

Речевые технологии на СеВIT 2008

М.В. Хитров,

*генеральный директор компании
«Центр речевых технологий»*

Анализ компаний, представленных на выставке СеВIT 2008, занимающихся разработкой и развитием речевых технологий, позволил выявить приоритетные направления этого рынка. Появляется всё больший спрос на системы голосовой биометрии. Этот интерес исходит прежде всего от интеграторов, которые уже сейчас представили на рынке системы верификации по отпечатку пальца и радужной оболочке глаза. Ниша голосовой биометрии относительно свободна и рынок только начинает формироваться. Те системы, которые существуют у конкурентов, явно проигрывают по качеству и надёжности системе Voice Key, разработанной компанией ЦРТ. Сотрудники Aixvox, посетившие стенд ЦРТ, отметили, что Voice Key отличается хорошей шумоустойчивостью, т.е. сохраняет заявленную надёжность даже в условиях повышенного шумового фона, и скоростью работы.

Ещё один быстроразвивающийся рынок речевых технологий — системы синтеза речи и системы распознавания команд на основе пофонемного распознавания. Сегодня существуют системы хорошего качества TTS (text-to-speech) и STT (speech-to-text) для основных европейских языков — английского, немецкого, французского, испанского, португальского: для этих языков рынок уже сформировался, и лидирующие позиции заняли компании, имеющие готовые решения для разработчиков. Лидирующий игрок — Nuance, постепенно упрочивает свои позиции компания Loquendo. На выставке СеВIT компания Nuance выставлялась в павильоне навигации и решений для автомобилей: у них был маленький стенд и отсутствие всякого демонстрационного материала. Это свидетельствует о том, что Nuance главным образом работает на рынке B2B, т.е. на рынке конечных потребителей, она практически не работает. В павильоне было много компаний, которые демонстрировали голосовое управление для GPS навигаторов и автомобильных компьютеров, вплоть до своего рода «мобильных офисов» (InPhoDrive. Talking information) <http://www.inphodrive.com/home.html>.

Сотрудники ЦРТ протестировали систему распознавания команд и синтеза речи у InPhoDrive — надёжность очень высокая (около 99%). При этом сама компания только появилась и выходит на рынок (10 человек, 1 год). Также на рынке речевых технологий для автопрома есть компании-интегратор, например, Voice Insight (<http://www.voice-insight.com>). Они стали интеграторами для компании, которая выпускает системы управления проигрывателем и навигатором для автомобиля. При этом их работа — это выбор движка (Nuance или Loquendo), встраивание в интерфейс. Представитель бельгийского подразделения отмечает, что компания готова работать с русским языком. По его словам, компания Nuance имеет движок для русского языка, но они его не демонстрировали.



Несмотря на ведущие позиции компаний, которые уже имеют TTS и STT, многие интеграторы ищут альтернативные решения. Компания LinguaTEC встраивает движок ViaVoice компании IBM в свои продукты — распознавание слитной речи (рынок медицинских и юридических услуг), синтез речи, переводчики. Синтез, который смотрел сотрудник ЦРТ для немецкого и английского языков, отмечен им очень высоко. Для распознавания слитной речи, по словам представителей компании, требуется около одного часа тренировки на диктора. Но немецкую речь нашего сотрудника программа не распознала.

В целом сложилось впечатление, что рынок для синтеза качества естественной речи, фонового распознавания голосовых команд и распознавание слитной речи уже сложился, многие компании разрабатывают и предлагают готовые решения. Ниша русского языка (а также других славянских языков) свободна, и будут успешны те компании, которые в скором времени предложат готовые решения. Интерес вызван большой диаспорой русскоязычного населения в Европе, Израиле и наличие хорошего рынка в России.

В области дикторонезависимого распознавания русского языка Центр речевых технологий обладает весьма продвинутыми разработками, что даёт возможность оформлять технологии распознавания в форме рыночного продукта (готового для встраивания в продукты для конечного пользователя).

Спрос на голосовое управление компьютером отмечается со стороны компаний и организаций, разрабатывающих и продающих продукты для людей с ограниченными возможностями.

Общая картина на рынке технологий распознавания характеризуется небольшим количеством компаний, имеющих собственные разработки алгоритмов распознавания. При этом такие компании практически не занимаются разработкой решений для конечного пользователя. Разработчики технологий распознавания речи поставляют на рынок компоненты, готовые для встраивания в разнообразные системы и программные продукты сторонних производителей. Наблюдаются ситуации, когда между компанией-разработчиком технологии распознавания и компанией-разработчиком готового решения существует компания-интегратор, задача которой — подбор технологий распознавания, удовлетворяющих требованиям продукта для конечного пользователя. Например, для встраивания системы голосового управления в автомобильные электронные устройства компания-производитель таких устройств пользуется услугами интегратора, который берёт на себя исследование технологий распознавания и встраивание их в продукт для конечного пользователя.

Отметим, что на современном рынке технологий распознавания почти отсутствует интерес к дикторозависимому распознаванию одиночных команд. Большинство демонстрируемых решений не предполагают обучения системы на конкретного диктора. Эта общая тенденция при очевидном повышении комфорта работы с системой накладывает и определённые ограничения — распознавание возможно только для определённого языка.