

ОЗВУЧИВАНИЕ ТРЕВОГ И СОБЫТИЙ В SCADA INDUSOFT WEB STUDIO С ПОМОЩЬЮ СИНТЕЗАТОРОВ РЕЧИ MICROSOFT SPEECH API

А.В. Барбот, С.А. Гусев (ЗАО «Первая миля»)

Представлен экономичный проект оснащения «речевым интерфейсом» системы управления помещением, базирующийся на использовании SCADA-системы Indusoft Web Studio, речевого синтезатора с открытым исходным кодом и программного интерфейса от Microsoft, предназначенного для синтеза (и распознавания) речи.

Ключевые слова: речевой интерфейс, синтез и распознавание речи, умный дом.

Использование обычного «человеческого» языка для взаимодействия с компьютером становится все более заметным трендом в последние годы. Мы уже почти привыкли к тому, что почти все используемые нами технические устройства, начиная с телефона и заканчивая автомобилем, могут общаться с нами на совершенно новом вербальном уровне. Голосовые ассистенты мобильных операционных систем уже вполне готовы не только помочь нам «пройти в библиотеку» или выдать нам наиболее популярные ссылки в Internet по запросу, но и поддержать разговор на общие темы, и даже иногда пошутить, конечно, в рамках дозволенного их разработчиками.

Мы давно не удивляемся тому, что наши автомобили уже лучше нас знают, где и куда нам лучше свернуть, чтобы быстрее достичь заданной цели, поскольку давно и незаметно для нас научились обмениваться информацией с навигационными спутниками, серверами Internet, обрабатывающими информацию о дорожном трафике и пробках, а в последнее время — и напрямую с другими участниками дорожного движения, а точнее с программами, работающими в их навигаторах или гаджетах их пассажиров. Естественно, что для принятия решения о том, куда повернуть на ближайшем перекрестке, навигационный компьютер получит информацию от других компьютеров по целому спектру протоколов, но результат будет озвучен специально для нас на понятном нам русском (или ином понятном нам) языке.

Неудивительно, что подобный стиль общения мы уже считаем «обычным» и для систем управления жилищами. Системы Building Automation все чаще оснащаются «речевыми интерфейсами» и, надеюсь, не далек тот день, когда мы в своих квартирах или домах, сказав некую кодовую фразу типа «Окей, гугл...» или «по шучьему велению, по моему хотению» (кто к чему привык), сможем изменить режим отопления (кондиционирования или вентиляции) или опустить жалюзи, или получить отчет о том, как работают инженерные системы, и не было-ли каких отказов или аварий...

Все это, как правило, доступно в достаточно дорогих системах управления зданиями. Но можно ли, используя более дешевые SCADA-системы, например InduSoft

Web Studio (IWS), и бесплатные (то есть распространяемые в виде исходных кодов) системы класса «text to speech», получить что-то подобное уже сегодня?

Рассмотрим пример реализации подобного приложения с использованием SCADA-системы InduSoft Web Studio, работающей в ОС Windows. Остановимся немного подробнее на SAPI (Speech Application Programming Interface) — программном интерфейсе Microsoft, предназначенном для синтеза (и распознавания) речи.

Распознавание речи — процесс преобразования произнесенных слов в печатный текст, включает: захват и оцифровку звука, произнесенного в микрофон; преобразование оцифрованного звука в фонемы; конструирование из фонем слов; анализ контекста, в котором было произнесено слово, и при необходимости замена слов на похожие по звучанию.

Синтез речи — процесс преобразования текста в произносимые слова — включает: разделение слов на фонемы; нахождение текста, требующего преобразование в символы, такого как числа, символы валюты и пунктуация; генерация цифрового звука для воспроизведения.

Механизмы преобразования текста в речь могут использовать один из двух способов синтеза голоса:

- генерация звуков, подобных издаваемым голосовыми связками, и использование различных фильтров для моделирования длины горла, формы ротовой полости, расположения губ и языка;
- сбор его из многочисленных кусков речи, записанных диктором.

Задача распознавания речи на порядок сложнее, чем синтез [1, 2]. Именно поэтому на сегодняшний день все более или менее доступные и отлаженные системы распознавания речи удовлетворительно работают только с одним единственным языком (английским), а все системы распознавания русского языка пока находятся в начальной стадии развития, не предлагая нам бесплатных и устойчиво работающих приложений. Поэтому в рамках данной статьи мы ограничимся только описанием системы синтеза речи, которая может быть использована любым из наших читателей.

В среде Indusoft Web Studio (IWS) в сценариях VBScript возможно использовать программный объект SpVoice интерфейса Microsoft SAPI, реализующий синтез речи из исходного текста. При этом доступны все его свойства и методы.

После анализа доступных на сегодняшний день вариантов реализации систем SAPI был выбран бесплатный речевой синтезатор RHVoice — свободно распространяемый речевой синтезатор с открытым исходным кодом, работоспособность которого была проверена, и качество оказалось более чем удовлетворительным¹.

Сначала необходимо установить основной пакет (на момент написания статьи RHVoice-v0.5-setup.exe), затем русский языковой пакет (RHVoice-language-Russian-v2.0-setup.exe), далее можно установить голоса (RHVoice-voice-Russian-Elena-v1.0-setup.exe, RHVoice-voice-Russian-Aleksandr-v1.0-setup.exe и др.).

Русскоязычные голоса "читают" текстовые строки, содержащие русские буквы, но неправильно читают английский текст, поэтому можно еще установить английский языковой пакет (RHVoice-language-English-v1.3-setup.exe) и голос CLB (RHVoice-voice-English-CLB-v1.0-setup.exe). После этого можно использовать парный голос, например, Elena+CLB для чтения смешанных текстов. В этом случае будет использован русскоязычный голос Elena и англоязычный CLB.

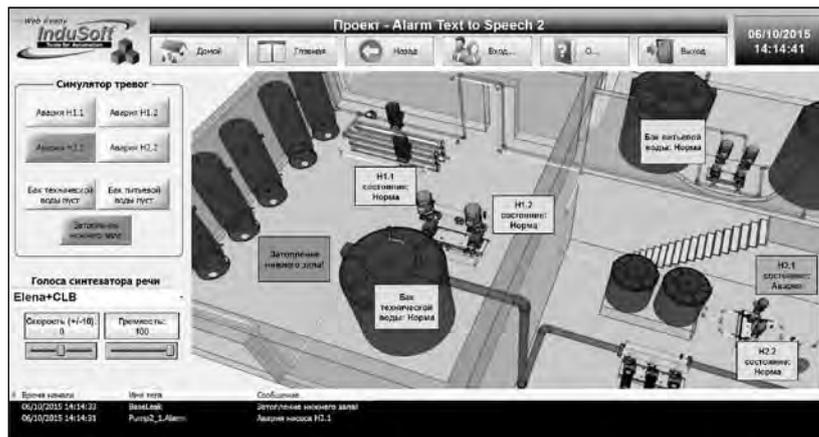
Проверить наличие установленных голосов можно в «Панели управления» ОС Windows в меню «Распознавание речи» → «Преобразование текста в речь». В поле «Выбор голоса» должны присутствовать установленные ранее голоса.

Alarm Text To Speech 2

Создадим проект, который содержал бы некоторое число событий и тревог, и «озвучим» эти события с помощью SAPI приложения.

За основу возьмем оригинальный пример проекта Alarm Text To Speech, которой можно найти на сайте <http://www.indusoft.com> в разделе примеров для изучения. Проект рекомендуется переименовать, например в Alarm Text To Speech 2, и начать его переработку и расширение. Основные изменения, которые следует в него внести, коснутся механизма получения сообщений активных тревог. Вместо VBScript класса SMA.Alarms будем использовать функции *GetAlarmCount()* и *GetAlarmInfo()* встроенного API Indusoft. Синтезатор речи озвучивает сообщения новых активных тревог.

¹ Скомпилированные установочные пакеты RHVoice доступны по адресу [https://github.com/Olga-Yakovleva/RHVoice/wiki/Downloads-\(Russian\)](https://github.com/Olga-Yakovleva/RHVoice/wiki/Downloads-(Russian))



В проекте присутствуют следующие глобальные функции, процедуры и фоновые сценарии:

GetVoicesList() — процедура возвращает список доступных голосов (для вывода в Combobox);

TextToSpeech(text) — функция "читает" переданные в качестве параметра текст;

Проигрывание новых тревог — фоновый сценарий, реализует наблюдение за активными тревогами и управляет вызовами функции TextToSpeech в отдельном асинхронном потоке (чтобы не тормозить выполнение фонового сценария).

В стартовом сценарии задается необходимый для нормальной работы скриптов формат даты Indusoft, и выбирается голос по умолчанию.

На мнемосхеме проекта можно выбрать голос синтезатора, скорость и громкость воспроизведения (рисунок).

Работа проекта проверялась на профессиональной версии ОС Windows 7 в Indusoft Web Studio v7.1 SP3.

Готовый проект можно скачать по ссылке http://www.firstmile.ru/indusoft/AlarmTextToSpeech2_IWS7.1.zip.

Специалистам несложно будет разобраться с примером использования технологии SAPI, и они смогут существенно «оживить» проекты управления зданиями (и не только) с помощью добавления речевого интерфейса. Возможно, в ближайшее время появятся также доступные и качественные подсистемы распознавания русской речи, и мы сможем создать для себя и своих близких полностью интерактивную среду обитания.

Список литературы

1. Burger S., Sloane Z., Yang J. Competitive Evaluation of Commercially Available Speech Recognizers in Multiple Languages. Pittsburgh: Carnegie Mellon University, 2006. 6 p.
2. Музычук Д.С., Медведев М.С. Сегментация, шумоподавление и фонетический анализ в задаче распознавания речи // Молодой ученый. 2013. №6. С. 86-96.

Барбот Александр Владимирович — главный инженер,
Гусев Сергей Анатольевич — генеральный директор ЗАО «Первая миля»
 Контактный телефон (495) 960-31-59.
 E-mail: info@firstmile.ru
<http://www.firstmile.ru>