Речевые технологии в биометрических системах идентификации и верификации

Перед руководителями многих предприятий сегодня остро стоит проблема защиты от несанкционированного доступа к своим материальным и интеллектуальным ресурсам. Для решения этой задачи применяются различные системы идентификации личности. Важным элементом такой системы является способ авторизации (подтверждения подлинности личности) пользователя.
Самым простым в реализации способом подтверждения прав доступа к ресурсу является использование пароля доступа. В связи с развитием Интернета, организации доступа к информационным ресурсам по телефонному каналу, по локальной сети все больший интерес вызывают голосовые биометрические системы.

Справка ТБ

Никифоров Сергей Никонорович.
Образование: высшее техническое.
1971—1977 гг. — Санкт-Петербургский государственный технический университет информационных технологий, механики и оптики. Специальность — «Специальные оптико-электронные приборы».
1989 г. — курсы повышения квалификации по специальности «Системы управления» (Институт повышения квалификации Министерства радиотехнической промышленности (г. С.-Петербург).
1985 г — Институт повышения квалификации по специальности «Инженер-программист» (г. Минск). Руководитель и автор ряда разработок в области речевых технологий: систем записи переговоров, оборудования и программного обеспечения для контакт-центров и т. д. Автор 5 патентов на изобретения. Сфера интересов: цифровая обработка сигналов, синтез и распознавание речи.

Использование голосовых биометрических данных, уникальных для каждого человека, предоставляет автоматизированным системам с удаленным доступом дополнительный уровень безопасности. Система голосовой биометрии не зависит от языка клиента и не использует пользовательские словари, поэтому позволяет выбрать легко запоминающееся ключевое слово. Таким образом, отсутствует необходимость в использовании PIN-кодов. При этом улучшается качество обслуживания клиентов, уменьшается необходимое время для стандартных операций. Кроме того, система обеспечивает персонализацию голосовых сервисов благодаря идентификации клиентов и предоставлению предпочитаемых ими опций.

Уменьшается вероятность фальсификаций и мошенничества при использовании базы данных подозрительных голосов и переключении подозрительных абонентов на службу безопасности.

Кроме того, система легко интегрируется в любую архитектуру.

Технология автоматического распознавания диктора делает возможным применение голоса в самых различных приложениях.

Это необходимо для обеспечения допуска в пропускные и контрольные системы ограниченного использования, в системы речевого ввода информации, в интерфейсы, например, доступ к банковским вкладам или совершение покупки по телефону и т. п.

На сегодняшний день известен целый ряд фирм и предприятий, ведущих разработки в области голосовой биометрии. Это известные за рубежом Neuance, в России — «Центр речевых технологий», «Вокорд» и другие.

В Беларуси работы по созданию системы идентификации ведутся группой специалистов под научным руководством заведующего лабораторией распознавания и синтеза речи Объединенного института проблем информатики НАН Беларуси (ОИПИ), доктора технических наук Б. М. Лобанова. Данная разработка основана на многолетнем опыте создания систем распознавания речи.

Структура системы показана ниже (рис. 1) и состоит из нескольких этапов:

- оцифровка речевого сигнала и фильтрация;
- выделение тональных участков речи и удаление пауз (сегментация);
- определение первичных признаков (высокий или низкий тон, т. е. разделение на мужской или женский голос);
- формирование вектора признаков (создание параметрического образа диктора);
- сравнение с эталонами и нахождение наиболее близкого образа;
- принятие решения и вывод результатов на экран компьютера в виде, удобном для анализа (рис. 2).

Программа построена по принципу «клиент—сервер» и конфигурируется в зависимости от назначения в нескольких вариантах:

- для служб безопасности;
- для центров обработки вызовов;
- для задач верификации (голосовой пароль).

При этом, в отличие от упомянутых выше систем,

упор в разработке делается не только на верификацию, т. е. подтверждение

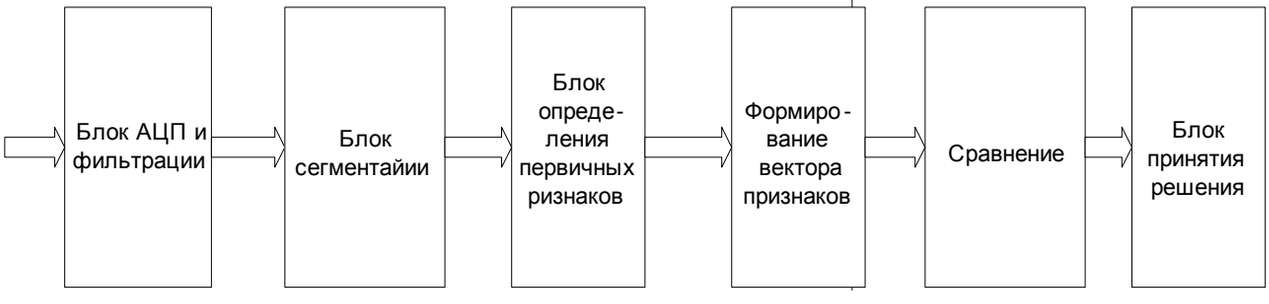

Рис.1 Структура системы идентификации.

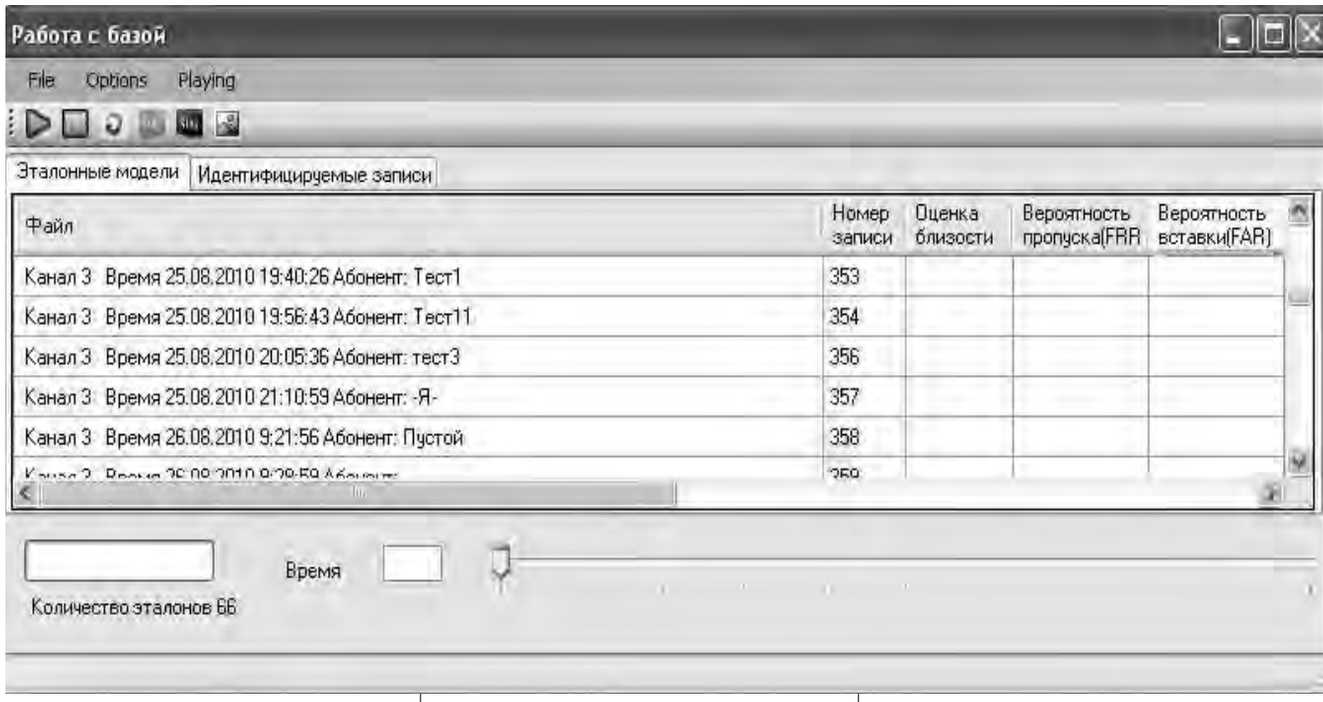

Рис. 2. Вид главного меню сервера тестовой системы.

диктора, произнесшего определенную фразу, но и на идентификацию и поиск диктора в большом массиве произвольных фонограмм речи (например, диалоги оператора контакт-центра и клиента).

Рассмотрим подробнее некоторые аспекты автоматического определения диктора по голосу.

При **идентификации** необходимо классифицировать неизвестного диктора в некотором эталонном множестве голосов, а при **верификации** требуется принять решение, принадлежит ли спорная фонограмма одному конкретному эталонному диктору из справочной базы голосов. Верификация значительно проще идентификации, поскольку тут принимается альтернативное решение — «да» или «нет», а диктор произносит одну и ту же парольную фразу.

Несмотря на некоторое сходство этих задач цели и способы применения их могут очень отличаться. Технически при этом задача компьютера заключается в том, чтобы сравнить параметрический код предъявляемого голоса с эталонным высказыванием заявленного лица (при верификации) или сравнить с каждым из конечного числа параметрических описаний (эталонов) зарегистрированных лиц (при идентификации). Если сравнение при верификации показывает достаточную по критериям системы близость, диктор считается системой «своим», а если близость превышает некий порог, то диктор объявляется «чужим». При идентификации компьютер, сравнивая рассчитанные параметры голоса абонента, выбирает наиболее близкое из числа эталонов, находящихся в базе.

Полная процедура обычно такова. Анализ речевого сигнала начинается с перевода его в цифровую форму. Производится сегментация сигнала на отдельные элементы. Затем акустический сигнал обрабатывается с помощью определенных алгоритмов — спектрального анализа, линейного предсказания, кепстральной обработки и других. В результате получается параметрическое описание сегментов речевого сигнала в виде вектора параметров. Следующий этап — сравнение с имеющимися эталонными описаниями зарегистрированного числа дикторов в базе данных компьютера. Это достигается путем использования метода динамического программирования, скрытых Марковских моделей (особенно для распознавания по слитной речи), искусственных нейронных сетей или комбинаций указанных методов. Качественным показателям любой системы идентификации служит вероятность правильного обнаружения, оцененная на представительной выборке реализаций. Вычисляются также вероятности пропуска «чужого» (когда разные люди ошибочно отождествляются) и отказа в идентификации «своего» (реализации голоса одного диктора опознаются как принадлежащие разным людям). Очевидно, что эти ошибки имеют «противоположный» характер. Снижая вероятность одной, мы неизбежно повышаем вероятность второй. Имеется множество разных критериев построения эффективной системы, и все они, так или иначе, «взвешивают» два типа ошибок в зависимости от степени их важности для системы. Сложность идентификации заключается в том, что идентифицируемый не всегда хочет быть узнанным, поэтому в ряде случаев он пытается изменить голос, что затрудняет процедуру опознания. При верификации же человек обычно заинтересован в сотрудничестве с компьютером. Предъявляя свое речевое высказывание в качестве голосового пароля, он старается помочь машине провести опознание. Потенциальный злоумышленник, правда, тоже будет стараться ей «помочь», но с этим можно успешно бороться, настраивая систему так, чтобы она надежно пропускала только «своих» ценой отказа кому-то из них в допуске, смещая статистику работы в область малых вероятностей пропуска «чужого». Здесь возможна ситуация, когда компьютеру вместо речи живого человека предъявляется голос, заранее записанный или вовсе скомпилированный из различных высказываний. Как правило, в кино система принимает фальшивку и допускает злодея, на практике все сложнее, не говоря о том, что возможны различные дополнительные средства контроля. Так, например, компьютер может запрашивать другую форму пароля или его иное грамматическое оформление.

До недавнего времени мировой рынок голосовой биометрии оценивался приблизительно в $ 30 млн. По оценкам агентства frost & sullivan, доходы мировых компаний, действующих на этом сегменте биометрического рынка, к 2011 г. достигнут около $ 500 млн. Потенциальный рынок данной технологии в мире постоянно растет.

Следовательно, применение современных речевых биометрических технологий является одним из перспективных и приоритетных направлений совершенствования систем контроля и управления доступом к информационным и другим ресурсам. ■