Санкт-Петербургский национальный исследовательский университет
информационных технологий, механики и оптики

На правах рукописи

Щемелинин Вадим Леонидович

МЕТОДИКА И КОМПЛЕКС СРЕДСТВ ОЦЕНКИ ЭФФЕКТИВНОСТИ АУТЕНТИФИКАЦИИ ГОЛОСОВЫМИ БИОМЕТРИЧЕСКИМИ СИСТЕМАМИ

Специальность 05.13.19 - Методы и системы защиты информации, информационная безопасность

Диссертация на соискание ученой степени
Кандидата технических наук

Научный руководитель
кандидат технических наук, доцент
Симончик Константин Константинович

Санкт-Петербург – 2015

СОДЕРЖАНИЕ

Перечень сокращений

ГБС	Голосовая биометрическая система
ВЛД FAR	False acceptance rate, вероятность ложного допуска злоумышленника
ВЛНД FRR	False rejection rate, вероятность ложного недопуска клиента
ВЛОИ FNIR	Вероятность ложно отрицательной идентификации, false-negative identification-error rate
ВЛПИ FPIR	Вероятность ложно положительной идентификации, false-positive identification-error rate
ВЛНС FNMR	Вероятность ложного несовпадения, false non-match rate
ВЛС FMR	Вероятность ложного совпадения, false match rate
ВОР FTE	Вероятность отказа регистрации, failure-to-enroll rate
ВОСД FTA	Вероятность отказа сбора данных, failure-to-acquire rate
КОО DET	Detection Error Trade-off, кривая компромисса определения ошибки
РВО EER (equalerror rate)	Уровень равной ошибки или точка совпадения вероятностей ошибок первого (пропуск цели) и второго рода

Введение

Актуальность темы исследования

Развитие компьютерных технологий в последние десятилетия дало возможность совершить прорыв в области обработки речевого сигнала. Современный мир уже сложно представить без повседневного использования речевых технологий. Системы распознавания речи позволяют не отвлекаться на управление мобильными устройствами во время движения за рулём, системы синтеза речи оповещают нас по телефону, в метро, на вокзалах и в офисах, голосовые биометрические системы обеспечивают решение задачи аутентификации при доступе к защищённым персональным данным или поиске нарушителей.

Исследования голосовых биометрических технологий занимают одно из ведущих мест в области обработки речевого сигнала. В первую очередь, следует отметить основополагающие работы авторов Douglas A. Reynolds, Patrick J. Kenny. Регулярные оценки эффективности аутентификации голосовыми биометрическими системами проводятся в виде конкурса Национальным Институтом Стандартов и Технологий США (NIST). Целью конкурса является определение доминирующих направлений в данной технологии. Однако, возникающие при обработке речевого сигнала задачи, в виду их комплексного характера и сложности, далеки от того, чтобы их можно было считать решенными как в практическом, так и в научном плане.

В последнее время все большее количество потребителей биометрических систем озабочено не только качеством непосредственно голосовой биометрии, но и противодействием различным видам атак, проводимых с целью получения доступа к защищённой информации (спуфинга).

Большую работу в направлении исследования спуфинг атак на голосовые биометрические системы провела группа исследователей под

руководством Tomi Kinnunen в Университете Восточной Финляндии. В 2015 году ими был организован первый в мире международный конкурс Automatic Speaker Verification Spoofing and Countermeasures (ASVspoof) Challenge 2015 при крупнейшей конференции по речевым технологиям Interspeech. Результаты их исследований показали не только уязвимость голосовых биометрических систем к простейшим атакам на основе записи речи на диктофон, но и к более сложным способам синтеза голоса, а также к преобразованию голоса злоумышленника к заданному голосу пользователя системы.

Таким образом, методы противодействия спуфинг атакам, позволяющие повысить степень защиты голосовых биометрических систем, являются на сегодняшний день крайне актуальными. Оценка эффективности аутентификации, используемой современными голосовыми системами безопасности, должна включать не только требования к надежности базовой технологии идентификации диктора, но и к защищенности такого рода систем от несанкционированных попыток доступа к ним.

Целью исследования является повышение эффективности аутентификации голосовыми биометрическими системами в условиях возможных спуфинг атак.

Для достижения поставленной цели были сформулированы и решены следующие основные **задачи:**

1. Анализ уязвимости современных голосовых биометрических систем к различным способам фальсификации индивидуальных голосовых биометрических характеристик человека.
2. Разработка методики оценки эффективности аутентификации голосовыми биометрическими системами с учётом возможного влияния различных видов спуфинг атак на модуль ввода биометрической информации.

3. Разработка комплекса программных средств оценки эффективности аутентификации голосовыми биометрическими системами.
4. Разработка метода противодействия спуфинг атакам, позволяющего повысить устойчивость голосовых биометрических систем к спуфинг атакам различного вида на модуль ввода биометрической информации.

Объект исследования. Голосовые биометрические системы и способы фальсификации индивидуальных голосовых биометрических характеристик человека.

Предмет исследования. Методика и комплекс средств оценки эффективности аутентификации голосовыми биометрическими системами, оценка защищённости и выбор средств защиты персональных данных, обрабатываемых в голосовых биометрических системах.

Методы исследования. В работе использованы методы теории вероятности и математической статистики, цифровой обработки сигналов, методы проектирования и разработки программного обеспечения ЭВМ.

Научная новизна диссертационного исследования заключается в следующем:

1. Предложенный метод имитации атак на голосовые биометрические системы отличается применением автоматической разметки речевых данных для создания модели синтезированного голоса целевого диктора.
2. Предложенная методика оценки эффективности аутентификации голосовыми биометрическими системами отличается учётом воздействия различных видов спуфинг атак на модуль ввода биометрической информации.

3. Реализованный комплекс программных средств оценки эффективности аутентификации голосовыми биометрическими системами отличается наличием модуля имитации атаки и модуля расчёта показателей эффективности аутентификации с учётом воздействия спуфинг атак.
4. Разработанный метод противодействия спуфинг атакам отличается комбинированием методов факторного анализа, сигнальной обработки и признакового описания сигнала.

Основные положения, выносимые на защиту.

1. Метод имитации атак на голосовые биометрические системы, обеспечивающий автоматическое создание модели голоса для синтеза голосовых биометрических характеристик.
2. Методика оценки эффективности аутентификации голосовыми биометрическими системами, обеспечивающая учёт воздействия различных видов спуфинг атак на модуль ввода биометрической информации.
3. Комплекс программных средств оценки эффективности аутентификации голосовыми биометрическими системами, позволяющий оценивать устойчивость к различным видам атак при проведении технологических испытаний.
4. Метод противодействия спуфинг атакам, позволяющий значительно повысить устойчивость голосовых биометрических систем к различным методам спуфинг атак на модуль ввода биометрической информации.

Обоснованность научных положений, выводов и практических рекомендаций, полученных в диссертационной работе, обеспечивается результатами экспериментальных исследований, успешным представлением

основных положений в ряде докладов на ведущих международных конференциях, в том числе, на международном конкурсе Automatic Speaker Verification Spoofing and Countermeasures (ASVspoof) Challenge 2015, а также результатами технологических испытаний реальных систем, при оценке которых были использованы предложенные методы, методика и комплекс программных средств. Практические рекомендации, сформулированные в диссертации, обоснованы проведенными исследованиями и могут служить руководством при решении практических задач.

Практическая значимость работы заключается в реализации предложенной методики в виде комплекса программных средств оценки эффективности аутентификации голосовыми биометрическими системами. Разработанные технические решения по совершенствованию защиты заняли второе место на международном конкурсе ASVspoof Challenge 2015 и могут быть встроены в коммерческие голосовые биометрические системы.

Внедрение результатов работы. Результаты диссертации использованы при выполнении следующих научно-исследовательских и опытно-конструкторских работ: НИР Министерства образования и науки «Исследование методов и алгоритмов многомодальных биометрических и речевых систем» (грант 074-U01); ОКР Федеральной службы безопасности, шифр «Ярмарка-ТМС»; ОКР Министерства внутренних дел, шифр «Кристалл-М (Флот)»; ОКР Федеральной службы по контролю за оборотом наркотиков, шифр «Этнос». Также результаты работы были внедрены в различные коммерческие продукты компаний ООО “ЦРТ”.

Апробация результатов исследования. Результаты, полученные в рамках работы над диссертацией, представлялись и обсуждались на следующих научно-методических конференциях: «15th International Conference on Speech and Computer SPECOM 2013» (Пльзень, Чехия, 2013), «XLIII научная и учебно-методическая конференция НИУ ИТМО» (Санкт-

Петербург, 2014), «III Всероссийский конгресс молодых ученых» (Санкт-Петербург, 2014) - диплом за лучший доклад на секции, «4th International Workshop on Spoken Language Technologies for Under-resourced Languages (SLTU'14)» (Санкт-Петербург, 2014), «16th International Conference on Speech and Computer SPECOM 2014» (Новый Сад, Сербия), «XVI Международная конференция по вопросам качества программного обеспечения SQA Days 16» (Санкт-Петербург, 2014), «XLIII научная и учебно-методическая конференция НИУ ИТМО» (Санкт-Петербург, 2014), «XLIV научная и учебно-методическая конференция НИУ ИТМО» (Санкт-Петербург, 2015), а также были представлены в виде системы детектирования атак, занявшей 2-ое место на международном конкурсе Automatic Speaker Verification Spoofing and Countermeasures (ASVspoof) Challenge 2015.

Личный вклад автора. Автором лично проведён анализ уязвимости современных голосовых биометрических систем к различным способам фальсификации индивидуальных голосовых биометрических характеристик человека. На основе проведённого анализа разработана методика оценки эффективности аутентификации голосовыми биометрическими системами с учётом возможного влияния различных видов атак на модуль ввода биометрической информации. Проведены исследования, демонстрирующие преимущества предложенной методики в сравнении с существующими аналогами. Разработан комплекс программных средств, позволяющий применить предложенную методику на практике. С учётом проведённых исследований разработан метод противодействия спуфинг атакам, позволяющий повысить устойчивость голосовых биометрических систем к различным методам спуфинг атак на модуль ввода биометрической информации. Подготовка основных публикаций проводилась с соавторами, при этом вклад автора был основным.

Публикации. По теме диссертации опубликовано десять печатных работ, шесть из которых - в изданиях из перечня рецензируемых научных журналов ВАК, в том числе, пять - в международных изданиях, индексируемых в базе данных Scopus.

Объем и структура диссертации

Диссертационная работа состоит из введения, четырёх глав, заключения и списка литературы. Материал изложен на 139 страницах, включает 8 таблиц, 29 рисунков и схем, а также одно приложение. Список использованной информации содержит 101 наименование.

ГЛАВА 1. Обзор современных методик и стандартов в области оценки эффективности аутентификации голосовыми биометрическими системами

1.1. Современные голосовые биометрические системы

Голосовые биометрические системы являются подмножеством систем, опирающихся на уникальность индивидуальных биометрических характеристик человека. Таким образом, области их применения пересекаются. Отличительной особенностью систем распознавания по голосовым биометрическим признакам является практически полное отсутствие специальных требований к оборудованию, используемому для получения биометрических данных. В большинстве случаев, в голосовых биометрических системах могут быть использованы стандартные микрофоны, используемые в мобильных и стационарных телефонах, гарнитурах для персональных компьютеров, ноутбуках или планшетах.

Благодаря возможности использования голосовых биометрических систем со стандартными устройствами ввода-вывода, они идеально подходят для решения задач биометрических систем, таких, как:

- Автоматический поиск разыскиваемого человека в открытых каналах связи.
- Обработка больших объёмов речевых баз данных.
- Предоставление доступа к информационным ресурсам по средствам мобильной аутентификации пользователя.

Помимо этого, голосовые биометрические системы могут быть использованы при проведении криминалистических экспертиз в качестве модулей для соответствующих аппаратно программных комплексов.

В международном стандарте ISO/IEC 2382-37:2012 Information technology — Vocabulary — Part 37: Biometrics дано следующее определение биометрическим системам. Биометрическая система это система, предназначенная для автоматического распознавания индивида (личности человека), основанного на его поведенческих и биологических характеристиках. Таким образом, голосовая биометрическая система это система, предназначенная для автоматического распознавания личности человека, основанного на его поведенческих и биологических характеристиках, содержащихся в голосе. Задача автоматического распознавания индивида, включает в себя:

1. Биометрическую верификацию - процесс подтверждения заявления, о том, что субъект сбора биометрических данных является или не является собственно источником установленного или неустановленного биометрического контрольного шаблона. при биометрическом сравнении.
2. Биометрическую идентификацию - процесс поиска по базе данных биометрической регистрации, направленный на поиск и возврат идентификатора (идентификаторов) биометрического контрольного шаблона, связанного с одним индивидом.

В том же стандарте зафиксировано определение аутентификации как действия, доказывающего или показывающего бесспорное происхождение или достоверность. А также дано указание на то, что данный термин используется в биометрии в качестве синонима для приложений и функций биометрической идентификации и биометрической верификации.

В соответствии с разделением задачи, голосовые биометрические системы делятся на системы верификации и системы идентификации. Обобщённая голосовая биометрическая система включает в себя два

основных функциональных процесса: процесс регистрации субъекта в системе и процесс верификации или идентификации субъекта.

При регистрации субъекта его данные обрабатывается системой для создания и сохранения регистрационного шаблона данного субъекта. Процесс регистрации состоит из следующих этапов:

- получение речевого образца;
- сегментация и выделение речевых признаков;
- проверка качества, в результате которой образец или признаки, непригодные для создания шаблона, могут быть отклонены, и будет сформирован запрос на получение дополнительных образцов;
- создание шаблона (может потребовать признаки нескольких образцов) с возможным преобразованием его в формат обмена биометрическими данными и хранения;
- попытки верификации или идентификации, чтобы гарантировать пригодность регистрации;
- попытки повторной регистрации, которые могут быть предоставлены, если первоначальная регистрация оказалась неудовлетворительной.

Верификация – это процедура подтверждения личности говорящего. В качестве результата система верификации диктора по голосу возвращает степень совпадения голоса диктора с шаблоном.

Процесс верификации включает в себя следующие этапы:

- получение речевого образца;
- сегментация и выделение речевых признаков;

– проверка качества, в результате которой образец или признаки, непригодные для создания шаблона, могут быть отклонены, и будет сформирован запрос на получение дополнительных образцов;

– сравнение признаков образца с признаками, извлеченными из шаблона, для определения степени схожести;

– формирование решения о соответствии признаков образца признакам, извлеченным из шаблона, которое принимают, если степень схожести образца превышает порог принятия решений;

– возвращение результата верификации, основанного на результате сравнения одной или более попыток в соответствии с политикой принятия решений.

Идентификация представляет собой определение личности из заданного, ограниченного и открытого списка людей. Результат идентификации предоставляет результат поиска текущего диктора среди списка кандидатов, наиболее близких к оцениваемому образцу голоса.

Процесс идентификации состоит из следующих этапов:

– получение образца;

– сегментация и выделение признаков;

– проверка качества (которая может отклонить образец или признаки, непригодные для сравнения и потребовать получения дополнительных образцов);

– сравнение с некоторыми или со всеми шаблонами базы данных, определяющее степень схожести для каждого сравнения;

– формирование решения об идентичности шаблонов, которое принимается, если степень схожести превышает порог принятия

решений и (или) находится среди первых значений k степеней схожести;

- возвращение результата идентификации одной или более попыток в соответствии с политикой принятия решений.

1.1.1. Обобщённая структура голосовой биометрической системы

Обобщённая структура голосовой биометрической системы, включает такие компоненты, как:

- устройство ввода;
- подсистема обработки речевых данных;
- подсистема хранения шаблонов;
- подсистема сравнения и принятия решения;
- интерфейс приложения; подсистема передачи данных.

Основные компоненты системы показаны на рисунке 1.1.

Отметим, что многие голосовые биометрические системы, основаны на одних и тех же биометрических комплектах средств разработки (Software Development Kit, SDK), являющихся ядром таких компонентов, как подсистема обработки данных, и подсистема сравнения и принятия решения.

Описанию теоретических основ обработки сигнала и методам идентификации диктора посвящено множество работ известных авторов. Сред них работы таких авторов, как Reynolds [71], Kenny [57], Bimbot [37], Dehak [44], Матвеев [18], Аграновский [1], Раев [28], Пеховский [20] и Симончик [32].

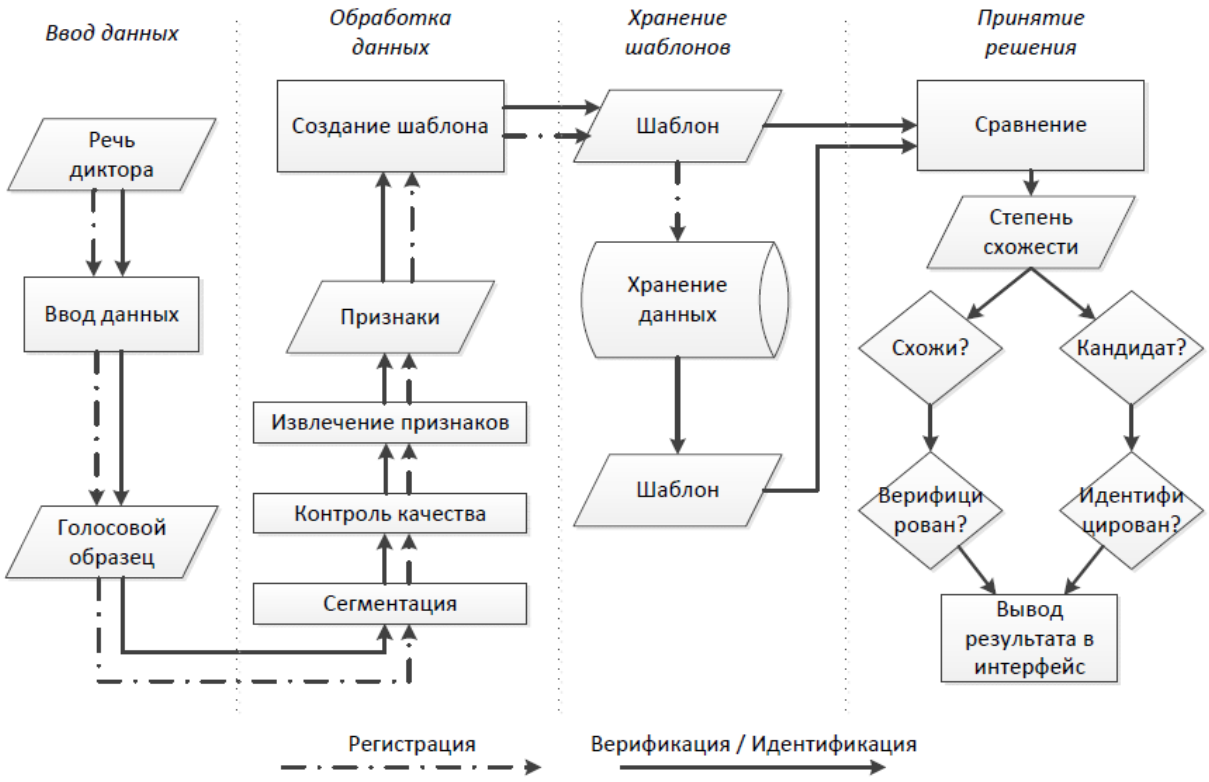

Рисунок 1.1 - Компоненты обобщённой голосовой биометрической системы

Рассмотрим современные автоматические методы распознавания диктора, лежащие в основе голосовых биометрических систем.

1.1.2. Метод сравнения статистик основного тона

Основной тон является одним из базовых параметров речевого сигнала, и при этом не сильно зависит от условий записи и типа канала. В связи с этим, метод распознавания дикторов, основанный на параметрической статистике основного тона (ОТ) является одним из базовых среди автоматических методов распознавания диктора.

Преимущество спектрального метода, используемого для вычисления частоты основного тона, заключается в том, что он позволяет оценить частоту основного тона, используя всю доступную частотную полосу сигнала.

Амплитудно-частотная характеристика (АЧХ) канала записи фонограммы всегда оказывает значительное влияние на форму спектра.

Данное влияние необходимо учитывать, т.к. в противном случае АЧХ канала, с одной стороны, может замаскировать индивидуальные параметры голоса диктора, а с другой стороны подавить часть спектра речевого сигнала и сделать его недоступным для дальнейшего биометрического распознавания.

К ключевым факторам, влияющим на эффективность аутентификации методом на основе анализа статистики основного тона, следует отнести следующие:

1. Длина образцовой и спорной фонограммы диктора. Эффективность системы заметно выше, если используется запись длительностью речи, достаточной для получения достоверных статистик основного тона.
2. Наличие образцовой и спорной записей, полученных при одном и том же эмоциональном состоянии диктора.
3. Наличие образцовой и спорной записей, полученных при одном и том же фоновом окружении (в случае, когда влияние окружения настолько велико, что меняется стиль речи диктора).
4. Отношение сигнал/шум.
5. Отсутствие реверберации на записях.

Идентификация дикторов методом сравнения голосов на основании параметрической статистики основного тона голоса включает в себя три этапа:

1. Выделение основного тона голоса при помощи универсального выделителя статистик основного тона для произвольных дикторов, каналов записи, языков, эмоциональных состояний и длительностей фонограмм.
2. Вычисление параметрических статистических характеристик ОТ.
3. Сравнение параметрических статистических характеристик ОТ.

Более подробное описание данного метода можно найти в работе [15].

1.1.3. Метод спектрально формантного анализа

Метод идентификации дикторов на основе спектрально-формантных признаков осуществляет сравнение спорных образцов естественной речи с аналогичными образцами из речевой базы данных эталонов путем анализа положения формант. Основными этапами работы данного метода являются:

1. Предобработка входного звукового файла, включающая: удаление пауз; нормализацию на канал; построение СГР моделей дикторов с использованием формантных векторов в качестве входных признаков; построение SVM модели дикторов.
2. Сравнение моделей дикторов. Алгоритм метода идентификации дикторов на основе сравнения спектрально-формантных представлений состоит из следующих блоков: нормализация на канал; вычисление идентификационных моделей дикторов; вычисление индивидуальных порогов принятия решения «свой/чужой» путем сравнения полученной идентификационной модели со стандартным набором эталонных моделей заведомо «чужих» дикторов; непосредственное сравнение полученной идентификационной модели с моделью из базы данных и принятие решения «свой/чужой» в соответствии с индивидуальными порогами и заданными вероятностями ложной тревоги и пропуска цели.

Детальное описание работы данного метода можно изучить в работах [16, 58].

1.1.4. Метод анализа смеси гауссовских распределений

Смеси гауссовых распределений, на сегодняшний день, являются одним из основных подходов при решении задачи голосовой биометрической аутентификации.

Модель гауссовой смеси голоса диктора обеспечивает вероятностную

модель основных звуков, содержащихся в речи диктора. Для представления большого объёма экспериментальных распределений, в качестве базиса, используется линейная комбинация гауссовых функций.

Основным достоинством метода СГР является возможность формирования гладких аппроксимаций экспериментальных распределений компонент акустического пространства, форма которых может иметь произвольную форму.

Основной сложностью при решении задачи биометрической аутентификации для систем, основанных на методе СГР, является нивелирование влияния рассогласования, вносимого помехами, содержащимися в канале, используемом при записи.

Причинами этого рассогласования могут быть:

- шумы окружающей среды при записи;
- искажения в каналах записи и передачи речевого сигнала;
- изменчивость голоса диктора с течением времени.

Под каналом, в данном случае, понимается следующая совокупность эффектов:

- искажения, вносимые записывающей аппаратурой;
- влияние микрофона устройства, используемого для извлечения индивидуальных голосовых биометрических характеристик диктора;
- влияние АЧХ канала соединения.

Более подробное описание данного метода можно найти в работах таких исследователей, как Белых И.Н. [3]. Матвеев Ю.Н. [19], Симончик К.К. [14].

1.1.5. Метод совместного факторного анализа

Для решения проблемы согласования, было найдено решение, ставшее традиционным к сегодняшнему дню. Данное решение заключается в применении совместного факторного анализа (Joint Factor Analysis, JFA), который позволяет в отдельном произнесении диктора эффективно отделять канальную информацию от дикторской информации. Это позволяет строить каналонезависимые GMM-модели речи диктора и подавлять эффекты канала в звуковых данных, по которым происходит построение модели диктора.

В качестве обязательного элемента, в JFA используется универсальная фоновая модель (universal background model, UBM). Данная модель строится для выделения общих «чужих» дикторов во всех возможных контекстах. Обучающая база для построения UBM формируется с учетом максимально большого объема речевых данных, сбалансированных по гендерному типу, каналам записи, акустическим условиям и т.д. Типовым подходом к построению UBM-модели является использование метода оценки максимального правдоподобия (Maximum Likelihood, ML) – ML-метод [69].

Основным недостатком JFA метода является большой размер шаблонов диктора. Кроме этого для построения полной JFA-модели требуется большой объем обучающей базы данных.

Более подробное описание JFA метода доступно в работе [56].

1.1.6. Метод матрицы полной изменчивости

Одним из вариантов решения проблему большого размера моделей диктора является применение низкоразмерных векторов признаков. Так, в модифицированной версии JFA для генерации векторов признаков используется матрица полной изменчивости (Total Variability, TV). Данная модифицированная версия JFA часто называется TV-методом автоматического распознавания дикторов. Данный метод получил широкое

распространение как наиболее перспективный метод распознавания дикторов, обеспечивающий приведение высокоразмерных входных данных к низкоразмерному вектору признаков, с обеспечением сохранения большей части полезной информации. Это позволяет снизить объем модели диктора до 2-3 кбайт, что чаще всего является приемлемым при решении задачи идентификации диктора на большом множестве или построении системы, требующей передачи моделей диктора по медленным каналам связи.

Подробное описание данного метода можно найти в работе [44].

1.1.7. Метод вероятностного линейного дискриминантного анализа

Вторым вариантом решения проблемы рассогласования, является модификация JFA метода, содержащая дополнительную операцию на основе вероятностного линейного дискриминантного анализа (Probabilistic Linear Discriminative Analysis, PLDA) [62]. Это позволяет более эффективно нивелировать эффект канальных искажений при решении задачи голосового распознавания диктора. Отличительными особенностями данного метода являются:

1. Представление каждого произнесения в виде низкоразмерного вектора в пространстве с базисом, представленным матрицей полной изменчивости.
2. Априорные распределения факторов вариативности тестовых и обучающих произнесений и описываются распределением Стьюдента с так называемыми «тяжелыми хвостами» (heavy tailed priors), что позволяет получить устойчивые оценки параметров модели.
3. На этапе сравнения, верификационная оценка производится, основываясь на симметричной, относительно сравниваемых произнесений, оценке PLDA.

Системы на основе данного метода, на данный момент занимают лидирующие позиции на международных соревнованиях NIST [82, 83] среди голосовых биометрических систем [17]. Более подробно про данный метод можно изучить в работе [77].

1.2. Возможные атаки на устройство ввода биометрической информации

Несмотря на то, что точность распознавания голосовых биометрических систем значительно выросла в последние несколько лет, на практике лишь небольшое количество людей доверяют системам безопасности, основанным на голосовой биометрии. Самый распространённый аргумент против использования таких систем состоит в том, что злоумышленник может легко обойти биометрическую систему контроля доступа, используя простые техники имитации, выдавая себя за другого диктора. Подтверждение этому можно найти в работах [35, 88, 90].

Рассмотрим виды имитационных техник, направленных на взлом голосовой биометрической системы.

1.2.1. Методы атак на основе имперсонализации

Имперсонализацией называется процесс выдачи себя за другого человека. Применительно к голосовым биометрическим системам, такой вид атаки представляет собой наиболее простой и очевидной метод, основанный на имитации и изменении голоса путём подражания, без использования компьютерных технологий или специальных устройств [59]. Однако согласно исследованиям, проведённым в работах [61, 101] такой вид атаки не несет большой опасности, так как даже профессионалы пародии не в состоянии стабильно взламывать голосовые биометрические системы, основанные на устаревающих методах распознавания диктора и не содержащие каких либо решений по противодействию подобным атакам.

1.2.2. Методы атак на основе записи и повтора БХЧ

Самым очевидным способом фальсификации биометрических характеристик является их запись и дальнейший повтор. В случае голосовых биометрических систем, злоумышленнику необходимо записать речь пользователя голосовой биометрической системы и произвести попытку входа с использованием полученной записи.

Вероятность успеха подобной атаки очень сильно зависит от конструктивных особенностей голосовой биометрической системы. Например, при использовании текстозависимой системы верификации с динамической парольной фразой, схема работы которой изображена на рисунке 1.2, данный метод атаки применить невозможно.

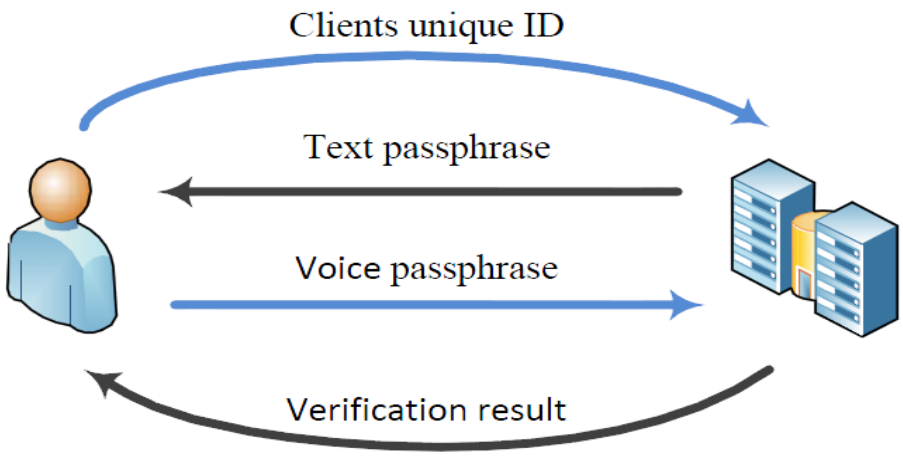

Рисунок 1.2 - Схема работы текстозависимой голосовой биометрической системы с применением динамической парольной фразы

Другим гарантированным способом защиты от «атаки повтором» является использование мультимодальных биометрических систем.

Некоторые методы противодействия подобным атакам представлены в работах [34, 86, 91].

1.2.3. Методы атак на основе преобразования БХЧ

Более серьёзную угрозу для голосовых биометрических систем представляет техника голосовой конверсии или преобразования речи злоумышленника, в речь пользователя голосовой биометрической системы. Техники голосовой конверсии модифицируют одно голосовое высказывание (источник), и звучат так, как будто произносятся другим диктором (целевым). Системы голосового преобразования состоят из этапа обучения и этапа преобразования. В обоих этапах речевой сигнал изначально разбивается на векторы признаков малой размерности. В обучающей фазе функции источника и целевого диктора вначале соединяются на уровне сегментов, обычно для этого используются параллельные обучающие высказывания. Функция преобразования, в соответствии со стохастической моделью Гауссовых смесей (GMM), далее испытывается с использованием парных векторов [81]. На этапе конверсии функция преобразования используется для отображения невидимых источников признаков, относящихся к целевому диктору. Конвертированное высказывание переделывается с использованием обратной параметризации.

Рассмотрим один из методов преобразования речи основанный, на моделях Гауссовых смесей [81].

Рассмотрим выровненную по сегментам последовательность обучаемых векторов из исходных (x) и целевых (y) дикторов:

$$x = \left[x_1^{\mathrm{T}}, x_2^{\mathrm{T}}, \dots, x_t^{\mathrm{T}}, \dots, x_T^{\mathrm{T}}\right]^{\mathrm{T}}$$

$$y = \left[y_1^{\mathrm{T}}, y_2^{\mathrm{T}}, \dots, y_t^{\mathrm{T}}, \dots, y_T^{\mathrm{T}}\right]^{\mathrm{T}}$$

где T означает транспонирование вектора.

Векторы укладываются по уровню сегментов в объединённые векторы $z_t = [x_t^{\mathrm{T}}, y_t^{\mathrm{T}}]^{\mathrm{T}}$. Объединённая плотность вероятности векторов исходных и целевых признаков моделируется при помощи GMM,

$$P(z_t|\lambda^{(z)}) = \sum_{m=1}^{M} \varpi_m^{(z)} N(z_t|\mu_m^{(z)}, \Sigma_m^{(z)})$$

где $\mu_m^{(z)} = \begin{bmatrix} \mu_m^{(x)} \\ \mu_m^{(y)} \end{bmatrix}$ и $\Sigma_m^{(z)} = \begin{bmatrix} \Sigma_m^{(xx)} & \Sigma_m^{(xy)} \\ \Sigma_m^{(yx)} & \Sigma_m^{(yy)} \end{bmatrix}$ это вектор средних и ковариационная матрица мультивариативной Гауссовой плотности $N(z_t|\mu_m^{(z)}, \Sigma_m^{(z)})$, соответственно.

Априорные вероятности $\varpi_m^{(z)}$ суммируются к единице. Параметры GMM $\lambda^{(z)} = \{\varpi_m^{(z)}, \mu_m^{(z)}, \Sigma_m^{(z)} | m = 1,2, \dots, M\}$ оцениваются в максимальной похожести смысла использования, хорошо известным алгоритмом максимального мат. ожидания. Здесь, используем $M = 8$ Гауссоиду с полноковариационными матрицами. В фазе конверсии, данный новый вектор исходного диктора (x), и обученная объединённая модель плотности используются для предсказания вектора целевого диктора $\hat{y}$ как,

$$\hat{y} = F(x) = E(y|x) = \sum_{m=1}^{M} p_m(x)(\mu_m^{(y)} + \Sigma_m^{(yx)} \left(\Sigma_m^{(xx)}\right)^{-1} (x - \mu_m^{(x)}))$$

где $p_m(x) = \frac{\varpi_m N(x|\mu_m^{(x)}, \Sigma_m^{(xx)})}{\Sigma_{k=1}^{K} \varpi_k N(\mu_k^x, \Sigma_k^{(xx)})}$ апостериорная вероятность возникновения исходного вектора x из m Гауссоиды.

Используем описанную выше процедуру для конвертации спектральных параметров. Для основной частоты $(F0)$, конверсия происходит выравниванием средних и дисперсий исходного и целевого $logF0$ распределения.

Согласно требованию структуры стохастической конвертации, обученные вектора необходимо выровнять. Как правило, для обучения используют параллельный набор произнесений из исходного и целевого диктора. Это означает, что один и тот же текст зачитывается обоими

дикторами. Эти обучающие произнесения затем должны быть выровнены по времени при помощи, например, метода динамического временного выравнивания (DTW). В случае если речевая база состоит из диалогов по телефону, без параллельных высказываний (что является более вероятным при подготовке атаки), можно воспользоваться методами не параллельного выравнивания [93].

Основным недостатком данного вида атаки является необходимость непосредственного участия злоумышленника при попытке взлома, что делает невозможной полную автоматизацию данного вида атаки и, как следствие, снижает угрозу для голосовых биометрических систем.

Другие методы атак на основе технологии преобразования, а также методы противодействия им, можно найти в работах [33, 36, 46, 90, 95].

1.2.4. Методы атак на основе синтеза БХЧ

Данный метод взлома заключается в создании синтезированного голоса пользователя системы верификации. Для обучения системы синтеза используется предварительно записанная спонтанная речь пользователя. На этапе текстозависимой верификации, при помощи полученного синтезированного голоса и перехваченного парольного текста, создаётся синтезированная парольная фраза, используемая далее при попытке аутентификации.

Рассмотрим схему применения системы синтеза речи на основе гибрида двух наиболее популярных подходов, изображённую на рисунке 1.3:

1. Алгоритм Unit Selection (выбор речевых элементов) позволяет достичь максимальной естественности синтезированной речи, при условии корректно отсегментированной на разных уровнях сбалансированной речевой базы данных большого объема.

2. Статистические модели (СММ-синтез) позволяют легко модифицировать характеристики голоса с помощью адаптации/интерполяции дикторов. Речь, полученная на основе СММ технологии, на слух менее естественна, однако в ней отсутствуют резкие, необусловленные контекстом перепады по частоте и энергии, обычно присущие конкатенативному синтезу. Кроме того, применение технологии СММ-синтеза позволяет разрабатывать новый голос за гораздо меньший период времени, а также требует значительно меньше памяти для хранения речевой базы.

Структурно, система разделена на части:

– Обучающая часть (подготовительная стадия).
– Синтезирующая часть.

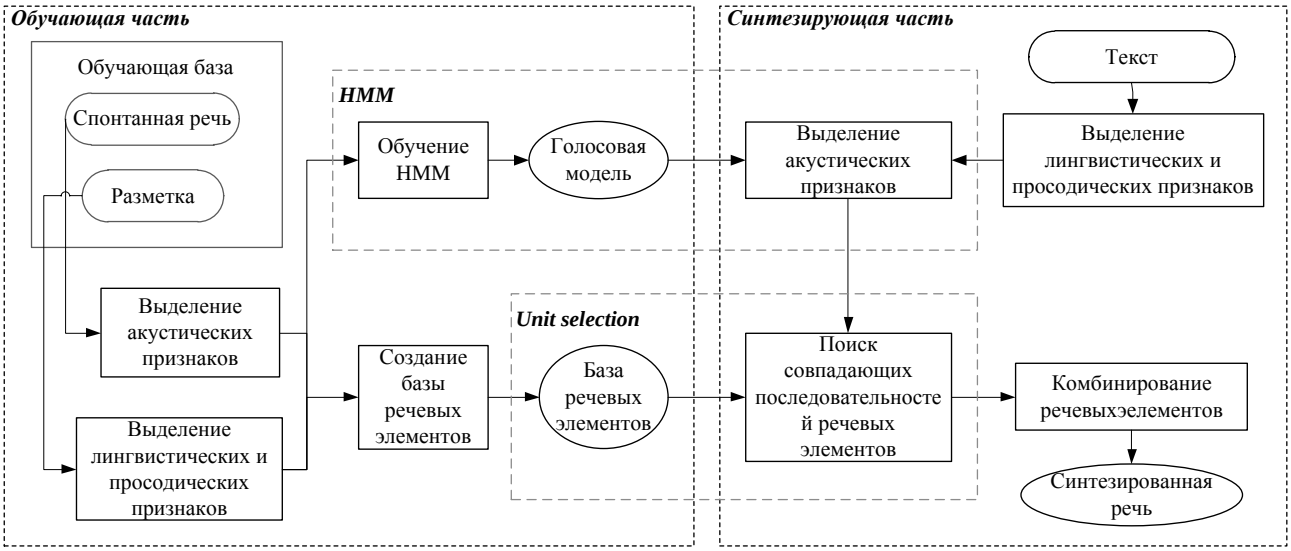

Рисунок 1.3 - Схема, иллюстрирующая основные шаги работы системы синтеза речи

Обучающая база подготавливается из «спонтанной речи», содержащей набор звуковых файлов (каждый файл содержит запись произвольного предложения) и набор соответствующих файлов разметки (они содержат информацию о речевых элементах для каждого звукового файла) [70].

Эксперименты [40] показывают, что натуральность звучания речи синтезированной на гибридной системе синтеза, улучшена относительно результатов работы систем основанных только на методе Unit Selection или

HMM. Детальное описание использованной системы синтеза представлено в [40].

Как видно из описания схемы атаки, данный способ может быть полностью автоматизирован. Это означает, что даже небольшое снижение надёжности голосовой биометрической системы под воздействием данного вида атаки, может привести к успешному взлому системы.

Другие методы атак на основе технологии синтеза, а также методы противодействия им, можно найти в работах [41, 42, 51, 73, 74, 85, 96, 98].

1.3. Анализ существующих методик оценки эффективности аутентификации ГБС

Понятие «эффективность» имеет большой разброс в трактовках применительно к системе в целом и голосовой биометрической системе в частности.

Например, в работе [13] проведён анализ стандартов и методик, содержащих различные определения понятия «эффективность». На основе проведённого анализа автор [13] предлагает в первую очередь трактовать эффективность как комплексную характеристику работы системы:

Эффективность – комплексная характеристика системы, отражающая степень ее соответствия потребностям и интересам ее заказчиков, пользователей, других заинтересованных лиц. [13]

Иными словами, эффективность может быть представлена как степень соответствия целей создания системы и фактических результатов работы системы.

Помимо этого, в работе [13] сформулированы следующие выводы:

- эффективность голосовой биометрической системы меняется при изменении потребностей сторон или других условий, ее оценка справедлива для конкретного периода времени;
- эффективность голосовой биометрической системы в большинстве случаев невозможно оценивать как «эффективность ее самой по себе», можно говорить только об оценке вклада этой голосовой биометрической системы в достижение целей организации, региона или даже страны.

Чтобы численно оценить эффективность аутентификации выполняемой биометрическими системами, необходимо определить показатели голосовой биометрической системы, поддающиеся измерению или расчету.

Как следует из определения эффективности системы в целом, численные показатели эффективности должны определять меру соответствия результатов работы системы задачам, для решения которых она была разработана. Данные обобщающие показатели должны характеризовать степень оптимальности работы системы.

Показателями, наиболее полно характеризующими оптимальность функционирования системы, являются экономические показатели эффективности системы. Данные показатели отображают зависимость полученной выгоды от затрат, потраченных на разработку и поддержку системы.

Согласно ГОСТ Р ИСО 9000-2001 «Системы менеджмента качества. Основные положения и словарь» эффективность функционирования информационной системы определяется соотношением затраченных ресурсов и результатов.

Установить стоимость затраченных на внедрение биометрической системы ресурсов легко. Гораздо сложнее оценить результаты внедрения.

Результатом может быть сокращение трудозатрат сотрудников и клиентов, уменьшение расходов на карты, ключи и другие материальные идентификаторы, наконец, повышение уровня безопасности. Кроме того, конечная цель внедрения может формулироваться не в количественных, а в качественных показателях.

Таким образом, необходимо определиться с набором показателей, позволяющих не только достаточно полно численно оценивать эффективность аутентификации голосовыми биометрическими системами, но и корректно сравнивать их между собой.

В настоящий момент исследователи активно используются такие показатели эффективности аутентификации выполняемой биометрическими системами, как:

1. Вероятность ложного доступа пользователя, не зарегистрированного в голосовой биометрической системе.
2. Вероятность ложного совпадения входного образца голоса с несоответствующим шаблоном в базе данных шаблонов.
3. Вероятность ложного отклонения подлинного биометрического образца зарегистрированного в ней пользователя.
4. Вероятность ложного несовпадения при сравнении входного образца и соответствующего ему зарегистрированного шаблона.
5. График рабочей характеристики, визуализирующий компромисс между характеристиками ложного совпадения и ложного несовпадения.
6. Равновероятную ошибку или вероятность переходных ошибок, при которых ошибка допуска совпадает с ошибкой отклонения.
7. Вероятность отказа в регистрации при попытке создать шаблон из входных данных.

Для оценки экономической эффективности биометрических систем можно использовать подходы и положения ГОСТ 24.702-85 «Единая система стандартов автоматизированных систем управления. Эффективность автоматизированных систем управления. Основные положения».

Данный стандарт распространяется на автоматизированные системы управления (далее – АСУ) всех видов и назначений, и их части, вводимые в эксплуатацию для всех уровней управления, кроме общегосударственного, и устанавливает основные положения по определению эффективности АСУ и принципы оценки экономической эффективности АСУ.

Для каждой конкретной АСУ цель ее создания состоит в обеспечении наиболее полного использования потенциальных возможностей объекта управления для решения, поставленных перед ним задач. Поскольку решение любой биометрической системы, в конечном счете, также используется объектом управления, то к ней правомерно применять показатели эффективности АСУ.

Согласно ГОСТ 24.702-85 эффективность АСУ определяют сопоставлением результатов от функционирования АСУ и затрат всех видов ресурсов, необходимых для ее создания и развития.

Критерий эффективности АСУ определяют на множестве (системе) показателей, каждый из которых описывает одну из сторон рассматриваемой системы. В зависимости от используемого математического аппарата критерий может быть выражен в виде целевой функции или порядковой меры, устанавливающей упорядоченную последовательность сочетаний показателей.

При определении результатов от функционирования АСУ задают универсальную систему обобщенных показателей, таких, как оперативность (своевременность), устойчивость, качество управления и др.

Используемые показатели должны быть развернуты применительно к характеристикам конкретной АСУ (например: оперативность – вероятностно-временные характеристики элементов процесса управления; устойчивость – показатели надежности, помехозащищенности и т. п.).

Согласно действующим нормативным документам биометрические устройства и системы могут быть подвергнуты различным испытаниям. Данные испытания согласно ГОСТ Р ИСО/МЭК ТО 24741 могут включать в себя оценку:

- эксплуатационных характеристик (в терминах вероятностей ошибок и производительности);
- надёжности, доступности и удобства эксплуатации;
- степени защищённости;
- безопасности;
- приемлемости системы для пользователя;
- влияния человеческих факторов;
- коэффициента эффективности затрат;
- степени соответствия правилам конфиденциальности.

В течение последних трех десятилетий оценка эксплуатационных характеристик является наиболее распространенной формой испытаний. Эксплуатационные испытания обычно проводят с целью прогноза эксплуатационных характеристик системы для целевой выборки и в целевых условиях применения, но исторически сложилось так, что экстраполяция результатов испытаний в тестовых условиях на «практику» вызывает много трудностей.

Для того чтобы результаты испытаний лучше соответствовали эксплуатационным характеристикам систем при практической эксплуатации,

разработаны стандарты, устанавливающие процедуры проведения испытаний.

В ГОСТ Р ИСО/МЭК 19795-1-2007 в качестве эксплуатационных характеристик определены следующие группы характеристик:

- вероятности появления ошибок;
- показатели пропускной способности.

Ошибки результатов верификации и идентификации возникают вследствие ошибок соответствия (т.е. ошибок ложного соответствия и ложного несоответствия) или ошибок получения образцов (т.е. отказов регистрации и отказов сбора данных).

Влияние сочетания данных базовых ошибок на появление ошибок принятия решения зависит от числа требуемых сравнений, от того, является ли истинным или ложным запрос идентичности, а также от политики принятия решения, т.е. допускает ли биометрическая система проведение нескольких попыток.

Несмотря на то, что описание эксплуатационных характеристик в биометрии традиционно проводилось в терминах вероятностей допуска, т.е. вероятностей ложного допуска и ложного недопуска, в литературе неявно возникают определения, противоречащие друг другу: в описании идентификационных голосовых биометрических систем встречается понятие «ложное отклонение», возникающее вследствие неправильного соответствия представленного образца шаблону, зарегистрированному другим пользователем.

В литературе по управлению доступом встречается понятие «ложное принятие», возникающее вследствие неправильного соответствия представленного образца шаблону, зарегистрированному другим пользователем.

Чтобы исключить путаницу, следует использовать стандартные определения. ВЛС (FMR) и ВЛНС (FNMR) в общем случае не являются синонимами ВЛД (FAR) и ВЛНД (FRR). ВЛС (FMR) и ВЛНС (FNMR) вычисляют относительно сравнений, а ВЛД (FAR) и ВЛНД (FRR) – относительно транзакций и относят к принятию или отклонению утверждаемых гипотез, положительных или отрицательных. ВЛД и ВЛНД также включают в себя отказы сбора данных.

Показатели пропускной способности устанавливают число пользователей, подвергаемых анализу в единицу времени, зависящее от скорости вычислений и взаимодействия человека с биометрической системой.

В общем случае данные показатели применяются во всех биометрических системах и устройствах. Достижение достаточного значения пропускной способности является важным показателем работы голосовой биометрической системы.

Показатели пропускной способности для системы идентификации, например для системы регистрации, в программе социального обеспечения, могут быть сильно занижены из-за потерь во времени, необходимых для сравнения зарегистрированного образца с образцами базы данных.

Показатели пропускной способности для системы верификации, например для системы управлении доступом, обычно связаны со скоростью взаимодействия пользователя с системой в процессе получения высококачественного биометрического образца.

Таким образом, в зависимости от типа системы целесообразно определить время взаимодействия пользователя с системой, а также время режима работы вычислительных аппаратных и программных средств.

Фактические экспериментальные измерения быстродействия вычислительной системы приведены в таких руководствах, как [63], и не рассматриваются в настоящей работе.

1.4. Выводы

В данной главе были рассмотрены современные голосовые биометрические системы, стандарты в области оценки эффективности аутентификации выполняемой подобными системами. Был приведен обзор возможных атак на голосовые биометрические системы, основанных на различных способах фальсификации индивидуальных биометрических характеристик человека.

Имеющиеся методики и стандарты не предусматривают оценку устойчивости голосовых биометрических систем к различным способам имитационных атак с целью взлома.

ГЛАВА 2. Анализ уязвимости голосовых биометрических систем

До недавнего времени, многие исследователи в области голосовой биометрии полагали, что реализация успешной атаки на голосовую биометрическую систему, путём фальсификации индивидуальных биометрических характеристик человека методами преобразования или синтеза речи, невозможна [31]. Основным доводом в таких выводах являлась необходимость разработки готовых систем, обладающих возможностью учёта индивидуальных характеристик диктора, которые далее используются при его аутентификации в системе. Однако, технологии не стоят на месте, и как было показано в первой главе, подобные системы уже разработаны, а значит могут быть получены злоумышленниками.

Учитывая активное внедрение голосовых биометрических технологий в повседневной жизни, становится очевидной необходимость проведения анализа уязвимости модулей обобщённой голосовой биометрической системы.

2.1. Анализ уязвимости модулей обобщённой голосовой биометрической системы

Классифицируем нарушителей по их доступу к системе:

- внутренние - имеют возможность считывать, записывать и модифицировать данные в биометрической системе на программном или аппаратном уровне;
- внешние - имеют возможность взаимодействовать с системой только на уровне устройства ввода, как и обычные пользователи системы.

Определим возможные цели, преследуемые нарушителем:

- получение доступа к защищённой информации за счёт срабатывания ложноположительной ошибки голосовой биометрической системы;
- деперсонификация т.е. исключение конкретного шаблона из результатов верификации или идентификации голосовой биометрической системы.

Опишем основные возможные сценарии воздействия нарушителя на голосовую биометрическую систему:

- замена зарегистрированного в системе шаблона на неправомочный шаблон;
- удаление зарегистрированного в системе шаблона;
- добавление в систему неправомочного шаблона;
- влияние на уровень порогового значения принятия решения;
- использование модифицированного, неавторизованного биометрического оборудования;
- ввод речевых данных в биометрическую систему нестандартным образом;
- воздействие на протоколы или каналы передачи данных.

Отметим, что реализация определённых ранее целей по описанным сценариям возможна как внутренними нарушителями, так и внешними.

Выделим основные компоненты обобщённой голосовой биометрической системы с их связями. В схеме, изображённой на рисунке 2.1, любой из элементов может быть подвергнут атаке с целью взлома системы.

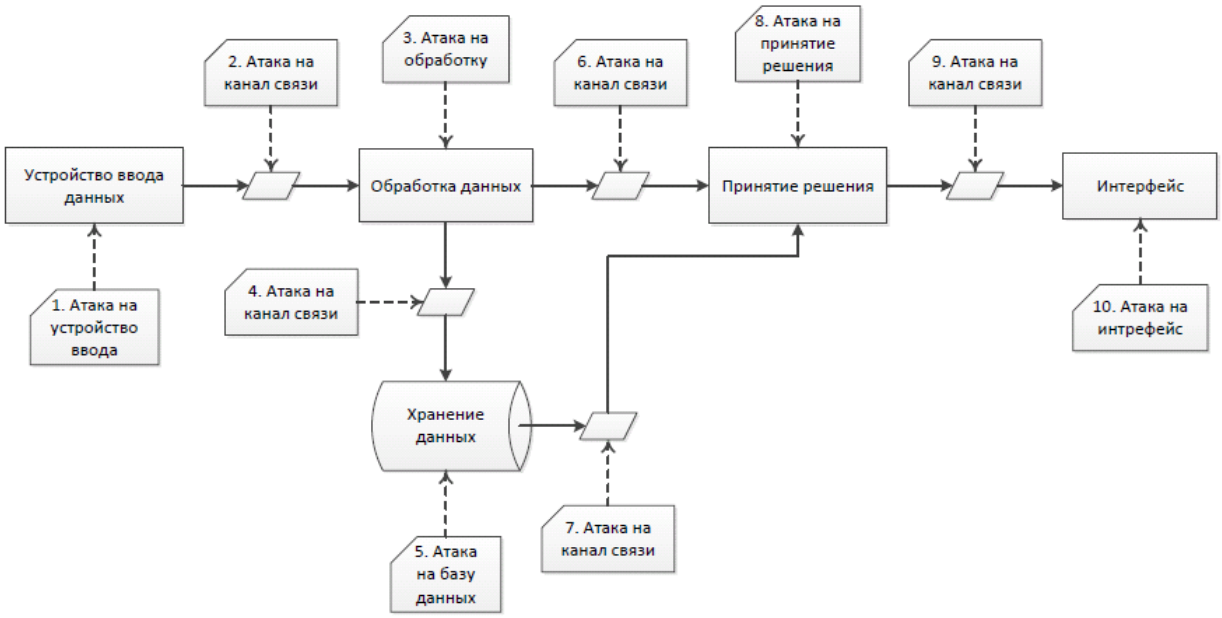

Рисунок 2.1 - Компоненты обобщённой голосовой биометрической системы с обозначенными атаками

Опишем типичные атаки на различные элементы обобщённой голосовой биометрической системы:

1. Атака на устройство ввода биометрической информации.
2. Атака на канал связи между устройством ввода и компонентом обработки данных.
3. Атака на компонент обработки данных.
4. Атака на канал связи между компонентом обработки данных и базой данных речевых шаблонов.
5. Атака на базу данных речевых шаблонов.
6. Атака на канал связи между компонентом обработки данных и компонентом принятия решения.
7. Атака на канал связи между базой данных и компонентом принятия решения.
8. Атака на компонент принятия решения.

9. Атака на канал связи между компонентом принятия решения и интерфейсом вывода результата работы голосовой биометрической системы.
10. Атака на интерфейс вывода результата работы голосовой биометрической системы.

В работах [4, 72] сделан вывод о том, что все перечисленные атаки, за исключением атаки на устройство ввода биометрической информации, являются общими, вне зависимости от модальности биометрической системы. Эффективное противодействие данным атакам достигается применением цифрового кодирования, шифрования открытого канала передачи данных и использованием временных меток.

Таким образом, наиболее уязвимым компонентом системы остаётся модуль ввода биометрической информации. Исключим из рассмотрения оценку устойчивости голосовых биометрических систем к методам атак, основанным на методах имперсонализации, а также методах записи и повтора, т.к. данные виды атак не зависят от развития технологии и уже достаточно изучены ранее [29].

2.2. Оценка устойчивости к спуфинг атакам на основе преобразования БХЧ

Для оценки устойчивости к спуфинг атакам на основе преобразования индивидуальных биометрических характеристик человека рассмотрим текстонезависимую систему верификации, основанную на одном из самых эффективных методов - методе i-векторов [66].

Последние соревнования NIST SRE 2012 [84] показали, что на сегодняшний день системы на базе представления модели голоса диктора в пространстве полной изменчивости (total variability, TV) являются доминирующими. Суть метода заключается в использовании смесей

гауссовых распределений (СГР) для моделирования голоса диктора, а затем редукции СГР до так называемого i-вектора в пространстве полной изменчивости низкой размерности.

Для улучшения качества распознавания диктора в данной системе используется модуль предварительной обработки сигнала. Данный модуль включает детектор речи, основанный на энергии сигнала, а также детекторы клиппирования сигнала, импульсных и тональных помех [2]. В качестве речевых признаков используются вектора мел-частотных кепстральных коэффициентов (mel-frequency cepstrum coefficients, MFCC)[49], а также их производные первого и второго порядка в количестве 13 элементов. Длина каждого речевого кадра для вычисления MFCC составляет 22мс и сдвигается при следующем шаге на 11мс. Для компенсации эффекта Гиббса, в тестируемых системах, используется взвешивание сигнала окном Хемминга. Компенсация эффектов канальных искажений на уровне признаков реализована путем вычитания кепстрального среднего (cepstral mean subtraction, CMS).

На этапе создании модели голоса диктора, в тестируемых системах, используется гендеро-независимая универсальная фоновая модель (universal background model, UBM), представленная 512-компонентной СГР. Обучение UBM рассматриваемой системы производилось с помощью стандартного EM-алгоритма на телефонной части речевых баз данных NIST SRE 1988-2010 [17]. Для ускорения вычислений применяется диагональная ковариационная матрица UBM. Общее количество дикторов в обучающих базах данных составляло около 4000.

Модуль оценки i-вектора рассматриваемой системы также обучен на более чем 60000 телефонных и микрофонных записях из речевых баз данных NIST SRE 1998-2010, включающих голоса более 4000 дикторов.

Основное выражение [32], определяющее представление модели СГР в низкоразмерном пространстве полной изменчивости, приведено ниже:

$$\mu = m + T\omega + \varepsilon$$

где μ- супервектор параметров СГР модели диктора,

m- супервектор параметров UBM,

T - матрица, задающая базис в редуцированном пространстве признаков,

ω - i-вектор в редуцированном пространстве признаков, $\omega \in N(0,1)$,

ε - вектор ошибки.

Модуль линейного дискриминантного анализа рассматриваемой системы также обучен на речевых базах данных NIST SRE 1988-2010.

Рассматриваемая система не содержит каких либо решений по противодействию спуфинг атакам на модуль ввода биометрической информации, как и большинство других современных систем.

2.2.1. Описание способа атаки

Общая схема имитируемой атаки представлена на рисунке 2.2.

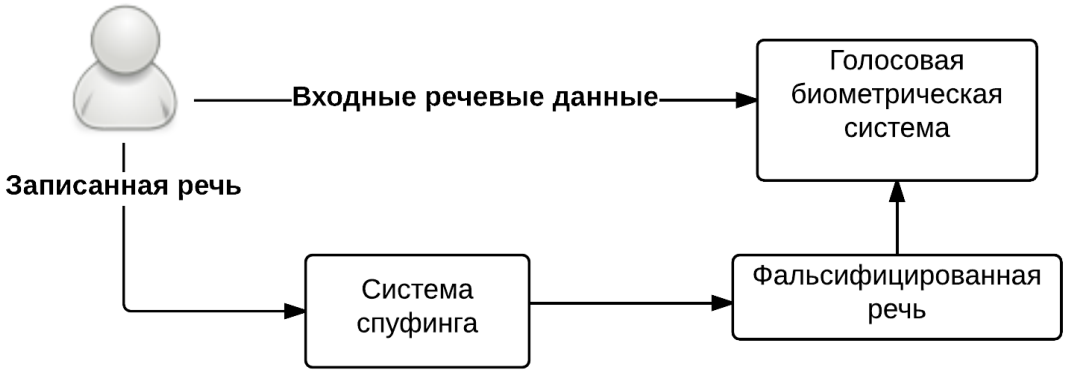

Рисунок 2.2 - Общая схема имитируемой атаки на модуль ввода биометрической информации текстонезависимой системы верификации

В качестве методов атаки использовались следующие методы преобразования индивидуальных биометрических характеристик человека:

1. Метод преобразования речи, основанный на простой реализации алгоритма подбора фрейма (frame selection) описанный в работах [46, 95]. Преобразование речи происходит путём замены фреймов речи "самозванца" на соответствующие фреймы речи пользователя голосовой биометрической системы.
2. Метод преобразования речи, основанный на простейшем алгоритме замены первого мел-кепстального коэффициента таким образом, чтобы максимально уменьшить различие в спектрах целевой и исходной речи [48].
3. Метод преобразования речи основанный на использовании системы Festvox [22].

Данные методы были представлены на конкурсе Automatic Speaker Verification Spoofing and Countermeasures (ASVspoof) Challenge 2015 [94] под номерами S1, S2 и S5 соответственно. Описание результатов данного конкурса будет частично приведено в четвёртой главе.

2.2.2. Результаты экспериментальной оценки

Для проведения экспериментальной оценки была подготовлена тестовая речевая база данных, содержащая 3497 записи оригинальной речи 35 дикторов, 20 из которых женщины, а также 29925 записей содержащих фальсифицированную речь данных дикторов. По 9975 для каждого метода фальсификации.

Полученные результаты отображены на рисунке 2.3 в виде четырёх кривых компромиссного определения ошибки. Первая отображает уровень надёжности системы, полученный только на оригинальных записях речи, вторая, третья и четвёртая для каждого из рассматриваемых методов спуфинга соответственно.

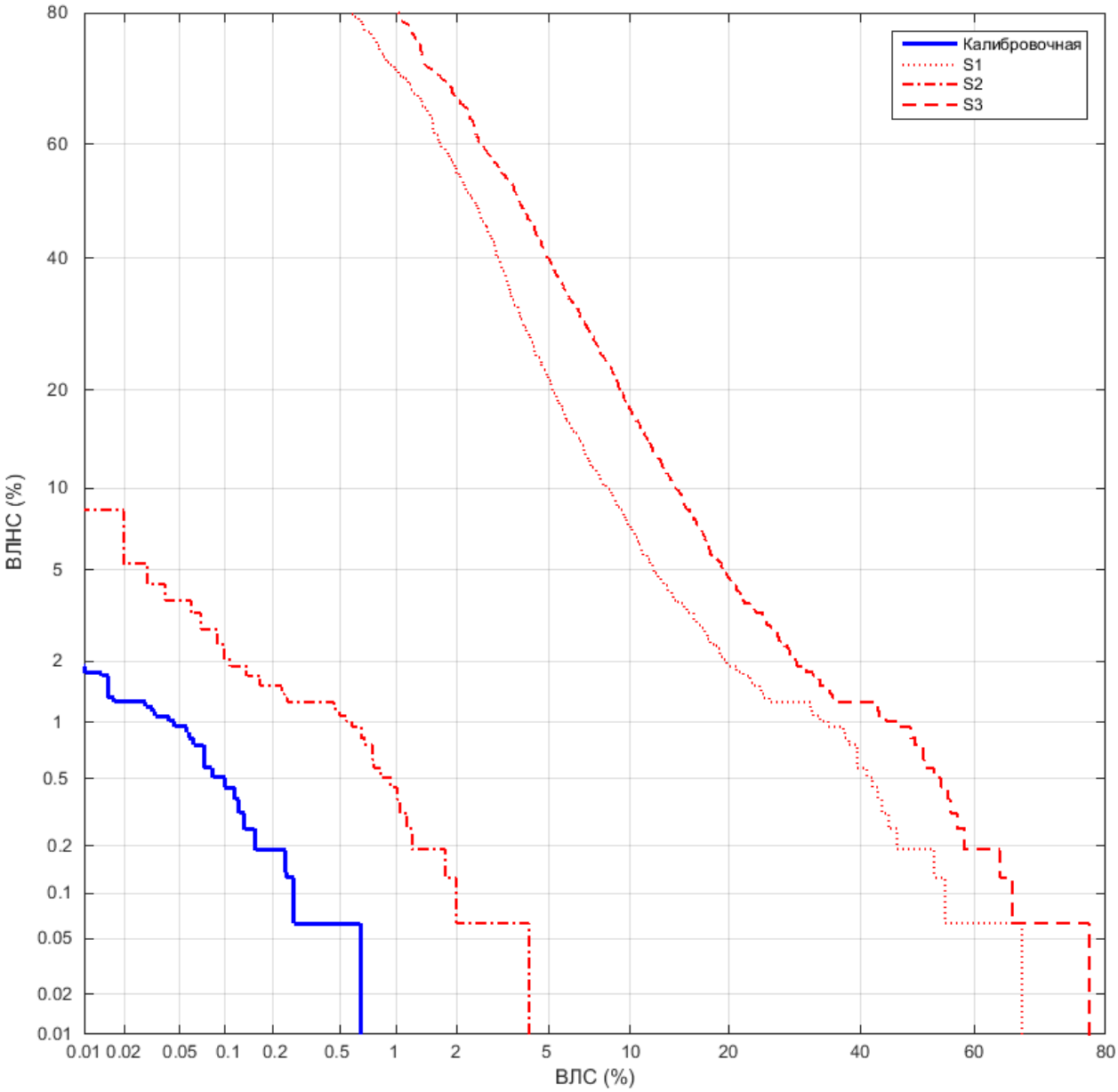

Рисунок 2.3 - Кривые КОО полученные при имитации трёх методов спуфинга основанных на технологии преобразования речи

Как видно из результатов, второй метод спуфинга, основанный на замене первого мел-кепстрального коэффициента не оказывает значительного влияния на уровень надёжности оцениваемой голосовой биометрической системы. В тоже время третий метод увеличил уровень ошибки системы более чем на десять процентов.

2.3. Оценка устойчивости к спуфинг атакам на основе синтеза БХЧ

Для оценки устойчивости к спуфинг атакам на основе синтеза индивидуальных биометрических характеристик человека рассмотрим туже текстонезависимую систему верификации, основанную на методе i-векторов.

2.3.1. Описание способа атаки

Общая схема имитируемой атаки остаётся прежней.

В качестве методов атаки использовались следующие методы синтеза индивидуальных биометрических характеристик человека:

1. Метод синтеза речи основанный на подходе использования Скрытых Марковских Моделей с применением алгоритма адаптации под целевого диктора [98] произведённого на двадцати адаптационных фразах зарегистрированного в системе диктора.
2. Метод синтеза речи основанный на подходе использования Скрытых Марковских Моделей с применением алгоритма адаптации под целевого диктора, произведённого на сорока адаптационных фразах зарегистрированного в системе диктора.

Данные методы были также представлены на конкурсе ASVspoof Challenge 2015 [94] под номерами S3 и S4 соответственно.

2.3.2. Результаты экспериментальной оценки

Для проведения экспериментальной оценки была подготовлена тестовая речевая база данных, содержащая 3497 записи оригинальной речи 35 дикторов, 20 из которых женщины, а также 19950 записей содержащих фальсифицированную речь данных дикторов. По 9975 для каждого метода фальсификации.

Полученные результаты отображены на рисунке 2.4 в виде трёх кривых компромиссного определения ошибки. Первая отображает уровень надёжности системы, полученный только на оригинальных записях речи, вторая и третья для каждого из рассматриваемых методов спуфинга соответственно.

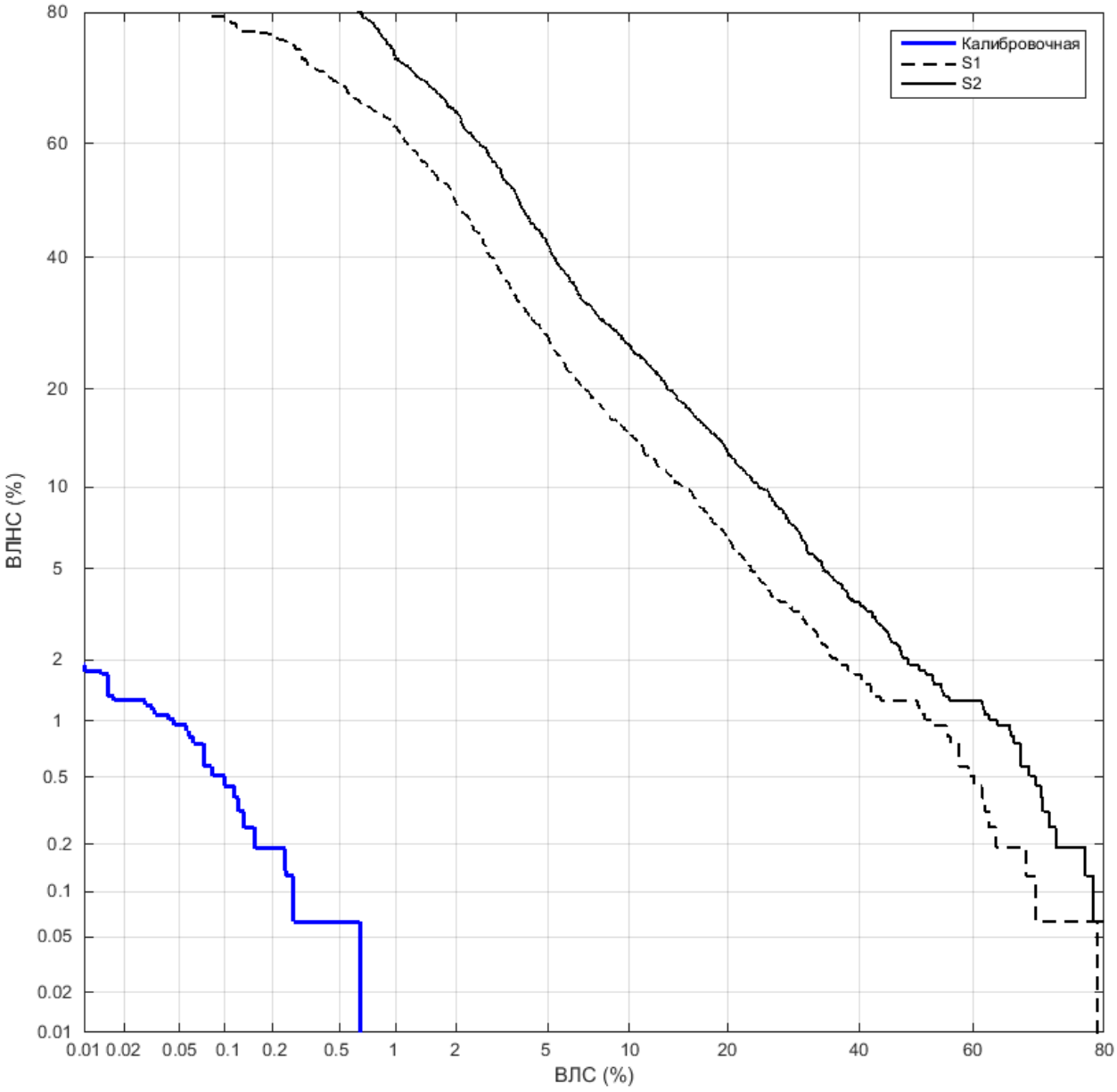

Рисунок 2.4 - Кривые КОО полученные при имитации двух методов спуфинга основанных на технологии синтеза речи

Как видно из результатов проведённой оценки, оба метода спуфинга, основанные на одном из самых простых алгоритмов синтеза увеличили уровень ошибки системы более чем на десять процентов.

Таким образом, можно сделать следующий вывод: системы спуфинга, основанные на технологии синтеза представляют угрозу большую, чем системы спуфинга, основанные на технологии преобразования речи. Особенно с учётом отсутствия необходимости непосредственного участия злоумышленника, при проведении атаки основанной на технологии синтеза.

2.4. Метод имитации спуфинг атак на основе автоматического создания модели голоса синтеза

Отметим, что методы спуфинга показали большую опасность для текстонезависимой системы верификации, даже без учёта возможности их полной автоматизации. В связи с этим, оценим влияние длительности и качества подготовки речевого материала, используемого для подготовки системы спуфинга на более защищённую голосовую биометрическую систему. Для этого будем использовать текстозависимую систему верификации, использующую сценарий работы с динамической парольной фразой. Данный сценарий даёт гарантированную защиту от попыток взломать систему методом записи и повтора индивидуальных биометрических характеристик человека.

Для оценки влияния объёма и качества, обучающих данных системы синтеза, на устойчивость ГБС будем использовать гибридный метод синтеза на основе использования метода Unit Selection и скрытых марковских моделей (СММ), который является наилучшим на сегодняшний день методом синтеза речи [40].

Рассматриваемый метод имитации атаки заключается в создании синтезированного голоса пользователя системы верификации. Для обучения системы синтеза используется предварительно записанная спонтанная речь пользователя. На этапе текстозависимой верификации, при помощи полученного синтезированного голоса и перехваченного парольного текста, создаётся синтезированная парольная фраза, используемая далее для

попытки верификации. Детальная схема процедуры атаки представлена на рисунке 2.5.

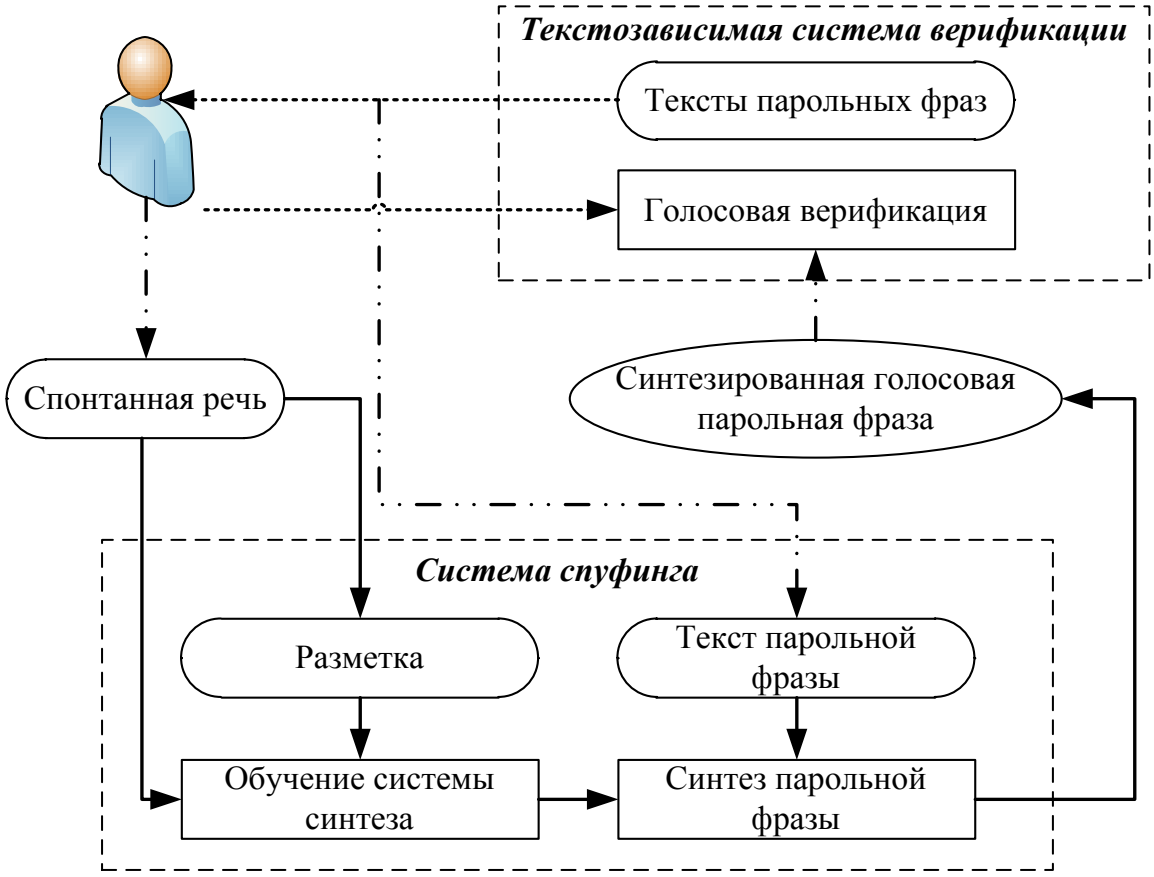

Рисунок 2.5 - Схема имитируемой атаки на текстозависимую систему верификации с динамической парольной фразой

Для имитации системы спуфинг-атаки будем использовать систему голосового синтеза, разработанную в ООО «ЦРТ» [39]. Данная система основана на комбинировании двух, наиболее популярных, подходов.

3. Алгоритм Unit Selection (выбор речевых элементов) позволяет достичь максимальной естественности синтезированной речи, при условии корректно отсегментированной на разных уровнях сбалансированной речевой базы данных большого объема.
4. Статистические модели (CMM-синтез) позволяют легко модифицировать характеристики голоса с помощью адаптации/интерполяции дикторов. Речь, полученная на основе CMM технологии, на слух менее естественна, однако в ней отсутствуют

резкие, необусловленные контекстом перепады по частоте и энергии, обычно присущие конкатенативному синтезу. Кроме того, применение технологии СММ-синтеза позволяет разрабатывать новый голос за гораздо меньший период времени, а также требует значительно меньше памяти для хранения речевой базы.

Эксперименты показали повышение естественности синтезируемой речи по сравнению с реализацией на Unit Selection или только на основе СММ-технологии. Детальное описание представленной системы синтеза и результатов экспериментальной оценки естественности синтезируемой речи доступно в статье [40].

2.4.1. Влияние объёма обучающих данных системы синтеза на устойчивость ГБС

Для проведения оценки была подготовлена тестовая речевая база русского языка, содержащая семь различных дикторов (двое мужчин и пять женщин), голоса которых использовались для обучения системы синтеза речи. Для каждого диктора было записано по девять парольных фраз длительностью 2-3 секунды речи. Примеры парольной фразы: «Город Екатеринбург, улица вокзальная, дом 22, вокзал», «заплатить три рубля и дать объявление в бюллетене» и т.п. Важно отметить, что записанные фразы не участвовали в дальнейшем при обучении системы синтеза речи. Итого было записано 63 фразы различных дикторов.

Целью данных экспериментов являлась установление зависимости ошибки ложного принятия верификации (FA) от длительности речевого материала, используемого при обучении системы синтеза голоса.

Для экспериментов была взята описанная ранее система верификации по голосу. Калибровка порогов срабатывания системы производилась на речевой базе YOHO [38] содержащей 138 дикторов (мужчины и женщины),

каждый из которых произносил фиксированную парольную фразу вида «36-24-36» длительностью около 1.5-2 секунды активной речи.

Было выставлено два порога системы верификации:

1. Порог по равновероятной ошибке пропуска-отклонения (equal error rate, EER), так называемый ThresholdEER. На калибровочной базе EER был равен 4%.
2. Порог при вероятности ложного принятия не более 1% - ThresholdFA1. Данный порог обычно используется в системах, где необходимо обеспечить максимальную защиту от доступа злоумышленника.

Далее для каждого диктора производились попытки доступа в систему верификации с помощью синтезированных парольных фраз, подготовленных системой синтеза голоса. При этом объем спонтанной речи использованной для обучения синтеза менялся от одной минуты до четырёх часов речи для каждого диктора. Результаты эксперимента представлены на рисунке 2.6.

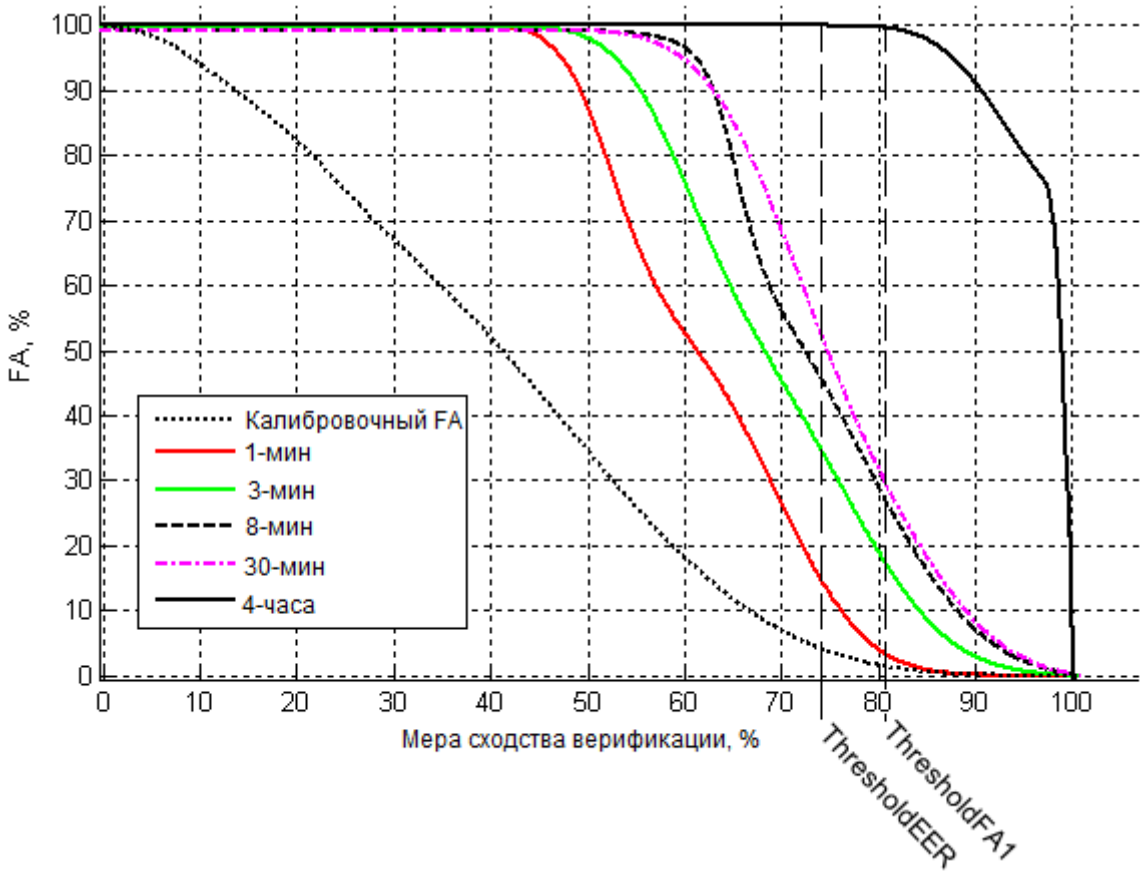

Рисунок 2.6 - Графики зависимости вероятности ложного совпадения от длительности спонтанной речи, использованной для обучения синтеза

Дополнительно, в таблице 2.1 представлены значения вероятности ложного принятия для двух порогов исследуемой системы верификации. Как видно из таблицы, надёжность системы верификации значительно ухудшается при использовании спонтанной речи длительностью от восьми минут и более для обучения системы синтеза. В случае если синтез был подготовлен на данных большого объема (четыре часа речи), синтезированная речь практически перестаёт отличаться от живой речи человека с точки зрения системы верификации.

Таблица 2.1 - Значения вероятности ложного совпадения для различных объёмов обучающих данных системы синтеза

Объем речи для обучения синтеза	Ошибка FA для порога ThresholdEER (%)	Ошибка FA для порога ThresholdFA1 (%)
1 минута	12,7% (8)	1,5% (1)

3 минуты	34,9% (22)	7,9% (5)
8 минут	44,4% (28)	19,1% (12)
30 минут	55,6% (35)	23,8% (15)
4 часа	100% (63)	98,4% (62)

На основании полученных результатов можно сделать выводы о том, что предлагаемый способ спуфинга позволяет не только серьезно ослаблять надежность системы верификации, но и обходить такую дополнительную защиту, как детектор присутствия диктора. В случае если система верификации сообщает пользователю пароль в виде звукового сообщения, возможно использование системы распознавания речи для полной автоматизации процесса спуфинга. В отличие от методов спуфинга путем преобразования признаков речи, предложенный метод, при использовании совместно с системой распознавания речи, позволяет исключить участие человека из диалога с системой верификации.

2.4.2. Влияние качества обработки обучающих данных системы синтеза на устойчивость ГБС

Учитывая высокую уязвимость ГБС к методам спуфинга основанным на технологии синтеза обученной даже на небольших объёмах речевого материала, необходимо оценить влияние качества обработки обучающих данных.

В предыдущем эксперименте, обучающие данные были размечены экспертами вручную.

В данном эксперименте будем использовать технологию автоматической разметки, состоящую из разбиения обучающих речевых данных на периоды основного тона (ЧОТ) и проведения аллофонной сегментации. Для выполнения разметки по ЧОТ используется автокорреляционный метод расчёта основного тона с предварительной фильтрацией и постобработкой

для уточнения положения меток основного тона. Низкочастотная фильтрация используется для снижения ошибки определения меток основного тона путём удаления из сигнала составляющих с частотой выше 500 Гц. Высокочастотная предварительная фильтрация используется для определения участков, на которых нет меток основного тона (невокализированные звуки).

Аллофонная сегментация выполняется автоматически с помощью модулей системы распознавания речи с использованием СММ. Сегментация проводится на основе force alignment транскрипции и звукового сигнала. Она состоит из трёх этапов:

1. Построение акустических моделей монофонов, т.к. именно монофоны наилучшим образом подходят для этой задачи;
2. Получение «идеальной» сегментации - в точности соответствует заданной транскрипции, и «реальной» сегментации - отличается более точным акустическим соответствием с фонограммой;
3. Автоматическая корректировка полученных на предыдущих этапах границ аллофонов на основе разметки частоты основного тона.

Детально, процесс автоматической подготовки синтезированного голоса описан в работе [79].

В качестве обучающих данных для системы синтеза была взята речевая база данных русской речи, записанная в телефонном канале связи. Для проведения оценки была подготовлена тестовая речевая база русского языка, содержащая пять различных дикторов, голоса которых использовались для обучения системы синтеза речи. Для каждого диктора было записано по девять парольных фраз длительностью 2-3 секунды речи. Примеры парольной фразы: «Город Екатеринбург, улица вокзальная, дом 22, вокзал»,

«заплатить три рубля и дать объявление в бюллетене» и т.п. Важно отметить, что записанные фразы не участвовали в дальнейшем при обучении системы синтеза речи. Итого было записано 95 фраз различных дикторов.

Система верификации, калибровочная база и пороги не отличались от предыдущего эксперимента. Полученные результаты приведены на рисунке 2.7.

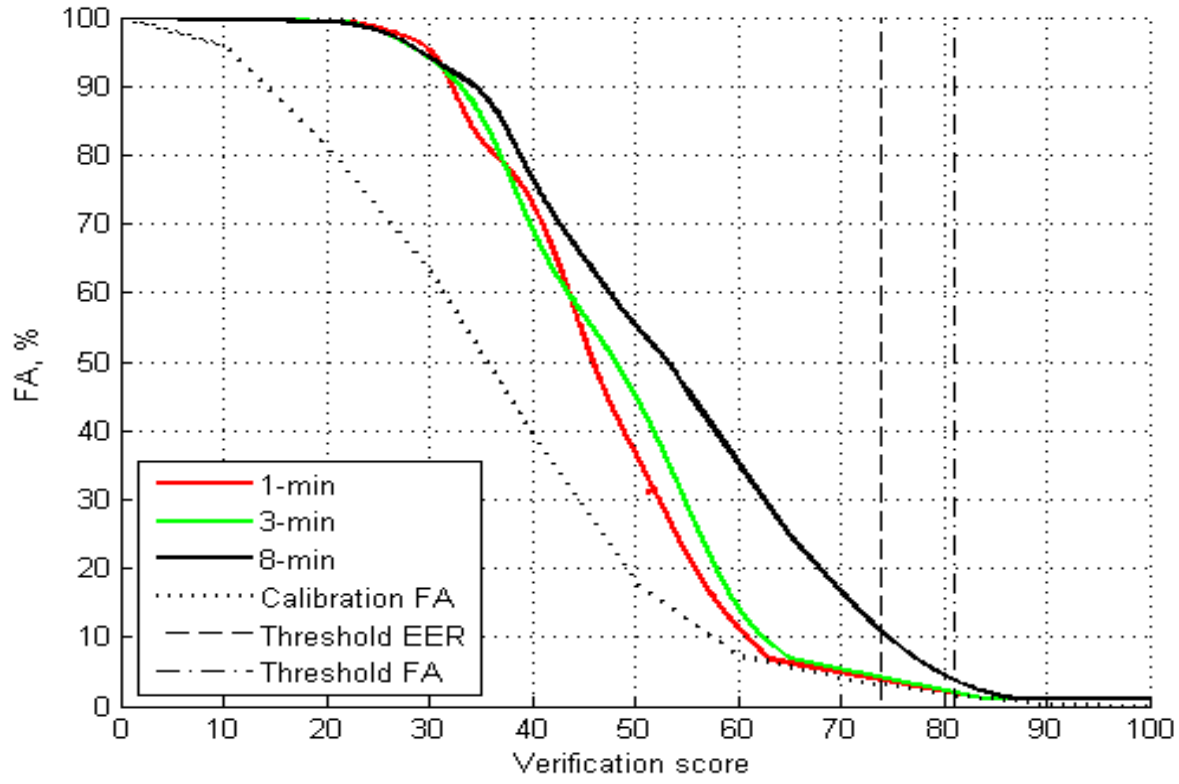

Рисунок. 2.7 - Графики зависимости вероятности ложного принятия от длительности спонтанной речи, использованной для обучения синтеза на автоматически размеченных данных

В таблице 2.2 приведены результаты обоих экспериментов для двух выбранных порогов. В первом эксперименте, в качестве обучающих данных для системы синтеза использовались экспертная разметка и записанная в студийном качестве речь. Во втором эксперименте применялась технология автоматической разметки без последующей корректировки экспертом, а в качестве обучающих данных использовалась речь из телефонного канала.

Таблица 2.2 - Значения вероятности ложного совпадения для различных объёмов обучающих данных системы синтеза размеченных экспертами и автоматически

Объем речи для обучения синтеза	Ошибка FA для порога ThresholdEER (%)		Ошибка FA для порога ThresholdFA1 (%)	
	Экспертная разметка	Автоматическая разметка	Экспертная разметка	Автоматическая разметка
1 минута	12,7%	1,1%	1,5%	1,1%
3 минуты	34,9%	4,6%	7,9%	1,8%
8 минут	44,4%	10,8%	19,1%	4,5%

Как показали эксперименты, метод спуфинга использующий в качестве обучающих данных для системы синтеза, речь из телефонного канала с последующей автоматической разметкой показывает значительно меньший уровень ошибки ложного совпадения, чем при обучении системы синтеза на речи, записанной в студийном качестве с последующей ручной экспертной разметкой. Так, при обучении на восьми минутах речи, метод спуфинга использующий автоматическую разметку для обучающих данных системы синтеза показал только 10% ошибки ложного совпадения, против 44% полученных ранее.

Тем не менее, полученный результат ещё раз показывает высокую необходимость проверки систем верификации на устойчивость к спуфингу различными методами и разработки технических решений по совершенствованию защиты голосовых биометрических систем от методов спуфинга основанных на преобразовании и синтезе индивидуальных биометрических характеристик человека. Т.к. даже 10% уровень ошибки ложного совпадения, при полной автоматизации атаки позволяет легко взломать систему голосовой верификации.

2.5. Выводы

Из результатов проведённого анализа можно сделать следующий вывод. Методы спуфинг атак на устройство ввода голосовой биометрической системы, основанные на технологии синтеза, представляют угрозу большую, чем аналогичные методы на базе технологии преобразования речи. При этом, несмотря на значительную редукцию вероятности успешной атаки при использовании не качественных данных для подготовки системы синтеза и полной автоматизации процесса, надёжность работы голосовой биометрической системы уменьшается значительно.

Таким образом, при разработке методики оценки эффективности аутентификации голосовыми биометрическими системами, учитывающей возможные атаки на модуль ввода биометрической информации, необходимо учесть следующие моменты:

- Возможность наличия неэффективных методов атаки на голосовую биометрическую систему.
- Возможный низкий уровень разборчивости синтезированной речи.
- Необходимость сбалансированного использования различных методов имитации атаки.

ГЛАВА 3. Методика и комплекс программных средств оценки эффективности аутентификации голосовыми биометрическими системами

Как было отмечено в первой главе, принятые в настоящее время стандарты оценки эффективности аутентификации голосовыми биометрическими системами, не включают в себя оценку устойчивости к спуфинг атакам на модуль ввода биометрической информации. Вместе с этим, высокая уязвимость существующих голосовых биометрических систем к уже существующим методам фальсификации индивидуальных биометрических характеристик человека требует разработки перспективных средств противодействия подобным методам атаки. Таким образом, создание методики оценки аутентификации выполняемой голосовыми биометрическими системами, включающей оценку устойчивости к спуфинг атакам на модуль ввода биометрической информации, и позволяющей корректно сравнивать между собой различные голосовые биометрические системы, является крайне важной задачей. Данной точки зрения придерживаются многие исследователи в области разработки голосовых биометрических систем [47]. Стоит также отметить, необходимость разработки комплекса программных средств, позволяющего проводить оценку различных систем в сжатые сроки, с минимумом необходимых трудозатрат.

Разработанная методика оценки эффективности аутентификации голосовыми биометрическими системами состоит из следующих основных этапов:

1. Планирование испытаний.
2. Оценка показателей аутентификации в нормальных условиях.

3. Оценка устойчивости к атакам на модуль ввода биометрической информации.

Ниже описаны рекомендации по выполнению каждого этапа оценки эффективности голосовых биометрических систем.

3.1. Планирование испытаний

Численная оценка эффективности аутентификации голосовыми биометрическими системами является сложным и многогранным процессом. Его проведение без тщательной подготовки работ и их планирования может привести к потере трудозатрат, календарного времени, а главное искажению результата. Предложенные основы планирования оценки эффективности аутентификации голосовыми биометрическими системами, соответствуют требованиям раздела 6 ГОСТ Р ИСО/МЭК 19795-1-2007.

3.1.1. Определение информации о системе

На первом этапе исследования необходимо определить:

- тип испытуемых систем, особенности их применения и условия испытания;
- эксплуатационные характеристики оцениваемых голосовых биометрических систем;
- необходимые для оценки эксплуатационных характеристик данные.

Информация о необходимых данных является основой для определения вида испытаний: технологического, сценарного, оперативного или в режиме реального времени. На основе этого разрабатывается соответствующий протокол испытания, определяются необходимые средства контроля условий испытаний, а также выборка испытуемых субъектов и объем испытаний.

Учитывая, что множество голосовых биометрических систем значительно больше множества применяемых в них SDK с алгоритмами распознавания диктора по голосу, в большинстве случаев можно ограничиться проведением технологических испытаний самих SDK. Это позволит существенно сократить трудозатраты и время, необходимое для проведения испытаний.

Заметим также, что высокая надёжность SDK не гарантирует отсутствия ошибок при его встраивании в голосовую биометрическую систему. Действительно, при наличии подобных ошибок, надёжность голосовой биометрической системы может быть в разы ниже надёжности используемого в ней SDK, что негативно скажется на эффективности аутентификации. Таким образом, необходимо в обязательном порядке проводить проверку корректности встраивания алгоритмов в конечную систему.

При планировании процедур сбора данных необходимо дать ответы на следующие вопросы о голосовой биометрической системе, подвергаемой испытанию:

а) Регистрирует ли система информацию о транзакциях.

Если нет, то данная информация должна быть записана самим испытуемым субъектом, оператором или наблюдателем за испытанием.

б) Сохраняет ли система записи или признаки образцов для каждой транзакции.

Это необходимо, если степень схожести определяется в режиме отложенного задания.

в) Выдает ли система информацию о степени схожести или только решения о допуске или недопуске.

Если доступна информация о степени схожести, то по каким параметрам. В противном случае данные могут быть собраны с различными параметрами настройки безопасности, которые также необходимо зафиксировать.

г) Предоставил ли изготовитель комплект программного обеспечения.

Создание степеней схожести для подлинных лиц и «самозванцев» в режиме отложенного задания потребует использования программных модулей:

- для создания зарегистрированных шаблонов от зарегистрированных образцов;
- для извлечения признаков образца из испытуемых образцов;
- для определения степеней схожести признаков образцов и шаблонов.

Степени схожести, определяемые программами в режиме отложенного задания, должны быть такими же, которые созданы активной системой. Для этого может потребоваться регулирование параметров.

д) Требует ли система проведения модификаций для испытания. Требуемые модификации могут влиять на эксплуатационные характеристики системы.

е) Создает ли система независимые шаблоны.

Если шаблоны зависимы, процедуры для сбора или создания транзакций «самозванцев» будут различными.

ж) Использует ли система алгоритмы адаптации, которые адаптируют шаблон после успешной верификации.

В этом случае необходимо исходить из количества адаптаций шаблона, которые должны произойти до измерения эксплуатационных характеристик,

и возможности неблагоприятного влияния испытания «самозванца» на шаблоны.

и) Каковы рекомендуемое качество биометрических образцов и пороги принятия решений для целевого применения. Данные параметры настройки влияют на качество представленных образцов и вероятности ошибок.

к) Известны ли ожидаемые вероятности ошибок. Данную информацию используют при проверке достаточности объема испытания.

л) Какие факторы будут влиять на эксплуатационные характеристики для этого типа системы. Они должны быть контролируемыми.

м) Зависят ли эксплуатационные характеристики от размера регистрационной базы данных.

Данная зависимость существует у большинства систем идентификации и у некоторых систем верификации, которые выполняют регистрацию группы или осуществляют поиск «один ко многим» во время процесса верификации.

Строгий контроль качества может привести к уменьшению числа ложных соответствий и ложных несоответствий, но к увеличению ВОСД.

Если результаты сравнения предоставляют пользователю, то порог принятия решения также должен быть установлен соответствующим образом, так как положительная или отрицательная обратная связь влияет на поведение пользователя.

Оптимальные условия и компромиссные параметры настройки могут быть рекомендованы изготовителем.

3.1.2. Подготовка тестовой речевой базы данных

При проведении технологических испытаний необходимо очень внимательно отнестись к подготовке тестовой речевой базе данных. Она

должна содержать записи достаточного числа дикторов, на достаточном числе каналов и при разных фоновых условиях записи.

Объем числа дикторов и числа попыток в тестовой речевой базе данных влияет на точность измерения вероятностей ошибок. Чем больше проводят попыток, тем выше точность результатов. Для уменьшения числа попыток, необходимых для конкретного уровня точности, применяют правило трех или правило тридцати [30].

Данные правила являются очень оптимистичными, поскольку предполагают, что вероятности ошибок являются следствием одного источника вариативности, что неверно в случае биометрических систем. Десять испытуемых зарегистрированных пар образцов от каждого из 100 человек статистически не эквивалентны одной испытуемой зарегистрированной паре образцов от каждого из 1000 человек и не обеспечивают ту же доверительную вероятность результатов.

По мере увеличения объема испытания дисперсия оценки уменьшается, но масштабный коэффициент зависит от источника дисперсии. Например, пользователи могут иметь отличающиеся вероятности ошибки [45], дающие компоненту дисперсии, которую вычисляют как единицу, разделенную на число испытуемых субъектов, вместо единицы, разделенной на число попыток.

Если стоимость и усилия по сбору данных от большого числа субъектов при подготовке тестовой речевой базы данных не являются важным условием, то для обеспечения независимости транзакций используют большое число испытуемых субъектов, каждый из которых создает отдельную транзакцию.

Между несколькими транзакциями существует некоторая корреляция, но часто при использовании нескольких транзакций от меньшего числа испытуемых субъектов возникает меньшая неопределенность результатов

испытания, чем при испытании аналогичного объема, использующем одну транзакцию от большего числа испытуемых субъектов.

При определении точности результатов испытания число испытуемых субъектов имеет большее значение, чем число проводимых попыток.

Количество дикторов должно быть настолько большим, насколько это практически осуществимо. Мерой целесообразности могут быть затраты на сбор базы, трудозатраты на её обработку и скорость обработки данных оцениваемой биометрической системой.

От каждого испытуемого диктора должно быть получено достаточное число образцов, чтобы общее число попыток превысило значение, требуемое по правилу трёх или правилу 30 соответственно.

После того, как данные будут собраны и проанализированы, должна быть проведена оценка неопределенности измерений рабочих характеристик, а также достаточности объема проведенных испытаний.

В этом случае применяют закон убывающей отдачи, т.е. цель достигнута, если число ошибок, возникающих из-за изменений условий использования или выборки испытуемых субъектов, превышает число ошибок, возникающих вследствие объема группы и числа опытов.

Принципиальным является требование того, чтобы записи из тестовой речевой базы не входили в число обучающих, использованных при обучении алгоритмов распознания диктора, применяемых в оцениваемой голосовой биометрической системе. В случае несоблюдения данного требования, будет получено ложное улучшение показателей эффективности аутентификации системой.

Речь дикторов для тестовой речевой базы данных должна быть собрана на различных микрофонах или каналах (например, различные микрофоны в телефонных трубках и/или различные линии), для получения полной

картины эффективности аутентификации голосовой биометрической системой. Кроме того, при формировании тестовой речевой базы данных очень важно учесть разделение дикторов по полу и языку. Например, если априори, известно, что верифицируемая речь принадлежит только дикторам-мужчинам, то хороший результат аутентификации на дикторах женщинах будет искажать картину.

К числу других факторов, влияющих на эксплуатационные характеристики, относятся:

- Возраст. Дети (характеризующиеся быстрым ростом) и пожилые люди (у которых биометрические характеристики могут долго восстанавливаться после повреждения) имеют тенденцию к большему количеству ложных несоответствий и отказов в сборе данных, чем в среднем по выборке.
- Социальное положение, пол и род занятий. Качество биометрической характеристики человека (для конкретной биометрической системы) может зависеть от социального положения, пола и рода занятий человека. Биометрическая система, настроенная на определенную целевую выборку, может работать хуже, если будет использоваться для других сочетаний социальных слоев или полов.
- Время между регистрацией и верификацией. Старение шаблона, то есть изменение биометрического образца пользователя, и метода представления данных будет происходить в период времени между созданием регистрируемого шаблона и попыткой верификации или идентификации.
- Время суток. Поведение и физиологические особенности пользователя могут изменяться в течение суток.

- Осведомленность пользователя. Когда пользователи понимают принципы работы биометрической системы, они более подготовлены к самостоятельному правильному взаимодействию с ней, а также могут самостоятельно решить многие из возможных проблем, возникающих при верификации.
- Мотивация пользователя. Пользователи будут действовать по-разному, в зависимости от значимости биометрической транзакции.
- Изменение состояния здоровья может происходить быстрее, чем нормальные эффекты старения.
- Диалект, акцент и родной язык.
- Выражение, интонация и громкость голоса.

Важно контролировать, чтобы тестовые данные были сбалансированы по различным длинам произнесения и прочим параметрам. Например, при использовании гендеронезависимых данных, нужно убедиться, что соблюдается баланс мужской и женской речи. Иначе полученный результат оценки эффективности будет смещен в сторону доминирующих мужских или женских фонограмм речи. Этот же аргумент справедлив и для других разбиений на типы фонограмм, например, речь в базе должна быть одинаково распределена по типу используемого микрофона (угольному, электретному) телефонной трубки.

Для упрощения дальнейшего анализа и расчёта численных показателей рекомендуется использовать следующее соглашение об именовании файлов фонограмм:

MXXXXST(SS)_r.wav

где M - идентификатор пола диктора (M - мужской пол, F - женский пол),

XXXX - идентификатор диктора,

ST - указание на то, что фонограмма содержит фальсифицированную речь диктора с использованием метода фальсификации номер T,

SS - номер сессии фонограммы,

r - идентификатор канала для разделённой по каналам стерео записи (r - правый канал, l - левый канал).

3.2. Оценка фундаментальных показателей эффективности

В данной методике, на втором её этапе, предлагается опираться на следующие действующие стандарты в области оценки эффективности аутентификации голосовыми биометрическими системами:

- ГОСТ Р ИСО/МЭК 19795-1-2007 “Автоматическая идентификация. Идентификация биометрическая. Эксплуатационные испытания и протоколы испытаний в биометрии. Часть 1. Принципы и структура”.
- ГОСТ Р ИСО/МЭК 19795-2-2008 “Автоматическая идентификация. Идентификация биометрическая. Эксплуатационные испытания и протоколы испытаний в биометрии. Часть 2. Методы проведения технологического и сценарного испытаний”.
- ГОСТ Р ИСО/МЭК ТО 19795-3-2009 “Автоматическая идентификация. Идентификация биометрическая. Эксплуатационные испытания и протоколы испытаний в биометрии. Часть 3. Особенности проведения испытаний при различных биометрических модальностях”.
- ГОСТ Р ИСО/МЭК ТО 19795-4-2010 “Автоматическая идентификация. Идентификация биометрическая.

Эксплуатационные испытания и протоколы испытаний в биометрии. Часть 4. Тестирование производительности биометрических систем”.

– ГОСТ Р ИСО 9000-2011 “Системы менеджмента качества. Основные положения и словарь”.

После подготовки тестовой речевой базы, в соответствии с ГОСТ Р ИСО/МЭК 19795-1-2007 “Автоматическая идентификация. Идентификация биометрическая. Эксплуатационные испытания и протоколы испытаний в биометрии. Часть 1. Принципы и структура” проводятся необходимые эксперименты и рассчитываются фундаментальные численные показатели аутентификации голосовой биометрической системой, не зависящие от её типа. Иными словами, общие эксплуатационные характеристики систем, решающих задачи идентификации на открытом или закрытом множестве, а также текстонезависимой или текстозависимой верификации. Перечислим данные показатели ниже.

3.2.1. Вероятность отказа регистрации

ВОР (FTE) – это доля выборки, для которой система не может закончить процесс регистрации.

ВОР включает в себя дикторов, которые:

– не могут предоставить запись голоса;

– не могут предоставить образец голоса с достаточным качеством;

– не могут получить результат оценки схожести со своим заново созданным шаблоном при процессе регистрации.

При технологическом испытании анализ основан на предварительно подготовленной тестовой речевой базе данных. Не смотря на это, даже в этом случае может произойти сбой в регистрации, например, в ситуации, когда

качество записи речевого образца имеет столь низкое значение, что извлечь из него необходимые признаки становится невозможным.

ВОР для целевой выборки следует определять как долю (или весовую долю) людей в испытуемой группе, которые не смогли зарегистрироваться в процессе регистрации.

$$\text{BOP}(\tau) = \frac{N_{fail}(\tau)}{N_{total}} \cdot 100\%$$

где N_{total} - общее количество попыток регистрации,

$N_{fail}(\tau)$ - количество не успешных попыток регистрации, в зависимости от порога.

ВОР зависит от политики регистрации, которая определяет уровень качества образца для регистрации, порог принятия решения для подтверждения применимости регистрации, а также число попыток или время, отведенное на регистрацию при транзакции регистрации. Политика регистрации должна быть описана наряду с наблюдаемой ВОР.

Более строгие требования к качеству регистрации увеличивают ВОР, но улучшают эксплуатационные характеристики сравнений в голосовой биометрической системе.

3.2.2. Вероятность отказа сбора данных

ВОСД (FTA) – это доля попыток верификации или идентификации, для которых биометрическая система не может получить или отобрать образец удовлетворительного качества.

ВОСД должна включать в себя:

- попытки, при которых голосовая биометрическая характеристика не может быть получена (например, из-за физического состояния диктора);

- попытки, при которых не удается произвести сегментацию или извлечение необходимых признаков;
- попытки, при которых извлеченные признаки не подходят по порогу проверки качества.

ВОСД можно определить для каждой транзакции, например, путем определения числа транзакций, в процессе которых ни при одной из попыток регистрации не был получен образец удовлетворительного качества для сравнения.

При технологическом испытании анализ основан на предварительно собранной базе данных.

ВОСД следует определять как долю (или весовую долю) записанных попыток подлинного лица (и, по возможности, любых пассивных попыток «самозванца» в режиме реального времени), которые не могут быть закончены из-за отказов в представлении (изображение не получено), сегментации, извлечении признаков или контроля качества.

$$\text{ВОСД}(\tau) = \frac{N_{fail}(\tau)}{N_{total}} \cdot 100\%$$

где N_{total} - общее количество попыток верификации или идентификации,

$N_{fail}(\tau)$ - количество не успешных попыток идентификации или верификации, в зависимости от порога.

ВОСД зависит от порога качества образца, а также от времени, установленного на получение образца, или допускаемого числа представлений.

Данные параметры настройки должны быть указаны в протоколе испытаний вместе со значением ВОСД.

Более строгий уровень качества для получения образцов увеличивает ВОСД, но улучшает эксплуатационные характеристики сравнений в голосовой биометрической системе.

Попытки, при которых исходный образец не был получен или не имеет удовлетворительного качества, не обрабатываются алгоритмом сравнения, а степень схожести не определяется. Такие отказы сбора данных должны быть исключены при вычислении ВЛС и ВЛНС, но должны быть включены в вычисление ВЛД и ВЛНД.

ВОСД, ВЛС и ВЛНС должны быть вычислены при одних и тех же значениях порога принятия решения.

3.2.3. Вероятность ложного несовпадения

ВЛНС (FNMR) – это доля образцов, полученных в результате попыток подлинного лица, которые ошибочно признаны несовпадающими с шаблоном зарегистрированного в системе пользователя.

ВЛНС следует определять как долю (или весовую долю) зафиксированных попыток подлинного лица, которые были переданы подсистеме сравнения, и для которых степень схожести была ниже соответствующего порога принятия решения о схожести.

$$\text{ВЛНС}(\theta) = \frac{N_{imposter\ (error\ )}(\theta)}{N_{target}} \cdot 100\%$$

где N_{target} - количество сравнений вида "свой-свой",

$N_{imposter\ (error\ )}(\theta)$ - количество сравнений вида "свой-свой", идентифицированных как "свой-чужой", в зависимости от порога.

ВЛНС зависит от порога принятия решения о схожести и должна быть указана вместе с наблюдаемой ВЛС при том же пороге принятия решения.

3.2.4. Вероятность ложного совпадения

ВЛС (FMR) – это доля образцов, полученных в результате пассивных попыток «самозванца», которые ошибочно признаны совпадающими с шаблоном зарегистрированного пользователя.

При пассивных попытках «самозванца» пользователи предоставляют свою собственную биометрическую характеристику, как будто они совершают попытку успешной верификации с собственным шаблоном. Например, в случае динамической верификации подписи «самозванец» при пассивной попытке поставил бы свою собственную подпись.

ВЛС следует определять как долю (или весовую долю) зафиксированных пассивных попыток «самозванца», которые были переданы подсистеме сравнения и для которых степень схожести не ниже соответствующего порога принятия решения о схожести.

$$\text{ВЛС}(\theta) = \frac{N_{target\ (error\ )}(\theta)}{N_{imposter}} \cdot 100\%$$

где $N_{imposter}$ - количество сравнений вида "свой-чужой";

$N_{target\ (error\ )}(\theta)$ - количество сравнений вида "свой-чужой", идентифицированных как "свой-свой", в зависимости от порога.

ВЛС зависит от порога принятия решения о схожести и должна быть указана наряду с наблюдаемой ВЛНС при том же пороге принятия решения.

3.2.5. Равновероятная ошибка

РВО (EER) это доля образцов, полученных в результате пассивных попыток «самозванца», которые ошибочно признаны совпадающими с шаблоном зарегистрированного пользователя, совпадающая с долей образцов, полученных в результате попыток подлинного лица, которые

ошибочно признаны несовпадающими с шаблоном зарегистрированного в системе пользователя.

РВО следует определять как долю (или весовую долю) зафиксированных пассивных попыток «самозванца», которые были переданы подсистеме сравнения и для которых степень схожести не ниже такого значения порога принятия решения о схожести, при котором значение РВО совпадает с долей (или весовой долей) зафиксированных попыток подлинного лица, которые были переданы подсистеме сравнения, и для которых степень схожести была ниже выбранного значения порога принятия решения о схожести.

$$\text{РВО} = \text{ВЛС}(\theta_{\text{РВО}}) = \text{ВЛНС}(\theta_{\text{РВО}})$$

где $\theta_{\text{РВО}}$ - порог принятия решения при котором значение ВЛС равно значению ВЛНС.

В ситуациях, когда количество измерений не позволяет подобрать такое значение порога принятия решения $\theta_{\text{РВО}}$, при котором значение ВЛС совпадает с ВЛНС, РВО рассчитывается следующим образом:

$$\text{РВО} = \sqrt{\frac{1}{4}\left(\left(\max_{\text{ВЛНС}(\theta)<\text{ВЛС}(\theta)} \text{ВЛНС}(\theta)\right)^2 + \left(\min_{\text{ВЛНС}(\theta)<\text{ВЛС}(\theta)} \text{ВЛС}(\theta)\right)^2 + \left(\max_{\text{ВЛНС}(\theta)\geq\text{ВЛС}(\theta)} \text{ВЛС}(\theta)\right)^2 + \left(\min_{\text{ВЛНС}(\theta)\geq\text{ВЛС}(\theta)} \text{ВЛНС}(\theta)\right)^2\right)}$$

где θ - порог принятия решения.

3.2.6. Кривая компромиссного определения ошибки и кривая рабочей характеристики

Измерения КОО (DET) должны быть проведены с использованием степеней схожести подлинного лица и «самозванца», полученных при сравнении одного образца одной попытки с одним зарегистрированным шаблоном.

По результату каждой попытки определяют степень схожести. Степени схожести, определенные по попыткам подлинного лица, должны быть упорядочены. Аналогично определяют степени схожести «самозванца».

Выбросы должны быть изучены для определения ошибок, возникающих при маркировке. Изъятие любых степеней схожести из результатов испытания должно быть отображено в документах и обеспечивать проведение внешней оценки испытания.

В основе построения кривых КОО (или РХ) лежит накопление упорядоченных степеней схожести подлинного лица и «самозванца». Поскольку степени схожести могут иметь любые значения, кривые КОО (или РХ) строят параметрически: каждая точка (x, y) соответствует ВЛС и ВЛНС, а степень схожести является варьируемым параметром.

Кривые строят, отображая ВЛС по абсциссе и ВЛНС по ординате. Оси могут иметь логарифмические масштабы. Пример кривой КОО отображён на рисунке 3.1.

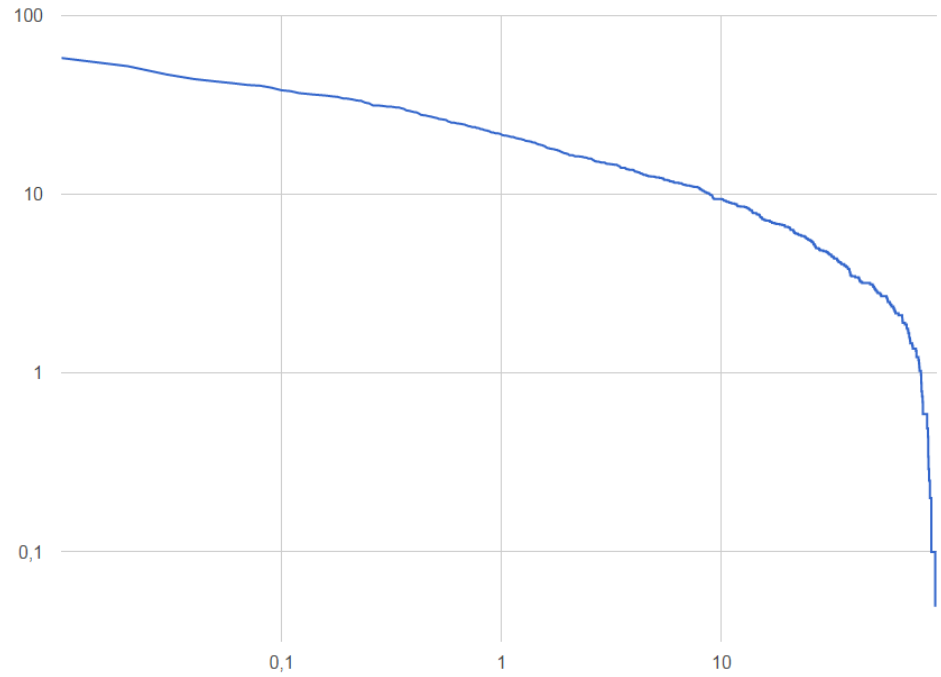

Рисунок 3.1 - Пример кривой КОО для PLDA системы идентификации диктора, полученный на базе телефонных переговоров

Рекомендуемый алгоритм эффективного получения данных для кривых КОО/РХ приведен в ГОСТ Р ИСО/МЭК 19795-1–2007, приложение F1.

Данный алгоритм имеет следующий вид:

а) сортировка значений степеней схожести подлинного лица в порядке возрастания: $s_1 < s_2 < s_3 < \ldots < s_k$;

б) вычисление частоты, с которой берется каждая степень схожести подлинного лица: $g_1, g_2, g_3, \ldots g_k$;

в) вычисление числа попаданий степеней схожести «самозванца» в каждый интервал: $(-\infty, s_1), (s_1, s_2), (s_2, s_3) \ldots (s_k, \infty)$: $h_0, h_1, h_2, \ldots h_k$;

г) для каждого значения степени схожести подлинного лица s_j по очереди:

1) расчет суммы степеней схожести «самозванцев», больше или равных s_j:

$$\sum_{i=j}^{k} h_i$$

2) деление на общее число попыток «самозванцев», определяющее ВЛС для данного порога степени схожести;

3) вычисление суммы степеней схожести подлинного лица, меньших s_j:

$$\sum_{t=1}^{j-1} g_i$$

4) деление на общее число попыток подлинного лица, определяющее ВЛНС для данного порога степени схожести.

Кривые КОО (или РХ) могут также использоваться для построения зависимости ВЛД от ВЛНД.

Транзакции нескольких попыток могут потребовать создания новой транзакционной степени схожести, основанной на степенях схожести части

попыток (например, максимальное значение степени схожести для лучшей из трех попыток, определенных политикой принятия решения).

3.2.7. График зависимости ВЛС и ВЛНС от порога

Основным недостатком кривых КОО (или РХ) является отсутствие информации о пороге принятия решения. Для наглядного отображения его влияния на значения ВЛС и ВЛНС рекомендуется строить кривые зависимости ВЛС и ВЛНС от порога принятия решения.

Кривые строят, отображая значения ВЛС и ВЛНС по абсциссе и соответствующее значение порога принятия решения по ординате. Пример графика зависимости ВЛС и ВЛНС от порога отображён на рисунке 3.2.

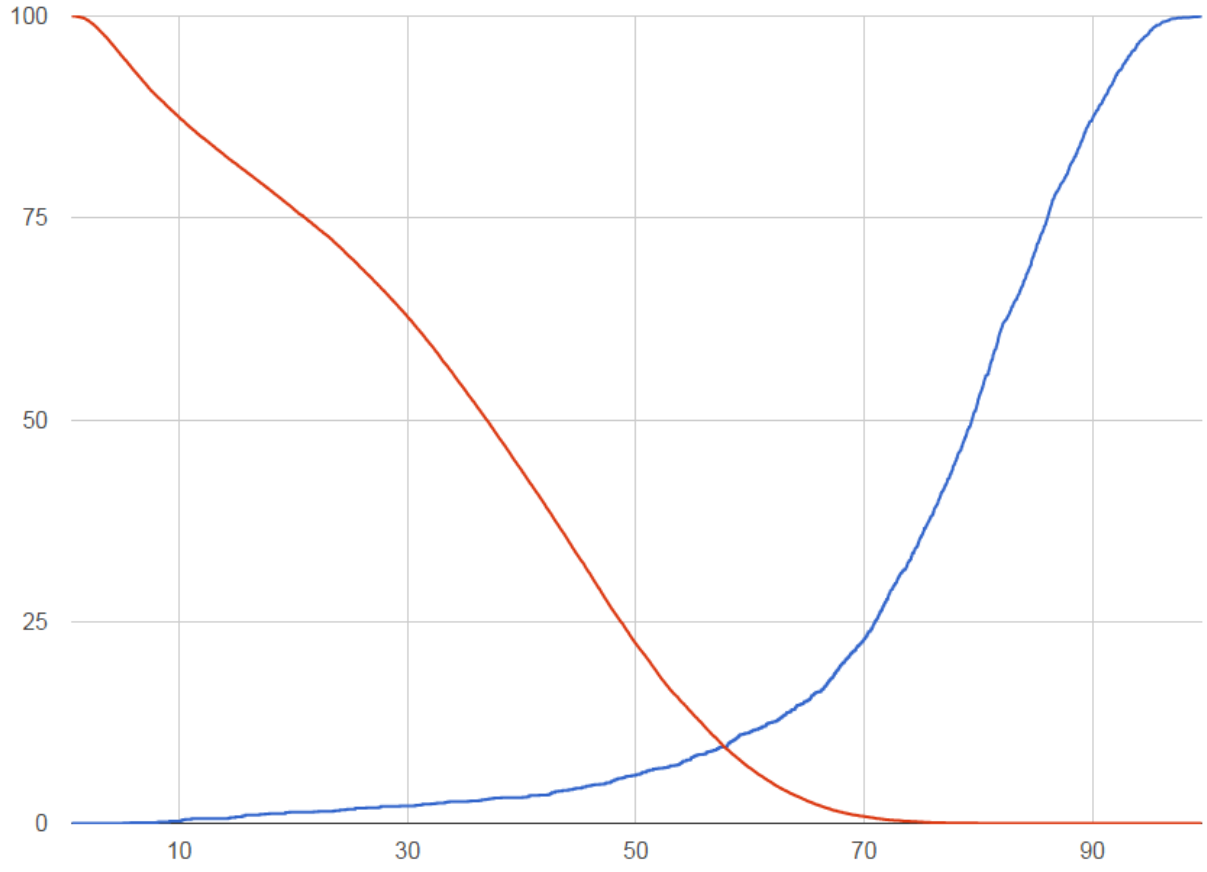

Рисунок 3.2 - Пример графика зависимости ВЛС и ВЛНС от порога для PLDA системы идентификации диктора, полученный на базе телефонных переговоров

Рекомендуемый алгоритм эффективного получения данных для графика зависимости ВЛС и ВЛНС от порога имеет следующий вид:

а) сортировка значений степеней схожести подлинного лица s_k и степеней схожести «самозванцев» f_i в порядке возрастания: $f_1 < f_2 < s_1 < \cdots < f_i < \cdots < s_k$;

б) вычисление шага для изменения ВЛНС: $\Delta\text{ВЛНС} = \frac{100}{k}$;

в) вычисление шага для изменения ВЛС: $\Delta\text{ВЛС} = \frac{100}{i}$;

г) для каждого значения степени схожести подлинного лица или «самозванца» θ_j по очереди:

1) расчет значения ВЛНС при пороге принятия решения θ_j:

$$\text{ВЛНС}(\theta_j) = \begin{cases} \text{ВЛНС}(\theta_{j-1}) + \Delta\text{ВЛНС}, & \theta_j \in \{s_1, \dots, s_k\} \\ \text{ВЛНС}(\theta_{j-1}), & \theta_j \in \{f_1, \dots, f_i\} \\ 0, & j = 0 \end{cases}$$

2) расчет значения ВЛС при пороге принятия решения θ_j:

$$\text{ВЛС}(\theta_j) = \begin{cases} \text{ВЛС}(\theta_{j-1}), & \theta_j \in \{s_1, \dots, s_k\} \\ \text{ВЛС}(\theta_{j-1}) - \Delta\text{ВЛС}, & \theta_j \in \{f_1, \dots, f_i\} \\ 100, & j = 0 \end{cases}$$

График зависимости ВЛС и ВЛНС от порога может также строиться для построения зависимости ВЛД и ВЛНД от порога.

3.3. Оценка показателей для систем верификации

При использовании в аутентификации голосовых биометрических систем, решающих задачу верификации, необходимо также рассчитать следующие эксплуатационные характеристики систем.

3.3.1. Вероятность ложного недопуска

ВЛНД (FRR) – это доля транзакций верификации подлинного лица, которые были ошибочно отвергнуты. В зависимости от политики принятия

решения транзакция может состоять из одной или более попыток подлинного лица.

ВЛНД следует определять как долю (или весовую долю) записанных транзакций подлинного лица, которые были ошибочно отвергнуты. Сюда также входят транзакции, отвергнутые из-за отказа сбора данных и ошибок соответствия.

Например, если транзакция верификации состоит из единственной попытки, то отказ сбора данных или ложное несоответствие вызовут ложный недопуск, и ВЛНД будет равна сумме ВОСД с произведением ВЛНС и значения обратного ВОСД

ВЛНД зависит от политики принятия решения, порога принятия решения о схожести и качества образца, в связи с этим ВЛНД должна быть указана в протоколе вместе с данными параметрами с оценкой ВЛД для тех же параметров.

3.3.2. Вероятность ложного допуска

ВЛД (FAR) – это доля транзакций верификации «самозванца», которые могут быть ошибочно приняты. Транзакция в зависимости от политики принятия решения может состоять из одной или более попыток «самозванца».

ВЛД следует определять как долю (или весовую долю) записанных пассивных транзакций «самозванца»,которые были ошибочно приняты.

Например, если транзакция верификации состоит из единственной попытки, то для ложного допуска необходимо, чтобы представленный образец не был отклонен при проверке качества (то есть не должно происходить отказа сбора данных) и произошла ошибка соответствия. Соответственно ВЛД будет равна произведению ВЛС и значения обратного ВОСД

ВЛД зависит от политики принятия решения, порога принятия решения о схожести и порога качества образца. Следовательно, ВЛД должна быть указана в протоколе вместе с данными параметрами с оценкой ВЛНД для тех же параметров.

3.4. Оценка показателей для систем идентификации

3.4.1. Вероятность истинно положительной идентификации

Вероятность истинно положительной идентификации (TPIR) на замкнутом множестве – это вероятность того, что транзакция зарегистрированного в системе пользователя включает в себя истинный идентификатор этого пользователя в N процентной доле первых возвращаемых соответствий от всех возвращаемых соответствий. При использовании в качестве аргумента, вместо процентной доли, количество первых возвращаемых соответствий следует ссылаться непосредственно на размер базы данных.

Определение эксплуатационных характеристик идентификации на замкнутом множестве следует оценивать при $N = 1\%$, а также изображать в виде кривой XCC (CMC), для которой вероятность истинно положительной идентификации представляет собой функцию от размера процентной доли N.

Рекомендуемый алгоритм эффективного получения данных для кривой XCC основан на представленном в ГОСТ Р ИСО/МЭК 19795-1–2007, приложение F2.

При предположении, что каждому человеку соответствует один шаблон биометрических данных, алгоритм получения данных для кривой XCC имеет следующий вид:

а) определение ранга идентификации для каждой попытки проводят следующим образом:

1) поиск степени подобия подлинного лица для данной попытки;

2) подсчет числа степеней подобия для данной попытки (в сравнении с несобственными шаблонами и собственным шаблоном), которые:

– больше степени схожести подлинного лица: x,

– равны степени схожести подлинного лица: y;

3) если при $y = 1$ попытка имеет ранг идентификации $x + 1$, то во всех других случаях ранг определяется диапазоном значений $(x + 1), \dots, (x + y)$;

б) для каждого ранга r:

1) вычисляют число попыток с рангом r и меньше. Попытки, которые имеют диапазон рангов, считаются как сумма рангов из диапазона, не превышающих r;

2) деление на общее число попыток увеличивает вероятность того, что для испытуемого образца среди r наиболее схожих шаблонов в базе данных зарегистрированных шаблонов будет найден правильный шаблон или модель. Эту вероятность строят в виде кривой на графике ХСС в зависимости от соответствующей рангу r процентной доли N.

3.5. Оценка показателей устойчивости к спуфинг атакам

На данном этапе методики производится исследование устойчивости к различным методам спуфинг атак на модуль ввода биометрической информации.

При выборе методов имитации спуфинг атак необходимо учитывать информацию о тестируемой голосовой биометрической системе, полученную на первом этапе. Действительно, в зависимости от сценария использования системы или наличия организационных решений по повышению уровня защиты, угроза от ряда методов спуфинг атак может оказаться минимальной, даже при отсутствии технических решений по противодействию данным

методам атак. Например, если голосовая биометрическая система установлена в колл-центре и выполняет задачу верификации диктора в фоновом режиме, во время его общения с оператором колл-центра. В этом случае методы фальсификации индивидуальных биометрических характеристик диктора, приводящие к низкой разборчивости речи, необходимо исключить из испытаний, т.к. они вносят грубые ошибки в оценку надёжности системы.

Для оценки разборчивости фальсифицированной речи рекомендуется использовать метод парных сравнений, описанный в ГОСТ Р 50840-95 - "Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости".

Оценка производится путём прослушивания не менее 35 пар одинаковых фраз длительностью не более трёх секунд, содержащих не менее пяти дикторов. Каждая пара содержит "контрольный" образец, содержащий речь "целевого" диктора, а также "испытуемый" образец, содержащий фальсифицированную речь того же диктора. Для каждой пары выставляется оценка по пяти бальной шкале с точностью до 0,1.

Полученные результаты суммируются и нормируются по формуле:

$$X = \frac{\sum_{i=1}^{N} X_{\text{и}i}}{\sum_{i=1}^{N} X_{\text{к}i}} \cdot 5$$

где $X_{\text{и}i}$ - результат единичного измерения испытуемого образца,

$X_{\text{к}i}$ - результат единичного измерения контрольного образца,

N - количество единичных измерений.

В случае, если значение итоговой оценки составит менее 3,6 оцениваемый метод фальсификации не рекомендуется применять при тестировании голосовой биометрической системы, сценарий которой, подразумевает прослушивание речи пользователя оператором системы.

При проведении технологических испытаний рекомендуется использовать заранее подготовленные речевые базы данных, содержащие натуральную и фальсифицированную речь одних дикторов. Такие базы можно найти в работах [78, 92]. При подготовке и использовании речевых баз данных необходимо придерживаться общих методических рекомендаций для подготовки тестовых речевых баз, данных в описании первого этапа методики.

В дополнение к данным ранее рекомендациям, необходимо также учитывать сбалансированность данных по методам фальсификации индивидуальных биометрических характеристик диктора.

Важно отметить, что все необходимые эксперименты и расчеты численных показателей устойчивости голосовой биометрической системы к спуфинг атакам, необходимо проводить для всей системы в целом, а не для модуля детектирования спуфинга атак (при наличии возможности его тестирования в отдельности).

Перечислим численные показателей устойчивости голосовой биометрической системы к спуфинг атакам на модуль ввода биометрической информации.

3.5.1. Вероятность ложного совпадения фальсифицированного образца

ВЛСФ – это доля образцов, полученных в результате активных попыток «самозванца», применяющего методы фальсификации индивидуальных биометрических характеристик, которые ошибочно признаны совпадающими с шаблоном зарегистрированного пользователя.

При активных попытках «самозванца» нарушители предоставляют фальсифицированную биометрическую характеристику зарегистрированного

пользователя, как будто они совершают попытку успешной верификации с собственным шаблоном.

ВЛСФ следует определять как долю (или весовую долю) зафиксированных активных попыток «самозванца», применяющего методы фальсификации индивидуальных биометрических характеристик, которые были переданы подсистеме сравнения и для которых степень схожести не ниже соответствующего порога принятия решения о схожести.

$$\text{ВЛСФ}(\theta) = \frac{N_{target\ (error\ )}(\theta)}{N_{spoofing}} \cdot 100\%$$

где $N_{spoofing}$ - количество сравнений вида "свой-чужой" при применении методов фальсификации БХЧ,

$N_{target\ (error\ )}(\theta)$ - количество сравнений вида "свой-чужой" при применении методов фальсификации БХЧ, идентифицированных как "свой-свой", в зависимости от порога.

ВЛСФ зависит от порога принятия решения о схожести и должна быть указана наряду с наблюдаемой ВЛНС при том же пороге принятия решения.

В качестве наглядных численных показателей необходимо указывать следующие значения:

- ВЛСФ($\theta_{\text{РВО}}$) - значение ВЛСФ, полученное при пороге, соответствующем точке РВО, полученной на предыдущем этапе методики. Так как значение ВЛНС при воздействии спуфинг атаки остаётся неизменным, данный показатель наглядно демонстрирует влияние спуфинг атаки на общий уровень ошибки системы, откалиброванной на минимальное значение ВЛС и ВЛНС.
- ВЛСФ($\theta_{0,01}$) - значение ВЛСФ, полученное при минимальном пороге, в котором выполняется следующее уравнение:

$$\text{ВЛС}(\theta_{0,01}) \leq 0{,}01\%$$

В случае если известно фиксированное значение порога принятия решения о схожести, на которое предполагается настраивать систему после внедрения, необходимо также оценить ВЛСФ и ВЛС при данном значении порога принятия решения о схожести.

3.5.2. Кривая компромиссного определения ошибки

В случае оценки устойчивости к спуфинг атакам, измерения КОО (DET) должны быть проведены с использованием степеней схожести подлинного лица и активных попыток «самозванца», полученных при сравнении одного образца одной попытки с одним зарегистрированным шаблоном.

По результату каждой попытки определяют степень схожести. Степени схожести, определенные по попыткам подлинного лица, должны быть упорядочены. Аналогично определяются степени схожести активных попыток «самозванца», использующего методы фальсификации индивидуальных биометрических характеристик человека.

Выбросы должны быть изучены для определения ошибок, возникающих при маркировке. Изъятие любых степеней схожести из результатов испытания должно быть отражено в документации и обеспечивать проведение внешней оценки испытания.

Кривые строят, отображая ВЛСФ по абсциссе и ВЛНС по ординате. Оси могут иметь логарифмические масштабы. При сравнении полученных кривых с кривыми, полученными на предыдущем этапе методики, легко оценить степень устойчивости голосовой биометрической системы к спуфинг атакам. Чем ближе кривые друг к другу, тем надёжнее система работает под воздействием атаки. Пример кривых КОО под воздействием атаки и без её воздействия, построенных для двух различных систем, изображён на рисунке 3.3.

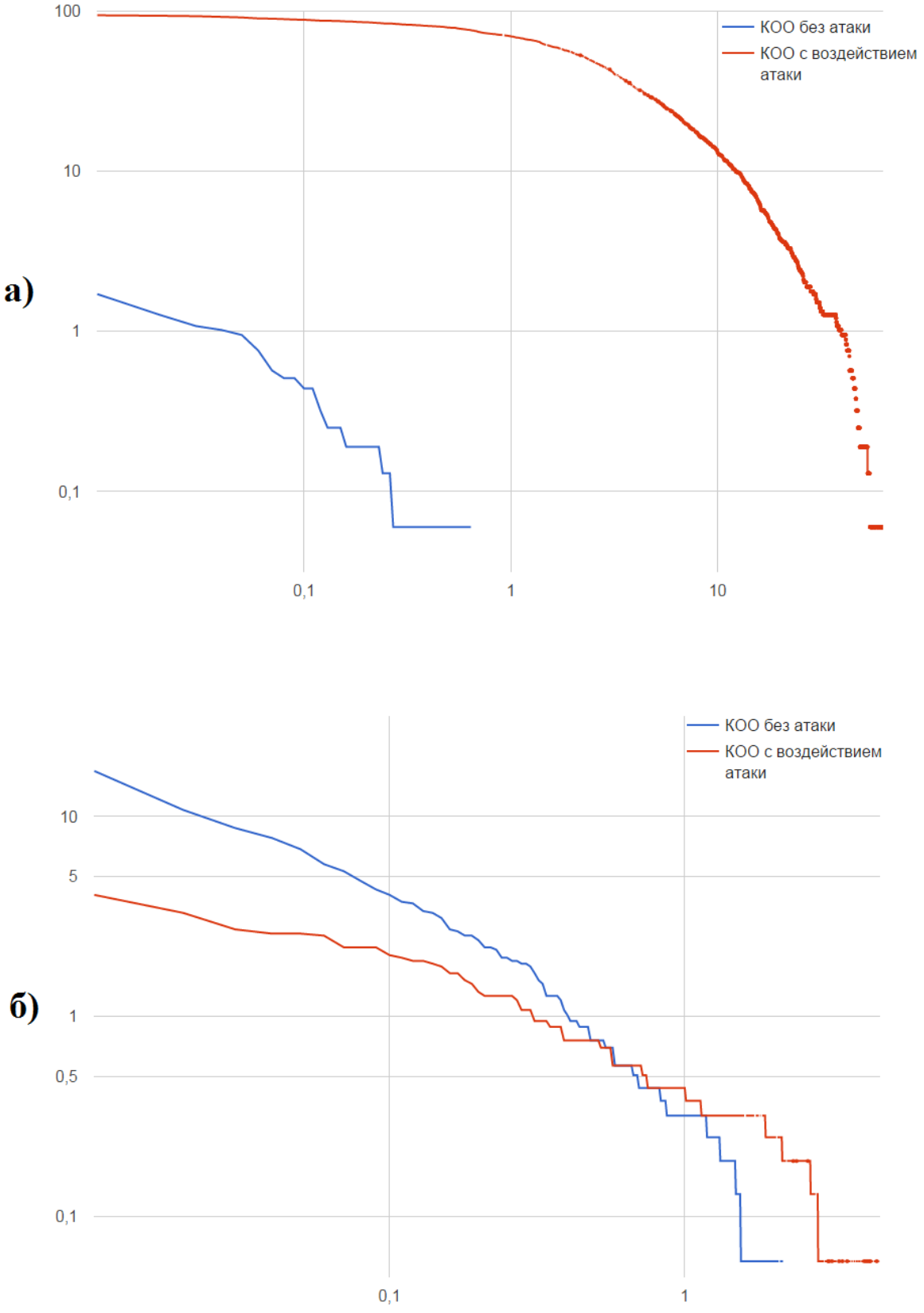

Рисунок 3.3 - Кривые КОО под воздействием спуфинг атак на модуль ввода биометрической информации и без воздействия, для систем: а) без детектора спуфинг атаки, б) с детектором спуфинг атаки

3.6. Комплекс программных средств оценки эффективности аутентификации ГБС

Нельзя не отметить возможность существенного снижение затрат на проведение оценки эффективности аутентификации голосовыми биометрическими системами при автоматизации данного процесса, особенно при проведении технологических испытаний.

Действительно, часть операций первого этапа методики может быть автоматизирована за счёт автоматизированного анализа тестовых данных, а на втором и третьем этапе возможно полная автоматизация процесса. Следовательно, трудозатраты на проведение оценки могут быть сведены к подготовке и настройке комплекса программ оценки эффективности, что особенно важно при регулярном проведении работ по оценке систем распознавания диктора по голосу.

Подобные работы необходимо проводить на регулярной основе как один из этапов при разработке голосовых биометрических систем, вместе с работами по функциональному, регрессионному и нагрузочному тестированию системы. Кроме этого, данные работы необходимо проводить как финальный этап при сдаче систем заказчику или, как предварительный этап при выборе системы перед её внедрением.

Таким образом, комплекс программных средств оценки эффективности аутентификации голосовыми биометрическими системами должен отвечать следующим требованиям:

1. Возможность сопряжения с различными голосовыми биометрическими системами при проведении технологических испытаний.

2. Возможность интеграции в инфраструктуру, применяемую при функциональном, регрессионном и нагрузочном тестировании систем.
3. Возможность указания пороговых значений для численных показателей эффективности. Такими значениями могут быть результаты предыдущей оценки эффективности.
4. Автоматическое уведомление оператора комплекса программных средств об обнаруженном не соответствии численных показателей и установленных порогов.
5. Возможность применения распределённых вычислений для экономии временных затрат.
6. Минимальные требования к составу дополнительных компонентов среды и библиотек для упрощения процесса установки комплекса и исключения возможности влияния сторонних модулей на работу тестируемой системы.
7. Автоматическая генерация протоколов испытаний и возможность настройки шаблона протокола.
8. Отсутствие зависимостей от типа, разрядности и версии операционных систем, включая мобильные ОС.

Учитывая описанные требования к комплексу программных средств, было принято решение использовать при его реализации языки программирования C++ и Python, т.к. они в большей степени позволяют добиться независимости от выбора операционной системы и уменьшить требования к составу дополнительных компонентов среды и библиотек.

В состав разработанного комплекса программных средств вошли следующие модули:

1. Модуль сопряжения с голосовой биометрической системой.

2. Модуль формирования протоколов по тестовым голосовым базам данных.
3. Модуль имитации спуфинг атак на голосовую биометрическую систему.
4. Модуль тестирования голосовой биометрической системы.
5. Модуль расчета показателей эффективности аутентификации голосовой биометрической системой.
6. Модуль генерации протоколов испытаний.

Общая схема комплекса программных средств оценки эффективности аутентификации голосовыми биометрическими системами представлена на рисунке 3.4.

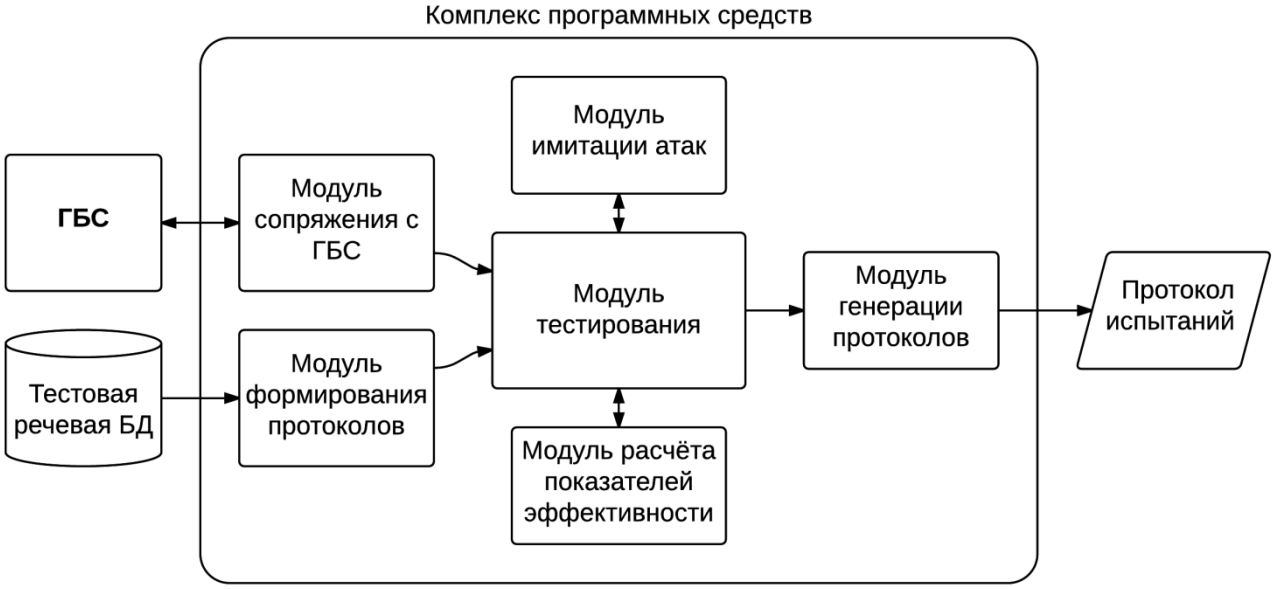

Рисунок 3.4 - Общая схема разработанного комплекса программных средств оценки эффективности аутентификации голосовыми биометрическими системами

Рассмотрим модули комплекса программных средств детальнее.

3.6.1. Модуль сопряжения с голосовой биометрической системой

Учитывая требование к возможности сопряжения комплекса программных средств оценки с различными голосовыми биометрическими системами при проведении технологических испытаний, а также

необходимость использования функций голосовой биометрической системы на разных этапах испытаний, становится очевидным, что функционал сопряжения с голосовой биометрической системой должен быть вынесен в отдельный модуль.

С архитектурной точки зрения, данный модуль должен содержать в себе набор классов, каждый из которых отвечает за сопряжение с соответствующим ему функционалом голосовой биометрической системы. К такому функционалу относятся:

- регистрация новых шаблонов в системе;
- создание нового образца голоса перед попыткой верификации или идентификации;
- сравнение зарегистрированного шаблона с образцом голоса и получение решения о степени сходства;
- и т.д.

Каждый из полученного набора классов наследуется от одного из двух базовых классов:

- Первый обеспечивает сопряжение с консольным процессом, управляя его потоками ввода и вывода. Это позволяет проводить испытание систем, у которых отсутствует графический интерфейс пользователя или SDK. Большинство производителей SDK комплектуют свой продукт примерами использования, обеспечивающими покрытие функционала SDK. В случаях, когда данные примеры не скомпилированы, их разработка не содержит больших трудозатрат при наличии грамотной документации или оперативной поддержи со стороны разработчиков. Отсутствие, как документации, так и возможности оперативной поддержки со

стороны разработчиков алгоритмов, ставит под сомнение целесообразность использования такого продукта.

– Второй обеспечивает сопряжение с графическим интерфейсом рабочего места оператора голосовой биометрической системы. Автоматизация процесса взаимодействия с графическим интерфейсом позволяет проводить сопряжение с голосовыми биометрическими системами, работающими в облаке, прямой доступ к которому не возможен. Этот способ также является одним из типовых, при предоставлении демонстрационного доступа к системе. Важно учитывать, что трудозатраты на сопряжение могут быть существенно снижены при наличии JSON, REST или аналогичного API по работе с облачной системой. В этом случае можно избежать настройки сопряжения с графическим интерфейсом пользователя.

Вся информация об особенностях сопряжения с тем или иным функционалом системы содержится в соответствующем ему классе модуля сопряжения. Общий функционал, не зависящий от типа функционала и версии системы, выносится в базовые классы сопряжения.

Таким образом, в остальных модулях комплекса достаточно вызывать унифицированный интерфейс класса сопряжения с необходимым функционалом, без необходимости его доработки при сопряжении с новой голосовой биометрической системой, т.к. все необходимые настройки проводятся в модуле сопряжения.

3.6.2. Модуль формирования протоколов по тестовым голосовым базам данных

Подготовленные на первом этапе методики тестовые речевые базы данных необходимо подавать в голосовую биометрическую систему в соответствии с протоколами. Такие протоколы могут содержать:

- указание на регистрацию шаблонов из определённых звуковых файлов (в том числе по несколько звуковых файлов для одного шаблона);
- указание на проведение адаптации указанного зарегистрированного шаблона с помощью указанных звуковых файлов;
- указание на проведение транзакции верификации или идентификации типа "свой-свой" с указанием звуковых файлов и зарегистрированных шаблонов;
- указание на проведение транзакции верификации или идентификации типа "свой-чужой" с указанием звуковых файлов и зарегистрированных шаблонов;
- и т.д.

Подготовка данных протоколов вручную является трудоёмкой задачей, но может быть автоматизирована при использовании соглашения об именовании речевых баз данных. Пример подобного соглашения был дан ранее, при описании процесса подготовки тестовых речевых баз данных. В этом случае модуль подготовки протоколов, опираясь на информацию о дикторе, номере сессии, наличии или отсутствии фальсифицированной речи, может в автоматическом режиме подготовить необходимые для испытаний протоколы.

3.6.3. Модуль имитации спуфинг атак на голосовую биометрическую систему

Модуль имитации спуфинг атак подключается при отсутствии возможности использования заранее подготовленной тестовой речевой базы данных, содержащей фальсифицированную речь дикторов.

В модуль заложена возможность использования различных методов имитации спуфинг атаки.

Схема варианта спуфинг атаки, имитируемой модулем, представлена на рисунке 3.5.

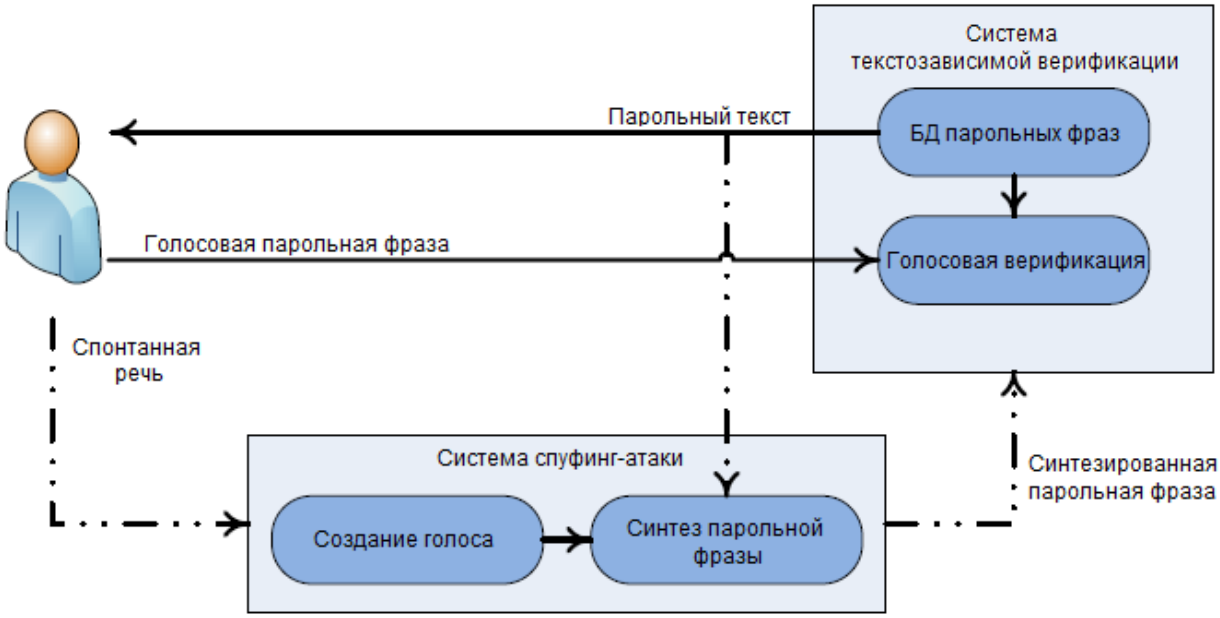

Рисунок 3.5 - Схема варианта спуфинг атаки, имитируемой модулем имитации спуфинг атаки

Данный метод имитации атаки заключается в создании синтезированного голоса пользователя голосовой биометрической системы. При этом заложенная возможность имитации сценария атаки на текстозависимую систему верификации позволяет использовать данный метод как для имитации атак на тестозависимые, так и для имитации атак на текстонезависимые голосовые биометрические системы. Для обучения системы синтеза используется предварительно подготовленная тестовая речевая база данных, содержащая речь дикторов, шаблоны которых зарегистрированы в системе. Для проведения транзакции идентификации или текстонезависимой верификации, содержащей активную попытку "самозванца", при помощи полученного голоса синтеза и произвольного текста подготавливается синтезированная парольная фраза, содержащая фальсифицированную речь зарегистрированного в системе диктора. Для

проведения транзакции текстозависимой верификации, содержащей активную попытку "самозванца", используется заранее подготовленный текст парольной фразы, имитирующий перехват "самозванцем" информации о тексте индивидуальной парольной фразы.

3.6.4. Модуль тестирования голосовой биометрической системы

При выборе варианта реализации данного модуля, ключевым являлось выполнение следующих требований к комплексу:

- возможность интеграции в инфраструктуру, применяемую при функциональном, регрессионном и нагрузочном тестировании системы;
- возможность указания пороговых значений для численных показателей эффективности;
- автоматическое уведомление оператора комплекса программных средств об обнаруженном не соответствии численных показателей и установленных порогов;
- возможность применения распределённых вычислений для экономии временных затрат.

Современные автоматизированные системы функционального, регрессионного и нагрузочного тестирования, в подавляющем большинстве, опираются на подход модульного тестирования (Unit-testing) [67]. При этом используются типовые решения автоматизации и программные пакеты, опирающиеся на единую парадигму [26]. Типовые решения по автоматизации тестирования включают в себя решения по автоматизации процесса разработки программного обеспечения такие, как системы контроля версий (Versions Control System, VCS) и системы непрерывной интеграции (Continues Integration System, CIS). Использование данных систем и программных пакетов при тестировании позволяет не только снизить

издержки на поддержку дополнительных систем, но и даёт возможность запускать тесты распределённо, мгновенно уведомляя команду разработчиков об обнаруженном несоответствии.

Таким образом, в качестве ядра для модуля тестирования голосовой биометрической системы был выбран стандартный пакет языка программирования Python под названием unittest [27].

Данный пакет изначально разрабатывался как инструмент для модульного тестирования кода, написанного с использованием языка программирования Python. Для этого в нём реализованы следующие сущности:

- assert - проверка совпадения ожидаемого и полученного значений;
- test - функция, являющаяся отдельным тестом, содержащая набор проверок;
- TestCase - класс, являющийся тестовым сценарием, содержит в себе набор тестов;
- TestSuite - модуль, содержащий набор тестовых сценариев.

Пакет позволяет найти все тестовые сценарии в указанных модулях и последовательно или параллельно запустить все тесты, содержащиеся в них, последовательно выполнив каждую проверку, имеющуюся в каждом из тестов.

При решении задачи тестирования голосовой биометрической системы для оценки эффективности аутентификации на технологическом испытании, пользователю системы необходимо последовательно подавать в систему различные тестовые речевые базы данных, соответствующие разным сценариям и условиям работы. При этом, в рамках одного сценария работы, необходимо измерять одни показатели эффективности. Например, при оценке эффективности аутентификации системой, работающей на

аналоговом телефоне, мы оцениваем те же показатели, что и при оценке эффективности с GSM каналом. Разница заключается в порогах и тестовых речевых базах данных.

Таким образом, для каждого сценария работы можно реализовать свой базовый класс тестового сценария. Он будет иметь следующий общий вид:

```
import unittest # импортирум пакет

def setUpSuite()
    """
    содержит код, который вызывается перед запуском тестов всех тестовых
    сценариев данного набора
    """
    pass

class SomeBaseTest(unittest.TestCase):

    @classmethod
    def setUpClass(cls):
        """
        содержит код, который будет вызван один раз перед запуском тестов
        данного тестового сценария, например запуск необходимой тестовой
        речевой базы данных на обработку ГБС по соответствующему сценарию
        """
        pass

    def setUp(self):
        """
        содержит код, который будет вызываться каждый раз перед запуском
        каждого теста данного тестового сценария
        """
        pass

    def test_FTE(self):
        """
        Содержит необходимые проверки, например получение значения ВОР - res_fte
        и её сравнение с пороговым - 0.1%
        """
        self.assertLess(res_fte, 0.1):

    def tearDown(self):
        """
        содержит код, который будет вызываться каждый раз после запуска
        каждого теста данного тестового сценария
```

```
            """
            pass

        @classmethod
        def tearDownClass(cls):
            """
            содержит код, который будет вызван один раз после запуска тестов
            данного тестового сценария, например сохранение полученных результатов
            в протокол
            """
            pass

def tearDownSuite()
    """
    содержит код, который вызывается после запуска тестов всех тестовых
    сценариев данного набора
    """
    pass
```

К очевидным плюсам предложенной реализации относятся:

- возможность наследования одного тестового сценария от другого и указания независимых пороговых значений;
- возможность параллельного запуска тестовых сценариев;
- отсутствие дополнительных трудозатрат на реализацию и поддержку функционала уведомления оператора комплекса о выявленных несоответствиях;
- лёгкая интеграция с существующей инфраструктурой автоматизированного регрессионного тестирования ПО.

Пример визуализации уведомления оператора комплекса о несоответствии численного показателя пороговому значению при интеграции с системой непрерывной интеграции TeamCity, разработанной компанией JetBrains [24] отображён на рисунке 3.6.

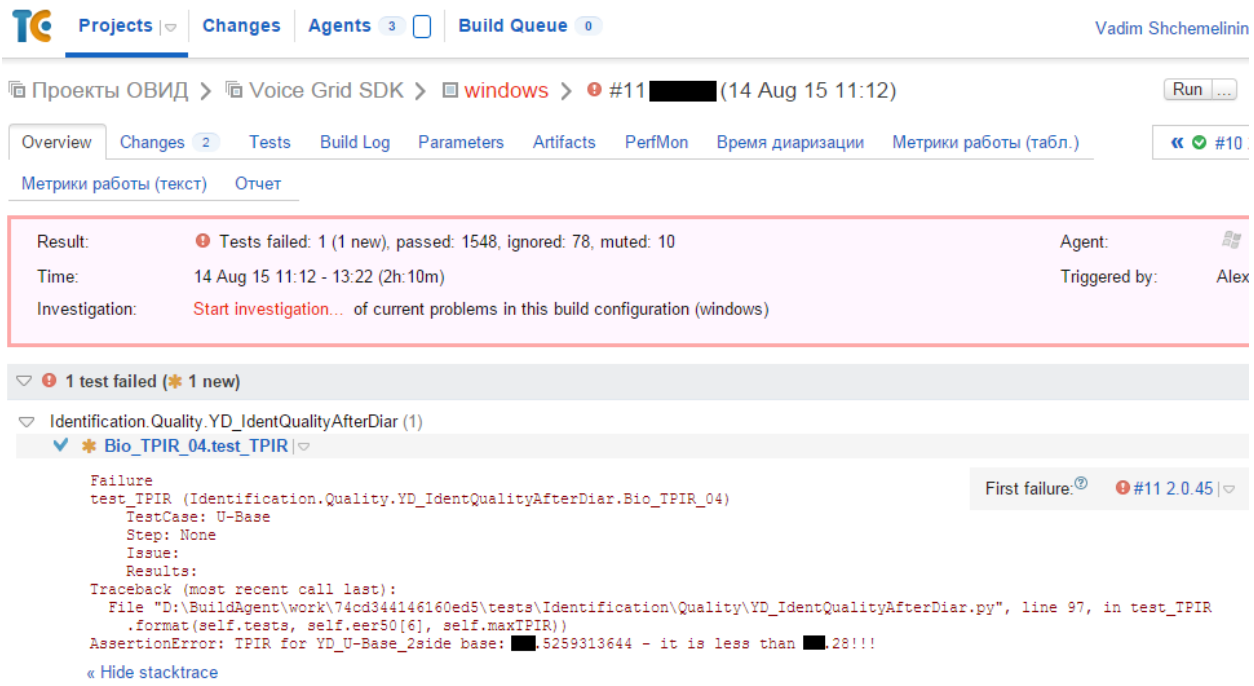

Рисунок 3.6 - Визуализация уведомления оператора комплекса о несоответствии численного показателя установленному пороговому значению

3.6.5. Модуль расчета показателей эффективности аутентификации голосовой биометрической системой

Учитывая необходимость использования функций расчёта показателей эффективности аутентификации голосовой биометрической системой на разных этапах испытаний, становится очевидным, что функционал расчёта показателей эффективности аутентификации голосовой биометрической системой должен быть вынесен в отдельный модуль.

Архитектурно, данный модуль представляет собой класс расчёта показателей эффективности в соответствии с предложенной ранее методикой и набор вспомогательных классов, реализованный в соответствии с принципами ООП.

Конструктор основного класса принимает на вход два набора:

1. Набор полученных значений степеней сходства для транзакций вида "свой-свой".

2. Набор полученных значений степеней сходства для транзакций вида "свой-чужой".

Используя полученные данные и информацию от модуля сопряжения с голосовой биометрической системой, модуль рассчитывает значения ВОР, ВОСД и точки графика зависимости ВЛС и ВЛНС от порога.

Остальные численные и графические показатели рассчитываются на основе уже полученных результатов, при вызове соответствующих методов класса.

Результаты расчёта для численных показателей возвращаются в виде структуры, содержащей описание показателя и его значение. Результаты графических показателей возвращаются в виде списка объектов "точек" на графике.

3.6.6. Модуль генерации протоколов испытаний

Вне зависимости от наличия или отсутствия интеграции комплекса оценки эффективности аутентификации голосовой биометрической системой с системами непрерывной интеграции, результатом работы комплекса программных средств должен стать протокол испытаний, оформленный в соответствии с заданным шаблоном или без него.

В связи с этим, было принято решение выводить протокол оценки в виде документа в формате языка гипертекстовой разметки (Hyper Text Markup Language, HTML). Данный формат поддерживается на всех операционных системах и легко конвертируется в более популярные форматы для файлов протоколов, такие, как межплатформенный формат электронных документов (Portable Document Format, PDF), файлы документов (Document, DOC) или открытый формат текста (Open Document Text, ODT). Для построения графиков применяется сервис Google Charts [23].

При наличии интеграции комплекса с системой непрерывной интеграции, сгенерированные отчёты легко добавляются в форму отчётности подобных систем. Пример отчёта изображён на рисунке 3.7.

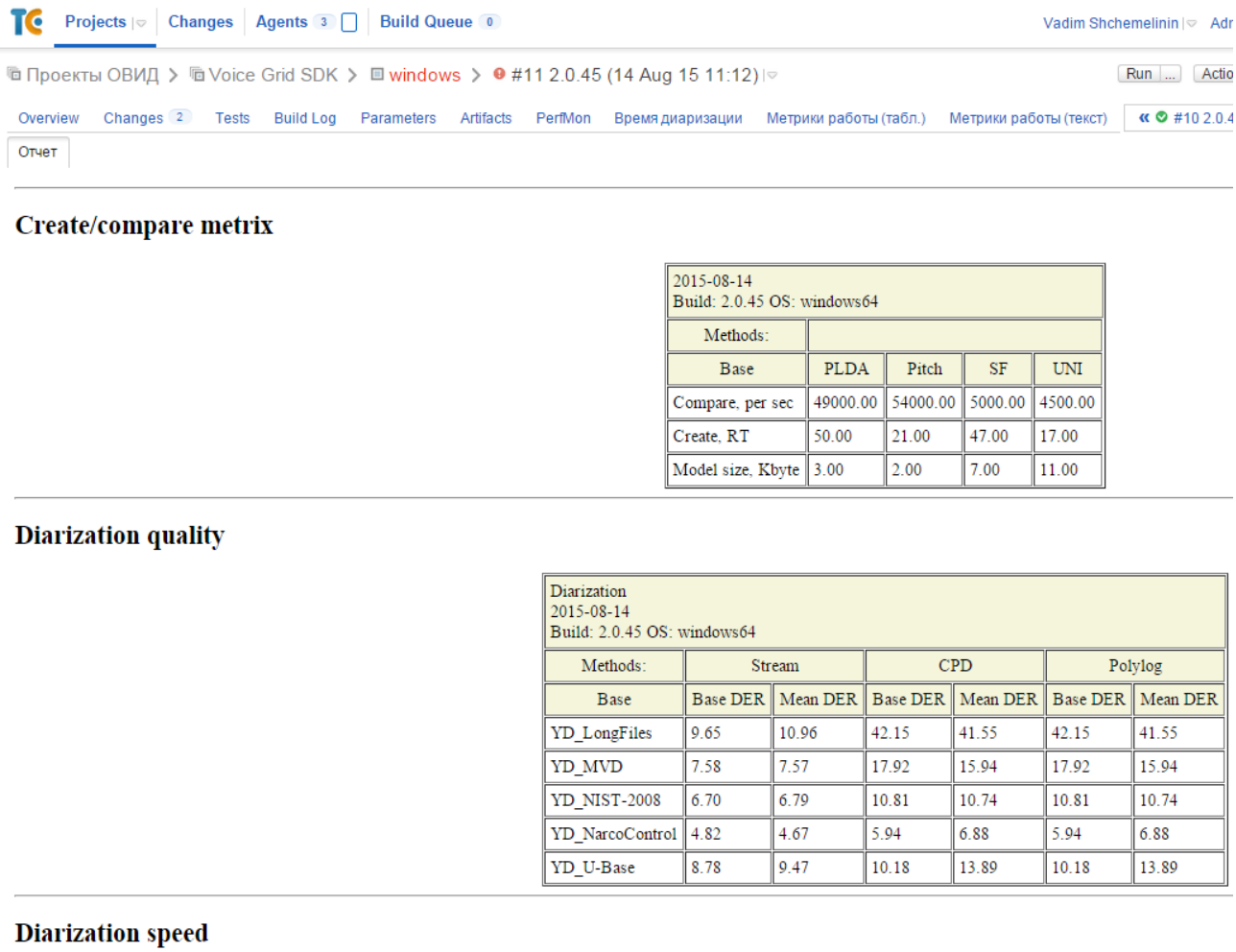

Рисунок 3.7 - Отображение отчёта об оценки эффективности ГБС при интеграции с системой непрерывной интеграции TeamCity

3.7. Выводы

Предложенная в главе методика оценки эффективности аутентификации голосовыми биометрическими системами, на данный момент является единственной методикой, в которой учена оценка устойчивости системы к возможным методам спуфинг атак на модуль ввода биометрической информации.

Описанный комплекс программных средств оценки эффективности аутентификации голосовыми биометрическими системами позволяет

полностью автоматизировать данный процесс при проведении технологических испытаний и может быть легко модернизирован при изменении методики оценки.

Кроме этого, базовые принципы и решения, заложенные в комплексе, позволяют применять его не только для оценки эффективности аутентификации ГБС, но и для проведения функционального и нагрузочного тестирования. Кроме того, практика показала, что при добавлении необходимых метрик и модулей сопряжения, базовые модули предложенного комплекса с успехом применяется при проведении оценки эффективности функций таких систем, как:

- системы распознавания по изображению;
- системы автоматического распознавания речи;
- системы синтеза речи;
- системы шумоочистки;
- системы преобразования сигнала.

Применение предложенного комплекса и методики является первым шагом на пути к разработке технических решений по увеличению устойчивости голосовых биометрических систем к спуфинг атакам модуль ввода биометрической информации.

ГЛАВА 4. Методы повышения устойчивости ГБС к спуфинг атакам

4.1. Увеличение устойчивости к спуфинг атакам фальсификации БХЧ

Проведённые в предыдущих главах исследования устойчивости голосовых биометрических систем к спуфинг атакам показывают, что для поддержания необходимого уровня надёжности под воздействием атак необходима разработка алгоритмов их детектирования. Большая часть существующих алгоритмов противодействия спуфинг атакам на голосовые биометрические системы обучена на небольших тренировочных базах и способна защитить лишь от крайне специфических атак, в то время, как реальные атаки будут основаны на неизвестных методах фальсификации индивидуальных биометрических характеристик человека. По этой причине был организован конкурс Automatic Speaker Verification Spoofing and Countermeasures (ASVspoof) Challenge 2015 [94], в котором соревновались алгоритмы детектирования известных и неизвестных видов спуфинга.

Конкурс ASVspoof Challenge 2015 нацелен на решение отдельной задачи детектирования атаки, без учёта влияния детектора на систему в целом.

В ходе работы были разработаны несколько детекторов спуфинг атак, которые были отправлены на конкурс ASVspoof Challenge 2015. В основе алгоритмов детекторов спуфинг атак был использован стандартный совместный факторный анализ в пространстве полной изменчивости (Total Variability Joint Factor Analysis, TV-JFA) для статистического моделирования акустических особенностей речевых сигналов. В качестве классификаторов применялись метод опорных векторов (Support Vector Machine, SVM) или, в качестве запасной альтернативы, нейронная сеть глубокого обучения (Deep Belief Network, DBN).

При разработке детекторов был сделан фокус на выборе наиболее подходящих акустических признаков в рассматриваемых системах детектирования спуфинга. В частности, были исследованы признаки, вычисляемые с использованием информации фазового спектра, а также с помощью вейвлет-преобразования [80]. Целью нашей работы являлось нахождение наиболее надёжного метода детектирования неизвестных спуфинг атак.

4.1.1. Детектор нулей

В рамках проведения экспериментов на обучающей базе конкурса ASVspoof Challenge 2015 было замечено, что в большинстве звуковых записей, являющихся результатом работы системы спуфинга, присутствуют длинные последовательности нулевых значений энергии. На основании этого, был сделан вывод о необходимости применения простого предварительного детектора нулей, как предварительного шага системы детектирования попыток взлома. Предварительный детектор рассчитывает максимальную длину последовательности нулевых значений энергии во входном звуковом сигнале и на основании её значения принимает решение о наличии или отсутствии спуфинг атаки.

На рисунке 4.1 представлены ошибки вероятности ложного совпадения и ложного несовпадения такого детектора спуфинг атаки в зависимости от максимальной длины непрерывной последовательности нулевых значений энергии, полученные на обучающей части речевой базы конкурса ASVspoof Challenge 2015. Данная часть базы содержит 3752 записи настоящей человеческой речи и 12625 записей полученных различными методами спуфинга, основанными на синтезе и подмене индивидуальных биометрических характеристик человека.

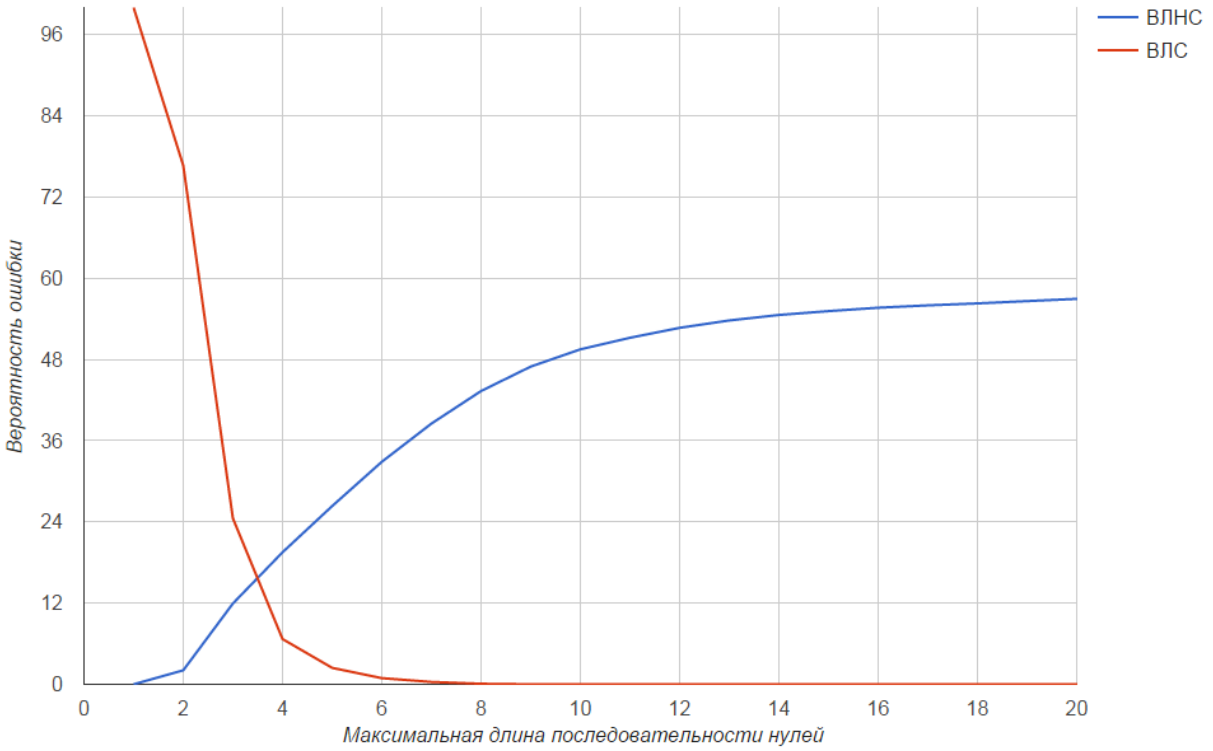

Рисунок 4.1 - Зависимость ошибок ложного совпадения и ложного несовпадения детектора спуфинг атаки на основе детектирования нулей, в зависимости от максимальной длины последовательности нулевых значений энергии для обучающей части базы конкурса ASVspoof Challenge 2015

Как видно из результата оценки ВЛС и ВЛНС, при установке значения порога принятия решения равным девяти, вероятность ложного срабатывания детектора становится нулевой, при 53% вероятности успешного детектирования атаки, что является не плохим результатом для такого элементарного алгоритма, работающего в роли предварительного детектора.

Таким образом, в случае обнаружения последовательности нулей некоторой длины, с 53% надёжностью принимается решение о том, что входной сигнал является спуфинг атакой, в противном случае (а также при работе без предложенного предварительного детектора) сигнал поступает на вход к модулю извлечения акустических признаков.

4.1.2.Детекторы спуфинг атак ООО "ЦРТ"

Все представленные на конкурс Automatic Speaker Verification Spoofing and Countermeasures (ASVspoof) Challenge 2015 системы автоматического

детектирования спуфинг атак, разработанные в ООО "ЦРТ", состоят из трёх основных компонентов:

- модуль извлечения информативных акустических признаков из аудиозаписи;
- модуль извлечения i-векторов в пространстве полной изменчивости;
- классификатор.

Модули извлечения акустических признаков, использованные в построенных системах, представляют собой комбинацию нескольких различных методов извлечения речевых характеристик из входного сигнала. Описание акустических признаков изученных и протестированных в рамках участия в конкурсе, описаны ниже.

Полученные векторы информативных акустических признаков подаются на вход модулю извлечения i-векторов для каждого типа акустических признаков. Они, в свою очередь, объединяются в один общий i-вектор, после чего он центрируется и нормируется по длине. Финальным модулем является классификатор, который вычисляет результирующую оценку принадлежности речевого сигнала к классу спуфинга или к классу подлинной речи. Общая схема разработанных детекторов атак представлена на рисунке 4.2.

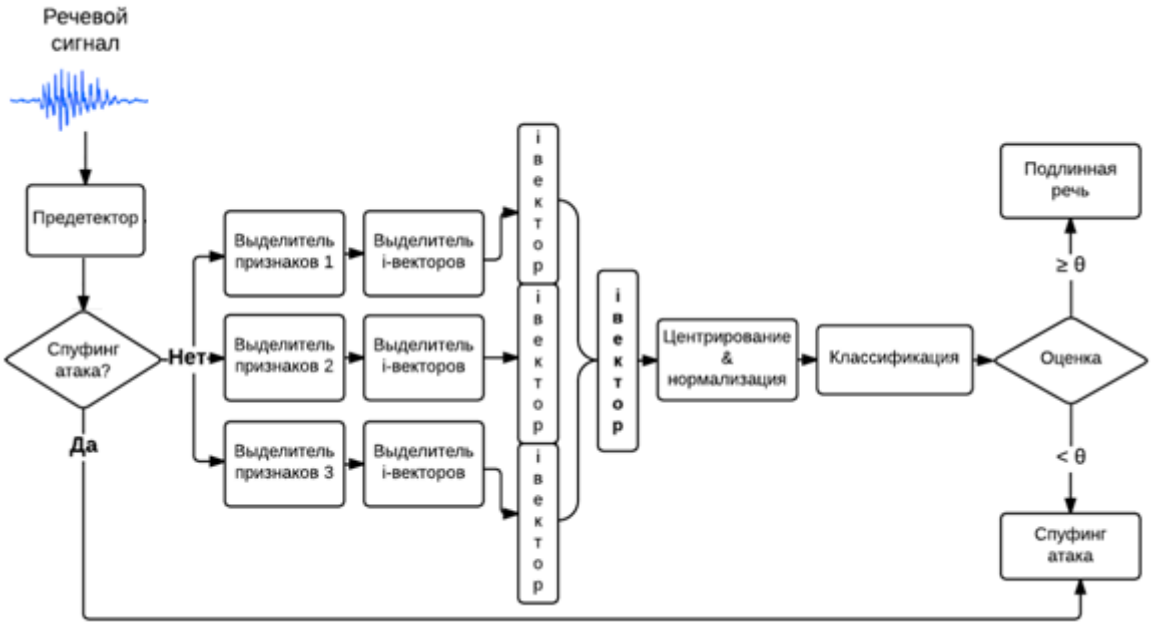

Рисунок 4.2 - Общая схема системы автоматического детектирования спуфинг атак

Для моделирования вероятностного пространства акустических признаков был использован стандартный совместный факторный анализ в пространстве полной изменчивости (Total Variability Joint Factor Analysis, TV-JFA), являющийся одним из самых современных в области голосовой верификации [43]. Модели голоса в JFA имеют вид:

$$M = m + Ux + Vy + Dz$$

где:

M – супервектор смеси гауссовских распределений (GMM-модели) фонограммы,

m – супервектор универсальной фоновой модели (Universal Background Model, UBM),

U, V, D–матрицы собственных каналов (Eigen Channel), собственных голосов (Eigen Voice) и остаточной изменчивости.

В пространстве полной изменчивости (Total Variability, TV) i-вектор извлекается посредством применения обычного Гауссова факторного анализатора, определенного на средних супервекторах универсальной

фоновой модели (Universal Background Model, UBM) и матрицы полной изменчивости T. Модель голоса в TV-методе имеет вид:

$$M = m + Tw$$

где:

w - низкоразмерный вектор,

T - матрица полной изменчивости.

В предложенных системах ООО "ЦРТ", UBM была представлена смесью гауссовых моделей описанных признаков. Для обучения Т-матрицы и UBM использовались признаки, полученные на обучающей базе конкурса. Диагональная ковариационная UBM была обучена с помощью EM-алгоритма (Expectation– Maximization) [50].

В исследованиях, в качестве классификатора применялся метод опорных векторов (Support Vector Machine, SVM) с линейным ядром. Разделяющая гиперплоскость была построена в нормализованном пространстве i-векторов обучающей базы конкурса ASVspoof Challenge 2015 для детектирования спуфинг атак. Для обучения SVM использовалась библиотека LIBLINEAR [25], обеспечившая необходимую точность и вычислительную скорость.

В качестве альтернативного классификатора использовалась нейронная сеть глубокого обучения (Deep Belief Network, DBN) с softmax выходным слоем и стохастическими бинарными скрытыми слоями [52]. DBN принимает на вход нормированные общие i-вектора, полученные от модуля извлечения i-векторов. Использовалось послойное предварительное обучение слоев ограниченными машинами Больцмана (Restricted Boltzmann Machines, RBM) [52], после чего применяли метод обратного распространения для обучения DBN с учителем для решения задачи детектирования речи и спуфинг атаки.

В исследовании были рассмотрены различные акустические признаки, для определения наиболее информативных, и, как следствие, эффективных

при решении задачи детектирования спуфинг атаки в рамках конкурса ASVspoof Challenge 2015. Опишем их подробнее.

4.1.2.1. Амплитудные спектральные признаки

В качестве краткосрочных амплитудных спектральных акустических признаков были выбраны мел-частотные коэффициенты, полученные двумя способами. Мел-частотные кепстральные коэффициенты (Mel Frequency Cepstral Coefficients, MFCC) [49] были получены с помощью дискретного косинусного преобразования, в то время, как мел-частотные кепстральные коэффициенты второго типа были получены с помощью метода главных компонент, вследствие чего назовём их мел-частотными главными коэффициентами (Mel Frequency Principal Coefficients, MFPC). Эти признаки являются представлением краткосрочной энергии спектра сигнала и хорошо отображают общие характеристики голосового тракта.

Для вычисления MFCC коэффициентов использовалась оконная функция Хэмминга с размером окна равным 256 и 50% перекрытием, рисунок 4.3.

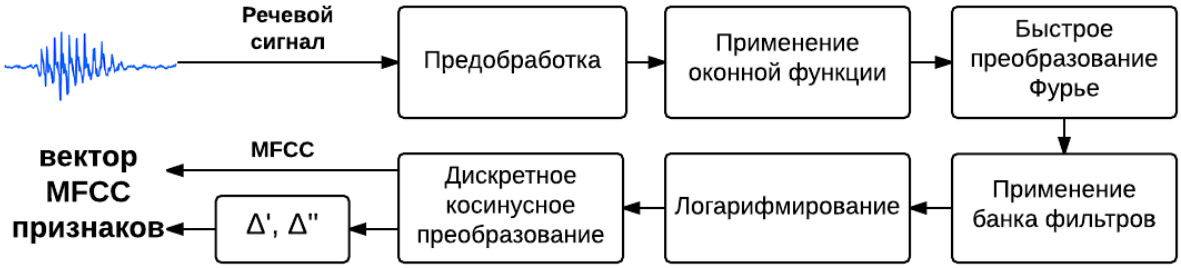

Рисунок 4.3 - Модуль извлечения MFCC признаков

Были использованы первые 12 коэффициентов вместе с их первыми и вторыми производными, как наиболее информативные признаки, таким образом, получился вектор признаков длиной 36 элементов.

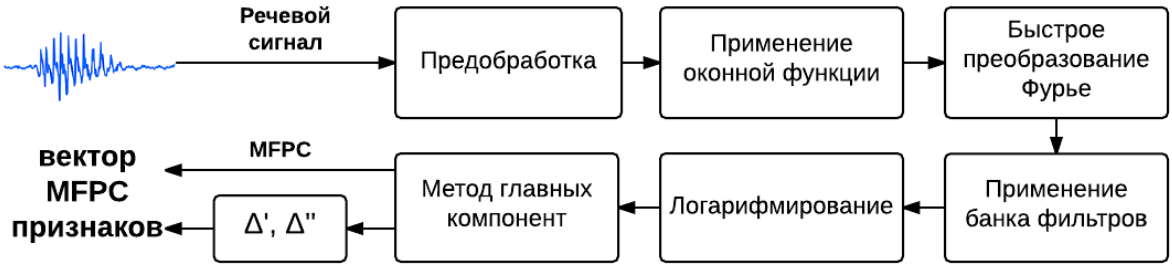

Рисунок 4.4 - Модуль извлечения MFPC признаков

MFPC коэффициенты были получены аналогично MFCC коэффициентам, но, используя метод главных компонент вместо дискретного косинусного преобразования для декорреляции информативных акустических признаков, рисунок 4.4.

4.1.2.2. Фазовые признаки

Желая добавить в рассмотрение фазовую информацию речевого сигнала, были использованы фазовые признаки Cos Phase, подробно описанные в работе [88]. Эти признаки были выделены из фазового спектра следующим способом:

1. Фазовый спектр сглаживался, для получения непрерывной функции от частоты.
2. Скорректированный фазовый спектр был нормирован функцией косинуса для ограничения его области значений до [-1; 1].
3. Для понижения размерности использовался метод главных компонент, базис которого был вычислен заранее на обучающем множестве.

Аналогично признакам, выделенным из амплитудного спектра, были оставлены только первые 12 коэффициентов с их первыми и вторыми производными, которые образовали результирующий вектор косинусно-фазовых главных коэффициентов (Cos Phase Principal Coefficients, CosPhasePC), рисунок 4.5.

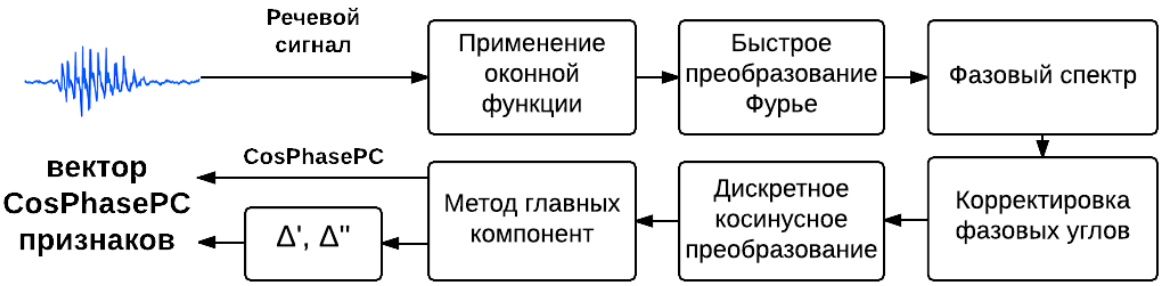

Рисунок 4.5 - Модуль извлечения CosPhasePC признаков

4.1.2.3. Вейвлет-признаки

С целью детального частотно-временного анализа речевых сигналов в работе предложено использовать информационные акустические признаки на основе вейвлет-пакетного преобразования [80], адаптированного к мел-шкале, рисунок 4.6.

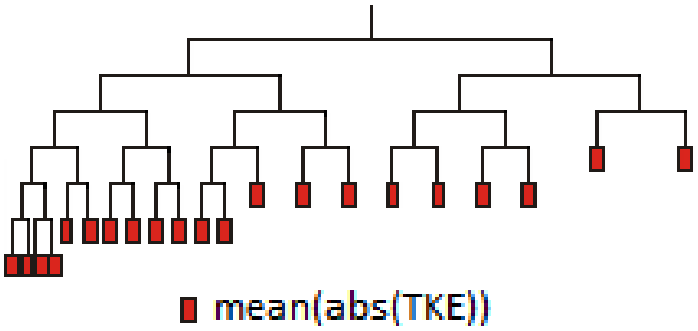

Рисунок 4.6 - Вейвлет-пакетное преобразование. TKE – энергия Тигера Кайзера

Вместо обычной энергии частотных подполос, применялся оператор энергии Тегера-Кайзера. Энергия Тегера-Кайзера (TKE) обладает большей информативностью, по сравнению с обычной энергией отсчета, и является помехоустойчивым параметром для речевых сигналов [100]. Использовалась следующая формула для расчета TKE:

$$\Psi\big(s(t)\big) = s(t)^2 - s(t-1)s(t+1)$$

где $s(t)$ – временной отсчет выходного сигнала рассматриваемой подполосы.

Для декорреляции полученных признаков последовательно применялся метод главных компонент, и получались 12 коэффициентов. Эти признаки для краткости назовём мел-частотными вейвлет-пакетными коэффициентами (Mel Wavelet Packet Coefficients, MWPC). Здесь также принимались во внимание первые и вторые производные, рисунок 4.7, а для выделения признаков использовали оконную функцию Хэмминга с размером окна равным 256 и 50% перекрытием.

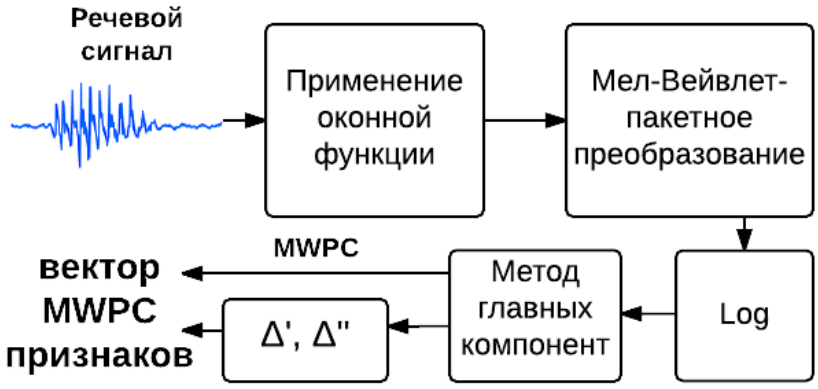

Рисунок 4.7 - Модуль извлечения MWPC признаков

4.1.2.4. Результаты конкурса ASVspoof Challenge 2015

Исследования были проведены при обучении всех параметров системы детектирования спуфинг атак на обучающем множестве речевой базы данных конкурса, а тестирование производилось на тестовом множестве речевой базы данных конкурса ASVspoof Challenge 2015. По условиям конкурса в обучающем и тестовом множествах были представлены пять вариантов [94] спуфинг атак S1-S5: S1, S2, S5 – варианты, основанные на алгоритмах преобразования речи, а S3, S4 – варианты, основанные на адаптированном под диктора HMM методе синтеза речи. Детальное описание этих вариантов спуфинг атак было дано во второй главе.

В качестве примера, на рисунке 4.8 представлены LDA проекции части i-векторов MWPC признаков, полученных на тестовом множестве, на первые три главные компоненты P1, P2, P3, которые были вычислены на обучающем

множестве. Для вейвлет разложения использовалась вейвлет функция Добеши.

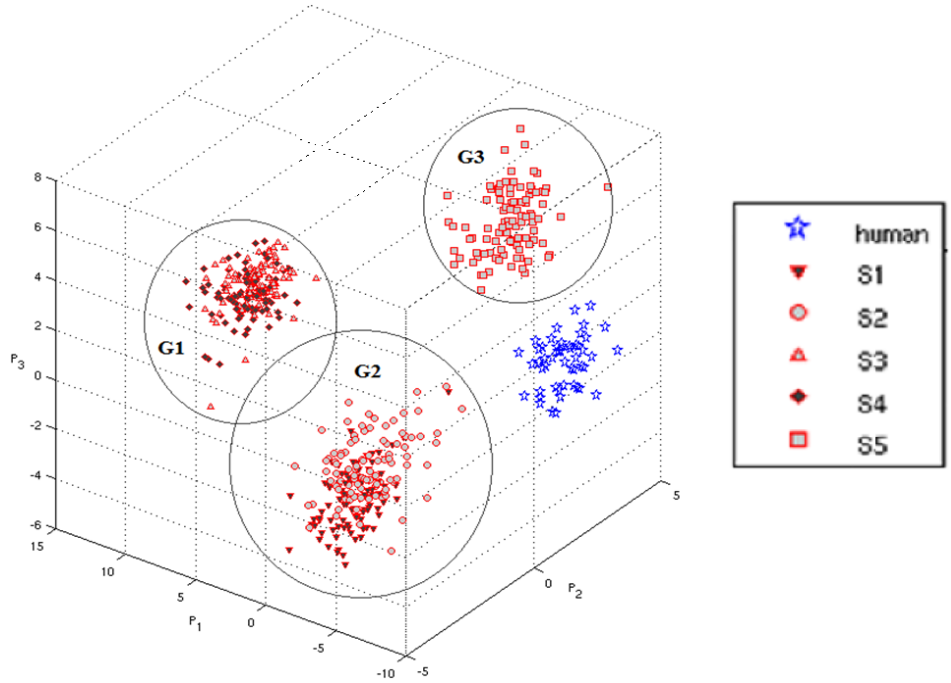

Рисунок 4.8 - LDA проекции MWPC i-векторов на оси главных компонент P1, P2, P3

На рисунке 4.8 видно, что класс естественной (человеческой) речи хорошо отделим от классов спуфинга. Это дает основания предполагать, что использование методов линейного разделения классов типа SVM является эффективным для решения задачи детектирования спуфинг-атак в этом пространстве. Заметим, что в этом пространстве признаков также хорошо дискриминируются три группы вариантов спуфинга: G1 – группа спуфинга на основе HMM синтеза речи (S3+S4); G2 – группа методов простого преобразования речи (S1+S2); G3 представляет Festvox [22] метод преобразования речи.

В таблице 4.1 представлены результаты оценки уровня равновероятностной ошибки EER(%) на тестовом множестве ASVspoof Challenge 2015 для систем верификации на основе TV-SVM метода при использовании различных акустических признаков, описанных ранее. Данные результаты получены с применением UBM с 256 компонентами и i-векторами размерностью 200.

Таблица 4.1 - Результаты для TV-SVM систем на различных спуфинг атаках, EER(%)

Признаки	S1	S2	S3	S4	S5	Все
MFCC	0,38	2,13	0,36	0,39	1,48	1,14
MFPC	0,13	0,29	0,09	0,09	0,37	0,23
CPPC	0,13	0,20	0,04	0,05	0,23	0,15
MWPC	0,03	0,11	0,00	0,00	0,08	0,05

Результаты показывают, что MFCC признаки уступают остальным рассматриваемым признакам. Следует отметить, что замена дискретно-косинусного преобразования в MFCC на декоррелирующий базис главных компонент и переход к MFPC демонстрирует существенное улучшение показателя EER для всех вариантов спуфинга. Использование фазовых признаков CosPhasePC дает небольшие улучшения EER по сравнению с MFPC. А наилучший результат демонстрируют признаки на основе мульти-разрешающего вейвлет-преобразования, достигая 0,05% EER для всех известных методов атак. Отметим, что в ходе экспериментов было выявлено, что применение оператора TKE в MWPC демонстрирует немного лучший результат по сравнению с обычной энергией. Результаты исследований показали, что для всех комбинаций признаков, для известных атак, равновероятная ошибка детектора близка к 0%. Заметим, что нулевую ошибку детектирования спуфинга на тестовом множестве удалось получить только с применением CosPhasePC и MWPC признаков.

Основываясь на результатах проведенных экспериментов, на конкурс ASVspoof Challenge 2015 были предложены три системы автоматического детектирования спуфинг атак.

Отметим, что эксперименты показали самую высокую ошибку на методе атаки S2. При этом, как было показано в предыдущих главах, данный метод представляет самую низку угрозу для системы распознавания из всех пяти рассматриваемых. Для учёта степени угрозы от различных методов атак, необходимо производить оценку надёжности детекторов вместе с

результатом работы системы распознавания диктора. Но в рамках конкурса ASVspoof Challenge 2015 это было не предусмотрено.

Основная система, согласно схеме изображённой на рисунке 4.2, содержала предварительный детектор и модули выделения MFCC, MFPC и CosPhasePC признаков. UBM была представлена смесью 1024 гауссовых моделей описанных признаков, а размерность TV пространства была равна 400. Для классификации использовался классификатор SVM.

В **первой альтернативной** системе не использовался предварительный детектор, а MFPC признаки были заменены MWPC.

Вторая альтернативная система также не использовала предварительный детектор, а для классификации использовала нелинейный DBN-классификатор. Для того, чтобы избежать переобучения, количество компонент UBM было понижено до 256 для всех акустических признаков, а размерность TV пространства - до 200.

Дополнительно к известным методам атак, конкурсная база содержала спуфинг атаки неизвестных методов S6-S10 [94]. Значения EER на конкурсной базе представлены в таблице 4.2.

Таблица 4.2 - Результаты экспериментов на конкурсной базе, EER(%)

Система	Известные атаки	Неизвестные атаки	Все
Основная	0,008	3,922	1,965
Альтернативная 1	0,009	4,891	2,450
Альтернативная 2	0,017	6,162	3,090
Основная (без предетектора)	0,008	5,151	2,579

Не смотря на хорошие результаты на известных видах атак, результаты, полученные для неизвестных видов спуфинга, оказались заметно хуже: даже лучшая предложенная (основная) система позволила достичь EER=3,92% и заняла второе место на конкурсе Automatic Speaker Verification Spoofing and Countermeasures (ASVspoof) Challenge 2015 [94]. Эта оценка подтверждает

необходимость улучшения мер по противодействию спуфинг атакам неизвестных методов. Подробные результаты, показанные системой на конкурсе доступны в работе [65]. Результаты работы некоторых других участников также доступны в опубликованных работах [54, 55, 60, 64, 68, 75, 76, 87, 97, 99].

Основная система показала наилучший результат, в частности, за счет применения предварительного детектора по энергии, что подтверждают результаты сравнения EER основной системы с предварительным детектором и без. Однако необходимо отметить, что описанный предварительный детектор не будет работать в условиях канальных искажений и аддитивных шумов.

В отличие от основной системы, первая альтернативная система не имела предварительного детектора, и показала относительно хорошую эффективность по результатам конкурса. Возможная причина этого заключается в использовании MWPC признаков. В основе этих признаков лежит вевлет-разложение, которое позволяет проводить детальный мультиразрешающий анализ сигналов, что дает дополнительный выигрыш в решении задачи анти-спуфинга.

Результат второй альтернативной системы, на основе DBN-классификатора, оказался наихудшим. Возможно, здесь не удалось избежать эффектов более сильного переобучения DBN-классификатора на обучающем множестве, по сравнению с SVM. На основании этих выводов можно заключить, что в предложенной системе анти-спуфинга целесообразнее использовать линейный SVM классификатор.

4.2. Результат экспериментальной оценки устойчивости к различным методам спуфинг атак при использовании детектора

Согласно методике, предложенной в третье главе, оценивать устойчивость системы распознавания речи к спуфинг атакам необходимо комплексно. Для демонстрации эффективности предложенной методики, проведём оценку аутентификации двумя системами текстонезависимого распознавания диктора. Обе системы, используемые для сравнения, основаны на i-векторах. Для улучшения качества распознавания в них используется модуль предварительной обработки сигнала. Данный модуль включает детектор речи, основанный на энергии сигнала, а также детекторы клиппирования сигнала, импульсных и тональных помех [2]. В качестве речевых признаков используются вектора мел-частотных кепстральных коэффициентов (mel-frequency cepstrum coefficients, MFCC), их производные первого и второго порядка в количестве 13 элементов. Длина каждого речевого кадра для вычисления MFCC составляет 22мс со сдвигом в 11мс. Для компенсации эффекта Гиббса, в тестируемых системах, используется взвешивание сигнала окном Хемминга. Компенсация эффектов канальных искажений на уровне признаков реализована путем вычитания кепстрального среднего (cepstral mean subtraction, CMS).

На этапе создании модели голоса диктора, в тестируемых системах, используется гендеро-независимая универсальная фоновая модель (universal background model, UBM), представленная 512-компонентной СГР. Обучение UBM производилось с помощью стандартного EM-алгоритма на телефонной части речевых баз данных NIST SRE 1988-2010 [17]. Для ускорения вычислений применяется диагональная ковариационная матрица UBM. Общее количество дикторов в обучающих базах данных составляло около 4000.

Модуль оценки i-вектора также обучен на более чем 60000 телефонных и микрофонных записях из речевых баз данных NIST SRE 1998-2010, включающих более 4000 голосов дикторов.

Основное выражение, определяющее представление модели СГР в низкоразмерном пространстве полной изменчивости, приведено ниже:

$$\mu = m + T\omega + \varepsilon$$

где μ- супервектор параметров СГР модели диктора,

m- супервектор параметров UBM,

T - матрица, задающая базис в редуцированном пространстве признаков,

ω - i-вектор в редуцированном пространстве признаков, $\omega \in N(0,1)$,

ε - вектор ошибки.

Модуль линейного дискриминантного анализа был обучен на тех же речевых базах данных NIST SRE 1988-2010.

Различие систем заключается в отсутствии модуля детектирования спуфинг атак в первой предложенной системе (система А), и его наличии во второй (система Б).

Схема подключения модуля детектирования спуфинг атак, описанного ранее, изображена на рисунке 4.9.

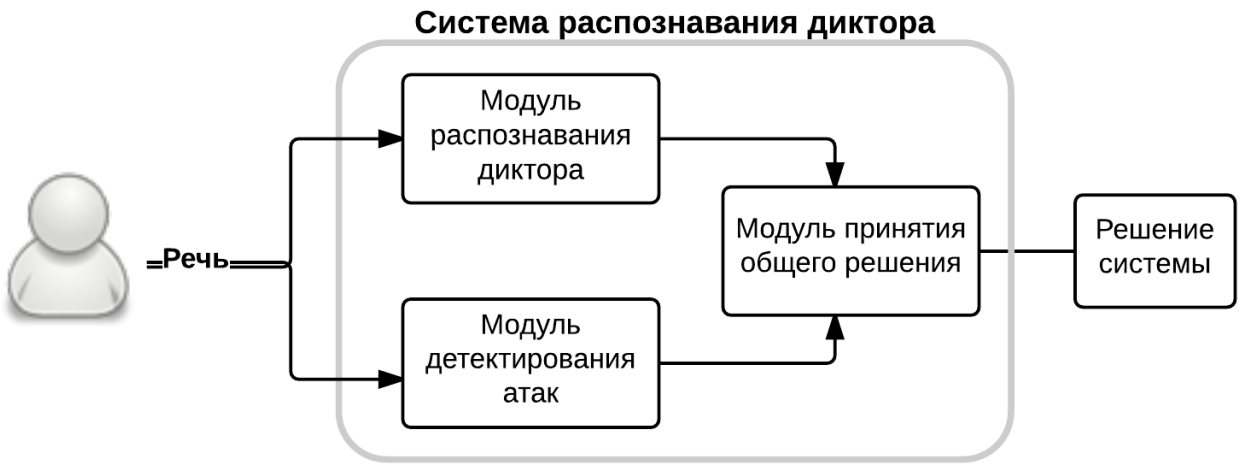

Рисунок 4.9 - Схема взаимодействия модуля детектирования атак с модулем распознавания диктора

Эксперименты показали, что данная схема подключения модуля детектирования спуфинг атак позволяет уменьшить количество ошибок ложного отклонения системой распознавания диктора, относительно последовательного подключения модуля распознавания диктора и модуля детектирования спуфинг атак.

Предположим, что случайное событие A - наличие целевого диктора на фонограмме и случайное событие B - наличие атаки на фонограмме, являются независимыми. Тогда $\mathrm{P}(A \cap \overline{B}) = \mathrm{P}(A) * \mathrm{P}(\overline{B})$.

Для объединения результатов детектора с модулем распознавания, в модуле принятия общего решения системы Б используется следующая формула:

$$P_{system} = P_{speaker} \cdot (1 - P_{spoofing})$$

где:

P_{system} - результат системы распознавания диктора,

$P_{speaker}$ - вероятность того, что диктор в записи совпадает с шаблоном,

$P_{spoofing}$ - вероятность того, что запись является искусственной.

Для преобразования результатов работы подмодулей системы из независимых степеней схожести к нормированным значениям вероятности схожести применяется комплекс программ BOSARIS [21].

На первом шаге предложенной оценки необходимо рассчитать фундаментальные показатели для обеих голосовых биометрических систем. В качестве тестовой речевой базы данных используем класс "естественной" человеческой речи из тестового множества данных конкурса ASVspoof Challenge 2015. Полученная тестовая база содержит 3497 записей 35 дикторов, 15 из которых мужчины, а 20 оставшихся - женщины.

Численные значения фундаментальных показателей эффективности системы А, не содержащей детектор спуфинга, приведены в таблице 4.3.

Таблица 4.3 - Фундаментальные численные показатели эффективности аутентификации для системы А

Показатель	Значение
ВОР	0,00%
ВОСД	0,00%
ВЛС (при пороге 50)	12,94%
ВЛНС (при пороге 50)	0,00%
РВО	0,19%

Численные значения фундаментальных показателей эффективности системы Б, содержащей детектор спуфинга, приведены в таблице 4.4.

Таблица 4.4 - Фундаментальные численные показатели эффективности аутентификации для системы Б

Показатель	Значение
ВОР	0,00%
ВОСД	0,00%
ВЛС (при пороге 50)	10,6%
ВЛНС (при пороге 50)	0,00%
РВО	0,57%

Кривые КОО полученные на описанной тестовой базе для сравниваемых систем текстонезависимого распознавания дикторов показаны на рисунке 4.10.

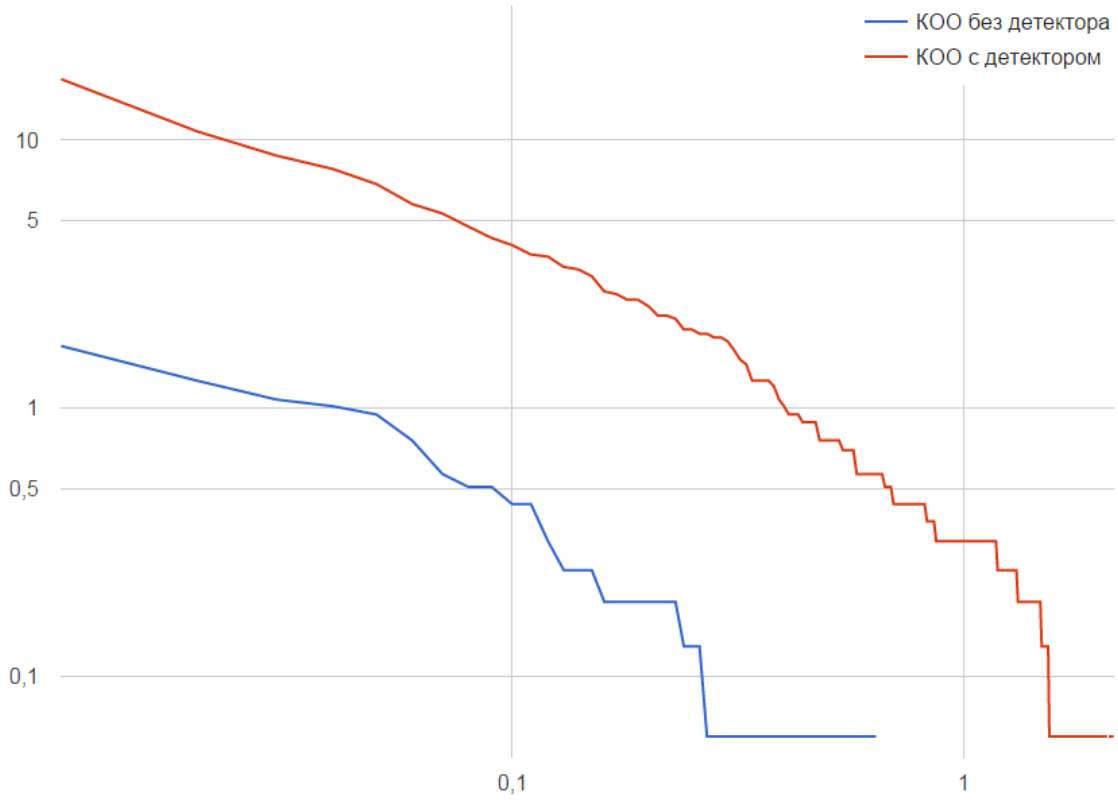

Рисунок 4.10 - Кривые КОО полученные для системы А (без детектора) и системы Б (с детектором), в условиях отсутствия атак

Графические представления зависимостей значений ВЛС и ВЛНС от порогов принятия решений, полученные на описанной тестовой базе для сравниваемых систем текстонезависимого распознавания дикторов, показаны на рисунке 4.11.

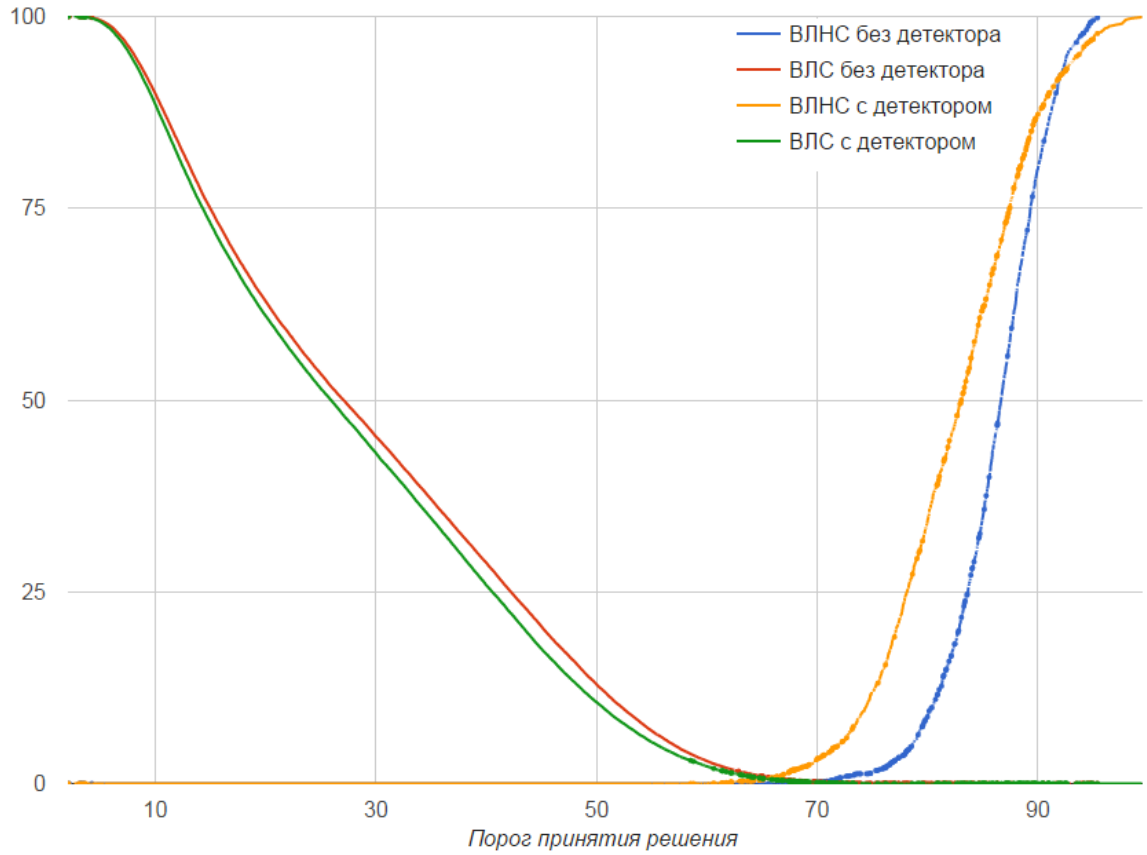

Рисунок 4.11 - Графики зависимости ВЛС и ВЛНС от порога полученные для системы А (без детектора) и системы Б (с детектором), в условиях отсутствия атак

Из полученных результатов напрашивается вывод, о более высокой эффективности аутентификации, выполняемой системой А, по сравнению с системой Б. Действительно, алгоритм детектирования спуфинг атак добавляет свою ошибку в общую ошибку системы, а следовательно, в нормальных условиях система Б будет работать чуть хуже системы А.

В соответствии со вторым шагом предложенной методики, перед принятием конечного решения об эффективности тестируемых систем, необходимо рассчитать показатели устойчивости обоих биометрических систем к воздействию различных атак на модуль ввода биометрических характеристик.

Для имитации требуемых атак воспользуемся речевой базой данных фальсифицированной речи из тестового множества данных конкурса ASVspoof Challenge 2015. Полученная тестовая база содержит 49875 записей

фальсифицированных голосов 35 дикторов, 15 из которых мужчины, и 20 - женщины. В качестве методов фальсификации голоса использованы пять различных спуфинг атак на базе технологий синтеза речи и преобразования речи. Более детальное описание методов фальсификации было дано в предыдущих главах.

Для подготовки эталонных моделей дикторов воспользуемся речевой базой данных, использованной на первом шаге методики.

Численные значения показателей устойчивости к спуфинг атакам системы А, не содержащей детектор спуфинга, приведены в таблице 4.5.

Таблица 4.5 - Численные показатели эффективности аутентификации для системы А под воздействием атаки

Показатель	Значение
ВЛСФ (при пороге EER)	51,29%
ВЛСФ (при пороге ВЛСФ=0,1%)	44,98

Численные значения показателей устойчивости к спуфинг атакам системы Б, содержащей детектор спуфинга, приведены в таблице 4.6.

Таблица 4.6 - Численные показатели эффективности аутентификации для системы Б под воздействием атаки

Показатель	Значение
ВЛСФ (при пороге EER)	0,81%
ВЛСФ (при пороге ВЛСФ=0,1%)	0,03%

Кривые КОО полученные на описанных тестовых базах для сравниваемых систем текстонезависимого распознавания дикторов показаны на рисунке 4.12.

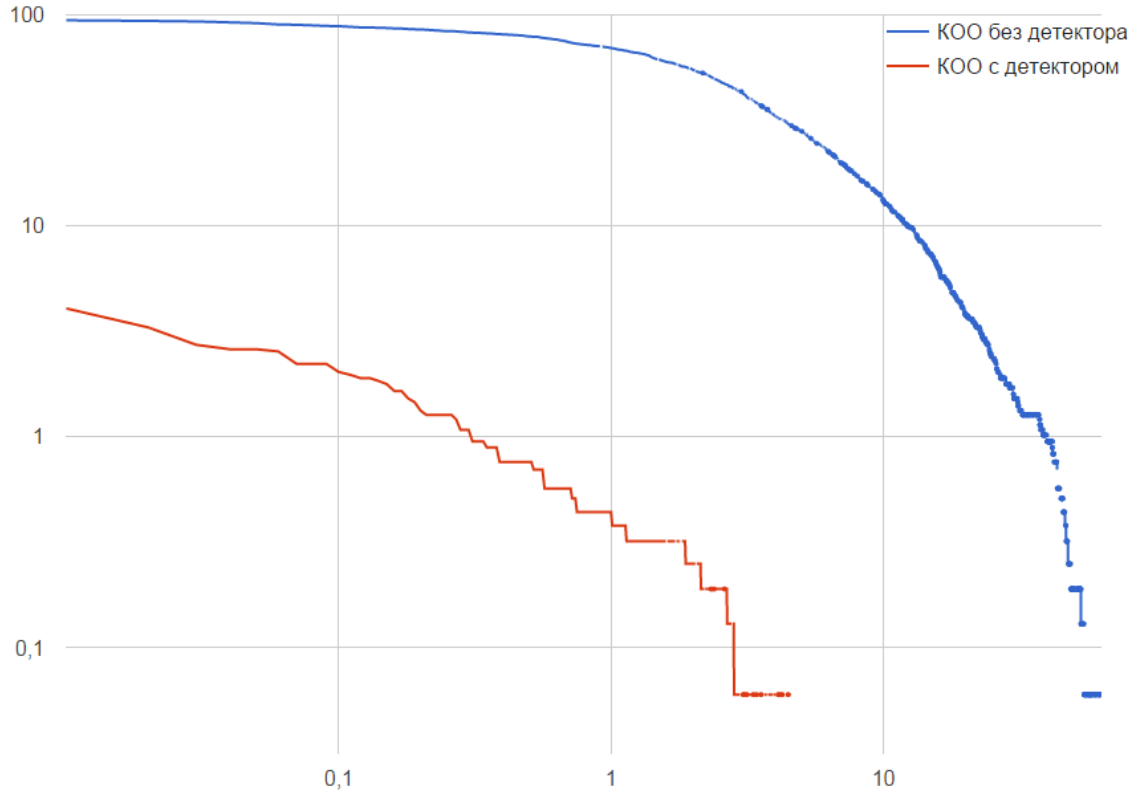

Рисунок 4.12 - Кривые КОО полученные для системы А (без детектора) и системы Б (с детектором), в условиях воздействия спуфинг атак на модуль ввода биометрической информации

Как видно из результатов, полученных на втором шаге предложенной методики, система Б, включающая предложенные технические решения по улучшению защищённости от спуфинг атак, показывает значительное превосходство над системой А, не обладающей модулем детектирования спуфинг атак.

Отметим, что при проведении оценки эффективности в соответствии с действующими стандартами NIST или ГОСТ, система Б была бы признана менее эффективной несмотря на её значительно большую устойчивость к различным методам атак на модуль ввода биометрических характеристик.

4.3. Выводы

Предложенный в главе метод противодействия спуфинг атакам, позволяющий повысить устойчивость голосовых биометрических систем к различным методам спуфинг атак на модуль ввода биометрической информации. показал свою высокую эффективность и занял второе место на

международном конкурсе подобных решений Automatic Speaker Verification Spoofing and Countermeasures (ASVspoof) Challenge 2015 [94]. Тем не менее, текущая надёжность современных технических решений не даёт говорить о полном решении данной задачи.

Проведённые в главе эксперименты ещё раз показывают необходимость комплексной оценки эффективности аутентификации голосовыми биометрическими системами как в нормальных условиях, так и в условиях имитации спуфинг атаки. Только такая комплексная оценка позволит корректно оценить эффективность аутентификации голосовой биометрической системой.

Очень важным является и то, что надёжность технических средств по совершенствованию защиты голосовых биометрических систем необходимо оценивать в рамках единой, цельной системы распознавания диктора. При оценке детекторов атак отдельно от системы распознавания диктора, неизбежны лишние усилия по борьбе с неэффективными методами атак (метод S2 из конкурса ASVspoof Challenge 2015 давал самую большую ошибку детектора, при фактическом отсутствии угрозы для системы распознавания диктора) и плохо контролируемое ухудшение таких фундаментальных показателей, как ВОСД и ВЛНС.

Заключение

Главный результат представленной работы заключается в исследовании и разработке методики оценки эффективности аутентификации голосовыми биометрическими системами, позволяющей комплексно оценивать различные голосовые биометрические системы и корректно сравнивать их между собой с учётом возможных атака на устройство ввода биометрической информации.

Наряду с этим в работе были получены следующие основные результаты:

1. Проведён анализ уязвимости современных голосовых биометрических систем к различным способам фальсификации индивидуальных голосовых биометрических характеристик человека. Показана необходимость совершенствования защиты модуля ввода биометрической информации от атак, использующих фальсификацию индивидуальных голосовых биометрических характеристик человека.
2. Разработан метод имитации атаки на устройство ввода биометрической информации, обеспечивающий автоматическое создание модели голоса для синтеза голосовых биометрических характеристик. Проведён численный эксперимент, показывающий необходимость дополнения существующих стандартов оценки эффективности аутентификации голосовыми биометрическими системами, оценкой устойчивости к спуфинг атакам.
3. В соответствии со сделанными выводами, разработана методика оценки эффективности аутентификации голосовыми биометрическими системами учитывающая воздействие различных видов спуфинг атак на модуль ввода биометрической информации. Проведены численные эксперименты, показывающие преимущества разработанной методики в сравнении с существующими аналогами.

4. На основе предложенной методики, с использованием языков программирования C++ и Python, разработан комплекс программных средств оценки эффективности аутентификации голосовыми биометрическими системами, позволяющий оценивать устойчивость к различным видам атак при проведении технологических испытаний или на этапе разработки системы.

5. Разработан метод противодействия спуфинг атакам, позволяющий значительно повысить устойчивость голосовых биометрических систем к различным методам спуфинг атак на модуль ввода биометрической информации. Проведены численные эксперименты, показывающие значительную редукцию уровня ошибки распознавания диктора при воздействии атаки на модуль биометрического ввода. Разработанный метод детектирования спуфинг атак занял второе место на международном конкурсе Automatic Speaker Verification Spoofing and Countermeasures (ASVspoof) Challenge 2015.

Список литературы

1. Аграновский А.В., Леднов Д.А. Теоретические аспекты алгоритмов обработки и классификации речевых сигналов. – Москва: Радио и связь, 2004. -164 с.
2. Алейник С.В., Матвеев Ю.Н., Раев А.Н. Метод оценки уровня клиппирования речевого сигнала // Научно-технический вестник информационных технологий, механики и оптики. - 2012. - №3 (79). - С. 79-83.
3. Белых И.Н., Капустин А.И., Козлов А.В., Лоханова А.И., Матвеев Ю.Н., Пеховский Т.С., Симончик К.К., Шулипа А.К. Система идентификации дикторов по голосу для конкурса NIST SRE 2010 // Информатика и её применения, 6:1 (2012), 91–98.
4. Болл Р.М. Руководство по биометрии / Р. М. Болл, Дж. Х. Коннел, Ш. Панканти, Н. К. Ратха, Э. У. Сеньор; пер. с англ. Н. Е. Агаповой. — М.: Техносфера, 2007. — 368 с.
5. ГОСТ 24.702-85. Единая система стандартов автоматизированных систем управления. Эффективность автоматизированных систем управления. Основные положения. М., 2009.
6. ГОСТ Р 50840-95. Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости. М., 1996.
7. ГОСТ Р ИСО 9000-2001. Системы менеджмента качества. Основные положения и словарь. М., 2004.
8. ГОСТ Р ИСО 9000-2011. Системы менеджмента качества. Основные положения и словарь. М., 2012.
9. ГОСТ Р ИСО/МЭК 19795-1-2007. Автоматическая идентификация. Идентификация биометрическая. Эксплуатационные испытания и

протоколы испытаний в биометрии. Часть 1. Принципы и структура. М., 2008.

10. ГОСТ Р ИСО/МЭК 19795-2-2008. Автоматическая идентификация. Идентификация биометрическая. Эксплуатационные испытания и протоколы испытаний в биометрии. Часть 2. Методы проведения технологического и сценарного испытаний. М., 2009.
11. ГОСТ Р ИСО/МЭК ТО 19795-3-2009. Автоматическая идентификация. Идентификация биометрическая. Эксплуатационные испытания и протоколы испытаний в биометрии. Часть 3. Особенности проведения испытаний при различных биометрических модальностях. М., 2010.
12. ГОСТ Р ИСО/МЭК ТО 19795-4-2010. Автоматическая идентификация. Идентификация биометрическая. Эксплуатационные испытания и протоколы испытаний в биометрии. Часть 4. Тестирование производительности биометрических систем. М., 2010.
13. Зиндер Е.З. Что такое «эффективность ИТ». Intelligent Enterprise, 2006, №8.
14. Капустин А.И., Симончик К.К. СИСТЕМА ВЕРИФИКАЦИИ ДИКТОРОВ ПО ГОЛОСУ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ CGP-SVM ПОДХОДА // Доклады 12-й Международной конференции «Цифровая обработка сигналов и ее применение», Инсвязьиздат, Выпуск XII-1, Москва, 2010, том 1, стр. 207-210.
15. Коваль С.Л., Лабутин П.В., Малая Е.В., Прощина Е.А. Идентификация дикторов на основе сравнения статистик основного тона голоса // Сборник трудов XV международной научной конференции «Информатизация и информационная безопасность правоохранительных органов». Москва, 2006.

16. Коваль С.Л., Раев А.Н., Лабутин П.В. Патент РФ 2230375 от 10.06.2004. «Метод распознавания диктора и устройство для его осуществления».

17. Козлов А.В., Кудашев О.Ю., Матвеев Ю.Н., Пеховский Т.С., Симончик К.К., Шулипа А.К. Система идентификации дикторов по голосу для конкурса NIST SRE 2012 // Тр. СПИИРАН, 25 (2013).

18. Матвеев Ю.Н. Технологии биометрической идентификации личности по голосу и другим модальностям // "Вестник МГТУ. Приборостроение", Спецвыпуск №2 "Биометрические технологии", 2012.

19. Матвеев Ю.Н., Симончик К.К. Система идентификации дикторов по голосу для конкурса NIST SRE 2010 // Труды 20-й межд. конф. по компьютерной графике и зрению «ГрафиКон’2010». Спб, 2010.

20. Пеховский Т.С., Лоханова А.И. Выбор UBM Модели с помощью Вариационного Байесовского Анализа для GMM-UBM Систем Распознавания Диктора // SPECOM-2011 Proceedings. 14th International conference on SPEECH and COMPUTER. - Kazan, Russia: Типография «ПАЛАДИН», 27-30 Sept. 2011, P. 327-332.

21. Проект BOSARIS Toolkit, [Электронный ресурс] - Режим доступа: https://sites.google.com/site/bosaristoolkit/, свободный. Яз. Англ. (дата обращения 03.02.2015).

22. Проект Festvox, Carnegie Mellon University, [Электронный ресурс] – Режим доступа: http://www.festvox.org/, свободный. Яз. Англ. (дата обращения 03.02.2015).

23. Проект Google Developers Charts, [Электронный ресурс] - Режим доступа: https://developers.google.com/chart/, свободный. Яз. Англ. (дата обращения 03.02.2015).

24. Проект JetBrains TeamCity, [Электронный ресурс] - Режим доступа: https://www.jetbrains.com/teamcity/, свободный. Яз. Англ. (дата обращения 03.02.2015).

25. Проект LIBLINEAR: A Library for Large Linear Classification. [Электронный ресурс] – Режим доступа: http://www.csie.ntu.edu.tw/~cjlin/liblinear/, свободный. Яз. Англ. (дата обращения 03.02.2015).

26. Проект OpenQuality. [Электронный ресурс] – Режим доступа: http://openquality.ru/, свободный. Яз. Англ. (дата обращения 03.02.2015).

27. Проект Python Unittest. [Электронный ресурс] – Режим доступа: https://docs.python.org/2/library/unittest.html, свободный. Яз. Англ. (дата обращения 03.02.2015).

28. Раев А.Н., Матвеев Ю.Н., Голощапова Т.И. Анализ влияния состояния наркотического опьянения на характеристики голосов дикторов // Научно-технический вестник информационных технологий, механики и оптики. - 2012. - №5 (81). - С. 87-93.

29. Рамишвили Г.С. Автоматическое опознавание говорящего по голосу. - М: Радио и связь, 1981. - 224 с.

30. Румшиский Л.З. Математическая обработка результатов эксперимента - Наука, 1971. - 192 с.

31. Ручай А.Н. Биометрика. Текстозависимая верификация диктора. Учебное пособие. 2012 - 105 с.

32. Симончик К. К. Метод и алгоритмы текстонезависимой верификации дикторов по голосу // LAP LAMBERT Academic Publishing GmbH & Co. KG, Saarbrucken, Germany, 2011, ISBN: 978-3-8433-1295-0, 188 с.

33. Alegre F., Amehraye A., and Evans N. Spoofing countermeasures to protect automatic speaker verification from voice conversion // Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), 2013.

34. Alegre F., Janicki A., and Evans N. Re-assessing the threat of replay spoofing attacks against automatic speaker verification // Proc. Int. Conf. of the Biometrics Special Interest Group (BIOSIG), 2014.

35. Alegre F., Vipperla R., Amehraye A., and Evans N. A new speaker verification spoofing countermeasure based on local binary patterns // Proc. Interspeech, 2013.

36. Alegre F., Vipperla R., Evans N. et al. Spoofing countermeasures for the protection of automatic speaker recognition systems against attacks with artificial signals // Proc. Interspeech, 2012.

37. Bimbot F., et al. A Tutorial on Text-Independent Speaker Verification // EURASIP Journal on Applied Signal Processing, 2004, №4, С.430-451.

38. Campbell J. and Higgins A. YOHO Speaker Verification database. [Электронный ресурс] – Режим доступа: http://www.ldc.upenn.edu/Catalog/catalogEntry.jsp?catalogId=LDC94S16, свободный. Яз. Англ. (дата обращения 23.09.2015).

39. Chistikov P., Korolkov E. Data-driven Speech Parameter Generation For Russian TexttoSpeech System. Computational Linguistics and Intellectual Technologies // Papers from the Annual International Conference ”Dialogue” (2012). Issue 11 (18) Volume 1 of 2. p. 103-111.

40. Chistikov P., Korolkov E., Talanov A. Combining HMM and unit selection tech nologies to increase naturalness of synthesized speech // Dialog-2013.

41. De Leon P., Hernaez I., Saratxaga I., Pucher M., and Yamagishi J. Detection of synthetic speech for the problem of imposture // Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), 2011.

42. De Leon P., Pucher M., Yamagishi J., Hernaez I., and Saratxaga I. Evaluation of speaker verification security and detection of HMM-based synthetic speech // IEEE Trans. Audio, Speech and Language Processing, vol. 20, no. 8, pp. 2280–2290, 2012.

43. Dehak N. Support Vector Machines versus Fast Scoring in the Low-Dimensional Total Variability Space for Speaker Verification // Proc. Interspeech, 2009, pp. 1559-1562.

44. Dehak N., Kenny P., Dehak R., Dumouchel P., Ouellet P. Front-end factor analysis for speaker verification // IEEE Transactions on Audio, Speech & Language Processing, 19:4(2011).

45. Doddington G., Liggett W., Martin A., Przybocki M. and Reynolds D. Sheep, goats, lambs and wolves: A statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation // ICSLP, November 1998.

46. Dutoit T., Holzapfel A., Jottrand M., Moinet A., Perez J., and Stylianou Y. Towards a voice conversion system based on frame selection // Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), 2007.

47. Evans N., Yamagishi J., and Kinnunen T. Spoofing and countermeasures for speaker verification: a need for standard corpora, protocols and metrics // IEEE Signal Processing Society Speech and Language Technical Committee Newsletter, 2013.

48. Fukada T., Tokuda K., Kobayashi T., and Imai S. An adaptive algorithm for mel-cepstral analysis of speech // Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), 1992.

49. Ganchev T., Fakotakis N., and Kokkinakis G. Comparative evaluation of various MFCC implementations on the speaker verification task // SPECOM 2005, Vol. 1, P. 191–194.

50. Hastie T., Tibshirani R., Friedman J. The EM algorithm, The Elements of Statistical Learning. // Springer, 2001,P. 236–243.

51. Helander E., Silen H., Virtanen T., and Gabbouj M. Voice conversion using dynamic kernel partial least squares regression // IEEE Trans. Audio, Speech and Language Processing, vol. 20, no. 3, pp. 806–817, 2012.

52. Hinton G., Osindero S., Teh Y. A fast learning algorithm for deep belief nets // Neural Computation, Vol. 18, pp. 1527–1554, Jul. 2006.

53. ISO/IEC 2382-37:2012 Information technology — Vocabulary — Part 37: Biometrics.

54. Jahangir A., Kenny P., Bhattacharya G., Stafylakis T. Development of CRIM System for the Automatic Speaker Verification Spoofing and Countermeasures Challenge 2015 // Proc. Interspeech, 2015.

55. Janicki A. Spoofing Countermeasure Based on Analysis of Linear Prediction Error // Proc. Interspeech, 2015.

56. Kenny P., Boulianne G., Ouelle P., Dumouchel P. Joint factor analysis versus eigenchannels in speaker recognition // IEEE Transactions on Audio, Speech and Language Processing, 15:4 (2007), C. 1435–1447.

57. Kenny P., Ouellet P., Dehak N., et al. A Study of Inter-Speaker Variability in Speaker Verification // IEEE Transactions on Audio, Speech and Language Processing, 16:5 (2008), C. 980-988.

58. Koval S., Labutin P., Raev A. Automatic Speaker Recognition using Formants-Based Nearest-Neighbour Distance Measure // Proceedings EUROSPEECH’95, Madrid, 1995.

59. Lau Y., Wagner M., Tran D. Vulnerability of speaker verification to voice mimicking // Intelligent Multimedia, Video and Speech Processing, 2004.

60. Longbiao W., Yoshida Y., Kawakami Y., Nakagawa S. Relative phase information for detecting human speech and spoofed speech // Proc. Interspeech, 2015.

61. Mariethoz J., Bengio S. Can a Professional Imitator Fool a GMM-Based Speaker Verification System? // IDIAP Research report IDIAP–RR 05-61. 2006.

62. Matejka P., Glembek O., Castaldo F., Alam J., Plchot O., Kenny P., Burget L., Cernocky J. FullCovariance UBM and Heavy-Tailed PLDA in i-vector Speaker Verification // Proc. ICASSP. (Prague, Czech Republic, May), 2011, C. 4828-4831.

63. Mensfield A.J. and Wayman J.L. Best Practices in Testing and Reporting Performance of Biometric Devices, Version 2.01 // NPL Report CMSC 14/02. 2002.

64. Nanxin C., Qian Y., Dinkel H., Chen B., Yu K. Robust Deep Feature for Spoofing Detection - The SJTU System for ASVspoof 2015 Challenge // Proc. Interspeech, 2015.

65. Novoselov S., Kozlov A., Lavrentyeva G., Simonchik K. and Shchemelinin V. STC Anti-spoofing Systems for the ASVspoof 2015 Challenge [Электронный ресурс] - Режим доступа: http://www.spoofingchallenge.org/asvspoof2015/STC.pdf, свободный. Яз. Англ. (дата обращения 23.09.2015).

66. Novoselov S., Pekhovsky T., Simonchik K. STC Speaker Recognition System for the NIST i-Vector Challenge // Proc. Odyssey 2014 - The Speaker and Language Recognition Workshop.

67. Osherove R., The Art of Unit Testing: with Examples in .NET - November 2013 ISBN 9781617290893 296 pages.

68. Patel T. , Hemant P. Combining Evidences from Mel Cepstral, Cochlear Filter Cepstral and Instantaneous Frequency Features for Detection of Natural vs. Spoofed Speech // Proc. Interspeech, 2015.

69. Pekhovsky T., Oparin I. Maximum Likelihood Estimations for Session Independent Speaker Modeling // SPECOM–2009. Proc. XIII Intern. Conf. «Speech and Computer». St.-Petersburg, 2009. P. 267–270.

70. Prodan A., Chistikov P., Talanov A. Voice building system for Russian TTS system “Vital Voice” // Proceedings of the Dialogue-2010 International Conference, No 9 (16), pp. 394-399.

71. Reynolds D. Experimental evaluation of features for robust speaker identification // IEEE Transactionson Speech and Audio Processing, 2:4 (1994), C. 639-643.

72. Roberts C. Biometric attack vectors and defences // Computers and Security. — 2007. Vol. 26, №1. — P. 14–25.

73. Saito D., Yamamoto K., Minematsu N., and Hirose K. One-to many voice conversion based on tensor representation of speaker space // Proc. Interspeech, 2011.

74. Sanchez J., Saratxaga I., Hernaez I., Navas E., and Erro D. A cross-vocoder study of speaker independent synthetic speech detection using phase information // Proc. Interspeech, 2014.

75. Sanchez J., Saratxaga I., Hernaez I., Navas E., Erro D. The AHOLAB RPS SSD Spoofing Challenge 2015 submission // Proc. Interspeech, 2015.

76. Shitao W., Chen S., Yu L., Wu X., Cai W., Liu Z., Li M. The SYSU System for the Interspeech 2015 Automatic Speaker Verification Spoofing and Countermeasures Challenge // arXiv:1507.06711, 2015.

77. Simonchik K., Pekhovsky T., Shulipa A., Afanasyev A. Supervized Mixture of PLDA Models for Cross-Channel Speaker Verification // Proc. Interspeech, 2012.

78. Simonchik K., Shchemelinin V. “STC SPOOFING” DATABASE FOR TEXT-DEPENDENT SPEAKER RECOGNITION EVALUATION // Proc. 4th International Workshop on Spoken Language Technologies for Under-resourced Languages (SLTU) - 2014, pp. 221-224.

79. Solomennik A., Chistikov P., Rybin S., Talanov A., Tomashenko N. Automation of New Voice Creation Procedure For a Russian TTS System // Vestnik MGTU. Priborostroenie, Special Issue 2 ”Biometric Technologies”, 29-32, 2013.

80. Stephane M. A Wavelet Tour of Signal // Proc. 3rd ed., Academic Press, Dec. 2008.

81. Stylianou Y. Voice transformation: A survey // Proc. Int. conference on acoustics, speech, and signal processing (ICASSP 2009), Taipei, Taiwan, April 2009, pp. 3585–3588.

82. 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, свободный. Яз. Англ. (дата обращения 23.09.2015).

83. The NIST Year 2010 Speaker Recognition Evaluation Plan. [Электронный ресурс] – Режим доступа:http://www.nist.gov/itl/iad/mig/upload/NIST_SRE10_evalplan-r6.pdf, свободный. Яз. Англ. (дата обращения 23.09.2015).

84. The NIST Year 2012 Speaker Recognition Evaluation Plan. [Электронный ресурс] – Режим доступа:

http://www.nist.gov/itl/iad/mig/upload/NIST_SRE12_evalplan-v17-r1.pdf, свободный. Яз. Англ. (дата обращения 23.09.2015).

85. Toda T., Black A., and Tokuda K. Voice conversion based on maximum-likelihood estimation of spectral parameter trajectory // IEEE Trans. Audio, Speech and Language Processing, vol. 15 no. 8, pp. 2222–2235, 2007.
86. Villalba J. and Lleida E. Preventing replay attacks on speaker verification systems // in IEEE Int. Carnahan Conf. on Security Technology (ICCST), 2011.
87. Villalba J., Miguel A., Ortega A., Lleida E. Spoofing Detection with DNN and One-class SVM for the ASVspoof 2015 Challenge // Proc. Interspeech, 2015.
88. Wu Z. , Kinnunen T., Chng E., Li H., and Ambikairajah E. A study on spoofing attack in state-of-the-art speaker verification: the telephone speech case // Proc. Asia-Pacific Signal Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012.
89. Wu Z., Chng E., Li H. Detecting Converted Speech and Natural Speech for anti-Spoofing Attack in Speaker Recognition // Proc. Interspeech, 2012.
90. Wu Z., Evans N., Kinnunen T., Yamagishi J., Alegre F., and Li H. Spoofing and countermeasures for speaker verification: A survey // Speech Communication, Vol. 66, no. 0, pp. 130– 153, 2015.
91. Wu Z., Gao S., Chng E., and Li H. A study on replay attack and anti-spoofing for text-dependent speaker verification // Proc. Asia-Pacific Signal Information Processing Association Annual Summit and Conference (APSIPA ASC), 2014.
92. Wu Z., Khodabakhsh A., Demiroglu C., Yamagishi J., Saito D., Toda T., and King S. SAS: A speaker verification spoofing database containing

diverse attacks // Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), 2015.

93. Wu Z., Kinnunen T., Chng E., and Li H. Text-independent F0 transformation with non-parallel data for voice conversion // Proc. Interspeech 2010, Makuhari, Japan, September 2010, pp.1732–1735.

94. Wu Z., Kinnunen T., Evans N., Yamagishi J., Sahidullah C., Sizov A. ASVspoof 2015: the First Automatic Speaker Verification Spoofing and Countermeasures Challenge, 2015, [Электронный ресурс] - Режим доступа: http://www.spoofingchallenge.org/is2015_asvspoof.pdf, свободный. Яз. Англ. (дата обращения 23.09.2015).

95. Wu Z., Virtanen T., Kinnunen T., Chng E., and Li H. Exemplarbased unit selection for voice conversion utilizing temporal information // Proc. Interspeech, 2013.

96. Wu Z., Xiao X., Chng E., and Li H. Synthetic speech detection using temporal modulation feature // Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), 2013.

97. Xiao X., Tian X., Du S., Xu H., Chng E., Li H. Spoofing Speech Detection Using High Dimensional Magnitude and Phase Features: the NTU Approach for ASVspoof 2015 Challenge // Proc. Interspeech, 2015.

98. Yamagishi J., Kobayashi T., Nakano Y., Ogata K., and Isogai J. Analysis of speaker adaptation algorithms for HMM-based speech synthesis and a constrained smaplr adaptation algorithm // IEEE Trans. Audio, Speech and Language Processing, vol. 17, no. 1, pp. 6683, 2009.

99. Yi L., Tian Y., He L., Liu J., Johnson M. Simultaneous Utilization of Spectral Magnitude and Phase Information to Extract Supervectors for Speaker Verification Anti-spoofing // Proc. Interspeech, 2015.

100. Ying G., Mitchell C., and Jamison L. Endpoint detection of isolated utterances based on modified teager energy measure // ICASSP, 1993, pp. 732-735.

101. Zetterholm E, Blomberg M, Elenius D A comparison between human perception and a speaker verification system score of a voice imitation // Proceedings of tenth australian international conference on speech science and technology, Macquarie University, Sydney, Australia, pp 393–397, 2004.

Приложение А. Акт внедрения

УТВЕРЖДАЮ

Исполнительный директор
ООО «Центр речевых технологий»,

И.В. Вересов

«6» октября 2015 г.

АКТ

о внедрении результатов диссертационной работы Щемелинина Вадима Леонидовича «Методика и комплекс средств оценки эффективности аутентификации голосовыми биометрическими системами»

Разработанные в диссертационной работе Щемелинина Вадима Леонидовича **«Методика и комплекс средств оценки эффективности аутентификации голосовыми биометрическими системами»** методика и комплекс программных средств оценки эффективности аутентификации голосовыми биометрическими системами, а также предложенные методы по совершенствованию защиты голосовых биометрических систем используются обществом с ограниченной ответственностью «Центр речевых технологий» в опытно–конструкторских разработках автоматических систем идентификации личности по голосу и производстве коммерческих продуктов: биометрическая платформа для подтверждения личности по голосу "VoiceKey", система поиска голосов мошенников в больших архивах аудиозаписей "VoiceGrid X", мультимодальная система биометрического поиска и криминалистического учёта "VoiceGrid".

Исключительные права на созданные результаты интеллектуальной деятельности (методы, алгоритмы, программное обеспечение и т.д.) принадлежат обществу с ограниченной ответственностью «Центр речевых технологий».

Директор НИД
ООО «Центр речевых технологий»,
кандидат технических наук К.Е. Левин