. Эффективный метод подавления импульсных помех в тропосферных системах связи [Текст] / Ю.В. Берноскуни, М.А. Быховский, В.В. Плеханов, В.В. Тимофеев // Электросвязь. 1984. № 9 С. II—14.
- 4. Богданов, Д.С. База речевых фрагментов русского языка «ISABASE» [Текст] / Д.С. Богданов, О.Ф. Кривнова, А.Я. Подрабинович, В.В. Фарсобина // Интеллектуальные технологии ввода и обработки информации. М., 1998.
- 5. Бондарко, Л.В. Звуковой строй современного русского языка: Учебное пособие для вузов [Текст] / Л.В. Бондарко. М.: Просвящение, 1977. 175 с.
- 6. Бондарь, Н.В. Моделирование анализирующего фильтра речевого кодека [Текст] / Н.В. Бондарь, Е.К. Левин // Проектирование и применение радиотехнических устройств и систем. Сборник научных трудов.- Владимир. : Владим. гос. техн. ун-т. 1996. С.74-77.
- 7. Быховский, М.А. Эффективность одного метода подавления импульсных помех [Текст] / М.А. Быховский // Электросвязь. 1985. № 12 С. 44-47.

- 8. Варакин, Л.Е. Системы связи с шумоподобными сигналами [Текст] / Л.Е. Варакин. М.: Радио и связь, 1985.- 384 с.
- 9. Винцюк, Т.К. Анализ, распознавание и интерпретация речевых сигналов. [Текст] / Т.К. Винцюк . Киев: Наук. думка, 1987. 264 с.
- 10. Воеводин, В.В. Матрицы и вычисления [Текст] / В.В. Воеводин, Ю.А. Кузнецов. М.: Наука, Главная редакция физико-математической литературы, 1984. 320 с.
- 11. Галкин, В.А. Цифровая мобильная радиосвязь: учебное пособие для вузов [Текст] / В.А. Галкин. М.: Горячая линия-Телеком, 2007. 432с.
- 12. Галунов, В.И Обеспечение помехоустойчивости при обработке информации в слуховой системе [Электронный ресурс] / В.И. Галунов, И.В. Королева // Сенсорные системы М., 1988. Т. 2 №2. С. 211 219.
 Режим доступа: http://www.auditech.ru/page/sensor.html Дата обращения: 10.01.2014.
- 13. Галунов, В.И. Состояние исследований в области речевых технологий и задачи, выдвигаемые государственными заказчиками [Электронный ресурс] / В.И.Галунов, Г.П. Кутуков, С.Н. Матюнин. Режим доступа: http://auditech.ru/page/galunov_d.html. Дата обращения : 24.11.12.
- 14. Галунов, В.И. Речь как система [Электронный ресурс] / В.И. Галунов // Сборник трудов XIII сессии Российского акустического общества. М., 2003. Т. 3 С. 19-21. Режим доступа: http://masters.donntu.edu.ua/2012/iii/akopyan/library/article2.htm Дата обращения: 10.01.2014
- 15. Гмурман, В.Е. Теория вероятностей и математическая статистика: Учебное пособие. 12-е изд., перераб. [Текст] / В.Е. Гмурман. М.: Высшее образование, 2006. 479 с.

- 16. ГОСТ 21655-87. Каналы и тракты магистральной первичной сети единой автоматизированной системы связи. Электрические параметры и методы измерений [Электронный ресурс] / М.: Издательство стандартов.- 1988.- 106 с. Режим доступа: http://standartgost.ru/%D0%93%D0%9E%D0%A1%D0%A2%2021655-
 - http://standartgost.ru/%D0%93%D0%9E%D0%A1%D0%A2%2021655-87#page-1 Дата обращения : 24.11.13
- 17. ГОСТ 7153-85. Аппараты телефонные общего применения. Общие технические условия [Текст] /М.: Издательство стандартов.- 1986.-38 с.
- 18. Гультяев, А.К. Имитационное моделирование в среде Windows: практическое пособие [Текст] / А.К. Гультяев. СПб.: КОРОНА принт, 2001. 400 с.
- 19. Давыденко, Ю.И. Дальняя тропосферная связь [Текст] / Ю.И. Давыденко.- М.: Воениздат, 1968. 211 с.
- 20. Дальняя тропосферная радиосвязь [Текст] / И.А. Гусятнинский, А.С. Немировский, А.В. Соколов и др. М.:Связь, 1968. 248 с.
- 21. Данилевский, Л.Н. Методы обработки сигналов в антенных решетках при наличии широкополосных помех [Текст] / Л.Н. Данилевский, О.В. Коробко, Б.И. Таурогинский // Зарубежная радиоэлектроника. 1987. № 3 С. 43-52.
- 22. Двайт, Г.Б. Таблицы интегралов и другие математические формулы [Текст] / Г.Б Двайт. М.: Наука, 1978. 224 с.
- 23. Дворянкин, С.В. Цифровая шумоочистка аудиоинформации [Текст] / С.В. Дворянкин. М.: РадиоСофт, 2011. 208с.
- 24. Егоров, Е.И. Использование радиочастотного спектра и радиопомехи [Текст] / Егоров Е.И., Калашников Н.И., Михайлов А.С. М.: Радио и связь, 1986. 304 с.

- 25. Ермоленко, Т.В. Методика подавления цветных шумов в речевом сигнале [Электронный ресурс] / Т.В. Ермоленко, Е.Е. Фёдоров // Речевые технологии. 2009. №3. С.3-13. Режим доступа: http://speechtechnology.ru/files/3-2009.pdf Дата обращения: 11.01.2014.
- 26. Загоруйко, Н.Г. Об исследованиях проблемы речевых технологий [Электронный ресурс] / Н.Г. Загоруйко // Речевые технологии. 2008. №3. С.81- 96 . Режим доступа: http://speechtechnology.ru/files/3-2008.pdf Дата обращения: 11.01.2014.
- 27. Заездный, А.М. Основы расчетов по статистической радиотехнике [Текст] / А.М Заездный. – М.: Связь, 1969. – 448 с.
- 28. Калашников, Н.И. Оценка устойчивости работы цифровых РРЛ в условиях частотно-селективных замираний [Текст] / Калашников Н.И., Каплунов П.Г. // -Электросвязь. 1985. № 11 С. 47- 50.
- 29. Карабко, Т.А. Применение электронносчетного частотомера для построения генератора синусоидальных колебаний, управляемого цифровым кодом [Текст] / Т.А. Карабко, В.А. Коньков, Е.К. Левин // Повышение эффективности и надежности РЭС: Межвуз. сб. науч. тр. Вып.3. Л.: ЛЭТИ, 1974. С. 81-84.
- 30. Кодзасов, С.В. Общая фонетика: учебник [Текст] / С.В. Кодзасов, Кривнова О.Ф. – М.: Рос. Гос. Гуманит. Ун-т, 2001. – 592 с.
- 31. Комашинский, В.И. Системы подвижной радиосвязи с пакетной передачей информации. Основы моделирования [Текст] / В.И. Комашинский, А.В.Максимов. М.: Горячая линия-Телеком, 2007. 176с.
- 32. Комптон, Р. Адаптивная антенная решетка в широкополосной системе связи [Текст] / Р. Комптон // ТИИЭР. 1978. Т. 66 №3 С. 23-34.

- 33. Коньков, В.А. Автоматизированный измеритель амплитудно-частотных характеристик [Текст] / В.А. Коньков, А.Б.Казаринов, Е. К. Левин, А.Д.Поздняков // Повышение эффективности и надежности РЭС: Межвуз. сб. науч. тр. Вып.б. Л.: ЛЭТИ,, 1976. С.85-87.
- 34. Коньков, В.А. Цифровой измеритель параметров амплитудно-частотных характеристик четырехполюсников [Текст] / В.А. Коньков, А.Б.Казаринов, Е.К. Левин, А.Д.Поздняков // Приборы и техника эксперимента. 1976. № 2. С..253
- 35. Копылов, С.Н. Программное обеспечение комплекса лабораторных работ по исследованию речевых сигналов [Текст] / С.Н. Копылов, Е.К. Левин // Проектирование и применение радиотехнических устройств. Тезисы докладов областной научно-технической конференции молодых специалистов и студентов. / Владим. гос. университет Владимир: ВлГУ 2000. С.22 23.
- 36. Коржик, В.И. Расчет помехоустойчивости систем передачи дискретных сообщений: справочник [Текст] / В.И. Коржик, Л. М. Финк, К.Н.Щелкунов; под ред. Л. М. Финка- М.: Радио и связь, 1981. 232с
- 37. Корн, Г. Справочник по математике для научных работников и инженеров [Текст] / Г. Корн, Т. Корн // М.: Наука, Главная редакция физико-математической литературы, 1973. 832 с.
- 38. Кривнова, О.Ф. Речевые корпусы (опыт разработки и использование) [Электронный ресурс] / О.Ф. Кривнова, Л.М. Захаров, Г.С. Строкин // Москва, МГУ. Режим доступа: http://www.dialog-21.ru/Archive/2001/volume2/2_33.htm Дата обращения : 10.01.2014.
- 39. Ксенофонтов, Р.Н. Исследование алгоритмов подавления помех в речевом сигнале при автоматическом распознавании голосовых команд [Текст] / Р.Н. Ксенофонтов, Е.К. Левин, К.Е. Левин // Перспективные

- технологии в средствах передачи информации. Материалы докладов шестой международной научно-технической конференции. Владимир, 2005. C. 275 276.
- 40. Левин, Б.Р. Теоретические основы статистической радиотехники.-3-е изд., перераб. и доп. [Текст] / Б.Р. Левин М.: Радио и связь, 1989. 656с.
- 41. Левин, Е. К. Повышение быстродействия цифровой системы автоматической подстройки частоты [Текст] / Е. К. Левин // Вопросы обработки сигналов: сб. науч. тр. Вып.1.- Л.: ЛЭТИ, 1976,. С. 11-16.
- 42. Левин, Е.К. Ввод в персональный компьютер и обработка речевой информации в медицинском учреждении [Текст] / Е.К. Левин // Физика и радиоэлектроника в медицине и биотехнологии. Материалы 2-ой международной научно-технической конференции. Владимир, 1996. С. 128-130.
- 43. Левин, Е.К. Влияние помех в телефонном канале связи на точность автоматического распознавания голосовых команд [Текст] / Е.К. Левин, К.Е.Левин // Обработка сигналов в системах телефонной связи и вещания. Материалы тринадцатой межрегиональной конференции. М.: МНТОРЭС им.А.С. Попова, 2004. С.90 92.
- 44. Левин, Е.К. Комплекс программных инструментальных средств для исследования речевых кодеков на основе линейного предсказания [Текст] / Е.К. Левин // Обработка сигналов в системах двусторонней телефонной связи. Пятая межрегиональная конференция. Тезисы докладов. М.: МНТОРЭС им.А.С. Попова, 1995. С.96-97.
- 45. Левин, Е.К. Методы подавления посторонних звуков и слов при автоматическом распознавании голосовых команд в телефонии [Текст] / Е.К. Левин // Молодые ученые 2008: Материалы международной научно-технической школы-конференции «Молодые ученые науке, технологиям и профессиональному образованию», 10-13 ноября 2008г.,

- г. Москва. М.: Энергоатомиздат, 2008, часть 4. С.221 226.
- 46. Левин, Е.К. Моделирование алгоритмов сжатия речевых сигналов на основе линейного предсказания [Текст] / Е.К. Левин, А.В.Николаев, Н.В.Бондарь // Обработка сигналов в системах двусторонней телефонной связи. Четвертая межрегиональная конференция. Тезисы докладов. М.: МНТОРЭС им.А.С. Попова, 1995. С.75-77.
- 47. Левин, Е.К. Моделирование системы распознавания речевых команд [Текст] / Е.К. Левин // Тезисы докладов 7-ой межрегиональной конференции «Обработка сигналов в системах двусторонней телефонной связи» МНТОРЭС им.А.С. Попова, М., 1997. С.198-199.
- 48. Левин, Е.К. Адаптивная компенсация помех в системах связи с сигналами, подверженными селективно-частотным замираниям [Текст] / Е.К. Левин. // Методы и устройства первичной обработки сигналов в радиотехнических системах: Межвуз. сб. научн. трудов. г. Горький, Горьк. политехн. ин-т., 1988 с.76-81.
- 49. Левин, Е.К. Средства исследования помехоустойчивости систем распознавания голосовых команд в телефонии. Монография. [Текст] / Е.К. Левин Владимир: Рост, 2014. 234с.
- Левин, Е.К. Адаптивная компенсация помех с использованием обратной связи по решению [Текст] /Е.К. Левин // Электросвязь, 1998. № 3. С. 32 34.
- 51. Левин, Е.К. Адаптивная компенсация помех с обратной связью по решению [Текст] / Е.К.Левин.- М.,1995. 10с. Деп. в ЦНТИ "Информсвязь" 20.08.95, №2067-св.95.
- 52. Левин, Е.К. Адаптивная компенсация широкополосных радиопомех [Текст] / Е.К. Левин // Проектирование и технология электронных средств. 2013. №1. С. 2-6.

- 53. Левин, Е.К. Анализ измерителя основного тона речевого кодека на основе линейного предсказания [Текст] / Е.К. Левин // Перспективные технологии в средствах передачи информации. Материалы второй международной научно-технической конференции. / Владим. гос. университет Владимир: ВлГУ. т.1- 1997. С.225 227.
- 54. Левин, К.Е. Моделирование детектора голосовых команд для автоматической справочной службы медицинского учреждения [Текст] / К.Е. Левин, О.Р. Никитин // Физика и радиоэлектроника в медицине и экологии: доклады 5-ой межд. науч.-техн. конф. В 2 т. Владимир: Фирма «Рост», 2002. –С. 148-149.
- 55. Левин, Е.К. Зависимость ложного срабатывания системы распознавания голосовых команд от параметров произнесений посторонних слов. [Текст] / Е.К. Левин, П.С. Рагузин // Перспективные технологии в средствах передачи информации. Материалы докладов девятой международной научно-технической конференции. / Владим. гос. университет Владимир: ВлГУ. т.1- 2011. С.185 188.
- 56. Левин, Е.К. Использование голосового управления в телефонной справочной системе медицинского учреждения [Текст] / Е.К. Левин, О.Р. Никитин // Биомедицинская радиоэлектроника. 2000. №6. С.22-25.
- 57. Левин, Е.К. Использование программного моделирующего комплекса для исследования точности оценки огибающей спектра речевого сигнала [Текст] / Е.К. Левин, К.Е. Левин, О.Р. Никитин // Проектирование и технология электронных средств. 2004. №1. С.49-53.
- 58. Левин, Е.К. Исследование влияния аддитивного шума на точность определения параметров речевых сигналов при их распознавании [Текст] / Е.К. Левин, К.Е. Левин // Физика и радиоэлектроника в медицине и экологии Материалы докладов шестой международной научнотехнической конференции. В 2 т. Владимир, 2004. Т. 1. С. 148-150.

- 59. Левин, Е.К. Исследование влияния аддитивных помех на точность определения параметров речевых сигналов при их распознавании [Текст] Е.К. K.E. // Левин, Левин Фундаментальные проблемы приборостроения (Intermatic-2004). радиоэлектронного Материалы докладов третьей международной научно-практической конференции. В 2 T. – M., 2004. – T. 2. - C. 105-108.
- 60. Левин, Е.К. Исследование изменчивости параметров речевых команд [Текст] / Е.К. Левин // Обработка сигналов в системах двусторонней телефонной связи. Восьмая межрегиональная конференция. Тезисы докладов. М.: МНТОРЭС им.А.С. Попова, 1998. С.83 84.
- 61. Левин, Е.К. Исследование процесса предварительной обработки речевых сигналов в системах автоматического распознавания голосовых команд [Текст] / Е.К. Левин, К.Е. Левин // Перспективные технологии в средствах передачи информации. Материалы докладов четвертой международной научно-технической конференции. Владимир, 2001. С.129-130.
- 62. Левин, Е.К. Исследование точности спектральной оценки речевого сигнала для автоответчика с голосовым управлением в медицинском учреждении [Текст] / Е.К. Левин, К.Е. Левин // Физика и радиоэлектроника в медицине и экологии. Материалы докладов пятой международной научно-технической конференции. Владимир, 2002. С.146-147.
- 63. 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.
- 64. Левин, Е.К. Система определения акустических признаков речевых сигналов [Текст] / Е.К. Левин, А.В.Рыбкин // Материалы 3-ей международной конференции «Перспективные технологии в средствах

- передачи информации», ПТСПИ'99. Владимир, 1999. С. 230-231.
- 65. Левин, Е.К. Метод подавления совокупности помех от источника излучения и пассивных переизлучателей [Текст] /Е.К. Левин, А.С. Немировский // Электросвязь, 1987. № 9. С. 48 51.
- 66. Левин, Е.К. Моделирование процесса создания эталонов голосовых команд для систем автоматического распознавания речи [Текст] / Е.К. Левин, К. Е. Левин // Цифровая обработка сигналов. 2002. №2. С. 21-22.
- 67. Левин, Е.К. Моделирование системы автоматического распознавания голосовых команд [Текст] / Е.К. Левин, К.Е. Левин // Обработка сигналов в системах телефонной связи и вещания. Материалы докладов одиннадцатой межрегиональной конференции. М., МТУСИ 2002. С.63-65.
- 68. Левин, Е.К. Моделирование устройств обработки речевых сигналов в телефонии [Текст] / Е.К. Левин // Перспективные технологии в средствах передачи информации. Материалы пятой международной научнотехнической конференции. / Владим. гос. университет Владимир: ВлГУ. 2003. С.102 103.
- 69. Левин, Е.К. О тестировании систем автоматического распознавания голосовых команд [Текст] / Е.К. Левин // 8-я Международная НТК «Физика и радиоэлектроника в медицине и экологии», 2008г. Владимир. Книга 2 С.140-145.
- 70. Левин, Е.К. Оценка вероятности ложного срабатывания системы автоматического распознавания голосовых команд [Текст] / Е.К. Левин, П.С.Рагузин // Проектирование и технология электронных средств.-2009.- №4. С.57-61.
- 71. Левин, Е.К. Оценка вероятности ложного срабатывания системы

- автоматического распознавания голосовых команд [Текст] / Е.К. Левин, П.С. Рагузин, М.Ю. Татарникова // Научно-технические ведомости СПбГПУ. Информатика, телекоммуникации, управление 2011. N = 5 С. 7 13.
- 72. Левин, Е.К. Оценка помехоустойчивости автоматического распознавания голосовых команд в телефонии [Текст] / Е.К. Левин, О.Р. Никитин // Радиотехнические и телекоммуникационные системы 2011. №4 С. 75-80.
- 73. Левин, Е.К. Оценка устойчивости системы автоматического распознавания голосовых команд к произнесениям посторонних слов [Текст] / Е.К. Левин, П.С. Рагузин, // 10-я Международная НТК «Физика и радиоэлектроника в медицине и экологии». Книга 2 Владимир-Суздаль, 2012, С.273 277.
- 74. Левин, Е.К. Паразитная частотная модуляция в цифровой системе частотной автоподстройки [Текст] / Е.К. Левин // 33-я Всесоюзная научная сессия, посвященная Дню радио: Тезисы докл. М.: 1976, с. 64.
- 75. Левин, Е.К. Разработка телефонной справочной системы с голосовым управлением для медицинского учреждения [Текст] / Е.К. Левин, О.Р.Никитин // Физика и радиоэлектроника в медицине и экологии Материалы докладов четвертой международной научно-технической конференции. В 2 т. Т. 2. Владимир, 2000. С. 118-120.
- 76. Левин, Е.К. Распознавание голосовых команд в телефонной справочной системе для медицинского учреждения [Текст] / Е.К. Левин, К.Е. Левин // Физика и радиоэлектроника в медицине и экологии Материалы докладов четвертой международной научно-технической конференции. В 2 т. Т. 2. Владимир, 2000. С. 121-122.

- 77. Левин, Е.К. Система для исследования алгоритма распознавания речевых сигналов [Текст] / Е.К. Левин, К.Е. Левин // Биомедицинские технологии и радиоэлектроника. 2004. №3. С.28-32.
- 78. Левин, Е.К. Система распознавания голосовых команд [Текст] / Е.К. Левин // Обработка сигналов в системах телефонной связи и вещания. Десятая межрегиональная конференция. Тезисы докладов. М.: МНТОРЭС им.А.С. Попова, 2000. С.121 125.
- 79. Левин, Е.К. Сравнение телефонных систем автоматического распознавания голосовых команд по достоверности распознавания [Текст] / Е.К. Левин, К. Е. Левин // Перспективные технологии в средствах передачи информации. Материалы 7-ой международной научно-технической конференции. / Владим. гос. университет Владимир: РОСТ, 2007. С.70 73.
- 80. Левин, Е.К. Удаление пауз из речевого сигнала [Текст] / Е.К. Левин, К.Е. Левин // Проектирование и применение радиотехнических устройств. Тезисы докладов областной научно-технической конференции молодых специалистов и студентов. Владимир, 2000. С.3.
- 81. Левин, Е.К. Уменьшение влияния вариаций коэффициента передачи подстраиваемого генератора на динамическую точность цифровой АПЧ [Текст] / Е.К. Левин // Проблемы теории чувствительности электронных и электронномеханических устройств и систем. Республиканское совещание. Тезисы докл. Владимир, 1976. С. 72.
- 82. Левин, Е.К. Уменьшение объема звуковой базы, используемой при тестировании систем автоматического распознавания голосовых команд [Текст] / Е.К. Левин // Проектирование и технология электронных средств. 2010. №2. С. 42-44.
- 83. Левин, Е.К. Устройство адаптивной компенсации комплекса помех, образованных источником излучения и пассивными переотражателями [Текст] / Е.К. Левин. М., 1986. -12 с. Деп. в ЦНТИ "Информсвязь"

- 1.09.86, № 925 cв.
- 84. Левин, Е.К. Флуктуации коэффициентов весового суммирования в устройствах адаптивной компенсации помех [Текст] / Е.К.Левин.- М.,1987.
 12с. Деп. в ЦНТИ "Информсвязь" 20.08.87, №1152-св.87.
- 85. Левин, Е.К. Экспериментальное исследование системы автоматического распознавания голосовых команд при наличии произнесений посторонних слов [Текст] / Е.К. Левин // Перспективные технологии в средствах передачи информации. Материалы 8-ой международной НТК. Т.2. / Владим. гос. ун-т Владимир, 2009.- С.156 158.
- 86. Левин, Е.К. Экспериментальное исследование системы дикторонезависимого распознавания голосовых команд [Текст] / Е.К. Левин, К.Е. Левин // Телекоммуникационные и вычислительные системы Материалы докладов конференции. М., 2004. С.159-162.
- 87. Левин, Е.К. Экспериментальные исследования системы распознавания голосовых команд [Текст] / Е.К. Левин, К.Е. Левин, О.Р. Никитин // Проектирование и технология электронных средств. 2005. —№ 3. С.38-44.
- 88. Левин, Е.К. Эффективность компенсации помех при замирании широкополосного сигнала [Текст] / Е.К. Левин // Научные исследования института техническому и культурному прогрессу. Материалы 25-й НТК Владим. политехн. ин-та. часть І. Владимир, 1990. С.94-95.
- 89. Левин, Е.К.. Анализ быстродействующей цифровой системы частотной автоподстройки [Текст] / Е.К. Левин. М., 1978. 10 с. Деп. в ЦНТИ Информсвязь 23.04.1978, №9.
- 90. Левин, К. Е. Анализ алгоритма подавления помех, используемого в системах автоматического распознавания речи [Текст] / К. Е. Левин, Е.К. Левин, О.Р.Никитин // Методы и устройства формирования и обработки

сигналов в связи и локации -2006. №2 -C20-22.

- 91. Левин, К. Е. Аппаратное обеспечение системы прямого доступа к абонентам АТС учреждения, управляемое голосовыми командами [Текст] / К.Е. Левин, Е.К. Левин // VII международная научнотехническая конференция «Физика и радиоэлектроника в медицине и экологии ФРЭМЭ 2006» Доклады. В 2 т. Владимир, Собор, 2006. Т. 1. С. 300-302.
- 92. Левин, К. Е. Система распознавания голосовых команд для устройств компьютерной телефонии [Текст] / К.Е. Левин, Е.К. Левин, О.Р. Никитин // VII международная научно-техническая конференция «Физика и радиоэлектроника в медицине и экологии ФРЭМЭ 2006» Доклады. В 2 т. Владимир, Собор, 2006. Т. 1. С. 303-306.
- 93. Левин, К.Е. Использование фильтра Винера для обработки речевых сигналов при автоматическом распознавании голосовых команд [Текст] / К.Е. Левин, Е.К. Левин // Молодые ученые 2005. Материалы докладов третьей международной научно-технической школы-конференции. М., 2005. С. 51-53.
- 94. Левин, К.Е. Система прямого доступа к абонентам АТС учреждения, управляемая голосом [Текст] // Е.К. Левин, К.Е. Левин, О.Р.Никитин // Методы и устройства передачи и обработки информации. Межвуз. Сб. научн. тр. Вып.7/ СПб.: Гидрометеоиздат, 2006. С.223-229.
- 95. Левин. Е. К. Определение характеристик случайной последовательности результатов измерения частоты [Текст] / Е. К. Левин // Повышение эффективности и надежности РЭС: Межвуз. сб. науч. тр. Вып.6. Л.: ЛЭТИ., 1976. С.82-85.
- 96. Маркел, Дж.Д. Линейное предсказание речи [Текст] / Д.Д. Маркел, А.Х.

- Грей. М.: Связь, 1980. 308с.
- 97. Мельник, О. Речевые технологии: на пороге важных событий [Электронный ресурс] / О. Мельник // CRN/RE. №3 (77), 2013г. Режим доступа: http://www.crn.ru/numbers/spec-numbers/detail.php?ID=79646
 Дата обращения: 07.01.2014.
- 98. Монзинго, Р.А. Адаптивные антенные решетки: Введение в теорию [Текст] / Р.А. Монзинго, Т.У. Миллер; пер. с англ. М.: Радио и связь, 1986. 448 с.
- 99. Назаров, М.В. Методы цифровой обработки и передачи речевых сигналов [Текст] / М.В. Назаров, Ю.Н. Прохоров. М.: Радио и связь, 1985. 176 с.
- 100. Немировский, А.С. Борьба с замираниями при передаче аналоговых сигналов [Текст] / А.С Немировский. М.: Радио и связь, 1984. 208 с.
- 101. Новоселов, С. А. Выделение и предобработка сигналов в системах автоматического распознавания речевых команд: автореф. Дис...канд. тех. наук: 05.12.04 [Текст]/ Новоселов Сергей Александрович. Владимир, 2011. 20с.
- 102. Полушин, П.А. Избыточность сигналов в радиосвязи [Текст] / Полушин П.А., Самойлов А.Г. М.: Радиотехника. 2007. 256с.
- 103. Потапова, Р.К. Новые информационные технологии и лингвистика [Текст] / Р.К. Потапова. М.: Библиотека лингвиста, 2002. 575 с.
- 104. Потапова, Р.К. Основы речевой акустики: учебное пособие [Текст] / Р.К. Потапова, В.Г.Михайлов. М.: ИПК МГЛУ «Рема», 2012.- 494 с.
- 105. Потапова, Р.К. Речь: Коммуникация, Информатика, Кибернетика [Текст] / Р.К. Потапова. М.: УРСС, 2001. 562 с.
- 106. Рабинер, Л. Р. Цифровая обработка речевых сигналов [Текст] / Л. Р. Рабинер, Р.В. Шафер; пер. с англ. М.: Радио и связь, 1981. 496 с.

- 107. Рагузин, П.С. Программа для оценки вероятности ложного срабатывания системы распознавания голосовых команд [Текст] / П.С. Рагузин, Е.К. Левин // Перспективные технологии в средствах передачи информации: Материалы 10-ой международной научно-технической конференции. / Владим. гос. ун-т Владимир: ВлГУ, Т.1. 2013.- С.175 177.
- 108. Распознавание речи [Электронный ресурс] / Центр речевых технологий.
 Санкт-Петербург. Режим доступа:
 http://www.speechpro.ru/technologies/recognition
 Дата обращения: 07.01.2014
- 109. Репина, О.И. Искажения в телефонном тракте [Текст] / О.И. Репина. М.: Связь, 1978. 176 с.
- Ронжин, А. Л. Автоматическое распознавание русской речи [Текст] / А. Л. Ронжин, И. В. Ли // Вестник Российской Академии Наук. 2007. №2. С.133-138
- 111. Ронжин, А.Л. Система автоматического распознавания русской речи SIRIUS [Электронный ресурс] / А.Л. Ронжин., А.А. Карпов, И.В. Ли // Санкт-Петербургский институт информатики и автоматизации РАН, Санкт-Петербург, 2005 12c. Режим доступа:

 http://www.spiiras.nw.ru/speech/intas/Papers/ii.pdf Дата обращения: 11.02.2013
- 112. Сергиенко, А.Б. Цифровая обработка сигналов: учеб. пособие.- 3-изд. [Текст] / А.Б. Сергиенко. СПб.: БХВ-Петербург, 2011. 768с.-
- 113. Скляр, Б. Цифровая связь. Теоретические основы и практическое применение [Текст] / Б. Скляр; перевод с англ. 2-е изд., испр. М. : Вильямс, 2003. 1104 с.
- 114. Соболев, В.Н. Информационные технологии в синтетической телефонии: монография [Текст] / В.Н. Соболев. М.: ИРИАС, 2007.- 360 с.

- 115. Соболев, В.Н. Структурные преобразования речевого сигнала: учеб. пособие [Текст] / В.Н.Соболев. М.: МТУСИ, 2005. 188 с.
- 116. Тимохин, О.Г. Программное обеспечение лабораторной работы «Распознавание речевых команд» [Текст] / О.Г. Тимохин, Е.К. Левин // Обработка сигналов в системах двусторонней телефонной связи. Шестая межрегиональная конференция. Тезисы докладов. М.: МНТОРЭС им.А.С. Попова, 1996. С.66-68.
- 117. Тимохин, О.Г. Речевое управление персональным компьютером [Текст] / О.Г. Тимохин, Е.К. Левин // Физика и радиоэлектроника в медицине и биотехнологии. Материалы 2-ой международной научно-технической конференции. Владимир, 1996. С. 121-123.
- 118. Уидроу, Б. Адаптивная обработка сигналов [Текст] / Б. Уидроу, С. Стирнз; перевод с англ. М.: Радио и связь, 1989. 440 с.
- 119. Устройство для автоматической подстройки частоты [Текст]: а.с. № 511669: МКИ 2 Н 03 В 3/04 / Е.К. Левин (СССР) №2072577/09; заявл. 04.11.74, опубл. 25.04.76 Бюл. № 15. -2 с. : ил.
- 120. Устройство приема сигналов с двукратным разнесением [Текст]: а.с. 1092741 СССР: МКИ³ Н 04 В 7/02; Н 04 В 7/08 /Е.К. Левин, А.А. Покровский, П.А. Полушин, А.Г. Самойлов (СССР). №3534593/18-09; заявл. 06.01.83; опубл. 15.05.84 Бюл. № 18. 8 с.: ил.
- 121. Устройство приема широкополосных сигналов с двукратным разнесением [Текст] : а.с. 1286079 СССР : МКИ³ Н 04 В 7/04 /Е.К. Левин, И.М. Покровская, П.А. Полушин, А.Г. Самойлов (СССР). №3880798/24-09; заявл. 05.04.85; опубл. 15.02.87 Бюл. № 3. 14 с. : ил.
- 122. Фант, Г. Акустическая теория речеобразования [Текст] / Г. Фант. М.: Наука, 1964. 284c.
- 123. Финк, Л. М. Теория передачи дискретных сообщений [Текст] / Л. М. Финк. М.: Сов. Радио, 1970. 728с.

- 124. Фланаган, Д. Анализ, синтез и восприятие речи [Текст] / Д. Фланаган,; перевод с англ; под редакцией А.А. Пирогова. М.: Связь, 1968. 396 с.
- 125. Хижинский, Д. Системы IVR: проблемы и пути их решения [Электронный ресурс] / Д. Хижинский // Мобильные системы. 10.11.2004. Режим доступа: http://ct.forte-it.ru/info/press/79. Дата обращения: 12.06.2013.
- 126. Цзинбинь, Я. Исследование характеристик системы поиска ключевых слов на основе минимального интервала редактирования и мер доверительности [Электронный ресурс] / Я. Цзинбинь, И.Э. Хейдоров, А.В. Ткаченя // Речевые технологии. 2009. №1. С.5-14. Режим доступа: http://speechtechnology.ru/files/4-2008.pdf Дата обращения: 11.01.2014.
- 127. Цифровое устройство частотной автоподстройки [Текст]: а.с. № 1104659 СССР: МКИ³ Н 03 L 7/00 / Е.К. Левин (СССР) №3539904/18-09; заявл. 11.01.83, опубл. 23.07.84 Бюл. № 27. 8 с. : ил.
- 128. Чучупал, В.Я. Цифровая фильтрация зашумленных речевых сигналов [Электронный ресурс] / В.Я. Чучупал, А.С. Чичагов, К. А. Маковкин // Сообщения по программному обеспечению ЭВМ, вычислительный центр Российской Академии Наук, Москва, 1998 52с. Режим доступа: http://www.ccas.ru/depart/chuchu/publics/Public_98.pdf Дата обращения: 11.01.2014
- 129. Шелухин, О.И. Цифровая обработка и передача речи [Текст] / О.И. Шелухин, Н.Ф. Лукьянцев; под ред. О.И. Шелухина. М.: Радио и связь, 2000. 456с.
- 130. Шумоочистка. [Электронный ресурс] / Центр речевых технологий. Санкт-Петербург. Режим доступа:

 http://www.speechpro.ru/technologies/noiseclear#tab2 Дата обращения : 07.01.2014

- 131. Эксплуатационные нормы на электрические параметры коммутируемых каналов сети ТфОП. Утверждено приказом Госкомсвязи России от 05.04.99 №54. [Электронный ресурс]. Режим доступа: http://www.aboutphone.info/lib/law/54.html. Дата обращения: 10.04.2012.
- 132. Этапы и работы по созданию голосового меню. [Электронный ресурс] / Центр речевых технологий. Режим доступа: http://vxml.ru/wiki Дата обращения : 07.01.2014.
- 133. Adami, A. Qualcomm –ICSI OGI Features for ASR [Электронный ресурс] / A. Adami, L. Burget, S. Dupont, H. Garuadadri, F. Grezl, H. Hermansky, P. Jain, S. Kajarekar, N. Morgan, S. Sivadas // ICSLP. 2002, P. 4-7. Режим доступа:

 http://www1.icsi.berkeley.edu/ftp/global/global/pub/speech/papers/icslp02-aurora.pdf. Дата обращения: 11.03.2010.
- 134. Ajmera, J. Keyword spotting using durational entropy [Электронный ресурс] / J. Ajmera, F. Metze // ICASSP 2007, P. 973-976. Режим доступа: http://www.cs.cmu.edu/~fmetze/interACT/Publications_files/publications/ajmera.pdf/. Дата обращения: 19.02.2010
- 135. Andersson, M. An Evaluation of Noise Robustness of Commercial Speech Recognition Systems [Электронный ресурс] / M. Andersson // Master Thesis. Stockholm, 2003. 46 р. Режим доступа:

 http://www.speech.kth.se/prod/publications/files/1641.pdf. Дата обращения: 19.02.2013.
- 136. Baghdasaryan, A. G. Automatic Phoneme Recognition with Segmental Hidden Markov Models [Электронный ресурс] / A. G. Baghdasaryan // Thesis submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering. Blacksburg, Virginia. 2010.-151p. Режим доступа: http://scholar.lib.vt.edu/theses/available/etd-02082010-174617/unrestricted/Baghdasaryan AG T 2010.pdf

Дата обращения: 15.03.2011

- 137. Bernard, A. Low-Bitrate Distributed Speech Recognition for Packet-Based Wireless Communication / A. Bernard, A. Alwan // IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 10, NO. 8, 2002. – P.570-579. - Режим доступа: http://www.seas.ucla.edu/spapl/paper/bernard_sap02.pdf Дата обращения: 11.02.2013
- 138. Bilmes, J. A Review of Graphical Model Architectures for Speech Recognition [Электронный ресурс] / J.Bilmes, C. Bartels // University of Washington, Seattle Department of Electrical Engineering. - 2005. -23p. -Режим доступа: http://melodi.ee.washington.edu/~bilmes/mypubs/bilmes2005-gmarch
 - asr.pdf. Дата обращения: 19.02.2010
- 139. Broun, C. Speaker Independent Speech Recognition with Robust Out-of-Vocabulary Rejection [Электронный ресурс] / С. Broun, W. Campbell // Motorola Human Interface Laboratory, Tempe, Arizona 85284, USA. -200, -4р. . - Режим доступа: http://www.eurasip.org/Proceedings/Eusipco/Eusipco2000/SESSIONS/THUA M/OR2/CR1403.PDF. Дата обращения: 10.04.2012.
- 140. Burget, L. Distibuted Speech Recognition [Электронный ресурс] / L. Burget, P. Motlĭček, F. Grézl, P. Jain // Radioengeneering, vol. 11, No. 4, December. -2002, Р.12-16. - Режим доступа: http://radioeng.cz/fulltexts/2002/02_04_12_16.pdf Дата обращения: 16.02.2010.

- 141. Cowling, M. Non-Speech Environmental Sound Classification System for Autonomous Surveillance [Электронный ресурс] / M. Cowling // Griffith University, Faculty of Engineering and Information Technology, School of Information Technology, Australia, 2004. 143p. . Режим доступа: https://www120.secure.griffith.edu.au/rch/items/58df9554-8699-90c2-bfbf-05529caf5280/1/ Дата обращения : 11.03.2013
- 143. Dimitriadis, D. Advanced Front-end for Robust Speech Recognition in Extremely Adverse Environments [Электронный ресурс] / D. Dimitriadis, J. C. Segura, L.Garcia, A. Potamianos, P.Maragos, V. Pitsikalis // Proc. of Intern. Conf. on Speech Communication and Technology Interspeech 2007, Antwerp, Belgium, 2007. 4p. Режим доступа: cvsp.cs.ntua.gr/projects/pub/HIWIRE/HiwirePublications/DMSP_HAFE_AS
 R_Interspeech07.pdf Дата обращения: 11.03.2013
- 144. Dupont, S. Robust feature extraction and acoustic modeling at Multitel: experiments on the Aurora databases [Электронный ресурс]/ S. Dupont, C. Ris // Eurospeech.-. 2003. 4p. Режим доступа: http://citeseerx.ist.psu.edu/viewdoc/download?rep=rep1&type=pdf&doi=10.1.
 1.130.1062 Дата обращения: 16.02.2010.
- 145. Ealey, D. Harmonic tunneling: tracking non-stationary noises during speech [Электронный ресурс] / D. Ealey, H. Kelleher, D. Pearce // Eurospeech. 2001.- 4p. Режим доступа:

- http://kom.aau.dk/~kvs/papers/harmonic_tunneling.pdf.
- Дата обращения: 10.03.2013.
- 146. ETSI ES 202 050 V1.1.5 (2007-01) Speech processing, Transmission and Quality aspects (STQ); Distributed speech recognition; Advanced front-end feature extraction algorithm; Compression algorithms. [Электронный ресурс]./ ETSI Standard, 2007. 45р. Режим доступа:

 http://www.etsi.org/deliver/etsi_es/202000_202099/202050/01.01.05_60/es_2_02050v010105p.pdf. Дата обращения: 10.03.2010.
- 148. Evans, N.W.D. Computationally Efficient Noise Compensation for Robust Automatic Speech Recognition Assessed under the Aurora 2/3 Framework [Электронный ресурс] / N.W.D. Evans, J.S.Mason // ICSLP 2002. CiteSeer.IST Scientific Literature Digital Library. 2002. 4р. Режим доступа:
 - http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.19.8476&rep=rep1
 &type=pdf Дата обращения: 11.03.2010.
- 149. Evans, N.W.D. LPC-Based, Temporal-Lateral Noise Estimation Evaluated on the AURORA Corpus [Электронный ресурс] / N. W.D. Evans, J. S. Mason // SPPRA 2002. CiteSeer.IST Scientific Literature Digital Library. -2002. 5р. Режим доступа:
 - http://wendang.baidu.com/view/bd3d2f6b7e21af45b307a878.html?from=related

- Дата обращения: 10.03.2010
- 150. Fisher A. Database and Online Adaptation for Improved Speech Recognition in Car Environments [Электронный ресурс] / A. Fisher, V. Stahl // ICASSP 1999, vol.1, P. 445-448. Режим доступа:

 http://www.mirlab.org/conference_papers/International_Conference/ICASSP

 %201999/PDF/SCAN/IC991449.PDF Дата обращения: 11.04.2010
- 151. Gales, M. Acoustic Modelling for Speech Recognition: Hidden Markov Models and Beyond? [Электронный ресурс] / M. Gales // Cambridge University Engineering Department, 2009. 36p. Режим доступа: http://www.asru2009.org/uploadedimages/talk/mg_talk.pdf Дата обращения: 11.01.2014
- 152. Gales, M. Model-based techniques for Noise robust speech recognition [Электронный ресурс] / M. J. F. Gales // Dissertation. Gonville and Caius College. University of Cambridge 1995. 132р.- Режим доступа: http://mi.eng.cam.ac.uk/~mjfg/thesis.pdf Дата обращения : 12.03.2010
- 153. Gales, M. The Application of Hidden Markov Models in Speech Recognition [Электронный ресурс] / M. Gales, S. Young // Foundations and Trends in Signal Processing Vol. 1, No. 3, 2007. P.195–304. Режим доступа: http://mi.eng.cam.ac.uk/~mjfg/mjfg_NOW.pdf Дата обращения: 11.01.2014
- 154. Gong, Y. Transforming HMMS for Speaker-Independent Hands-Free Speech Recognition in the Car [Электронный ресурс] / Y. Gong, J. Godfrey // C. ICASSP. -1999, vol.1 P. 297 300 Режим доступа:

 http://www.mirlab.org/conference_papers/International_Conference/ICASSP
 %201999/PDF/AUTHOR/IC991721.PDF. Дата обращения: 10.03.2012.
- 155. Heck, M. Segmentation of Telephone Speech Based on Speech and Non-speech Models [Text] / M. Heck, C. Mohr, S. St¨uker, M. M¨uller, K. Kilgour, J. Gehring, Q. B. Nguyen, V. H. Nguyen, A. Waibel // SPECOM'2013. Faculty of Applied Sciences, University of West Bohemia

- in Plze n (Pilsen), Czech Republic, 2013. P. 286-293.
- 156. Hermansky, H. Perceptual Linear Predictive (PLP) Analysis of Speech [Электронный ресурс] / H. Hermansky //Journal of Acoust. Soc. Am. 1990. №4. Р. 1738-1752. Режим доступа: http://seed.ucsd.edu/mediawiki/images/5/5c/PLP.pdf. Дата обращения : 10.02.2011.
- 157. Hermansky, H. RASTA processing of Speech [Электронный ресурс] / H. Hermansky, N.Morgan // IEEE Transactions on speech and audio processing. Vol.2, No. 4. 1994, P. 578 589. Режим доступа: http://www.ee.columbia.edu/~dpwe/papers/HermM94-rasta.pdf. Дата обращения: 11.03.2010.
- 158. Hirsch, G. Aurora -5 Experimental Framework for the Performance Evaluation of Speech Recognition in Case of Hands-free in Noisy Environments [Электронный ресурс] / G.Hirsch // Aurora 11/07.- 2007. 7р. Режим доступа: http://aurora.hsnr.de/download/aurora5_v21.pdf. Дата обращения : 15.03.2010.
- 159. Huang, X. Spoken language Processing: A guide to theory, algorithm and system development [Text] / X. Huang, A. Acero, H.-W. Hon // Prentice Hall PTR, Englewood Cliffs, NJ 07632, 2002. 960p.
- 160. Juang, B.H.Automatic Speech Recognition A Brief History of the Technology Development [Электронный ресурс] / B.H. Juang, L. R. Rabiner // Georgia Institute of Technology, Atlanta , Rutgers University and the University of California, Santa Barbara, 2006 24p. Режим доступа: http://www.ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/354_LALI-ASRHistory-final-10-8.pdf Дата обращения : 16.03.2013
- 161. Karnjanadecha, Montri and Stephen A. Zadorian Signal Modeling for Isolated Word Recognition [Электронный ресурс] / М. Karnjanadecha, S. A. Zadorian // ICASSP 1999, vol.1, 1999. -. P. 293-296. Режим доступа:

- http://ww2.odu.edu/engr/speechlab/paper11.pdf Дата обращения: 16.03.2011
- 162. Kim, N. S. Time-Varying Noise Compensation using Multiple Kalman Filters [Электронный ресурс] / N.S.Kim // ICASSP 1999, vol.1, P. 429-432. Режим доступа:
 - http://www.mirlab.org/conference_papers/International_Conference/ICASSP
 http://www.mirlab.org/conference_papers/International_Conference/ICASSP
 http://www.mirlab.org/conference_papers/International_Conference/ICASSP
 http://www.mirlab.org/conference_papers/International_Conference/ICASSP
 http://www.mirlab.org/conference_papers/International_Conference/ICASSP
 http://www.mirlab.org/conference_papers/International_Conference/ICASSP
 http://www.mirlab.org/conference/ICASSP
 <
- 163. Lacouturel, R. Detection of ambiguous portions of signal corresponding to OOV words or misrecognized portions of input [Электронный ресурс] / R. Lacouturel, Y. Normandin // ICSLP 1996, P.2071 2074. . Режим доступа:
 - http://www.asel.udel.edu/icslp/cdrom/vol4/156/a156.pdf
 Дата обращения :
 11.03.2011
- 164. Lee, A. Recent Development of Open-Source Speech Recognition Engine Julius [Электронный ресурс] // A.Lee, T. Kawahara // Nagoya Institute of Technology, Nagoya, Aichi 466-8555, Kyoto University, Kyoto 606-8501, Japan. 2009. 7р. Режим доступа: http://www.ar.media.kyoto-u.ac.jp/EN/bib/intl/LEE-APSIPA09.pdf Дата обращения: 10.01.2014
- 165. Lee, C-H. A frame-synchronous network search algorithm for connected word recognition [Электронный ресурс] / C-H. Lee, L.R. Rabiner //IEEE Trans.Acoust., Speech, Signal Processing, vol.37, No.11, Nov., 1989, P. 1649-1658. Режим доступа:
 - http://www.ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/frame%20synchron ous%20network%20search.pdf Дата обращения : 17.03.2011.
- 166. Leonard R. G. A Database for Speaker-Independent Digit Recognition [Электронный ресурс] / R. G. Leonard // ICASSP, San Diego, California, 1984, vol. 9, P. 328-331. Режим доступа:
 - http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=1172716&url=http%3
 A%2F%2Fieeexplore.ieee.org%2 Дата обращения : 17.03.2011.

- 167. Levin, E. K. The Noise Suppression for Automatic Speech Commands Recognition [Text] / E. K Levin, K.E. Levin // SPECOM'2005. University of Patras, Patras, Greece, 2005. P. 507-509.
- 168. Levin, E. K. Using a simulation program complex for researches of speech recognition devices [Text] / E. K Levin, K.E. Levin // SPECOM'2003. Moscow State Linguistic University, Moscow, Russia, 2003. P.216-219.
- 169. Lin, H. OOV detection by joint word/phone lattice alignment [Электронный ресурс] / H. Lin, J. Bilmes, D. Vergyri, K. Kirchhoff // ASRU .— 2000, P. 478 483. Режим доступа:

 http://wenku.baidu.com/view/eb6494e8172ded630b1cb6a3.html/
 Дата обращения: 11.02.2010.
- 170. Liu, W. M. Morphological filtering of Spectrograms for automatic Speech recognition [Электронный ресурс] / W. M. Liu, V.J.R. Bastante, F. R. Rodriguez, N.W.D.Evans, , J.S.D.Mason // School of Engineering, University of Wales Swansea, UK, 2003. 4p.. -Режим доступа:

 http://galilee.swan.ac.uk/publications/W.M.Liu/iasted04.pdf Дата обращения : 11.02.2012.
- 171. Maher, A.G. Adaptive noise reduction techniques for speech recognition in telecommunications environments [Электронный ресурс] / A.G. Maher, R.W.King, J.M.Song // Speech Technology Group, Department of Electrical Engineering, The University of Sydney,1992. 6p. Режим доступа: http://www.assta.org/sst/SST-92/cache/SST-92
 SpeechAnalysisandRecognition-p2.pdf Дата обращения: 16.01.2014
- 172. Milner, B. A comparison of Front-end Configurations for Robust Speech Recognition [Электронный ресурс] / B. Milner // ICASSP 2002, P. 797-800. Режим доступа:

 http://whale.cse.nsysu.edu.tw/~jzye/A%20COMPARISON%20OF%20FRONT-END%20CONFIGURATION%20FOR%20ROBUST%20SPEECH.pdf.

 Дата обращения: 10.04.2012.

- 173. Mokbel C. Incremental Enrollment of Speech Recognizers [Электронный ресурс] / C. Mokbel, O.Collin // ICASSP 1999, vol.1, P. 453-456. Режим доступа:
 - http://www.mirlab.org/conference_papers/International_Conference/ICASSP %201999/PDF/AUTHOR/IC991468.PDF Дата обращения : 10.04.2012.
- 174. Murphy, K. Hidden Markov Model (HMM) Toolbox for Matlab [Электронный ресурс] / K. Murphy // 2005. Режим доступа: http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html Дата обращения : 10.04.2012.
- 175. Nadeu, C. Improving the robustness of the usual FBE-based ASR front-end [Электронный ресурс] / C. Nadeu, D. Macho, J. Hernando // TALP Research Center, Universitat Politècnica de Catalunya.- 2000.- 20р. Режим доступа: http://nlp.lsi.upc.edu/papers/hernando_improv.pdf Дата обращения : 11.04.2012
- 176. O'Shaughnessy, D. Towards a Robust/Fast Continuous Speech Recognition System Using a Voiced-Unvoiced Decision [Электронный ресурс] / D. O'Shaughnessy, H. Tolba // ICASSP. 1999, vol.1, C. 413-416 Режим доступа:
 - http://www.researchgate.net/publication/3793847_Towards_a_robustfast_cont inuous_speech_recognition_system_using_avoiced-unvoiced_decision. Дата обращения: 10.03.2013.
- 177. Parada, C. Contextual, Information Improves OOV Detection in Speech [Электронный ресурс] / C. Parada, M. Dredze, D. Filimonov, F. Jelinek // Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL. 2010, P. 216–224. Режим доступа: http://aclweb.org/anthology//N/N10/N10-1025.pdf. Дата обращения : 15.03.2010.

- 178. Pearce, D. The Aurora Experimental Framework for the Performance Evaluation of Speech Recognition Systems under Noisy Conditions [Электронный ресурс] / D.Pearce, G.Hirsch, // 6th International Conference on Spoken Language Processing ICSLP. 2000. 4р.- Режим доступа: http://dnt.kr.hs-niederrhein.de/papers/icslp2000_final_footer.pdf. Дата обращения: 15.03.2010.
- 179. Phonexia Acoustic Keyword Spotting [Электронный ресурс] // Phonexia, Czech Republic.- Режим доступа: http://www.phonexia.com/technologies/kws Дата обращения: 15.01.2014.
- 180. Pollak, P. Long Recording Segmentation Based on Simple Power Voice Activity Detection with Adaptive Threshold and Post-Processing [Text] / P. Pollak, J.Rajnoha // Proceedings of the 13-th International Conference "Speech and Computer" SPECOM'2009. St.Petersburg: SUAI, 2009. P.55 60.
- 181. Povey, D. The Kaldi Speech Recognition Toolkit [Электронный ресурс] / D. Povey, A. Ghoshal, G. Boulianne, L.Burget, O.Glembek, N. Goel, M. Hannemann, P. Motlı cek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, K. Vesely // IEEE 2011 Workshop on Automatic Speech Recognition and Understanding},2011. 4p. Режим доступа:

 http://publications.idiap.ch/downloads/papers/2012/Povey_ASRU2011_2011.

 pdf Дата обращения: 11.01.2014.
- 182. Rabiner, L. A model-based connected-digit recognition system using either hidden Markov models or templates [Электронный ресурс] / L.Rabiner, J.G.Wilpon, B.H.Juang. // Computer Speech & Language, vol.1, no.2, 1986. P.167-197 Режим доступа:

 http://www.sciencedirect.com/science/article/pii/S0885230886800213 Дата обращения: 17.08.2011
- 183. Rabiner, L. Continuous Training Procedure for Connected Digit Recognition [Электронный ресурс] / L. Rabiner, J. Wilpon, B. Juang // ICASSP.— 1986,

- P. 1065-1068. Режим доступа:
- http://www.ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/247_ICASSP_86.p df/ Дата обращения : 19.02.2010
- 184. Rabiner, L. Fundamentals of speech recognition [Text] / L. Rabiner, B.-H. Juang // Prentice Hall PTR, Englewood Cliffs, NJ 07632, 1993.—507p.
- 185. Rabiner, L. High performance connected digit recognition using hidden Markov models. [Электронный ресурс] /L. Rabiner, J.G.Wilpon, F.K.Soong. // IEEE Trans.Acoust., Speech, Signal Processing, vol.37, No.6, Aug., 1989, P. 1214-1225. Режим доступа: http://cronos.rutgers.edu/~lrr/Reprints/high%20performance%20connected%2 Odigit%20recognition.pdf Дата обращения: 17.08.2011
- 186. Rabiner, L. HMM clustering for connected word recognition system [Электронный ресурс] / L. Rabiner, C. Lee, B. Juang, J. Wilpon // ICASSP.— 1989, P. 405-408. Режим доступа:

 http://www.ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/hmm%20clustering
 %20for%20connected%20word%20recognition.pdf Дата обращения:
 19.02.2010
- 187. Rabiner, L. R. Isolated and Connected Word Recognition Theory and Selected Applications [Электронный ресурс] / L. R. Rabiner, S. E. Levinson // IEEE Trans. on Communications, Vol. COM-29, No. 5, 1981. Р. 621-659. Режим доступа:

 http://www.ece.ucsb.edu/Faculty/Rabiner/ece259/publications.html

 Дата
- 188. Samouelian, A. Connected digit recognition using inductive inference [Электронный ресурс] / A. Samouelian // Speech Technology Research Laboratory Department of Electrical and Computer Engineering University of Wollongong, 1997.- 7р. Режим доступа:

обращения: 19.02.2013

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.52.818&rep=rep1&t

- <u>уре=pdf</u> Дата обращения 10.03.2013
- 189. Saon, G. Large-vocabulary Continuous Speesh Recognition Systems [Электронный ресурс] / G. Saon, J.Chien // IEEE Signal Processing Magazine 2012. № 6. Р.18 33. Режим доступа: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6296522 Дата обращения 10.03.2013
- 190. Sarikaya R. Analysis of the Root-Cepstrum for Acoustic Modeling and Fast Decoding in the Speech Recognition [Электронный ресурс] / R. Sarikaya, J.Hansen // Eurospeech.-. 2001. Р. 687-690. Режим доступа:

 http://perso.telecom-paristech.fr/~chollet/Biblio/Congres/Audio/Eurospeech01/CDROM/papers/page687.pdf Дата обращения 10.12.2013.
- 191. Skowronski M. D. Increased MFCC Filter Bandwidth for Noise-Robust Phoneme Recognition [Электронный ресурс] / M. D. Skowronski, J. G. Harris // ICASSP 2002, P. 801-804. Режим доступа: http://www.cnel.ufl.edu/~markskow/papers/mfccVW.pdf Дата обращения 10.12.2013.
- 192. Stolbov, M. Speech and Crosstalk Detection for Robust Speech Recognition Using a Dual Microphone System [Text] / M. Stolbov, M. Tatarnikova // SPECOM'2013. Faculty of Applied Sciences, University of West Bohemia in Plze'n (Pilsen), Czech Republic, 2013. P. 310-318.
- 193. Tangruamsub, S. Thai Speech Keyword Spotting using Heterogeneous Acoustic Modeling [Электронный ресурс] / S. Tangruamsub, P. Punyabukkana, A. Suchato. (IEEE, 2007), pp.253 260. Режим доступа: http://www.researchgate.net/publication/4251751 Thai Speech Keyword Spotting using Heterogeneous Acoustic Modeling Дата обращения: 11.03.2011.

- 194. The Spoken Term Detection (STD) 2006 Evaluation Plan [Электронный pecypc] / The National Institute of Standards and Technology (NIST), US, 2006 13р. Режим доступа:

 http://www.itl.nist.gov/iad/mig/tests/std/2006/docs/std06-evalplan-v10.pdf

 Дата обращения: 11.03.2013.
- 195. Tsuge, S. Speaker Normalized Spectral Subband Parameters for Noise Robust Speech Recognition [Электронный ресурс] / S. Tsuge, T.Fukada, H.Singer // ICASSP 1999, vol.1, P. 285-288. Режим доступа:

 http://www.researchgate.net/publication/3793941_Speaker_normalized_spectral_subband_parameters_for_noise_robustspeech_recognition

 Дата обращения 10.12.2013.
- 196. Veprek, P. Analysis, enchancement and evaluation of five pitch determination techniques [Электронный ресурс] / P.Veprek, M.S. Scordilis // Speech Communication 37. 2002, P. 249-270. Режим доступа: http://rmcet.com/lib/E-Journals/Speech%20Communication/1-s2.0-S0167639301000176-main.pdf Дата обращения 10.12.2013
- 197. Villarrubia, L. Rejection techniques for digit recognition in telecommunication applications [Электронный ресурс] / L. Villarrubia, A. Acero // Speech Technology Group. Madrid, Spain. (IEEE ICASSP'1993), P.II-455 II-458. Режим доступа: http://research.microsoft.com/pubs/78422/1993-villarrubia-icassp.pdf Дата обращения 10.12.2013
- 198. Wilpon, J. Automatic recognition of keywords in unconstrained speech using hidden markov models [Электронный ресурс] / J. G. Wilpon, L. R. Rabiner, C.-H. Lee, E.R.Goldman // IEEE Trans.Acoust., Speech, Signal Processing .— 1990, vol.38, No 11, P. 1870-1878. Режим доступа: http://ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/automatic%20recognition%20of%20keywords.pdf Дата обращения: 19.02.2010

- 199. Wilpon, J.G. Application of hidden Markov models to automatic speech endpoint detection [Электронный ресурс] / J.G. Wilpon, L.R. Rabiner. // Comput. Speech and Language, vol.2, no.3/4, 1987. Р. 321-341. Режим доступа:
 - http://cronos.rutgers.edu/~lrr/lrr%20papers/262_Application%20of%20HMM %20to%20Endpoint%20Detection.pdf Дата обращения : 15.03.2011
- 200. Wong S. Channel and Noise Adaptation via HMM Mixture Mean Transform and Stochastic Matching [Электронный ресурс] /S. K. Wong, B. Shi // ICASSP 1999, vol.1. P. 301-304. Режим доступа:

 http://www.researchgate.net/publication/3793954 Channel and noise adaptat ion via HMM mixture mean transform and stochastic matching Дата обращения 10.12.2013.
- 201. Wu, J. A Noise-Robust ASR Front-End using Wiener Filter Constructed from MMSE Estimation of Clean Speech and Noise [Электронный ресурс] / J. Wu, J. Droppo, L. Deng, A. Acero // Automatic Speech Recognition and Understanding.— 2003, P. 321-326. Режим доступа:

 http://research.microsoft.com/pubs/63606/2003-droppo-asru.pdf Дата обращения: 19.02.2010
- 202. Yao, K. Noise Adaptive Speech Recognition In Time-Varying Noise Based On Sequential Kullback Proximal Algorithm [Электронный ресурс] / K. Yao , K. Paliwal, S. Nakamura // ICASSP, 2002,- P. 189-192. Режим доступа: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.59.1390&rep=rep1 &type=pdf. Дата обращения: 18.01.2010
- 203. Yasui, H. Noise robust speech recognition using spectral subtraction and F0 information extracted by Hough transform [Электронный ресурс] / H. Yasui, K. Shinoda, S. Furui, K. Iwano // APSIPA ASC 2009, P.631-634. Режим доступа: http://eprints.lib.hokudai.ac.jp/dspace/bitstream/2115/39775/1/TP-P1-4.pdf Дата обращения: 11.12.2013.

- 204. Yazgan, A. Hybrid language models for out of vocabulary word detection in large vocabulary conversational speech recognition [Электронный ресурс] / A. Yazgan, M. Saraclar// ICASSP'2004. P. 1-745 1-748. Режим доступа: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.163.8907&rep=rep1 &type=pdf. Дата обращения: 11.03.2010.
- 205. Young S. The HTK book (for HTK Version 3.4.1) [Электронный ресурс]. / S.Young., G. Evermann, D.Kershaw, G.Moore, J.Odell, D.Ollason, V.Valtchev, P.Woodland // Speech group, Cambridge University Engineering Department, March, 2009. 277р. Режим доступа: http://htk.eng.cam.ac.uk/ Дата обращения: 11.04.2014.
- 206. Yuk, D. Telephone Speech Recognition using Neural Networks and Hidden Markov Models [Электронный ресурс] / D. Yuk, J. Flanagan // ICASSP 1999, vol.1, P. 157-160. Режим доступа:

 http://www.researchgate.net/publication/2654329_Telephone_Speech_Recogn_ition_using_Neural_Networks_and_Hidden_Markov_Models

 Дата обращения: 10.04.2012.
- 207. Zhan, P. Speaker Normalization and Speaker Adaptation a Combination for Conversational Speech Recognition [Электронный ресурс] / P. Zhan, M.Westphal, M. Finke, A.Waibe // EuroSpeech, 1997. 4р. Режим доступа:

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.45.9787&rep=rep1
&type=pdf
Дата обращения: 10.01.2014.

Приложения П1. Акты внедрения

УТВЕРЖДАЮ Генеральный директор 900 «Телеком-Экспресс» Д.В. Витчинка Our 0/10 2011r.

АКТ ВНЕДРЕНИЯ

результатов диссертационной работы

«Разработка средств исследования и повышения помехоустойчивости систем автоматического распознавания голосовых команд в телефонии» доцента Владимирского государственного университета Левина Е.К.

Мы, нижеподписавшиеся, представители организации ООО «Телеком-Экспресс»: Директор по информационным технологиям Матаков В.А., начальник отдела технологий и контроля качества Абухович М.А составили настоящий акт в том, что при разработке системы автоматического контроля качества были использованы результаты диссертационной работы Е.К. Левина. Система используется для проведения лексико-семантического анализа в составе комплекса SmartLogger II, предназначенном для автоматизированной оценки качества предоставляемых услуг в телефонном контакт-центре.

Smart Logger II отличается низким уровнем ложных срабатываний при поиске лексико-семантических шаблонов в речевых записях.

Директор ИТ Матаков В.А. Начальник ОТиКК Абухович М.А.

(C)) UPT

Общество с ограниченной ответственностью «Центр речевых технологий» а/я 515, Санкт-Петербург, 196084
Тел. (812) 325-8848
Факс (812) 327-9297
Е-mail: info@speechpro.com http://www.speechpro.ru
ОКПО 20502206, ОГРН 1027810243295, ИНН/КПП 7805093681/783901001
10.10.2011 N 854-244

om

Ha №

УТВЕРЖДАЮ

Сентр Речевых технологий»

М.В.Хитров

« _____ » октября 2011г.

АКТ ВНЕДРЕНИЯ

результатов диссертационной работы

«Разработка средств исследования и повышения помехоустойчивости систем автоматического распознавания голосовых команд в телефонии» доцента Владимирского государственного университета Левина Е.К.

Мы, нижеподписавшиеся, представители ООО «Центр речевых технологий»: директор научно-исследовательского департамента Раев А.Н., старший научный сотрудник отдела распознавания речи Татарникова М.Ю. — составили настоящий акт в том, что результаты диссертационной работы Левина Е.К. использованы при разработке системы поиска ключевых слов VoiceDigger. Система используется для поиска задаваемого набора слов в записях телефонных разговоров, получаемых с помощью многоканальных систем записи. Применение VoiceDigger позволяет значительно ускорить процесс поиска необходимой информации.

В частности, был применен метод сравнения систем распознавания речи по достоверности распознавания, который снижает вероятность ошибки сравнения систем по результатам их тестирования. Метод основан на определении выборочного среднего и выборочного стандартного отклонения значений разности метрик, полученных при сопоставлении множества произнесений голосовой команды с акустическими моделями команд.

Директор научно-исследовательского департамента

Старший научный сотрудник

Раев А.Н.

/Татарникова М.Ю.

ООО «НПП «Акустика»»

УТВЕРЖДАЮ

196084, Санкт-Петербург, ул. Красуцкого, д.4, корп,4, пом. 103 ОКПО 80479560, ОГРН 1077847277309, ИНН / КПП 7839355416 / 783901001

> Тел.: (812) 309-52-02 e-mail: <u>it@it-akustika.ru</u>

12.10.2011 N ____ Ha № 0T Генеральный директор
ООО «НПП Акустика»
М. А. Досов
« Де» октября 2011г.

АКТ ВНЕДРЕНИЯ

результатов диссертационной работы

«Разработка средств исследования и повышения помехоустойчивости систем автоматического распознавания голосовых команд в телефонии» доцента Владимирского государственного университета Левина Е.К.

Мы, нижеподписавшиеся, представители ООО «НПП «Акустика»»: технический директор Камешков А.А., и менеджер продукта Исаев С.А. – составили настоящий акт в том, что при разработке системы голосовой навигации «Барышня» были использованы результаты диссертационной работы Е.К. Левина. Система используется для предоставления информации по телефону посредством голосового меню.

В частности, при разработке системы был применен метод оценки вероятности ложного срабатывания систем распознавания голосовых команд на произнесения посторонних слов. Данный метод обеспечивает значительное ослабление нежелательной зависимости оценки от состава произнесений слов тестовой звуковой базы, что позволяет спрогнозировать число ложных срабатываний системы на звуковом материале с составом произнесенных слов, который отличается от состава слов тестовой базы.

Метод основан на учете созвучности произнесений посторонних слов голосовым командам.

Технический директор

K

Камешков А.А.

Менеджер продукта

M

Исаев С.А.

УТВЕРЖДАЮ

Первый проректор, проректор по учебной работе Федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Владимирский косударственной университет имени Александра Григо

Глокошев

Акт

об использовании результатов диссертационной работы «Разработка средств исследования и повышения помехоустойчивости систем автоматического распознавания голосовых команд в телефонии» в учебном процессе Владимирского государственного университета

Мы, нижеподписавшиеся, заведующий кафедрой радиотехники и радиосистем д.т.н.,профессор Никитин О.Р., д.т.н., профессор Бернюков А.К., заведующая лабораториями Королева О.В. составили настоящий акт в том, что результаты диссертационной работы Левина Е.К. внедрены в учебный процесс на кафедре радиотехники.

Материалы диссертации используются в лекционном курсе и при выполнении лабораторных работ в рамках учебных дисциплин «Компьютерная телефония», и «Цифровая обработка сигналов изображения и звука», предназначенных, соответственно, для магистрантов направления «Радиотехника» и студентов специальности «Радиофизика и электроника».

Зав.кафедрой радиотехники

и радиосистем д.т.н., профессор

Д.т.н., профессор

Заведующая лабораториями

Никитин О.Р.

Бернюков А.К.

Королева О.В.

П2. Анализ влияния шума на оценку периода помехи

Проанализируем влияние шума на погрешность определения периода детерминированной составляющей u(t) помехи $\xi(t)$, которая, кроме того, содержит белый гауссов шум n(t).

$$\xi(t) = u(t) + n(t).$$
 (II2.1)

Считаем, что период детерминированной составляющей помехи равен целому числу периодов дискретизации, и погрешность за счет дискретизации отсутствует. Если период T_p определяется по аргументам максимумов оценки автокорреляционной функции помехи, то необходимо проанализировать значение случайной величины $\Delta \hat{R}(1) = \hat{R}(LB) - \hat{R}(LB+1)$, где $\hat{R}(LB)$ и $\hat{R}(LB+1)$ оценки (ковариационный метод) [106,112] автокорреляционной функции

$$\widehat{R}(LB) = \frac{1}{N - LB} \sum_{i=LB}^{N-1} \xi(i) \ \xi(i - LB)$$

$$\widehat{R}(LB + 1) = \frac{1}{N - LB - 1} \sum_{i=LB+1}^{N-1} \xi(i) \ \xi(i - LB - 1)$$
(II2.2)

Здесь B — период детерминированной составляющей помехи, выраженный в виде числа периодов T_d дискретизации; L - порядковый номер рассматриваемого максимума автокорреляционной функции; N — длительность паузы, выраженная через число периодов T_d дискретизации.

Если случайная величина $\Delta \hat{R}(1) > 0$, то при значении дискретного времени равном LB оценка автокорреляционной функции периодической составляющей помехи принимает максимальное значение, и величина периода детерминированной составляющей помехи определяется без погрешности. Если же $\Delta \hat{R}(1) < 0$, то оценка автокорреляционной функции принимает максимальное значение при значении дискретного времени равном LB+1. В этом случае величина периода детерминированной составляющей помехи определяется с погрешностью равной периоду дискретизации.

Подставляя (П2.1) в выражение (П2.2), получаем

$$\hat{R}(LB) = \frac{1}{N - LB} \sum_{i=LB}^{N-1} [u(i) u(i - LB) + n(i)u(i - LB) + u(i)n(i - LB) + n(i)u(i - LB)]$$
(II2.3)

Если учесть, что u(i-LB)=u(i) и считать, что процессы n(i) и n(i-LB), а также n(i) и u(i) — не коррелированы, то из полученного выражения следует, что $\widehat{R}(LB)$ - случайная величина с математическим ожиданием

$$\mu(LB) = \frac{1}{N - LB} \sum_{i=LB}^{N-1} u^2(i) = \frac{1}{N - LB} \sum_{i=0}^{N-1-LB} u^2(i)$$
 (II2.4)

Если N-LB>>LB, то $\mu(LB)\approx R_u(LB)=R_u(0)$, где $R_u(LB)$ — значение максимума автокорреляционной функции периодического компонента помехи. Дисперсия случайной величины $\hat{R}(LB)$

$$D(LB) = \frac{2\sigma_n^2}{(N - LB)^2} \sum_{i=LB}^{N-1} u^2(i) + \frac{\sigma_n^4}{N - LB} =$$

$$= \frac{2\sigma_n^2}{(N - LB)^2} \sum_{i=0}^{N-1-LB} u^2(i) + \frac{\sigma_n^4}{N - LB}$$
(II2.5)

Здесь $\sigma_n^2 = P_n$ — дисперсия шума. Рассмотрим теперь оценку автокорреляционной функции, когда ее аргумент отличается от аргумента максимума функции на один период дискретизации

$$\widehat{R}(LB+1) = \frac{1}{N-LB-1} \sum_{i=LB+1}^{N-1} \xi(i) \ \xi(i-LB-1). \tag{\Pi2.6}$$

Рассуждая так же, как и в предыдущем случае, получаем

$$\widehat{R}(LB+1) = \frac{1}{N-LB-1} \left[\sum_{i=LB+1}^{N-1} u(i) u(i-LB-1) + \sum_{i=LB+1}^{N-1} u(i-LB-1) n(i) + \sum_{i=LB+1}^{N-1} u(i) n(i-LB-1) + \sum_{i=LB+1}^{N-1} n(i) n(i-LB-1) \right]$$

Математическое ожидание оценки $\hat{R}(LB+1)$

$$\mu(LB+1) = \frac{1}{N-LB-1} \sum_{i=LB+1}^{N-1} u(i) u(i-LB-1) =$$

$$= \frac{1}{N-LB-1} \sum_{i=0}^{N-2-LB} u(i) u(i+1)$$
(II2.8)

Если N-1-LB>>LB, то $\mu(LB+1)\approx R_u(LB+1)=R_u(1)$. Дисперсия случайной величины $\hat{R}(LB+1)$

$$D(LB+1) = \frac{\sigma_n^2}{(N-LB-1)^2} \left[\sum_{i=LB+1}^{N-1} u^2(i-LB-1) + \sum_{i=LB+1}^{N-1} u^2(i) \right] + \frac{\sigma_n^4}{N-LB-1} = \frac{\sigma_n^2}{(N-LB-1)^2} \left[\sum_{i=0}^{N-2-LB} u^2(i) + \sum_{i=0}^{N-2-LB} u^2(i+1) \right] + \frac{\sigma_n^4}{N-LB-1}$$
(II2.9)

Математическое ожидание случайной величины $\Delta \hat{R}(1)$ равно разности математических ожиданий случайных величин: $\hat{R}(LB)$, $\hat{R}(LB+1)$

$$\mu_{\Delta} = \mu(LB) - \mu(LB + 1) =$$

$$= \frac{1}{N - LB} \sum_{i=0}^{N-1-LB} u^{2}(i) - \frac{1}{N - LB - 1} \sum_{i=0}^{N-2-LB} u(i)u(i+1) \approx$$

$$\approx R_{u}(0) - R_{u}(1)$$
(II2.10)

Приближенное равенство тем точнее выполняется, чем в большей степени соблюдается условие N-LB>>LB, и чем больше L. Рассмотрим теперь разность случайных составляющих оценок корреляционной функции

$$\Delta n = \frac{1}{N - LB} \sum_{i=LB}^{N-1} [n(i)u(i - LB) + u(i)n(i - LB) + n(i)n(i - LB)] - \frac{1}{N - LB - 1} \sum_{i=LB+1}^{N-1} [u(i - LB - 1)n(i) + u(i)n(i - LB - 1) + (\Pi 2.11) + n(i)n(i - LB - 1)] =$$

$$= \sum_{i=0}^{N-LB-2} [\frac{n(i + LB + 1)u(i + 1)}{N - LB} - \frac{n(i + LB + 1)u(i)}{N - LB} + \frac{n(LB)u(0)}{N - LB}] + \frac{n(N - LB - 1)u(N - 1)}{N - LB} + \frac{n(i)u(i)}{N - LB} - \frac{n(i)u(i + LB)}{N - LB} - \frac{n(i)n(i + LB + 1)}{N - LB}] + \frac{n(N - LB - 1)u(N - 1)}{N - LB} + \frac{n(N - LB - 1)n(N - 1)}{N - LB}$$

Из данного выражения при условии, что шум является белым, получаем выражение для дисперсии

$$D_{\Delta} = \sigma_n^2 \sum_{i=0}^{N-LB-2} \left[\left(\frac{u(i+1)}{N-LB} - \frac{u(i)}{N-LB-1} \right)^2 + \left(\frac{u(i)}{N-LB} - \frac{u(i+1)}{N-LB-1} \right) \right]^2 +$$

$$+ \sigma_n^2 \frac{u^2(N-1) + u^2(0)}{(N-LB)^2} + 2\sigma_n^4 \frac{(N-LB)^2 - (N-LB) + 1}{(N-LB)^2 * (N-LB-1)}$$
(II2.12)

ПЗ. Результаты экспериментов по оценке вероятности ложного срабатывания

ПЗ.1. Границы доверительных интервалов оценок вероятности по результатам тестирования системы (разработка ЦРТ) базами 3 и 4

Таблица ПЗ.1.1 (База 3)

N	1	2	3	4	5	6	7	8	9	10	11
слово/	была	была	была	была	говорила	говорила	даже	даже	даже	даже	меня
структура	010	01010	020	02010	50	060	010	01010	020	030	010
p1	0.0055	0.1017	0.1028	0.4229	0.6762	0.2531	0.0086	0.0289	0.029	0.1794	0.0023
p2	0.0073	0.1336	0.1491	0.6627	0.9229	0.4348	0.011	0.0543	0.0447	0.3369	0.0034
pm	0.0064	0.1176	0.1259	0.5428	0.7996	0.344	0.0098	0.0416	0.0369	0.2582	0.0028
err	0.1427	0.1356	0.1841	0.2208	0.1543	0.2641	0.121	0.3045	0.2139	0.305	0.1803

Таблица ПЗ.1.1 (продолжение)

N	12	13	14	15	16	17	18	19	20	21	22	23
слово/	надо	надо	надо	ничего	ничего	очень	очень	подумал	тебе	тебе	тебе	тебе
структура	010	0110	020	010	030	010	030	020010	010	020	02010	030
p1	0.0025	0.0169	0.0238	0.0014	0.077	0.0017	0.3241	0.0829	0.0028	0.0789	0.4298	0.5562
p2	0.0035	0.0265	0.0388	0.0022	0.1238	0.0026	0.4772	0.136	0.0041	0.1165	0.589	0.7323
pm	0.003	0.0217	0.0313	0.0018	0.1004	0.0022	0.4006	0.1094	0.0034	0.0977	0.5094	0.6443
err	0.1685	0.2214	0.2388	0.2054	0.233	0.2157	0.1911	0.243	0.1935	0.1926	0.1563	0.1367

Таблица ПЗ.1.2 (База 4)

N	1	2	3	4	5	6	7	8	9	10	11
слово/	была	была	была	была	говорила	говорила	даже	даже	даже	даже	меня
структура	010	01010	020	02010	050	060	010	01010	020	030	010
p1	0.0063	0.1124	0.1016	0.5583	0.7171	0.4806	0.0054	0.044	0.0458	0.2153	0.0016
p2	0.0087	0.1578	0.1886	0.8034	0.9342	0.6517	0.0079	0.081	0.0703	0.3858	0.0028
pm	0.0075	0.1351	0.1451	0.6809	0.8257	0.5661	0.0066	0.0625	0.0581	0.3005	0.0022
err	0.1631	0.1682	0.2996	0.18	0.1315	0.1512	0.1868	0.2962	$0.2\overline{116}$	0.2837	0.2597

Таблица П3.1.2 (продолжение)

N	12	13	14	15	16	17	18	19	20	21	22	23
слово/	надо	надо	надо	ничего	ничего	очень	очень	подумал	тебе	тебе	тебе	тебе
структура	010	0110	020	010	030	010	030	020010	010	020	02010	030
p1	0.0033	0.0139	0.0348	0.0019	0.1163	0.002	0.1901	0.1066	0.0035	0.0446	0.3091	0.5175
p2	0.0049	0.0283	0.0561	0.0032	0.1964	0.0034	0.3213	0.1888	0.0055	0.0804	0.5205	0.6901
pm	0.0041	0.0211	0.0455	0.0025	0.1564	0.0027	0.2557	0.1477	0.0045	0.0625	0.4148	0.6038
err	0.1908	0.3393	0.2346	0.263	0.2563	0.2596	0.2567	0.2784	0.2233	0.2864	0.2548	0.143

Таблица ПЗ.1.3

N	W	e	S	d	Sum-i-o	Sum-g-o	р1-о	р2-о	pm-o	err-o
1	4	0	0	1	611	146549	0.0039	0.0045	0.0042	0.0663
2	4	0	0	2	406	15807	0.0237	0.0278	0.0258	0.0804
3	4	0	2	0	367	8830	0.0382	0.0452	0.0417	0.0838
4	4	0	2	1	150	676	0.1967	0.2493	0.223	0.1178
5	4	1	2	0	158	468	0.3027	0.3744	0.3385	0.1059
6	6	0	0	1	64	44482	0.0012	0.0018	0.0015	0.2023
7	6	0	0	2	90	18393	0.0041	0.0058	0.005	0.1711
8	6	0	2	0	36	6050	0.0045	0.0078	0.0062	0.2659
9	6	1	2	0	54	836	0.052	0.08	0.066	0.2126
10	7	0	0	1	22	83650	0.0002	0.0004	0.0003	0.3354
11	7	0	2	1	45	521	0.0682	0.1088	0.0885	0.2294
12	7	4	2	0	84	117	0.6452	0.7809	0.713	0.0951
13	8	3	2	0	22	26	0.698	0.929	0.8135	0.142
14	8	4	2	0	25	73	0.2582	0.438	0.3481	0.2582

Таблица ПЗ.1.4

N	w	e	S	d	Sum-i-r	Sum-g-r	p1-r	p2-r	pm-r	err-r
1	4	0	0	1	374	87032	0.0039	0.0047	0.0043	0.0846
2	4	0	0	2	202	7764	0.0232	0.0292	0.0262	0.1137
3	4	0	2	0	208	4579	0.0406	0.0508	0.0457	0.1109
4	4	0	2	1	76	286	0.2251	0.3108	0.2679	0.1598
5	4	1	2	0	120	357	0.2964	0.3784	0.3374	0.1215
6	6	0	0	1	40	32221	0.001	0.0016	0.0013	0.2536
7	6	0	0	2	39	8599	0.0035	0.0059	0.0047	0.2562
8	6	0	2	0	22	4050	0.0038	0.0077	0.0058	0.3346
9	6	1	2	0	37	276	0.1039	0.1714	0.1376	0.2453
10	7	0	0	1	21	61311	0.0002	0.0005	0.0004	0.3426
11	7	0	2	1	38	435	0.0676	0.1123	0.0899	0.2486
12	7	4	2	0	39	48	0.7041	0.8875	0.7958	0.1152
13	8	3	2	0	24	28	0.7171	0.9342	0.8257	0.1315
14	8	4	2	0	50	88	0.4806	0.6517	0.5661	0.1512

Здесь pm=0.5(p1+p2) — середина доверительного интервала; err = (p1-p2)/ (p1+p2) — отношение половины ширины доверительного интервала к его среднему значению; p1, p2 — границы доверительных интервалов; буква «о» соответствует базе 3, а буква «r» - базе 4. Результаты представлены для числа i > 19 ложных срабатываний.

ПЗ.2. Относительная частота ложного срабатывания (FAR)по результатам тестирования системы (разработка фирмы Phonexia) базами 1 и 2

Таблица ПЗ.2.1

]	База	1							
Слово	Структура помехи	W	a	e	S	d	e/w	s/w	d/w	i	FAR
Москва	040	6	4	2	2	0	0,333	0,333	0	6	1
программа	070	8	7	5	2	0	0,625	0,25	0	8	1
свобода	060	7	6	4	2	0	0,571	0,286	0	3	1
Россия	040	6	4	2	2	0	0,333	0,333	0	7	0,88
государство	090	11	9	7	2	0	0,636	0,182	0	6	0,86
Россия	050	6	5	3	2	0	0,5	0,333	0	10	0,77
история	060	8	6	4	2	0	0,5	0,25	0	3	0,75
Россия	0200010	6	3	0	2	1	0	0,333	0,167	4	0,21
город	020	5	2	0	2	0	0	0,4	0	3	0,034
человек	020	6	2	0	2	0	0	0,333	0	3	0,018
]	База	2							
автомобилистов	0C0	14	12	10	2	0	0,714	0,143	0	56	0,95
автомобилисты	0C0	13	12	10	2	0	0,769	0,154	0	36	0,92
автомобилисту	0C0	13	12	10	2	0	0,769	0,154	0	55	0,92
водителю	070	8	7	5	2	0	0,625	0,25	0	32	0,76
водителей	070	9	7	5	2	0	0,556	0,222	0	25	0,59
парковок	070	8	7	5	2	0	0,625	0,25	0	23	0,58
парковок	0102010	8	4	0	2	2	0	0,25	0,25	11	0,52
парковок	050	8	5	3	2	0	0,375	0,25	0	12	0,3
парковки	050	8	5	3	2	0	0,375	0,25	0	9	0,15

В Таблице П3.2.1 использованы обозначения для чисел больших девяти: A-10, B-11, C-12, D-14 и т.д. Величина FAR определялась, как отношение i/g.

Таблица ПЗ.2.2

			База 1			
w	e	S	d	i	g	i/g
5	0	2	0	5	637	7,85E-03
5	1	2	0	5	20	0,25
6	0	0	2	4	5147	7,77E-04
6	0	0	3	5	337	1,48E-02
6	0	2	0	6	2037	2,94E-03
6	0	2	1	7	384	1,82E-02
6	1	2	0	5	215	2,32E-02
6	2	2	0	20	51	0,39
6	3	2	0	15	19	0,79
7	4	2	0	7	8	0,88

8	5	2	0	10	10	1
11	7	2	0	6	7	0,86
			База 2			
8	0	2	2	11	21	0,52
8	3	2	0	21	100	0,21
8	5	2	0	55	82	0,67
9	5	2	0	25	42	0,60
13	10	2	0	91	99	0,92
14	10	2	0	56	59	0,94

В Таблице ПЗ.2.2 представлены точечные оценки (i/g) вероятности ложной тревоги в зависимости от значений набора параметров {w, e, s, d}. В обеих таблицах представлены данные, которые соответствуют числу ложных срабатываний равному или большему трех.

Таблица ПЗ.2.3

q	Кол-во «созвучных» фрагментов	Кол-во ложных тревог	Диапазон q	Кол-во «созвучных» фрагментов в диапазоне q	Кол-во ложных тревог в диапазоне
			База	1	
0.000	-	1			
0.125	937	1	0.125	6788	36
0.143	2034	2	0.375		
0.167	996	2			
0.200	2146	4			
0.250	233	3			
0.286	383	8			
0.333	268	10			
0.375	24	6			
0.400	277	10	0.4	432	40
0.429	15	3	0.571		
0.455	1	1			
0.500	127	18			
0.556	1	1			
0.571	11	7			
0.600	61	21	0.6	116	60
0.625	3	3	0.8		
0.667	32	20			
0.714	5	4			
0.750	1	1			
0.800	14	11			
0.818	7	6	0.818	52	45
0.833	22	18	1		

0.857	12	10			
0.875	9	9			
0.889	1	1			
1.000	1	1			
			База 2		
0.923	101	91	0.857	243	202
0.875	82	55	0.923		
0.857	60	56			
0.778	42	25	0.778	183	59
0.667	20	2	0.625		
0.625	121	32			

Здесь q=a/w

П4. Листинги т-функций

П4.1. Моделирование адаптивного компенсатора помех

function

```
[snr,W,N_out,inputN1,Rn]=compens10(B,N1,N2,acp_tau_1,acp_tau_2,Fs,t_end,T)
h1 = dfilt.delay(round(Fs*acp_tau_1(1)));
N1 acp tau1=filter(h1,N1);
h2 = dfilt.delay(round(Fs*acp_tau_1(2)));
N1 acp tau2=filter(h2,N1);
h3 = dfilt.delay(round(Fs*acp_tau_2(1)));
N2_acp_tau1=filter(h3,N2);
h4 = dfilt.delay(round(Fs*acp_tau_2(2)));
N2_acp_tau2=filter(h4,N2);
inputN0=[N1_acp_tau1',N1_acp_tau2',N2_acp_tau1',N2_acp_tau2'];
x=t_end-acp_tau_1(2)-acp_tau_2(2)-T;
k=floor(x/T);
n_T=Fs*k*T;
inputN1=inputN0(end-n_T+1:end,:);
[a,b]=size(inputN1);
Rn=(1/a)*(inputN1'*inputN1);
mu=1e6;
E=diag(ones(4,1));
R_{obr}=inv(E+mu*Rn);
W=R obr*B;
N_out=real(inputN1*W);
snr=(W'*W)/var(N out);
function [N1,N2,s]=pomexa_sim12(t_end,Fs,f0,T,bw,bwr,tau1,tau2,v1,v2)
t=0:1/Fs:t end;
d=0:T:t_{end};
s1=pulstran(t,d,'gauspuls',f0,bw,bwr);
s=hilbert(s1);
L1=length(tau1);
L2=length(tau2);
Ls=length(s);
N01=zeros(Ls,L1);
N02=zeros(Ls,L2);
  for k1=1:L1
   h1=dfilt.delay(round(Fs*tau1(k1)));
   N01(1:Ls,k1)=filter(h1,s)';
  end
N1=v1*N01';
```

```
for k2=1:L2
   h2 = dfilt.delay(round(Fs*tau2(k2)));
   N02(1:Ls,k2)=filter(h2,s)';
  end
N2=v2*N02';
function [s1,s2]=sig_psk7(Fs,f0,Td,psp)
%N=100:
\%a=randint(1,N,2);
hmod=modem.pskmod('M',2);
FsFd=round(Td*Fs);
c=repmat(psp,FsFd,1);
c1=c(:);
y=real(modulate(hmod,c1));
t=(1:length(y))/Fs;
s01=y.*cos(2*pi*f0*t)';
h1 = dfilt.delay(FsFd);
s02=filter(h1,s01);
s1=s01(FsFd:end);
s2=s02(FsFd:end);
function
[s1_out,s2_out,inputS1,Rs,Ps_out]=sig_out4(s1,s2,Fs,acp_tau_1,acp_tau_2,W)
h1 = dfilt.delay(round(Fs*acp_tau_1(1)));
s1 acp tau1=filter(h1,s1);
h2 = dfilt.delay(round(Fs*acp_tau_1(2)));
s1 acp tau2=filter(h2,s1);
h3 = dfilt.delay(round(Fs*acp_tau_2(1)));
s2 acp tau1=filter(h3,s2);
h4 = dfilt.delay(round(Fs*acp_tau_2(2)));
s2 acp tau2=filter(h4,s2);
inputS0=[s1_acp_tau1,s1_acp_tau2,s2_acp_tau1,s2_acp_tau2];
x=length(s1)-round(Fs*max(acp_tau_1(2),acp_tau_2(2)));
inputS1=inputS0(end-x:end,:);
s1 out=inputS1(:,1)*W(1)+inputS1(:,2)*W(2);
s2_{out}=inputS1(:,3)*W(3)+inputS1(:,4)*W(4);
[a,b]=size(inputS1);
Rs=(1/a)*(inputS1'*inputS1);
Ps out=W'*Rs*W;
```

П4.2. Определение вероятности ошибки при сравнении САРГК

```
function [a,alpha,p]=StatParamShort(n,mu1,sko,porog)
%генерируется п норм.распр.чисел г
%с заданными мат.ож.mu1 и ско sko.
%Если r(k)<porog, то фиксир. появл-е события.
%Определ-ся относит. частота р появл-я события,
%среднее знач-е а для r, ско и alpha=a/ско.
m=0;
r=normrnd(mu1,sko,1,n);
for k=1:n
if r(k)<porog
m=m+1;
end
end
p=m/n;
sigma=std(r);
a=mean(r);
alpha=a/sigma;
function [err_a,err_alpha,err_p,delta_s]=ErrA_P_Alpha2(mu,mu1,sko,sko1,n,M)
%Генерир. М раз две норм. распр. выборки объемом п с заданн.
%mu,mu1,sko,sko1. Для каждой выб-ки определ. сренее a, ско, alpha и
% относит. частота р. Определ относит. частоты событий:
%a1-a<0,alpha1-alpha<0,p-p1<0. Кроме того, формируется матрица
%разностей a1-a,alpha1-alpha,p-p1.
porog=0;
delta_s=zeros(M,3);
err a=0;
err_alpha=0;
err_p=0;
for m=1:M
 [a,alpha,p]=StatParamShort(n,mu,sko,porog);
 [a1,alpha1,p1]=StatParamShort(n,mu1,sko1,porog);
 if a1-a<0 err a=err a+1;
 end
 if alpha1-alpha<0 err_alpha=err_alpha+1;
 if p-p1<0 err_p=err_p+1;
 end
delta_s(m,:)=[a1-a,alpha1-alpha,p1-p];
err_a=err_a/M;
```

```
err_alpha=err_alpha/M;
err_p=err_p/M;
function p1 = VerOshibki(alpha)
p1=0.5-0.5*erf(alpha./sqrt(2));
function [VerOshResh,p1,p2] = OshResh(mu1,sigma1,mu2,sigma2,n)
%Определяются вер-сть ош.реш-я и вер-сти ошибок сравниваемых САР по
заданным
%объему тестирующей выб-ки и оценкам мат.ожид. и СКО
alpha1=mu1/sigma1;
alpha2=mu2/sigma2;
p1=VerOshibki(alpha1);
p2=VerOshibki(alpha2);
alphadelta=sqrt(n)*abs(alpha2-alpha1)./sqrt(2+0.5*alpha1.^2+0.5*alpha2.^2);
VerOshResh=VerOshibki(alphadelta);
function
[err_alpha_n,err_p_n]=VerOshibResh(mu,mu1,sko,sko1,n_min,n_max,delta_n,M)
N=floor((n_max-n_min)/delta_n);
err_alpha_n=zeros(N,1);
err_p_n=zeros(N,1);
cnt=0:
for n=n_min:delta_n:n_max
 cnt=cnt+1;
 [err_a,err_alpha,err_p,delta_s]=ErrA_P_Alpha2(mu,mu1,sko,sko1,n,M);
 err alpha n(cnt,1)=err alpha;
  err_p_n(cnt,1)=err_p;
end
      П4.3. Сравнение по помехоустойчивости результатов распознавания команд
function outputData = drawAlphaToSNRGraph (AllModels,noiseType)
  global countOfFiles;
  global countOfCoefficients;
  global wavsRootCatalog;
  NoiseFrequency = 2 * pi * 1500;
  definedSNRs = 1:2:50;
commandsToCheck = \{ (0 \le 5', (1 \le 5', (25', (3 \le 5', (4 \le 5', (5 \le 5', (6 \le 5', (7 \le 5', (9 \le 5')))) \} \}
```

commandColumns = [1,2,3,4,5,6,7,8,9,10];

```
commandsColors = \{'-+b', '-og', '-*r', '-xk', '-sc', '-dm', '--^b', '--vg', '-->r', '--< k'\};
  for command = 1:length(commandColumns)
     % Считываем файлы
    dataOfReadSignal = cell (1, countOfFiles);
    melCepstralCoeffsWithNoise = cell (1, countOfFiles);
    SampleRate = 0;
    graphDiscretes = zeros (1,length (definedSNRs));
    nameTemplate = '%replace%.wav';
    cd (commandsToCheck{command});
    for i=1:countOfFiles
    [WaveData,SampleRate] = wavread (strrep(nameTemplate,'%replace%',int2str(i)));
       dataOfReadSignal{i} = WaveData;
    for currentSNR = 1:length(definedSNRs)
       for file = 1:countOfFiles;
         switch noiseType
           case 'sine'
              SignalWithNoiseData = mixSignalWithNoise (dataOfReadSignal{file},
SampleRate, NoiseFrequency, definedSNRs(currentSNR)); % считываем очередной
файл
           case 'white noise'
              SignalWithNoiseData = mixSignalWithWhiteNoise
(dataOfReadSignal{file}, definedSNRs(currentSNR) );
           otherwise
              disp ('Тип шума не выставлен или неверен');
         end
         K = melcepst(SignalWithNoiseData,SampleRate); %вычисляем MFCC для
текущего файла
         K = K(:,1:countOfCoefficients);
         melCepstralCoeffsWithNoise{file} = K'; % заполнение массива ячеек
данными К
       end
       logLikehoodMatrix = makeComparsionTable (melCepstralCoeffsWithNoise,
AllModels);
% Рассчитываем ошибку
       [er,alpha,mu,sko,h,otnos_chas,rasch_er,raznost] = calculateError
(commandColumns(command),countOfFiles,logLikehoodMatrix);
       graphDiscretes(currentSNR) = alpha;
```

```
alpha = 0;
end
outputData(command,:) = graphDiscretes;
plot(definedSNRs, graphDiscretes, commandsColors{command});
% Удерживаем первый график
if (command == 1)
hold;
end
cd (wavsRootCatalog);
end
outputData(command+1,:) = definedSNRs;

title ('Альфа от сигнал/помеха');
legend ('Ноль','Один','Два','Три','Четыре','Пять','Шесть','Семь','Восемь','Девять');
hold off;
end
```

Обозначения переменных

countOfFiles — глобальная переменная, содержащая количество анализируемых файлов.

WaveData — значения звукового сигнала.

SampleRate — частота дискретизации анализируемого звукового сигнала.

DataOfReadSignal — переменная, хранящая значения звукового сигнала всех анализируемых файлов.

noiseType – тип шума, с которым будет смешан звуковой синал исследуемой ГК.

SignalWithNoiseData — значения огибающей сигнала смешанного с шумом. definedSNRs — массив дискретных значений сигнал/шум, для которых будет проводиться анализ.

currentSNR - номер значения сигнал/шум из definedSNRs, для которого проводится анализ в данный момент.

logLikehoodMatrix — матрица результатов сопоставления звуковых файлов исследуемой ГК с моделями САРГК.

alpha - критерий достоверности.

commandColumns – массив с номерами столбцов исследуемых команд в logLikehoodMatrix

command — номер команды, для которой строится график.

function [err,alpha,mu,sko,h,otnos_chast,rasch_err,raznost] = calculateError (num_model,num_file,Tablica_sravn)

- % Функция для вычисления ошибок
- % функция обеспечивает сопоставление произнесений команды с заданным

```
% номером (num model) со всеми моделями, входящими в состав словаря команд
% num_model - номер модели (команды), произнесения которой предъявляются
ДЛЯ
% распознавания
% num_file - число произнесений команды с заданным номером
% Tablica_sravn - массив ячеек, в котором хранятся данные сопоставления
% указанных произнесений со всеми моделями команд из словаря команд
% err - число ошибок распознавания
% alpha - критерий достоверности распознавания
% otnos chast - относительная частота ошибок распознавания
% rasch_err - вероятность ошибки, рассчитанная по критерию
% достоверности распознавания
% raznost - массив разностей между метриками сопоставления произнесений
% со своей моделью и максимально похожей на нее чужой моделью
A=Tablica sravn; % выделяем из массива ячеек
% в отдельный массив строки с числовыми данными
A voiceComand=A(:,num model); % выделяем в отдельный массив столбец
% с результатми сопоставления модели команды (с заданным номером) с
% произнесениями, которые соответствуют этой команде
A ost=A;
% оставляем столбцы с остальными моделями
A_ost(:,num_model)=[];
% определяем результаты сопоставления произнесений с чужими моделями,
% которые макимально похожи на модель с заданным номером
[m,n]=size(A);
if n>2
ost_max=max(A_ost');
else
ost_max= A_ost';
% определяем разности между метриками сопоставления произнесений
% со своей моделью и максимально похожей на нее чужой моделью
raznost=A_voiceComand-ost_max';
%--- Добавили отключатель графиков
if 0
  figure
  hist(raznost);
end
h=hist(raznost);
% определяем критерий достоверности распознавания
mu=mean(raznost);
sko=std(raznost);
```

```
alpha=mu/sko;

% определяем число ошибок распознавания err=0;
for k=1:num_file
if raznost(k)<0
err=err+1;
end
end

% определяем относительную частоту ошибок распознавания otnos_chast=err/num_file;
% рассчитываем вероятность ошибки rasch_err=0.5-0.5*(erf(alpha/sqrt(2)));
end
```