

ПРОГРАММНЫЕ СРЕДСТВА СИНТЕЗА РЕЧИ

И.В. САВЧЕНКО

На данный момент существуют различные программные средства синтеза речи. Данные программные средства используют различные подходы для генерации речевого сигнала. Для разработки программного средства генерации речеподобных помех, необходима интеграция с модулем, позволяющим непосредственно генерировать речь на основании текста. Рассмотрим различные существующие программные средства предоставляющие возможность интеграции.

Festival — обобщенная многоязычная система синтеза речи. Она предлагает полную систему синтеза речи с различными API, а также среду для разработки и исследования методов синтеза речи. Система написана на C++ с командным интерпретатором для общей настройки и расширения. Достоинствами данного программного средства являются: поддержка нескольких языков, свободная лицензия, наличие специального API, дополнительные утилиты для работы с системой. К недостаткам можно отнести сложность конфигурации системы под конкретную задачу и ее платформазависимость.

VoiceXML (Voice eXtensible Markup Language) — один из открытых стандартов W3C, который представляет собой интерпретатор, преобразующий текстовую фразу в синтезированную речь. Стандарт предназначен для разработки интерактивных голосовых приложений. VoiceXML имеет теги, являющиеся командами для голосового браузера, который синтезирует, распознает речь, предоставляет диалоговое управление. Достоинствами данного стандарта являются: простота и удобство в использовании, поддержка технологии крупными компаниями, наличие различных реализаций стандарта, использование XML языка разметки, кроссплатформенная реализация. Главными недостатками стандарта являются отсутствие единого главного стандарта и тонкой настройки системы для конкретной задачи.

eSpeak — это компактный свободный программный синтезатор речи, поддерживающий Speech Synthesis Markup Language — SSML. Версии eSpeak существуют под такие операционные системы, как Microsoft Windows, Mac OS X, Linux, RISC OS, Android, Maemo, а также доступен его исходный код на языке C++ Windows Mobile. Слова входного текста для синтеза в данной программе проходят два этапа обработки. На первом этапе слово в буквенном представлении преобразуется в последовательность фонем. На втором этапе на основе полученной последовательности генерируется звуковой сигнал. eSpeak является open source проектом, благодаря этому, некоторые разработчики интегрировали его в свои продукты. К достоинствам данного синтезатора речи относятся: поддержка большого числа языков, открытый исходный код, кроссплатформенность, наличие дополнений и специальных услуг, широкие возможности для тонкой настройки, использование эффективных алгоритмов синтеза речи. К недостаткам можно отнести необходимость перекомпиляции программы для каждой отдельно взятой платформы.

Таким образом наиболее подходящими для интеграции являются программные средства, имеющие открытый исходный код либо предоставляющие специальное API.

Литература

1. Лобанов Б.М., Цирульник Л.И. Компьютерный синтез и клонирование речи. Минск, 2008. 316 с.
2. Сорокин В.Н. Синтез речи. М., 1992. 392 с.
3. Манахов П. Обзор мобильных Text-To-Speech движков / [Электронный ресурс]. Режим доступа: <http://habrahabr.ru/post/102199/>. Дата доступа: 19.05.2012.