ДМИТРИЙ ДЫРМОВСКИЙ
ddv@speechpro.com

ЮРИЙ МАТВЕЕВ
matveev@speechpro.com

ЛАРИСА БАЛЫКИНА
balykina@speechpro.com

СОВРЕМЕННЫЙ РЫНОК РЕЧЕВЫХ ТЕХНОЛОГИЙ

В статье рассматриваются наиболее востребованные речевые технологии и решаемые с их помощью задачи в различных отраслях и сферах нашей жизни.

Сегодня речевые технологии прочно вошли в жизнь современного человека, делая ее намного удобнее и проще. С их помощью можно озвучивать книги, sms-сообщения, документы и целые веб-сайты, строить запросы в поисковых системах без помощи клавиатуры, изучать языки, получать информацию с личного счета без использования паролей и даже давать указания персональному автомобилю.

КЛЮЧЕВЫЕ РАЗРАБОТКИ

В области современных речевых технологий, помимо трех основных задач — распознавания речи [1], синтеза речи по тексту [2], а также голосовой биометрии [3], — наиболее востребованными разработками, как в России, так и за рубежом, являются: запись звука и речи; шумоочистка и улучшение разборчивости речевого сигнала; интеллектуальный анализ и обработка речевых данных. Если технологии распознавания и синтеза речи зависят от языка, то остальные являются «языконезависимыми» технологиями. Рассмотрим подробнее основные речевые технологии.

Запись звука и речи

Устройства записи речевой информации и звука (с телефонных линий, микрофонов или линейных выходов аппаратуры) могут выступать в качестве автономных регистраторов или цифровых диктофонов. Среди основных достоинств автономных регистраторов можно выделить возможность работы с фонограммами через веб-интерфейс, низкое энергопотребление, централизованную базу данных и централизованный мониторинг всех записывающих устройств. Цифровые диктофоны, так же, как и автономные регистраторы, отличаются наличием безопасного доступа к аудиозаписям, высоким качеством записей, что делает их пригодными для распознавания речи и голоса.

Шумоочистка

Шумоочистка — обработка сигнала, которая позволяет повысить разборчивость речевого сигнала, уменьшить долю шумов и компенсировать искажения, вызванные как акустическими, так и технологическими причинами.

Современные технологии позволяют проводить шумоочистку в реальном и отложенном времени, применять различные фильтры. В основу большинства алгоритмов обработки речевых сигналов [4] положена идея адаптации, суть которой заключается в использовании текущей информации о сигнале для автоматической подстройки режима его обработки к типу помехи.

Распознавание речи

При изучении технологии распознавания речи, как правило, выделяют:

- Распознавание отдельных команд. Эта технология лежит в основе голосовой навигации по сайтам. Она построена на раздельном произнесении и последующем распознавании слова или словосочетания из небольшого заранее заданного словаря.
- Распознавание по грамматике. Суть технологии — распознавание фраз, соответствующих определенным заданным правилам (грамматике). Чтобы ее реализовать, для задания грамматик используются стандартные XML-языки (VoiceXML), а обмен данными между системой распознавания и приложением, как правило, осуществляется по протоколу управления медиаресурсами (Media Resource Control Protocol, MRCP). Технология широко применяется в системах голосового самообслуживания (СГС): пользователя могут попросить произнести дату, какие-либо номера, фамилии, адреса, подтвердить какое-либо действие словами «да» или «нет».
- Поиск ключевых слов (ПКС). Он строится на основе распознавания отдельных участков речи. В этом случае речь может быть как спонтанной, так и соответствующей определенным правилам. Произнесенная речь не полностью преобразуется в текст — в ней автоматически находятся лишь те участки, которые содержат заданные слова или словосочетания. ПКС применяется в поисковых системах, а также в системах мониторинга речи.
- Распознавание слитной речи на большом словаре. (Large

Vocabulary Continuous Speech Recognition, LVCSR). Самая сложная технология: она переводит речь в текст, не ограничиваясь при этом какой-либо наперед заданной грамматикой. Иногда ее называют STT (speech-to-text), поскольку данная технология больше других приближает человека к мечте о его взаимодействии с компьютером. Задача полноценного распознавания слитной речи не решена нигде в мире, однако достоверность распознавания уже достаточно высока для использования технологии на практике: например, на телевидении (для создания скрытых субтитров) или в медицине (для ввода данных в электронные карты пациентов).

Синтез речи

Синтез речи — это технология, которая дает возможность прочитать текст (документ, письмо, sms) голосом, приближенным к естественному. Чтобы синтезированная речь звучала натурально, необходимо решить целый комплекс задач, связанных как с обеспечением естественности голоса на уровне тембра, плавности звучания и интонации, так и с правильной расстановкой ударений и пауз, расшифровкой сокращений, чисел, аббревиатур и специальных знаков.

На практике технологии синтеза речи применяются для озвучивания новостных RSS-каналов, субтитров, собственного контента, а также при создании голосовых открыток. Более того, синтез речи не ограничивается использованием определенных голосов. Есть возможность реализовать уникальный голос «на заказ», например воссоздать голос великого актера Юрия Юрьева и реконструировать все его монологи, как это было сделано в Александринском театре в рамках программы сохранения культурного наследия России. Как правило, на создание нового голоса необходимо три-четыре месяца, в зависимости от требуемого качества звучания, а для создания голоса на новом языке — до полугода.

Голосовая биометрия

Согласно ГОСТ ISO/IEC 24713-1-2013, биометрия есть автоматизированное распознавание личности человека, основанное на его поведенческих или биологических характеристиках. Соответственно, голосовая биометрия есть автоматизированное распознавание личности по фонограммам речи. Основными режимами распознавания являются верификация (подтверждение личности диктора) и идентификация (установление (поиск) диктора). Термин «диктор» введен тем же стандартом и означает говорящего человека.

Уникальность голоса человека обусловлена множеством физиологических причин — строение голосовых связок, трахеи, носовых полостей, манера произношения звуков, расположение зубов и др. Комбинация всех этих характеристик так же индивидуальна, как и отпечатки пальцев. Однако на практике ни одна из унимодальных биометрических систем, в том числе и голосовая, не может гарантировать 100% правильной идентификации. Использование бимодальной биометрии (по голосу и лицу) имеет свои преимущества: повышение точности идентификации, возможность работы с большими базами данных с сохранением эффективности поиска, повышение устойчивости к атакам нарушителей и фальсификациям [6].

Технология, лежащая в основе голосовой биометрии, применима в любой стране мира, так как является независимой от перечисленных выше характеристик: не имеют значения язык речи, акцент диктора, используемый диалект, содержание произносимой речи и т. д.

Анализ и обработка речи

К технологиям анализа и обработки речи относят быстрый поиск ключевых слов в аудиозаписях, автоматический анализ и оценку телефонных переговоров, интеллектуальный анализ речевой информации. Данная технология отличается простотой использования и точностью поиска в фонограммах, которая определяется размером поискового словаря. Так, для словаря из пяти слов надежность поиска составляет не менее 95%, для словаря из 100 слов — 81%.

Интеллектуальный анализ речевой информации позволяет автоматически определять тематику телефонных переговоров. В основе анализа лежат технологии распознавания слитной речи на большом словаре LVCSR и извлечения информации с помощью кластерного анализа данных (Data Mining Clustering). В результате автоматического распознавания речь дикторов преобразуется в текстовый индексированный файл, пригодный для автоматического лексико-семантического анализа. Решение о принадлежности аудиозаписи к абстрактному тематическому кластеру проводится с учетом частотности и связности слов и словосочетаний, употребляемых дикторами в ходе телефонной беседы (рис. 1).

СОВРЕМЕННЫЕ ПРИНЦИПЫ АНАЛИЗА И СИНТЕЗА РЕЧИ

Информация, заключенная в речевом сигнале, может быть разделена на основную (языковую), заключающуюся в передаче смыслового содержания речи, а также дополнительную (неязыковую), к которой относят информацию о характеристиках передающей среды и паралингвистическую (экстралингвистическую) информацию и др.

К характеристикам передающей среды обычно относятся уровень и тип шума окружающей среды (офисные шумы, шумы улицы, фоновая музыка, голоса других людей и т. д.), уровень реверберации (степень наложения на речевой сигнал его отражений от различных поверхностей), шумы и искажения в канале передачи (микрофоны, усилители, АЦП, кодеки и т. д.).

Знание характеристик передающей среды помогает решать задачи шумоочистки и улучшения качества речевых сигналов, а также оценивать их пригодность для последующего использования в системах автоматического распознавания речи и голоса. Так, например, точность большинства систем автоматического распознавания речи и голоса резко ухуд-

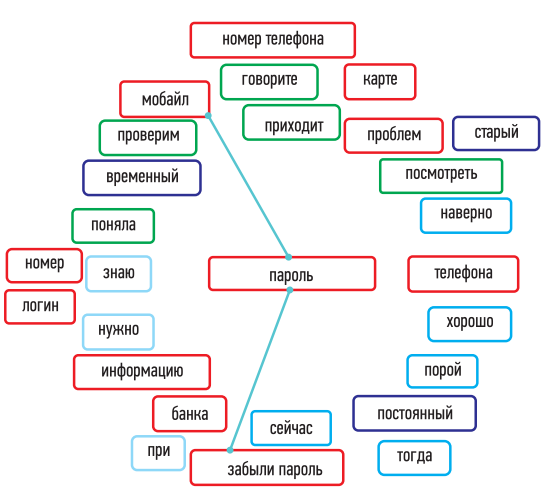


РИС. 1. Пример семантического облака темы «Восстановление пароля»

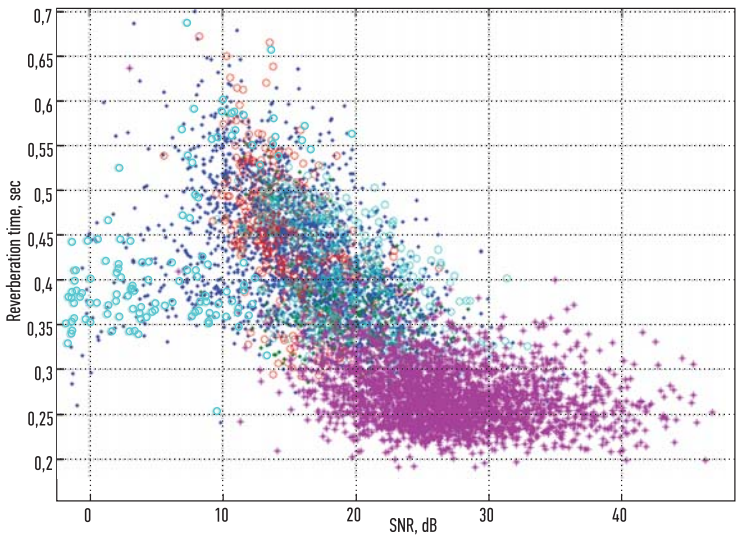


РИС. 2. ▶ Скаттерограмма корпусов речевых данных NIST: фиолетовый цвет — сотовый корпус, остальные — корпуса речевых данных в акустике помещений

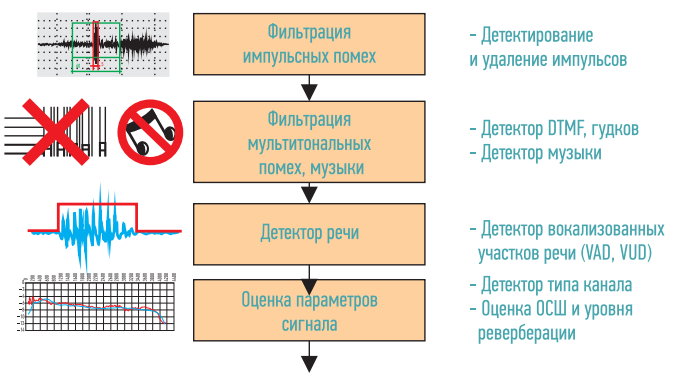


РИС. 3. ▶ Предобработка и оценка качества речевого сигнала

шается при снижении отношения сигнал-шум менее 15 дБ, увеличении уровня реверберации более 0,4 с. Речевые сигналы с «пригодными» параметрами характерны, в основном, для каналов телефонной связи (рис. 2). Речевые сигналы в акустике помещений имеют значительно худшие параметры, что приводит к низкой точности распознавания речи и голоса на таких данных.

Предобработка и оценка качества речевого сигнала (рис. 3) предполагает разделение фонограмм на сегменты с полезным речевым сигналом и сегменты с шумом, паузами, телефонными и музыкальными сигналами. Кроме того, выполняется оценка качества речевого сигнала для оценки его пригодности для распознавания речи и голоса.

Анализ и обработка речевых сигналов обычно производится не во временной, а в частотно-временной области. Для этого осуществляется кратковременное преобразование Фурье по квазистационарным фрагментам речевого сигнала длительностью 20–25 мс со сдвигом на половину фрагмента. В результате получается так называемая сонограмма (спектрограмма) речи — визуальное отображение речи как функции времени (горизонтальная ось), частоты (вертикальная ось) и энергии голоса (степень зачернения, цвет). Наиболее темные горизонтальные полосы частот показывают спектральные максимумы (рис. 4, [7]).

Речевой сигнал является по своей природе полиинформативным, что проявляется в многообразии типов информации, передаваемой с помощью речи. Так, на сонограмме можно выделить непрерывно следующие друг за другом сегменты различного уровня: фонемы, слова, фразы.

Фонема — наименьшая смыслоразличительная единица речи. Фонема не есть физическая реализация звука, а является представлением звука в сознании. «Фон» (phone) — конкретная реализация фонемы. Фоны, принадлежащие к одной фонеме, называются аллофонами. Звуковое окружение искажает форму фонемы, т. е. фонема в разных местах слова может быть не похожа сама на себя. Например, в похожих между собой словах «Даша» и «Маша» звук «а» звучит по-разному, так как речевой аппарат по-разному произносит одну и ту же гласную после звуков «д» и «м». Кроме того, разные люди в принципе по-разному произносят одни и те же звуки (рис. 5, [7]).

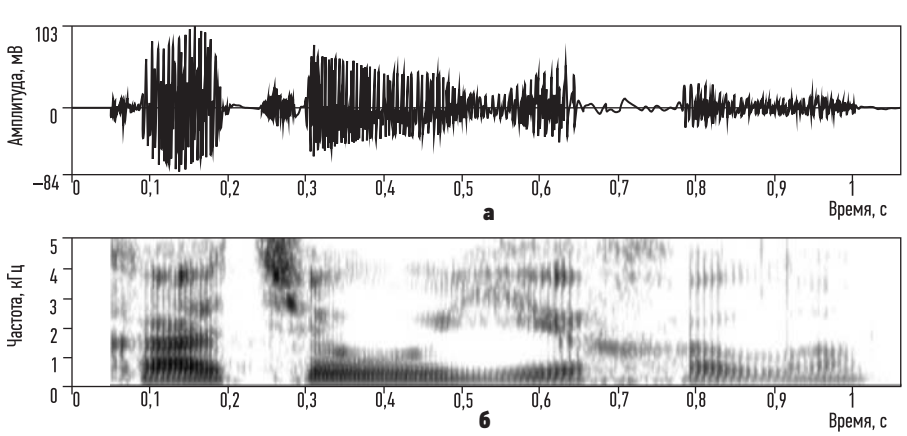


РИС. 4. ▼ Примеры представления речевого сигнала в виде: а) осциллограммы; б) сонограммы

Распознавание слитной речи

Одной из кардинальных задач распознавания речи является обеспечение устойчивости и стабильности распознавания фонов в условиях их огромной акустической вариативности. При этом слитная спонтанная речь труднее поддается автоматическому распознаванию по сравнению со слитной диктовочной речью из-за большей лингвистической («свободный» стиль речи, редукции, жаргонизмы, оговорки,

неканонические транскрипции, неправильная структура фраз), канальной (искажения и шумы в акустике помещений и каналах связи) и диктор-ской (индивидуальные особенности голосов дикторов, различный акцент, диалект, возраст и психофизическое состояние дикторов и др.) вариативности.

В таблице 1 приведены данные по точности современных систем распознавания слитной речи. Для сравнения: пословная ошибка распознавания речи человеком составляет 2–4%.

Самой современной технологией является распознавание слитной речи на основе многослойных нейронных сетей (Deep Neural Network, DNN). Сегодня ее используют все лидеры рынка речевых технологий. Эта технология имитирует работу человеческого мозга и позволяет распознавать несколько тысяч фонов. Факторы успеха: много/очень много качественных/не очень качественных обучающих данных (от сотен до десятков тысяч часов речи в реальных ситуациях), эффективный «тюнинг» модели и процедуры обучения.

На текущий момент имеются ресурсы консорциума лингвистических данных (Linguistic Data Consortium, LDC), речевые базы данных компаний-разработчиков (Google, Yandex, Baidu, ЦРТ и др.). Фонд перспективных исследований РФ определил, что одним из условий успешной реализации будущих систем распознавания речи является формирование речевых баз данных и словарей большого объема силами добровольцев (технологии краудсорсинга).

Синтез речи

Основными направлениями в разработке технологии синтеза являются:

- компилятивный синтез с использованием технологии Unit Selection (выбор звуковых элементов из речевой базы);
- синтез, генерирующий звуковой сигнал по параметрам, предсказанным на основе скрытых марковских моделей (Hidden Markov Models, HMM-синтез [2]).

Метод Unit Selection является разновидностью конкатенативного синтеза речи, т. е. в процессе синтеза речевого сигнала используются заранее сделанные звукозаписи естественной речи.

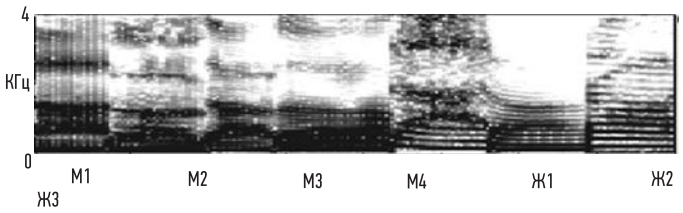


РИС. 5. ◄
Сонограмма звука «а», произнесенного четырьмя мужчинами и тремя женщинами

В процессе акустического синтеза алгоритм строит оптимальную последовательность звуковых единиц, учитывая одновременно и то, насколько кандидат подходит под описание необходимых характеристик звука (стоимость замены), и то, насколько хорошо выбранные элементы будут конкатенироваться с соседними (стоимость связи). Такой подход позволяет минимизировать модификации речевого сигнала, что повышает естественность синтезируемой речи.

В случае HMM-синтеза производится описание звуковой базы данных параметрической моделью. Параметры (например, спектральные характеристики, частота основного тона, длительность и т. д.) обобщаются множеством статистических моделей, которые содержат в себе шаблоны речевых элементов. Определение параметров речевого сигнала происходит на основе критерия максимального правдоподобия применительно к этим моделям. Синтез речи, основанный на моделях, реализован в компаниях Microsoft и Whistler.

В качестве схем, объединяющих HMM-синтез и Unit Selection, могут применяться следующие: генерация физических параметров звуковых элементов на основе скрытых марковских моделей для последующего вычисления стоимости замены для метода Unit Selection; использование статистических моделей для вычисления стоимости связи между элементами и т. п.

Голосовая биометрия

Для извлечения характеристик голоса диктора сначала осуществляется разделение дикторов на фонограмме [5]: выделяется речь на фоне акустических помех (создаваемых телевизором, радио и т. п.); разделяется речь на участках, содержащих речь нескольких дикторов, которая может налагаться друг на друга, образуя «голосовой коктейль». Выделенные участки речевого сигнала размечаются по принадлежности различным дикторам.

Далее в выделенных участках речевого сигнала производится автоматическое извлечение биометрических признаков голоса и речи. Экспертами традиционно используются акустические признаки: частота основного тона диктора (частота смыкания/размыкания голосовых связок), формантные частоты (резонансные частоты голосового тракта) и их траектории. В автоматических методах используются различные кепстральные признаки, таких как MFCC, LFCC, LPCC и т. д. [3].

В статистических методах верификации/идентификации модель голоса диктора представляет собой аппроксимацию распределения извлеченных признаков смесью гауссовых распределений (GMM-модель) [3].

Процедура распознавания диктора заключается в автоматическом попарном сравнении «голосовых моделей», в которых закодированы индивидуальные (биометрические) характеристики голоса и речи дикторов. Следует отметить, что совсем недавно системы распознавания по голосу обладали значительно худшими рабочими характеристиками (точность распознавания, размер биометрической модели и т. д.) по сравнению с системами других биометрических модальностей.

ТАБЛИЦА 1. ПОСЛОВНАЯ ОШИБКА РАСПОЗНАВАНИЯ WER (%) В РАЗЛИЧНЫХ СОВРЕМЕННЫХ СИСТЕМАХ РАСПОЗНАВАНИЯ СЛИТНОЙ РЕЧИ

Задача	Английский	Русский	Французский
Новости, интервью в медиаканалах	7–15	20–40	~15
Спонтанная телефонная речь	10–25	25–45	~20

ТАБЛИЦА 2. УРОВНИ ОШИБОК ИДЕНТИФИКАЦИИ ДЛЯ РАЗЛИЧНЫХ БИОМЕТРИЧЕСКИХ МОДАЛЬНОСТЕЙ

Биометрический признак	Тест	Условия тестирования	FRR, %	FAR, %
Отпечатки пальцев	FVC 2006	неоднородная популяция, включая работников ручного труда и пожилых людей	2,2	2,2
Лицо	MBE 2010	полицейская база фотографий; база фотографий с документов	4,0 0,3	0,1
Голос	NIST 2012	текстонезависимое распознавание	2–3	1
Радужная оболочка глаз	ICE 2006	контролируемое освещение, широкий диапазон качества	1,1–1,4	0,1

Однако за последние 5–7 лет в области голосовой биометрии были достигнуты значимые успехи [3], которые позволили приблизить рабочие характеристики голосовой модальности к другим модальностям, в особенности к лицевой (табл. 2, 3).

Основными режимами распознавания диктора являются текстозависимый, по комбинации из ключевых фраз или по наборам из 10 цифр, текстонезависимый по читаемому тексту или по разговорной речи. Первые три режима обеспечивают высокий уровень точности распознавания, но требуют произнесения заранее подготовленного текста. Эти режимы не всегда удобны для пользователя и не обеспечивают должного уровня защиты в системах безопасности.

На практике наиболее востребован текстонезависимый режим, когда пользователь общается с системой на естественном языке. Однако основной проблемой при решении задачи текстонезависимого распознавания диктора является проблема рассогласования, вызванная изменчивостью сессий записи голоса для отдельного диктора. Причинами этого рассогласования могут быть шумы окружающей среды при записи, искажения в каналах записи и передачи речевого сигнала, а также изменчивость голоса самого диктора. Учет эффектов канала является самым значимым фактором среди перечисленных.

Для решения указанной проблемы традиционным стало применение совместного факторного анализа (Joint Factor Analysis, JFA) [3], который позволяет эффективно расщеплять дикторскую и канальную информацию в отдельном произнесении диктора, что, в свою очередь, позволяет строить канало-независимые GMM-модели диктора и подавлять эффекты канала в тестовом произнесении. Дополнительно к порождающему методу GMM в системах голосовой биометрии популярным является дискриминантный метод распознавания диктора — машины опорных векторов (Support Vector Machine, SVM). Гибридные системы SVM-GMM и GMM-JFA-SVM [5] обладают лучшей эффективностью как по параметрам точности (более робастны к различного рода шумам, а также к межсессионной и внутридикторской вариативности), так и по параметрам быстродействия.

ОТРАСЛЕВЫЕ РЕШЕНИЯ

Разработки в области речевых технологий пользуются нарастающим спросом во многих отраслях: государственный сектор, финансы, здравоохранение, правовая и судебная системы, медиакоммуникации, военная промышленность. Основной причиной, подстегивающей научно-исследовательский и бизнес-интерес к данному направлению, является рост спроса на решения для оптимизации рутинных процессов на производстве и в бизнесе.

ТАБЛИЦА 3. РАЗМЕРЫ МОДЕЛЕЙ ДЛЯ РАЗЛИЧНЫХ БИОМЕТРИЧЕСКИХ МОДАЛЬНОСТЕЙ

Биометрическая модальность	Размер модели, кбайт
Отпечатки пальцев	0,25–1,2
Лицо	0,1–2,0
Голос	2–3
Радужная оболочка глаз	0,25–0,5

Правительство и государственные структуры

Одним из драйверов роста российского рынка речевых технологий выступают государственные и силовые структуры [8]. Помимо устройств шумоочистки и записи речевых сигналов, начинают внедряться системы криминалистического учета по голосу и лицу [9]. В некоторых органах законодательной и исполнительной власти РФ сейчас проходит внедрение систем подготовки стенограмм заседаний с использованием технологии слитного распознавания русской речи.

Контактные центры

На рынке речевых технологий настоящий «бум» переживают так называемые «системы голосового самообслуживания» (IVR), которые активно внедряются в контактные центры различных компаний и в контактные центры, работающие на аутсорсинге.

Технологии, лежащие в основе систем голосового самообслуживания, постоянно развиваются: помимо предоставления справочной информации и обработки типовых запросов в автоматическом режиме, перед контактным центром ставятся задачи по созданию виртуальных консультантов с возможностью искусственного интеллекта.

Растет число проектов по внедрению в контактные центры систем речевой аналитики, систем управления качеством работы операторов и оценки удовлетворенности клиентов. Использование этих систем открывает возможности по определению уровня стрессоустойчивости и психофизического состояния операторов, анализу причин повторного

обращения клиентов, определению уровня их лояльности и удовлетворенности.

Финансовый сектор

Существенный рост числа мошеннических кредитов за последние несколько лет привел к тому, что банки стали активно внедрять решения на базе речевых технологий для снижения рисков мошенничества, защиты существующих клиентов и повышения доверия к банку.

Исходя из необходимости удаленного обслуживания клиентов, например при подтверждении личности оператором контактного центра, а также на сайте, при входе в личный кабинет, или в мобильном приложении, банки все больше склоняются к использованию технологий голосовой биометрии. В условиях удаленного обслуживания голос человека становится наиболее надежным способом верификации пользователя, поскольку его, в отличие от любой другой информации, нельзя украсть или подделать.

Здравоохранение

На рынке речевых технологий существуют решения для здравоохранения, которые нацелены на повышение производительности труда медиков. Технология распознавания речи врача и автоматического занесения информации в медицинскую систему может применяться при заполнении карты при осмотре пациента, а также при работе в операционной. В качестве эффекта от внедрения данной технологии наблюдается рост количества обследований, экономия денежных средств за счет возможности отказаться от услуг медсестер и помощников, а также экономия времени врача.

Службы безопасности

В коммерческих организациях, а также на объектах промышленного и гражданского назначения, в городском общественном транспорте, в образовательных учреждениях и учреждениях развлекательной сферы реализуются проекты по внедрению в службы безопасности систем эффективной охраны периметра и эвакуации за счет организации сплошного видеонаблюдения и автоматического оповещения ответственных лиц по различным каналам связи.

Для профилактики утечки информации и разбора происшествий службами безопасности применяется анализ речи и эмоционального состояния и централизованная система регистрации переговоров сотрудников, диспетчеров, операторов.

Судебная система

В правительственных структурах, так же как и в судебных органах, используется система стенографирования для повышения мобильности и скорости подготовки стенограмм одновременно с нескольких заседаний. Современные речевые технологии позволяют осуществлять синхронную запись речи выступающего с его видеоизображением, а также провести подготовку протокола в автоматизированном режиме с использованием технологии распознавания слитной речи.

Интернет и телевидение

При ежедневном использовании Интернета и телевидения есть большая вероятность столкнуться с применением речевых технологий. Например, организация онлайновых трансляций спортивных игр с субтитрами строится на базе технологии распознавания речи [14], технология распознавания отдельных команд используется для внедрения сервиса голосовой навигации по сайту в Интернете, а проверка личности по голосу — для развлекательных Интернет-ресурсов и корпоративных порталов.

Автомобильная промышленность

Применение речевых технологий в автомобильной промышленности открывает новые возможности для использования развлекательных и сервисных функций, которые были недоступны автопроизводителям ранее. Универсальные голосовые решения для автомобилей упрощают способы взаимодействия водителя и пассажиров с мультимедиа и навигационной системой, тем самым снижая аварийную опасность, не отвлекая водителя от управления автомобилем, в отличие от привычных бортовых компьютеров, требующих повышенного внимания. Применение голосовой биометрической аутентификации водителя позволяет снизить риск угона автомобиля.

РАЗРАБОТКИ И ДОСТИЖЕНИЯ ЦРТ

«Центр речевых технологий» (ЦРТ) вырос из небольшой команды единомышленников до крупной ИТ-компании, занимающей ключевые позиции на рынке речевых технологий и мультимодальной биометрии, как в России, так и за рубежом.

Первыми крупными заказчиками ЦРТ стали правоохранительные органы, для которых компания

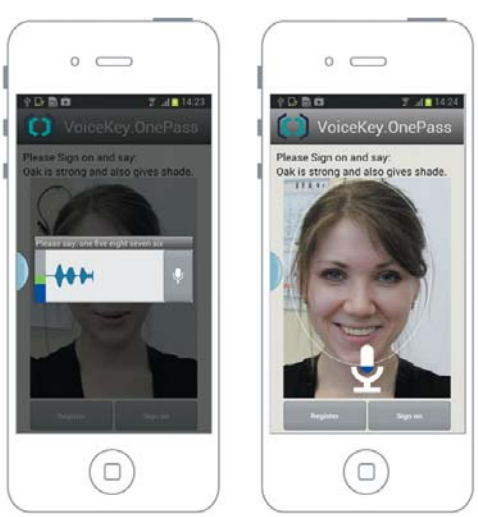
РИС. 6. ◄
Профессиональные устройства записи аудио- и видео-сигналов: интерактивная система аудиовидеонаблюдения AVIDIUS и диктофоны серии «Гном»

РИС. 7. Программный комплекс шумоочистки аудиозаписей Sound Cleaner

РИС. 8. Приложение «Читатель» для озвучивания электронных книг

разработала специализированный звуковой редактор для экспертов-криминалистов SIS, а также устройства шумоочистки серии «Золушка». После выхода на международный рынок в 1997 г. началось активное сотрудничество с такими известными компаниями, как Intel (США), Samsung (Корея), SWATCH (Швейцария).

РИС. 9. Система VoiceKey. ONEPASS — бимодальное решение для защиты мобильных приложений от несанкционированного доступа



Накопив научный потенциал, а также опыт ведения крупных проектов, в том числе международных, ЦРТ занялся созданием системы многоканальной записи, обработки и анализа вызовов «Незабудка», разработкой системы стенографирования «Нестор», серийным производством диктофонов серии «Гном» (рис. 6), продажами программного комплекса шумоочистки Sound Cleaner (рис. 7), выводом на рынок систем идентификации по голосу и лицу.

В начале 2000-х годов сотрудников ЦРТ начинают приглашать в качестве аудиоэкспертов для участия в расследовании крупных катастроф, например аварии на АПЛ «Курск» или дела о захвате заложников «Норд-Ост».

Первые шаги по созданию программ синтеза и распознавания русской речи были сделаны в 2007 г., когда ЦРТ получил премию Мининформсвязи в области качества, а также был признан лучшей компанией в области технологий шумоочистки на конгрессе AES в Денвере.

В 2010 г. ЦРТ успешно завершил проект по внедрению первой и крупнейшей в мире биометрической системы национального масштаба по заказу правительства Мексики, а в 2012 г. в МВД Эквадора была внедрена первая в мире интегрированная система биометрического поиска и национального криминалистического учета по голосу и лицу для поиска преступников.

Следуя общей тенденции развития мобильных приложений, ЦРТ расширил линейку продуктов и предложил рынку несколько приложений: для озвучивания новостных RSS-каналов — Radio RSS, «Читатель» (рис. 8) для озвучивания электронных книг, приложение-караоке Sing&Fly, а также выпустил уникальное решение для защиты мобильных приложений от взлома — VoiceKey. ONEPASS (рис. 9).

За прошедший, 2014 г. ЦРТ может гордиться не одним крупным проектом: трансляция закрытия Паралимпийских игр в Сочи с онлайн-субтитрами, создание виртуального консультанта «Елена» для службы клиентского сервиса «МегаФона», внедрение первой в России системы биометрической идентификации болельщиков на стадионе «Петровский» — SmartTracker.Arena.

В копилку побед ЦРТ добавилось первое место на всемирном конкурсе NIST i-vector Machine Learning Challenge 2014 за разработанную технологию идентификации диктора. Кроме этого, компания «ЦРТ-инновации» стала третьей российской организацией, представленной в сообществе FIDO Alliance (Fast Identity Online Alliance), в числе которых такие международные гиганты, как Microsoft, Google, LG Electronics и др.

Следуя мировым тенденциям, ЦРТ также вносит свой вклад в создание искусственного интеллекта.

***

Особенности современной ситуации на рынке речевых технологий:

- рынок речевых технологий и средств компьютерной обработки речи — один из самых быстрорастущих на сегодня;
- использование современных речевых решений позволяет оптимизировать внутренние процессы компаний и снизить затраты практически во всех отраслях;
- в основном компании вкладывают средства в разработку автоматического распознавания речи, технологию преобразования текста в речь и систему верификации спикера;
- лидерами среди разработчиков являются США, Великобритания, Япония, Израиль и Россия, однако по прогнозируемым темпам роста впереди находятся страны Азиатско-Тихоокеанского региона. ⬣

Полный текст статьи размещен на сайте журнала