

UDC 51-76; 004.11

IRSTI 50.37.23; 50.41.25

ОБРАБОТКА ГОЛОСОВОЙ ИНФОРМАЦИИ ДЛЯ БИОМЕТРИЧЕСКОЙ ИДЕНТИФИКАЦИИ ЛИЧНОСТИ

М.С. Элиасқар², Т.Ж. Мазаков^{1,2}, Г.С. Байрбекова¹, Н.Т. Исимов¹, Т.С.Шорманов²,
Д.К. Мухаев²

¹Институт информационных и вычислительных технологий КН МОН РК, Казахстан

²Казахский национальный университет имени аль-Фараби, Казахстан

87019931011@mail.ru

²ORCID ID: <https://orcid.org/0000-0001-9345-5167>

Abstract. Потребность в проверке личности людей с помощью надежных методов и параметров быстро возрастает во всем мире в эту информационную эпоху из-за желания сообщества выполнять повседневные процессы, такие как перевод денег или запрос услуги с использованием удаленных систем или колл-центров. Традиционные методы аутентификации, такие как использование паролей и показ удостоверений личности, теперь заменяются нашими физическими или поведенческими функциями, которые нельзя заменить или имитировать. Используя наши биометрические функции, методы аутентификации теперь могут выполняться только с использованием смартфонов, камер или только микрофона. Биометрические функции обеспечивают разные уровни шкалы доверительного интервала, у каждой из них есть свои плюсы и минусы, поэтому их можно выбрать по способности отвечать на запросы аутентификации, требующие разных уровней точности. Голосовая биометрия или технология распознавания говорящего обеспечивает эффективный и надежный способ аутентификации заявителя без необходимости находиться в одном и том же физическом месте и без особых усилий говорящего, более того, это можно сделать быстро во время телефонного звонка. Принимая во внимание рост числа мошеннических атак в последнее время из года в год и возможность его аутентификации, не сталкиваясь лицом к лицу с истцом, неудивительно, что использование технологии голосовой аутентификации растет в сфере банковских услуг и центров обработки вызовов. Это исследование представляет собой краткий обзор основных методов и исследований в области биометрии для личной идентификации.

Ключевые слова: биометрические технологии, голосовая аутентификация, извлечение голосовых характеристик, моделирование динамик.

Введение

В современном мире все больше проявляется интерес к голосовым технологиям, в частности, к идентификации личности. Это объясняется, с одной стороны, появлением высокопроизводительных вычислительных систем на базе персональных компьютеров и аппаратных средств, позволяющих производить ввод сигнала в компьютер, а, с другой стороны, высокой потребностью систем аутентификации в разных областях.

Метод опознавания личности по голосу существует с тех пор, как человек научился говорить. Поэтому достоинства и недостатки этого метода известны всем. Как не всегда по ответу на вопрос «Кто там?» мы можем определить, что за дверью стоит знакомый человек, и приходится развеивать свои сомнения, заглянув в дверной глазок, так и техническая система идентификации может ошибаться в силу изменения голоса отдельного человека.

Привлекательность данного метода - удобство в применении. Метод проверки голоса имеет два положительных отличия от остальных биометрических методов. Во-первых, это идеальный способ для телекоммуникационных приложений. Во-вторых, большинство современных компьютеров уже имеют необходимое аппаратное обеспечение. Продукты с проверкой голоса сейчас предлагают банки и другие компаний.

Основная проблема, связанная с этим подходом, - точность идентификации. Однако это

не является серьезной проблемой с того момента, как устройства идентификации различают характеристики человеческой речи. Голос формируется из комбинации физиологических и поведенческих факторов.

Идентификация по голосу удобный, но в тоже время не такой надежный, как другие биометрические методы. Например, человек с простудой или ларингитом может испытывать трудности при использовании данных систем. Существует также возможность воспроизведения звукозаписи.

Методы исследования

Биометрическая система, по существу, распознает или извлекает шаблоны определенных физиологических или поведенческих характеристик, которыми обладает целевой человек, и использует эти шаблоны для сопоставления с заданными данными.

К настоящему моменту у нас и за рубежом реализованы системы автоматической идентификации по голосу, большинство из которых строятся по единой концептуальной схеме:

- производится регистрация пользователя и вычисляется шаблон;
- выбираются участки речевого потока для дальнейшего анализа;
- осуществляется первичная обработка сигнала;
- вычисляются первичные параметры;
- строится «отпечаток» (шаблон) голоса;
- производится сравнение «отпечатков» голосов и формируется решение по идентичности голосов или «близости» голоса к группе голосов [1].

Гипотетически биометрическая система состоит из двух основных модулей: регистрации и идентификации. Система сканирует особенности или, другими словами, биометрические характеристики человека с помощью биометрического датчика, получает цифровое представление этих функций и сохраняет его в качестве шаблона в центральной базе данных или на смарт-карте на этапе регистрации.

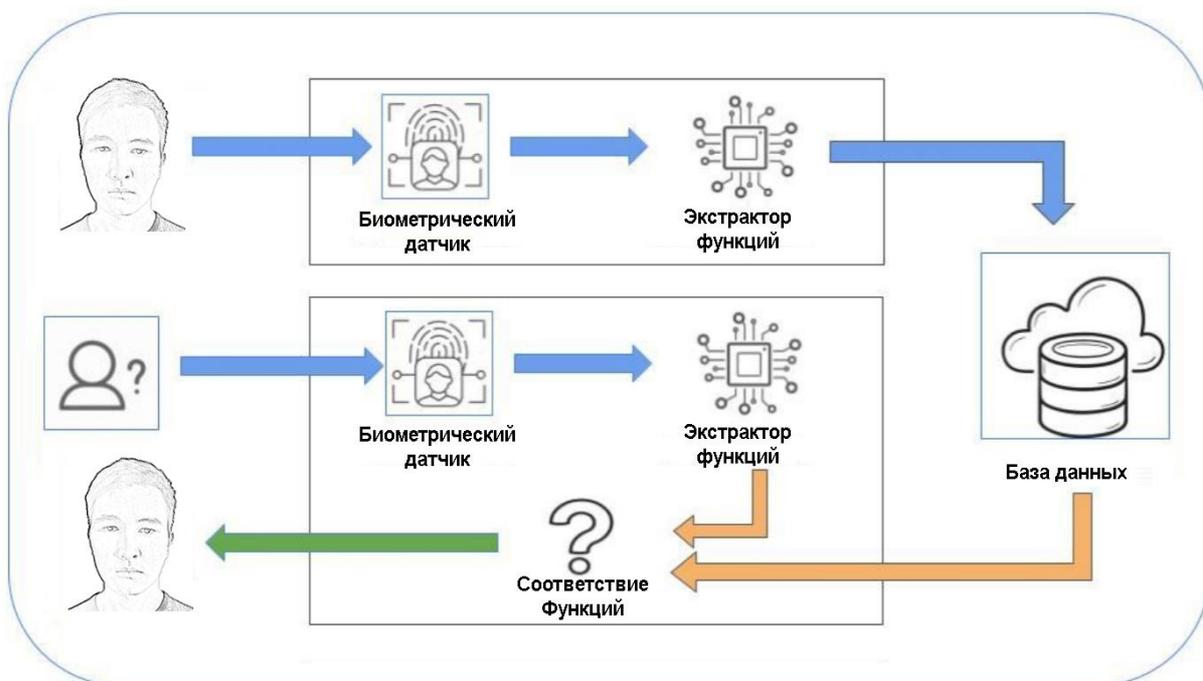


Рисунок 1 – Гипотетическая биометрическая система

Модуль распознавания снова сканирует биометрические характеристики целевого человека и преобразует их в цифровой формат, который используется на этапе регистрации. Затем система использует эти данные для сопоставления характеристик, что означает сравнение полученных данных с шаблонами для определения личности человека.

Общая производительность биометрической системы оценивается по ее показателю точности, скорости и хранению. Биометрические системы могут принять мошенника как действительного человека, и это называется «ложным совпадением» или, наоборот, система может отказать действительному лицу, и это называется «ложным несоответствием» [2].

Образец голоса – это индивидуальная характеристика, и невозможно найти двух людей с точно такой же спектрограммой. Спектрограмма – это визуальная запись речи, проанализированная по частоте, продолжительности и амплитуде. Поскольку голосовой отпечаток уникален для своего владельца, его можно использовать для уникальной идентификации.

Необходимое оборудование для голосовой аутентификации – это просто микрофон, поэтому его стоимость настолько мала, им легко пользоваться, и технология распознавания голоса сегодня присутствует в наших смартфонах. Но, конечно, он не идеален, как другие биометрические характеристики. Самый большой недостаток голоса в том, что он меняется в зависимости от возраста, и при переходе от детства к подростковому возрасту могут произойти резкие изменения. Также на голос могут негативно повлиять болезни и мгновенные эмоции, акустика помещения и окружающий шум [3].

Голосовая аутентификация рассматривается как проблема распознавания образов, которая является ветвью машинного обучения. Основная задача при разработке системы голосовой аутентификации - извлечение речевого сигнала. Для извлечения речевых характеристик из речевого сигнала доступны некоторые методы. Модели динамиков создаются для отдельных лиц и сохраняются в виде голосовых баз данных после извлечения функций. Для создания моделей динамиков могут использоваться методы моделирования, включая векторное квантование (VQ), GMM (модели гауссовой смеси), НММ (скрытые марковские модели) и т.д.

Задача аутентификации говорящего состоит из двух компонентов: извлечения функций и сопоставления их, как упоминалось ранее. После извлечения уникальных характеристик из входной речи функции сравниваются с шаблонами выступающих, которые создаются с помощью различных методов моделирования и сохраняются [4].



Рисунок 2 – Блок-схема системы распознавания говорящих

Методы сопоставления функций используются как на этапе обучения, так и на этапе тестирования. Чтобы построить модель динамика, система обучается с использованием извлеченных функций на этапе обучения. Эта модель проверяется системой на этапе тестирования посредством распознавания выступающих.

Наиболее часто используемые методы построения моделей: векторное квантование (VQ), скрытая марковская модель (НММ) и модель гауссовой смеси (GMM) [5].

Векторное квантование (VQ): векторное квантование, пожалуй, самый простой алгоритм для сопоставления признаков и создания моделей с использованием векторов признаков. Тип VQ – это алгоритм обучения без учителя. Он создает кластеры, которые представляют модели зарегистрированных спикеров.

Во-первых, полученные векторы признаков классифицируются в разные кластеры. Кластеры включают в себя векторы признаков со схожими свойствами, и эти похожие векторы моделируют атрибуты говорящих. Сначала получают векторы признаков, затем они сравниваются с центроидами уже существующих кластеров. Для этого используется расчет евклидова расстояния. Алгоритм помещает вектор признаков в кластер с минимальным расстоянием. После назначения вектора кластеру центроид кластера

обновляется. Другой термин для центроидов – это «кодовые слова», а термин «кодовая книга» используется для набора кодовых слов.

Скрытая марковская модель (НММ): НММ может моделировать статистические вариации характеристик, давать статистическое представление о том, как динамик воспроизводит звук [6]. Следовательно, НММ является более эффективным алгоритмом сопоставления признаков по сравнению с традиционной моделью VQ.

Модели НММ – это статистические модели, основанные на идее, которая является «распределением вероятностей состояния для времени x , условно независимого от всех предыдущих состояний, когда состояние во время $x-1$ задано. Кроме того, мы не можем напрямую определить состояния, для которых они невидимы. Используйте или, другими словами, они «скрыты», но мы можем использовать параметры наблюдения для прогнозирования состояний. Цель состоит в том, чтобы определить эти скрытые значения состояния с помощью наблюдаемых параметров.

НММ - один из лучших методов анализа голоса, потому что он имеет четко определенную математическую структуру, не требует экспертных знаний о речевом сигнале, не накапливаются ошибки в анализе и не требует большого количества шаблонов.

Модель гауссовой смеси (GMM): модель гауссовой смеси - один из наиболее распространенных и успешных методов моделирования динамиков. Речевые атрибуты легко могут быть распределены по Гауссу, поэтому GMM - очень эффективный способ моделирования [7]. Плотность гауссовой смеси представляет собой взвешенную сумму M плотностей компонентов, где M представляет собой количество гауссианов. Каждого говорящего можно эффективно смоделировать и представить с помощью GMM, и на него ссылается модель, связанная с ним / ней.

Выводы

Биометрические – это общее название процессов, используемых для проверки и идентификации людей путем измерения их физиологических и поведенческих характеристик. Биометрические идентификаторы более надежны по сравнению с методами, основанными на знаниях и токенах, поскольку они обладают такими сильными характеристиками, как универсальность, уникальность, постоянство и возможность сбора. Наиболее часто используемые биометрические параметры – это отпечатки пальцев, геометрия руки, распознавание лиц, сетчатка, радужная оболочка глаза, голос, и каждая из них имеет свои плюсы и минусы по сравнению с другими.

Голосовая аутентификация может быть объяснена как распознавание говорящего с использованием информации, уникальной для человека. Системы распознавания громкоговорителей сначала извлекают из речевого сигнала определенные характеристики, а затем создают модели или шаблоны громкоговорителей для хранения этих функций. Наконец, система сравнивает голос, который должен быть аутентифицирован, с сохраненными шаблонами или голосовыми распечатками и пытается определить, совпадают ли они.

Системы голосовой аутентификации все чаще используют методы биометрии, потому что они не требуют никакого оборудования, кроме микрофона, взаимодействие пользователя не является обязательным и реализуется с помощью методов, основанных на сильной математической основе. Системы распознавания голоса, которые в настоящее время используются в таких областях, как судебная экспертиза и банковское дело.

Благодарность

Работа выполнена за счет средств программы целевого финансирования на основе договора по оказанию услуг субподряда «Разработка национального электронного банка данных по научной зоологической коллекции Республики Казахстан, обеспечивающего их эффективное использование в науке и образовании».

References

- [1] Singh N., Agrawal A., Khan R.A. Голосовая биометрия: технология аутентификации на основе голоса, передовая наука, инженерия и медицина. Июль 2018.
- [2] Todkar S.P., Babar S.S., Ambike R.U., Suryakar P.V., Prasad J.R., Методы распознавания докладчиков. Материалы 3-ий Междунар. конф. по конвергенции в технологиях. 2018.
- [3] Ahmad, K. S., Thosar, A. S., Nirmal, J. H., Pande, V. S. Уникальный подход к распознаванию говорящего независимо от текста с использованием наборов функций MFCC и вероятностной нейронной сети. Материалы 8-ой Междунар. конф. (ICAPR), 2015.
- [4] Dehak N., Kenny P., Dehak R., Dumouchel P., Ouellet P., Front-end факторный анализ для проверки докладчика. IEEE Transactions on Audio, Speech, and Language Processing, 2011.
- [5] Y. Lei, N. Scheffer, L. Ferrer, and M. McLaren. Новая схема распознавания говорящего с использованием фонетически осведомленной глубокой нейронной сети. Материалы Междунар.конф. по акустике, речи и обработке сигналов (ICASSP), 2014г.
- [6] Bahmaninezhad F., Hansen J., Распознавание динамиков i-Vector / PLDA с использованием опорных векторов с дискриминантным анализом. Материалы Междунар. конф. IEEE по акустике, обработке речи и сигналов (ICASSP), 2017.
- [7] Garcia-Romero D., McCree A., Анализ глубинных нейронных сетей для распознавания говорящего. Interspeech 2015.