

СИНТЕЗАТОР РУССКОЙ РЕЧИ ПО ТЕКСТУ НОВОГО ПОКОЛЕНИЯ

Н. Вольская, А. Коваль, С. Коваль, И. Опарин, Е. Погарева, П. Скрелин, Н. Смирнова, А. Таланов

Центр Речевых Технологий, Санкт-Петербург

andre@speechpro.com

skrelin@phonetics.pu.ru

В статье представлена новая система синтеза русской речи по тексту *Оратор*, разработанная в Центре Речевых Технологий (ЦРТ). Компилятивный аллофонный синтез обеспечивает высокое качество речевого сигнала. Для описания интонации используется 26 типов мелодической кривой. Морфо-грамматический словарь синтезатора охватывает 2.5 млн. словоформ русского языка.

Введение

Задача создания новой системы синтеза русской речи возникла в связи с возникновением новых возможностей, предоставляемых активным развитием компьютерных технологий. Система *Оратор* использует большую аллофонную базу размером около 4500 звуковых единиц, новейшие достижения в области алгоритмов модификации сигнала, а так же ряд концептуально новых разработок, таких как, например, применение специальной интерфейсной грамматики для определения границ синтагм. С технической стороны, *Оратор* существует не только как Windows-приложение, но и в виде SDK, состоящего из динамических библиотек, соответствующих основным модулям системы (препроцессор, словари, просодический модуль, транскриптор), которые могут использоваться в других речевых разработках. Кроме того, существует версия *Оратора* в виде стандартного интерфейса Speech API 5.1.

Структура препроцессора

Главной задачей препроцессора является нормализация входного текста. При этом нормализация подразумевает не только стандартное выделение предложений и избавление от ненужных пробелов и прочих ошибок форматирования, но так же и осуществление ряда специализированных операций:

- Расшифровка аббревиатур. Она осуществляется на основании словаря «известных» системе аббревиатур, либо при помощи правил чтения аббревиатур;
- Расшифровка числового написания числительных и преобразование их в текстовую форму;
- Транслитерация слов, написанных латиницей;
- Контекстно-зависимая расшифровка сокращений (г. – город? господин? гора?);
- Расшифровка специальных знаков (напр., 99% - 99 процентов).

Определение места ударения

Определение места ударения необходимо для правильной работы транскриптора, преобразующего последовательность графем в последовательность аллофонов для синтеза. Место ударения определяется при помощи большого морфо-грамматического словаря, охватывающего 85 тыс. различных основ, на основании которых может быть образовано около 2.5 млн. разных словоформ. При этом в системе реализован высококачественный алгоритм контекстного снятия омографии. В тексте общей тематики количество слов с неправильно определенным местом ударения не превышает 2%.

Сегментный транскриптор

Сегментный транскриптор разработан совместно со специалистами кафедры фонетики Санкт-Петербургского государственного университета [3]. Транскрипция соответствует петербургскому варианту русской произносительной нормы. В транскрипторе предусмотрена возможность согласования транскриптора с темпом речи на уровне выбора соответствующих аллофонов для быстрого и медленного темпов речи, послонного произнесения и спеллинга.

Просодика

Модуль просодики был разработан совместно со специалистами кафедры фонетики СПбГУ. Для осуществления качественного интонационного оформления различных высказываний используется 26 интонационных контуров, определяемых большим количеством выверенных параметров, позволяющих максимально адекватно моделировать поведение мелодической кривой, темпорального и динамического контуров [1]. Каждый интонационный контур включает вариант просодического оформления сверхкоротких (1-2 слова) синтагм. Кроме того, в систему внедрена специально разработанная интерфейсная грамматика, обеспечивающая высокие результаты

автоматической сегментации текста на интонационные единицы [2]. С ее помощью также происходит окончательное снятие неразрешенной на предшествующих этапах омографии.

База звуковых единиц (аллофонов)

База звуковых единиц была создана совместно со специалистами кафедры фонетики СПбГУ. Набор звуковых единиц для синтеза речи представляет собой звуковые реализации аллофонов в их физических границах [3]. Выбор аллофонов в качестве базовых единиц для синтеза определялся требованиями достижения высокого качества синтезированного сигнала, на которое существенное влияние оказывает количество и место склеек базовых единиц при компиляции. Число аллофонов в базе составляет около 4500 единиц, что покрывает практически все теоретически возможные в русском языке комбинаторные и позиционные варианты фонем. Треть из общего числа базовых единиц представляет собой аллофоны, необходимые для моделирования быстрого и медленного темпов произнесения на сегментном уровне, а также для возможного моделирования диалектных и региональных вариантов русского языка при генерации соответствующих инструкций сегментным транскриптором.

В состав базы входят также дескрипторы каждой звуковой единицы, в которых находятся физические адреса границ периодов основной частоты (для гласных, сонантов и звонких согласных), а также служебная информация для изменения физических параметров звуков (частоты основного тона, длительности и интенсивности).

Обработка звукового сигнала

Модификация звуковых единиц по заданным просодическим параметрам производится на основе метода, подобного известному подходу PSOLA. Алгоритм был в значительной степени доработан в ЦРТ для того, чтобы сохранить большую естественность тембра голоса и обеспечить более широкий диапазон изменений частоты основного тона. Кроме того, были разработаны и внедрены в систему оригинальные методы формирования огибающей уровня синтезированного речевого сигнала, что позволило достичь еще более высокого уровня естественности речи.

Заключение

Система синтеза русской речи *Оратор*, разработанная в ЦРТ, отвечает всем современным требованиям, предъявляемым к такой системе как со стороны разборчивости и качества звучания синтезированной речи, так и со стороны ее реализации как программного продукта. При создании *Оратора* использовались не только широко известные современные технологии, но так же и оригинальные разработки ЦРТ в области

обработки сигнала. К определению фонетических параметров были привлечены как специалисты ЦРТ, так и ведущие фонетисты СПбГУ, что также во многом способствовало итоговому высокому качеству разработанного продукта.

Список литературы:

- 1) Вольская Н.Б., Скредин П.А., Таланов А.О. // Интегральное моделирование звуковой формы естественных языков. СПб., 2005. С. 64-74.
- 2) Опарин И.В. Автоматическое интонационное членение для решения прикладных задач // Интегральное моделирование звуковой формы естественных языков. СПб., 2005. С. 54-64.
- 3) P.Skrelin. Allophone-Based Concatenative Speech Synthesis System for Russian // Proceeding of TSD'99, Berlin, 1999.