УДК 004.001

МЕТОДЫ ЗАЩИТЫ В СОВРЕМЕННЫХ СИСТЕМАХ ГОЛОСОВОЙ АУТЕНТИФИКАЦИИ

Статья поступила в редакцию 15.04.2022, в окончательном варианте – 29.08.2022.

Евсюков Михаил Витальевич, Кубанский государственный технологический университет, 350072, Российская Федерация, г. Краснодар, ул. Московская, 2,
аспирант, ORCID: 0000-0001-7101-6251, e-mail: michael.evsyukov@gmail.com
Путято Михаил Михайлович, Кубанский государственный технологический университет, 350072, Российская Федерация, г. Краснодар, ул. Московская, 2,
кандидат технических наук, доцент, ORCID: 0000-0003-0414-6034, e-mail: putyato.m@gmail.com
Макарян Александр Самвелович, Кубанский государственный технологический университет, 350072, Российская Федерация, г. Краснодар, ул. Московская, 2,
кандидат технических наук, доцент, ORCID: 0000-0002-1801-6137, e-mail: msanya@yandex.ru
Немчинова Валерия Олеговна, Кубанский государственный технологический университет, 350072, Российская Федерация, г. Краснодар, ул. Московская, 2,
ассистент, ORCID: 0000-0002-4428-7128, e-mail: nemchinova.valeriya@yandex.ru

Вместе со стремительным развитием и широким распространением голосовых интерфейсов всё более актуальной становится проблема повышения безопасности систем голосовой аутентификации. В то время как алгоритмы распознавания личности по голосу хорошо изучены и демонстрируют высокую надёжность при проверке их эффективности живыми людьми, современные системы голосовой аутентификации подвержены ряду уязвимостей. В первую очередь, это связано с повсеместной распространённостью недорогой и высококачественной техники, предназначенной для записи и воспроизведения звука. Данный факт предоставляет злоумышленникам мощные инструменты для реализации атак на системы голосовой аутентификации. Как правило, цель злоумышленника состоит в прохождении аутентификации в системе под видом другого лица. Действия, направленные на достижение этой цели, называются спуфингом. В данной статье описаны основные голосовые характеристики, применяемые при реализации систем голосовой аутентификации, приведена актуальная классификация алгоритмов распознавания личности по голосу, описаны существующие метрики оценки эффективности систем голосовой аутентификации и изложены существующие подходы к классификации методов спуфинга. Кроме того, усовершенствована классификация контрмер против спуфинга и выделены перспективные направления будущих исследований в области аутентификации по голосу.

Ключевые слова: биометрия, искусственный интеллект, машинное обучение, информационная безопасность, защита информации, аутентификация, спуфинг

PROTECTION METHODS IN MODERN VOICE AUTHENTICATION SYSTEMS

The article was received by the editorial board on 15.04.2022, in the final version – 29.08.2022.

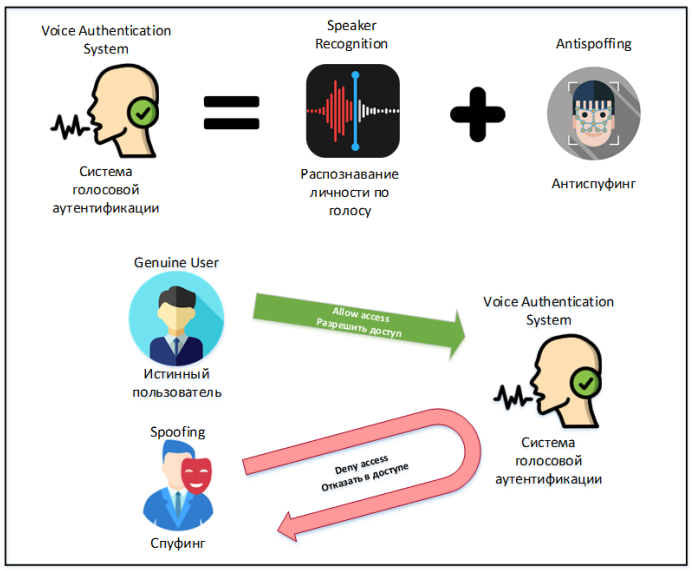
Evsyukov Michael V., Kuban State Technological University, 2 Moskovskaya St., Krasnodar, 350072, Russian Federation,
graduate student, ORCID: 0000-0001-7101-6251, e-mail: michael.evsyukov@gmail.com
Putyato Michael M., Kuban State Technological University, 2 Moskovskaya St., Krasnodar, 350072, Russian Federation,
Cand. Sci (Engineering), Associate Professor, ORCID: 0000-0001-9974-7144, e-mail: putyato.m@gmail.com
Makaryan Alexander S., Kuban State Technological University, 2 Moskovskaya St., Krasnodar, 350072, Russian Federation,
Cand. Sci (Engineering), Associate Professor, ORCID: 0000-0002-1801-6137, e-mail: msanya@yandex.ru
Nemchinova Valeriya O., Kuban State Technological University, 350072, Russian Federation, 2 Moskovskaya St., Krasnodar,
Assistant, ORCID: 0000-0002-4428-7128, e-mail: nemchinova.valeriya@yandex.ru

The problem of improving the security of voice authentication systems is becoming increasingly important due to the rapid development and widespread use of voice interfaces. Although voice recognition algorithms are well-studied and demonstrate high reliability when tested by live people, modern voice authentication systems are subject to a number of vulnerabilities. First of all, this is due to the access to affordable and high-quality devices for recording and playing sound. This fact provides attackers with powerful tools to implement attacks on voice authentication systems. As a rule, the attacker's purpose is to authenticate in the system under the guise of another person. Actions

aimed at achieving this goal are called spoofing. This article describes the main voice characteristics used in the implementation of voice authentication systems, provides an up-to-date classification of voice recognition algorithms, describes existing metrics for evaluating the effectiveness of voice authentication systems, and outlines existing approaches to classifying spoofing methods. In addition, the classification of countermeasures against spoofing has been improved and promising directions for future research in the field of voice authentication have been identified.

Keywords: biometrics, artificial intelligence, machine learning, information security, authentication, spoofing, liveness detection

Graphical annotation (Графическая аннотация)



Введение. Согласно данным Google, 500 миллионов пользователей ежемесячно используют Google Assistant [1]. Apple утверждает, что голосовой помощник Siri ежемесячно обрабатывает 25 миллиардов запросов [2]. Простота применения и экономия времени – основные причины, по которым голосовые помощники набирают популярность. Кроме того, появление широкого ассортимента умных устройств и стремительное развитие интернета вещей (IoT) делают голосовой интерфейс ещё более востребованным, поскольку он способен предоставить наиболее комфортный пользовательский опыт. Возможность управления голосом реализована, например, в смарт-колонке «Яндекс.Станция», автомобилях Tesla, а также в различных системах типа «умный дом».

Таким образом, голосовые помощники вошли в повседневную жизнь многих пользователей, и следующим естественным шагом является их внедрение в платёжные системы и банкинг. Основным драйвером развития голосовых решений является персонализация. Это связано с тем, что голосовое взаимодействие способно предоставить ценные сведения о потребностях и поведении клиентов, что позволяет банкам и FinTech-компаниям предложить услуги, наиболее полно соответствующие ожиданиям конкретного пользователя.

В результате опроса, проведённого Business Insider Intelligence в 2017 году в США, 8 % респондентов заявили, что использовали голосовые команды для покупки товаров, оплаты счетов и выполнения P2P-транзакций. Согласно прогнозам, к 2022 году количество пользователей голосовых интерфейсов вырастет до 31 % взрослого населения США [3].

Высокая потребительская ценность голосовых платежей стимулирует банки и таких платёжных провайдеров, как PayPal, Amazon, Apple и Google к развитию технологий искусственного интеллекта, специализированных на обработке голоса.

Однако проблемы информационной безопасности – основное препятствие, которое не позволяет голосовым платежам завоевать полное доверие со стороны банков и пользователей. Для того чтобы они стали такими же естественными, как взаимодействие с продавцом или сотрудником банка, необходимо усовершенствовать существующие методы защиты и аутентификации [3].

Алгоритмы подтверждения личности человека по голосу хорошо изучены, удобны в использовании и применимы как для непрерывной, так и для разовой аутентификации. Однако из-за широкого распространения недорогих устройств записи и воспроизведения звука они подвержены спуфингу, т.е. уязвимы к действиям злоумышленников, направленным на выдачу себя за другого человека. В связи с этим разработка и изучение способов противодействия спуфингу является основным направлением развития систем голосовой аутентификации.

Целью данной статьи является рассмотрение современного состояния исследований в области голосовой аутентификации. Далее будет описан концептуальный подход к голосовой аутентификации, перечислены основные голосовые характеристики, применяемые при реализации систем голосовой аутентификации, приведена актуальная классификация алгоритмов распознавания личности по голосу, описаны существующие метрики оценки эффективности систем голосовой аутентификации и изложены существующие подходы к классификации методов спуфинга. Кроме того, в рамках данного исследования усовершенствована классификация контрмер против спуфинга и выделены перспективные направления будущих исследований в области голосовой аутентификации.

Общая характеристика голосовой аутентификации. Голосовая аутентификация – динамический метод биометрической аутентификации, использующий уникальные характеристики человеческого голоса в качестве признака, позволяющего распознать субъекта и подтвердить его личность [4].

Концептуальная схема современного механизма голосовой аутентификации представлена на рисунке 1.

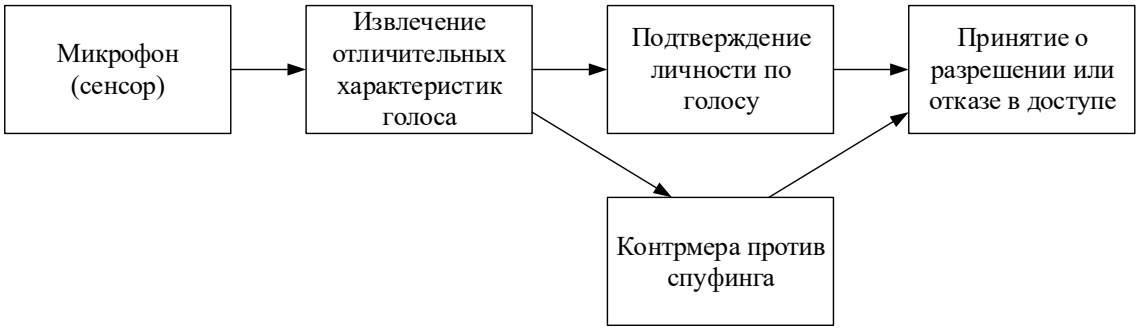


Рисунок 1 – Концептуальная схема современного механизма голосовой аутентификации

Как было упомянуто выше, голосовая аутентификация хорошо изучена и демонстрирует высокую эффективность, если при её тестировании проводить испытания только с живыми людьми. Однако её основным недостатком является уязвимость к спуфингу. В связи с этим система голосовой аутентификации должна включать в себя дополнительный механизм противодействия спуфингу, который называется контрмерой. Задача контрмеры – зафиксировать факт того, что система подвергается спуфинг-атаке.

Цикл функционирования любой системы биометрической аутентификации, в том числе голосовой, состоит из двух режимов: регистрация и верификация пользователя.

На этапе регистрации происходит сбор голосовых характеристик пользователя и формирование «голосового отпечатка», т.е. эталона биометрических характеристик, идентифицирующего пользователя.

На этапе верификации происходит предъявление пользователем его голосовых характеристик и сравнение их с хранимым в базе эталоном. Если в ходе сравнения подтверждается, что предъявляемые характеристики принадлежат заявленному пользователю, то пользователю предоставляется доступ. В противном случае ему отказывается в доступе.

В зависимости от ограничений, накладываемых на фразу, произносимую пользователем в процессе аутентификации, выделяют два вида подтверждения личности по голосу: текстонезависимое и текстозависимое.

Текстонезависимое подтверждение личности позволяет использовать произвольные фразы при регистрации и верификации пользователя. Преимуществом данного подхода является гибкость, однако для корректной работы он требует использования более длинных фраз по сравнению с текстозависимым подтверждением личности. Кроме того, как будет описано ниже, данный подход более эффективен против некоторых видов спуфинга.

Текстозависимое подтверждение личности предусматривает использование фиксированной фразы. Основное преимущество таких методов заключается в том, что они позволяют использовать фразы меньшей длины при регистрации и верификации пользователя.

Текстозависимая аутентификация тесно связана с такой задачей обработки голоса как распознавание речи, которая подразумевает выделение текста из речи. Решение данной задачи реализуется во всех системах голосового управления.

Оценка эффективности систем голосовой аутентификации. Для систем голосовой аутентификации применяются общие метрики оценки эффективности биометрических систем: вероятность ошибочного допуска, вероятность ошибочного отказа, кривая компромисса обнаружения ошибок и равная вероятность ошибки [5].

Вероятность ошибочного допуска (FAR) отражает долю спуфинговых атак, которые ошибочно верифицируются системой как истинные пользователи:

$$FAR = \frac{FA}{TA}, \quad (1)$$

где *FA* – число ошибочных допусков; *TA* – общее число попыток аутентификации.

Вероятность ошибочного отказа (FRR) отражает долю истинных пользователей, которым система ошибочно отказала в доступе:

$$FRR = \frac{FR}{TA}, \quad (2)$$

где *FR* – число ошибочных отказов в допуске; *TA* – общее число попыток аутентификации.

Обычно в процессе работы система аутентификации оценивает степень уверенности (вероятность) в том, что предъявленные биометрические характеристики принадлежат заявленному субъекту. В связи с этим имеется возможность настройки порогового значения степени уверенности. Если при попытке аутентификации степень уверенности системы больше порогового значения, то человеку разрешается доступ, а иначе – запрещается.

В зависимости от выбранного порогового значения степени уверенности, вероятность ошибочного допуска и вероятность ошибочного отказа меняют свои значения. Перечень возможных соотношений этих значений представлен кривой компромисса обнаружения ошибок.

Например, на рисунке 2 представлены кривые компромисса обнаружения ошибок для одной из систем голосовой аутентификации, участвующей в ASVSpoof 2015, при воздействии на систему каждым из 10 способов проведения спуфинг-атаки, используемом в конкурсе [6].

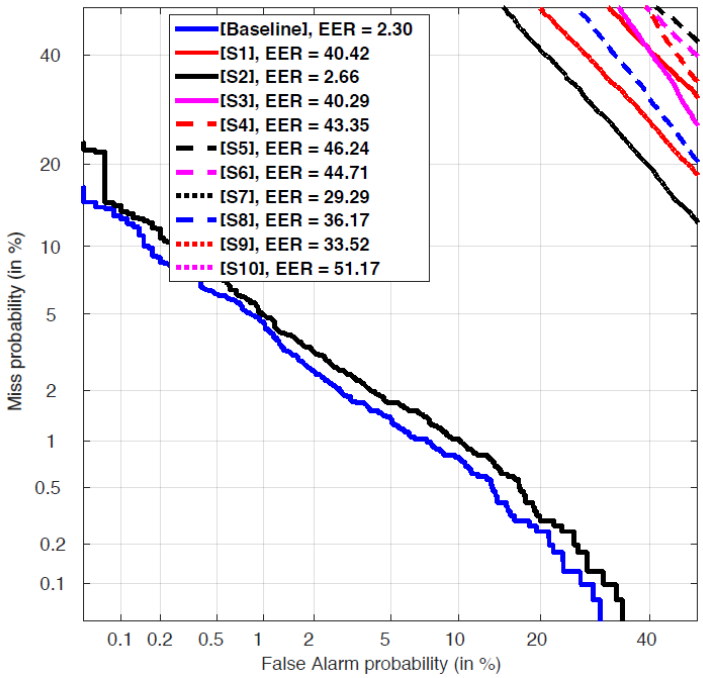


Рисунок 2 – Примеры кривых компромисса обнаружения ошибок

В качестве численного значения оценки эффективности системы аутентификации используется вероятность равной ошибки (EER) – которая соответствует точке на кривой, при котором вероятность ошибочного допуска равна вероятности ошибочного отказа.

Однако, как показано на рисунке 1, специфика работы систем голосовой аутентификации состоит в совместной работе двух классификаторов, обрабатывающих исходные данные: подсистемы подтверждения личности пользователя и контрмеры. Изначально их производительности оценивались независимо друг от друга, но в работе [7] был предложен более эффективный метод их совместной оценки – функция тандемной оценки стоимости обнаружения (t-DCF). Данная метрика хорошо зарекомендовала себя в ходе конкурса ASVSpoof, в котором она используется, начиная с 2019 года.

Наравне с выбором метрики оценки важное значение имеют условия проведения экспериментального исследования эффективности системы голосовой аутентификации.

В зависимости от возможностей доступа злоумышленника в систему выделяют два вида испытаний:

- испытания с логическим доступом (over-the-wire), при проведении которых не предусматривается использование сенсора (микрофона), и данные загружаются в цифровом виде напрямую в систему. Данный подход упрощает проведение атаки условному злоумышленнику и преимущественно используется для оценки эффективности таких видов спуфинга, как синтез и преобразование речи;
- испытания с физическим доступом (over-the-air), при проведении которых предусматривается взаимодействие с системой через сенсор (микрофон).

Также важно учитывать, что результат оценки эффективности системы голосовой аутентификации с физическим доступом подвержен влиянию следующих факторов:

- соотношение сигнал – шум;
- ревебрация;
- характеристики используемого микрофона;
- свойства помещения, в котором проводится испытание;
- качество записывающей и воспроизводящей аппаратуры, используемой при спуфинге.

Отличительные голосовые характеристики. Отличительные голосовые характеристики – это значимые особенности, извлекаемые из необработанного голосового сигнала, идентифицирующие человека.

Имеющиеся исследования свидетельствуют о том, что выбор отличительных голосовых характеристик имеет не меньшее влияние на результат работы системы голосовой аутентификации, чем выбор классификатора [6].

Для эффективного использования при реализации аутентификации и контрмер извлекаемые характеристики должны обладать следующими свойствами:

- большая вариативность у разных пользователей и малая вариативность у одного пользователя;
- устойчивость к искажениям и шуму;
- частая встречаемость в речи;
- лёгкость измерения;
- сложность подделки;
- независимость от состояния здоровья человека;
- неизменяемость у человека с течением времени.

Наибольшую распространённость получили кратковременные спектральные характеристики, при расчёте которых используются фрагменты речевого сигнала длиной 20–30 миллисекунд. Данные характеристики заключают в себе информацию о тембре и особенностях голосового тракта человека.

Кратковременные спектральные характеристики обладают следующими преимуществами:

- простота извлечения;
- потребность в небольшом объёме данных;
- независимость от текста и языка;
- возможность эффективной обработки.

По сравнению с более высокоуровневыми поведенческими характеристиками речи, кратковременные спектральные характеристики менее устойчивы к шуму и канальным искажениям, однако они гораздо лучше подходят для практической реализации системы голосовой аутентификации.

Механизм извлечения большинства таких характеристик основывается на дискретном (DFT), а точнее, на быстром преобразовании Фурье (FFT) [8]. Однако информация, которую содержит амплитудный спектр голосового сигнала, полученный при помощи FFT, избыточна. Поэтому при голосовой аутентификации используют другие характеристики, которые содержат наиболее существенную для задачи обработки голоса информацию, но имеют меньшую размерность и тем самым обеспечивают более простую обработку.

Наибольшее распространение получили мел-кепстральные частотные коэффициенты (MFCC) [9], использующие фильтр, учитывающий особенности восприятия звуков человеком (психоакустику), логарифмическое сжатие и дискретное косинусное преобразование.

Также используются методы оценки спектра, альтернативные DFT, например, линейные предсказательные частотные коэффициенты (LPCC) [10], основанные на вычислительной процедуре линейного предсказания.

Существует большое количество исследований, направленных на оценку эффективности применения различных характеристик. При этом установлено, что совместное использование характеристик, основанных на разном математическом аппарате, позволяет повысить эффективность работы системы.

Также существуют труды, рассматривающие возможность использования временных спектральных характеристик и просоидальные характеристики, однако они не получили такого широкого распространения, как кратковременные спектральные характеристики.

Алгоритмы подтверждения личности по голосу. При регистрации пользователя полученные голосовые характеристики используются для тренировки распознавателя голоса. Распознаватель голоса – это математическая модель, используемая для сравнения голоса диктора, проходящего верификацию, с эталонными характеристиками заявленного субъекта [11].

В зависимости от подхода к тренировке моделей их можно разделить на генеративные (классические) и дискриминативные. В то время как генеративные модели моделируют распределение характеристик речи конкретного пользователя, дискриминативные модели аппроксимируют границу между голосами разных людей в гиперпространстве характеристик.

В свою очередь, генеративные модели можно разделить на шаблонные (непараметрические) и вероятностные (параметрические) модели.

Шаблонные модели рассматривают предъявляемый вектор голосовых характеристик как неточную копию эталонного вектора пользователя. Исходя из этого рассчитывается степень отличия между этими векторами, на основании которой и определяется успешность аутентификации.

Вероятностные модели рассматривают человеческий голос как некоторое распределение характеристик, имеющее определённую функцию плотности вероятности. На этапе обучения производится аппроксимация параметров данной функции. На этапе верификации выполняется оценка вероятности того, что параметры речи верифицируемого пользователя соответствуют эталонной модели [11].

В таблице 1 представлены наиболее распространённые генеративные модели.

Среди перечисленных алгоритмов модель гауссовой смеси стала де факто стандартом, и эффективность других алгоритмов распознавания личности по голосу сравнивается именно с ней.

Таблица 1 – Классификация наиболее широко распространённых генеративных моделей, используемых для голосовой аутентификации

	Текстозависимая аутентификация	**Текстонезависимая аутентификация**
Шаблонные модели	Динамическая трансформация временной шкалы	Векторное квантование
Вероятностные модели	Скрытая марковская модель	Модель гауссовой смеси

К дискриминативным моделям относятся нейронные сети и машины опорных векторов. Преимущество нейронных сетей заключается в том, что они способны объединить процесс извлечения характеристик и распознавания личности. Машины опорных векторов – также широко применяемый инструмент подтверждения личности по голосу, обладающий хорошей обобщающей способностью [12]. Некоторые системы аутентификации объединяют в себе сразу несколько алгоритмов подтверждения личности по голосу, что позволяет повысить общую эффективность [13].

Разновидности спуфинга. Под спуфингом понимаются действия злоумышленника, направленные на успешную аутентификацию в системе под видом другого лица. Благодаря широкому распространению качественного записывающего и воспроизводящего звукового оборудования, системы голосовой аутентификации в существенной мере подвержены спуфингу.

Существуют следующие основные виды спуфинга [12]:

1. Выдача себя за другое лицо. Данный вид спуфинга реализуется посредством подражания одним человеком голосовым характеристикам другого человека. Выдача себя за другое лицо отличается от других видов спуфинга тем, что для его реализации злоумышленник не использует вспомогательных технических средств и методов. В связи с этим противодействие этому виду спуфинга не требует дополнительных контрмер и реализуется за счёт качественной работы системы голосовой верификации.

2. Запись речи (атака повторным воспроизведением). Запись речи – простой и эффективный вид спуфинга, который, по мнению многих исследователей, представляет наиболее серьёзную угрозу системам голосовой аутентификации. Его реализация заключается в записи фрагмента речи человека с целью его последующего предъявления системе аутентификации.

3. Преобразование речи. Преобразование речи подразумевает использование специализированных программных средств, изменяющих речь человека таким образом, чтобы она стала похожей на речь другого человека.

Оценка сопротивляемости различных контрмер данному виду спуфинга была предметом конкурса ASVSpoof 2015, в ходе которого использовались следующие алгоритмы преобразования речи [13]:

- выбор фрагментов речи на основе образца для преобразования голоса с использованием временной информации;
- подстраивание первого мел-кепстрального коэффициента под значение человека-цели (один из простейших алгоритмов);

- алгоритмы, использующие модель гауссовой смеси (самые распространённые);
- алгоритм, основанный на тензорном представлении пространства признаков пользователя;
- алгоритм, использующий регрессию частичным методом наименьших квадратов, основанный на ядре пространства.

4. Синтез речи. Данный метод подразумевает генерацию искусственного голоса на основе произвольного текста, обладающего характеристиками голоса определённого человека.

Оценка сопротивляемости различных контрмер данному виду спуфинга была предметом конкурса ASVSpoof 2015, в ходе которого использовались следующие алгоритмы синтеза речи [13]:

- синтез речи, основанный на выборе и конкатенации отдельных речевых элементов (данный вид спуфинга оказался наиболее эффективным);
- статистический синтез речи, основанный на скрытой марковской модели.

Перспективным методом синтеза и преобразования речи является технология дипфейк, подразумевающая использование генеративно-состязательных нейронных сетей. Оценка способности контрмер противостоять дипфейк-атакам является одной из задач конкурса ASVSpoof 2021.

5. Замаскированные атаки на системы обработки человеческого голоса, использующие особенности восприятия звуков человеком.

В исследовании [14] рассматриваются 4 способа преобразования записи голоса таким образом, чтобы она стала непонятной для человека, но чтобы её существенные акустические голосовые признаки остались неизменными и запись могла пройти систему голосовой аутентификации или быть обработанной системой распознавания речи.

Контрмеры против спуфинга. Общая классификация контрмер против спуфинга представлена в работе [12]. Мы предлагаем её дополненную версию.

1. Интерактивные контрмеры. Интерактивные контрмеры подразумевают явное взаимодействие пользователя с системой в ходе аутентификации. Как правило, при их использовании, система генерирует случайный текст, который нужно прочитать пользователю. Для аутентификации пользователя используется алгоритм текстонезависимого подтверждения личности, а правильность прочтения текста проверяется алгоритмом распознавания речи.

Данный тип контрмер показывает высокую эффективность противодействия наиболее опасному виду спуфинга – атакам повтором. Это связано с тем, что у злоумышленника, как правило, отсутствует возможность заранее записать речь пользователя таким образом, чтобы из её фрагментов можно было оперативно составить случайную фразу.

2. Контрмеры, использующие акустические особенности сгенерированного или синтезированного голоса.

Данное семейство контрмер концентрируется на извлечении из записи голоса несовершенств, свидетельствующих о том, что отрывок речи был получен при помощи методов синтеза или преобразования речи. Именно такие контрмеры являлись объектом исследования ASVSpoof 2015 [13].

Методы, относящиеся к этому типу, так же как и методы верификации по голосу, преимущественно опираются на использование кратковременных спектральных характеристик. Наибольшее распространение получили такие классификаторы, как модель гауссовой смеси, машины опорных векторов и искусственные нейронные сети.

Одной из особенностей данных контрмер является то, что оценку их эффективности можно проводить в формате исследования с логическим доступом.

3. Методы обнаружения живого голоса, основанные на особенностях речевого тракта человека. Поскольку спуфинг подразумевает использование звуковоспроизводящей аппаратуры, задачу голосовой аутентификации можно представить, как совокупность двух следующих задач:

- подтвердить личность человека по голосовым характеристикам (верификация);
- подтвердить, что источником голоса является живой человек (контрмера).

Данное семейство контрмер опирается на особенности речевого тракта человека, приводящие к возникновению акустических эффектов, которые затруднительно записать и воспроизвести при помощи искусственных средств.

Примеры методов обнаружения живого голоса, основанных на особенностях речевого тракта человека:

- обнаружение живого голоса, основанное на хлопающем шуме, вызванном дыханием человека [14];
- обнаружение живого голоса для аутентификации на смартфонах, основанное на локализации фонем;
- обнаружение живого голоса, основанное на артикулярных жестах.

4. Методы обнаружения живого голоса, основанные на особенностях воспроизведения звука громкоговорителем. На данный момент широкому кругу пользователей доступны устройства записи и

воспроизведения звука, способные копировать голос человека с очень высоким качеством, которые продолжают развиваться. При использовании приведённых ранее голосовых характеристик (например, мел-кепстральных частотных коэффициентов) искусственный голос крайне затруднительно отличить от живого, о чём свидетельствуют результаты конкурса ASVSpoof 2017 [15].

В связи с этим высказываются предложения по применению других демаскирующих признаков, позволяющих понять, что при попытке аутентификации звук воспроизводится искусственным громкоговорителем. Например, в работе [16] предлагается использовать магнитное поле для того, чтобы отличить громкоговоритель от живого человека.

5. Совместное использование разных биометрических методов. Данный подход подразумевает повышение эффективности системы аутентификации и устойчивости к спуфингу за счёт использования двух или более несвязанных биометрических характеристик. Например, в работе [17] предлагается бимодальная система подтверждения личности, использующая модель гауссовой смеси с универсальной фоновой моделью для голосовой аутентификации и систему верификации лица, при помощи признаков Гэбора и линейного дискриминантного анализа.

Кроме того, в качестве дополнительной категории контрмер можно выделить методы, повышающие качество распознавания личности по голосу за счёт использования множества микрофонов.

Заключение. В данной статье были перечислены основные отличительные голосовые характеристики, используемые при реализации систем голосовой аутентификации, и описаны ключевые особенности их применения. Приведена актуальная классификация алгоритмов распознавания личности по голосу в зависимости от фиксации конкретной фразы при аутентификации и вида применяемого математического аппарата. Описаны существующие метрики оценки эффективности систем голосовой аутентификации: показаны как общие метрики оценки эффективности систем аутентификации, так и уникальная метрика, используемая для систем голосовой аутентификации. Изложены существующие подходы к классификации методов спуфинга.

Кроме того, усовершенствована классификация контрмер против спуфинга и выделены перспективные направления будущих исследований в области голосовой аутентификации.

Библиографический список

1. Eadicco, L. Google just revealed that half a billion people around the world are using the Google Assistant as it battles with Amazon to conquer the smart home / L. Eadicco // Insider. – Режим доступа: https://www.businessinsider.com/google-assistant-500-million-users-challenges-amazon-alexa-2020-1, свободный. – Заглавие с экрана. – Яз. англ. (дата обращения: 20.01.2022).

2. Kinsella, B. Apple Still in Holding Pattern on Voice, Siri Used 25 Billion Times Per Month But New Features Limited / B. Kinsella // Voicebot.ai. – Режим доступа: https://voicebot.ai/2020/06/22/apple-still-in-holding-pattern-on-voice-siri-used-25-billion-times-per-month-but-new-features-limited/, свободный. – Заглавие с экрана. – Яз. англ. (дата обращения: 20.01.2022).

3. Dyke, D. V. Soon nearly a third of US consumers will regularly make payments with their voice / D.V. Dyke // Insider. – Режим доступа: https://www.businessinsider.com/the-voice-payments-report-2017-6#:~:text=Voice%20payments%20are%20catching%20on,of%20US%20adults%20by%202022, свободный. – Заглавие с экрана. – Яз. англ. (дата обращения: 20.01.2022).

4. Ravika, N. An Overview of Automatic Speaker Verification System / N. Ravika // Intelligent Computing and Information and Communication : Proceedings of 2nd International Conference, ICICC 2017, 2–4 August 2017. – Pune, India, 2017. – P. 603–610.

5. El-Abed, M. Evaluation of Biometric Systems / M. El-Abed, C. Charrier // New Trends and Developments in Biometrics. – 2012. – P. 149–169.

6. Wu, Z. ASVspoof 2015: the First Automatic Speaker Verification Spoofing and Countermeasures Challenge / Z. Wu, T. Kinnunen, N. Evans, J. Yamagishi, C. Hanilçi, M. Sahidullah, A. Sizov // 16th Annual Conference of the International Speech Communication Association (Interspeech 2015). – Dresden, Germany, 2015.

7. Kinnunen, T. t-DCF: a Detection Cost Function for the Tandem Assessment of Spoofing Countermeasures / T. Kinnunen, K. Lee, H. Delgado, N. Evans, M. Todisco, et al. // Speaker Odyssey 2018. The Speaker and Language Recognition Workshop, 17th August 2018. – Les Sables-d'Olonne, France, 2018. – Режим доступа: https://hal.inria.fr/hal-01880306/document, свободный. – Заглавие с экрана. – Яз. англ. (дата обращения: 20.01.2022).

8. Oppenheim, A. Discrete-Time Signal Processing. Second edition / A. Oppenheim, R. Schafer, J. Buck. – New Jersey : Prentice Hall, 1999. – 893 p.

9. Deller, J. Discrete-Time Processing of Speech Signals. Second edition / J. Deller, J. Hansen, J. Proakis. – New York : IEEE Press, 2000. – 936 p.

10. Huang, X. Spoken Language Processing: a Guide to Theory, Algorithm, and System Development / X. Huang, A. Acero, H.-W. Hon. – New Jersey : Prentice-Hall, 2001. – 935 p.

11. Kinnunen, T. An Overview of Text-Independent Speaker Recognition: from Features to Supervectors / T. Kinnunen, H. Li // Speech Communication. – 2010. – № 52. – P. 12–40.

12. Hao, B. Voice Liveness Detection for Medical Devices / B. Hao, X. Hei // Design and Implementation of Healthcare Biometric Systems. – 2019. – P. 109–136.

13. Wu, Z. ASVspoof: the Automatic Speaker Verification Spoofing and Countermeasures Challenge / Z. Wu, J. Yamagishi, T. Kinnunen, C. Hanilc, M. Sahidullah, A. Sizov, N. Evans, M. Todisco // IEEE Journal of Selected Topics in Signal Processing. – 2017. – Vol. 11, № 4. – P. 588–604.

14. Abdullah, H. Practical Hidden Voice Attacks against Speech and Speaker Recognition Systems / H. Abdullah, W. Garcia, C. Peeters, P. Traynor, K. Butler, J. Wilson // The Network and Distributed System Security Symposium, NDSS 2019, 24–27 February 2019. – San Diego, USA, 2019. – Режим доступа: https://hal.inria.fr/hal-01880306/document, свободный. – Заглавие с экрана. – Яз. англ. (дата обращения: 20.01.2022).

15. Kinnunen, T. // The ASVspoof 2017 Challenge: Assessing the Limits of Replay Spoofing Attack Detection / T. Kinnunen, M. Sahidullah, H. Delgado, M. Todisco, N. Evans, J. Yamagishi, K. A. Lee // 19th Annual Conference of the International Speech Communication Association (Interspeech 2018). – Stockholm, Sweden, 2018.

16. Li, L. A study on replay attack and anti-spoofing for automatic speaker verification / L. Li, Y. Chen, D. Wang // 18th Annual Conference of the International Speech Communication Association (Interspeech 2017). – Stockholm, Sweden, 2018. – P. 92–96.

17. Usoltsev, A. Full Video Processing for Mobile Audio-Visual Identity Verification / A. Usoltsev, D. Petrovska-Delacrétaz, K. Houssemeddine // Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2016). – Rome, Italy, 2016. – P. 552–557.

References

1. Eadicco, L. Google just revealed that half a billion people around the world are using the Google Assistant as it battles with Amazon to conquer the smart home. *Insider.* Available at: https://www.businessinsider.com/google-assistant-500-million-users-challenges-amazon-alexa-2020-1 (accessed 20.01.2022).

2. Kinsella, B. Apple Still in Holding Pattern on Voice, Siri Used 25 Billion Times Per Month But New Features Limited. *Voicebot.ai.* Available at: https://voicebot.ai/2020/06/22/apple-still-in-holding-pattern-on-voice-siri-used-25-billion-times-per-month-but-new-features-limited (accessed 20.01.2022).

3. Dyke, D. V. Soon nearly a third of US consumers will regularly make payments with their voice. *Insider.* Available at: https://www.businessinsider.com/the-voice-payments-report-2017-6#:~:text=Voice%20payments%20are%20catching%20on,of%20US%20adults%20by%202022 (accessed 20.01.2022).

4. Ravika, N. An Overview of Automatic Speaker Verification System. *Intelligent Computing and Information and Communication. Proceedings of 2nd International Conference. ICICC.* Pune, India, 2017, pp. 603–610.

5. El-Abed, M., Charrier, C. *Evaluation of Biometric Systems. New Trends and Developments in Biometrics*, 2012, pp. 149–169.

6. Wu, Z., Kinnunen, T., Evans, N., Yamagishi, J., Hanilçi, C., Sahidullah, M., Sizov, A. ASVspoof 2015: the First Automatic Speaker Verification Spoofing and Countermeasures Challenge. *16th Annual Conference of the International Speech Communication Association (Interspeech 2015).* Dresden, Germany, 2015.

7. Kinnunen, T., Lee, K., Delgado, H., Evans, N., Todisco, M., et al. t-DCF: a Detection Cost Function for the Tandem Assessment of Spoofing Countermeasures *Speaker Odyssey 2018. The Speaker and Language Recognition Workshop, 17th August 2018.* – Les Sables-d'Olonne, France, 2018. Available at: https://hal.inria.fr/hal-01880306/document (accessed 20.01.2022).

8. Oppenheim, A., Schafer, R., Buck, J. *Discrete-Time Signal Processing.* Second edition. New Jersey, Prentice Hall, 1999. 893 p.

9. Deller, J., Hansen, J., Proakis, J. *Discrete-Time Processing of Speech Signals.* Second edition. New York, IEEE Press, 2000. 936 p.

10. Huang, X., Acero, A., Hon, H.-W. *Spoken Language Processing: a Guide to Theory, Algorithm, and System Development.* New Jersey, Prentice Hall, 2001. 935 p.

11. Kinnunen, T., Li, H. An Overview of Text-Independent Speaker Recognition: from Features to Supervectors. *Speech Communication*, 2010, no. 52, pp. 12–40.

12. Hao, B., Hei, X. Voice Liveness Detection for Medical Devices. *Design and Implementation of Healthcare Biometric Systems,* 2019, pp. 109–136.

13. Wu, Z., Yamagishi, J., Kinnunen, T., Hanilc, C., Sahidullah, M., Sizov, A., Evans, N., Todisc,o M. ASVspoof: the Automatic Speaker Verification Spoofing and Countermeasures Challenge. *IEEE Journal of Selected Topics in Signal Processing,* 2017, vol. 11, no. 4, pp. 588–604.

14. Abdullah, H., Garcia, W., Peeters, C., Traynor, P., Butler, K., Wilson, J. Practical Hidden Voice Attacks against Speech and Speaker Recognition Systems. *The Network and Distributed System Security Symposium. NDSS.* San Diego, USA, 2019. Available at: https://hal.inria.fr/hal-01880306/document (accessed 20.01.2022).

15. Kinnunen, T., Sahidullah, M., Delgado, H., Todisco, M., Evans, N., Yamagishi, J., Lee, K. A. The ASVspoof 2017 Challenge: Assessing the Limits of Replay Spoofing Attack Detection. *19th Annual Conference of the International Speech Communication Association (Interspeech 2018).* Stockholm, Sweden, 2018.

16. Li, L., Chen, Y., Wang, D. A study on replay attack and anti-spoofing for automatic speaker verification. *18th Annual Conference of the International Speech Communication Association (Interspeech 2017).* Stockholm, Sweden, 2018, pp. 92–96.

17. Usoltsev, A., Petrovska-Delacrétaz, D., Houssemeddine, K. Full Video Processing for Mobile Audio-Visual Identity Verification. *Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2016).* Rome, Italy, 2016, pp. 552–557.