

КИБЕРСЛУХ

Есть у Voisi необычный клиент — фирма из Германии, которая проводит турниры по киберспорту. В ходе турниров участники команд постоянно ведут переговоры, для этого есть специальный голосовой чат. После матчей переговоры принято анализировать, чтобы понять, что было сделано правильно, а что — нет. От этого зависит качество следующих турниров. «Наша речевая аналитика позволяет тренеру и команде быстрее проводить разбор матчей», — говорит Фирун. — Мы даже не догадывались о таком применении нашей технологии, но у него хорошие перспективы. Наш продукт востребован в мире киберспорта. После того как мы стали работать с ребятами из Германии, к нам с похожим запросом обратились разработчики из США».

но в нем намного больше ядер, что позволяет распараллеливать процесс намного эффективнее, чем на обычном процессоре», — объясняет Фирун. Благодаря видеокартам обработка звука шла быстрее, чем у конкурентов. В результате получилась промышленная технология, которая существенно ускоряла и удешевляла процесс распознавания речи. «Нашей идеей впоследствии воспользовались куча компаний, но мы были первыми, — подчеркивает Константин. — Мы неплохо зарабатывали на оцифровке медиаархивов зарубежных телерадиокомпаний, однако потребителей в России наш продукт не нашел».

Смысл, а не буквы

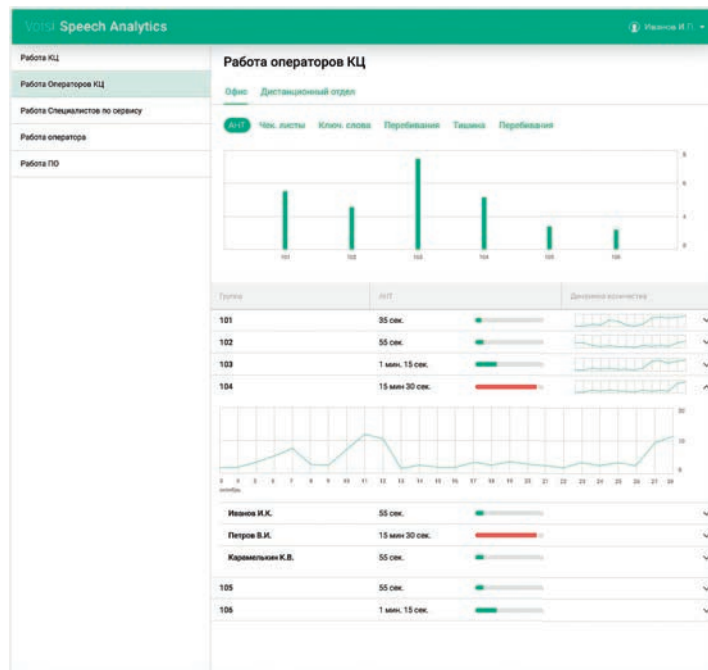
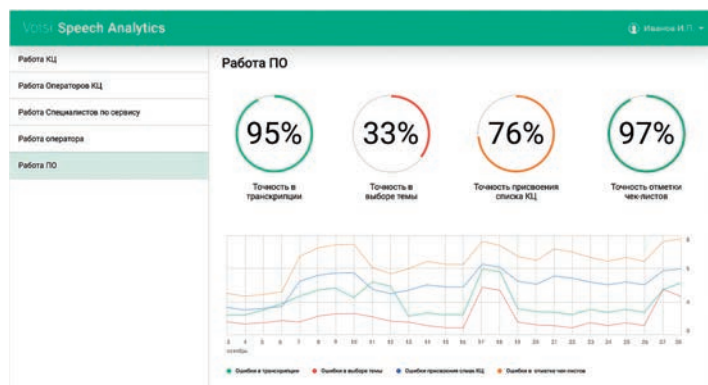
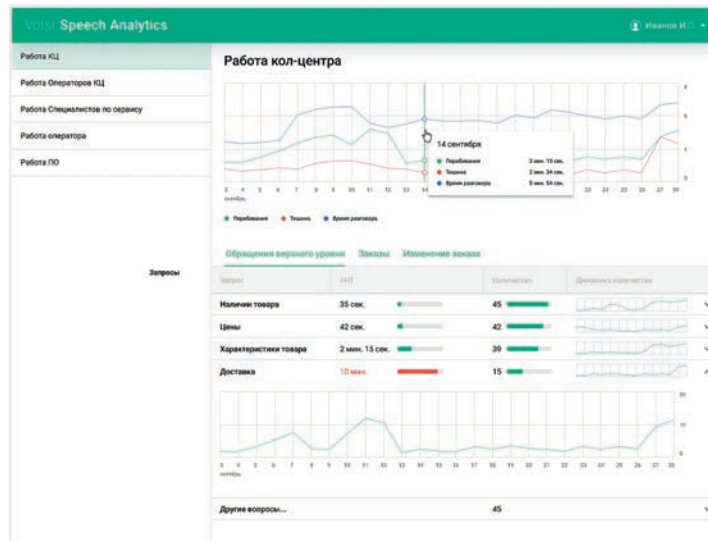
С 2011 года Константин Фирун разрабатывает собственную технологию распознавания речи, позволяющую автоматизировать контроль за работой контакт-центров. «В начале 2014-го нас взял под крыло выпускник нашего университета. Он инвестировал в наши речевые технологии и продукты на их базе. Так родилась компания Voisi». Сегодня технология уже существует в виде программного продукта, пользователь может получить комплексную услугу.

Большая часть алгоритмической базы распознавания речи известна с середины 1980-х и хорошо описана в научной литературе. В последние годы исследователи вели гонку за улучшение качества и скорости работы этих технологий, но результаты улучшались не так значительно, как в конце XX века.

«Мы сфокусировались на библиотеках с открытым исходным кодом (программное обеспечение, которое предоставляется с открытым исходным кодом и которое каждый может отредактировать под свои цели — примерно как «Википедия», которую каждый может дополнить), которые стали стандартом де-факто для разработчиков речевых технологий. Нам удалось ускорить работу этих библиотек и сделать их стабильнее. А для ускорения их работы мы использовали вычисления на видеокартах. По сути, это был технологический прорыв: десять лет назад такой способ казался экзотикой, хотя сегодня использование видеокарт для просчета задач, не связанных с компьютерной графикой, становится все более популярным», — объясняет Константин.

Похожий сервис предлагают и другие компании, в том числе интернет-гиганты: письменное распознавание речи есть и у Google, и у Apple, с прошлого года у Amazon есть платная услуга Transcribe, переводящая речь в текст, существует множество небольших бесплатных сервисов вроде dictation.io или speechpad.ru. Но у Voisi есть важные отличия. Во-первых, набор тематических словарей: в автосалоне, например, совершенно иной набор терминов, нежели в больнице. Со словарем, адаптированным под определенные темы, качество распознавания речи существенно повышается. Сейчас компания использует только русский и английский языки, но уже идет работа над интеграцией европейских: Voisi планирует выходить на мировой рынок.

Во-вторых, Voisi обрабатывает данные не только по облачной модели на собственных серверах, но и на



СТЕНОГРАФИСТ-АВТОМАТ

Технологию Voisi можно приспособить к расшифровке стенограмм — аудиозаписей, которые ведутся во время всевозможных заседаний и конференций. И журналистские интервью тоже можно с ее помощью достаточно точно переводить в письменный текст. Правда, пока такой текст не получится в форме диалога, потому что в телефонном разговоре два канала и в каждом отдельный голос, а в записи интервью — только один с несколькими голосами. Но в будущем появятся технологии, которые смогут разделять их автоматически. «Мы разрабатываем такие технологии, — говорит Константин Фирун. — Думаю, что создание устойчивых алгоритмов, которые смогут различить разные голоса на одной записи, появятся у нас и других разработчиков через два-три года».

сервера заказчика. Это важно для банков, которые должны хранить персональные данные клиентов в своей компьютерной сети. Кроме того, Voisi не требует для своих услуг высокопроизводительных серверов — достаточно обычных персональных компьютеров, которых много в любом офисе. А это значит, не нужно закупать дорогое оборудование.

В-третьих, не все конкуренты Voisi предлагают заказчику полнотекстовую расшифровку. Многие ограничиваются тем, что вычлениают в разговорах ключевые слова. Voisi полностью расшифровывает телефонные разговоры, превращая их в текст в форме диалога. По этим текстам можно вести поиск.

«Мы считаем, что заказчики должны проверить, как работает наш продукт, причем бесплатно, без ограничения по срокам и размеру записей, — указывает еще на одно преимущество своей компании Константин Фирун. — Технологии распознавания образов, в частности речи, — это вероятностные технологии, поэтому при работе с ними невозможно прогнозировать, какой будет точность. Недавно, например, нам загрузили запись, где звонят узбеки, которые плохо говорят по-русски. Программа выявила в их речи только ключевые слова. Поэтому мы всегда проводим тестирование, чтобы и мы, и клиенты были уверены, что все работает в соответствии с их требованиями».

Сейчас Voisi ведет несколько тестирований в автосалонах, строительных компаниях, риэлторских агентствах.

«Через полтора-два квартала мы планируем выйти на самоокупаемость. У нас уже есть клиенты, которые нам платят, в частности компания iGoods — сервис доставки продуктов в Москве, Санкт-Петербурге и других городах России, банк «Оранжевый», работающий в Москве и Петербурге. Нашей системой заинтересовались крупные ритейловые и логистические сети, перед которыми остро стоит задача речевой аналитики. Им не нужно объяснять, зачем нужен такой продукт, потому что они сами его ищут», — заключает Константин Фирун.

ЕЛЕНА ТУЕВА

1. Сводная статистика работы колл-центра ритейлера за сентябрь 2018 года. На графике можно увидеть динамику колл-центра по основным метрикам: времени разговора, тишине, числу перебиваний оператора клиентом и наоборот. Наиболее частые запросы клиентов тут тоже можно увидеть. Тишина, то есть молчание клиента и оператора, — важная метрика, потому что молчание в телефонном разговоре — это впустую потраченные деньги бизнеса. Подсчет перебиваний необходим для оценки эмоциональной составляющей разговора
2. Детальная информация о работе групп операторов в колл-центре ритейлера. На графике показаны основные метрики: Average Handling Time (общая длительность разговора), соответствие чек-листам / скриптам, детектированные ключевые слова (то есть найденные в расшифровке разговора), перебивания и молчание. Программа сравнивает распознанную речь с эталонным скриптом (сценарием), по которому должен общаться сотрудник, и выставляет сотруднику оценку. Для сравнения используется разработанный Voisi алгоритм с поддержкой морфологии русского языка, а также учетом редакционного расстояния (необходимо для операций над текстом расшифровки, чтобы сверять слова в расшифровке с теми, которые написаны в сценарии сотрудника)
3. Телеметрия работы программного обеспечения Voisi, доступная администратору колл-центра. На экран выведены данные о точности распознавания речи, собранные с пользователей программы. Телеметрия рассчитывается на основе данных, которые генерируют пользователи: сколько времени они проводят на каждой странице, какую оценку качеству распознавания выставляют, как много ошибок исправляют в расшифровке
4. Так выглядит на экране текстовая расшифровка телефонных переговоров