Министерство образования и науки РФ

ФГБОУ ВПО «Алтайский государственный университет»

На правах рукописи

Малинин Петр Владимирович

ПРИМЕНЕНИЕ ПРОЕКЦИОННЫХ МЕТОДОВ АНАЛИЗА МНОГОМЕРНЫХ ДАННЫХ К ГОЛОСОВОЙ ИДЕНТИФИКАЦИИ ЛИЧНОСТИ

Специальность 05.13.19 – Методы и системы защиты информации, информационная безопасность

Диссертация на соискание ученой степени
кандидата технических наук

Научный руководитель:
доктор физико-математических наук,
профессор, В.В. Поляков

Барнаул – 2014

СОДЕРЖАНИЕ

ВВЕДЕНИЕ 3

ГЛАВА 1. ПРОБЛЕМЫ ИДЕНТИФИКАЦИИ ЛИЧНОСТИ ПО ГОЛОСУ 15

1.1 Голосовая идентификация личности в задачах защиты информации 15

1.2 Методы обработки, анализа и классификации данных при голосовой идентификации 27

1.3 Постановка задачи 35

ГЛАВА 2. ВЫБОР ИНФОРМАТИВНЫХ ГОЛОСОВЫХ ПРИЗНАКОВ ДЛЯ ПРОЕКЦИОННЫХ МЕТОДОВ АНАЛИЗА 40

2.1 Методы главных компонент и проекций на латентные структуры 40

2.2 Получение исходных голосовых характеристик 47

2.3 Выбор информативных голосовых признаков 54

ГЛАВА 3. АНАЛИЗ И КЛАССИФИКАЦИЯ ГОЛОСОВЫХ ДАННЫХ 63

3.1 Разработка алгоритмов голосовой идентификации дикторов 63

3.2 Анализ голосовых сигналов на основе метода ближайших соседей 69

3.3 Анализ голосовых сигналов на основе иерархического подхода 78

ГЛАВА 4. ВЛИЯНИЕ ПОМЕХ НА ИДЕНТИФИКАЦИЮ ЛИЧНОСТИ ПО ГОЛОСУ 84

4.1 Влияние искажений голосового сигнала на идентификацию 84

4.2 Влияние внешнего акустического шума 90

4.3 Оценка ошибок идентификации при искажении и наложении акустического шума 96

ЗАКЛЮЧЕНИЕ 108

СПИСОК ЛИТЕРАТУРЫ 110

ВВЕДЕНИЕ

Актуальность темы исследования.

Влияние информационной сферы на развитие современного общества непрерывно возрастает. В связи с этим обеспечение информационной безопасности становится одним из приоритетов национальной политики государства [1]. Содержание проблем, объединяемых понятием «информационная безопасность», в последний период времени определяется прежде всего быстрым распространением новых информационных технологий. Эти технологии, приведшие к проникновению компьютеризации во все области человеческой деятельности, увеличивают зависимость от информационных систем и услуг и создают новые виды угроз для интересов отдельной личности, предприятий и организаций, общества в целом. Уязвимость частных лиц, организаций и государства по отношению к угрозам информационной безопасности особенно возрастает при использовании информационных сетей – как общего пользования, так и корпоративных. Этому способствует также расширяющаяся тенденция к распределенной обработке данных, связанная с использованием дистанционного режима и телекоммуникационных технологий (в частности, расширяется область деятельности сотрудников и привлеченных лиц вне соответствующей организации). Все большие масштабы принимают криминальные направления компьютерной деятельности. Согласно унификации Комитета министров Европейского Совета, к этим направлениям можно отнести компьютерное мошенничество, несанкционированный доступ к информации, подделку компьютерной информации, несанкционированный перехват данных и другие виды преступных действий [2]. В связи с этим важнейшей задачей становится создание и применение новых эффективных методов и средств защиты информации.

Развитие новых методов и средств обеспечения информационной безопасности призвано прежде всего предотвратить угрозы доступа к информационным ресурсам посторонних лиц, не имеющих доступа. Для решения этой задачи необходимо наличие идентификаторов и создание процедур идентификации для всех пользователей [3]. Современные идентификация и аутентификация включают в себя различные системы и способы биометрической идентификации личности. Развитие систем идентификации личности, основанных на биометрических измерениях, связано с целым комплексом преимуществ: такие системы более надежны поскольку биометрические показатели сложнее подделать; современные микропроцессорная техника делает биометрические методы (неотъемлемые биометрические идентификаторы) более удобными по сравнению с обычными методами идентификации (электронными идентификаторами); наконец, они значительно проще подвергаются автоматизации измерений.

Методы биометрической идентификации обычно разделяют на две группы: физиологические и поведенческие (учитывающие подсознательные действия человека). К физиологическим методам идентификации относится использование таких характеристик, как отпечатки пальцев рук или ладоней, радужной оболочки или сетчатки глаза, 2-мерных и 3-мерных изображений лица и т.д. [4–7]. Та или иная конкретная биометрическая технология может обладать определенными преимуществами в зависимости от различных конкретных условий и решаемых задач. Одной из наиболее распространенных биометрических характеристик человека является его голос, обладающий набором индивидуальных особенностей, относительно легко поддающихся измерению (например, частотный спектр голосового сигнала). К достоинствам голосовой идентификации относятся также удобство применения и использования, достаточно невысокая стоимость устройств, применяемых для идентификации (например, микрофонов).

Возможности идентификации личности по голосовым данным захватывают весьма широкий спектр задач, что выделяет их среди других биометрических систем. Прежде всего, голосовая идентификация достаточно давно и широко используется в различных системах разграничения доступа к физическим объектам и информационным ресурсам [8,9]. Голосовая идентификация является частью отдельного научного направления – теории речеобразования [10,11]. Перспективным представляется ее новое применение в системах, основанных на телекоммуникационных каналах связи. В качестве примера, в мобильной связи с помощью голоса можно осуществлять управление услугами, причем внедрение голосовой идентификации способствует защите от мошенничества [12,13].

Большая роль голосовой идентификации обусловлена также решением такой важной задачи, как защита речевой информации. Эта идентификации применяется при создании новых технических средств и программно-аппаратных устройств защиты речевой информации [14], в частности, от утечки по акустическим, виброакустическим и другим каналам.

Особое место идентификация личности по голосу занимает при расследовании преступлений, в том числе в сфере компьютерной информации [15–17], и при формировании доказательной базы такого расследования. В этих случаях часто возникает необходимость проведения идентификации неизвестной голосовой записи. Проведение голосовой идентификации – важная практическая задача при поиске подозреваемого по записи голоса в телекоммуникационных каналах связи. Определение таких характеристик по голосу диктора, как пол, возраст, национальность, диалект, эмоциональную окраску речи, также важны в области криминалистики и антитеррористических действий [18–20]. Результаты идентификации важны при проведении фоноскопических экспертиз, при осуществлении экспертного

криминалистического исследования на основе теории криминалистической идентификации [21,22].

Существенный интерес представляет развитие методов голосовой идентификации для смежных направлений, именно, для новых речевых технологий, связанных с распознаванием устной речи [23], управлением компьютерными системами с помощью голосовых команд [24].

Отдельную важную составляющую голосовой идентификации личности представляет формирование баз голосовых данных. Роль таких баз данных существенно возросла в связи с развитием новых технических средств обработки и хранения голосовой информации [25]. Базы голосовых данных необходимы, в частности, при апробации новых методов оценки защищенности речевой информации, а также при проверке надежности технических устройств ее защиты.

Нужно отметить также, что при выполнении особо важных работ необходимо использовать сочетание различных технологий и методов, обеспечивающее наиболее надежную идентификацию и аутентификацию. Другими словами, в этих случаях целесообразно сочетать биометрические голосовые методы идентификации со специальными физическими устройствами доступа с памятью (token) и с микропроцессорными картами (смарт-карты) [3,26].

Несмотря на широкую применимость и перечисленные выше достоинства, используемые методы идентификации личности по голосовым данным обладают рядом серьезных недостатков. К ним относятся, прежде всего, невысокая различающая способность методов и значительный процент ошибок как первого рода (ложно отвергнутые лица, имеющие право на допуск), так и наиболее опасного второго рода (ложно допускаемые к конфиденциальной информации лица, права на допуск к ней не имеющие) [27,28]. Особо усложняет ситуацию проведение идентификации в

реальных условиях, сопровождающихся набором неблагоприятных внешних факторов.

Идентификация личности по голосу, проводимая в реальных условиях, встречается со следующими серьезными затруднениями. Во-первых, при такой идентификации возникают всевозможные аппаратные искажения и помехи [29], обусловленные особенностями аппаратуры и устройств для записи, обработки и хранения информации. Во-вторых, на голосовой сигнал неизбежно накладываются внешние акустические шумы, которые могут существенно искажать индивидуальные информативные характеристики. В силу этого системы идентификации, демонстрировавшие достаточно высокую эффективность в лабораторных условиях, при анализе речевой информации с внешними шумами могут показать надежность, значительно более низкую [30]. Наконец, в ряде задач приходится проводить идентификацию в весьма сложных условиях наложения голосов нескольких дикторов, в том числе с близкими акустическими характеристиками. Отметим, что исследования возможностей голосовой идентификации для этого наиболее сложного случая практически не проводились.

Проведение голосовой идентификации включает в себя комплекс технических, алгоритмических и математических методов, охватывающих все этапы, начиная с записи голоса и заканчивая классификацией голосовых данных. Рассмотренные трудности и недостатки приводят к выводу, что дальнейшее развитие систем голосовой идентификации настоятельно требует разработки новых подходов, направленных на обработку больших массивов экспериментальных акустических сигналов, их эффективный анализ и надежную классификации. Это свидетельствует об актуальности исследований по созданию новых математических методов обработки, анализа и классификации голосовых данных, обеспечивающих надежность и достоверность идентификации личности.

Настоящая работа проводилась в соответствии с проектом Аналитической ВЦП «Развитие научного потенциала высшей школы на 2009–2011 г.г.» (рег. номер НИР 01200959307) и проектом № 8.1664.2011 по государственному заданию Минобрнауки РФ на 2012–2014 г.г.

Цель и задачи исследования. Основной **целью** диссертационного исследования являлось повышение надежности голосовой идентификации личности в условиях внешнего механического шума.

В соответствии со сформулированной целью в работе были поставлены следующие **задачи**:

1. Выбрать наиболее информативные признаки акустических голосовых сигналов для применения математического аппарата многомерного анализа данных.

2. Разработать алгоритмы применения проекционных методов многомерного анализа, ориентированных на обработку голосовых сигналов.

3. Сформировать базу данных голосовых сигналов дикторов, предназначенную для тестирования систем голосовой идентификации личности.

4. Исследовать влияния искажений голосового сигнала, различных видов и величины внешнего шума при идентификации дикторов в реальных условиях.

5. Рассчитать ошибки первого и второго рода для оценки надежности разработанного подхода.

Научная новизна.

1) В диссертации впервые проведено комплексное исследование возможности технологии голосовой идентификации пользователей на основе проекционных методов анализа многомерных данных. Предложены и апробированы математические решения по совершенствованию голосовых средств защиты информации, отличающиеся использованием в проекционных

методах коэффициентов мел-кепстрального разложения, полученных при математической обработке голосов дикторов. (пункт 11 паспорта специальности 05.13.19)

2) Предложено использование в технологиях голосовой идентификации алгоритмов обработки данных, отличающихся применением методов главных компонент и проекций на латентные структуры. Алгоритмы основаны на математических решениях, отличающихся классификацией идентифицируемых сигналов с помощью метрики Махаланобиса. Для случая идентификации пользователей с близкими голосовыми характеристиками предложено решение, отличающееся иерархическим разбиением кластеров голосовых сигналов с близкими акустическими характеристиками (пункт 13 паспорта специальности 05.13.19).

3) Предложены решения по совершенствованию средств голосовой защиты информации и разработана архитектура системы защиты по голосовым данным на основе проекционных методов, отличающиеся уменьшением влияния величины внешнего шума и искажений акустических сигналов на результат голосовой идентификации (пункты 11, 13 паспорта специальности 05.13.19).

Практическая значимость результатов.

Результаты работы могут применяться при разработке систем и программно-аппаратных устройств биометрической идентификации личности, в различных системах разграничения доступа, в том числе использующих телекоммуникационные каналы связи.

Предложенные и апробированные алгоритмы построения баз голосовых данных могут использоваться при создании технических систем голосовой идентификации, при оценке надежности работы таких систем. Результаты исследования влияния внешнего шума могут быть применены при апробации

и проверке методов и технических средств оценки защищенности речевой информации от утечки по различным каналам.

Предложенный в диссертации иерархический подход обеспечивает практическое проведение достоверной классификации данных и последующую идентификацию личности в наиболее сложных случаях наложения записей голосов большого числа дикторов с близкими частотными характеристиками голоса. Это может быть использовано в задачах криминалистической экспертизы при расследовании компьютерных преступлений, когда возникает необходимость проведения идентификации неизвестных голосовых записей.

Результаты работы внедрены в ряде специализированных организаций, они использованы также при подготовке учебных дисциплин для студентов, обучающихся по специальности «Комплексное обеспечение информационной безопасности автоматизированных систем» и направлению магистратуры «Информационная безопасность». На Международной выставке научно-технических и инновационных разработок «Измерение, мир, человек – 2011» подготовленный по результатам работы экспонат был удостоен серебряной медали.

Достоверность и обоснованность полученных результатов. Достоверность полученных результатов подтверждается применением современных математических методов исследования, статистической обработкой результатов классификации, анализом решения задачи с помощью различных способов, а также соответствием полученных данных и результатов исследований других авторов.

Положения, выносимые на защиту:

1. Технология голосовой идентификации, основанная на применении коэффициентов мел-кепстрального разложения акустических голосовых сигналов в качестве признаков в проекционных методах многомерного

анализа данных и позволяющая повысить разделимость голосов дикторов при их идентификации. (пункт 11 паспорта специальности 05.13.19)

2. Решения по совершенствованию средств защиты информации, заключающиеся в применении алгоритмов анализа голосовых сигналов с помощью методов главных компонент и проекций на латентные структуры, обеспечивают повышение защиты информации при идентификации личности в условиях внешнего шума и близких голосовых характеристик идентифицируемых дикторов. (пункт 13 паспорта специальности 05.13.19)

3. Предложенный метод идентификации и разработанная архитектура системы защиты позволяют уменьшить влияние величины внешнего механического шума различного вида и аппаратных искажений голосовых сигналов на результаты идентификации пользователей. (пункты 11, 13 паспорта специальности 05.13.19)

Апробация результатов.

Материалы диссертационной работы докладывались на следующих научных конференциях: VI Всероссийская научно-практическая конференция молодых ученых (г. Оренбург, 2007 г.); Межрегиональный семинар «Проблемы правовой и технической защиты информации» (г. Барнаул, 2008 г.); VIII научно-практическая конференция студентов, аспирантов и молодых ученых «Безопасность информационного пространства» (г. Челябинск, 2009 г.); X конференция молодых ученых «Наука и инновации XXI века» (г. Сургут, 2009 г.); XI Всероссийская научно-практическая конференция «Проблемы информационной безопасности государства, общества и личности, безопасность нанотехнологий» (г. Томск, 2009 г.); XII Всероссийская научно-практическая конференция «Проблемы информационной безопасности государства, общества и личности» (г. Белокуриха, 2010 г.); XIII Всероссийская научно-практическая конференция «Проблемы информационной безопасности государства,

общества и личности» (г. Новосибирск, 2012 г.), Всероссийская научно-практическая конференция «Многоядерные процессоры, параллельное программирование, ПЛИС, системы обработки сигналов» (г. Барнаул, 2013 г.), Междисциплинарная межвузовская конференция студентов, магистрантов и аспирантов «Проблемы правовой и технической защиты информации – 2014» (г. Барнаул, 2014 г.), а также на Международной выставке научно-технических и инновационных разработок «Измерение, мир, человек – 2011» (г. Барнаул, 2011 г.).

Личный вклад автора. Автор лично участвовал в проведении теоретических и экспериментальных исследований по теме диссертации, их обсуждении, в анализе и интерпретации полученных результатов; в совместных работах автор лично участвовал в обсуждении результатов исследований с соавторами.

Соответствие диссертации паспорту специальности. Диссертационная работа по своим целям, задачам, содержанию и научной новизне соответствует п. 11 «Технологии идентификации и аутентификации пользователей и субъектов информационных процессов. Системы разграничения доступа» и п. 13 «Принципы и решения (технические, математические, организационные и др.) по созданию новых и совершенствованию существующих средств защиты информации и обеспечения информационной безопасности» паспорта специальности 05.13.19 – Методы и системы защиты информации, информационная безопасность.

Публикации. Содержание диссертации изложено в 18 публикациях, в том числе в 6 публикациях из перечня ВАК, получены свидетельство о государственной регистрации программы для ЭВМ и свидетельство о государственной регистрации базы данных.

Структура и объем диссертации. Диссертационная работа состоит из введения, 4 глав, заключения и списка литературы. Диссертация содержит 125 страницы, 53 рисунок. Список цитируемой литературы включает 132 наименования.

В первой главе «Проблемы идентификации личности по голосу» рассматривается применение голосовой идентификации личности в задачах защиты информации. Описываются различные виды биометрической идентификации. Каждая из биометрических технологий, обладая своими преимуществами и недостатками, может быть наиболее подходящей для определенной практической задачи. Более подробно описывается голосовая идентификация. Рассматриваются задачи защиты информации и других областей в которых используется голосовая идентификация. Приведены характеристики по которым оцениваются системы голосовой идентификации. Среди них наиболее значимые надежность, быстродействие и помехоустойчивость. Сделан обзор работ, посвященных проблеме воздействия шума и помех на голосовую идентификацию. Описаны методы получения и предварительная обработка речевого сигнала, извлечение признаков. Далее проводится обзор методов обработки, анализа и классификации данных при голосовой идентификации. Рассмотрены методы классификации и принятия решения. Были поставлены основные задачи.

Во второй главе «Выбор информативных голосовых признаков для проекционных методов анализа» описываются проекционные методы: метод главных компонент и метод проекций на латентные структуры. Рассмотрен математический аппарат обоих методов. Приведены описания процедур центрирования и шкалирования для указанных методов. Далее следует описание процесса получения исходных голосовых характеристик. Описана аппаратная часть. Указаны факторы негативно влияющие на голосовую идентификацию, а также способы противодействия им. Приводится выбор

информативных голосовых признаков, рассматривая при этом несколько преобразований.

В третьей главе «Анализ и классификация голосовых данных» приводится описание алгоритмов голосовой идентификации дикторов. Описано предварительное тестирование разработанных алгоритмов. Далее приводится анализ голосовых сигналов на основе метода ближайших соседей. Описан математический аппарат метода ближайших соседей и его модификация. Описывается иерархический подход для голосовой идентификации. Анализируются голосовые сигналы на основе иерархического подхода.

В четвертой главе «Влияние помех на идентификацию личности по голосу» приводится описание влияния искажения голосового сигнала на идентификацию, а также влияния внешнего акустического шума. Для последнего случая приводится результаты идентификации от уровня и вида шума. Описывается оценка ошибок идентификации при искажении и наложении акустического шума.

В заключении обобщены итоги диссертационной работы и сформулированы основные результаты и выводы.

ГЛАВА 1. ПРОБЛЕМЫ ИДЕНТИФИКАЦИИ ЛИЧНОСТИ ПО ГОЛОСУ

1.1 Голосовая идентификация личности в задачах защиты информации

С быстрыми темпами роста информатизации общества растет потребность в защите конфиденциальной информации. Одной из распространенных мер защиты информации является разграничение доступа. Разграничение доступа включает в себя аутентификацию – процедуру проверки подлинности, и идентификацию – присвоение идентификатора и (или) сравнение предъявленного идентификатора с перечнем идентификаторов [31]. Существует более простая задача – верификации, подразумевающая сравнение предъявляемых признаков, заявленных как соответствующие известному объекту, с известными признаками того же объекта.

Среди множества различных методов разграничения доступа особое место занимают биометрические технологии. Растущий интерес к биометрическим технологиям связан в основном с удобством их применения. Часто под биометрическими технологиями, в контексте разграничения доступа, понимают методы биометрической идентификации. Задача идентификации подразумевает сравнение признаков одного объекта с признаками других известных объектов.

Большинство методов биометрической идентификации можно разделить на две группы: физиологические (или статические) и поведенческие (или динамические, учитывающие подсознательные действия). Физиологические методы идентификации: по отпечаткам пальцев, по геометрии рук, по отпечатку ладони, по венам руки, по лицу (в 2-мерном и 3-мерном пространстве), по радужной оболочке и сетчатке глаза, по ушной

раковине, по зубам, по отпечатку и геометрии стопы, по дезоксирибонуклеиновой кислоте. Поведенческие методы идентификации, или поведенческая биометрия (behavioral biometrics): по голосу, по подписи (почерку), по динамике работы с клавиатурой или мышью, по походке, исследуется также возможность идентификации по электромагнитным волнам мозга.

Ведущую роль среди биометрических технологий занимают технологии, основанные на распознавании отпечатков пальцев, лица, радужной оболочки глаз и голоса. Для определения эффективности технологий оценивают такие критерии, как надежность, устойчивость к подделке, устойчивость к окружающей среде (помехоустойчивость), стабильность признака от времени, скорость, стоимость и удобство применения. Рассмотрим достоинства и недостатки некоторых из биометрических технологий.

К достоинствам идентификации по отпечаткам пальцев можно отнести то, что отпечатки не меняются с возрастом, обладают высокой надежностью, стоимость системы идентификации относительно низкая. К недостаткам же можно отнести то, что папиллярный узор отпечатков пальцев очень легко повреждается, например, механическим, химическим воздействиями, также недостатком является подверженность пропуску по поддельным изображениям отпечатков.

Идентификацию по лицу разделяют на два вида: 2-мерную и 3-мерную. Достоинствами 2-мерной является то, что возможна идентификация на некотором расстоянии и стоимость системы низкая, а к недостаткам можно отнести то, что лицо меняется с возрастом, низкая надежность, чувствительность к внешним факторам. Для 3-мерной идентификации существенно ниже чувствительность к внешним факторам и высокий уровень

надежности. Недостатком 3-мерной идентификации является высокая стоимость оборудования.

Биометрическая технология, основанная на радужке оболочки глаз, имеет высокую надежность, со временем радужные оболочки глаз практически не изменяются и устойчивость к подделке может обеспечиваться различными методами защиты. Существенным недостатком является высокая стоимость.

Все оценки критериев эффективности и надежности той или иной технологии в большой мере зависят от используемых баз данных. В лабораторных условиях созданные базы данных биометрических признаков при идентификации могут обеспечивать высокую надежность, но в реальных условиях, где влияют различные внешние помехи, надежность может оказаться значительно ниже заявленной. Несмотря на разную эффективность этих технологий, каждая биометрическая технология может быть определенно лучше остальных для какой-либо специфической задачи.

Голосовая идентификация обладает такими достоинствами, как удобство применения и невысокая стоимость. Недостатком такой идентификации является низкая надёжность. Одним из перспективных путей повышения надёжности голосовой идентификация является привлечение характеристик динамики подсознательных движений, что активно используется при идентификации по подписи [32]. С другой стороны, есть области применения, в которых голосовая идентификация является наиболее удобной, например, удалённый доступ к телекоммуникационным каналам связи по анализу голосовых данных.

В современных системах голосовой идентификации для повышения надежности применяют тексто-зависимую идентификацию, например, произнесение парольной фразы, которая каждый раз генерируется случайным образом. Использование индивидуальных признаков и совпадение

сгенерированной и распознанной парольных фраз повышает надежность. Тексто-независимая идентификация подразумевает использование только индивидуальных признаков.

Немаловажной характеристикой системы голосовой идентификации является скорость (быстродействие) определения личности. Быстродействие особенно существенно для приложений, обрабатывающих большие базы голосовых данных и работающих в реальном режиме времени. Повышение быстродействия может быть достигнуто за счет использования новых быстрых алгоритмов обработки данных. Таким образом, голосовая идентификация личности, несмотря на указанные в настоящей работе недостатки, в определенных условиях обладает существенными преимуществами, которые необходимо развивать.

Повышение надежности голосовой идентификации важно не только для такого направления, как разграничение доступа к физическим и информационным объектам, например, доступа к операционной системе персонального компьютера или удаленного доступа к телекоммуникационным каналам связи по анализу голосовых данных. Определенный интерес имеется и для смежных направлений речевых технологий: распознавание устной речи, управление голосовыми командами и другие. На сегодняшний день широкое распространение получила электронно-цифровая подпись для защиты конфиденциальных документов в виде защищенного электронного устройства (token), в связи с этим перспективным направлением является разработка защиты конфиденциальных документов на основе речевой подписи [33]. Кроме того, практические применения таких исследований полезны для правоохранительных органов, например, поиск подозреваемого по голосу в телекоммуникационных голосовых каналах связи.

Поиск диктора по голосу может успешно применяться для обнаружения диктора (обнаружение признаков диктора в аудио-потока) [34] и протоколирования дикторов (определение участков устной речи дикторов по отдельности в аудио-потоке) [35].

Подобно отпечаткам пальцев в криминалистическом учении – дактилоскопии, в фоноскопии используется свой криминалистический объект – фонограмма [36,37]. В связи с этим для фоноскопической экспертизы используют свои специфические методы и технические средства. Существуют два основных направления криминалистического исследования фонограмм: первое – исследование технических характеристик звукозаписи, второе – исследование непосредственно самой устной речи человека. В основе экспертного криминалистического исследования лежит теория криминалистической идентификации и диагностики. Целью идентификации и диагностики является установление объекта. Идентификация подразумевает сопоставление звуковых признаков нескольких конкретных объектов, и результатом этого сопоставления является сходство или различие объектов. При диагностике используют известные признаки класса или группы объектов и сравнивают их с признаками конкретного объекта, в результате чего определяется принадлежность его к данному классу или группе. Для достижения цели эксперты решают различные виды задач. Это исследование источника звука, среды распространения (закрытое помещение, открытое пространство), а также приемника (запись, хранение, воспроизведение). Основной задачей идентификационных исследований фонограмм является идентификация личности по фонограмме устной речи.

Существуют много инструментов исследования речевых сигналов, основанных на математической обработке, среди которых основным является спектральный анализ. Использование временной зависимости и спектрального анализа дает дополнительную информацию и называется

спектрально-временной анализ устной речи. Другой вид анализа, лингвистический анализ устной речи, дополняет спектральный дополнительной более высокоуровневой информацией. До сих пор в криминалистике используется сопоставление спектрограмм речевого сигнала путем сравнения чисто внешнего сходства. Наиболее простой способ исследования устной речи с целью идентификации – ее исследование на слух непосредственно или по фонограмме. Признаков, которые используются при исследованиях устной речи, существует большое количество, но четкой систематизации признаков нет.

Для принятия решения о сходстве или различии, либо принадлежности или непринадлежности к известному классу объектов в задаче идентификации личности применяют статистические методы. Статистические методы обычно выдают ответ с некоторой вероятностью, которая, в общем случае, не может служить доказательством сходства или принадлежности. Вычислительные мощности современных компьютеров позволяют обрабатывать большое количество данных, что способствует созданию более надежных автоматизированных систем идентификации личности. С другой стороны, необходимость построения автоматизированных систем идентификации обусловлена влиянием субъективного фактора при криминалистической экспертизы.

Немаловажной характеристикой системы голосовой идентификации является помехоустойчивость. Под помехами понимаются искажения, шумы, импульсные помехи и т. п. Современные методы классификации, используемые в системах голосовой идентификации, очень чувствительны к шуму, что приводит к снижению надежности при воздействии шума.

Помехи могут исходить от разных источников, например, от двигателя автомобиля, от звука ветра, от голосов других дикторов, от реверберации звука. В работе [38] источники искажений разделяют на два вида: аддитивный

шум и искажения канала. Аддитивный шум может быть стационарным и нестационарным. Стационарный шум имеет спектральную плотность мощности, не изменяющуюся с течением времени, например, шум двигателя автомобиля. Нестационарный шум имеет статистические свойства, которые меняются с течением времени, например, голоса других дикторов. Примером искажения канала может быть реверберация звука, то есть звук многократно отражается от предметов в помещении. Окружающий фон, в некоторых случаях являясь помехой, может оказывать значительное влияние на голосовую идентификацию. Так, чтобы оценить, как сильно влияет окружающий фон, в работе [39] использовали записи различных реальных внешних условий: в машине, в поезде, в ресторане, в беседе. В телефонном канале помехами могут быть щелчки, перегрузки, музыкальные сигналы, гудки (тональные сигналы). Например, в работе [40] представлен алгоритм детектирования музыкальных фрагментов, который может использоваться для борьбы с одной из помех телефонного канала.

В статье [41] представляется новый метод для распознавания диктора с очень ограниченными речевыми данными обучения и при наличии фонового шума. Предлагается использовать одну из форм косинусного сходства в качестве меры расстояния между векторами речевых признаков. Каждый речевой кадр моделируется с помощью поддиапазона признаков, в этих рамках подготовки вводятся обучение с многими состояниями (multicondition training) и оптимальный выбор признака, в результате чего система способна выполнять распознавание диктора в присутствии реального, изменяющегося во времени шума, который неизвестен во время обучения. Эксперименты по идентификации диктора проводились с использованием SPIDRE базы данных [42]. Проводились также эксперименты для тестирования новой модели идентификации диктора, учитывая ограниченные данные обучения и с

различными уровнями и типами реалистичных фоновых шумов. Полученные результаты показали надежность новой системы.

Работа [43] посвящена проблеме верификации диктора, где аддитивный шум присутствует при регистрации и тестировании высказывания. При регистрации шумных сигналов речи наблюдалось понижение достоверности стандартной системы верификации диктора. В этой работе использовался речевой корпус с зашумленными условиями, на основе данных NIST SRE 2008 года [44] и NIST SRE 2010 года [45], построенный с использованием открытого исходного кода и свободно доступных образцов шума.

В работе [46] исследуется проблема идентификации и верификации дикторов в зашумленных условиях, предполагая, что речевые сигналы были подвержены шуму окружающей среды, характеристики которого неизвестны. Это исследование частично мотивировано возможным применением технологий распознавания диктора на мобильных устройствах или в сети Интернет. В этой статье описывается метод, который сочетает в себе обучение модели со многими состояниями (multicondition model training) и теории неполных данных (missing-feature theory) к модели шума с неизвестными спектрально-временными характеристики. Обучение со многими состояниями проводится с использованием моделирования зашумленных данных с ограниченным изменением шума, обеспечивая некоторую компенсацию, а теория неполных данных применяется для уточнения компенсации, игнорируя изменение внешнего шума, который обеспечивает условия обучения, тем самым снижая ошибки обучения и тестирования. Новый алгоритм был протестирован с помощью двух баз данных, моделированной и реалистичных зашумленных речевых данных. Первая база данных представляла собой реконструкцию базы данных TIMIT [47] при перезаписи в присутствии различных типов шума, используется для проверки модели при идентификации диктора с акцентом на

разновидности шума. Вторая база данных представляла собой базу данных, собранных портативным устройством в реалистичных зашумленных условиях, используемых для дальнейшей проверки модели при реальных верификациях диктора. Предложенная в статье модель по сравнению с исходными системами обладает более низким процентном ошибки.

В работе [48] исследуется использование микрофонных массивов в системах распознавания диктора «свободные руки» (handsfree). Громкая связь предпочтительнее во многих потенциальных применениях распознавания диктора, однако распознавание с одним удаленным микрофоном является проблематичным в реальных условиях шума. Возможным решением этой проблемы является использование микрофонных массивов, которые имеют потенциал для улучшения качества сигнала, основанный исключительно на знании направления его прихода. Использование микрофонного массива для повышения надежности систем распознавания речи была хорошо изучена в последнее время, однако мало исследований было проведено в области распознавания диктора. В статье обсуждается применение микрофонных массивов для приложений распознавания диктора, что представляет собой экспериментальную оценку применения технологии «свободные руки» при верификации диктора в зашумленных условиях.

В задаче голосовой идентификации применяют различные математические, алгоритмические, технические методы, начиная с этапа записи голоса и заканчивая этапом классификации. Практически каждая система идентификации содержит четыре основных этапа: получение сигнала, предварительная обработка сигнала, извлечение признаков и классификация признаков. Рассмотрим эти этапы для нашей задачи.

Этап получения сигнала. Метод получения или записи голосового сигнала, в большинстве случаев, представляет собой запись сигнала с помощью микрофона и представление сигнала в цифровом виде с помощью

аналого-цифрового преобразователя. В качестве аналого-цифрового преобразователя обычно используют звуковую карту персонального компьютера либо цифровой диктофон. Цифровые данные кодируются импульсно-кодовой модуляцией и помещаются в формат файла-контейнера (Waveform Audio File Format) для хранения записи оцифрованного аудиопотока. Параметры звуковой записи обычно выбирают следующие: битность отсчета – 16 бит, частота дискретизации – 22500 Гц. Так как современные цифровые мобильные устройства обычно имеют встроенный микрофон и производительные аппаратные средства, то создание система аутентификации по голосу с привлечением более затратных по вычислениям методов вполне решаемая задача для мобильных платформ. Тем не менее, обеспечить минимальные вычислительные затраты при сохранении точности, помехоустойчивости к различным видам помех и достаточную надежность при распространенных аппаратных средствах все же необходимо.

Этап предварительной обработки. Полученные цифровые сигналы, как и аналоговые, содержат в себе некоторую долю искажений и помех. Под искажениями понимаются искажения речеобразующего тракта (например, болезнь горла) и речепередающего канала (например, искажения телефонного канала).

В работах [49–51] описываются некоторые способы борьбы с искажениями речепередающего канала. В качестве помех выступают не только внешние шумы, но и речь посторонних людей. Внешние шумы обычно подавляют на этапе предварительной обработки с помощью различных фильтров: полосовых, медианных, адаптивных, на основе вейвлет-преобразований и других.

Например, в работе [52] для борьбы с шумом используют метод спектрального вычитания. Также для уменьшения влияния негативных

факторов (искажений и помех) применяют различные виды нормировок [53,54].

Этап извлечение признаков. Извлечение признаков обычно происходит с помощью Фурье-преобразования, вейвлет-преобразований, линейного предсказания и других. Коэффициенты преобразований выступают в качестве признаков. В настоящее время точно не определены голосовые признаки, по которым можно однозначно идентифицировать личность человека.

Выбор признаков влияет также на надежность идентификации. Существуют, методы которые описывают интегральные характеристики человеческого голоса и служат для извлечения тонов, динамики речи, просодических характеристик. Такими методами являются преобразование Фурье (амплитудно-частотное распределение), кепстральное преобразование (амплитудно-временное распределение), преобразование линейного предсказания (амплитудно-частотное распределение). Существуют также формантные методы и методы выделения фонем. Использование вейвлет-коэффициентов [55,56] не обеспечивает значительного преимущества, к тому же учет частотной и временной составляющих влечет дополнительные вычислительные затраты. Для эффективного использования частотной и временной составляющих голосовых сигналов в задаче идентификации личности по голосу необходимо проводить дополнительные исследования. Преимущества тех или иных голосовых признаков проявляются в конкретных случаях при определенных условиях и на определенной речевой базе данных.

В работе [57] сравнивались несколько векторов признаков, которые строились на основе коэффициентов: быстрого преобразования Фурье, линейного преобразования, кепстрального преобразования. Кроме того, к коэффициентам дополнительно применялось мел-преобразование. Для

различных методов классификации эффективнее могут оказаться наборы коэффициентов различных преобразований.

Из вышеперечисленных голосовых признаков, в разных комбинациях, формируется вектор признаков в виде последовательности чисел. Формирование векторов признаков для одного или нескольких дикторов входит в этап калибровки или обучения системы идентификации диктора. На выходе этапа калибровки строится общая модель одного или нескольких дикторов, которая соотносится с исходными речевыми данными. На этапе тестирования (проверки) вектор признаков извлекается из звуковой волны и сравнивается с построенной моделью.

Каждый вектор признаков должен формироваться оптимально, учитывая вычислительные затраты, с одной стороны, и информативность признаков – с другой. Так, при высокой информативности вектора признаков, а именно, большом количестве признаков, вычислительные затраты возрастают, и наоборот, при малом количестве признаков вычислительные затраты убывают. Однако при увеличении количества признаков в векторе информативность не всегда увеличивается. Некоторые оценки информативности признаков можно получить на этапе тестирования с помощью ошибки классификации (идентификации), то есть минимум ошибки будет соответствовать большей информативности. Поэтому часто проводят отдельные исследования по формированию оптимального вектора признаков для каждого диктора, то есть исследуют зависимость сочетаний различных наборов признаков из исходных данных (коэффициентов каких-либо преобразований) от ошибки классификации. Сочетания различных наборов признаков может исследоваться как простым перебором, так и с использованием методов оптимизации (например, генетических алгоритмов). С другой стороны, сами признаки изменяются в некотором диапазоне, что вызвано либо влиянием внешних шума и помех, либо влиянием внутренних

искажений голосового аппарата человека. Поэтому различие тестовых и калибровочных условий классификации может привести к формированию разных векторов признаков для одного и того же диктора, а следовательно – к понижению достоверности методов классификации.

Этап классификации признаков. В этот этап входит применение математических методов классификации, с помощью которых осуществляется принятие решения, а также расчет ошибок классификации. Более подробно составляющие этого этапа описаны в следующем разделе.

1.2 Методы обработки, анализа и классификации данных при голосовой идентификации

В задаче голосовой идентификации используют те же методы классификации, что и в области распознавания образов, а именно, методы статистического моделирования, которые строят определенные модели векторов акустических признаков. Наиболее распространенными из них являются модели гауссовых смесей и скрытые марковские модели. Однако другие модели, например, многослойные персептроны или машина опорных векторов, также успешно используются в данной задаче. Кроме того, в последнее время наблюдается тенденция использования комбинаций нескольких моделей.

Модели гауссовых смесей часто используют для тексто-независимой верификации дикторов [20,58,59], оценивая плотности вероятностей изменчивости речевых данных. Обладая невысокими вычислительными затратами и малой чувствительностью к временным изменчивостям речи, модели гауссовых смесей хорошо проявили себя в условиях близких к условиям тихого окружения, применения высококачественных микрофонов и т. п. В реальных условиях присутствия фонового шума, использования

различных микрофонов и каналов передачи эффективность моделей гауссовых смесей ухудшается.

Из-за ограниченности данных, доступных для тренировки модели диктора, востребованы технологии адаптации: EM-алгоритм (Expectation-Maximization) [60], максимум апостериорной вероятности (Maximum a Posteriori Probability) или максимум правдоподобия линейной регрессии (Maximum Likelihood Linear Regression) [61].

Скрытые марковские модели являются статистическими моделями, в которых система моделируется как марковский процесс с неизвестными параметрами [62]. Целью является определение наиболее вероятного состояния последовательности тестового набора относительно предварительно тренировочных моделей. Для приложений распознавания диктора каждое состояние скрытой марковской модели может представляться различными элементами речи [63]. Временная информация кодируется переходом из одного состояния в другое относительно разрешенных переходов. В этом случае метод идентификации на основе скрытых марковских моделей заключается в определении для каждого диктора наилучшего положения между последовательностью тестового речевого вектора и скрытой марковской моделью, связанной с определенным словом или фразой.

Современные тексто-независимые системы верификации диктора, использующие модели гауссовых смесей, не учитывают временное упорядочение векторов признаков. Лингвистическая и временная структура речевого сигнала в виде счетов и всех звуков, используемых в представлении, не дает уникальной модели. Скрытые марковские модели обладают определенными преимуществами. Временные и лингвистические знания могут быть зарегистрированы с помощью скрытых марковских моделей. В задаче текстозависимого распознавания диктора используются априорные

знания содержания текста при этом скрытые марковские модели [64] точнее, чем модели гауссовых смесей. В работе [65] показаны некоторые преимущества применения комбинации распознавания речи на слоговой скрытой марковской модели и дикторо-ориентированого распознавания на основе модели гауссовых смесей.

Многослойные персептроны являются разновидностью обучаемых нейронных сетей [66]. Применение нейросетевых методов в задаче верификации диктора показано в работах [67,68]. Для систем верификации диктора многослойные персептроны могут быть двоичными классификаторами, которые выделяют классы «своего» и «чужого». Многослойный персептрон обычно состоит из нескольких слоев, каждый из которых имеет несколько вершин. Каждая вершина вычисляется как сумма линейных весов всех входных соединений, где веса суммы являются подстраиваемыми параметрами. Нелинейная функция перехода применяется для результата вычисления выхода вершины. Веса сети оцениваются через градиент наклона, основанный на алгоритме обратного распространения. Многослойный персептрон будет классифицировать доступ «своего» и «чужого», считая каждый кадр тестового высказывания. Окончательно счета высказывания усредняются по всем выходам многослойного персептрона для всех кадров высказывания. Недостатками многослойного персептрона является относительная сложность выбора оптимальной конфигурации, а также необходимость большого количества данных для этапов обучения и проверки (кросс-проверки).

Машина опорных векторов является двоичным классификатором [69]. Основным принципом является проекция нелинейно разделимых многомерных данных в гиперпространство, где они могут быть линейно разделимы. Учитывая, что набор векторов признаков принадлежит к двум классам, которые разделяются гиперплоскостью, машина опорных векторов

будет пытаться найти гиперплоскость с максимальным краем. Другими словами, расстояние между ближайшими помеченными векторами к гиперплоскости будет максимальным. Эта гиперплоскость может быть в дальнейшем быть использована (на этапе проверки) для определения, к какому классу принадлежит неизвестный вектор признаков. В последние годы машина опорных векторов считается одним из эффективных методов дискриминации [70]. В задаче верификации диктора машина опорных векторов может использоваться отдельно, либо в комбинации с другими методами классификации. Например, в работе [71] используется комбинирование методов на основе модели гауссовых смесей и машины опорных векторов. В ней делается попытка использовать компенсации изменчивости диктора и канала, а именно, с помощью модели гауссовых смесей сформировать некоторый средний вектор, который далее обрабатывается машиной опорных векторов. Комбинирование методов приводит к повышению точности классификации.

Существенной проблемой для систем голосовой идентификации, основанных на вышеперечисленных методах, является сильное воздействие внешнего окружающего шума на исходные голосовые записи, из которых выделяются информативные признаки. Обусловленное шумом искажение этих признаков вызывает высокий уровень ошибок идентификации. Данная проблема исследуется в следующих работах.

В работе [46] исследовалось влияние шума на распознавание диктора. В качестве основного метода классификации для распознавания диктора использовались гауссовы смеси. Для обучения и тестирования использовался речевой корпус TIMIT [72], к которому аддитивно добавлялся шум. Речевой корпус TIMIT содержал микрофонные записи голосов 630 дикторов (438 мужских, 192 женских), для каждого диктора имелись 10 высказываний, каждое высказывание было в среднем с продолжительностью около 3 секунд.

Для обучения системы распознавания диктора использовались 8 высказываний каждого диктора, а для тестирования – оставшиеся 2 высказывания. Вектор признаков строился на основе выделенных полос спектра. Воздействие накладываемого шума на голосовые записи определялось отношением «полезный сигнал/шум».

Воздействие шума исследуется и в работе [41]. В качестве основного метода классификации для распознавания диктора используются гауссовы смеси. Речевой материал был взят из речевого корпуса «SPIDRE» [73], а именно, использовались голосовые записи 45 дикторов (27 мужских, 18 женских). Для обучения использовались речевые отрезки длительностью 30 секунд, а для тестирования – отрезки длительностью 10 секунд. Признаки формировались на основе кепстральных коэффициентов. Отношение «полезный сигнал/шум» задавался равным 15 дБ и 20 дБ.

В работе [74] также исследовалось влияние шума на идентификацию диктора по голосу. Базовый метод идентификации диктора основывался на гауссовых смесях. Речевой материал был взыт из речевой базы «TIDIGITS» [75], а именно, использовались записи голосов 31 диктора (21 мужских и 10 женских). Для каждого диктора имелись 77 высказываний, причем 50 высказываний использовались для обучения, а 27 высказываний – для тестирования. Признаки формировались на основе кепстральных коэффициентов.

Результаты влияния белого шума приведены в работе [76], где в качестве базового метода идентификации использовался многослойный персептрон. В качестве речевого материла выступали телефонные записи голосов корпуса «The 2002 NIST Speaker Recognition Evaluation corpus» [77], из которого использовались записи 50 дикторов (20 мужских и 30 женских). Для каждого диктора имелось около 2 минут речевого материала, который разбивался на отрезки длительностью 5 секунд. При тестировании

использовались 4 отрезка, а для обучения – остальные отрезки. Признаки формировались на основе кепстральных коэффициентов. В качестве шума выступали различные его виды, среди которых были белый и речеподобный шум. Шум добавлялся аддитивно, определенный уровень которого обеспечивал отношение «полезный сигнал шум» от -12 дБ до 18 дБ с интервалов 6 дБ.

Воздействие речеподобного шума исследуется в работе [78]. В качестве речевого материала использовались телефонные записи 300 дикторов из базы данных «The 2008 NIST Speaker Recognition Evaluation» [79]. На каждого диктора приходилось около 5 минут речевого материала. Предварительно выделенная информативная составляющая записи делилась на отрезки длительностью 5 секунд, из них 2 отрезка использовались для тестирования, а для обучения – остальные отрезки. Величины отношения «полезный сигнал шум» принимали значения от 0 дБ до 24 дБ с интервалом 6 дБ. В качестве базового метода идентификации применялся один из методов нейронных сетей «deep neural network». Признаки формировались на основе кепстральных коэффициентов.

В задаче идентификации диктора существует проблема **принятия решения**, которое по набору речевых данных, в ходе этапа проверки, относится к заявленной модели или нет. При заданном речевом сегменте X и заявленной личности S, система идентификации диктора должна выбрать одну из следующих гипотез:

H_S: X определяется как S

$H_{\bar{S}}$: X не определяется как S

Решение между двумя гипотезами часто основано на отношении правдоподобия:

$$\Lambda(X) = \frac{p(X|H_S)}{p(X|H_{\bar{S}})} = \begin{cases} > \theta, \text{принимается } H_S \\ < \theta, \text{принимается } H_{\bar{S}} \end{cases}, \quad (1.1)$$

где $p(X|H_S)$ и $p(X|H_{\bar{S}})$ – функции плотности вероятности (называемые также правдоподобием), связанные с диктором S («свой) и «не диктором» $H_{\bar{S}}$ («чужой). Величина θ вводится как порог принятия или отклонения гипотезы H_S.

На практике, гипотеза H_S представляет собой модель, которая оценивается при обучающей речи предполагаемого диктора S, и $H_{\bar{S}}$ представляет собой модель, оцениваемую набором других дикторов, который охватывает как можно большее вероятное пространство альтернативных гипотез. Среди различных способов построения модели «не целевого диктора» можно выделить два основных подхода. Первый подход заключается в выборе для каждого диктора S своего набора дикторов $\bar{S_1},\ldots,\bar{S_N}$ [80,81]. Этот набор будет соответствовать модели «нецелевого диктора», причем такой набор формируется для каждого диктора. Второй подход заключается в выборе уникальных или зависимых от пола дикторов, которые выступают в роли «не целевого диктора». При этом формируется одна модель «нецелевого диктора», а для случая зависимости от пола - две модели для всех дикторов. Формирование заключается в обучении модели на большом количестве характерных дикторов. Такую модель называют универсальной фоновой моделью (Universal Background Model (UBM)) [82–84]. Отметим, что второй последний подход оказался более практичным.

После обучения модели система верификации диктора должна принять решение о принятии или отклонении заявленной личности, учитывая некоторый порог θ (формула (1.1)). Выбор соответствующего порога для конкретного приложения верификации диктора остается сложной задачей. Порог обычно выбирается на этапе разработки и является дикторо-независимым. Однако, такого рода пороги не учитывает такие факторы, как индивидуальные особенности дикторов и собственную изменчивость диктора. Кроме того, несоответствие данных на этапах разработки и тестирования

может привести к отличию заданного и оптимального пороговых значений. Поэтому на практике может использоваться дикторо-зависимый порог, который должен компенсировать эффекты изменчивости. Другой подход заключается в применении методов нормализации, которые могут учитывать вышеперечисленные факторы. Такой подход, являясь дикторо-независмым, не требует установки порога для каждого диктора.

Эффективность любой системы распознавания диктора оценивается **ошибками классификации**. Существуют два типа ошибок, имеющих место в задаче верификации: ложноотвергнутые (ошибка 1-го рода), когда системы отвергла «своего», и ложнопринятые (ошибка 2-го рода), когда система приняла «чужого». Оба типа ошибки зависят от уровня принятия решения. При высоком уровне система будет делать мало ложных принятий, но много ложных отказов. Если уровень фиксирован низким значением, то система будет более удобна для пользователей и будет делать мало ложных отказов, но много ложных принятий. Ошибки ложных принятий R_{FA} (false accept rate) и ложных отказов R_{FR} (false reject rate) вычисляются по следующим формулам:

$$R_{FA} = \frac{\text{число ложнопринятых}}{\text{число принятых чужих}}, \tag{1.2}$$

$$R_{FA} = \frac{\textit{число ложноотвергнутых}}{\textit{число принятых своих}}. \tag{1.3}$$

Эти ошибки являются нормальной оценкой разрабатываемого набора и далее будут использоваться для определения функции стоимости (Detection Cost Function). Эта функция стоимости является весовой мерой ошибок ложного принятия и ложного отказа:

$$DCF = C_{FR}P_{tar}R_{FR}+C_{FA}P_{imp}R_{FA} \tag{1.4}$$

где $C_{FR}-$ стоимость ложного отказа, $C_{FA}-$ стоимость ложного принятия, P_{tar} и P_{imp} являются априорными вероятностями «своих» и

«чужих» соответственно. DCF будет нормализована значением: $C_{FR}P_{tar}+C_{FA}P_{imp}$.

На практике в качестве меры оценки эффективности системы верификации диктора часто используют функцию стоимости *DCF*. Кроме того, минимизация функции *DCF* может оказаться полезной для оптимизации системы. Для оценки эффективности системы при динамике некоторого параметра обычно используют рабочую характеристическую кривую (Receiver Operating Characteristic) или кривую компромисса определения ошибки (Detection Error Trade-off), где вводится такая характеристика, как эквивалентная ошибка классификации (Equal Error Rate), которая появляется на пересечении кривой компромисса определения ошибки с первой биссектрисой [85]. Меры эффективности в динамике некоторого параметра позволяют достаточно просто сравнивать различные системы верификации диктора.

Проведенный обзор показывает, что все методы классификации в задаче идентификации диктора обладают своими недостатками: одни чувствительны к шуму, другие не обладают наглядностью, третьи черезмерно сложны для практического применения. В связи с этим практичские реализации и внедрение систем идентификации дикторов не достаточно распространены. В значительной степени лишены отмеченных недостатков проекционные методы анализа многомерных данных [86–88]. Использование этих методов в задаче голосовой идентификации подробно рассматривается далее в настоящей работе.

1.3 Постановка задачи

Разработка новых методов и средств для предотвращения несанкционированного доступа к информационным ресурсам посторонних лиц требует развития систем и процедур идентификации личности. Особое

место занимают системы и способы биометрической идентификации, обладающие целым рядом достоинств и преимуществ. Биометрические показатели, основанные на уникальных индивидуальных особенностях человека, наиболее надежно характеризуют отдельного пользователя. Они достаточно легко подвергаются автоматизированной обработке и хранению в базах данных, что обеспечивает удобство в использовании и относительно небольшие затраты на применение. Среди наиболее перспективных систем биометрической идентификации, основанных на физиологических особенностях человека, большую роль играет голосовая идентификация. Это обусловлено целым рядом преимуществ: акустический амплитудно-частотный спектр речевых сигналов, уникальный для каждого человека, достаточно легко записывается, эти записи удобно хранить в компьютерных базах данных, подвергая при необходимости анализу и обработке.

Возрастающее значение систем идентификации личности, основанных на анализе голосовых данных, связано с быстрым распространением передачи информации по телекоммуникационным каналам связи. Появление систем, позволяющих по каналам беспроводной мобильной связи осуществлять управление различными услугами, напрямую связано с возможностями голосовой идентификации. Близкими задачами является управление компьютерными системами с помощью голосовых команд и создание систем автоматического распознавания речи.

Появление новых и все более совершенных технических средств, предназначенных для незаконного доступа к голосовой информации (например, к конфиденциальным переговорам), совершенствование различных технических каналов утечки речевой информации обуславливает ряд других аспектов, обеспечивающих важность исследования и развития голосовой идентификации. Именно, новые системы и методы такой идентификации оказываются востребованными при разработке и апробации

устройств защиты речевой информации от утечки по акустическим и виброакустическим каналам, а также при оценке защищенности от этой утечки.

Голосовая идентификация личности тесно переплетается с предшествовавшими ей криминалистическими задачами, связанными с анализом фонограмм. Эти задачи появляются при проведении расследования преступлений, когда возникает потребность в идентификации неизвестной голосовой записи (например, телефонных переговоров). Математические и технические методы и приемы идентификации оказываются вполне применимыми и полезными при проведении фоноскопических экспертиз, они способствуют важной практической деятельности по поиску преступника при наличии записей его голоса. Особо нужно подчеркнуть, что современная компьютерная преступность связана прежде всего с удаленным доступом к информации по беспроводным каналам связи. Это обстоятельство определяет ее латентный характер и весьма высокую сложность расследования компьютерных преступлений, оставляющих, как правило, не материальные, а так называемые виртуальные следы. В этой связи системы голосовой идентификации могут оказать значительную помощь в расследовании и предотвращении компьютерных преступлений, в частности, они могут привлекаться к анализу голосовых сигналов в телекоммуникационных каналах связи.

В то же время дальнейшее развитие идентификации личности по голосу сталкивается со значительными трудностями. Это вызвано достаточно серьезными недостатками, изначально присущими данному виду биометрической идентификации. К таким недостаткам относится невысокая различающая способность метода, которая может приводить к значительному числу ошибок при идентификации. Эти ошибки могут стать недопустимыми при проведении идентификации в сложных условиях.

Прежде всего, реальные условия могут сопровождаться наложением на анализируемый голосовой сигнал внешнего акустического шума, подавляющего этот сигнал или искажающего его информативные характеристики. Голосовые сигналы могут существенно искажаться за счет аппаратных помех, а также помех, обусловленных передачей по различным техническим каналам (например, телекоммуникационным). Нужно отметить, что обычно системы идентификации личности, предполагающие разграничение доступа по голосу, предназначены к использованию в идеализированных условиях малых помех и искажений. Влияние накладывающегося шума, роль и значение величины этого шума исследованы совершенно недостаточно. Практически не изучена, в частности, роль вида шума (например, речеподобных помех) в задачах идентификации. Итогом таких исследований должно явиться создание новых подходов, математических методов, способов и методик, обеспечивающих высокую надежность идентификации и достоверность ее результатов в условиях наложения шума и различных помех на обрабатываемые голосовые сигналы.

Отдельно нужно выделить проблемы, возникающие при идентификации в наиболее сложных условиях наложения друг на друга нескольких голосовых сигналов. В случаях, когда накладывающиеся голоса дикторов обладают близкими акустическими спектрами частот или таких голосов достаточно много, возможности идентификации резко снижаются. Подчеркнем, что исследования особенностей идентификации для этих сложных случаев практически отсутствуют. Как следствие, не разработаны подходы, призванные за счет новых методов математической обработки акустических голосовых сигналов повысить надежность результатов идентификации.

В соответствии с изложенным в настоящей работе были поставлены следующие основные **задачи**:

1. Выбрать наиболее информативные признаки акустических голосовых сигналов для применения математического аппарата многомерного анализа данных.

2. Разработать алгоритмы применения проекционных методов многомерного анализа, ориентированных на обработку голосовых сигналов.

3. Сформировать базу данных голосовых сигналов дикторов, предназначенную для тестирования систем голосовой идентификации личности.

4. Исследовать влияния искажений голосового сигнала, различных видов и величины внешнего шума при идентификации дикторов в реальных условиях.

5. Рассчитать ошибки первого и второго рода для оценки надежности разработанного подхода.

В качестве **объекта** исследования в работе выступает идентификация личности по голосовым данным. **Предметом** исследования являются акустические характеристики голосов дикторов.

ГЛАВА 2. ВЫБОР ИНФОРМАТИВНЫХ ГОЛОСОВЫХ ПРИЗНАКОВ ДЛЯ ПРОЕКЦИОННЫХ МЕТОДОВ АНАЛИЗА

2.1 Методы главных компонент и проекций на латентные структуры

Привлечение математического аппарата анализа многомерных данных обусловлено необходимостью повышения надежности голосовой идентификации личности, а именно, минимизацией влияния шума и искажений, а также разделением голосов дикторов с близкими характеристиками голоса.

Проекционные методы хорошо зарекомендовали себя при работе со спектральными данными и в задачах классификации [89–91]. В настоящей работе в качестве методов анализа многомерных данных используются метод главных компонент и метод проекций на латентные структуры. Другие методы относятся к области регрессионного анализа [92–94]. Общая идея проекционных методов [95,96] заключается в замене больших массивов экспериментальных данных их геометрическими проекциями на подпространство, которое ориентируется таким образом, чтобы выявить исследуемое свойство, тем самым удается восстановить латентные закономерности, существующие в исходных данных. При обработке голосов дикторов в качестве экспериментальных данных выступают характеристики частотных спектров соответствующих акустических сигналов, как правило, искаженных экспериментальными факторами и зашумленных. В качестве исследуемых свойств выступает принадлежность голосового сигнала (голосовых характеристик индивидуальности диктора) тому или иному диктору.

Для предварительного выявления из калибровочных данных скрытых закономерностей эффективен метод главных компонент, позволяющий

разделять эти данные на отдельные группы. Такой способ анализа и классификации (регрессия на главные компоненты) является распространенным в анализе многомерных данных. Опишем применение метода главных компонент к нашей задаче.

Пусть исходные данные составляют матрицу X размерностью nxp, где n – число «образцов», p – число переменных. В нашем случае n – число измеренных голосовых сигналов, p – число информативных характеристик, описывающих голосовые сигналы. Суть метода главных компонент в том, что выделяются вектора, направленные вдоль наибольших изменений (вариации) данных X в исходном многомерном пространстве. Первый вектор (первая главная компонента) направлен вдоль максимальной вариации, второй вектор (вторая главная компонента) – вдоль следующего по значению максимума вариации и т. д. Проводимая таким образом замена исходных данных их геометрическими проекциями на соответствующим образом ориентированные подпространства, позволяет разделить исходные данные на две части:

- структурная часть X' – данные, непосредственно связанные с исследуемым свойством,
- остатки (шум) E – данные, непосредственно с исследуемым свойством несвязанные.

Число главных компонент, описывающих основную вариацию данных, мало и гораздо меньше числа исходных переменных. В качестве наглядного инструмента, используемого в проекционных методах анализа и классификации, используются так называемые графики счетов и нагрузок. Счетами называют проекции исследуемых объектов («образцов») на подпространство главных компонент. График счетов описывает расположение счетов в этом пространстве. В качестве точки выступает

совокупность характеристик соответствующего объекта, в нашем случае в качестве точки выступает совокупность характеристик, описывающих голосовой сигнал. Близость точек на этом графике означает, что соответствующие объекты («образцы») имеют схожие характеристики. Например, принадлежат одному диктору или дикторам с похожими голосами. Соответственно точки, существенно удаленные друг от друга, описывают объекты, несвязанные между собой. На рис 2.1 приведен качественный пример использования метода главных компонент, где x_i, x_j – исходные экспериментальные данные, которые изображают двумерную проекцию исходных образцов, X_1, X_2 – исходные координаты, ГК1, ГК2 – направления первой главной компоненты, ориентированной вдоль максимальной вариации данных, и перпендикулярной к ней второй главной компоненты соответственно.

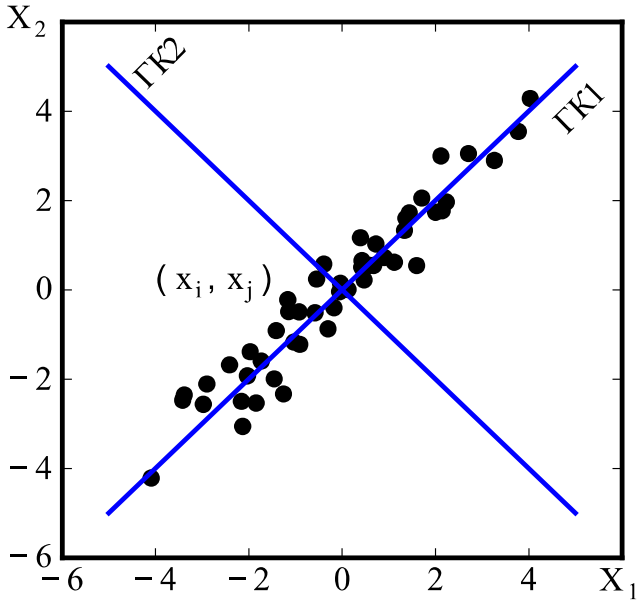


Рис. 2.1. Качественный пример использования метода главных компонент.

Кроме счетов, вводится понятие нагрузок, которые определяются как матрица перехода из исходного пространства переменных в пространство

главных компонент. Наглядность интерпретации усиливается при использовании так называемых графиков нагрузок, которые показывают взаимное расположение исходных переменных в пространстве главных компонент. Анализ этих графиков позволяет выяснить, какие переменные связаны, а какие независимы.

Преобразование метода главных компонент в матричном виде записывается следующим образом:

(2.1)

где T – счета, P – нагрузки (символ * означает транспонирование), они определяют пространство главных компонент, E – матрица остатков. Исходные данные представляются в виде некоторой оценки и остатков (шума). Формулу (2.1) можно пояснить с помощью рис. 2.2.

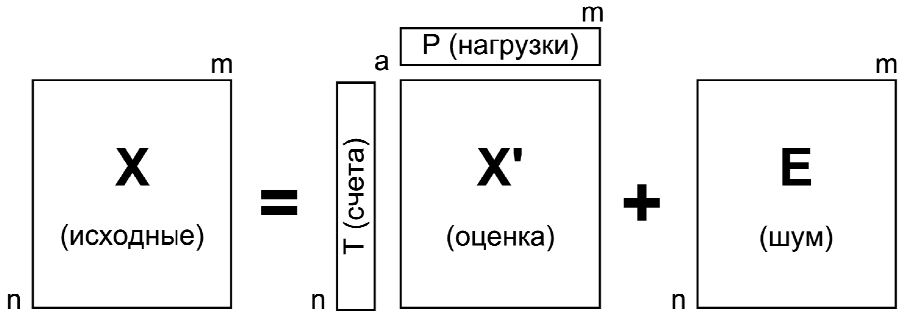


Рис. 2.2. Представление исходных данных в виде некоторой оценки и шума.

В случае, когда возможность учитывать априорные данные предсказанных переменных, более надежным является метод проекций на латентные структуры. При голосовой идентификации априорными данными являются данные о принадлежности отдельных речевых сигналов конкретному диктору. Другими словами, кроме матрицы X в методе проекций на латентные структуры используется матрица откликов Y размерностью nxk, где n – число «образцов», как в методе главных компонент, k – число принадлежностей тому или иному диктору. В нашем случае число k

определяется как априорно известное число соответствий того или иного сигнала конкретному диктору. В частности, если каждый диктор характеризуется одним голосовым сигналом, то k равняется числу дикторов. В матричной форме метод проекций на латентные структуры может быть описан выражением:

$$X = T \cdot P^* + E,\ \ Y = U \cdot Q^* + F,\ \ T = X \cdot W \cdot (P^* \cdot W)^{-1}, \qquad (2.2)$$

здесь P – матрица нагрузок данных (символ * означает транспонирование), T – матрица счетов, X – матрица данных (число столбцов равно числу признаков диктора, число строк равно количеству записей дикторов), Y – матрица откликов, матрицы U и Q – счета и нагрузки для Y-пространства, E и F – матрицы остатков для данных X и откликов Y соответственно, W – матрица взвешенных нагрузок. На рис. 2.3 изображено схематическое представление метода проекций на латентные структуры.

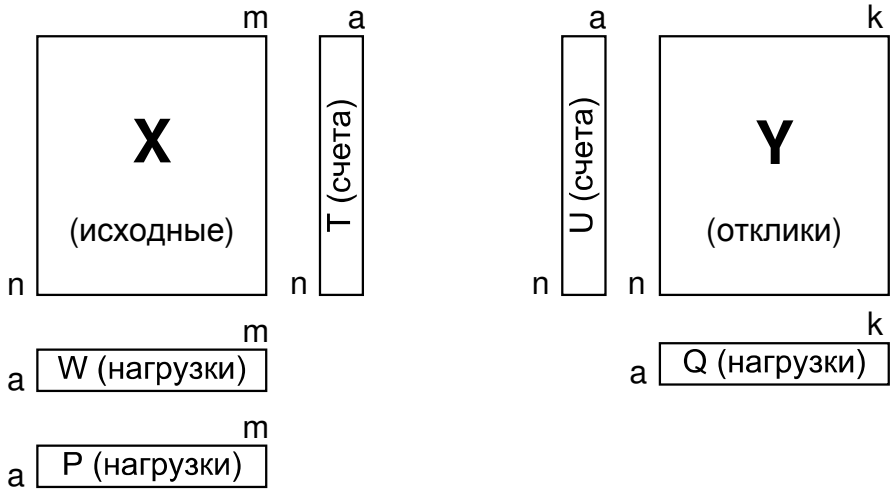


Рис. 2.3. Графическое пояснение метода проекций на латентные структуры.

Для предсказания Y на основе X необходимо предварительное получение калибровочных зависимостей между X и Y, которые строятся по известным априорным соотношениям между признаками и откликами. Соотношения матрицы X и соответствующей ей значения Y рассматриваются

как калибровочная или обучающая выборка. Для задачи идентификации личности эти соотношения означают калибровочную голосовую базу данных идентифицируемых дикторов. Идентификация неизвестных голосовых данных проводится с помощью известных соотношений между *X* и *Y*.

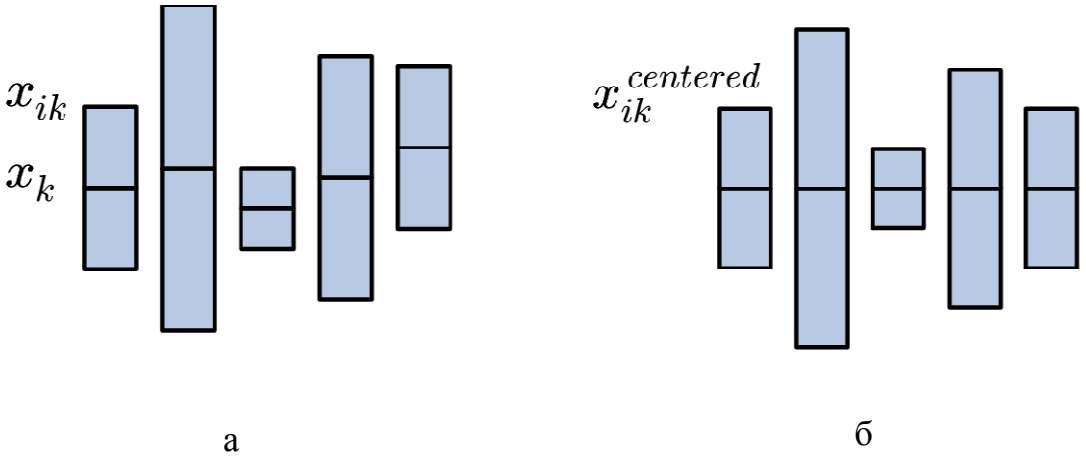
Число главных компонент, необходимых для классификации и анализа, может быть оценено с помощью так называемого графика остатков, который позволяет оценить разницу между исходными данными и пересчитанными. Величина остатка рассчитывается по формуле:

$$e_{tot}^2 = \sum_{i=1}^{n} e_i^2 \ , \ e_i^2 = \sum_{k=1}^{p} e_{ik}^2 \ , \qquad (2.3)$$

где – общая дисперсия остатков, – дисперсия остатков для образца *i*, *n* – число образцов, *p* – число переменных.

При обработке исходных данных в проекционных методах используются две процедуры: центрирование и шкалирование. Под **центрированием** понимается вычитание среднего по каждому признаку от соответствующего признака:

$$x_{ik}^{centered} = x_{ik} - x_k \ , \qquad (2.4)$$



а б

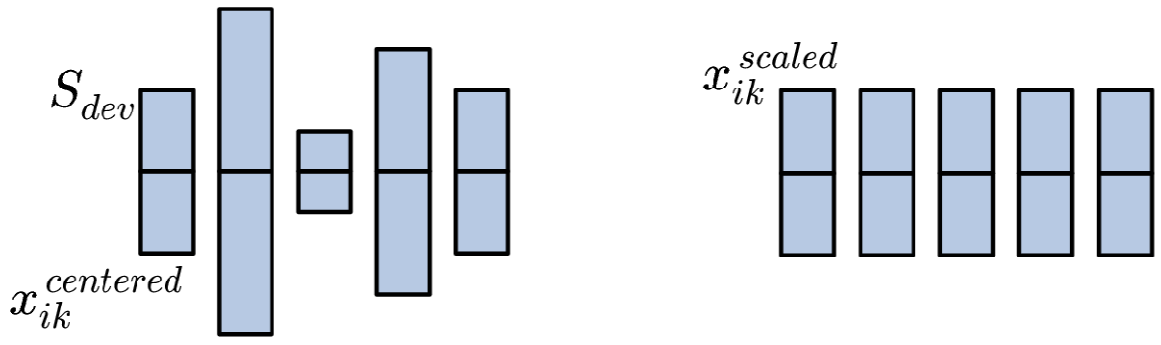
Рис. 2.4. Пояснение процедуры центрирования.
а – исходные данные, б – центрированные данные.

где x_{ik} – i-я переменная k-го образца, x_k – среднее значение k-го образца. Тем самым переменные образцов приводятся к нулевому среднему. На рис. 2.4 схематично пояснена процедура центрирования. Горизонтальная линия в середине прямоугольников означает среднее значение соответствующей переменной, сам вертикальный прямоугольник задает диапазон ее изменения.

При **шкалировании,** или взвешивании, все экспериментальные данные приводятся к одному масштабу. Это процедура необходима для снижения влияния на результат анализа сильно выраженных переменных. Существуют различные способы шкалирования [97], в настоящей работе был применена стандартизация, так как она наиболее изучена и апробирована. В стандартизации используется стандартное отклонение S_{dev}, которое является одним из наиболее часто используемых весовых коэффициентов. При этом каждый элемент матрицы X умножается на величину $\frac{1}{S_{dev}}$:

$$x_{ik}^{scaled} = x_{ik} \frac{1}{S_{dev}}, \tag{2.5}$$

где x_{ik} – i-я переменная k-го образца, x_{ik}^{scaled} – i-я шкалированная переменная



а б

Рис. 2.5. Пояснение процедуры шкалирования.
а – центрированные данные, б – шкалированные данные.

k-го образца, S_{dev} – стандартное отклонение образца.

На рис. 2.5 приводится схематичное пояснение процедуры шкалирования, где вертикальный прямоугольник (для случая шкалированных данных) означает нормированный на стандартное отклонение диапазон соответствующей переменной.

Для апробации методов необходимо провести **тестирование**, то есть проверку с помощью нового набора данных (тестового набора). Для тестового набора, как и для калибровочного массива, должны быть известны массивы Y. Процедура тестирования заключается в том, чтобы с помощью используемого проекционного метода предсказать значения Y для данного тестового набора и сравнить результат предсказания с известными данными.

2.2 Получение исходных голосовых характеристик

Первым этапом для голосовой идентификации личности является запись исходных голосовых сигналов дикторов. С позиций проекционных методов анализа это означает получение компонент матрицы признаков X как для калибровочной базы данных, так и для идентифицируемого голосового сигнала.

В задаче идентификации диктора не всегда используются высококачественные записи голосов, например, для фоноскопической экспертизы могут быть предоставлены звукозаписи телефонных переговоров с бытового записывающего плеера, где качество записи может быть невысоким. Качество звукозаписи обычно оценивают амплитудно-частотными характеристиками, коэффициентом нелинейных искажений, разборчивостью и другими. Для исследований голосовой идентификации необходимо иметь звукозаписывающую аппаратуру с запасом качества, которое можно использовать для оценки достоверности идентификации при искусственном наложении шума и искажений. Характеристики аппаратной

части звукозаписи определяются в основном качеством микрофона, усилителя низкой частоты и аналого-цифрового преобразователя.

Для записи голоса в настоящей работе использовался комплекс технических устройств, блок-схема которого изображена на рис. 2.6. Эта блок-схема включает в себя: микрофон 1, усилитель низкой частоты 2, аналого-цифровой преобразователь 3, программное обеспечение 4 и блок компьютера 5, обеспечивающий запись амплитудно-временных и расчет частотных зависимостей сигнала.

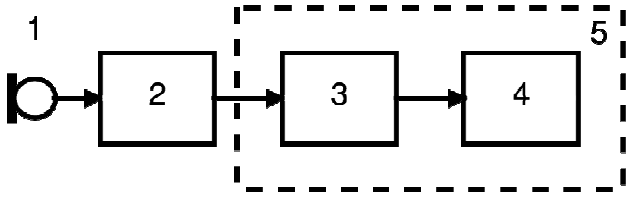


Рис. 2.6. Блок-схема аппаратной части системы идентификации личности по голосу.

В настоящей работе в качестве микрофона 1 использовался электретный микрофон типа МКЭ-3. В качестве усилителя низкой частоты 2 использовался усилитель комплекса виброакустических измерений «VNK-012GL», усиливающий слабый сигнал с микрофона до необходимого уровня в частотном диапазоне от 100 до 11200 Гц. Аналого-цифровой преобразователь 3 представлял из себя звуковую карту компьютера Creative «X-Fi Xtreme Audio» (параметры аналого-цифрового преобразователя: максимальная разрядность – 24 бит, максимальная частота – 96 кГц). Параметры записи для аналого-цифрового преобразователя следующие: битность отсчета – 16 бит, частота дискретизации – 44100 Гц. Запись сохранялась на компьютер в wav–формате. На начальном этапе использовался звуковой редактор Audacity [98].

Для проведения идентификации неизвестной голосовой записи была составлена и зарегистрирована **база данных дикторов** [99]. Используя вышеуказанную аппаратуру производилась запись голосовых сигналов, которые формировали базу данных.

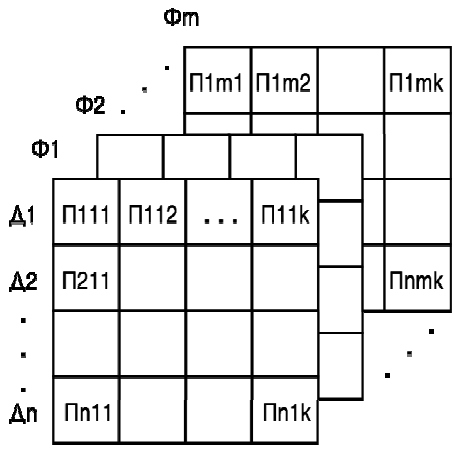
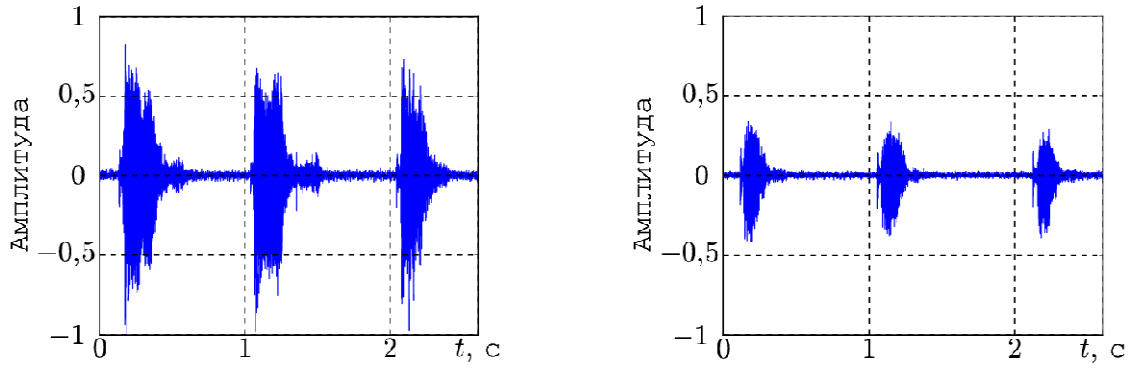


Рис. 2.7 Структура базы голосовых данных.

Структура базы данных изображена на рис. 2.7, где $П_{ijk}$ – повторение фразы, i – номер диктора, j – номер фразы, k – номер повторения фразы (длительность фразы) [100]. Каждая ячейка многомерной базы данных содержит одну запись. При построении голосовой базы данных и анализа голоса традиционно используются фонемы, представляющие из себя наименьший элемент речи. В качестве фраз на этапе разработки алгоритмов использовались короткие слова, представлявшие из себя сочетание цифр. В качестве дикторов выступали группы людей до 50 человек разного возраста и пола. Записи их голосов проводились в различных условиях. Для составления калибровочной базы данных использовались в основном фразы, представляющие из себя цифры от 0 до 9. Такого набора достаточно [101], чтобы охватить основные информативные фонемы, участвующие в узнавании голосов.

Запись фраз при формировании базы данных на компьютер производилась в реальном режиме времени. Типичные сигналы амплитудно-временной зависимости *a(t)* изображены на рис. 2.8 при разной громкости, а на рис. 2.9 при разных темпах речи (изображены записи одного диктора, который произносил слово «пять» три раза). Рис. 2.8 а относится к записи с

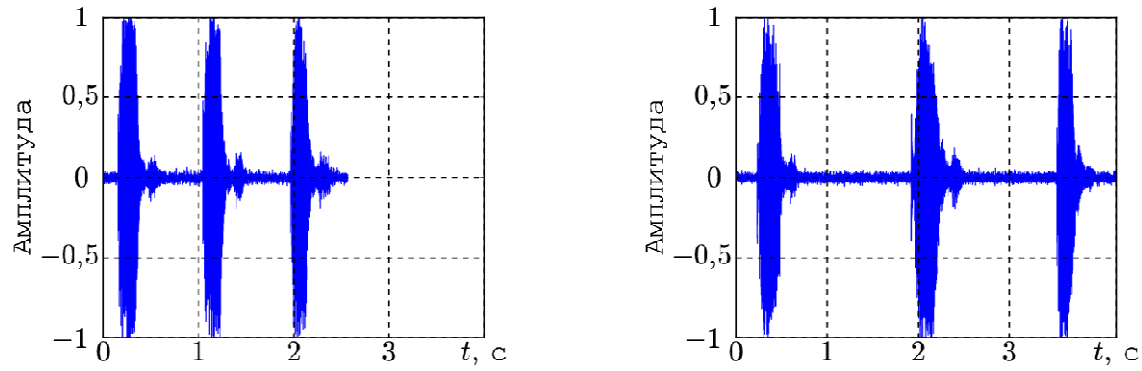


а б

Рис. 2.8. Голосовые сигналы одного диктора.
а – обычная громкость, б – низкая громкость.

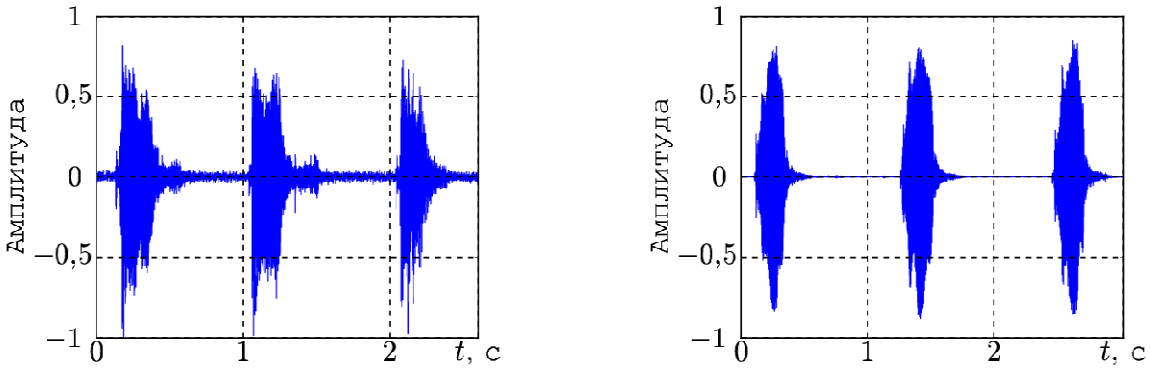
относительно высокой громкостью, а рис. 2.8 б – с громкостью в два раза ниже (при одинаковом темпе речи). На рис. 2.9 а темп речи выше, то есть слово произносилось быстрее, чем 2.9 б (при одинаковой громкости). Как видно из рис. 2.8 и рис. 2.9, исходные данные зависят от амплитуды и продолжительности сигнала (темпа речи).

Идентификация диктора зависит от вида индивидуальных особенностей голоса, прежде всего – от его частотных характеристик, совокупность которых определяет тембр голоса. В качестве примера на рис. 2.10 приведены записи голосов мужчины (рис. 2.10 а) и женщины (рис. 2.10 б).



а б

Рис. 2.9. Голосовые сигналы одного диктора.
а – обычнй темп речи, б – медленный темп речи.



а б

Рис. 2.10. Примеры записей дикторов.
а – мужской голос, б – женский голос.

Для решения задачи идентификации личности основным представляется анализ индивидуального частотного спектра голосовых сигналов. В такой постановке первые два фактора (амплитуда и продолжительность сигнала) являются случайными и от них нужно избавиться. Для этого все сигналы приводились к одной амплитуде, то есть производилось нормирование по амплитуде:

$$a_i^{norm} = \frac{a_i}{a_{\max}} \tag{2.6}$$

где a_i – измеряемая амплитуда, a_{max} – максимальная амплитуда, a_i^{norm} – нормированная амплитуд, i=0,1,…,k.

Для того, чтобы избавиться от второго фактора (темпа речи) производилась нормировка по времени. Учет второго фактора производился программно путем использования одинакового числа отсчетов. Непосредственно анализировались амплитудно-частотные характеристики в виде спектра записанного звукового сигнала. Частотные спектры имели вид зависимости амплитуды A от частоты f. Спектры рассчитывались с помощью стандартной функции программной среды MatLab [102]. Типичные спектры голосовых сигналов и нормированный спектр по амплитуде и времени изображены на рис. 2.11 для одного диктора с разными громкостью и темпом

речи. На рис. 2.11 а, б сопоставляются спектральные разложения сигналов с разными уровнями громкости, соответствующие рис. 2.8, а на рис. 2.11 а, в сопоставлены спектральные разложения сигналов с разными темпами речи, соответствующие рис. 2.9. На рис. 2.11 г приведено спектральное разложение, нормированное по амплитуде и времени.

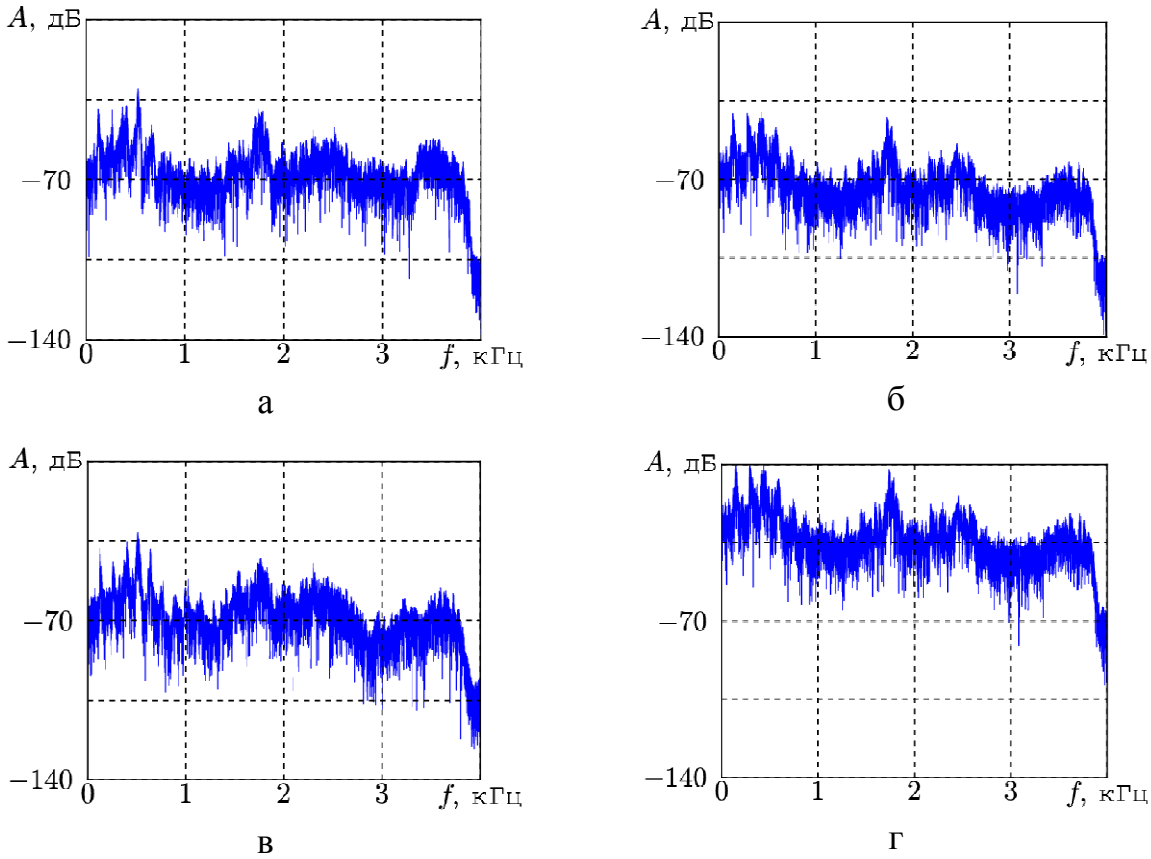


Рис. 2.11. Спектры голосовых сигналов одного диктора.
а – при обычной громкости, б – при низкой громкости,
в – при медленном темпе речи, г – нормированный спектр.

При записи голосовых сигналов в реальных условиях возможно наложение случайных факторов, включающих в себя как внешние механические шумы, так и аппаратные помехи. Для их подавления использовалась медианная фильтрация [103], заключавшаяся в исключении из исходного сигнала выбросов. В настоящей работе использовалась стандартная

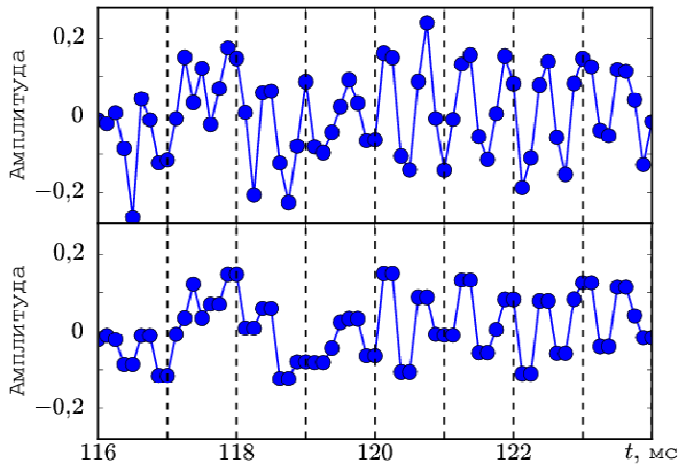


Рис. 2.12. Пример действия медианного фильтра третьего порядка.
а – исходный сигнал с выбросами, б – выбросы заменены
медианными значениями.

обработка в среде Matlab. На рис. 2.12 показан пример действия медианного фильтра третьего порядка.

Кроме того, отдельно исследовался вопрос о соотношении «полезный сигнал шум». На стадии отработки метода и составлении базы данных голосовые сигналы записывались в специальных стационарных условиях

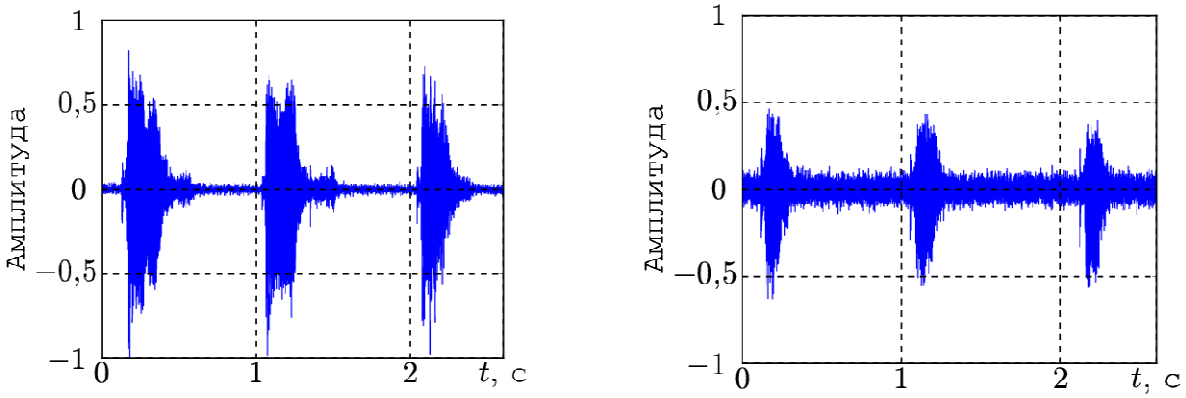


Рис. 2.13. Голосовые сигналы одного диктора с разными пороговыми значениями отношения «полезный сигнал шум».
а – уровень шума мал, б – уровень шума значимый.

изолированного помещения, при которых влияние шумов было практически исключено. На рис. 2.13 для иллюстрации приведены сигналы с существенно разными значениями отношения «полезный сигнал шум».

2.3 Выбор информативных голосовых признаков

Для применения методов анализа многомерных данных к задаче идентификации личности по голосу необходимо выбрать соответствующие характеристики, которые выполняют роль признаков, заполняющих матрицу *X*, введенную в разделе 2.2. Одним из основных требований для этих признаков является максимальная информативность, что должно обеспечить эффективность результата. В настоящем разделе осуществлен выбор признаков для исходных сигналов.

Отсутствие в настоящее время четкой систематизации голосовых признаков, а также существование большого количества голосовых характеристик различного уровня, например, таких, как основной тон [104], формантные частоты [105,106] и другие [107–109], представляет определенную сложность в выборе наиболее информативных признаков и характеристик для конкретного метода идентификации и требует отдельного исследования. В данном разделе приведены качественные и количественные оценки для выбора информативных голосовых характеристик.

Различие в тембрах голосов описывается разными частотными спектрами голосовых сигналов. Естественным математическим аппаратом для анализа частотного спектра является Фурье-разложение. Обработка данных, представляющих из себя численные амплитудно-временные зависимости, может быть проведена с помощью дискретного преобразования Фурье:

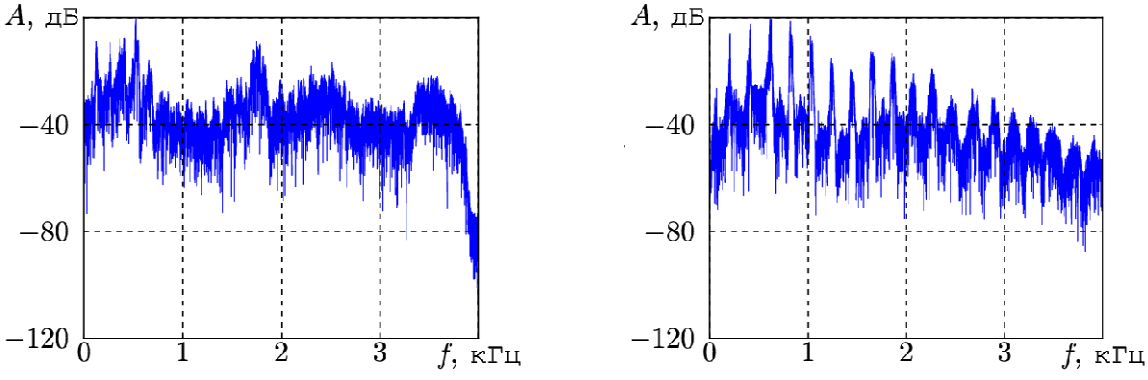
$$A(k) = \sum_{n=0}^{N-1} a(n) \cdot e^{-j2\pi\frac{k}{N}n}, \quad k = 0,1,2,...,N-1 \tag{2.7}$$

$$a(n) = \frac{1}{N}\sum_{k=0}^{N-1} A(k) \cdot e^{j2\pi\frac{k}{N}n}, \quad n = 0,1,2,...,N-1 \tag{2.8}$$

где $A(k)$, $a(n)$ – прямое и обратное дискретные преобразования Фурье соответственно, k и n – номера отсчетов, N – число отсчетов. Именно коэффициенты $A(k)$ могут быть использованы в качестве элементов матрицы X, формируя строки в этой матрице.

На рис. 2.14 приведены типичные спектры для записей голосов двух дикторов (мужчины и женщины), проведенных в соответствии с предыдущим разделом 2.2. Расчеты произведены на основе дискретного преобразования Фурье с помощью программы MatLab. По вертикальной оси отложена относительная амплитуда A (нормированная на максимальное значение амплитуды A_{max}), а по горизонтальной – частота f. Видно, что основной пик спектра (основной тон, первая форманта) диктора приходится на промежуток частот от 200 до 500 Гц. Следующие пики в диапазоне от 500 до 5000 Гц соответствуют обертонам. Другими словами, первая форманта лежит в диапазоне от 200 до 500 Гц, а последующие форманты в диапазоне от 500 до 5000 Гц. Форманты, описывающие обертоны, отличаются по форме пиков и набору частот (рис. 2.14 а и рис. 2.14 б). Это означает, что Фурье–разложение, вообще говоря отражает индивидуальные особенности голоса.

В то же время Фурье–разложение не достаточно эффективно разделяет



а б

Рис. 2.14. Типичные спектры дискретного преобразования Фурье.
а – спектр для мужского голоса, б – спектр для женского голоса.

голоса дикторов, поскольку различие между ними определяется набором дальних гармоник, вклад которых относительно мал и на которых сказываются случайные искажения сигнала [110]. Более предпочтительным для настоящей задачи представляется использование разложения, усиливающего относительную роль особенностей исходных сигналов, связанных с дальними гармониками, и нивелирующего вклад случайных искажений. В качестве такого разложения может быть использовано мел-кепстральное разложение [111,112], учитывающее нелинейные свойства человеческого слуха. В мел-кепстральном разложении Фурье-разложение заменяется логарифмическими зависимостями, при этом более успешно выделяются наиболее информативные для голосовой идентификации особенности сигнала, что обусловливает преимущество коэффициентов этого разложения по сравнению с коэффициентами Фурье [57].

Применение мел-кепстрального преобразования для получения голосовых признаков, которые используются в современных системах распознавания диктора, включает в себя два ключевых преобразования – мел-частотное и кепстральное.

Мел-частотное преобразование является нелинейным преобразованием частоты. Оно позволяет учитывать свойства человеческого слуха, связанные с нелинейным восприятием звуковых частот, а именно, связанные с наличием так называемых критических полос слуха [113], в которых сигналы с любой частотой из критической полосы неразличимы. В работах [57,114,115] показано, что применение мел-частот позволяет повысить надежность распознавания дикторов. Типичный вид рассчитанного нами мел-частотного преобразования представлен на рис. 2.15.

Кепстральное преобразование включает в себя несколько этапов, среди которых основными являются логарифмирование амплитудно-частотного спектра и вычисление обратного косинусного преобразования.

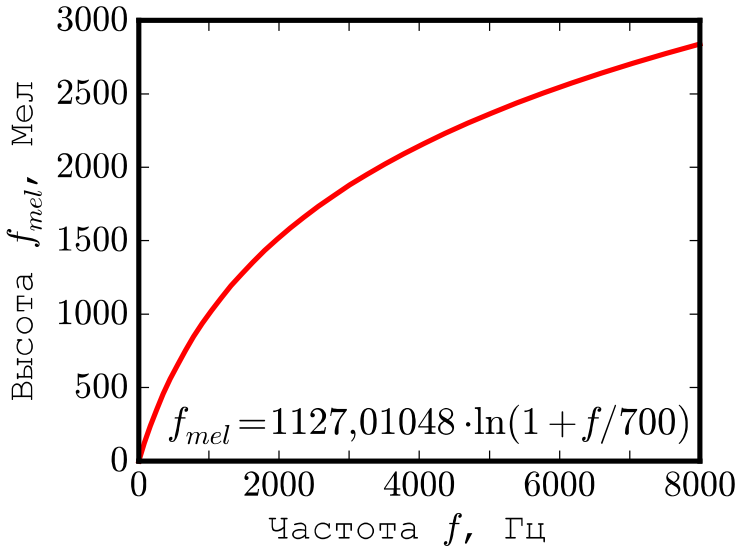


Рис. 2.15. Вид зависимости мел-частотного преобразования.

Получение мел-кепстральных коэффициентов в данной работе производилось следующим образом. Сначала вычислялись коэффициенты Фурье-разложения *A(k)* по формуле (2.7). Далее к амплитудно-частотному спектру |*A(k)*| применялся банк из 20 треугольных фильтров, которые обеспечивали наибольший вес центральным коэффициентам каждого фильтра. Треугольные фильтры рассчитывались по следующим формулам:

$$H_m(n) = \begin{cases} 0, & f(n) < f_b(m-1) \\ \dfrac{f(n) - f_b(m-1)}{f_b(m) - f_b(m-1)}, & f_b(m-1) < f(n) < f_b(m) \\ \dfrac{f(m+1) - f_b(n)}{f_b(m+1) - f_b(m)}, & f_b(m) < f(n) < f_b(m+1) \\ 0, & f(n) > f_b(m+1) \end{cases} \quad (2.9)$$

где *f(n)* – частота, соответствующая определенной гармонике, *m* – порядковый номер фильтра, f_b – частоты, соответствующие границам интервалов линейного диапазона частот. Из рассчитанных треугольных фильтров формировался банк фильтров. На рис. 2.16 приведен типичный вид рассчитанного нами банка 20 треугольных фильтров.

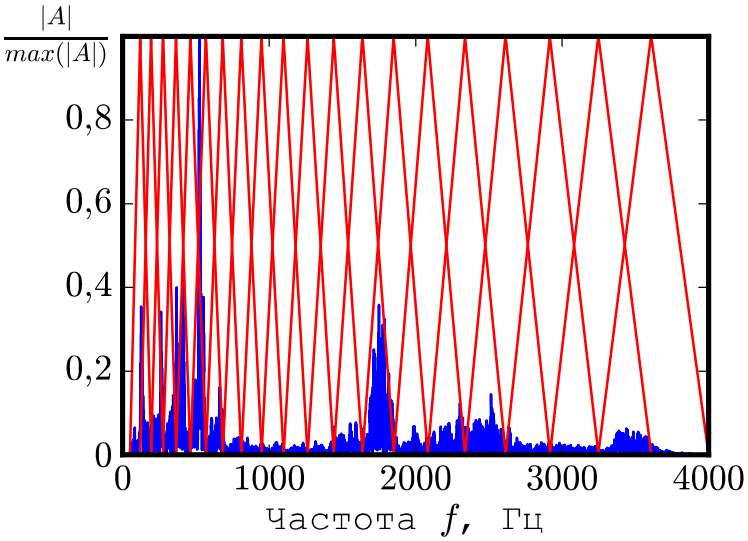


Рис. 2.16. Вид банка 20 треугольных фильтров.

Полосы частот фильтров соответствовали полосам мел-частот, задаваемых путем равномерного разбиения мел-частотного диапазона на участки и преобразования этих мел-частотных участков обратно в линейные участки частот. Таким образом, треугольные фильтры располагались нелинейно во всем диапазоне линейных частот (рис. 2.16). После этого для каждого треугольного фильтра вычислялась логарифмическая энергия:

$$B(m) = \ln\left(\sum_{k=0}^{N-1} |A(k)|^2 H_m(k)\right), \quad m = 0,1,2,\ldots,I-1 \qquad (2.10)$$

где $B(m)$ – логарифмическая энергия участка частот спектра (рис. 2.17), $H_m(k)$ – треугольный фильтр, соответствующий m-ой мел-частотной полосе, m – порядковый номер треугольного фильтра, I – задаваемое количество треугольных фильтров ($I = 20$).

На последнем шаге вычислялось обратное дискретное косинусное преобразование:

$$C(l) = \sum_{m=0}^{I-1} B(m) \cdot \cos\left(\frac{\pi}{I}\left(m + \frac{1}{2}\right)l\right), \quad l = 0,1,2,\ldots,I-1 \qquad (2.11)$$

где *C(l)* – коэффициенты мел-кепстрального разложения, *l* – порядковый номер мел-кепстрального коэффициента.

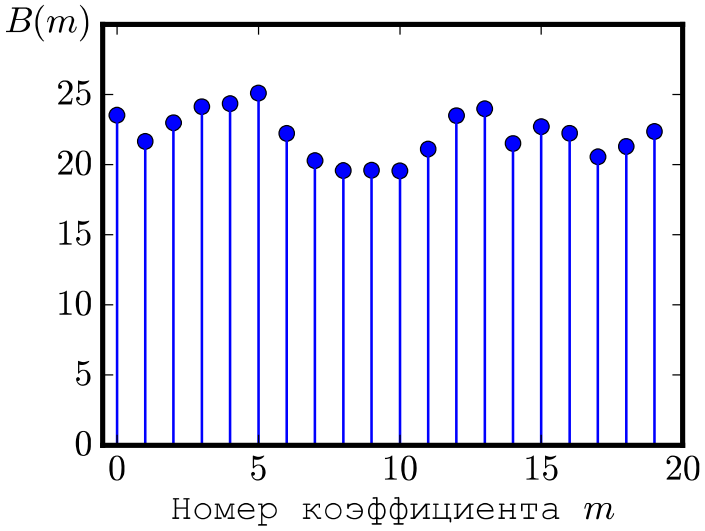


Рис. 2.17. Типичный вид коэффициентов логарифмической энергии частотных участков спектра.

Полученные таким образом коэффициенты *C(l)* использовались далее в качестве элементов в матрице признаков *X*, при этом номер коэффициента *l* задает номер столбца, а каждый сигнал, описываемый своим мел-кепстральным разложение *C(l),* задает номер строки, число которых определяется числом голосовых сигналов.

Полученные коэффициенты мел-кепстрального разложения *C(l)* в зависимости от номера фильтра *l,* харакетризующие особенности голоса диктора в разных частотных диапазонах, приведены на рис 2.18. Отметим, что это разложение не допускает такой наглядной интерпретации, как Фурье–разложение.

Для выбора признаков, обеспечивающих наилучшую разделимость голосов дикторов, были проведены расчеты остаточных дисперсий в методе проекций на латентные структуры с разными возможными информативными характеристикми: коэффициентами фурье-разложения, мел-частотными

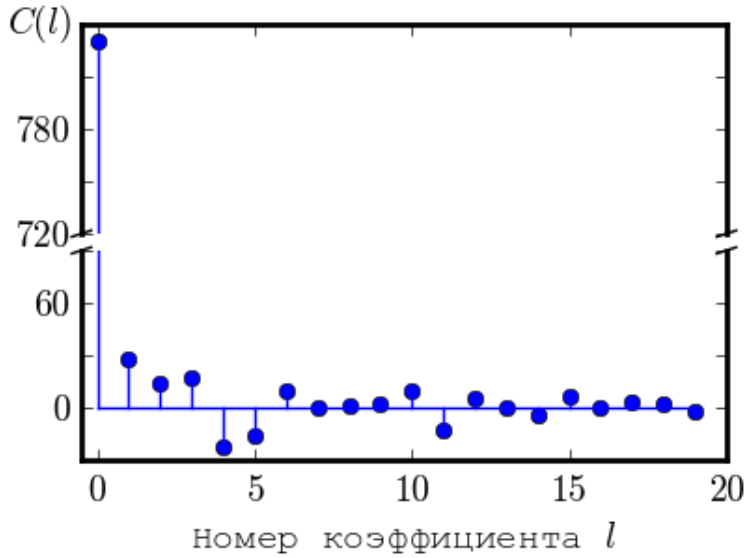


Рис. 2.18. Типичный вид мел-кепстральных коэффициентов.

полосами преобразования Фурье, коэффициентами кепстрального разложения, а также коэффициентами мел-кепстрального разложения. С этой целью рассчитывались матрицы остатков:

$$E_C = Y_C - Y_C', \qquad E_T = Y_T - Y_T' \tag{2.12}$$

где E_C и E_T – матрицы остатков, Y_C *и* Y_T – исходные матрицы откликов, Y_C' *и* Y_c' – матрицы откликов при декомпозиции в случае калибровки и тестирования соответственно. Остаточные дисперсии вычислялись по следующим формулам:

$$D_C = \frac{1}{I_C J} \sum_{i=1}^{I_C} \sum_{j=1}^{J} e_{C\,ij}^2 \tag{2.13}$$

где D_C – дисперсия остатков образцов при калибровке, $e_{C\,ij}^2$ – элемент матрицы E_C, I_C – число образцов при калибровке, J – число переменных (коэффициентов какого-либо преобразования);

$$D_T = \frac{1}{I_T J} \sum_{i=1}^{I_T} \sum_{j=1}^{J} e_{T\,ij}^2 \qquad (2.14)$$

здесь величины D_T, $e_{T\,ij}^2$, I_T, J описываются аналогично величинам формулы (2.13), индекс T указывает на тестирование.

Конкретные расчеты проводились для записей голосов дикторов, произносивших одну и ту же фразу (цифры от 0 до 9). Для голосовых сигналов составлялись наборы векторов признаков из коэффициентов соответствующих преобразований. Полученные результаты приведены на рис. 2.19. Значение остаточной дисперсии при номере главной компоненты,

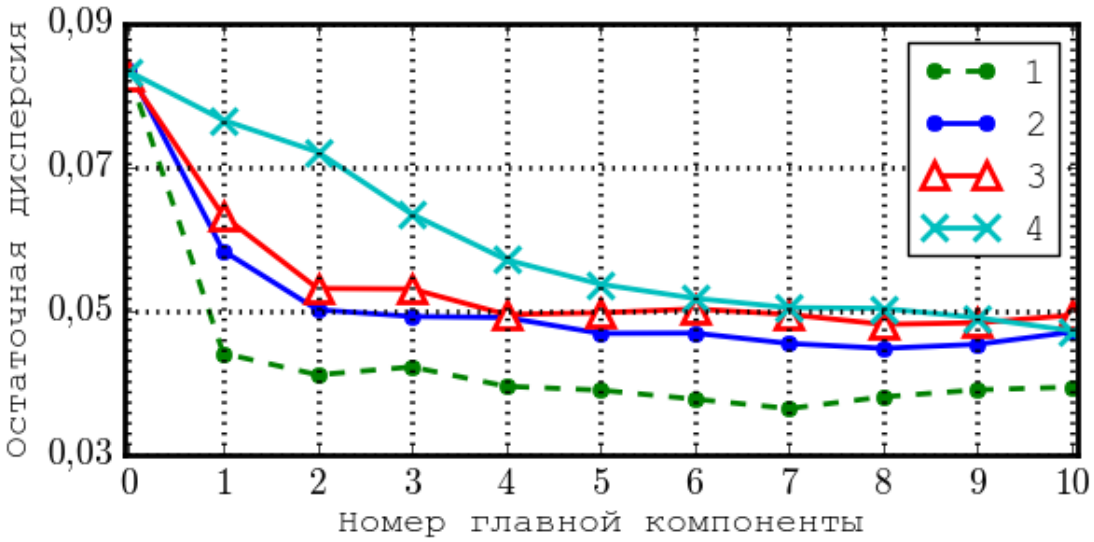


Рис. 2.19. Зависимость величины остаточной дисперсии от числа главных компонент.
1 – мел-кепстральные коэффициенты, 2 – кепстральные коэффициенты, 3 – мел-частотные спектральные коэффициенты, 4 – спектральные коэффициенты.

равном нулю (рис. 2.19), означает, что вычисляется общая остаточная дисперсия, которая представляет собой центрированную матрицу X (исходные данные). Из рис. 2.19 следует, что мел-кепстральные коэффициенты в первых 10 главных компонентах (кривая 1) обладают меньшей дисперсией по сравнению с кепстральными коэффициентами, мел-

частотными спектральными коэффициентами и обычными спектральными коэффициентами (кривые 2–4), что свидетельствует о превосходстве мел-кепстральных коэффициентов. Кроме того, мел-кепстральные коэффициенты раньше достигают минимума остаточной дисперсии, то есть необходимых компонент для описания голосового сигнала требуется меньше.Это означает, что мел-кепстральные коэффициенты являются наболее информативными признаками для используемых проекционных методов, обеспечивающими лучшую разделимость голосовых сигналов. Расчеты коэффициентов корреляции между значениями откликов при калибровке и тестировании, проведенные в работе [116], а также данные работы [57] привели к аналогичным выводам.

ГЛАВА 3. АНАЛИЗ И КЛАССИФИКАЦИЯ ГОЛОСОВЫХ ДАННЫХ

3.1 Разработка алгоритмов голосовой идентификации дикторов

Голосовая идентификация дикторов представляет из себя процесс, состоящий из ряда последовательных этапов. Реализация этих этапов требует построения соответствующих алгоритмов. В настоящем разделе рассматриваются алгоритмы идентификации, в которых используются предложенные в предыдущей главе методы получения, обработки и анализа голосовых сигналов.

Разработанный алгоритм идентификации, обеспечивающий получение калибровочной базы данных голосов дикторов, изображен на верхней части рис. 3.1. Аналогичный алгоритм для тестирования голосовых данных представлен на нижней части этого же рисунка.

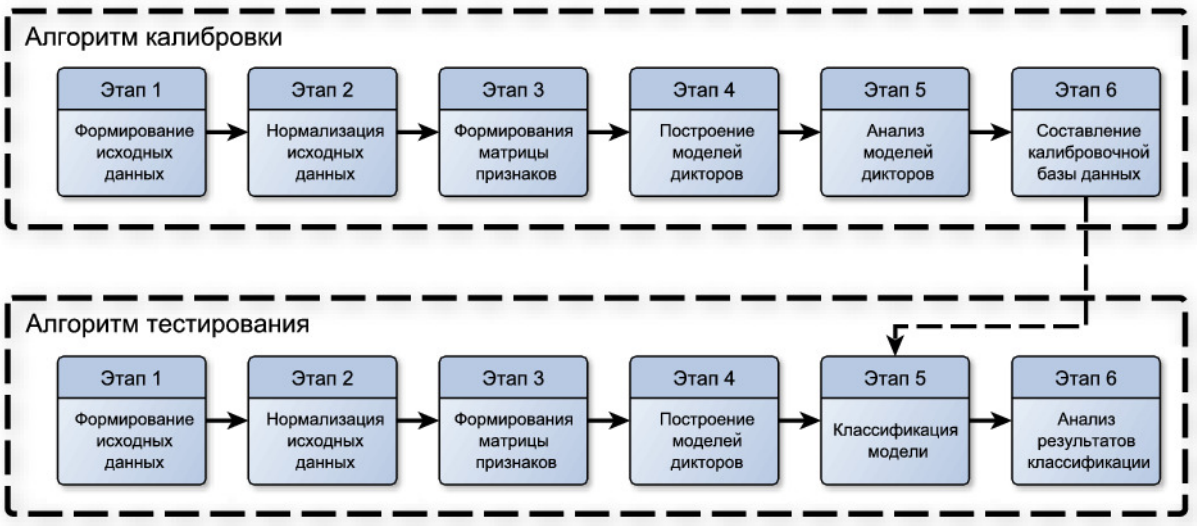


Рис. 3.1. Схема алгоритмов калибровки и тестирования.

Алгоритм калибровки включает в себя следующие этапы:

• Этап 1 – формирование исходных данных по записанным голосовым сигналам путем их записи, усиления и предварительной обработки, создание базы записей голосов дикторов.

• Этап 2 – нормализация исходных данных путем центрирования, шкалирования и других процедур.

• Этап 3 – формирование матрицы признаков *X*.

• Этап 4 – применение конкретного проекционного метода для построения моделей дикторов (получение графиков счетов и нагрузок в первых главных компонентах).

• Этап 5 – анализ и интерпретация моделей дикторов с помощью конкретного метода классификации и получение набора калибровочных данных для идентификации дикторов.

• Этап 6 – составление калибровочной базы данных.

Для процедуры тестирования, то есть идентификации неизвестной голосовой записи с помощью полученной калибровочной базы данных, последовательность этапов алгоритма следующая.

• Этап 1 – Формирование исходных данных по идентифицируемым записям голосовых сигналов путем их записи, усиления и предварительной обработки.

• Этап 2 – нормализация исходных данных, центрирование, шкалирование и другие процедур.

• Этап 3 – формирование матрицы признаков *X*.

• Этап 4 – применение конкретного проекционного метода для построения модели идентифицируемого диктора (получение графиков счетов и нагрузок).

• Этап 5 – классификация модели идентифицируемого диктора с калибровочным набором данных.

• Этап 6 – анализ результатов классификации и расчет ошибок классификации.

Аппаратно–программная реализация этапа 1 подробно описана в разделе 2.2. Для удобства предварительной обработки данных специально составлялась программа в среде Matlab, обеспечивающая быстрый доступ к файлам по требуемому списку. Составленная нами база данных [91] представляла из себя микрофонные записи длительностью от долей до свыше 10 секунд. Записи представляли из себя наборы слов, содержащих основные фонемы, в соответствии с разделом 2.2.

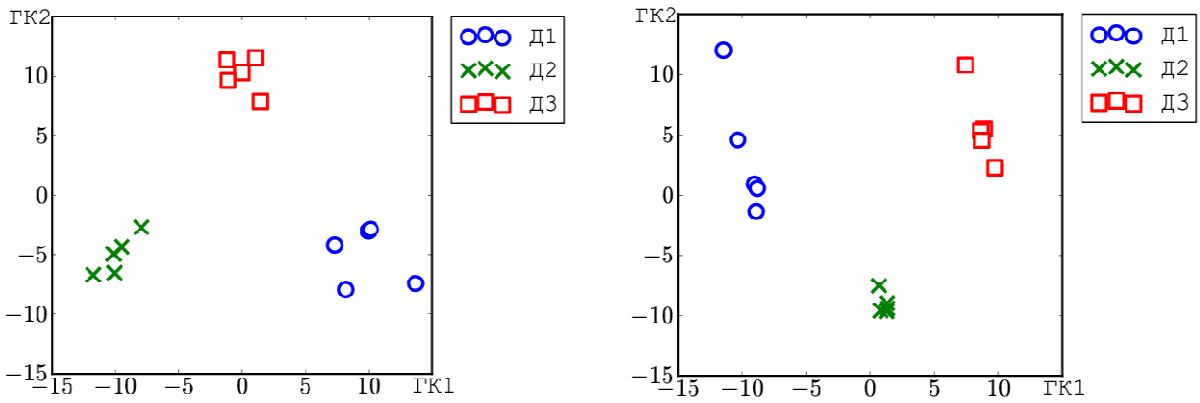
При реализации этапов 2, 3, 4 использовалась специально составленная программа «Классификация голосовых сигналов для разграничения доступа к конфиденциальной информации», получившая свидетельство о государственной регистрации программы для электронной вычислительной машины [117].

На этапе 5 проводился выбор методов классификации, позволявший наиболее эффективно сопоставить результаты расчетов в пространстве главных компонент с конкретным диктором [118,119]. В качестве указанных методов использовались как ранее известные, именно, метод ближайших соседей, так и специально разработанные новые методы. Программное обеспечение этапа разрабатывалось отдельно для каждого метода классификации.

Этап 6 заключался в получении базы данных в виде матрицы, в которой компоненты многомерных векторов признаков соотносились с соответствующими дикторами. Эта база данных использовалась при идентификации «неизвестных» дикторов.

В случае процедуры тестирования, на этапе 1, дополнительно использовалась идентификационная запись «неизвестного» диктора произвольной продолжительности и содержания. Этапы 2, 3, 4

реализовывались с помощью указанной выше программы «Классификация голосовых сигналов для разграничения доступа к конфиденциальной информации». Этап 5 при тестировании являлся основным этапом, при выполнении которого проводилось сопоставление «неизвестной» голосовой записи конкретного диктора с записями из калибровочной базы данных. Этап 6 включал в себя расчет ошибок 1-го и 2-го рода и позволял оценить надежность идентификации.

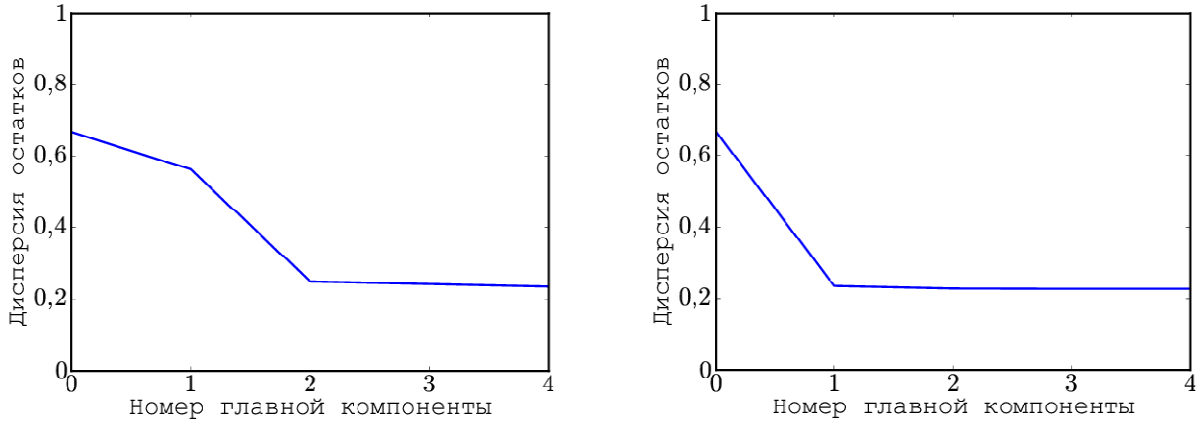


а б

Рис. 3.2. График счетов для трех дикторов.

а – метод главных компонент, б – метод проекций на латентные структуры.

Предварительная апробация предложенного алгоритма проводилась для двух проекционных методов – главных компонент и проекций на латентные структуры. Типичные результаты расчета на примере 3 дикторов, каждый из которых произносил пять раз слово «пять», приведены на рис. 3.2 в виде графика счетов в пространстве двух первых главных компонент (обозначения Д1, Д2, Д3 здесь и далее означают точки , описывающие голосовые сигналы соответствующих дикторов). Каждая из точек отражает одну запись одного диктора (записи одного и того же диктора обозначены одинаковыми значками).



а б

Рис. 3.3. График дисперсий остатков.

а – метод главных компонент, б – метод проекций на латентные структуры.

Как видно из рис. 3.2, для обоих проекционных методов все точки объединены в компактные области не пересекающиеся друг с другом, каждая область соответствует записям одного диктора. Это свидетельствует о том, что оба метода обеспечили надежное разделение дикторов по голосовым данным.

Отметим, что непосредственное сопоставление приведенных графиков счетов для методов главных компонент (рис. 3.2 а) и проекций на латентные структуры (рис. 3.2 б) не позволяет количественно оценить эти варианта расчета. Преимущество одного метода над другим можно оценить по графику дисперсии остатков. Дисперсии остатков рассчитывались по формуле (2.3), полученные данные приведены на рис. 3.3. Эти данные показывают, сколько необходимо рассчитать главных компонент, чтобы модель описывала существенную часть исходных данных. На рис. 3.3 а изображена дисперсия остатков для метода главных компонент, а на рис. 3.3 б – для метода проекций на латентные структуры. Излом на графике дисперсии остатков указывает необходимое количество главных компонент. Видно, что определенное преимущество имеет метод проекций на латентные структуры [120], так как соответствующая модель описывает существенную часть

данных одной главной компонентой, в то время как метод главных компонент – описывает двумя главными компонентами.

Отметим, что метод проекций на латентные структуры может работать на множестве калибровочных объектов только с известными откликами, тогда как метод главных компонент в общем случае – без откликов. Это означает, что при некотором множестве неизвестных голосовых записей метод проекций на латентные структуры не обеспечит построение модели. Поэтому, несмотря на указанное выше преимущество метода проекций на латентные структуры, далее в работе будет использоваться в основном метод главных компонент.

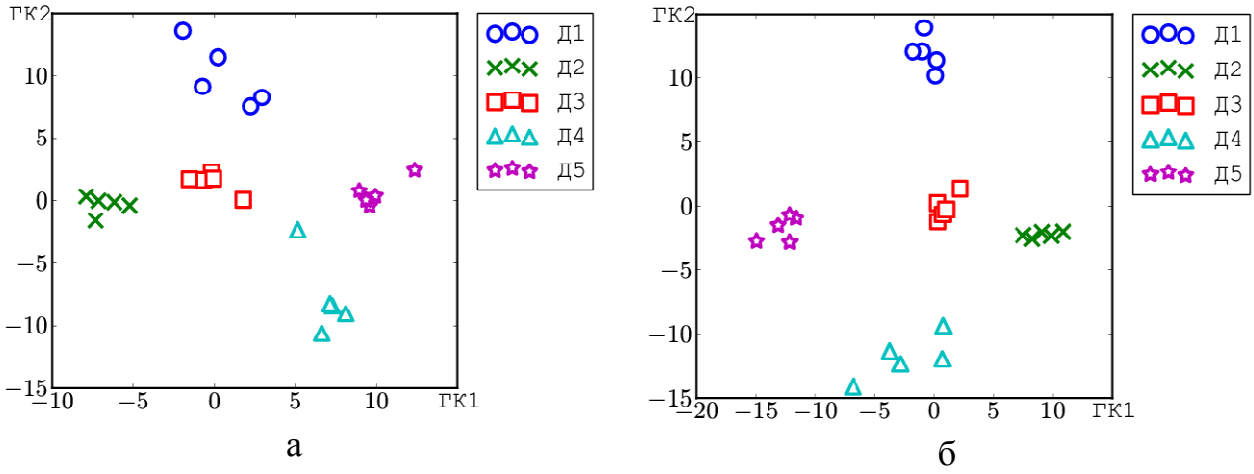


Рис. 3.4. Влияние длительности фразы на голосовую идентификацию.
а – слово «пять», б – фраза «один два три четыре пять».

Для исследования возможностей предложенного подхода было изучено влияние длительности фразы на голосовую идентификацию. Для этого использовались голосовые записи с разными фразами из разработанной базы данных. На рис. 3.4 сопоставлены рассчитанные графики счетов для двух случаев, в первом (рис. 3.4 а) в качестве калибровочных данных использовались известные голосовые сигналы пяти дикторов, произносивших короткую фразу (слово «пять») пять раз, во втором (рис. 3.4 б) эти же дикторы произносили фразу большей длительности: «ноль один два три четыре». Как

видно из рис. 3.4 влияние количества данных на голосовую идентификацию значительны. В случае более длиной фразы области, соответствующие разным дикторам, расположены компактнее и отстоят дальше друг от друга. Это означает, что увеличения длины фразы (голосового пароля) из базы данных увеличивает надежность идентификации.

3.2 Анализ голосовых сигналов на основе метода ближайших соседей

Для практической реализации изложенного в предыдущем разделе алгоритма идентификации необходима разработка количественных методов, позволяющих соотнести неизвестные голосовые сигналы с тем или иным диктором. Такие методы должны допускать удобную математическую и программную реализацию и достаточно высокую наглядность анализа графиков счетов.

Для преодоления субъективного фактора анализа графика счетов и автоматизации процесса идентификации личности по голосу представляется эффективным привлечение метода, обеспечивающего количественную оценку принадлежности точки своему классу, то есть фразы – своему диктору. В настоящей работе для этого используется метод ближайших соседей (метод *k* ближайших соседей). Суть этого метода состоит в классификации неизвестного объекта по количеству расположенных рядом известных объектов (соседей).

В нашей задаче в качестве неизвестного объекта выступают голосовые сигналы неизвестного диктора, а в качестве ближайших соседей – сигналы известных дикторов, взятые из калибровочной базы данных. Идентификация заключается в соотнесении на графике точек, изображающих неизвестные сигналы, с тем диктором, голосовые сигналы которого изображаются наиболее близкими точками.

Проведем математическое описание метода ближайших соседей [121]. Пусть $X \in R^n$ – множество объектов (голосовых сигналов), n – число переменных, в качестве которых выступают коэффициенты кепстральных разложений; Y – множество допустимых откликов, то есть принадлежностей конкретному диктору. Пусть известна калибровочная выборка $\{(x_i, y_i)\}_{i=1}^{l}$, представленная калибровочной базой голосовых данных, здесь l – количество записей. Задано множество объектов $X^m = \{x_i\}_{i=1}^{m}$ (m – количество записей тестовой выборки), в качестве которых выступают голосовые сигналы неизвестных дикторов (одного или нескольких). Требуется найти множество откликов $\{y_i\}_{i=1}^{m}$, то есть соотнести неизвестные сигналы с известными дикторами для объектов $\{x_i\}_{i=1}^{m}$.

Введем многомерное расстояние Евклида между объектами которое выражается координатами x и x':

$$\rho(x, x') = \sqrt{\sum_{i=1}^{n} (x_i - x'_i)^2} . \qquad (3.1)$$

Для произвольного объекта $x \in \{x_i\}_{i=1}^{m}$ расположим объекты обучающей выборки x_i в порядке возрастания расстояний до объекта с координатой x:

$$\rho(x, x_{1;x}) \le \rho(x, x_{2;x}) \le \ldots \le \rho(x, x_{m;x}) ,$$

где через $x_{i;x}$ обозначается тот объект обучающей выборки, который является i-м соседом объекта x. Аналогичное обозначение введем и для отклика на i-м соседе – $y_{i;x}$. Таким образом, произвольный объект порождает свою перенумерацию выборки. В наиболее общем виде принадлежность ближайшего соседа (ближайшей точки на графике счетов) соответствующему отклику (номеру диктора) можно записать в следующем виде:

$$a(x) = arg\max_{y \in Y} \sum_{i=1}^{m} [x_{i;x}=y] w(i,x), \qquad (3.2)$$

где $w(i,x)$ – заданная весовая функция, которая оценивает степень важности i-го соседа для классификации объекта, в настоящем разделе и далее $w(i,x)=1$. При $w(i,x)=1$, $i<k$ метод соответствует k ближайшим соседям:

$$a(x) = arg\max_{y \in Y} \sum_{i=1}^{k-1} [x_{i;x}=y]. \qquad (3.3)$$

Выбор числа ближайших соседей неоднозначно зависит от конкретных условий и представляет определенную сложность. Увеличение k с неизбежностью вовлечет в сопоставление данные разных дикторов и может ухудшить результат. Для иллюстрации этого обстоятельства на рис. 3.5 приведен качественный пример, где изображены данные для трех дикторов (Д1, Д2, Д3), а также неизвестные голосовые сигналы (X1, X2). «Образец» X1 имеет три соседа, принадлежащих диктору Д1, это означает, что с

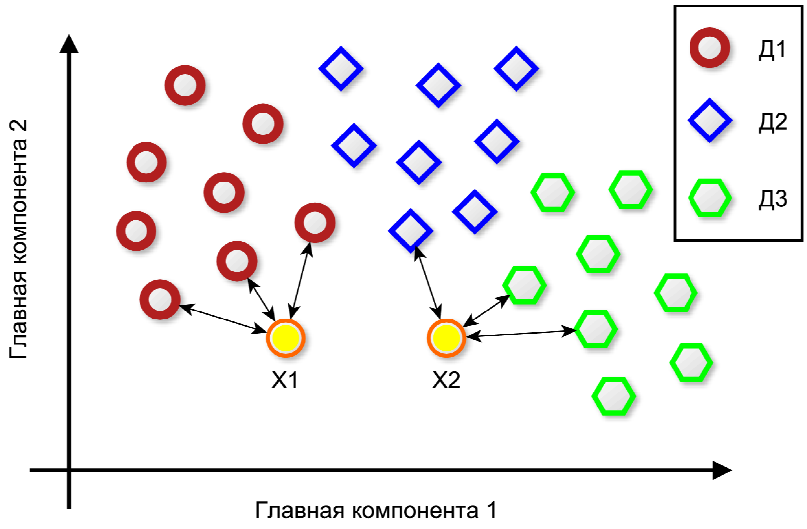


Рис. 3.5. Метод ближайших соседей для трех дикторов Д1, Д2, Д3. X1, X2 – «неизвестные» сигналы.

вероятностью 100% голосовой сигнал, соответствующий X1, относится к диктору Д1. Для образца X2 на рис. 3.5 видно, что два из трех ближайших соседей соответствуют диктору Д3, один – диктору Д2, то есть с вероятностью 67% голосовой сигнал, соответствующий X2, относится к диктору Д3.

Модифицированный метод ближайших соседей.

Изложенный выше метод ближайших соседей обладает тем недостатком, что выбор числа ближайших соседей не определен однозначно, поэтому результат идентификации может быть различным. Кроме того, этот метод не обеспечивает наглядность представления результата идентификации, что неудобно для интерпретации. В настоящей работе предложена модификация метода ближайших соседей, заключавшаяся в следующем. Вся многомерная совокупность объектов (сигналов, относящихся к одной группе) представлялась в виде многомерного эллипсоида, включавшего в себя все эти объекты и строящегося по максимальному разбросу точек. Таким образом, все точки одной группы располагались в объеме этого эллипсоида. Вся совокупность объектов представлялась в виде нескольких эллипсоидов в многомерном пространстве, при этом каждый эллипсоид описывал данные конкретного диктора. Идентификация неизвестных голосовых данных осуществляется по принадлежности описывавших эти данные точек объему соответствующего эллипсоида.

Для математического представления описанного подхода могут быть использованы различные метрики – Евклида, Махаланобиса, Хэмминга и другие [122]. Применение той или иной метрики определяется в основном данными, используемыми в конкретной задаче. В нашем случае задания описываемых точек эллипсоидами используется расстояние Махаланобиса, учитывающее ковариацию между объектами. Расстояние Махаланобиса рассчитывалось по следующей формуле:

$$\rho = \sqrt{(X'-\mu)^{*} C^{-1} (X'-\mu)} \tag{3.4}$$

где C – матрица ковариации, μ – математическое ожидание (множество со средним значением), $X' = \{ x'_i \}_{i=1}^{n}$ – многомерный вектор признаков, символ * обозначает транспонирование. Матрица ковариации вычислялась по следующей формуле:

$$\mathrm{C} = \frac{1}{\mathrm{m}} \sum_{i=1}^{m} (x_i - \bar{x})(y_i - \bar{y}) \tag{3.5}$$

где средние величины

$$\bar{x} = \frac{1}{\mathrm{m}} \sum_{i=1}^{m} x_i , \quad \bar{y} = \frac{1}{\mathrm{m}} \sum_{i=1}^{m} y_i . \tag{3.6}$$

В качестве величины μ выступает центр масс эллипсоида, координаты которого находились по следующей формуле:

$$\mu = \frac{1}{\mathrm{m}} \sum_{i=1}^{m} X_i , \tag{3.7}$$

здесь m – число многомерных точек X_i (векторов признаков) одной группы, описываемой эллипсоидом. Сама величина ρ представляла из себя диагональную матрицу, первые компоненты которой соответствовали расстоянию вдоль первых главных компонент.

Поверхности эллипсоидов строились по максимальным значениям расстояния Махаланобиса, рассчитывавшегося по формуле (3.4). Для идентифицируемого голосового сигнала также определялись расстояние Махаланобиса ρ_0 до центра масс эллипсоида и расстояние Махаланобиса ρ_i до ближайшей точки эллипсоида относительно идентифицируемой. По результатам сопоставления расстояний ρ_0, ρ_i определялась принадлежность идентифицируемой голосовой записи диктору (формула (3.8)), чьи данные характеризовались ближайшим эллипсоидом:

$$\begin{cases} \textit{принадлежит, если } \rho_0 \le \rho_i, \\ \textit{не принадлежит, если } \rho_0 > \rho_i. \end{cases} \quad (3.8)$$

Расчеты проводились с помощью специальной программы, которая позволяла автоматизированно соотносить неизвестный голосовой сигнал с конкретным диктором из базы данных.

В качестве иллюстрации на рис. 3.6 приведены результаты идентификации некоторого «неизвестного» голосового сигнала X с известными данными для двух дикторов (Д1, Д2). Рис. 3.6 а изображает расчеты в пространстве трех первых главных компонент, а рис. 3.6 б,в,г – в пространстве пар главных компонент: ГК1-ГК2, ГК2-ГК3, ГК1 и ГК3 соответственно. Как видно из рис. 3.6 а, данные, описывающие голосовые сигналы от двух дикторов, изображаются двумя неперекрывающимися эллипсоидальными «облаками» точек. Идентифицируемая точка (изображенная залитым кружком), как видно из рис. 3.6 а, соответствует диктору Д1. Важно отметить, что на рис. 3.6 б – 3.6 г, изображающих проекции эллипсоидов на соответствующие плоскости, попадание этой точки заметно только в компонентах ГК1, ГК3 (рис. 3.6 в) и в компонентах ГК2, ГК3 (рис. 3.6 г). Для графика на рис. 3.6 б в первых главных компонентах ГК1 и ГК2 области счетов дикторов Д1 и Д2 перекрываются и определить принадлежность идентифицируемой записи тому или иному диктору не представляется возможным.

Приведенные данные свидетельствуют о высокой наглядности предложенного способа описания. Проанализируем на основе модифицируемого метода ближайших соседей влияние на результат идентификации таких факторов, как число дикторов и число повторений фразы одним и тем же диктором в базе данных.

При расчетах использовалась база данных, описанная в разделе 2.2 и состоящая из записей 50 дикторов разного пола и возраста, произносивших

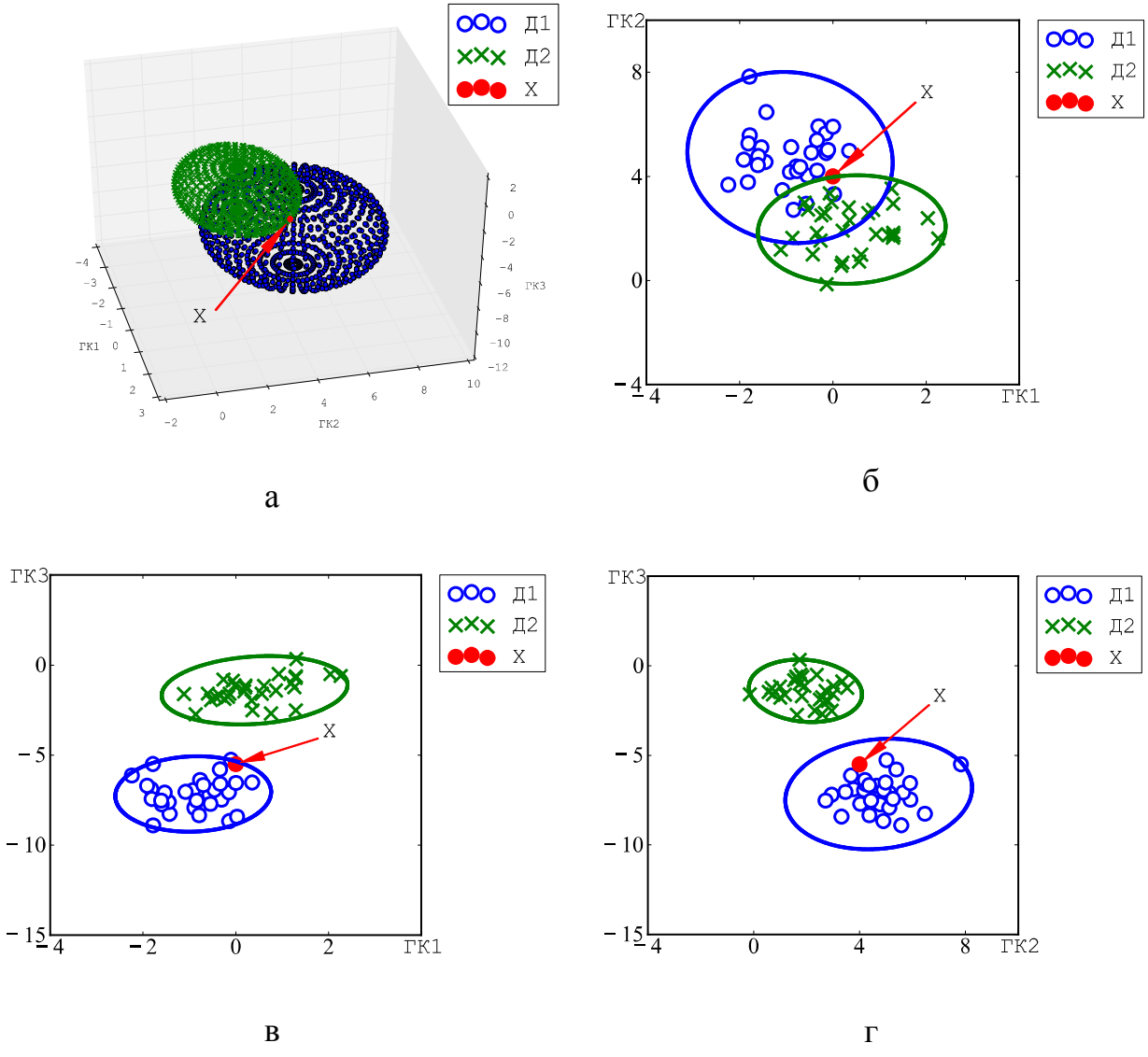
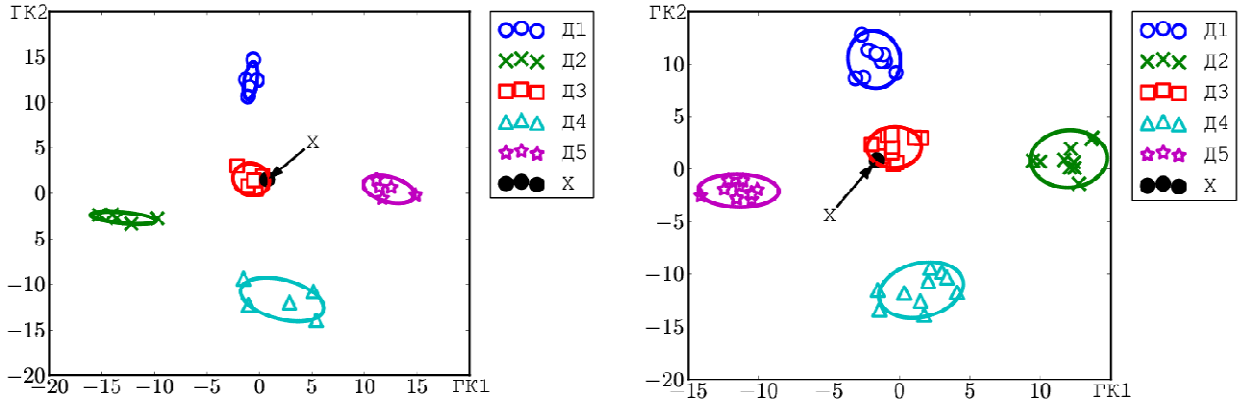


Рис. 3.6. Применения модифицированного метода ближайших соседей.
Д1 – счета 1-го диктора, Д2 – счета 2-го диктора,
залитая точка X – «неизвестный» идентифицируемый диктор;
а – для трех первых главных компонент (ГК1,ГК2,ГК3),
б – для главных компонент ГК1,ГК2,
в – для главных компонент ГК2,ГК3,
г – для главных компонент ГК1,ГК3.

специальные фразы определенное количество раз.

Число повторений фразы в базе данных. На рис. 3.7 приведены результаты идентификации для двух случаев: на рис. 3.7 а калибровочный набор состоит из записей голосов пяти дикторов (точки Д1, Д2, Д3, Д4, Д5),

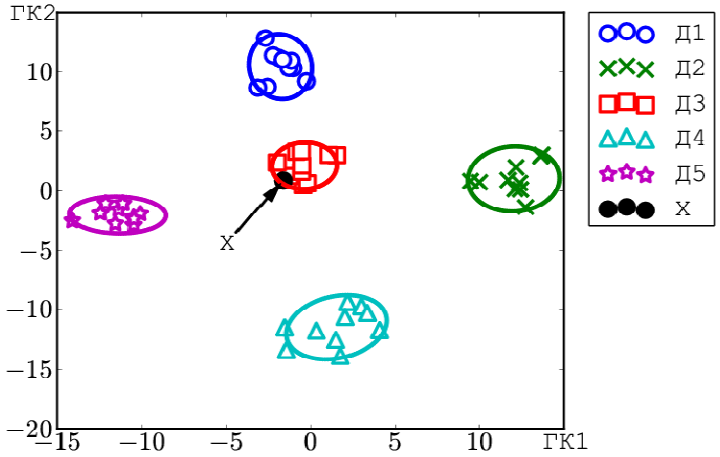


а б

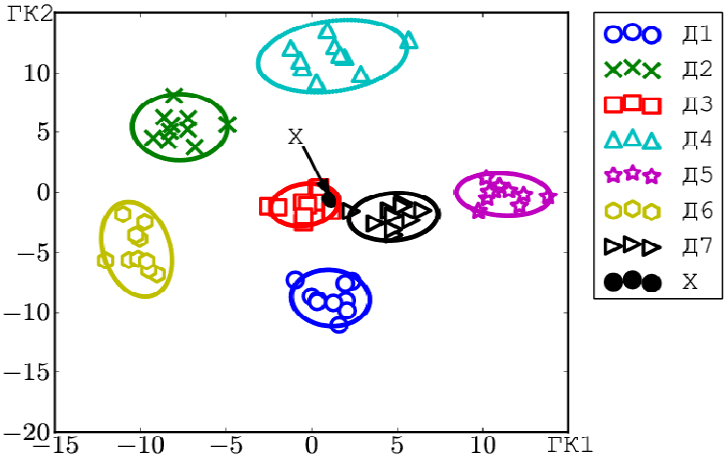
Рис. 3.7. Влияние числа повторений фразы.
а – 5 повторений, б – 10 повторений.

произносивших одну и ту же фразу («ноль один два три четыре») пять раз, а на рис. 3.7 б – 10 раз. В качестве идентифицируемого сигнала выступала запись голоса «неизвестного» диктора, изображаемая точкой X. Из рис. 3.7 видно, что увеличение числа повторений приводит к уменьшению области каждого эллипса, то есть более компактному расположению точек, относящихся к одному и тому жедиктору, и к увеличению расстояния между этими группами точек. Другими словами, увеличение надежности идентификации может быть достигнуто увеличением числа повторений фраз в базе данных.

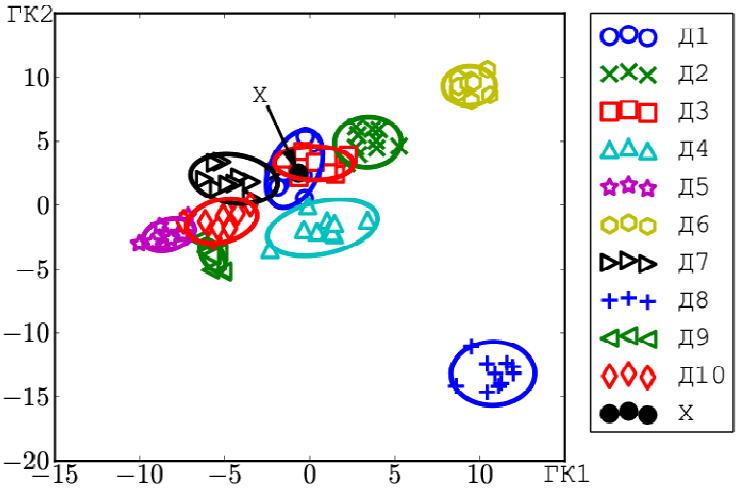
Число дикторов в базе данных. Рассчитывались графики счетов для трех групп из 5, 7, 10 дикторов, произносивших одну и ту же фразу 10 раз. На рис. 3.8 приведено сопоставление графиков счетов, полученных для этих групп. Видно, что увеличение числа дикторов в базе данных достаточно быстро приводит к перекрыванию соответствующих областей. В случае



а



б



в

Рис. 3.8. Влияние числа дикторов в базе данных.
а – 5 дикторов, б – 7 дикторов, в – 10 дикторов.

5 дикторов (рис. 3.8 а) области не перекрываются, для случая 7 дикторов (рис. 3.8 б) некоторые области заметно перекрываются, а для случая 10 дикторов (рис. 3.8 в) большая часть областей перекрывается друг с другом. Это является естественным результатом, потому что для достаточно большой группы часть дикторов неизбежно будет обладать схожими амплитудно-частотными характеристиками голосов. Результаты проведенного расчета показывают, что для такого случая метод ближайших соседей не позволяет провести надежную идентификацию и требуется разработка нового подхода. Содержание этого подхода рассматривается в следующем разделе.

3.3 Анализ голосовых сигналов на основе иерархического подхода

Описанный в предыдущем разделе 3.2 подход дает достаточно хорошие результаты в случае относительно небольшого числа дикторов, как используемых для формирования базы данных, так и идентифицируемых. В ситуации, когда происхождение идентифицируемого сигнала неизвестно и его приходится сравнивать с записями большого числа дикторов. Еще более сложной может оказаться ситуация, когда идентифицируемые сигналы включают в себя записи голосов нескольких дикторов. С позиции изложенных методов это означает, что многомерные области данных, описывающие голоса разных дикторов, будут перекрываться и достоверность идентификации будет существенно снижаться. В связи с этим в настоящем разделе предложен подход, призванный обеспечить разделение перекрывающихся голосовых сигналов.

Сущность это подхода заключалась в следующем. Перекрывающиеся области на графике счетов, соответствующие записям голосов различных дикторов с близкими амплитудно-частотными характеристиками, рассматривались как кластеры, возникающие из объективной близости голосов этих дикторов. К каждому из таких кластеров последовательно

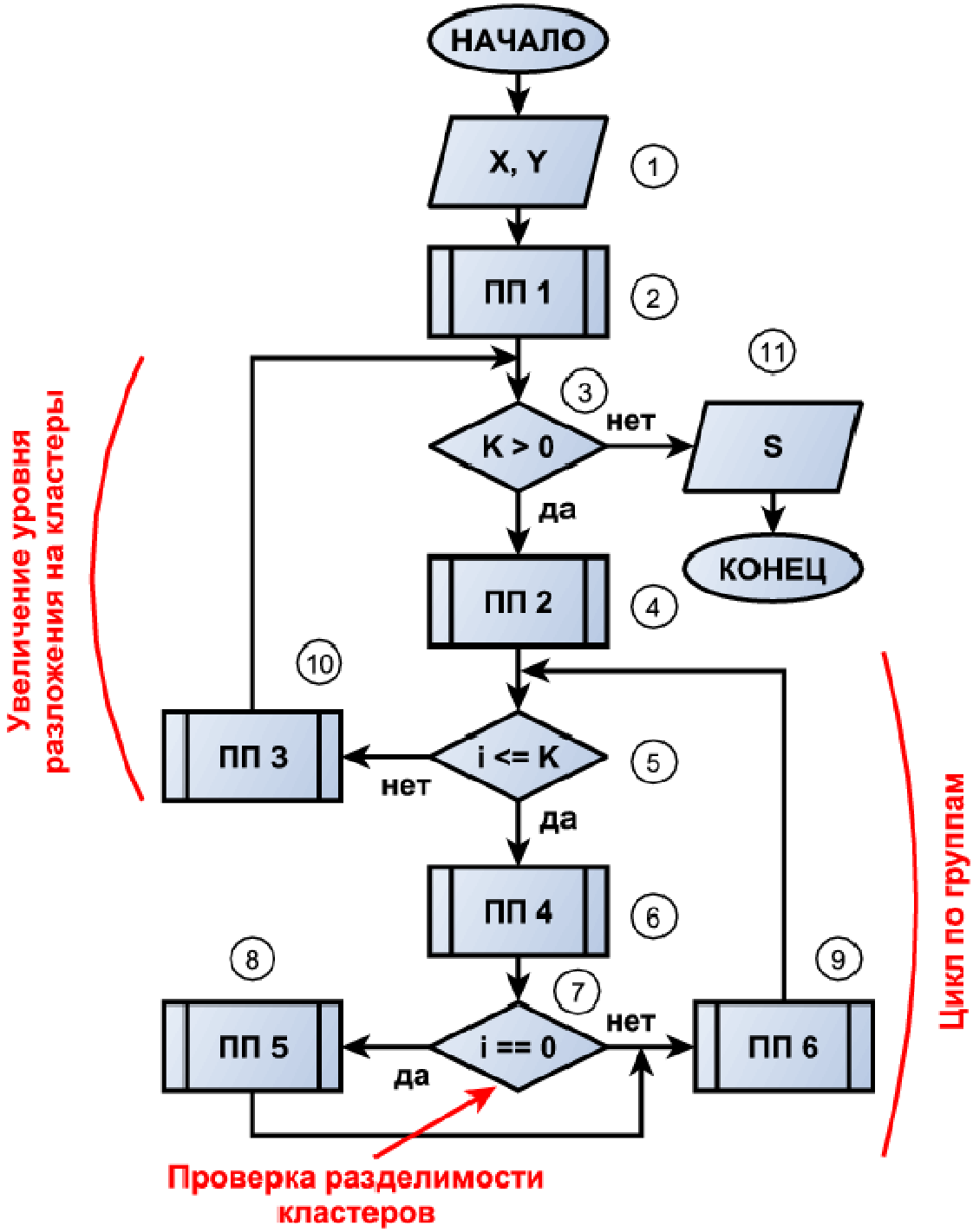


Рис. 3.9. Блок-схема алгоритма иерархического подхода для этапа калибровки.

применялся тот же алгоритм, что и для всех данных в целом. В результате отдельные кластеры распадались на меньшие кластеры, но уже образованные меньшим числом голосовых сигналов. Вся процедура продолжалась до тех пор, пока не происходило разделение на сигналы, идентифицируемые как принадлежавшие одному конкретному диктору, или пока достоверная идентификация признавалась невозможной. Этот подход, фактически означающий иерархичность классификации данных, может быть назван иерархическим подходом.

Схема предложенной алгоритмической процедуры приведена на рис. 3.9 для этапа калибровки. Здесь блок 1 – ввод данных: X – матрица признаков, Y – отклики, i и j индексы при итерации. Блок 2 производит вычисление индексов элементов M в группах, номера N и количества K групп. Блок 3 проверяет условие количества групп. Блок 4 задает параметры для цикла по группам. Блок 5 проверяет текущую группу: если условие выполняется, то следующий блок 6, иначе – блок 10. Блок 6 предназначен для вычисления главных компонент. Блок 7 проверяет разделимость кластеров, если условие выполняется, то блок 8, иначе – блок 9. Блок 8 сохраняет номера групп и соответствующие номера главных компонент. В блоке 9 сохраняются главные компоненты и увеличивается счетчик i на единицу. Далее выполняется проверка в блоке 5. В блоке 10 производится сохранение номеров групп и увеличивается счетчик j, после чего осуществляется переход в блок 3. Блок 11 осуществляет вывод переменной S, в которой сохранялись все необходимые для дальнейшей классификации данные.

Аналогично блок-схеме алгоритма иерархического подхода для этапа калибровки строилась блок-схема для этапа тестирования, в которую блок сравнения «неизвестной» голосовой записи с калибровочным набором. Достоинством двух приведенных алгоритмов является автоматизация

расчетов, что было реализованно с помощью специально разработанной программы.

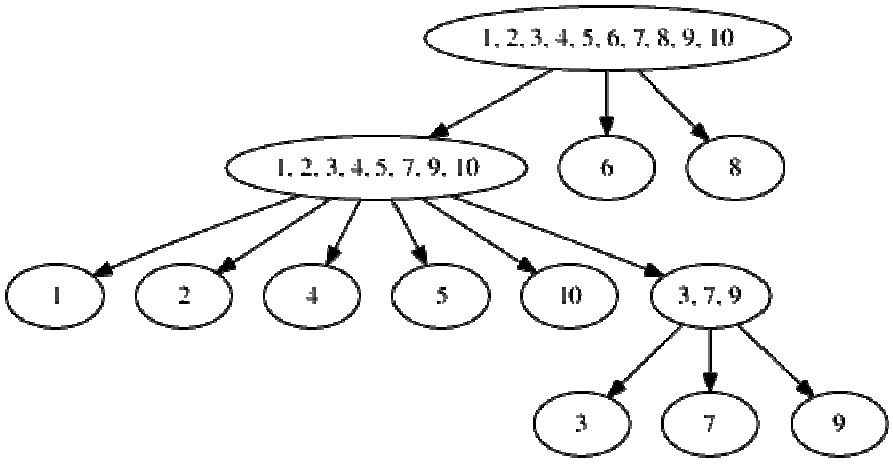
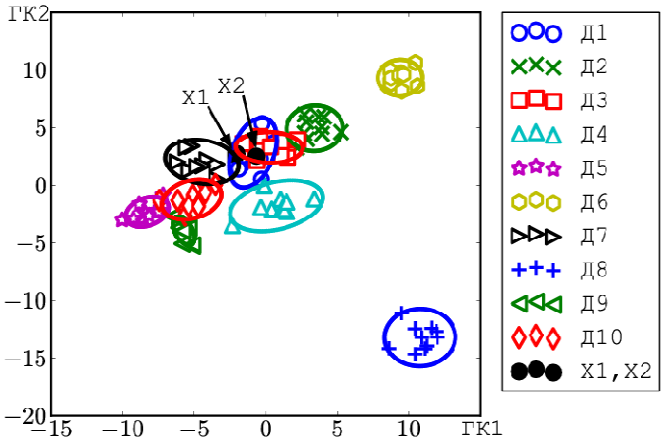


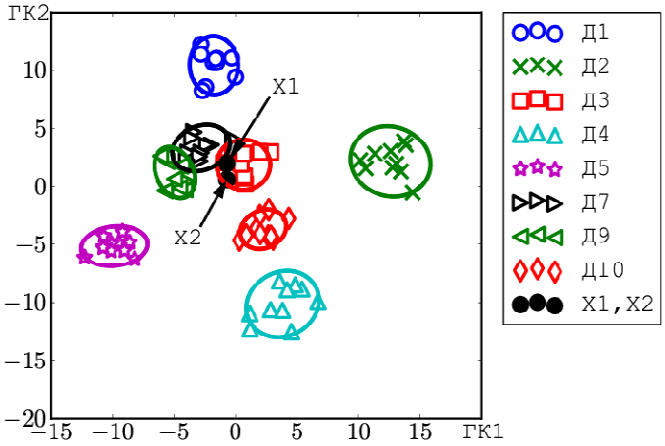
Рис. 3.10. Схема разбиения на кластеры счетов.

Предложенный метод был применен к исследованию влияния на результат идентификации числа дикторов, формировавших калибровочную базу данных [123]. В качестве примера рассмотрим графики счетов для 10 дикторов, произносивших одну и ту же фразу. Использованная в данном конкретном примере схема разбиения для группы из 10 дикторов приведена на рис. 3.10. Точки , относившиеся к одному диктору, для наглядности выделялись эллипсами, которые рассчитывались по расстоянию Махаланобиса согласно разделу 3.2.

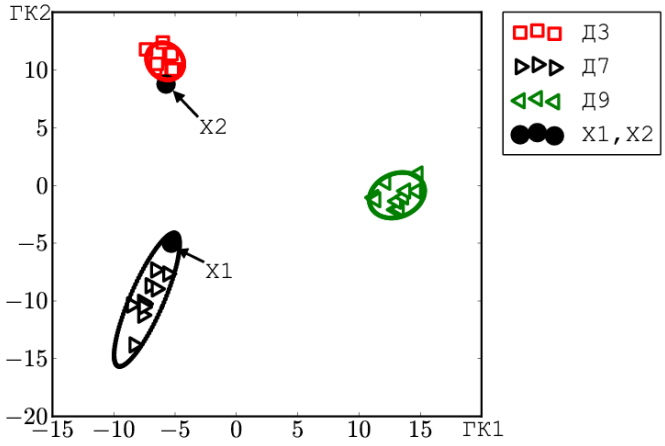
После применения процедуры идентификации область графика счетов (рис. 3.11 а) отчетливо разделялась на группы кластеров: первый кластер состоит из сигналов дикторов 1, 2, 3, 4, 5, 7, 9, 10, остальные описывали записи голосов отдельных дикторов. Идентификация голосов «неизвестных» дикторов X1 и X2 практически невозможна. На следующем этапе к каждому кластеру, образованному перекрывающимися данными, вновь применялась процедура идентификации. Видно, что кластер дикторов 1, 2, 3, 4, 5, 7, 9, 10, отчетливо разбивался на отдельные области, описывавшие записи дикторов 1,



а



б



в

Рис. 3.11. Разбиение счетов на кластеры с помощью иерархического подхода.

а – начальное состояние, б – первое разбиение, в – второе разбиение.

2, 4, 5, 10 (рис. 3.10 б), и общий кластер дикторов 3, 7, 9. Далее аналогично кластер дикторов 3, 7, 9 разбивался на неперекрывавшиеся области, соответствовавшие отдельным дикторам (рис. 3.11 в). Из рис. 3.11 в следует, что при завершении процедуры «неизвестные» голосовые сигналы X1 и X2 вполне доспускают идентификацию.

Отметим, что иерархический подход обладает еще и тем достоинством, что позволяет существенно уменьшить число главных компонент, необходимых для идентификации. Фактически все разделение производится с использованием первых двух главных компонент. Например, в случае большого числа дикторов, когда при модифицированном методе ближайших соседей приходилось использовать 8 главных компонент, при иерархическом подходе можно успешно разделить данные, используя 4 главные компоненты. Большое число главных компонент не позволяет контролировать процесс алгоритмизации, поскольку это число часто определяют по отдельно рассчитываемому графику остатков. В случае выделения кластеров голосовых сигналов в иерархическом подходе по близким голосовым сигналам такой контроль более эффективен и нагляден.

ГЛАВА 4. ВЛИЯНИЕ ПОМЕХ НА ИДЕНТИФИКАЦИЮ ЛИЧНОСТИ ПО ГОЛОСУ

4.1 Влияние искажений голосового сигнала на идентификацию

Идентификация личности по голосу, проводимая в реальных условиях, встречается с рядом серьезных затруднений. Во-первых, возможны искажения, связанные непосредственно с диктором и обусловленные особенностями его психофизического состояния, заболеванием и т. п. Эти искажения с помощью любой автоматизированной системы обработки и классификации исключить невозможно, можно лишь уменьшить их влияние. Во-вторых, возникают аппаратные искажения на различных участках прохождения речевого сигнала при его записи, обработке и хранении. В-третьих, на голосовой сигнал неизбежно накладываются внешние механические шумы, которые могут существенно его искажать. Важнейшей задачей систем голосовой идентификации является уменьшение негативного влияния второго и третьего факторов. Именно решению этой задачи с использованием развиваемого в диссертации подхода, основанного на аппарате анализа многомерных данных, посвящена настоящая глава.

На рис. 4.1 схематично изображены места воздействия внешнего шума и помех на разные участки прохождения речевого сигнала.

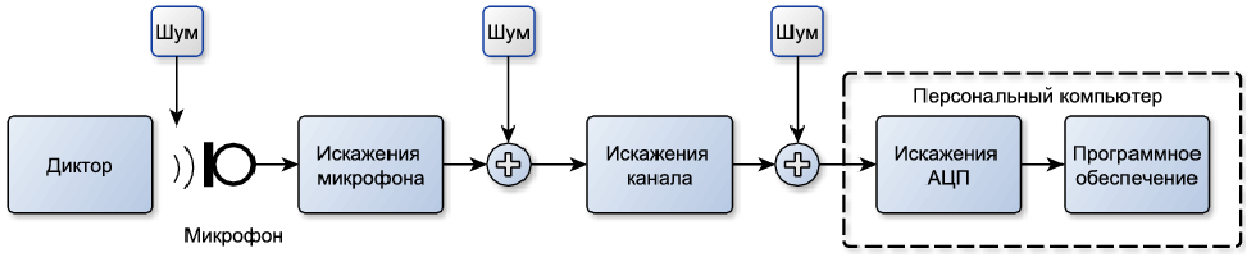


Рис. 4.1. Воздействие шума и помех на различные участки прохождения речевого сигнала.

Обычно выделяют: искажения сигнала, связанные с самим диктором, с шумом окружающей среды, с искажением микрофонной системы (в том числе электромагнитные помехи), искажения, возникающие в канале записи при передаче сигнала, и искажения при программной обработке сигнала в компьютере. В настоящем разделе будут рассмотрены искажения, обусловленные изменением амплитудно-частотного спектра самого исходного сигнала. Роль внешнего шума будет описана в следующем разделе 4.2.

Помехи, возникающие в аппаратной части системы идентификации, в конечном счете, сводятся к частотным и амплитудным искажениям исходного спектра и сигнала. Это может быть вызвано недостатком микрофонных устройств, обладающих нелинейными амплитудно-частотными характеристиками, применением различных фильтров при записи сигналов, а также искажениями при аналого-цифровом преобразовании. Для изучения влияния искажения голосового сигнала на идентификацию диктора применен подход, изложенный в предыдущей главе 3.

Для математического моделирования искажения голосовых сигналов был применен алгоритм передискретизации, основанный на использовании дискретного преобразования Фурье и позволявший повышать частоту дискретизации сигнала в задаваемое целое или дробное число раз [124]. Моделирование конкретного искажения [125] осуществлялось следующим образом. Пусть исходный сигнал характеризовался конечным числом отсчетов *a(n)*. На первом шаге алгоритма проводилось вычисление коэффициентов *A(k)* прямого преобразования Фурье:

$$A(k)=\sum_{n=1}^{N} a(n)\cdot e^{-j2\pi\frac{k}{N}n}, \quad k=1,2,\ldots,N \tag{4.1}$$

На втором шаге в область возле отсчета с номером *N/2* спектра вставлялись нулевые компоненты, количество которых задавалось

значениями начального числа отсчетов N и числа отсчетов в передискретизованном сигнале M. Коэффициенты $H(i)$ передискретизованного спектра в случае нечетных чисел N определялись формулами:

$$\begin{cases} H(i) = A(i), 1 \le i \le \dfrac{N+1}{2}, \\ H(i) = 0, \dfrac{N+1}{2} + 1 \le i \le \dfrac{N+1}{2} + M - N, \\ H(i) = A(i - M + N), \dfrac{N+1}{2} + M - N \le i \le M, \end{cases} \tag{4.2}$$

в случае четных N – формулами:

$$\begin{cases} H(i) = A(i), 1 \le i \le \dfrac{N}{2}, \\ H(i) = \dfrac{A(N/2+1)}{2}, i = \dfrac{N}{2} + 1, \\ H(i) = 0, \dfrac{N}{2} + 2 \le i \le \dfrac{N}{2} + M - N, \\ H(i) = \dfrac{A(N/2+1)}{2}, i = \dfrac{N}{2} + M - N + 1, \\ H(i) = A(i - M + N), \dfrac{N}{2} + M - N + 2 \le i \le M. \end{cases} \tag{4.3}$$

На заключительном шаге алгоритма вычислялись отсчеты $h(m)$ обратного дискретного преобразования Фурье с нормировкой:

$$h(m) = \frac{1}{M} \sum_{k=1}^{M} H(k) \cdot e^{-j2\pi \frac{k}{M} m}, \quad m = 1,2,\ldots,M \tag{4.4}$$

На этом формирование искаженного сигнала заканчивалось.

В качестве величины, количественно характеризующей искажения, был использован коэффициент нелинейных искажений K, который вводился как отношение среднеквадратичной суммы спектральных компонентов выходного сигнала, отсутствующих в спектре входного сигнала, к среднеквадратичной сумме спектральных компонентов входного сигнала:

$$K = \frac{\sqrt{\frac{1}{L}\sum_{l=1}^{L} H^2(l)}}{\sqrt{\frac{1}{N}\sum_{k=1}^{N} A^2(k)}} \qquad (4.5)$$

где $H(l)$ – спектральные компоненты выходного сигнала, отсутствующие в спектре входного сигнала $A(k)$, L – количество спектральных компонент $H(l)$.

Искажения генерировались таким образом, что частота дискретизации:

$$F = \frac{M}{t}, \qquad (4.6)$$

(здесь t – длительность исходного сигнала) увеличивалась за счет изменения количества отсчетов от N до M, при этом для каждого искаженного сигнала рассчитывался коэффициент нелинейных искажений K.

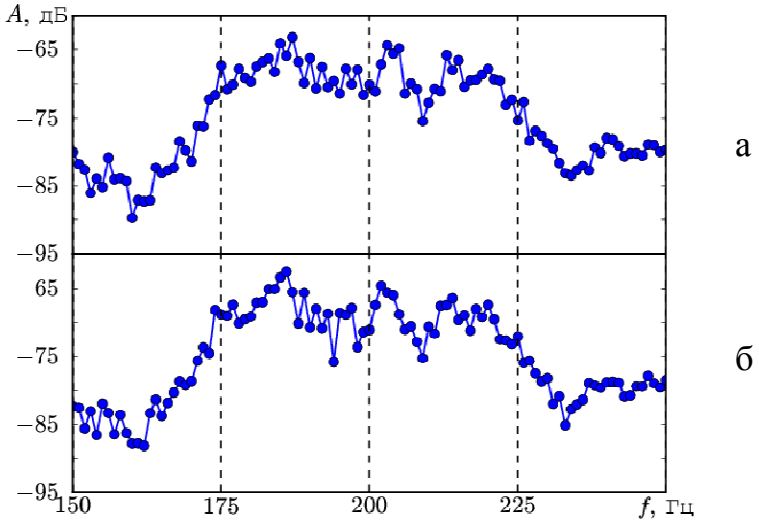


Рис. 4.2. Участки частотных спектров.
а – спектр исходного сигнала, б – частотный спектр искаженного сигнала.

На рис. 4.2 для иллюстрации приведен участок частотного спектра исходного сигнала и соответствующий частотный спектр искаженного сигнала (K=0,3) для интервала частот от f=150 Гц до f=250 Гц, для которого искажения оказались наиболее заметными (здесь A – амплитуда звуковых колебаний).

Далее исходный и искаженный акустические сигналы подвергались идентификации с помощью подходов, описанных в главе 3. Полученные при расчетах данные представлялись в виде графиков счетов в пространстве первых главных компонент [125,126], позволявших наглядно интерпретировать результат идентификации.

Для апробации подхода в качестве исходных сигналов использовались записи голоса диктора, представлявшие из себя фразу «один два три четыре пять шесть семь восемь девять», с частотой дискретизации 8 кГц и разрядностью 16 бит. Искажения в исходный сигнал вносились согласно формулам (4.1) – (4.4) и заключались в добавлении узкого участка спектра отсчетами, имевшими равную нулю амплитуду. В качестве графической иллюстрации на рис. 4.3 приведены участки передискретизованных спектров, подвергавшихся искажениям, соответствовавшим коэффициентам $K = 0{,}15$; 0,20; 0,25; 0,30.

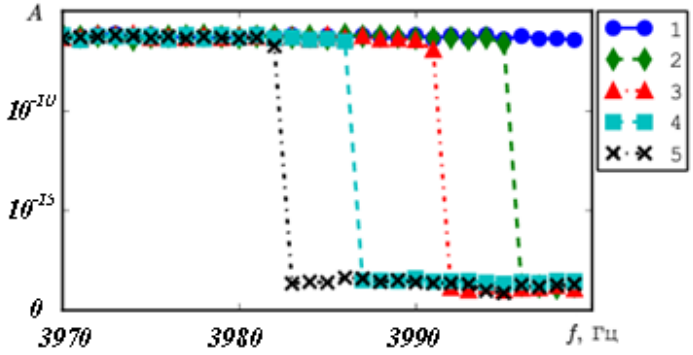


Рис. 4.3. Участки передискретизованных спектров.
1 – K=0, 2 – K=0,15, 3 – K=0,20, 4 – K=0,25, 5 – K=0,30.

Рассчитанный для исходного и искаженных сигналов график счетов представлен на рис. 4.4. Здесь отдельная точка, как и раньше, соответствует одной голосовой записи, исходные сигналы (данные одного диктора) изображены незалитыми точками **Д**, искаженные сигналы с разными значениями коэффициента искажения K – залитыми точками **И** (точка 1 –

K= 0,15, 2 – K= 0,20, 3 – K= 0,25, 4 – K = 0,30). Все исходные (неискаженные) сигналы представлены компактной областью, выделенной эллипсом, который строился (согласно разделу 3.2) по максимальному разбросу соответствующих точек. Попадание точек, описывавших искаженные сигналы, в выделенную область означало правильную идентификацию диктора. Случаи, отвечавшие точкам вне выделенной области, означали, что диктор воспринимался как «чужой», то есть искажения сигнала оказывались настолько значительными, что идентификация не достигалась. Из рис. 4.4 видно, что при коэффициентах искажения $K = 0{,}15$ и $K = 0{,}20$ (точки 1 и 2) диктор, несмотря на искажения сигнала, идентифицировался правильно. При коэффициентах $K = 0{,}25$ и $K = 0{,}30$ (точки 3 и 4, расположенные вне эллипса) идентификация уже не достигалась.

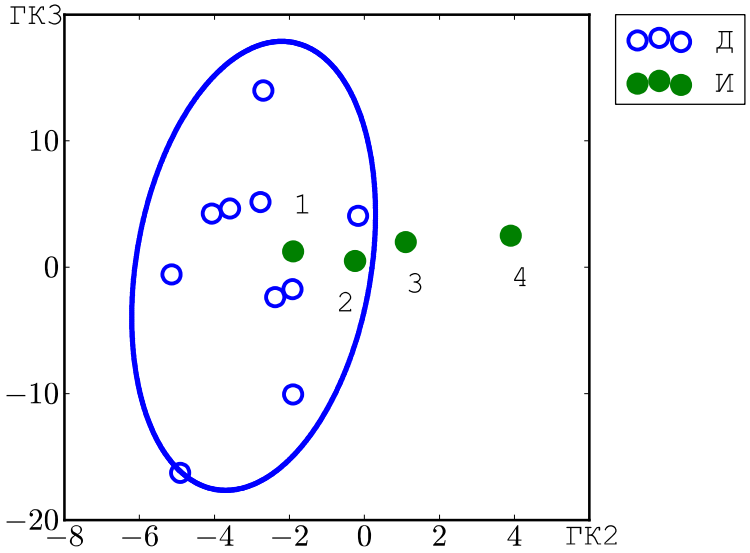


Рис. 4.4. Влияние искажений на голосовую идентификацию.
Д – неискаженные данные; И – данные с искажениями.
1 – K=0, 2 – K=0,15, 3 – K=0,20, 4 – K=0,25, 5 – K=0,30.

Таким образом, проведенное математическое моделирования искажений голосового сигнала дало возможность провести количественную оценку величины этих искажений, при которых возможна правильная идентификация личности [127]. Это показывает, что предложенный в настоящем разделе

подход к оценке влияний искажений может использоваться для анализа надежности методов голосовой идентификации.

4.2 Влияние внешнего акустического шума

Внешний акустический шум, накладывающийся в реальных условиях на идентифицируемый голосовой сигнал, является одним из основных источником ошибок при идентификации. При достаточно высокой величине этого шума идентификация может оказаться вообще невозможной.

В используемых на практике системах идентификации личности по голосу существуют различные методы подавления шума, которые применяются на различных этапах прохождения сигнала. В большинстве случаев подавление осуществляется на аппаратном уровне путем фильтрации помех и на этапе предварительной обработки речевого сигнала. В большинстве систем идентификации по голосу мало внимания уделяется подавлению шума на этапе классификации данных. В данном разделе влияние шума рассматривается именно для этапа классификации.

При анализе использовались следующие виды шума, отличавшиеся по частотным распределениям амплитуд: «белый» шум, «розовый» шум, «коричневый» шум, шум улицы, речеподобный шум.

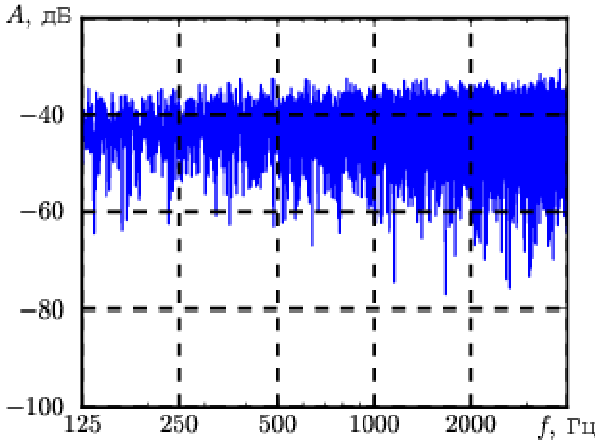


Рис. 4.5. Спектр «белого» шума.

«Белый» шум характеризуется равномерным распределением частот:

$$A \approx const \quad , \tag{4.7}$$

вид шума изображен на рис. 4.5.

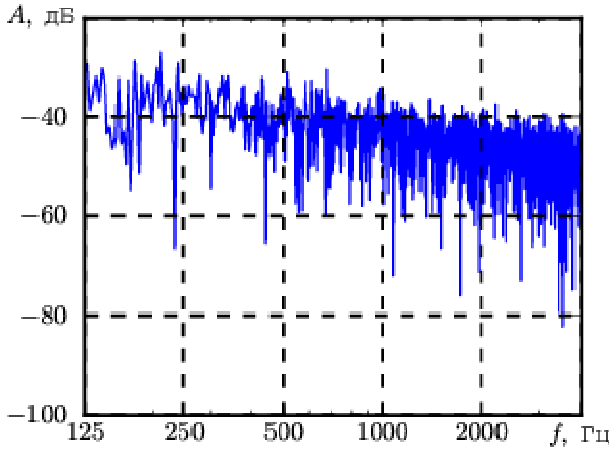


Рис. 4.6. Спектр «розового» шума.

Шум с распределением, обратно пропорциональным частоте и равномерным в логарифмической шкале частот, называют **«розовым» шумом**, для него:

$$A \approx 1 / f \quad , \tag{4.8}$$

график такого шума приведен на рис. 4.6.

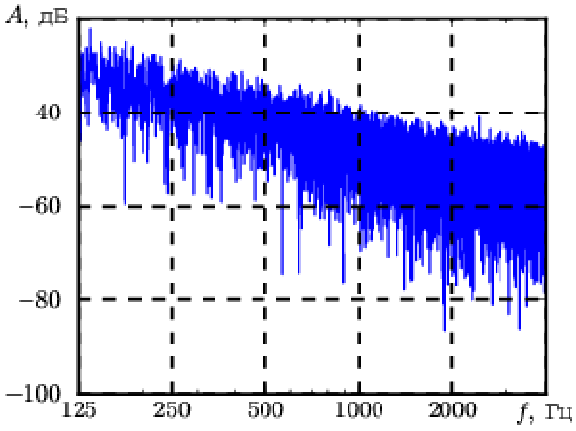


Рис. 4.7. Спектр «коричневого» шума.

Для **«коричневого» шума** (рис. 4.7) характерное распределение амплитуд обратно пропорционально квадрату частоты:

$$A \approx 1/f^2. \tag{4.9}$$

Особый интерес представляет исследование **речеподобного шума**, который специально создается путем наложения отрезков записей голосов

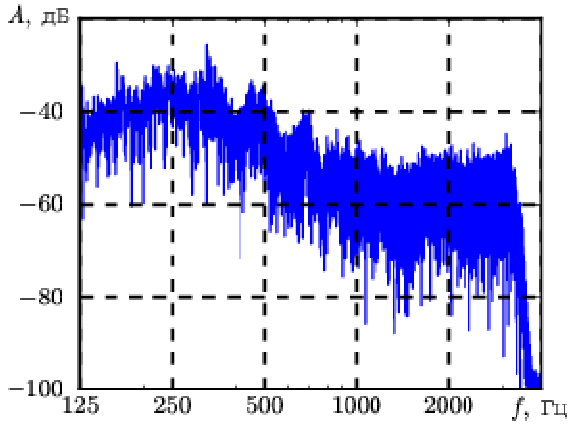


Рис. 4.8. Спектр речеподобного шума.

дикторов, поскольку такой шум обычно используется в системах защиты акустической информации. Вид спектра этого шума приведен на рис. 4.8.

Кроме того, для моделирования реальных условий использовался **шум**

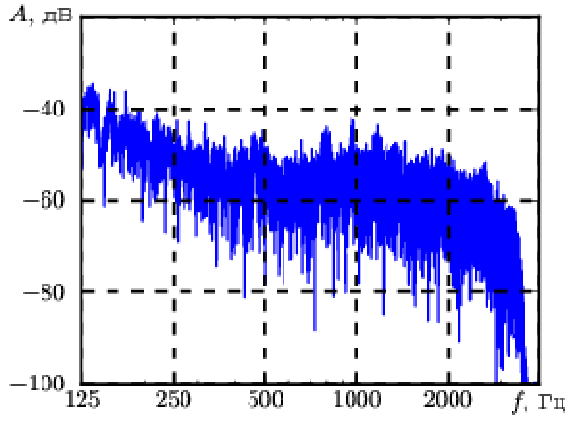


Рис. 4.9. Спектр шума улицы.

улицы, записывавшийся в реальных условиях на оживленной автомагистрали (рис. 4.9).

Для количественной оценки воздействия шума рассчитывались уровни его воздействия, которые определялись соотношением «полезный сигнал шум»:

$$R_{SN} = 20 \log_{10} \left(\frac{A_S}{A_N} \right) \tag{4.10}$$

где A_S – среднеквадратичное значение амплитуды сигнала, A_N – среднеквадратичное значение амплитуды шума. Величина отношения «полезный сигнал/шум» задавалась от 15 дБ (пренебрежимо малый шум) до -15 дБ, что соответствовало значительному превышению шума над уровнем сигнала.

Влияние уровня шума на идентификацию.

При расчетах влияние шума оценивалось путем аддитивного наложения шумового сигнала на исходные голосовые данные. В качестве таких данных были использованы записи голосов 10 дикторов, произносивших фразу,

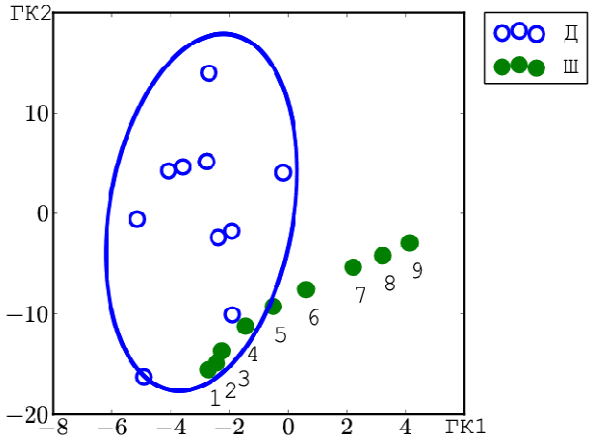


Рис. 4.10. Влияние уровня «белого» шума на результат идентификации.
Д – неискаженные данные для одного диктора,
Ш – данные с наложенные шумом.
1 –R_{SN}=15 дБ, 2 –R_{SN}=10 дБ, 3 –R_{SN}=7 дБ, 4 –R_{SN}=5 дБ, 5 –R_{SN}=0 дБ,
6 –R_{SN}=−5 дБ, 7 –R_{SN}=−7 дБ, 8 –R_{SN}=−10 дБ, 9 –R_{SN}=−15 дБ.

состоящую из цифр от 0 до 9 с десятикратным повторением. Оценка влияния шума на результат идентификации проводилась графиком счетов, рассчитывававшихся с помощью иерархического подхода. Это влияние подробно исследовалось на примере «белого» шума. Результаты проведенных расчетов представлены на рис. 4.10 в виде графика счетов для первых двух главных компонентн на примере одного диктора [128]. На представленном графике счетов калибровочные данные образуют в плоскости двух первых главных компонент компактную область. Эта область для наглядности выделена эллипсом, рассчитанным с помощью расстояния Махаланобиса в соответствии с разделом 3.2. Идентифицируемые голосовые сигналы с различным уровнем шума изображались в виде серии залитых точек (точки Ш, R_{SN} = 15; 10; 7; 5; 0; -5; -7; -10; -15 (дБ)). Идентифицирующим правилом для определяемой голосовой записи являлось попадание или непопадание соответствующей точки в выделенную область, описывающую неискаженные данные (точки Д). Как видно из рис. 4.10, точки 1–4, соответствующие уровням шума вплоть до R_{SN} = 5 дБ, оказывались внутри области и тем самым правильно идентифицировались. При меньших значениях отношения «полезный сигнал/шум» (R_{SN} < 5 дБ) точки 5–9 воспринимались как чужие, то есть диктор идентифицировался неверно. Полученные результаты позволяют количественно оценить уровень шума, не допускающий достоверную идентификацию по голосу.

Влияние вида шума на идентификацию.

Исследование роли вида шума представляет несомненный интерес как при создании систем по защите информации от утечек по акустическому каналу, так и для разработки методов контроля защищенности от таких утечек. Отметим, что роль вида действующего шума для задачи голосовой идентификации была фактически не исследована, что обусловлено

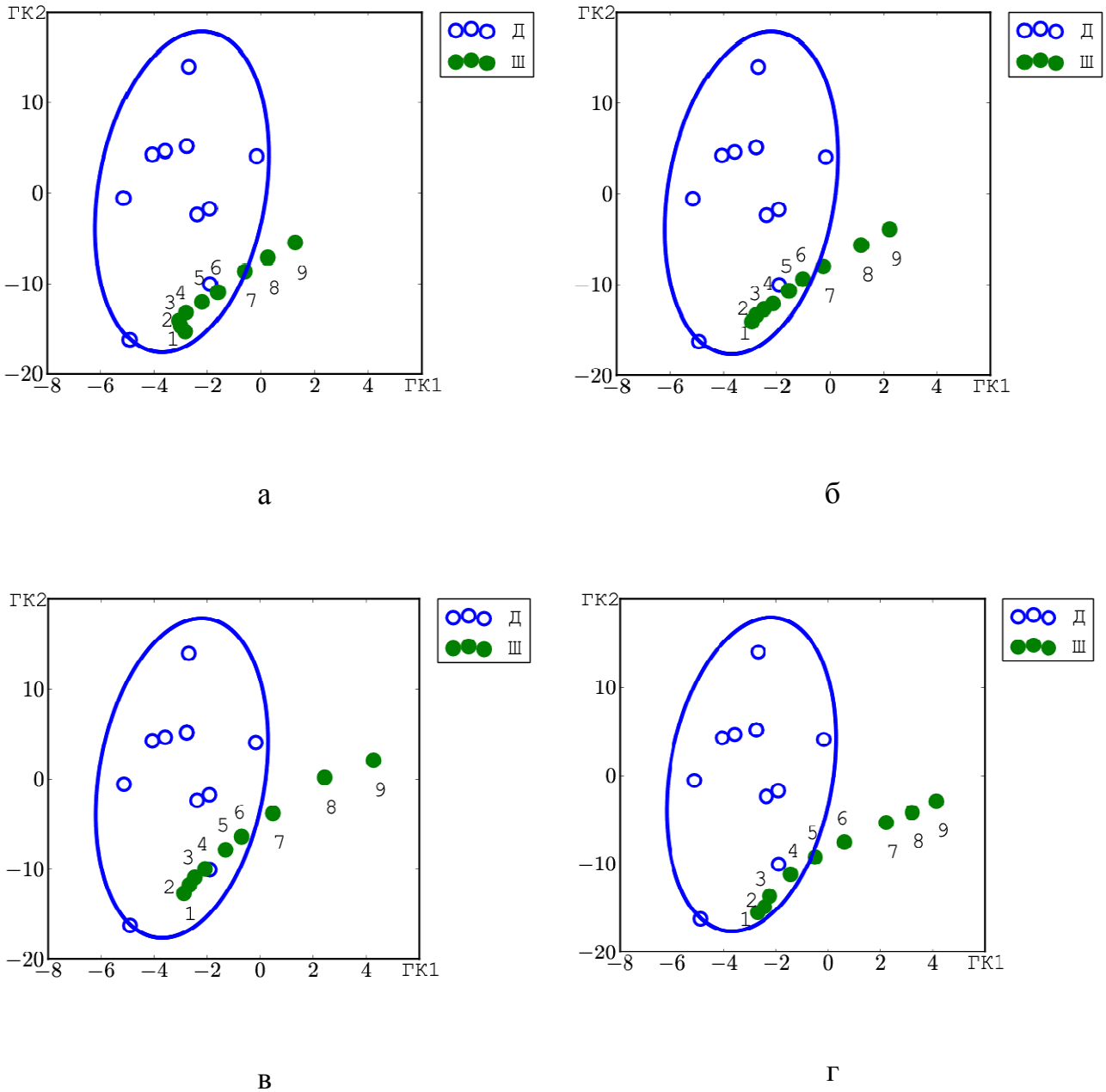


Рис. 4.11. Влияние вила шума на результат идентификации.
а – «розовый» шум, б – «коричневый» шум,
в – шум улицы, г – речеподобный шум.
Д – неискаженные данные для одного диктора,
Ш – данные с наложенные шумом.
1 –R_{SN}=15 дБ, 2 –R_{SN}=10 дБ, 3 –R_{SN}=7 дБ, 4 –R_{SN}=5 дБ, 5 –R_{SN}=0 дБ,
6 –R_{SN}=–5 дБ, 7 –R_{SN}=–7 дБ, 8 –R_{SN}=–10 дБ, 9 –R_{SN}=–15 дБ.

недостатками методов классификации голосовых данных, не обеспечивавших количественную оценку различий между разными видами шума.

На рис. 4.11 приведено сопоставление для случая одного диктора графики счетов при наложении «розового» (рис. 4.11 а), «коричневого» (рис. 4.11 б), речеподобного (рис. 4.11 г) шумов и шума улицы (рис. 4.11 в), «белый» шум был приведен на рис. 4.10. Из рис. 4. 11, например, следует, что серия точек с наложением шума на исходный сигнал в случае речеподобного шума (рис. 4.11 г) раньше оказывается за пределами эллипса, чем в случае шума улицы (рис. 4.11 в), то есть вносимые искажения при одном и том же значении отношения «полезный сигнал/шум» для речеподобного шума сильнее. Аналогично может быть проанализировано влияние других видов шума.

4.3 Оценка ошибок идентификации при искажении и наложении акустического шума

Для выяснения надежности и достоверности предложенных подходов, определения границ их применимости, а также выяснения наиболее значимых для идентификации факторов были проведены количественные расчеты ошибок идентификации. Под ошибками 1-го рода понимались ложноотвергнутые голосовые сигналы, а под ошибками 2-го рода – ложно принятые сигналы. Отметим, что для практических целей более опасными являются ошибки 2-го рода, так как при этом достигается допуск к конфиденциальной информации лиц, не имеющих на это право.

Ошибка 1-го рода, показывающая долю отказов в идентификации, рассчитывалась по формуле:

$$\delta^I = \frac{1}{M}\sum_{i=1}^{M}\delta_i^I \qquad (4.11)$$

$$\delta_i^{\,I} = \frac{1}{N_i}\sum_{k=1}^{N_k} x_{ik}\ , \quad x_{ik} = \begin{cases} 1, \text{если } a_{ijk} = 0, \text{при } i = j \\ 0, \text{если } a_{ijk} = 1, \text{при } i = j \end{cases} \tag{4.12}$$

где i,j – индекс группы одного диктора, k – индекс внутри группы одного диктора, a_{ijk} – результат идентификации диктора, x_{ik} – величина, характеризующая k-ый результат идентификации i-го диктора, N_k – количество «образцов» в группе одного диктора, M – количество дикторов, δ_i^I – ошибка I рода диктора под номером i. Пояснение вычисления ошибки 1-го рода приведено на рис. 4.12, где используются четыре группы (Гр. 1 – Гр. 4) по три образца в каждой группе. Величина x содержит значения, на основе которых вычисляется x_{ij} из формулы (4.12).

a	Гр. 1	Гр. 2	Гр. 3	Гр. 4
1	0	1	0	0
1	1	0	0	0
1	1	0	0	0
2	0	0	1	0
2	0	0	1	0
2	0	1	0	0
3	1	0	0	0
3	1	0	0	0
3	0	0	0	1
4	0	0	1	0
4	1	0	0	0
4	0	0	1	0

x
1
0
0
1
1
0
1
1
1
1
1
1

Рис. 4.12. Пояснение вычисления ошибок 1-го рода.

Ошибка 2-го рода, показывающая долю дикторов, ошибочно идентифицируемых как «свои», рассчитывалась по формуле:

$$\delta^{II} = \frac{1}{M}\sum_{i=1}^{M} \delta_i^{II} \tag{4.13}$$

$$\delta_i^{II} = \frac{1}{M-1}\sum_{l=1,l\neq i}^{M-1}\left(\frac{1}{N_j}\sum_{j=1}^{N_j} x_{ilj}\right), \quad x_{ilj} = \begin{cases} 1, \text{если } a_{ijk} = 1 \\ 0, \text{если } a_{ijk} = 0 \end{cases} \quad (4.14)$$

где δ_i^{II} – ошибка *II* рода диктора под номером *i*.

a	Гр. 1	Гр. 2	Гр. 3	Гр. 4
1	0	1	0	0
1	1	0	0	0
1	1	0	0	0
2	0	0	1	0
2	0	0	1	0
2	0	1	0	0
3	1	0	0	0
3	1	0	0	0
3	0	0	0	1
4	0	0	1	0
4	1	0	0	0
4	0	0	1	0

x			
	1	0	0
	0	0	0
	0	0	0
0		1	0
0		1	0
0		0	0
1	0		0
1	0		0
0	0		1
0	0	1	
1	0	0	
0	0	1	
Гр. 1	Гр. 2	Гр. 3	Гр. 4

Рис. 4.13. Пояснение вычисления ошибок 2-го рода.

Аналогично рис. 4.12 для вычисления ошибки 2-го рода приведено пояснение на рис. 4.13.

При расчетах итоговое значение ошибки находилось как среднее значение ошибок, соответствовавших разным калибровочным и тестовым наборам дикторов.

Были проведены расчеты ошибок 1-го и 2-го рода в зависимости от числа дикторов, входящих в калибровочную базу данных, длительности и повторяемости составляющих эту базу словосочетаний (фраз). В качестве калибровочного и тестового набора выступали записи из нашей речевой базы данных, а именно, записи 50 человек с 20 кратным повторением цифр от 0 до 9. Для калибровки использовалось 80% исходной базы данных, а для тестирования – 20% (рис. 4.14).

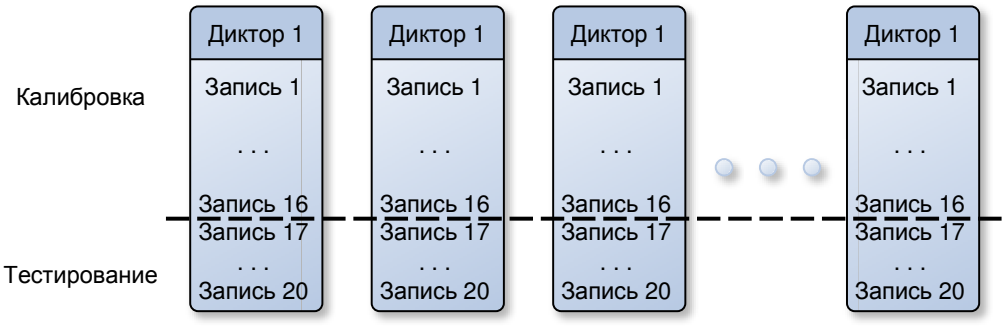


Рис. 4.14. Разбиение базы данных для калибровки и тестирования.

На рис. 4.15 приведены графики зависимостей ошибок от различного числа дикторов. Здесь кривая 1 изображает ошибки 1-го рода, кривая 2 – ошибки 2-го рода, расчет производился по методу ближайших соседей. На рис. 4.16 приведены аналогичные данные для иерархического подхода. Для наглядности отдельные точки соединены прямыми линиями, что позволяет

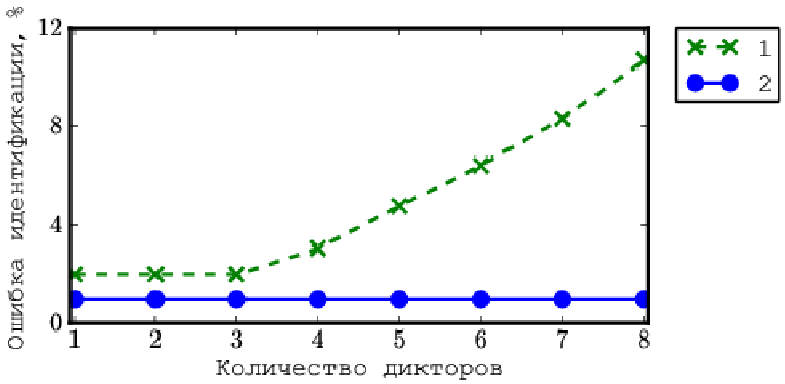


Рис. 4.15. Влияние количества дикторов в калибровочной базе (метод ближайших соседей).

выявить тенденцию в изменении величины ошибки. Как видно из рис. 4.15 и рис. 4.16, ошибка 1-го рода возрастает при увеличении числа дикторов, что связано, как было показано ранее (раздел 3.2), с перекрыванием амплитудно-частотных характеристик голосов. Ошибка 2-го рода пренебрежима мала, это обусловлено жестким заданием границ областей счетов дикторов.

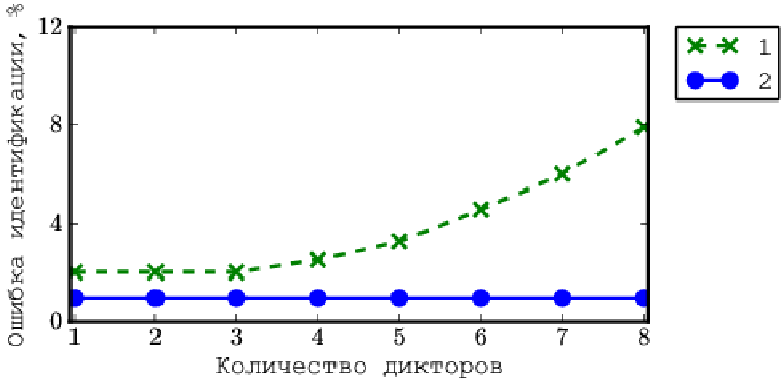


Рис. 4.16. Влияние количества дикторов в калибровочной базе (иерархический подход).

Сопоставление рис. 4.15 и рис. 4.16 показывает преимущество иерархического метода перед модифицированным методом ближайших соседей в случае большого числа дикторов.

На рис. 4.17 исследуется влияние числа повторений фраз калибровочного набора на ошибки идентификации (на примере произношения одного слова). Видно, что ошибки 1-го рода уменьшаются при увеличении числа повторений фраз, ошибки 2-го рода пренебрежимо малы. Расчеты проводились в рамках иерархического метода, результаты расчета для модифицированного метода ближайших соседей качественно аналогичны.

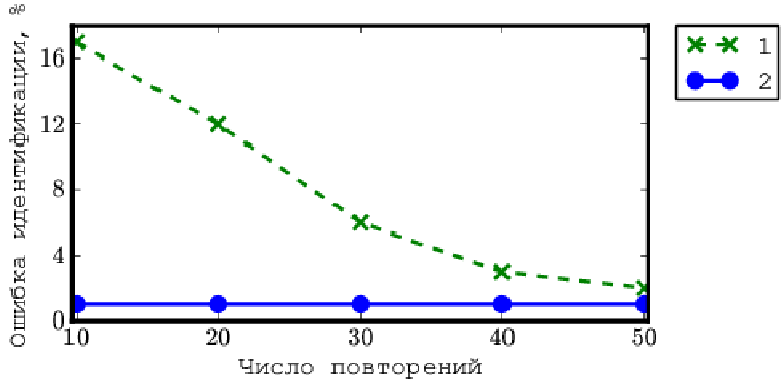


Рис. 4.17. Влияние числа повторений фразы в калибровочной базе.

На рис. 4.18 выясняется роль длительности фраз, а именно, при увеличении длительности фразы ошибки 1-го рода в целом уменьшаются в

отношении разброса точек, ошибки 2-го рода по-прежнему пренебрежимо малы.

Для использования системы идентификации в реальных условиях необходимо провести количественную оценку влияния шумов и искажений на

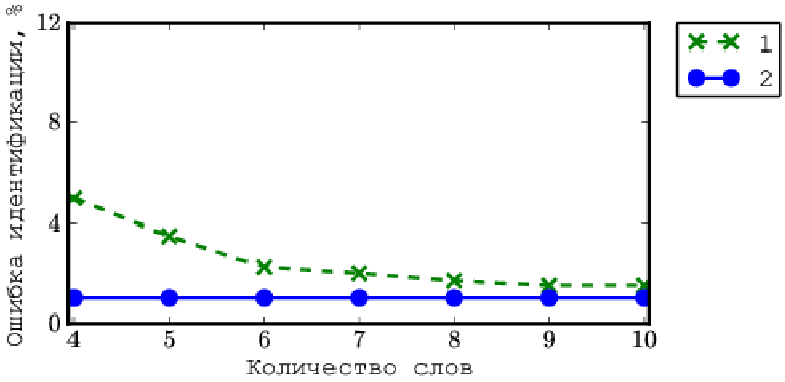


Рис. 4.18. Влияние длительности фразы в калибровочной базе.

результат идентификации. Расчет ошибок 1-го и 2-го рода в случае искажений, задавшихся согласно разделу 4.1, осуществлялся с помощью выражений (4.11), (4.12), (4.13), (4.14). На рис. 4.19 приведены зависимости ошибок 1-го рода (точки 1) и 2-го рода (точки 2) от числа добавляемых нулевых отсчетов искажения. Из рис. 4.18 видно, что увеличение числа добавляемых нулевых отсчетов ведет к быстрому росту ошибки 1-го рода, а ошибка 2-го рода при этом почти не изменяется.

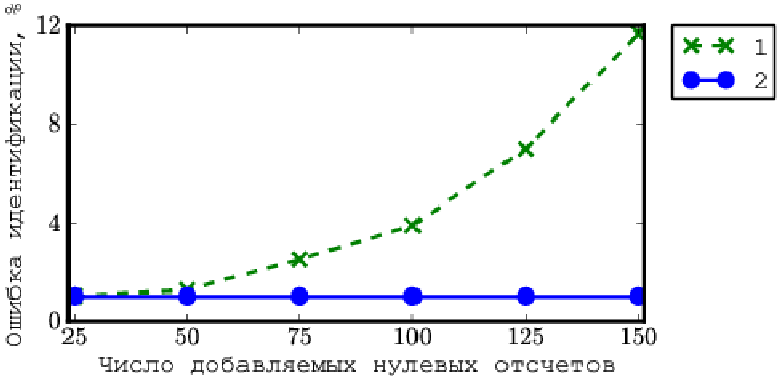


Рис. 4.19. Ошибки 1-го и 2-го рода в случае искажения спектра голосового сигнала.

В случае, когда исходные калибровочные данные содержат шум, то ошибка идентификации растет медленнее, чем при отсутствии шума в исходных данных. На рис. 4.20 изображены зависимости доли правильно идентифицируемых от величины воздействия шума улицы для вышеуказанных случаев, а именно, точки 1 – исходные данные без шума, точки 2 – исходные калибровочные данные содержат шум улицы (R_{SN} = 10 дБ). В дальнейших расчетах для сопоставления используется первый случай (исходные данные без шума).

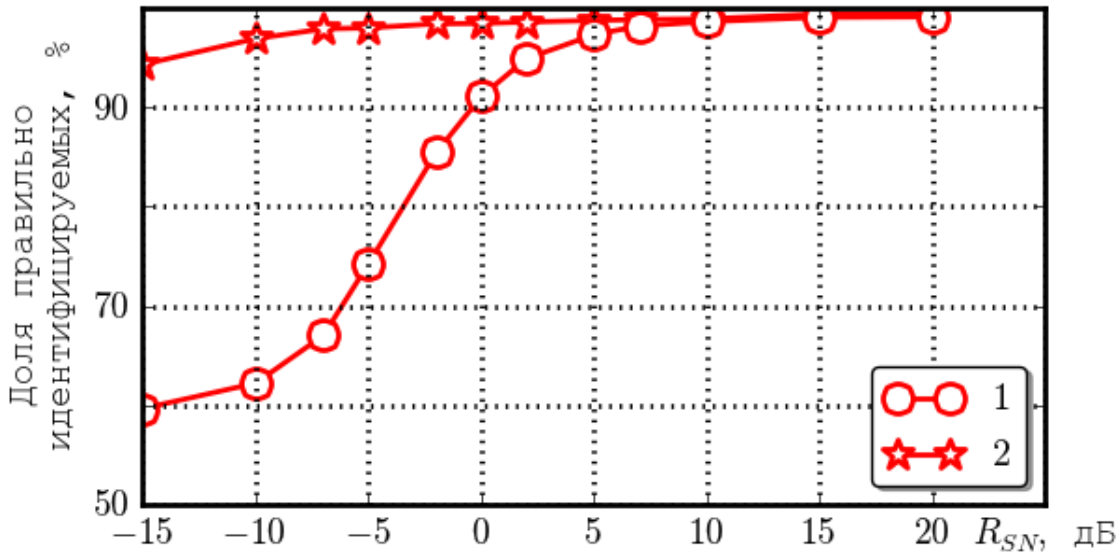


Рис. 4.20. Влияние исходных данных на результат идентификации.
1 – исходные данные без шума, 2 – исходные калибровочные данные содержат шум улицы (R_{SN} = 10 дБ).

Проведенные расчеты ошибок идентификации в условиях внешнего шума использовались для выявления надежности и преимуществ предложенного подхода по сравнению с другими возможными методами классификации голосовых сигналов. Для качественного сопоставления привлекались имеющиеся работы других авторов, в которых идентификация по голосовым данным проводилась в аналогичных условиях с использованием таких методов, как гауссовы смеси, нейронные сети и опорные вектора, описанных в разделе 1.2.

На рис. 4.21 приведены результаты расчетов доли правильно идентифицируемых дикторов в условиях наложения шума улицы, полученные с помощью метода гауссовых смесей в работах [46] (кривая 2), [41] (кривая 3), [129] (кривая 4) и в настоящей работе (кривая 1). Расчет доли правильно идентифицируемых дикторов производился с учетом ошибок как первого, так и второго рода согласно формулам (4.13), (4.14). В работах [41,46,129] идентификация в связи со значительным ухудшением результатов проводилась до уровня шума, равного исходному голосовому сигналу. Видно, что кривые 1–4 ведут себя качественно одинаковым образом, при этом результаты наших расчетов обладают определенными преимуществами. Именно, предложенный в настоящей работе метод обеспечивает большую надежность идентификации при увеличении уровня шума до $R_{SN} \sim 0$ дБ (кривая 1), тогда как сопоставляемые данные, основанные на другом методе, свидетельствуют о заметно более быстром ухудшении идентификации с ростом шума (кривые 2, 3, 4).

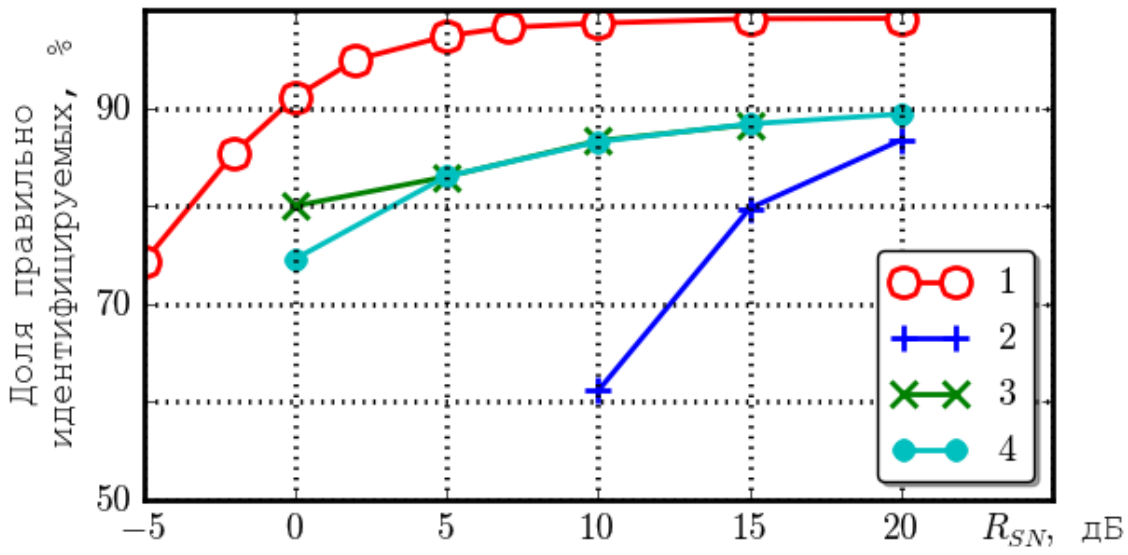


Рис. 4.21. Результаты расчетов доли правильно идентифицируемых дикторов в условиях наложения шума улицы.
1 – настоящая работа, 2 – [46], 3 – [41], 4 – [129].

На рис. 4.22 приведены результаты аналогичного сопоставления для условий наложения на исходный сигнал «белого» шума, при этом доля правильно идентифицируемых дикторов рассчитывалась как методом гауссовых смесей (работа [76], кривая 2), так и методом нейронных сетей (работа [74] , кривая 3). Из рис. 4.21 следует, что результаты наших расчетов (кривая 1) обеспечивают в интервале до $R_{SN} \sim 10$ дБ преимущества, аналогичные полученным для случая шума улицы. Так, ошибка идентификации при $R_{SN} \sim 10$ дБ для наших расчетов составляет 6 %, для расчетов по методу гауссовых смесей – порядка 8%, по методу нейронных сетей – порядка 20%. Отметим, что в то же время при более высоком уровне шума результаты расчетов уступают данным работы [76]. Это связано с тем, что в работе [76] при идентификации использовался метод, позволявший разделить зашумленный речевой сигнал на речь и шум, после чего применялся метод восстановления поврежденных данных.

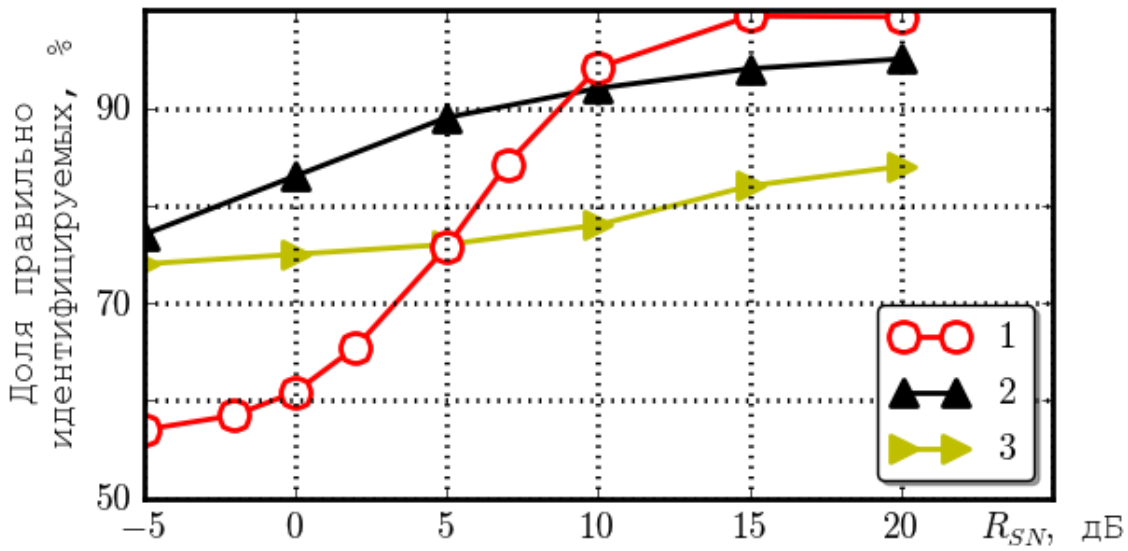


Рис. 4.22. Результаты расчетов доли правильно идентифицируемых дикторов в условиях наложения «белого» шума.
1 – настоящая работа, 2 – [76], 3 – [74].

На рис. 4.23 приведены результаты аналогичных расчетов для случая речеподобного шума, сопоставление наших расчетов (точки 1) проводилось с

работами [76] и [78] (точки 2 и 3), основанными на использовании нейронных сетей разного вида, и с работой [129] (точки 4). Данный случай представляет наибольший интерес, поскольку речеподобный шум вызывает наиболее сильное влияние на голосовую идентификацию. Это проявляется в том, что на рис. 4.23 точки 1–4 быстрее спадают с ростом уровня шума, чем на рис. 4.22 и рис. 4.21. В то же время, как и в предыдущих случаях, рассчитанные по предложенному в настоящей работе методу точки 1 демонстрируют преимущества перед точками 2–4, отражающие более надежную идентификацию при одинаковом уровне шума, то есть уровень ошибки меньше, а также обеспечивает большую помехоустойчивость.

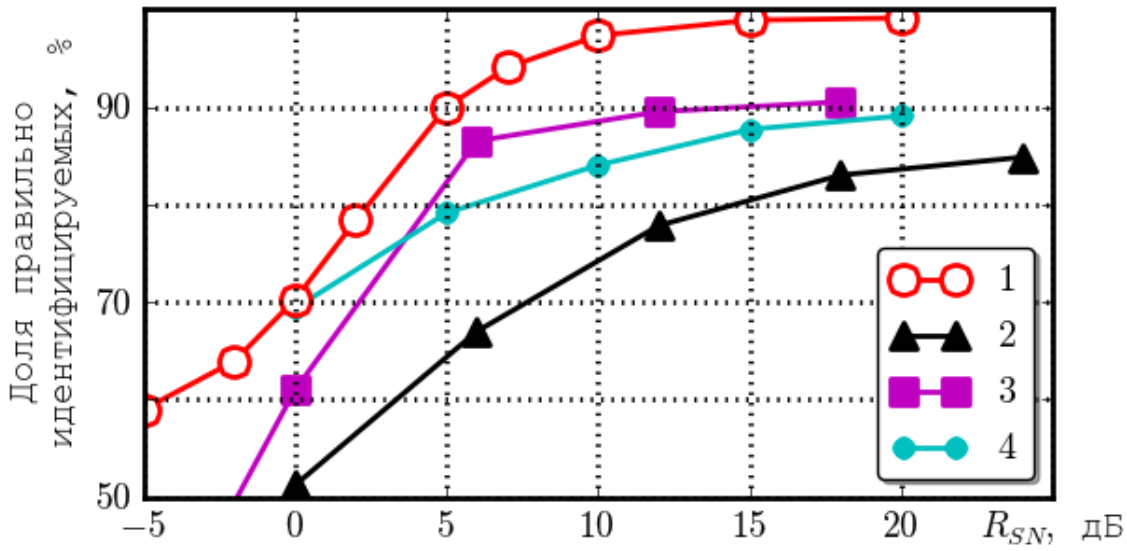


Рис. 4.23. Результаты расчетов доли правильно идентифицируемых дикторов в условиях наложения речеподобного шума.
1 – настоящая работа, 2 – [76], 3 – [78], 4 – [129].

Ошибки 1-го и 2-го рода, на основе которых рассчитывались доли правильно идентифицируемых дикторов, возрастали с увеличением уровня шума по-разному, при этом наиболее опасные ошибки 2-го рода увеличивались значительно медленнее, чем ошибки 1-го рода. Это является достоинством предложенного метода. Так, при $R_{SN} = 5$ дБ ошибки первого рода были порядка 12 %, а второго рода – не превышали 3%.

Кроме того, был произведен расчет доли правильно идентифицируемых дикторов с использованием методов настоящей работы, гауссовых смесей, нейронных сетей и машины опорных векторов для одной и той же базы речевых данных. Именно, для всех случаев наложения различных видов шума использовалась специально разработанная нами база [99] и стандартные реализации вышеуказанных методов в программе MatLab. При практически отсутствующем шуме ($R_{SN} = 20$ дБ) доля правильно идетифицируемых составляла от 97% до 99%. При увеличении шума до уровня $R_{SN} = 5$ дБ методы настоящей работы обеспечили долю правильно идетифицируемых дикторов, на 12–15% большую, чем остальные методы. Кроме того, при дальнейшем увеличении уровня шума доля правильно идетифицируемых дикторов при использовании методов настоящей работы снижалась медленнее, чем реализации других методов.

Для количественной оценки надежности и преимуществ развитого метода проводилось также сопоставление с используемыми на практике системами голосовой идентификации личности, при этом использовалась одна и та же созданная нами база голосовых данных дикторов [99], описанная в главе 2. Результаты такого сопоставления полностью подтвердили сделанные выше выводы о преимуществах развитого в настоящей работе подхода в условиях наложения внешнего шума.

Отметим, что отдельный интерес представляет количественное сопоставление влияния разных видов шума [130,131]. Расчеты ошибок показывают, что наибольшее влияние на результат голосовой идентификации оказывает, как и следовало ожидать, речеподобный шум, наименьшее – шум улицы; «белый» «розовый» и «коричневый» шумы занимают промежуточное положение. Эти могут оказаться полезными для анализа защищенности речевой информации от утечек по акустическому каналу.

Укажем также, что результаты расчетов выявили важность наличия надежной базы голосовых данных для идентификации. В качестве примера были проведены расчеты, позволявшие сравнить базу данных, сформированную в настоящей работе, с базой данных «Речевой корпус данных TeCoRus», предоставленной ВЦ РАН (г. Москва) [132]. База данных «TeCoRus» представляла из себя две части – акустико-фонетическую и речевую. В первой части использовались записи голосов 6 дикторов и содержались 3050 предложений, во вторую часть были включены 100 дикторов, произносивших различные фразы. Достоинством базы данных «TeCoRus» являлось большое число дикторов, а недостатком – отсутствие повторений голосового сигнала одного диктора. Тестовые расчеты показали близость получавшихся с использованием обеих баз данных. Кроме того, было установлено, что надежность идентификации существенно зависит от числа повторов голосовых данных одного диктора. Это обстоятельство необходимо учитывать при формировании голосовых баз данных.

В целом приведенные результаты расчетов ошибок 1-го и 2-го рода свидетельствуют о достаточной надежности и достоверности развитого метода, прежде всего для наиболее опасных ошибок 2-го рода. Кроме того, полученные данные могут быть использованы для выработки методических рекомендаций в построении надежных голосовых баз данных, используемых при идентификации.

ЗАКЛЮЧЕНИЕ

В работе развит подход, основанный на применении математического аппарата анализа многомерных данных к обработке и классификации акустических голосовых сигналов, характеризующих индивидуальные биометрические особенности дикторов. Проведено исследование возможностей голосовой идентификации, проводимой с помощью проекционных методов анализа. Предложены методики, позволяющие проводить классификацию голосовых сигналов при наложении нескольких голосов. Изучено влияние вида и величины внешнего шума, различных помех и искажений на идентифицируемые речевые сигналы.

Полученные результаты позволяют сделать следующие **основные выводы**.

1. Выбраные информативные признаки акустических голосовых сигналов на основе коэффициентов мел-кепстрального разложения позволили при применении проекционных методов повысить разделимость голосов идентифицируемых дикторов.

2. Разработаные алгоритмы обработки и классификации голосовых сигналов, основанные на использованиие методов главных компонент и проекций на латентные структуры, обеспечили повышение надежности идентификации, проявившееся в уменьшении ошибок 1-го и 2-го рода.

3. Предложенный метод классификации акустических голосовых сигналов, использующий метрику Махаланобиса для построения многомерных эллипсоидов, характеризующих сигналы отдельных дикторов, обеспечил высокую наглядность результатов идентификации. Разработанное иерархическое разбиение кластеров голосовых сигналов дало возможность позволило разделить голоса дикторов с близкими физическими характеристиками.

4. Разработанная база данных с повторами голосовых сигналов дикторов позволила установить влияние основных параметров голосовых записей на результат идентификации. Эта база данных может быть использована при тестировании аналогичных методов и систем голосовой идентификации.

5. Разработанный подход дал возможность повысить число правильно идентифицируемых дикторов в условиях значительного уровня внешнего акустического шума и уменьшить влияние роста этого шума на величину ошибок идентификации.

6. Проведенное с единых позиций систематическое исследование позволило выявить влияние вида внешнего акустического шума на голосовую идентификацию, что может быть использованы при разработке и тестировании систем защиты от утечек по акустическому каналу.

СПИСОК ЛИТЕРАТУРЫ

1. Доктрина информационной безопасности РФ // Российская газета. – 2000. – 28 сент.

2. Головин, А.В. Уголовно-правовые и криминологические проблемы защиты информации / Головин А.В., Исаев А.А., Мазуров В.А., Поляков В.В., Сидоренко Т.В. – Алматы : Изд. центр ОФППИ Интерлигал, 2008. – 338 с.

3. ГОСТ Р–ИСО/МЭК 17799-2005. Информационная технология. Практические правила управления информационной безопасностью. – М. : Стандартинформ, 2006.

4. Minkyu, C. Biometric Authentication: A Review / C. Minkyu, F. Alisherov A., R. Rahul [at al.] // International Journal of u- and e- Service, Science and Technology. – 2009. – vol. 2 (3). – pp. 13–28.

5. Chollet, G. Guide to Biometric Reference Systems and Performance Evaluation / G. Chollet, B. Dorizzi, D. Petrovska-Delacr´ etaz. – London : Springer, 2009. – 412 p.

6. Малинин, П.В. Применение корреляционно-признакового метода в аппаратной системе идентификации личности по радужной оболочке глаза / П.В. Малинин, И.Н. Свистун, В.В. Поляков // Современные информационные технологии в науке, образовании и практике : матер. VI Всерос. науч.-практ. конф. г. Оренбург, 27–28 ноября 2007 г. – Оренбург : Изд-во ИПК ГОУ ОГУ, 2007. – С. 447–448.

7. Малинин, П.В. Идентификация личности на основе комбинированных данных отпечатка пальцев и изображения лица / П.В. Малинин // Проблемы правовой и технической защиты информации: сб. ст. / под ред. В.В. Полякова, В.А. Мазурова. – Барнаул: Изд-во Алт. ун-та, 2008. – С. 163–166.

8. Иконин, С.Ю. Система автоматического распознавания речи SPIRIT ASR Engine / С.Ю. Иконин, Д.В. Сарана // Цифровая обработка сигналов. – 2003. – №4. – С. 5–13.

9. Mariethoz, J. Speaker Verification Based on User-Customized Password / J. Mariethoz, B. Herve, M.F. BenZeghiba // IDIAP Research Report 01-13. – Martigny, 2001. – 22 p.

10. Фант, Г. Акустическая теория речеобразования / Г. Фант; Пер. с англ. – М. : Наука, 1964. – 284 с.

11. Фланаган, Дж.Л. Анализ, синтез и восприятие речи / Дж.Л. Фланаган; пер. с англ. А.А. Пирогова. – М. : Связь, 1968. – 396 с.

12. Pellandini, F. GSM Speech Coding And Speaker Recognition / F. Pellandini, M. Ansorge, A. Dufaux [at al.] // International Conference on Acoustics, Speech, and Signal Processing (ICASSP): Book of abstracts. – Istanbul, 2000. – vol. 2. – pp. 1085–1088.

13. Amrouche, A. Effect of GSM speech coding on the performance of Speaker Recognition System / A. Amrouche, A. Krobba, M. Debyeche // 10th International Conference on Information Sciences Signal Processing and their Applications (ISSPA): Book of abstracts. – Kuala Lumpur, 2010. – pp. 137–140.

14. Дворянкин, С.В. О необходимости новых подходов к оценке эффективности технических средств акустозащиты / С.В. Дворянкин // Информация и безопасность. – 2002. – №2. – С. 244–245.

15. Андреев, Б.В. Расследование преступлений в сфере компьютерной информации / Б.В. Андреев, П.Н. Пак, В.Н. Хорст. – М. : Юрлитинформ, 2001. – 152 с.

16. Вехов, В.Б. Компьютерные преступления: способы совершения и раскрытия / В.Б. Вехов. – М. : Право и Закон, 1996. – 182 с.

17. Усов, А.И. Судебно-экспертное исследование компьютерных средств и систем: основы методического обеспечения / А.И. Усов. – М. : Изд-во «Экзамен», изд-во «Право и закон», 2003. – 368 с.

18. Сорокин, В.Н. Определение пола диктора по голосу / В.Н. Сорокин, И.С. Макаров // Акустический журнал. – 2008. – Т. 54. – № 4. – С. 659–668.

19. Галунов, В.И. О возможности определения эмоционального состояния говорящего по речи / В.И. Галунов // Речевые технологии. – 2008. – № 1. – С. 60–66.

20. Ромашкин, Ю.Н. Распознавание пола диктора на основе gmm-модели голоса / Ю.Н. Ромашкин, Ю.О. Петров // Речевые технологии. – 2009. – № 2. – С. 31–38.

21. Neustein, A. Forensic Speaker Recognition: Law Enforcement and Counter-Terrorism / A. Neustein, H.A. Patil. – New York : Springer, 2012. – 540 p.

22. Каганов, А.Ш. Криминалистическая идентификация личности по голосу и звучащей речи / А.Ш. Каганов. – М. : Юрлитинформ, 2009. – 291 с.

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

24. Гребнов, С.В. Разработка и реализация двухуровневого метода голосового управления на основе скрытых марковских моделей / С.В. Гребнов // Информационные технологии. – 2009. – № 9. – С. 40–46.

25. Кривнова, О.Ф. Области применения речевых корпусов и опыт их разработки / О.Ф. Кривнова // Труды xviii сессии Российского акустического общества РАО. – Таганрог, 2006. – С. 81–84.

26. Дшхунян, В.Л. Электронная идентификация. Бесконтактные электронные идентификаторы и смарт-карты / В.Л. Дшхунян, В.Ф. Шаньгин. – М. : ООО «Издательство АСТ», Издательство «НТ Пресс», 2004. – 695 с.

27. Orság, F. Speaker Dependent Coefficients for Speaker Recognition / F. Orság // International Journal of Security and Its Applications. – 2010. – vol. 4. – pp. 31–47.
28. Meng, H. The HKCUPU system for the NIST 2010 speaker recognition evaluation / W. Jiang, Man-Wai Mak, W. Rao, H. Meng // International Conference on Acoustics, Speech and Signal Processing (ICASSP): Book of abstracts. – Prague, 2011. – pp. 5288–5291.
29. Stolbov, M. Channel compensation for forensic speaker identification using inverse processing / M. Stolbov, P. Ignatov, S. Koval, A. Barinov // AES 39th International Conference. – Hillerød, 2010.
30. Diaz-de-Maria F. Robust ASR using Support Vector Machines / F. Diaz-de-Maria, C. Peláez-Moreno, A. Gallardo-Antol in [at al.] // Speech Communication. – 2007. – vol. 49 (4). – pp. 253–267.
31. Руководящий документ Гостехкомиссии России. «Защита от несанкционированного доступа к информации. Термины и определения». – М. : ГТК РФ, 1992. – 13 с.
32. Иванов, А.И. Биометрическая идентификация личности по динамике подсознательных движений / А.И. Иванов. – Пенза : Изд-во Пенз.гос.ун-та, 2000. – 187 с.
33. Дворянкин, С.В. О практическом применении технологии речевой подписи / С.В. Дворянкин // Конференция «Телекоммуникационные и вычислительные системы» в рамках международного конгресса Коммуникационные технологии и сети» (CTN-2002), международного форума информатизации (МФИ 2002): матер. конф. – М., 2002. – С. 159.
34. Huang, Q. Speaker Detection In Broadcast Speech Databases / Q. Huang, S. Parthasarathy, I. Magrin-Chagnolleau, A. E. Rosenberg // International Conference on Spoken Language Processing: Book of abstracts. – Sydney, 1998. – vol. 4. – pp. 1339–1342.

35. Torres-Carrasquillo, P. Approaches and applications of audio diarization / P. Torres-Carrasquillo, D. A. Reynolds // International Conference on Acoustics, Speech, and Signal Processing (ICASSP): Book of abstracts. – Philadelphia, 2005. – vol. 5. – pp. 953–956.

36. Галяшина, Е.И. Идентификация дикторов по цифровым фонограммам / Е.И. Галяшина // Речевые технологии. – 2010. – № 3. – С. 23–28.

37. Женило, В.Р. Компьютерная фоноскопия / В.Р. Женило. – М. : Изд-во Академия МВД России, 1995. – 206 с.

38. Huang, X. Spoken language processing: a guide to theory, algorithm, and system development / X. Huang, A. Acero, H.W. Hon. – New Jersey : Prentice Hall, 2001. – 1008 p.

39. Kim, H. Noise-Robust Speaker Recognition Using Subband Likelihoods and Reliable-Feature Selection / H. Kim, M. Ji, S. Kim // ETRI Journal. – 2008. – vol. 30 (1). – pp. 89–100.

40. Козлов, А.В. Алгоритм детектирования музыкальных фрагментов в задачах речевой обработки / А.В. Козлов, А.И. Лоханова, К.К. Симончик // «Научно-технические ведомости СПбГПУ». (103), – 2010. – № 4 (103). – С. 7–11.

41. Crookes, D. Speaker recognition in noisy conditions with limited training data / D. Crookes, J. Ming, N. McLaughlin // 19th European Signal Processing Conference (EUSIPCO 2011): Book of abstracts. – Barcelona, 2011. – pp. 1294–1298.

42. Reynolds, D.A. The effects of handset variability on speaker recognition performance: experiments on the switchboard corpus / D.A. Reynolds // ICASSP-96. – 1996. – vol. 1. – pp. 113.

43. Scheffer, N. Towards noise-robust speaker recognition using probabilistic linear discriminant analysis / N. Scheffer, M. Graciarena, L. Ferrer, L. Burget, Y. Lei //

International Conference on Acoustics, Speech, and Signal Processing (ICASSP): Book of abstracts. – Kyoto, 2012. – pp. 4253–4256.

44. The NIST Year 2008 Speaker Recognition Evaluation Plan [Электронный ресурс] – Режим доступа: http://www.itl.nist.gov/iad/mig/tests/sre/2008/sre08_evalplan_release4.pdf – Загл. с экрана.

45. The NIST Year 2010 Speaker Recognition Evaluation Plan [Электронный ресурс] – Режим доступа: http://www.nist.gov/itl/iad/mig/upload/nist_sre10_eval plan-r6.pdf – Загл. с экрана.

46. Reynolds, D.A. Robust Speaker Recognition in Noisy Conditions / D.A. Reynolds, J.R. Glass, T.J. Hazen, J. Ming // Audio, Speech, and Language Processing. july – 2007. – vol. 15 (5). – pp. 1711–1723.

47. Zue, V. Speech database development at MIT: TIMIT and beyond / V. Zue, S. Seneff, J. Glass // Speech Communication. – 1990. – vol. 9. – pp. 351–356.

48. Sridharan, S. Robust Speaker Recognition using Microphone Arrays / S. Sridharan, J. Pelecanos, I. McCowan // A Speaker Odyssey – The Speaker Recognition Workshop: Book of abstracts. – Crete, 2001. – pp. 101–106.

49. Reynolds, D.A. Channel robust speaker verification via feature mapping / D.A. Reynolds // International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hong Kong, China, 06 Apr – 10 Apr 2003. – 2003. – vol. 2. – pp. 53–56.

50. Laface, P. Channel factors compensation in model and feature domain for speaker recognition / P. Laface, E. Dalmasso, F. Castaldo [at al.] // Odyssey: Speaker and Language Recognition Workshop, San Juan, 28-30 June 2006 .

51. Sridharan, S. Modeling session variability in text-independent speaker verification / Sridharan S., Baker B., Vogt R. // 9th European Conference on

Speech Communication and Technology, September 4-8 in Lisbon, Portugal. – 2005. – pp. 3117-3120.

52. Крак, Ю.В. Система распределённого автоматизированного документирования речевых сигналов / Ю.В. Крак, А.С. Загваздин // Речевые технологии. – 2012. – № 2. – С. 43-53.

53. Gauvain, J. Feature and score normalization for speaker verification of cellular data / J. Gauvain, C. Barras // International Conference on Acoustics, Speech, and Signal Processing (ICASSP): Book of abstracts. – Hong Kong, 2003. – vol. 2. – pp. 49–52.

54. Morgan, N. RASTA processing of speech / H. Hermansky, N. Morgan // Transactions on Speech and Audio Processing. – 1994. – vol. 2 (4). – pp. 578–589.

55. Tiwary, U.S. Text independent speaker identification using wavelet transform / U.S. Tiwary, G.K. Verma // International Conference on Computer and Communication Technology (ICCCT): Book of abstracts. – Allahaba, 2010. – pp. 130–134.

56. S.C. Woo, C.P. Lim, and R. Osman. Development of a speaker recognition system using wavelets and artificial neural networks / R. Osman, C.P. Lim, S.C. Woo // International Symposium on Intelligent Multimedia, Video and Speech Processing: Book of abstracts. – Hong Kong, 2001. – pp. 413–416.

57. Ахмад, Х.М. Сравнительное исследование эффективности различных методов кепстрального описания речевых сигналов в задачах распознавания / Х.М. Ахмад // Вестник ТГТУ. – 2007. – Том 13, № 4. – С. 887–891.

58. Reynolds, D.A. Speaker Verification Using Adapted Gaussian Mixture Models / D. A. Reynolds, T. F. Quatieri, R. B. Dunn // Digital Signal Processing. – 2000. – vol. 10. – pp. 19–41.

59. Tutorial on Text-Independent Speaker Verification / D.A. Reynolds, D. Petrovska-Delacr´ etaz, J. Ortega-Garcá [at al.] // EURASIP Journal on Applied Signal Processing. – 2004. – vol. 4. – pp. 430–451.

60. Zhang, Y. Optimization of GMM Training For Speaker Verification / Y. Zhang, M. Scordilis // Odyssey 2004 -- The Speaker and Language Recognition Workshop, Toledo, Spain May 31 -- June 3, 2004. -- Toledo, 2004. -- pp. 231--236.
61. MLLR techniques for speaker recognition / M. Ferras, C. C. Leung, C. Barras, J.-L. Gauvain // Odyssey 2008 -- The Speaker and Language Recognition Workshop, Stellenbosch, South Africa January 21--24, 2008. -- Stellenbosch, 2008. -- pp. 21--24.
62. Rabiner, L. R. A tutorial on hidden markov models and selected applications in speech recognition / L. R. Rabiner // Proceedings of the IEEE. 1989. – vol. 77 (2). – pp. 257–286.
63. Carey, M. Speaker Recognition using a Trajectory-Based Segmental HMM / M. Carey, M. Russell, Y. Liu // Proceedings of Odyssey – Speaker and Language Recognition Workshop: Book of abstracts. – Toledo, 2004. – pp. 45–50.
64. Arslan, L.M. HMM-based text-dependent speaker recognition with handset-channel recognition / L.M. Arslan, O. Bvyük // Signal Processing and Communications Applications Conference: Book of abstracts. – Diyarbakir, 2010. – pp. 383–386.
65. Nakagawa, S. Text-Independent/Text-Prompted Speaker Recognition by Combining Speaker-Specific GMM with Speaker Adapted Syllable-Based HMM / S. Nakagawa, W. Zhang, M. Takahashi // IEICE Transactions on Information and Systems. March – 2006. – vol. 89 (3). – pp. 1058–1065.
66. Хайкин, С. Нейронные сети: полный курс / С. Хайкин; пер. с англ. Н.Н. Куссуль, А.Ю. Шелестова, под ред. Н.Н. Куссуль. – М. : Издательский дом «Вильямс», 2006. – 1104 с.
67. Геппенер, В.В. Разработка систем автоматической верификации дикторов с использованием нейронных сетей / В. В. Геппенер, К. К. Симончик // Нейрокомпьютеры: разработка, применение. – 2006. – № 7. – С. 14–23.

68. Novakovic, J. Speaker identification in smart environments with multilayer perceptron / J. Novakovic // Telecommunications Forum (TELFOR): Book of abstracts. – Belgrade, 2011. – pp. 1418–1421.

69. Вапник, В.Н. Теория распознавания образов (статистические проблемы обучения) / В.Н. Вапник, А.Я. Червоненкис. – М. : Издательство «Наука», 1974. – 416 с.

70. Renals, S. Speaker verification using sequence discriminant support vector machines / V. Wan, S. Renals // Speech and Audio Processing. – 2005. – vol. 13 (2). – pp. 203-210.

71. Reynolds, D.A. Support vector machines using GMM supervectors for speaker verification / D.A. Reynolds, D.E. Sturim, W.M. Campbell // Signal Processing Letters. – 2006. – vol. 13 (5). – pp. 308–311.

72. Речевой корпус данных TIMIT [Электронный ресурс] – Режим доступа: http://catalog.ldc.upenn.edu/LDC93S1 – Загл. с экрана.

73. Речевой корпус данных SPIDRE [Электронный ресурс] – Режим доступа: http://catalog.ldc.upenn.edu/LDC94S15 – Загл. с экрана.

74. Togneri, R. Robust Speaker Identification using combined feature selection and missing data recognition / R. Togneri, M. Kuhne, D. Pullella // International Conference on Acoustics, Speech, and Signal Processing (ICASSP). – 2008. – pp. 4833–4836.

75. Речевой корпус данных TIDIGITS [Электронный ресурс] – Режим доступа: http://catalog.ldc.upenn.edu/LDC93S10 – Загл. с экрана.

76. Wang, D. Robust Speaker Identification using a casa front-end / D. Wang, Y. Shao, X. Zhao // International Conference on Acoustics, Speech, and Signal Processing (ICASSP). – 2011. – pp. 5468–5471

77. Речевой корпус данных 2002 NIST [Электронный ресурс] – Режим доступа: http://catalog.ldc.upenn.edu/LDC2004S04 – Загл. с экрана.

78. Wang, D. Robust Speaker Identification in Noisy and Reverberant Conditions / D. Wang, Y. Wang, X. Zhao // IEEE Transactions on Audio, Speech and Language Processing 2014.
79. Речевой корпус данных 2008 NIST [Электронный ресурс] –
Режим доступа: http://catalog.ldc.upenn.edu/LDC2011S08 – Загл. с экрана.
80. Soong, F.K. The Use of Cohort Normalized Scores for Speaker Verification / F.K. Soong, B. Juang, C. Lee [at al.] // International Conference on Spoken Language Processing (ICSLP): Book of abstracts. – Banff, 1992. – pp. 599–602.
81. Cohen, A. On Cohort Selection for Speaker Verification / A. Cohen, Y. Zigel // European Conference on Speech Communication and Technology: Book of abstracs. – Geneva, 2003. – pp. 2977–2980.
82. Hansen, J.H.L. A Study on Universal Background Model Training in Speaker Verification / J.H.L. Hansen, T. Hasan // Transactions on Audio, Speech, and Language Processing. – 2011. – vol. 19 (7). – pp. 1890–1899.
83. Wu. W. A tree-based kernel selection approach to efficient Gaussian mixture model–universal background model based speaker identification / Z. Xiong and T. F. Zheng and Z. Song and F. Soong and W. Wu // Speech Communication. – 2006. – vol. 48 (10) – pp. 1273–1282.
84. Morrison, G.S. A comparison of procedures for the calculation of forensic likelihood ratios from acoustic–phonetic data: Multivariate kernel density (MVKD) versus Gaussian mixture model–universal background model (GMM–UBM) / G.S. Morrison // Speech Communication. – 2011. – vol. 53 (2). – pp. 242–256.
85. Przybocki, M. The DET curve in assessment of detection task performance / M. Przybocki, M. Ordowski, T. Kamm [at al.] // European Conference on Speech Communication and Technology: Book of abstracts. – Rhodes, 1997. – pp. 1895–1898.
86. Hair, Jr. J. F. Multivariate data analysis (7th ed.) / J.F. Hair, R.E. Anderson, R.L. Tatham, W.C. Black. – New Jersey : Prentice Hall, – 2010. – 816 p.

87. Многомерный анализ изображений в медицине и биологии / И.А. Беляев, С.В. Кучерявский, О.Е. родионова, А.Л. Померанцев // Энциклопедия инженера-химика. – 2008. – № 10. – С. 27–36.
88. Brereton, R.G. Introduction to multivariate calibration in analytical chemistry / R.G. Brereton // Analyst. – 2000. – vol. 125. – pp. 2125–2154.
89. Беляев, И.А. Применение принципов формального моделирования для классификации медицинских изображений / И.А. Беляев, С.В. Кучерявский // XIV международная конференция «Математика. Компьютер. Образование»: сб. трудов. – Дубна, 2007. – т. 2. – С. 355–361.
90. Поляков, В.В. Применение многомерного анализа данных к физико-механическим характеристикам биоминеральных образований / В.В. Поляков, Г.Г. Устинов, Е.В. Петрухно // Биомедицинские технологии. – 2011. – № 12. – С. 56–62.
91. Дмитриев, А.А. Применение методов многомерного анализа данных для диагностики ранних стадий болезни Альцеймера по томографии мозга / А.А. Дмитриев, С.В. Кучерявский // Известия АлтГУ. – 2010. – №1/2. – С. 112–115.
92. Дрейпер, Н. Прикладной регрессионный анализ. Множественная регрессия / Н. Дрейпер, Г. Смит. – М. : «Диалектика», 2007. – 912 с.
93. Себер, Дж. Линейный регрессионный анализ / Дж. Себер. – М. : Мир, 1980. – 187 с.
94. Hoskuldsson, A. Prediction methods in Science and Technology / A. Hoskuldsson. – Denmark : Thor Publ, 1996. – 157 p.
95. Эсбенсен, К. Анализ многомерных данных. Избранные главы / К. Эсбенсен; под ред. О.Е. Родионовой; пер. с англ. С.В. Кучерявского. – Черноголовка : Изд-во ИПХФ РАН, 2005. – 157 с.
96. Померанцев, А.Л. Хемометрика в аналитической химии / А.Л. Померанцев, О.Е. Родионова. – М. : Институт химической физики им. Н.Н. Семенова, 2003. – 61 с.

97. Айвазян, С.А. Прикладная статистика: Классификация и снижение размерности / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин. – М. : Финансы и статистика, 1989. – 607 с.

98. Звуковой редактор audacity [Электронный ресурс] – Режим доступа: http://audacity.sourceforge.net – Загл. с экрана.

99. Малинин, П.В. Акустические речевые сигналы для системы идентификации личности по голосовым данным. Свидетельство о государственной регистрации базы данных №2013620132, 9 января 2013 г.

100. Малинин, П.В. Алгоритмы обработки звуковых сигналов при формировании баз голосовых данных / П.В. Малинин, В.В. Поляков // Многоядерные процессоры, параллельное программирование, ПЛИС, системы обработки сигналов : сб. статей Всерос. науч.-практ. конф. г. Барнаул, 28 февраля 2013 г. – Барнаул : изд-во Алт. ун-та, 2013. – С. 85–89.

101. Рамишвили, Г.С. Автоматическое опознавание говорящего по голосу / Г.С. Рамишвили. – М. : Радио и связь, 1981. – 224 с.

102. Программный пакет матричных вычислений MatLab [Электронный ресурс] – Режим доступа: http://www.mathworks.com/ – Загл. с экрана.

103. Vaseghi, S.V. Advanced Digital Signal Processing and Noise Reduction / S.V. Vaseghi. – New York : John - Wiley & Sons, 2006. – 453 p.

104. Григорян, Р.Л. Сравнение различных способов оценки схожести распределений частоты основного тона в задаче идентификации диктора по его речи / Р.Л. Григорян, С.С. Коршунов, С.А. Репалов, М.Ю. Хрящев // Речевые технологии. – 2010. – № 3. – С. 35–38.

105. Голубинский, А.Н. Метод оценки формантных частот, основанный на полигармонической математической модели речевого сигнала / А.Н. Голубинский, О.М. Булгаков // Речевые технологии. – 2010. – № 3. – С. 29–34.

106. Конев, А.А. Об одном алгоритме оценки формантных частот на интервале сомкнутых голосовых складок / А.А. Конев, Р.В. Мещеряков, С.В. Жевуров, В.С. Хлебников // Речевые технологии. – 2010. – № 3. – С. 50–53.
107. Голубинский, А.Н. Модель речевого сигнала в виде импульса АМ-колебания с несколькими несущими для верификации личности по голосу / А.Н. Голубинский // Системы управления и информационные технологии. – 2007. – № 4. – С. 86–91.
108. Ролдугин, С.В. Модели речевых сигналов для идентификации личности по голосу / С.В. Ролдугин, А.Н. Голубинский, Т.А. Вольская // Радиотехника. – 2002. – № 11. – С. 79–81.
109. Азаров, И.С. Вычисление мгновенных гармонических параметров речевого сигнала / И.С. Азаров, А.А. Петровский // Речевые технологии. – 2008. – № 1. – С. 67–77.
110. Kekre, Dr. H. B. Speaker identification by using power distribution in frequency domain / Dr. H. B. Kekre, Vaishali Kulkarni // <<Thinkquest~2010>> Proceedings of the First International Conference on Contours of Computing Technology 13th and 14th March, 2010. – Mumbai: Springer, 2011. – pp. 279–283.
111. Md. Rahman, S. Speaker Identification using Mel Frequency Cepstral Coefficients / Md. R. Hasan, M. Jamil, Md. G. Rabbani, Md. S. Rahman // 3rd International Conference on Electrical & Computer Engineering: Book of abstracts. – Dhaka, 2004. – pp. 565–568.
112. Reynolds, D.A. Experimental evaluation of features for robust speaker identification / D.A. Reynolds // Speech and Audio Processing. – 1994. – vol. 2 (4). – pp. 639–643.
113. Beigi, Homayoon. Fundamentals of Speaker Recognition / H. Beigi. – Springer, – 2011. – 942 p.
114. Хайдер А.С. Сравнение методов параметризации речевых сигналов при решении задачи распознавания дикторов. // Известия СПбГЭТУ "ЛЭТИ",

серия "Информатика, управление и компьютерные технологии" вып.1, 2004, с.36-42

115. Заковряшин, А.С. Применение распределений мел-частотных кепстральных коэффициентов для голосовой идентификации личности / А.С. Заковряшин, П.В. Малинин, А.А. Лепендин // Известия АлтГУ. – 2014. – №1/1. – С. 156–160.

116. Малинин, П.В. Применение методов анализа многомерных данных в задаче идентификации личности / П.В. Малинин, В.В. Поляков // Проблемы информационной безопасности государства, общества и личности, безопасность нанотехнологий: Материалы 11-й Всероссийской научно-практической конференции. г. Томск, 3–5 июня 2009 г. – Томск : В-Спектр, 2009. – С. 29–33.

117. Малинин, П.В. Классификация голосовых сигналов для разграничения доступа к конфиденциальной информации. Свидетельство о государственной регистрации программы для ЭВМ № 2011611627, 17 февраля 2011 г.

118. Малинин, П.В. Применение методов многомерного анализа данных в системе разграничения доступа по голосу / П.В. Малинин // Проблемы правовой и технической защиты информации: сб. ст. / под ред. В.В. Полякова, В.А. Мазурова. – Барнаул: Изд-во Алт. ун-та, 2008. – С. 124–127.

119. Малинин, П.В. Идентификация личности по голосу на основе методов анализа многомерных данных / П.В. Малинин // Проблемы информационной безопасности государства, общества и личности : матер. IV Пленума СибРОУМО по образованию в области информационной безопасности и XII Всерос. научно-практ. конф. Томск – Барнаул – Белокуриха, 8–13 июня 2010 г. – Томск : Изд-во «В-Спектр», 2010. – С. 180–181.

120. Малинин, П.В. Метод ПЛС в задаче идентификации личности по голосу / П.В. Малинин // Наука и инновации XXI: мат-лы х конф. молодых ученых г.

Сургут, 26–27 ноября 2009 г. – Югры–Сургут : ИЦ СурГУ, 2010. – Т.1. – С. 34–36.

121. Журавлев, Ю.И. Распознавание. Математические методы. Программная система. Практические применения / Ю.И. Журавлев, В.В. Рязанов, О.В. Сенько. – М. : ФАЗИС, 2006. – 176 с.

122. Мандель, И.Д. Кластерный анализ / И.Д. Мандель. – М. : Финансы и статистика, – 1988. – 176 с.

123. Малинин, П.В. Иерархический подход в задаче идентификации личности по голосу с помощью проекционных методов анализа многомерных данных / П.В. Малинин, В.В. Поляков // Доклады ТУСУР. – 2010. – часть 1 (21). – С. 128-130.

124. Рабинер, Л.Р. Цифровая обработка речевых сигналов / Л.Р. Рабинер, Р.В. Шафер; под ред. М.В. Назарова, Ю.Н. Прохорова. – М. : Радио и связь, 1981. – 224 с.

125. Малинин, П.В. Математическое моделирование искажений спектра акустического сигнала при голосовой идентификации личности / П.В. Малинин, В.В. Поляков // Известия АлтГУ. – 2012. – № 1/2. – С. 87–89.

126 П.В. Малинин. Региональные аспекты технической и правовой защиты инфомации : монография / В.В. Поляков, В.А. Трушин, И.А. Рева, Вит.В. Поляков, П.В. Малинин и др. – Барнаул : Изд-во Алт. ун-та, 2013. – 176 с.

127. Малинин, П.В. Влияние искажений голосового сигнала на идентификацию диктора / П.В. Малинин, В.В. Поляков // Известия АлтГУ. – 2010. – № 1/2. – С. 172–174.

128. Малинин, П.В. Исследование влияния шума на идентификацию личности по голосу / П.В. Малинин, В.В. Поляков // Проблемы информационной безопасности государства, общества и личности : доклады V Пленума СибРОУМО по образованию в области информационной безопасности и XIII

Всерос. научно-практ. конф. Томск – Новосибирск, 5–9 июня 2012 г. – Томск : Изд-во «В-Спектр», 2012. – С. 96–98.

129. Noise Robust Speaker Identification for Spontaneous Arabic Speech / M. Graciarena, S. Kajarekar, A. Stolcke, E. Shriberg // ICASSP 2007. IEEE International Conference – 2007. – pp. 245–248.

130. Малинин, П.В. Влияние вида акустического шума на голосовую идентификацию личности / П.В. Малинин, В.В. Поляков // Известия АлтГУ. – 2013. – № 1/2. – С. 168–169.

131. Малинин, П.В. Влияние вида шума на голосовую идентификацию личности / П.В. Малинин, В.В. Поляков // Проблемы информационной безопасности государства, общества и личности : доклады V Пленума СибРОУМО по образованию в области информационной безопасности и XIII Всерос. научно-практ. конф. Томск – Новосибирск, 5-9 июня 2012 г. – Томск : Изд-во «В-Спектр», 2012. – С. 98–101.

132. Речевой корпус данных [Электронный ресурс] / Вычислительный центр им. А.А. Дородницына Российской академии наук. – М., 2000. – Режим доступа : http://www.ccas.ru/depart/chuchu/doc_ru/projects/tecorus_info_ru.pdf. – Заглавие с экрана.