УДК 004.934.8'1

АНАЛИЗ СУЩЕСТВУЮЩИХ ТЕХНОЛОГИЙ АУТЕНТИФИКАЦИИ ЛИЧНОСТИ ПО ГОЛОСОВОМУ СИГНАЛУ

Кулемзин Д.В., Данилюк С.С., Селезнев Д.В.

ФГБОУ ВО «Ростовский государственный экономический университет (РИНХ)», Ростов-на-Дону, e-mail: kulemzin8@mail.ru, vin.90@mail.ru, ghost-nimizis@ya.ru

В настоящее время продолжают уверенно формироваться новые методы, а также подходы к вопросам защиты информации в информационных системах, функционирующих с целью предоставления государственных услуг, а также в сфере безопасности. Одним из самых перспективных направлений является развитие биометрических систем аутентификации. Отдельно среди существующих методов биометрической аутентификации выделяется распознавание диктора по голосу. Этот метод отличает простота в использовании и относительная по сравнению с другими видами аутентификации дешевизна. В статье раскрыты главные принципы работы алгоритмов распознавания личности по голосу, а также рассмотрены и проанализированы основные имеющиеся на сегодняшний день методики, используемые для аутентификации личности по голосовому сигналу. Представлены как основные недостатки приведенных методов и алгоритмов, так и их преимущества. В ходе проведенного исследования выделены как наиболее перспективные для дальнейшего глубокого изучения и совершенствования такие методы голосовой аутентификации, как марковские модели HMM (статистические модели, в которых моделируемая система рассматривается как марковский процесс), а также распознавание SVM (метод опорных векторов, использующий для классификации линейный алгоритм). Указанные в статье методы голосовой аутентификации и проведенный анализ их функционирования указывает на актуальность проведения дальнейшего исследования с целью разработки новых алгоритмов и методов голосовой аутентификации по индивидуальным артикуляционным особенностям диктора.

Ключевые слова: биометрические системы, голосовая аутентификация, защита информации, диктор, идентификация, марковские модели HMM, распознавание SVM

ANALYSIS OF EXISTING TECHNOLOGIES PERSONALITY AUTHENTICATION BY VOICE SIGNAL

Kulemzin D.V., Danilyuk S.S., Seleznev D.V.

Rostov State University of Economics, Rostov-on-Don, e-mail: kulemzin8@mail.ru, vin.90@mail.ru, ghost-nimizis@ya.ru

Currently, new methods continue to be confidently formed, as well as approaches to the issues of information security in information systems that function to provide public services, as well as in the field of security. One of the most promising areas is the development of biometric authentication systems. Separately, among the existing methods of biometric authentication, speaker recognition by voice stands out. This method is easy to use and relatively cheap compared to other types of authentication. The article reveals the main principles of operation of algorithms for identifying a person by voice, and also considers and analyzes the main methods available today that are used to authenticate a person by a voice signal. Both the main disadvantages of the above methods and algorithms and their advantages are presented. In the course of the study, such methods of voice authentication as HMM Markov models (statistical models in which the simulated system is considered as a Markov process) and SVM recognition (support vector method that uses linear algorithm). The methods of voice authentication indicated in the article, and the analysis of their functioning indicates the relevance of further research in order to develop new algorithms and methods of voice authentication based on the speaker's individual articulatory features.

Keywords: biometric systems, voice authentication, information security, speaker, identification, HMM Markov models, SVM recognition

Биометрические системы аутентификации личности по состоянию на сегодняшний день стремительно развиваются и уверенно внедряются с каждым годом в работу как государственных, так и частных организаций. Примером этому служит то, что данные системы надежно закрепились в банковской сфере, а также в сфере предоставления государственных услуг.

Биометрические признаки уникальны для каждого человека, поэтому биометрические системы аутентификации личности нашли широкое применение. Распознавание диктора по голосу является одним из методов биометрической аутентификации и в свою очередь имеет большое преимущество перед другими методами за счет простоты и дешевизны используемых средств.

Цель исследования – выделение наиболее перспективных для дальнейшего изучения методов аутентификации личности по голосовому сигналу.

Материалы и методы исследования

Голосовая аутентификация состоит в характеристике личности по голосу, при которой предлагаемый образец голоса сопоставляется с имеющейся базой [1]. Основные характеристики (ошибки первого и второго рода) систем голосовой аутентификации, в значительной мере определяются

отношением сигнал/шум обрабатываемых материалов регистрации. Ввод речевого сигнала пользователя осуществляется, как правило, на фоне внешних мешающих сигналов, например акустических волн, обусловленных работой коммутационной аппаратуры, серверов, кондиционеров. Задача компенсации этих сигналов осложняется, если они действуют в полосе частот речевого сигнала. Экспериментальные исследования спектра сигнала, обусловленного работой коммутационной аппаратуры (коммутаторы, маршрутизаторы), показали, что он занимает большую полосу частот с максимумами в области 1; 3; 4,2; 5 кГц [2]. Аутентификация диктора в сигнале, который был модифицирован, закодирован, сжат или обработан другими видами, аналоговыми или цифровыми преобразователями значительно усложняется. При подаче речевого сигнала по каналам связи анализу подвергается данный сигнал. Кодирование позволяет исключить из предлагаемого речевого сигнала черты аутентификации.

Также необходимо принимать во внимание повышенный интерес к системам голосовой аутентификации мошенников и злоумышленников, чья цель состоит во взломе системы и получении доступа к данным пользователей. В частности, предъявление диктофонной или любой цифровой записи голоса человека является одной из распространенных форм атаки на системы распознавания.

Результаты исследования и их обсуждение

Из вышеперечисленных факторов очевидно, что без решения вопросов о степени алгоритмов цифровой обработки на каждую группу признаков, используемых для задачи аутентификации, процедура защиты данных не может быть защищенной, надежной и достоверной.

Основу большинства методов голосовой аутентификации составляют методы анализа лингвистических или акустических характеристик голоса, к которым относят дефекты речи, высоту и силу голоса, тембр, продолжительность фонации. В настоящее время распознавание диктора производится на основе экспертных и автоматических методик, которые также именуются объективными и субъективными методиками.

Экспертные методики в процессе фоноскопического анализа позволяют уточнить результаты применения автоматических средств распознавания. Экспертные методы являются наиболее предпочтительными в том случае, если в сложившихся условиях (например, при повышенном шуме, помехах), затруднено использование автоматических методов.

На сегодняшний день задействуются следующие методики:

− мелодический контур;

− выравнивание;

− микроанализ диапазона гласных [3].

Для задачи идентификации и аутентификации человека по голосовому сигналу могут использоваться практически все методы анализа этого сигнала. Однако анализировать весь голосовой сигнал даже для современных электронно-вычислительных систем является сложной задачей, что отражено в таблице.

Поэтому для задач защиты автоматической идентификации и аутентификации голосовой сигнал в большинстве случаев параметризованный, то есть представляется в виде малого количества информационно значимых параметров.

Сравнительный анализ показателей биометрических методов аутентификации [4]

Биометрическая технология Показатель	Признание пользователями	Устойчивость к подделкам и атакам	Стоимость	Простота использования	Frr	Far	Время распознавания объекта	Размер шаблона	Стабильность работы при болезнях человека
Отпечаток пальца	5	5	7	8	5	5	6	5	9
Геометрия руки	5	6	4	8	5	5	8	9	4
Геометрия лицо	9	3	7	9	1	6	8	5	3
Радужная оболочка	4	6	5	6	7	7	7	7	8
Динамика подписи	7	4	6	8	8	7	9	7	6
Голос	9	1	9	9	3	5	6	2	3

К значимым относятся следующие характеристики голосового сигнала:

− амплитудные;

− фазовые;

− временные;

− частотные;

− энергетические.

Традиционными для решения таких задач защиты информации являются алгоритмы, основанные на преобразованиях Фурье. Самыми часто используемыми среди алгоритмов являются:

− выделение мел-частотных коэффициентов (MFCC);

− выделение коэффициентов линейного предсказания [5].

У указанных методов имеются определенные преимущества. Это можно объяснить тем, что итоговый вектор не находится в зависимости от первоначального образца, а также тем, что во внимание принимаются индивидуальные особенности голоса исследуемого субъекта [6].

Опубликованные результаты практических исследований систем, используемых для параметризации этих алгоритмов, свидетельствуют, что процент точной аутентификации дикторов превышает 98 % [7]. Однако, по данным других источников, задействование образцов из реальных каналов не позволяет получить абсолютно точные данные (их точность в большинстве случаев составляет не более 90 %) [8].

На основе представленных алгоритмов имеется возможность с большой точностью установить максимальные значения спектра голоса человека для отдельных звуков и формальной частоты. Указанные значения будут зависеть от анатомических и артикуляционных характеристик голоса субъекта, который подлежит аутентификации, и возможных дефектов его речи [9].

Указанные признаки подвержены искажению речи под воздействием шума, изменения физического состояния человека [10].

Основанные на спектральном анализе методы имеют явные недостатки при их использовании:

− измерение интенсивности линий сужено видимой областью спектра, поэтому отображение структуры сигнала является весьма ограниченным;

− если некорректно подобрать параметры, то могут возникнуть сложности с поиском формантных частот;

− невысокая скорость;

− возможность искажений сигнала при прохождении его непосредственно по каналу связи [11].

Методики линейного предсказания и кепстральных коэффициентов характеризуются большей скоростью выполнения аутентификации, но также сопровождаются рядом недостатков, основными из которых являются следующие:

− сложности, которые имеют место при выборе порядка модели;

− полученные в результате использования таких методик данные не всегда являются стабильными;

− низкий уровень надежности идентификации при аутентификации одновременно нескольких голосов;

− возможны ситуации, когда модели не отражают влияние фазовых характеристик языка.

С целью вычисления фазовых и амплитудных параметров также используются преобразования Гильберта. Их применение удобно, но использование прямого подхода к вычислениям начальной фазы речевого сигнала осложняется ввиду неопределенности прямой фазы в широком диапазоне частоты.

В практической деятельности наиболее предпочтительными являются такие методики аутентификации голоса, результаты которых не зависят от личности диктора.

К таким методикам относятся:

− марковские модели HMM (статистические модели, в которых моделируемая система рассматривается как марковский процесс);

− распознавание SVM (метод опорных векторов, использующий для классификации линейный алгоритм).

Стандартным считается применение скрытых марковских моделей, которые представляют собой цепочку с присущей ей итоговой совокупностью состояний. Последовательность состояний не прослеживается, в связи с чем модель именуется скрытой.

В настоящее время формирование HMM осуществляется на основе следующих подходов:

− создание образцов минимальных языковых единиц;

− создание различных типов моделей голосов основывается на эффекте коартикуляции;

− распознавание отграниченных слов осуществляется на основе HMM для каждого из них;

− поток речи распознается на основе единой HMM, через определенные промежутки производится соединение.

Использование HMM имеет помимо положительных характеристик и определенные недостатки. Основным из таких недостатков является то, что добиться точного распознавания можно только в том случае, если первоначально имеется строго установленная совокупность фиксированных условий. Если при использовании данной

модели оказывают воздействие шумы, помехи или канал связи, то такая модель утрачивает собственную устойчивость.

Модель SVM при обработке речевых сигналов наиболее предпочтительна в том случае, если обработке подвергается естественный язык, например, когда производится анализ искаженной текстовой информации при аутентификации. Предпосылкой использования SVM может быть следующее утверждение: в пространстве существуют два вида объектов, разделенных с помощью гиперплоскости таким образом, что объекты разных классов в итоге окажутся по разные стороны от гиперплоскости. Вполне очевидным является то, что возможно существование нескольких плоскостей. Для того чтобы классы были как можно лучше разделены между собой, необходимо увеличивать расстояние между гиперплоскостями.

В большинстве случаев для того, чтобы найти параллельные гиперплоскости, при использовании метода опорных векторов до минимума снижается квадратичная функция. Решение указанной задачи состоит в том, что за основу принимаются координаты опорных векторов. Если сложится такая ситуация, что классы будут являться линейно неразделимыми в исходном пространстве, то отображение формируется на пространство большего размера с линейной разделенностью образцов классов, что именуется пространством вторичных признаков.

Заключение

После анализа существующих технологий аутентификации личности по голосовому сигналу становится ясным, что все они имеют свои существенные недостатки, но можно выделить два метода которые имеют потенциал для более глубокого изучения и совершенствования.

Приоритетными характеристиками HMM можно считать следующие из них:

− используемая для анализа математическая структура отличается своей простотой;

− модель имеет возможность выстраивать сложную цепь последовательных наблюдений;

− параметры, которые положены в основу описания совокупности данных, могут подбираться автоматически.

Рядом преимуществ также обладает SVM алгоритм. К таким преимуществам относятся:

− гарантированным является получение единственного решения поставленной задачи, что отличает в лучшую сторону данный алгоритм от нейронных сетей, в которых может иметь место множество решений или ответ может быть не определен вообще;

− алгоритм успешно справляется с излишними шумами и помехами, которые имеются во входном сигнале, что повышает распознаваемость речи;

− в рамках алгоритма имеется возможность обработки данных больших размеров, что необходимо при распознавании речи [11].

Данные методы возможно успешно использовать для построения современных безопасных систем голосовой аутентификации [12] и интегрировать как составляющую биометрической системы аутентификации в глобальную информационную инфраструктуру. Разработка новых алгоритмов и методов идентификации голосового сигнала по индивидуальным артикуляционным особенностям диктора без вышеназванных недостатков представляет собой актуальную задачу.

Список литературы

1. Мешков А.Ю., Новиков О.О. Алгоритмы анализа голосовых сигналов человека для задачи идентификации и диагностики физического состояния: сборник статей участников тридцать шестой международной научно-практической конференции «Инновационный потенциал мировой науки XXI века». Том 2. Естественные и точные науки (29.12.2018 – 05.01.2019). С. 26–28.

2. Файзулаева О.Н., Невлюдов И.Ш. Пути улучшения качества речевого сигнала пользователя систем голосовой аутентификации // Научно-технический вестник информационных технологий, механики и оптики. 2014. № 2 (90). С. 118–123.

3. Кравченко А.П., Крамарь Н.М., Морозов И.В. Автоматизированная компьютерная система голосового управления автомобилем // Автомобильный транспорт: сб. науч. тр. 2019. Вып. 25. С. 46.

4. Ляшенко Е., Астраханцев А.А. Исследование эффективности методов биометрической аутентификации // Системы обработки информации. 2017. № 2 (148). С. 112.

5. Levinson Stephen C. Mathematical models for speech technology. University of Illinois at Urbana-Champaign. Wiley – 2005. P. 19–20.

6. Заковряшин А.С., Малинин П.В., Лепендин А.А. Применение распределений мелкочастотных кепстральных коэффициентов для голосовой идентификации личности // Известия АлтГУ. 2014. № 1 (81). С. 159.

7. Matsui T., Furui S. Comparison of text-independent speaker recognition methods using VQ-distortion and discrete. Proc. ICSLP. 1992. P. 158.

8. Сорокин В.Н., Цыплихин А.И. Верификация диктора по спектрально-временным параметрам речевого сигнала // Информационные процессы. 2010. Вып. 10. № 2. С. 90.

9. Васильев Р.А. Исследование особенностей фонетического строя речи и идентификация дикторов по голосу // Современная наука: актуальные проблемы теории и практики. Естественные и технические науки. 2012. № 8–9. С. 19–22.

10. Фролов А.А. Алгоритм текстонезависимой идентификации человека по голосу // Известия ВолгГТУ. 2013. № 14 (117). С. 64.

11. Иванов И.И. Анализ метода мел-частотных кепстральных коэффициентов применительно к процедуре голосовой аутентификации // Актуальные проблемы гуманитарных и естественных наук. 2015. № 10–1. С. 107–108.

12. Берштейн С.И., Колокольцев Н.К., Ермолаева В.В. Голосовая аутентификация // Молодой ученый. 2018. № 25 (211). С. 93–94.