

УДК 621.391

В.В. Киселев

## КОМПЬЮТЕРНАЯ МОДЕЛЬ АЛЛОФОННО-ВОЛНОВОГО СИНТЕЗА РУССКОЙ РЕЧИ

*Описывается система синтеза речи на основе аллофонно-волнового представления речевого сигнала. Приводится общая схема предлагаемой системы. Даны функциональные характеристики составляющих модулей синтезатора речи. Описываются лингвистический, фонетический, просодический и акустический процессоры в составе системы синтеза речи. Одной из особенностей системы является ее модульность и многоуровневая организация, что позволяет реализовать многоязычный и многоголосовой синтез речи по тексту.*

### Введение

Речь является одним из основных способов общения между людьми и относится к высшей интеллектуальной деятельности человека. Поэтому не удивителен тот факт, что в рамках создания систем искусственного интеллекта исследователи пытаются моделировать человеческую речь. Однако до недавнего времени эти системы синтеза не находили своего широкого применения. Одна из причин такого положения – недостаточное качество синтезированной речи, ее роботизированность и неестественность. Стремительный рост мощностей вычислительной техники и научный потенциал лаборатории синтеза и распознавания речи ОИПИ НАН Беларуси [1] создали предпосылки для разработки высококачественной системы синтеза речи нового поколения, в которой реализованы новейшие технологии мирового уровня.

Из существующих моделей синтеза русской речи необходимо отметить разработки фонетической группы филологического факультета МГУ [2], а также исследователей кафедры экспериментальной фонетики СПбГУ [3], на базе которой реализована коммерческая версия Digalo французской компании Elan [4]. Английские версии синтезаторов речи представлены более широко [5, 6]. В качестве одной из самых современных систем можно отметить синтезатор речи компании AT&T [5].

В настоящей работе описывается разработанная под научным руководством Б.М. Лобанова и реализованная автором система синтеза русской речи по тексту. Большое внимание уделено алгоритмам обработки орфографического текста, а также лингвистическому, фонетическому, просодическому и акустическому процессорам, реализующим аллофонно-волновой принцип синтеза речи.

### 1. Общая схема системы синтеза речи

По функциональным признакам систему синтеза речи можно представить в виде совокупности четырех процессоров (рис. 1): лингвистического, фонетического, просодического и акустического [7, 8]. Входной орфографический текст подвергается ряду последовательных обработок в каждом из процессоров. Лингвистический процессор обрабатывает входной орфографический текст в следующей последовательности: очистка текста, преобразование его символов (аббревиатур, сокращений, чисел и др.), затем происходит расстановка словесных ударений и членение текста на синтагмы. Размеченный на синтагмы текст поступает на входы двух процессоров: просодического и фонетического. В результате работы просодического процессора текст делится на акцентные единицы (АЕ), которые далее размечаются на элементы акцентных единиц (ЭАЕ): интонационные предъядро, ядро и заядро. Последняя функция просодического процессора – установка значений амплитуды (А), длительности фонем (Т) и частоты основного тона (F0) для каждого ЭАЕ. Задача фонетического процессора заключается в преобразовании орфографического текста в фонемный, а также в генерации позиционных и комбинаторных аллофонов. Акустический процессор на основе информации о том, какие аллофоны необходимо синтезировать, а также какие просодические характеристики должны быть приписа-

ны каждому аллофону, генерирует речевой сигнал путем компиляции отрезков естественных звуковых волн соответствующих аллофонов.

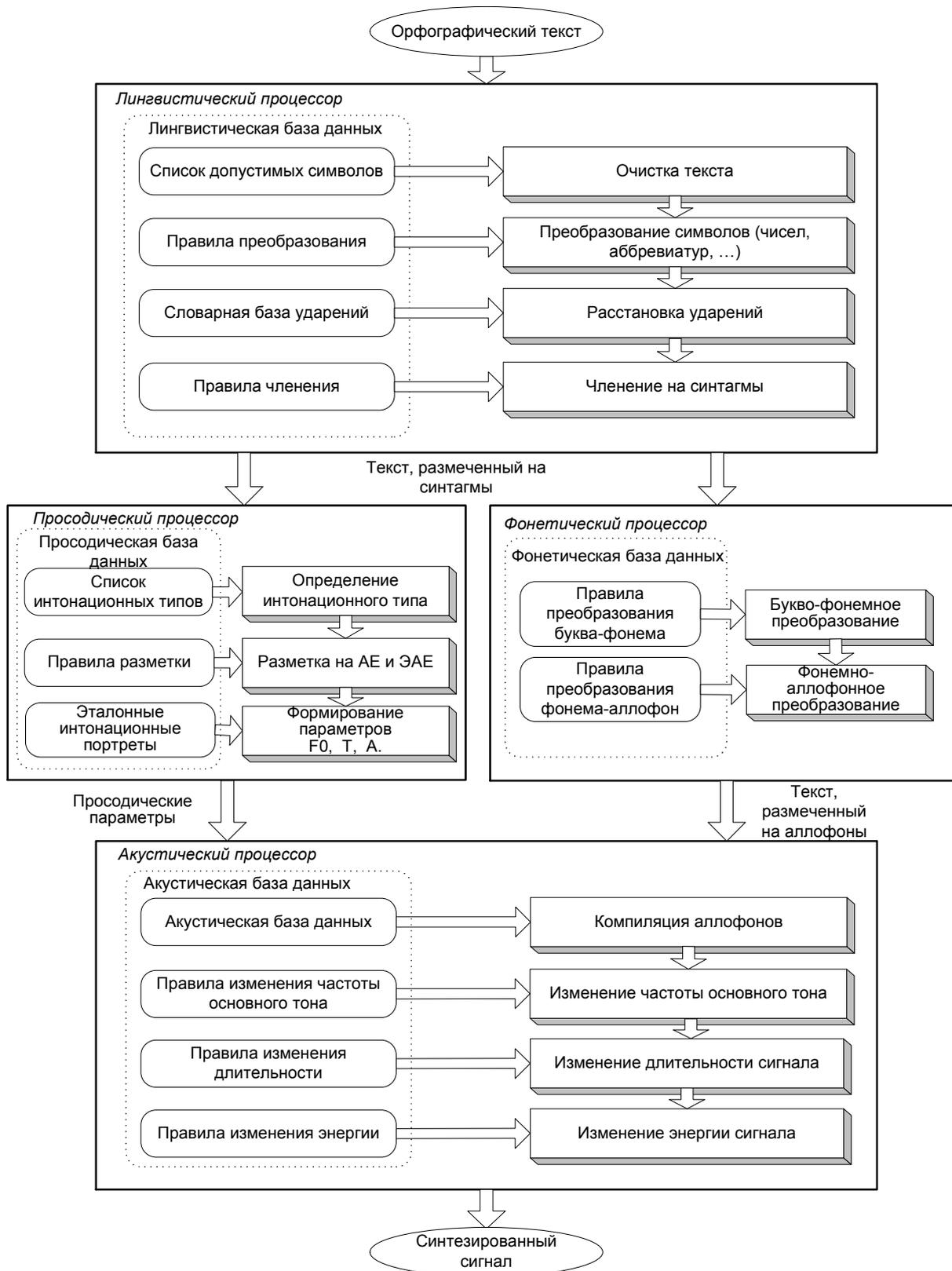


Рис. 1. Общая схема системы синтеза речи

## 2. Лингвистический процессор

Лингвистический процессор, как видно из рис. 1, состоит из лингвистической базы данных и последовательности четырех блоков. Рассмотрим более подробно каждый блок.

### 2.1. Очистка текста

Орфографический текст подается на вход лингвистического процессора в виде множества символов. На первом шаге поступившее множество необходимо очистить от элементов, не принадлежащих списку допустимых символов. Список допустимых символов состоит из следующих конечных множеств (границы множеств определены скобками):

- русских заглавных и прописных букв (А Б В Г Д Е Ё Ж З И Й К Л М Н О П Р С Т У Ф Х Ц Ч Ш Щ Ъ Ы Ь Э Ю Я а б в г д е ё ж з и й к л м н о п р с т у ф х ц ч ш щ ъ ы э ю я);
- латинских заглавных и прописных букв (A B C D E F G H I J K L M N O P Q R S T U V W X Y Z a b c d e f g h I j k l m n o p q r s t u v w x y z);
- математических знаков (– + / \* = < > %);
- знаков пунктуации (. , ; : ( ) ! ? –);
- специальных знаков (~ ` № # @ \$ ^ & | \ { } [ ] " ' ).

### 2.2. Преобразование символов

В орфографическом тексте, особенно в электронном виде из ресурсов сети Интернет, встречается множество специальных символов, при этом система должна произнести их в соответствии с правилами русской орфоэпии. Разработанный блок осуществляет преобразование в текст следующих символов:

- многозначных и дробных чисел;
- римских цифр;
- времени и даты;
- телефонных номеров;
- математических знаков;
- знаков пунктуации;
- аббревиатур;
- сокращений;
- интернет-адресов;
- иностранных слов.

На рис. 2 приведен пример работы программы в блоке преобразования. В нижней части показан входной орфографический текст: СТО, НЭП, НКВД и т. д., в верхней части в столбце Syntagma – результат преобразования.

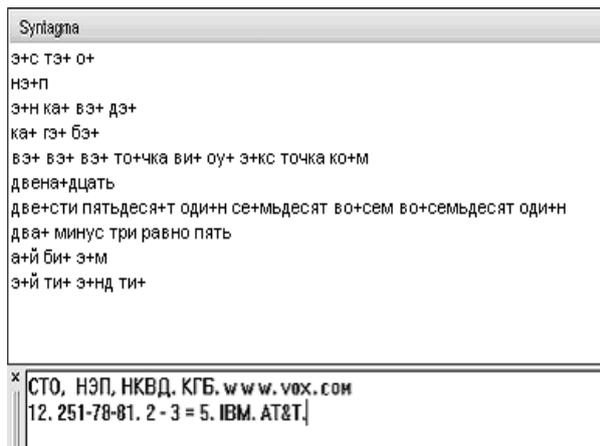


Рис. 2. Пример работы блока преобразования

Преобразование цифровых символов можно разделить на несколько этапов: преобразование количественных и порядковых числительных, десятичных и дробных целых чисел, а также преобразование времени и даты, телефонных номеров.

Опознать порядковое числительное в тексте можно по черточке, стоящей после цифры, и характерному окончанию, например: 1-й класс, 1-го класса, 1-му классу, 1-м классе, 1-е классы, 1-х классах, 1-ми классами; 2-й, 2-го и т. д. Для расшифровки порядковых числительных создан специальный набор базовых числительных и список родовых окончаний. Кроме того, порядковое числительное может быть опознано по маркеру даты и времени. В тексте информационный маркер времени можно определить по признакам:

[число] – массив целых десятичных чисел.

[ключевое слово] – группа *часов*: час, часы, часов; час.; ч.;

– группа *минут*: минут, минута, минуты; м., мин.

Данный алгоритм начинает работать, если текущее слово принадлежит массиву целых чисел, а следующее слово – группе ключевых слов.

В тексте информационный маркер даты можно определить по следующим признакам:

[число] – массив целых десятичных чисел;  
– массив римских чисел в диапазоне I – XXX.

[ключевое слово] – группа слова:

- названия месяцев – январь, февраль, март и т. д. (и различные их модификации);
- год (г.);
- век (в.);
- наша эра, до нашей эры (н. э., д. н. э.);
- столетие (столетия, ст.);
- век, века.

Если число опознано, но оно не принадлежит ни к одному из перечисленных выше классов, система автоматически преобразует его как количественное числительное. Расшифровка количественных числительных происходит с помощью набора базовых числительных: 0, 1, ..., 9, 10, 11, 12, ..., 19, 20, 30, ..., 90, 100, 200, ..., 900, 1000, 1000 000, 1000 000 000.

В пользовательских системах синтеза речи расшифровка знаков пунктуации является излишней, однако для людей со слабым зрением такая возможность оказывается достаточно полезной. При включении данной опции система будет автоматически конвертировать все знаки препинания в произносительный вариант, который также задан в лингвистической базе.

Интернет-ресурсы опознаются по списку маркеров, к которым относятся ключевые слова: www, ftp, url, @, http. Каждому такому маркеру ставится в соответствие вариант русского произношения, например www – *вэ вэ вэ*, ftp – *эф ти ни* и т. д. После информационного маркера следуют, как правило, английские буквы, отражающие положение интернет-ресурса (исключение составляет @, когда маркер стоит в середине). Английские буквы в данном случае читаются отдельно, т. е. каждой английской букве сопоставлено русское произношение, например www.vox.com = *вэ вэ вэ точка ви оу эк точка ком*.

Одной из ключевых проблем при автоматическом синтезе речи является опознание и озвучивание аббревиатур. Существующие в современном русском языке аббревиатуры можно разделить на следующие основные группы: буквенные, звуковые, буквенно-звуковые, аббревиатуры смешанного типа [10].

Структурная схема расшифровки аббревиатур и сокращений показана на рис. 3, а. В базе аббревиатур содержится 1000 аббревиатур и 500 сокращений, найденных в интернет-источниках [11].

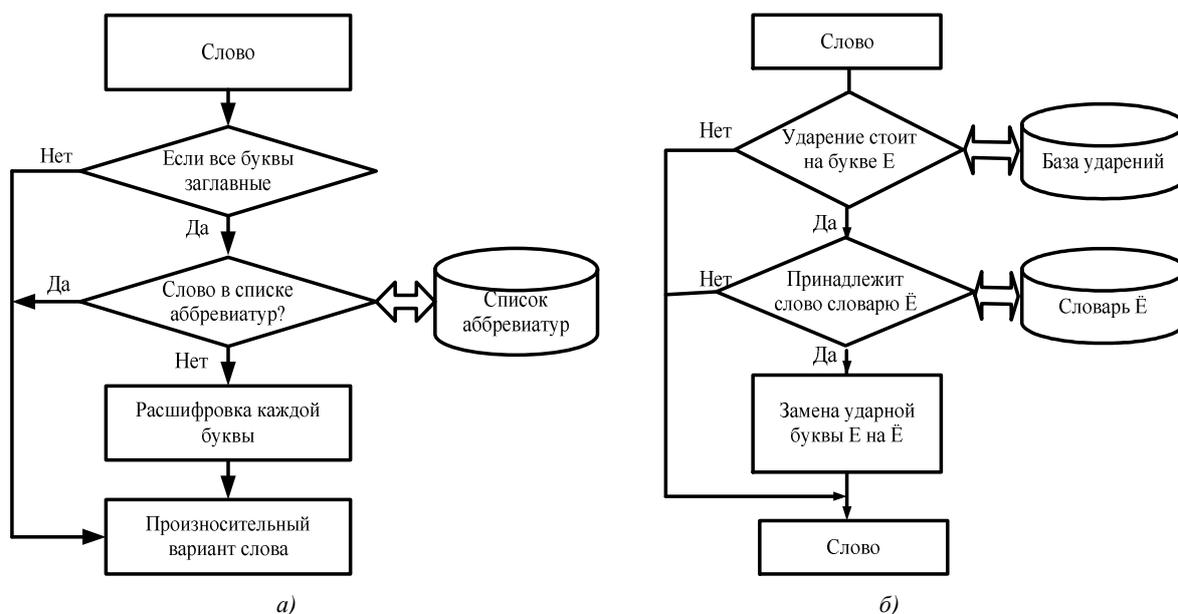


Рис. 3. Структурные схемы: а) расшифровки аббревиатур; б) определения буквы Ё

Все латинские буквы и слова система идентифицирует как иностранные слова, расшифровка которых происходит по следующей схеме. Если иностранное слово присутствует в лин-

гвистической базе, то вместо него ставится русский произносительный вариант. Как правило, в такой список входят часто используемые иностранные слова (например, наименования фирм, информационно-компьютерные термины и др.). В случае если слово не присутствует в базе, оно расшифровывается побуквенно. Каждая латинская буква заменяется на соответствующее ей русское произношение: Q=кю; W=вэ; R=эр; T=ти; P=пи; S=эс; D=ди; F=эф; G=джи; H=эйч; J=джей; K=кей; L=эл; Z=зет; X=икс; C=си; V=ви; B=би; N=эн; M=эм; E=и; U=ю; I=ай; O=оу; A=эй; Y=уй.

### 2.3. Расстановка ударений

При разработке систем синтеза русской речи очень важным является определение места словесного ударения. Русский язык принадлежит к типу языков со свободным ударением. Это означает, что позиция ударения в русском языке подвижна, поэтому алгоритмизировать процесс расстановки ударения практически невозможно. В разработанной системе ударение расставляется посредством поиска необходимого слова в словарной базе ударений, которая насчитывает порядка 450 000 словоформ. Для работы с базой ударений автором статьи разработана и реализована система под названием Ling Master (рис. 4).

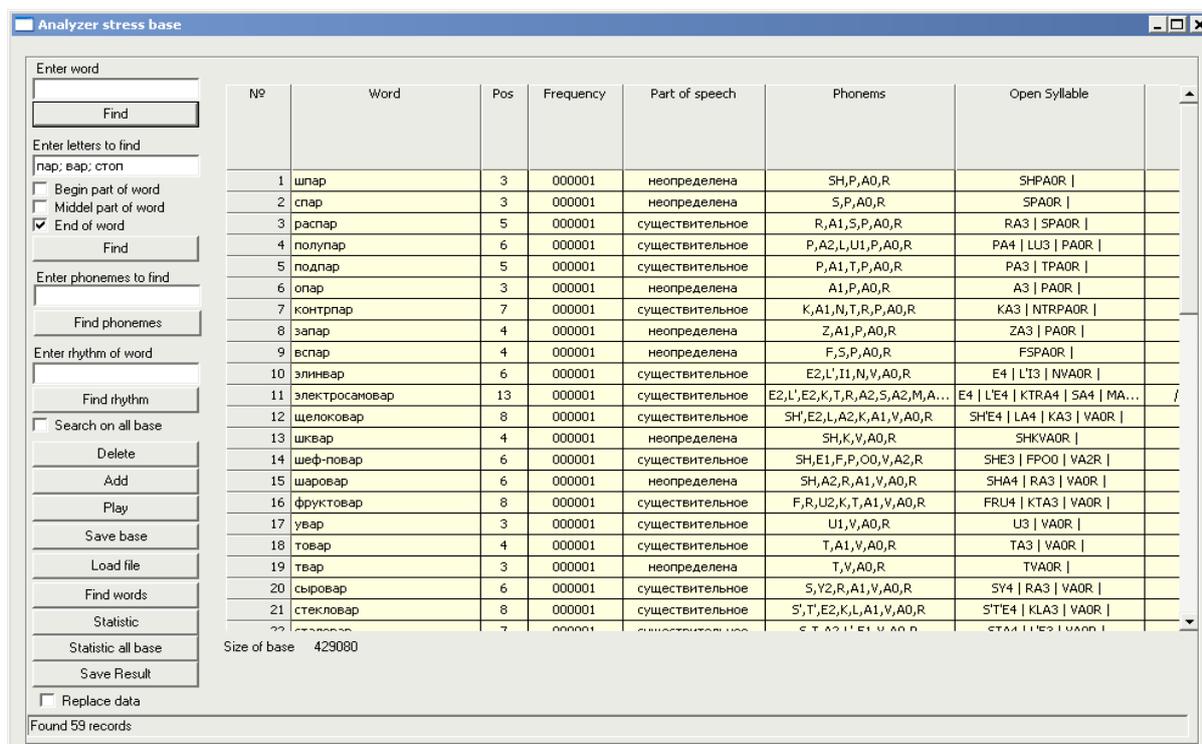


Рис. 4. Внешний вид интерфейса системы Ling Master

В электронном варианте русского текста достаточно часто встречается ситуация, когда вместо буквы Ё стоит буква Е. По правилам русской орфографии и пунктуации буква Ё обязательно пишется только в следующих случаях:

1. Когда необходимо предупредить неверное чтение и понимание слова, например: *узнаём* в отличие от *узнаем*; *всё* в отличие от *все*; *совершённый* (причастие) в отличие от *совершенный* (прилагательное) и т. д.
2. Когда надо указать произношение малоизвестного слова, например: *река Олёкма*.
3. В специальных текстах: букварях, учебниках русского языка для иностранцев, учебниках орфоэпии и т. п., а также в словарях для указания места ударения и правильного произношения.

В остальных случаях написание буквы Ё необязательно. В разработанной системе проблема решается следующим образом. Создан специальный словарь, содержащий слова с бук-

вой Ё̂ (количество слов в словаре составляет 6800). Обращение к специальному словарю происходит в случаях, когда в основном словаре под ударением находится буква Е и это слово содержится в словаре Ё̂. В этом случае замена происходит автоматически, в противном случае буква Е остается без изменения. Схема алгоритма замены буквы Е на Ё̂ изображена на рис. 3, б.

В сложных случаях, когда под ударением возможна буква Е и буква Ё̂, необходимо производить дополнительный грамматический анализ целой группы слов. В реализованной системе такой анализ пока не производится.

#### **2.4. Членение на синтагмы**

Согласно определению, данному в работе [12], под синтагмой понимаются элементы фразы, которые обладают определенной самостоятельностью, т. е. определенной ритмической и интонационной структурой, и допускают некоторую паузу после того, как они были произнесены. Поэтому при автоматическом синтезе речи важным является определение положения границ синтагм. Данный вопрос на сегодняшний день изучен недостаточно. Обзор работ по сегментации текста на простые составляющие и интонационные типы [12–14] позволил сформировать нейтральный алгоритм, заключающийся в следующем.

Вначале образуются орфографические синтагмы [2]. Таковыми считаются последовательности слов, объединенные по правилам:

- от начала текста до первого знака препинания;
- от знака препинания до следующего знака препинания;
- от знака препинания до конца текста.

Затем происходит формирование фонетической синтагмы в зависимости от количества слов в орфографической синтагме. Если количество полноударных слов меньше или равно четырем, орфографическая синтагма остается без изменений. Если количество полных ударений больше четырех, необходимо разбить орфографическую синтагму на фонетические. Первым информативным маркером для членения являются союзы *и*, *или*. Если данные союзы присутствуют в орфографической синтагме, то перед ними ставится раздел синтагмы. После этого определяется количество полноударных слов в каждой из синтагм. Если в какой-либо синтагме их количество больше четырех, необходимо продолжить дальнейшее разбиение на синтагмы. Для этого вначале расставляются маркеры в тех местах, где слова не могут быть разделены. К классу таких слов относятся прилагательные. Тип слова определяется через словарную базу ударений. Если в словарной базе ударений отсутствует информация о типе слова, тогда прилагательное определяется с большой степенью вероятности по окончаниям *-ая*, *-ея*, *-его*, *-ей*, *-ему*, *-ею*, *-ие*, *-ий*, *-ими*, *-их*, *-ою*, *-ого*, *-ое*, *-ой*, *-ому*, *-ою*, *-ую*, *-ый*, *-ые*, *-ым*, *-ыми*, *-ых*, *-юю*, *-яя*. После того как расставлены маркеры неделимых слов, происходит формирование фонетической синтагмы. Конец первой фонетической синтагмы ставится после третьего слова с полным ударением, если за ним не стоит маркер неделимости. Если маркер имеется, то конец фонетической синтагмы ставится после второго слова и т. д. Данный цикл происходит итерационно до тех пор, пока в фонетической синтагме останется не более четырех полных ударений.

### **3. Просодический процессор**

Просодический процессор состоит из просодической базы данных и трех последовательных блоков: определения интонационного типа синтагмы, разметки текста на акцентные единицы и формирования интонационных параметров (амплитуды, длительности, частоты основного тона).

#### **3.1. Определение интонационного типа**

Интонационный тип синтагмы определяется ее синтаксическим статусом [12]. Главными информаторами для определения просодического типа синтагмы являются знаки препинания и символ абзаца. В разработанной системе реализуются 14 различных интонационных типов: один тип восклицания, два типа вопроса, пять типов завершенности, шесть типов незавершенности. Тип восклицания определяется только по знаку «!». Если фонетическая синтагма содержит вопросительное слово, то она принадлежит к первому типу вопроса, иначе – ко второму. Список вопроси-

тельных слов содержит порядка 20 слов. Если фонетическая синтагма находится перед абзацем, ей присваивается первый тип завершенности, если не перед абзацем, то второй. Типы третий, четвертый и пятый присваиваются, если в конце синтагмы стоят знаки пунктуации «; , : , , » соответственно. Типы незавершенности также частично определяются по знакам пунктуации. Незавершенность первого типа имеет знак пунктуации «,» , второго – «-», третьего – «(» . Четвертый интонационный тип незавершенности присваивается синтагме, если она стоит перед союзами *и*, *или*. Если же фонетическая синтагма находится в начале орфографической, то это пятый тип, в середине орфографической – шестой.

### **3.2. Разметка на акцентные единицы и элементы акцентных единиц**

При просодической обработке важным является определение акцентных единиц и их элементов. Акцентной единицей называется слово или группа слов, объединенных одним полным словесным ударением [7]. Если в синтагме присутствуют слова, маркированные частичным ударением, то они объединяются в одну акцентную единицу с полноударным словом, стоящим справа. Если справа такого слова нет (конец синтагмы), то они объединяются в одну акцентную единицу с полноударным словом, стоящим слева. В процессе интонирования речи акцентные единицы рассматриваются как единое слово, ядром которого является полноударная гласная, а в заядро или предъядро включаются также слова с частичным ударением.

Акцентные единицы ранжируются относительно границ синтагмы на начальные, срединные, предконечные и конечные. Эти категории акцентных групп вносят различный по значимости вклад в формирование просодического контура синтагмы. Основное разнообразие мелодических контуров реализовано на конечной акцентной единице, несущей синтагматическое ударение. Мелодические, ритмические и энергетические характеристики акцентных единиц являются минимальными элементами, из которых складывается интонация синтагмы, фразы и текста в целом [8].

### **3.3. Формирование параметров мелодики, ритмики и энергетики**

Формирование просодических параметров происходит путем считывания интонационных портретов из базы данных просодического процессора. Для каждого интонационного типа синтагмы заданы три контура: мелодики (F0), ритмики (T) и энергетики (A). Рассмотрим кратко принципы мелодического, ритмического и энергетического оформления речи для различных интонационных типов.

Интонация синтагмы складывается последовательным соединением указанных классов акцентных единиц, на каждой из которых в соответствии с требуемым интонационным типом формируются необходимые контуры мелодики, ритмики и энергетики. При наличии в синтагме более четырех акцентных единиц добавляется необходимое количество срединных единиц, а при наличии менее четырех единиц исключаются вначале срединная, затем начальная акцентные единицы и предконечная.

Мелодика, ритмика и энергетика акцентной единицы задаются нормированными значениями частоты, длительности и интенсивности на трех ее участках: ядре, предъядре и заядре [9]. Нормированные значения извлекаются из соответствующих просодических контуров акустической базы данных.

Ядром акцентной единицы является ударный аллофон (первый индекс равен 0), предъядром и заядром – соответственно предшествующие ему и следующие за ним аллофоны акцентной единицы.

### **3.4. Фонетический процессор**

Основное назначение фонетического процессора заключается в автоматическом преобразовании текста в фонемный вид, а затем фонемного текста в аллофонный.

Блок букво-фонемного преобразования работает пословно, обрабатывая последовательно акцентные единицы. На вход блока поступает слово, размеченное ударением (знак «+» или «-»). На первом этапе слово сравнивается со списком исключений. Если слово присутствует в списке, то оно заменяется соответствующим исключением и далее подвергается обработке по общему алгоритму. Ударные гласные маркируются индексом 0, предударные – индексом 1, за-

ударные – индексом 2. В служебных словах ударная гласная маркируется индексом 5, а порядок индексирования заударных и преударных, при их наличии, остается прежним.

Главной задачей блока преобразования фонемы – аллофон является присвоение каждой фонеме трех последовательных индексов (например, A043) в зависимости от ее типа, т. е. от позиционных и комбинаторных особенностей расположения в слове. Численное значение каждого индекса строго определено этими условиями. Детальное описание процесса индексирования можно найти в работах [3, 4].

### 3.5. Акустический процессор

На вход акустического процессора поступает текст, размеченный на аллофоны. Например, *S001 A011 SH001 A211 \_ K102 U031 SH001 A211 L204 \_ K103 A031 SH001 U210 #F1*, где #F1 – интонационный тип, который означает завершенность первого типа. Далее процессор сопоставляет строковое описание каждого аллофона с его акустическим вариантом, соединяет и обрабатывает в соответствии с каждым интонационным контуром.

Управление мелодикой (основным тоном) речевого сигнала осуществляется за счет изменения длительности сегментов, соответствующих периоду основного тона у вокализованных аллофонов. Вокализованными считаются аллофоны всех гласных фонем, а также твердых согласных *v, l, m, n* и мягких *v', j', l', m', n'*. Для этих аллофонов предварительно осуществляется специальная разметка, цель которой – выделить границы сегментов, соответствующих периоду основного тона. Для плавного соединения амплитуда сигнала на границах сегментов должна быть близка к нулю. В качестве алгоритма изменения основного тона автором был разработан алгоритм, значительно повышающий качество синтезированной речи. Алгоритм является модификацией известного алгоритма PSOLA [9], однако он более точно учитывает индивидуальные характеристики голоса диктора.

Управление ритмикой речевого сигнала (т. е. длительностью аллофонов) осуществляется путем повторения или удаления необходимого числа отдельных периодов для вокализованных аллофонов, а для невокализованных – путем добавления (удаления) участка звука требуемой длительности. Относительная длительность каждого участка синтагмы считывается из ритмического контура. Управление энергией речевого сигнала производится с помощью умножения каждого синтезированного отсчета сигнала на коэффициент усиления, взятый из энергетического контура.

Одним из важных факторов изменения длительности звуков является регулировка темпа речи. Для ориентировочной оценки пределов изменения длительности фразы и составляющих ее звуков и пауз был проведен следующий эксперимент. Четырехсинтагменная фраза, содержащая примерно 100 звуков, была прочитана в нормальном (среднем) темпе, замедленном (в стиле «диктант») и в максимально быстром. Результаты эксперимента по модификации темпа речи выявили относительные коэффициенты длительности (табл. 1). Таким образом, для системы синтеза были выделены три типа темпа: нормальный, медленный и быстрый.

Таблица 1

Относительные длительности звуков речи

Измеряемые звуковые единицы	Темп, %		
	медленный	нормальный	быстрый
Вся фраза	170	100	70
Пауза	250	100	20
Ударные гласные	200	100	50
Преударные гласные	200	100	70
Заударные гласные	200	100	80
Сонанты	130	100	80
Звонкие взрывные и щелевые	130	100	80
Глухие взрывные	120	100	80
Глухие щелевые	110	100	85

Пример обработки ритмического и энергетического контуров показан на рис. 5. В верхней части рисунка (рис. 5, а) изображен исходный сигнал, предварительно размеченный на периоды основного тона, в нижней части (рис. 5, б) – соответственно результат обработки контуров, вначале ритмического, а затем энергетического.

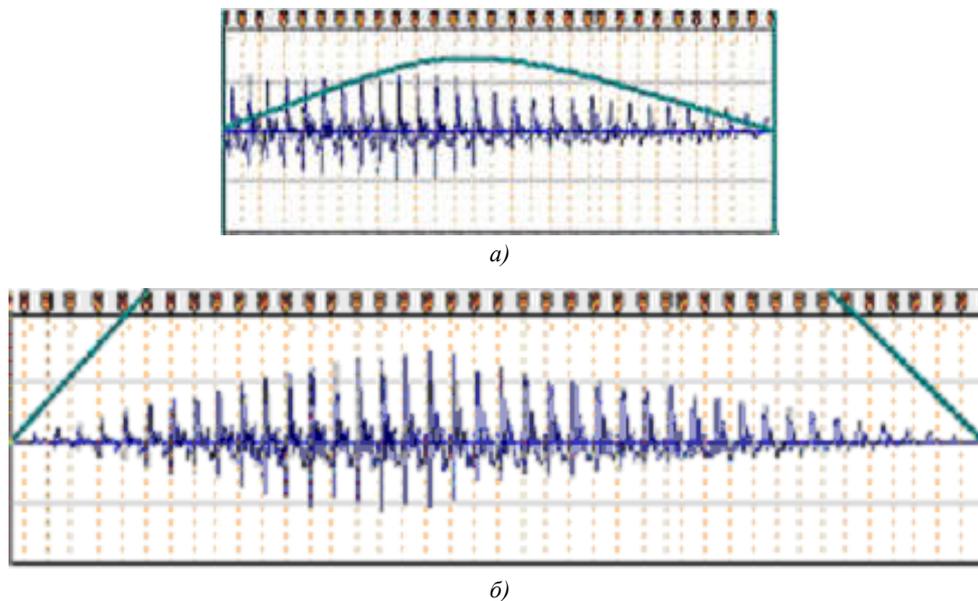


Рис. 5. Пример обработки энергетического и ритмического контуров: а) оригинальный сигнал; б) обработанный

### 3.6. Программная реализация системы синтеза речи

Система синтеза русской речи реализована на языке программирования C++ в среде Microsoft Visual Studio C++ 6.0. Ядро системы состоит из четырех компонентов, которые соответствуют главным процессорам системы синтеза речи (лингвистическому, фонетическому, просодическому и акустическому). Каждый процессор реализован с использованием COM-технологии и представляет собой динамически подключаемую библиотеку функций (DLL). Многокомпонентная реализация ядра обеспечивает достаточно простую переносимость системы синтеза и встраиваемость ее в различные прикладные приложения. Так, например, были реализованы следующие системы: экспертное место Synth Station, предназначенное для тестирования и отладки синтезатора; речевой толковый словарь Ожегова; система «виртуальный собеседник» [18] и ряд других прикладных программ.

### Заключение

Разработанная система синтеза русской речи по тексту основывается на компиляционном методе, в котором в качестве минимального акустического сегмента используется аллофон. Аллофонно-волновой принцип организации синтеза речи позволяет достичь высокого качества и разборчивости синтезированной речи. Ниже приведены результаты измерения звуковой разборчивости синтезированной речи, которые проводились по таблицам звуко сочетаний в соответствии с ГОСТ 16600-72 [15]. Были протестированы три наиболее распространенных синтезатора русской речи: Russian TTS 3000 Lernout & Hauspie Engine [16], Russian Elan TTS Digalo [4], Russian TTS Sakrament [17] – и разработанный синтезатор речи. Разборчивость определялась как процентное отношение количества правильно принятых звуков к количеству переданных звуков. Результаты измерения звуковой разборчивости речи для различных синтезаторов и класс качества в соответствии с ГОСТом приведены в табл. 2.

Из табл. 2 и рис. 6. видно, что разработанный синтезатор речи обеспечивает разборчивость, соответствующую первому классу качества, а процент ошибок восприятия речи существенно ниже, чем для наиболее распространенных синтезаторов.

Кроме того, разработанная модель синтезатора позволяет осуществить многоголосый синтез речи по тексту. Для русского языка в акустической базе хранится база данных для синтеза трех мужских и двух женских голосов. При этом в значительной степени сохраняются индивидуальные особенности, присущие голосу каждого из дикторов, участвовавших в создании акустической базы данных.

Таблица 2

Измерения звуковой разборчивости речи различных синтезаторов

Класс качества	Характеристика класса качества	Норма разборчивости звуков, %	Синтезатор речи	Разборчивость, %
Первый	Понимание речи без малейшего напряжения внимания	свыше 90	Russian SSRLab TTS	95,25
Второй	Понимание речи без затруднений	от 85 до 90	Russian Elan TTS Digalo	88
Третий	Понимание речи с напряжением внимания без переспросов и повторений	от 78 до 85	Russian TTS Sakrament	84,33
Четвертый	Понимание речи с большим напряжением внимания, переспросами и повторениями	от 60 до 78	Russian TTS 3000 Lernout & Hauspie	70,75
Пятый	Полная неразборчивость текста	до 60	—	—

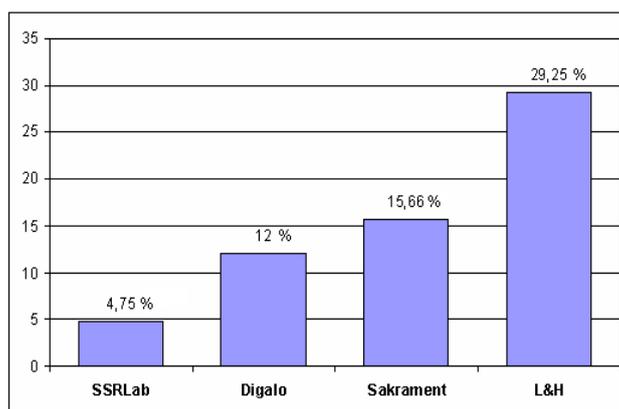


Рис. 6. Процент ошибок восприятия звуков речи

### Список литературы

1. Лобанов Б.М. История создания и развития в Беларуси компьютерных систем распознавания и синтеза речи // Информатика. – 2004. – № 4. – С. 85–95.
2. Опыт разработки системы автоматического синтеза речи для русского языка / О.Ф. Кривнова, Л.М. Захаров, Н.В. Зиновьева и др. // Тр. IX сессии РАО. – М.: АКИН, 1999. – С. 120–126.
3. Звуковая система русского языка в свете задач компилятивного синтеза русской речи / Л.В. Бондарко, В.И. Кузнецов, П.А. Скредин и др. // Бюллетень фонетического фонда русского языка. – № 6. – СПб.: СПбГУ, 1997. – С. 60–85.
4. Рекламный материал компании Elan // <http://www.elan.com>
5. Общее описание синтезатора речи компании Scansoft // <http://www.scansoft.com>
6. Рекламное описание системы синтеза речи // <http://www.naturalvoices.att.com>
7. Синтезатор персонализированной речи по тексту «ЛобаноФон-2000» / В.В. Киселев, Б.М. Лобанов, Т.В. Левковская и др. // Тр. Междунар. конф, посвящённой 100-летию российской экспериментальной фонетики. 1–4 февраля 2001 г., Россия, Санкт-Петербург. – СПб.: СПбГУ, 2001. – С. 101–104.

8. Лобанов Б.М., Карневская Е.Б., Левковская Т.В. Синтезатор речи по тексту как компьютерное средство клонирования персонального голоса // Тр. Междунар. конф. по компьютерной лингвистике и ее приложениям «Диалог-2001». 10–15 июня, Россия, Москва. – М.: Наука, 2001. – С. 265–272.
9. Babkin A.B., Krivnova O.F. F0 contour generation in TTS system for Russian language // Intern. Workshop «Prosody 2000: speech recognition and synthesis». Poland, Krakov. – Krakov: ICPHS, 2000. – С. 190–195.
10. Кривнова О.Ф. Обработка инициальных аббревиатур при автоматическом синтезе речи // Тр. Междунар. семинара по компьютерной лингвистике и ее приложениям «Диалог-99». 4–10 июня, Россия, Москва. – М.: Наука, 1999. – С. 254–260.
11. Электронная база аббревиатур и сокращений // <http://www.sokr.ru/>
12. Брызгунова Е.А. Интонация // Русская грамматика. В 2 т. – М.: Мир, 1980. – Т. 1.
13. Лобанов Б.М. Синтез речи по тексту // Тр. Четвертой Междунар. летней школы-семинара по искусственному интеллекту. 13–17 июля, Россия, Минск. – Мн.: Изд-во БГУ, 2000. – С. 57–76.
14. Кривнова О.Ф. Смысловая значимость просодических швов в тексте // Проблемы фонетики: сб. науч. тр. – М.: Изд-во МГУ, 1999. – С. 170–178.
15. ГОСТ 16600–72. Передача речи по трактам радиотелефонной связи. Требования к разборчивости речи и методы артикуляционных измерений.
16. Синтез русской речи компании L&H // <http://www.coyee.com/ttsengines.htm>
17. Синтез русской речи компании Sakrament // [www.sakrament.com](http://www.sakrament.com)
19. Речевой интерфейс виртуального собеседника / Б.М. Лобанов, А.Г. Давыдов, В.В. Киселев и др. // Тр. Междунар. конф. по компьютерной лингвистике и ее приложениям «Диалог-2005». 10–15 июня, Россия, Москва. – М.: Наука, 2005. – С. 347–356.

Поступила 11.01.05

*Объединенный институт проблем  
информатики НАН Беларуси,  
Минск, Сурганова, 6  
e-mail: vkiselov@mail.ru*

**V.V. Kiselev**

## **COMPUTER MODEL OF ALLOPHONE-WAVE SYNTHESIS OF RUSSIAN SPEECH**

The system of speech synthesis is described based on allophone-wave presentation of a signal. General scheme of the system is given. Functional features of speech synthesizer module components are described. Special attention is devoted to preprocessing stage of text processing. Linguistic, phonetic, prosodic and acoustic blocks of Russian language synthesizer are described. The peculiarity of the system is the modularity and multilayer structure to realize multilingual and multivoiced speech synthesis by text.