

## **СОСТАВ ФОНЕТИЧЕСКОЙ БАЗЫ ДАННЫХ В СИСТЕМАХ СИНТЕЗА РЕЧИ**

**А. А. Сняк**

Факт того, насколько быстро развивается вычислительная техника и как стремительно она внедряется в жизнь современного человека, заставляет искать наиболее удобные формы общения между компьютером и непосредственным его пользователем. Бесспорно, самой простой и удобной формой для человека является речевой диалог на естественном языке. Синтез речи – это технология, которая позволяет осуществить такой диалог между компьютером и пользователем. Благодаря этой технологии любая текстовая информация получает возможность быть озвученной.

В изучении синтеза речи особое положение занимает понятие фонетического периода, или фонетической единицы речи. При фонетическом членении выделяют фразы, синтагмы, акцентные единицы, фонетические слова, слоги, фонемы [1]. Фраза – это самая крупная фонетическая единица, законченное по смыслу высказывание, объединенное особой интонацией и отделенное от других фраз паузой [2]. Каждая фраза может состоять из одной или нескольких синтагм. В письменном тексте нет точного эквивалента синтагмы. Тем не менее, определёнными индикаторами границы синтагмы могут являться знаки препинания внутри

предложения. Синтагмы могут состоять из одной или нескольких акцентных единиц. Под акцентной единицей понимаются одно или несколько слов, объединенных общим сильным ударением. Акцентная единица, в свою очередь, состоит из фонетических слов. Фонетическое слово включает знаменательное слово и присоединённые к нему служебные слова – предлоги или частицы. Фонетическое слово подразделено на слоги. Слоги, в свою очередь, на фонемы, фонемы – на аллофоны, или оттенки фонем, которые реализуются в речи в виде акустических колебаний – звуков [3].

Непосредственно синтез устной речи по тексту осуществляется на основе лексико-грамматического анализа входного текста путём моделирования процессов речеобразования с учётом правил произношения звуков и интонирования, свойственных тому или иному языку. Орфографический текст документа поступает на вход синтезатора и далее подвергается последовательной обработке рядом специализированных процессоров в соответствии с общей структурой синтезатора речи по тексту. Синтезатор включает четыре основных модуля: текстовый процессор, просодический процессор, фонетический процессор и акустический процессор. Каждый из этих модулей поддерживается наборами соответствующих баз данных и правил [4].

Текстовый процессор обрабатывает входной орфографический текст в следующей последовательности: для начала происходит очистка самого текста, т.е. преобразование символов (аббревиатур, сокращений, чисел и др.). К примеру, некоторые аббревиатуры читаются не в соответствии со стандартными правилами дешифровки: «США» по правилам расшифровывается как «эс-ше-а», тем не менее общепринятое произношение – «сэ-ше-а». Во-вторых, осуществляется расстановка словесных ударений и грамматических признаков словоформ.

Преобразованный текст поступает на входы просодического процессора. В результате его работы, текст делится на синтагмы, акцентные единицы (АЕ) и их элементы (ЭАЕ). И, наконец, последняя функция просодического процессора – установка в соответствии с БД просодических так называемых «портретов» синтагм – это значения амплитуды (А), длительности фонем (Т) и частоты основного тона (F0) для каждого элемента АЕ.

Задача фонетического процессора заключается в преобразовании орфографического, текста в фонемный, а так же в генерации позиционных и комбинаторных аллофонов.

Акустический процессор на основе информации о том, какие аллофоны необходимо синтезировать, а также какие просодические характеристики должны быть приписаны каждому аллофону, генерирует речевой сигнал путём компиляции отрезков естественных звуковых волн соответствующих аллофонов [5].

Орфографическое письмо по своей природе консервативно. Оно отражает, как правило, произносительную форму языка в период становления письменности. А фонематическая запись должна отражать норму произношения, сложившуюся к настоящему времени. Поэтому при чтении орфографического текста человек наряду со знанием общих правил преобразования «буква – фонема» использует знания о целом ряде сложившихся исключений, таких как замена «ч» на «ш» в словах «что», «чтобы», устранение непроизносимых согласных в словах «солнце», «лестный», замена окончаний прилагательных «-ого», «-его» на «ово», «ево» и др.

Рассматривая, к примеру, русскую орфографию, можно заметить, что в ней в два раза больше гласных букв, чем гласных фонем, и в два раза меньше согласных букв, чем согласных фонем. Для целого ряда согласных фонем в русском языке нет специальных букв, и фонема определяется несколькими буквами или положением определённой буквы в слове. Например, мягкая согласная фонема может записываться соответствующей буквой и буквой «ь», стоящей за согласной. Мягкость согласной при написании может выражаться также стоящими за согласной гласными буквами «и», «е», «ё», «ю», «я» [6].

Процедура буквенно-фонемного преобразования (ПБФ), происходящая в фонетическом процессоре, в целом опирается на регулярные правила внутрисловного преобразования буква-фонема [7], при этом для каждой текущей буквы текста учитывается её левое и правое окружение. Слова, не подчиняющиеся регулярным правилам преобразования, и их фонемная транскрипция содержатся в списке исключений (например, Амстердам – *Амс[тэ]рдам*, мужчина – *му[щ]ина*) [8].

В ходе применения процедуры ПБФ вначале проверяется каждое слово поступившей синтагмы на наличие его в списке исключений и, в случае нахождения, осуществляется его замена на соответствующую цепочку фонем. Если же слово в списке не найдено, к нему применяется регулярные правила ПБФ. Эти правила учитывают категорию букв, основанную на классификации фонем по месту и способу образования [7]. Так, например, при обработке гласных букв первое, что проверяется, не является ли предшествующий знак пробелом, Ъ, Ь или одной из гласных. Если «да», то гласные Я, Ё, Ю, Е преобразуются, соответственно, в фонемы /Ј'А, Ј'О, Ј'У, Ј'Е/, а гласные А, О, У, Э, Ы, И в фонемы /А, О, У, Е/ (пьяный, подъём, дают, ель). Что касается обработки согласных, парных по глухости – звонкости, то, к примеру, если буква Г стоит перед К, то она переходит в фонему /Н/ (легко, лёгкий).

В потоке речи в зависимости от конкретных условий характеристики фонем изменяются, что приводит к появлению оттенков фонем, или аллофонов [9]. Аллофоны подразделяются на позиционные и комбинаторные. Позиционные аллофоны определяются положением фонемы по от-

ношению к ударному слогу в слове. Комбинаторные аллофоны определяются ближайшим контекстом фонемы и обусловлены наложением в потоке речи артикуляции одного звука на другой в результате действия эффектов коартикуляции и ассимиляции [10].

Описание фонем в потоке речи в виде последовательности аллофонов может быть осуществлено с различной степенью подробности. В общем случае невозможно дать точную оценку количества аллофонов, т.к. она напрямую зависит от степени детализации учёта влияния позиционных и комбинаторных факторов. При этом общее количество получаемых аллофонов может колебаться от нескольких сотен до нескольких десятков тысяч [11].

Таким образом, на основе фонетической базы данных осуществляется деятельность фонетического процессора в системах синтеза речи, основными функциями которого являются преобразование орфографического текста в фонемный, а так же генерация позиционных и комбинаторных аллофонов. Переоценить или недооценить важность фонетического процессора невозможно, так как это неотъемлемый элемент в системе, без участия которого работа синтезатора не представляется возможной.

#### Литература

1. Русская разговорная речь / Академия Наук СССР. Институт русского языка; под ред. Е.А.Земской. – М.: Наука, 1973. С. 485.
2. Интернет-адрес: <http://www.hi-edu.ru/e-books/xbook107/01/part-020.htm>
3. Лобанов Б.М. Компьютерный синтез и клонирование речи // Б.М. Лобанов, Л.И. Цирульник / Минск, Белорусская Наука, 2008. С. 24.
4. O.F. Krivnova. Automatic synthesis of Russian speech // Proceedings of the XIV International Congress of Phonetic Sciences, San Francisco, 1999. Vol.1, pp. 507–510.
5. Лобанов Б.М., Система синтеза белорусской речи по тексту / Ю.С. Гецевич, Б.М. Лобанов // Речевые технологии. – 2010. – № 1. – С. 92-101.
6. Панов М.В. Русская фонетика. М., 1967.
7. Лобанов Б.М. Компьютерный синтез и клонирование речи // Б.М. Лобанов, Л.И. Цирульник / Минск, Белорусская Наука, 2008. С. 34-37.
8. Лобанов Б.М. Компьютерный синтез и клонирование речи // Б.М. Лобанов, Л.И. Цирульник / Минск, Белорусская Наука, 2008. С. 296-298.
9. Чистович Л. А. и др. Речь. Артикуляция и восприятие. Л., "Наука", 1965.
10. Матусевич М. И. Современный русский язык. Фонетика. М., 1976.
11. Цирульник Л. И., Лобанов Б.М. Экспериментальная оценка вклада элементов компиляции в правдоподобие синтезированного речевого клона // Компьютерная лингвистика и интеллектуальные технологии: труды междунар. конф. Диалог'2006, Бекасово, 31 мая – 4 июня 2006г. / Институт проблем информатики РАН; отв. ред.: Н.И. Лауфер [и др.]. – М.: Наука, 2006. С. 545–551.