

УДК 395.521

Л.И. Цирульник

АЛГОРИТМ ГЕНЕРАЦИИ ФОНЕМНОЙ ПОСЛЕДОВАТЕЛЬНОСТИ ПО ОРФОГРАФИЧЕСКОМУ ТЕКСТУ В СИСТЕМЕ СИНТЕЗА РЕЧИ

Показываются особенности преобразования буква-фонема для русской речи, приводятся логические функции генерации фонем по орфографическому тексту. Описываемый алгоритм, в отличие от существующих, учитывает как внутрисловные, так и межсловные фонетические явления. Программная реализация алгоритма используется в многоязычной системе синтеза речи по тексту на славянских языках.

Введение

Преобразование входного орфографического текста в фонемную последовательность (преобразование буква-фонема) является одной из функций систем синтеза речи по тексту [1–5]. При этом модели и методы, используемые для преобразования буква-фонема, весьма разнообразны. Хороший обзор существующих подходов дан в работе [6], в которой выделяются три типа преобразования: по правилам, управляемое данными и основанное на вероятностных моделях.

Преобразование по правилам [7, 8] предполагает, что контекстно-зависимые правила отображения букв в последовательность фонем должны быть записаны «вручную» экспертом-фонетистом. Особенностью данного подхода является то, что детерминированные правила могут описывать также и фонетические исключения.

Управляемые данными преобразования [9, 10] реализуются, как правило, нейросетевыми моделями, самообучающимися на входных данных. Такой подход в большинстве реализаций принципиально отличается от предыдущего лишь способом получения правил: если в первом случае правила явно определяются экспертом, то здесь они устанавливаются путем автоматической настройки модели. К основным недостаткам данного подхода можно отнести необходимость наличия большого объема обучающих данных и зависимость от архитектуры модели.

Преобразования, основанные на вероятностных моделях [11, 12], также обучаются на входных данных, но, в отличие от предыдущего подхода, используют статистические методы обучения. Недостатком такого подхода является невысокий процент корректного преобразования (например, в работе [12] – 92,28 % для фонем и 61,08 % для слов) при достаточно больших объемах обучающих данных.

Правила преобразования буква-фонема для русской речи к настоящему времени хорошо изучены и описаны, в частности, в работах Л.В. Златоустовой [13], Б.М. Лобанова [14], Н.В. Зиновьевой и О.Ф. Кривновой [15], Л.М. Захарова [16]. Эти описания, однако, учитывают лишь внутрисловные фонетические закономерности изменения звуков русской речи.

Для высококачественного синтеза слитной речи по тексту необходимо учитывать и специфические явления на стыках слов, которые иногда существенно отличаются от подобных внутрисловных явлений. Фундаментальным исследованиям фонетики слитной речи, в том числе стыковых явлений, посвящена монография [17]. Однако описанные в ней правила часто являются противоречивыми и неоднозначными. Проведенные исследования [18] позволили систематизировать некоторые из этих правил и выявить особенности межсловных фонетических явлений.

На основании результатов исследований разработан алгоритм, описываемый в настоящей работе. Отличительной особенностью предлагаемого алгоритма являются уточнение и формализация внутрисловных правил преобразования буква-фонема, изложенных в фундаментальной работе [19], с использованием алгебры логики, а также их расширение с учетом исследованных межсловных фонетических явлений.

В разд. 1 описываются особенности внутрисловного и межсловного преобразования буква-фонема, в разд. 2 – процедура предварительной обработки текста, общая структура блока генерации фонемной последовательности показана в разд. 3, в разд. 4 приводятся логические функции генерации фонем по русскому орфографическому тексту.

1. Особенности внутрисловного и межсловного преобразования буква-фонема

Как показали исследования внутрисловных и межсловных фонетических явлений естественной речи [18], в правилах преобразования буква-фонема необходимо учитывать три типа позиции генерируемой фонемы: внутри слова, на стыке служебного и знаменательного слов, на стыке двух знаменательных слов.

Разделение двух последних ситуаций связано с тем, что служебные слова (предлоги и частицы) произносятся, как правило, без ударения и в слитной речи присоединяются к знаменательным словам. В связи с этим фонемы, находящиеся на стыке служебного и знаменательного слов, подвержены большей ассимиляции, чем фонемы, находящиеся на стыке двух знаменательных слов. Выявленные особенности внутрисловных и межсловных правил преобразования показаны на рисунке.

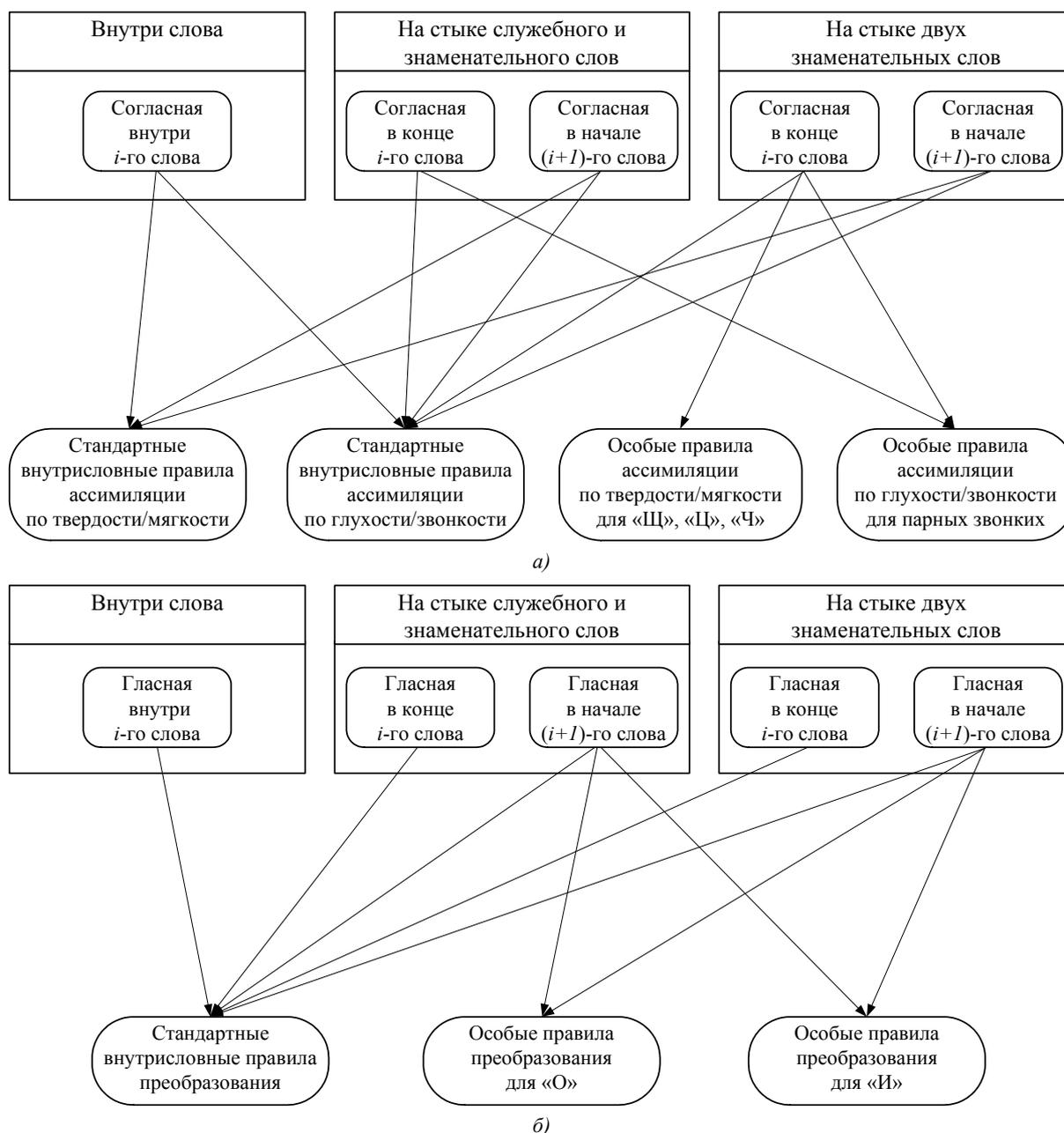


Рис. Особенности внутрисловных и межсловных правил преобразования буква-фонема: а) для согласных; б) для гласных

2. Предварительная обработка входного орфографического текста

Орфографический текст, подаваемый на вход системы синтеза речи, проходит предварительную обработку, которая включает следующие этапы:

- приведение текста к каноническому виду: преобразование число-числительное, расшифровка аббревиатур и сокращений, замена символов на соответствующие текстовые эквиваленты, например «5 %» – «пять процентов», «МВД» – «эмвэдэ»;
- замену слов – фонетических исключений – на эквиваленты, например «ЮНЕСКО» – «ЮНЭСКО», «нарочно» – «нарошно»;
- расстановку полных и частичных словесных ударений, при этом полноударный гласный в тексте помечается символом «+», следующим после него, а частично ударный гласный – символом «=», следующим после него, например «э=та статья+»;
- присоединение служебных слов к знаменательным, при этом граница слов помечается буквой «Ъ», например «э=таЪстатья+ вы+шла вЪна=шем журна+ле»;
- разбиение текста на последовательность синтагм, где под синтагмой понимается самостоятельная в интонационном смысле часть фразы или вся фраза.

Затем каждый символ «пробел» заменяется на символ «_», в начало и конец каждой синтагмы добавляется символ «#», и подготовленный таким образом текст посинтагменно подается на блок генерации фонемной последовательности.

3. Структура блока генерации фонемной последовательности

Входные данные блока – предварительно обработанный текст синтагмы – это последовательность символов L_1, L_2, \dots, L_n , где n – количество символов, при этом каждый символ L_i является элементом множества T , объединяющего множество букв русского языка $B = \{A, Б, В, Г, Д, Е, Ж, З, И, Й, К, Л, М, Н, О, П, Р, С, Т, У, Ф, Х, Ц, Ч, Ш, Щ, Ъ, Ы, Ь, Э, Ю, Я\}$ и множество вспомогательных символов $S = \{_, \#, +, =\}$.

Выходные данные – это последовательность разделенных символом «,» элементов множества P , являющегося объединением множества фонем русской речи: $N = \{u, e, o, a, y, i, ch', sh', c, h, h', j', m, m', n, n', r, r', l, l', b, p, b', p', v, f, v', f', d, t, d', t', g, k, g', k', z, s, z', s', zh, sh\}$ и ранее определенного множества S . Пояснение принятых обозначений фонем приведено в табл. 1, где символом «'» обозначена мягкость фонемы.

Таблица 1

Обозначения фонем русской речи

Обозначение	Значение								
<i>u</i>	[y]	<i>h</i>	[x]	<i>l</i>	[л]	<i>f'</i>	[ф']	<i>z</i>	[з]
<i>e</i>	[э]	<i>h'</i>	[x']	<i>l'</i>	[л']	<i>d</i>	[д]	<i>s</i>	[с]
<i>o</i>	[o]	<i>j'</i>	[й]	<i>b</i>	[б]	<i>t</i>	[т]	<i>z'</i>	[з']
<i>a</i>	[a]	<i>m</i>	[м]	<i>p</i>	[п]	<i>d'</i>	[д']	<i>s'</i>	[с']
<i>y</i>	[ы]	<i>m'</i>	[м']	<i>b'</i>	[б']	<i>t'</i>	[т']	<i>zh</i>	[ж]
<i>i</i>	[и]	<i>n</i>	[н]	<i>p'</i>	[п']	<i>g</i>	[г]	<i>sh</i>	[ш]
<i>ch'</i>	[ч]	<i>n'</i>	[н']	<i>v</i>	[в]	<i>k</i>	[к]		
<i>sh'</i>	[щ]	<i>r</i>	[р]	<i>f</i>	[ф]	<i>g'</i>	[г']		
<i>c</i>	[ц]	<i>r'</i>	[р']	<i>v'</i>	[в']	<i>k'</i>	[к']		

Генерация фонемной последовательности происходит циклически. На каждой итерации цикла, определяемой позицией i , где $1 \leq i \leq n$, на блок генерации поступает символ L_i и окружающие его символы $L_{i-2}, L_{i-1}, L_{i+1}, L_{i+2}$, если они существуют.

Символы подаются на вход системы функций F_{ϕ_j} , где ϕ_j – элементы множества P , причем система включает 42 функции генерации фонем и 4 функции генерации вспомогательных символов, каждая из которых принимает значение «истина» или «ложь» в зависимости от позиции i и значения текущего входного символа. Фактически каждая из функций распознает фонему, которая будет генерироваться, по последовательности входных символов.

Фонемы ϕ_j , для которых на i -й итерации функции F_{ϕ_j} приняли значение «истина», передаются в блок принятия решения, который в случае поступления более одной фонемы определяет порядок их генерации.

В процессе работы блок генерации фонемной последовательности использует подмножества множества T , которые описаны в табл. 2.

Таблица 2

Подмножества символов, используемые для генерации фонемной последовательности

Элементы	Обозначение	Элементы	Обозначение
{ }	M_1	{Б, В, М}	M_{12}
{#}	M_2	{Т, С}	M_{13}
{+, =}	M_3	{Н, Д, З, Л}	M_{14}
{Ъ}	M_4	{Х, Ц, Ч, Щ}	M_{15}
{Ь}	M_5	{Б, Г, Д, Ж, З}	M_{16}
{О, У, А, Ы, Э}	M_6	{К, П, С, Т, Ф, Ш}	M_{17}
{Е, Е, Ю, Я, И}	M_7	{Ц, Ш, Ж}	M_{18}
{Б, В, Г, Д, Ж, З, К, Л, М, Н, П, Р, С, Т, Ф, Х, Ц, Ч, Ш, Щ}	M_8	{С, Ш, Ж}	M_{19}
{К, П, С, Т, Ф, Х, Ц, Ч, Ш, Щ}	M_9	{Б, В}	M_{20}
{Б, В, Г, Д, Ж, З, Л, М, Н, Р}	M_{10}	{Д, З}	M_{21}
{П, Ф}	M_{11}	{К, Х}	M_{22}

4. Логические функции генерации фонем русской речи

4.1. Функции генерации вспомогательных символов

Все вспомогательные символы входного орфографического текста переносятся в формируемую фонемную последовательность, поскольку используются на последующих этапах обработки текста:

$$F_{|_|} = L_i \equiv \langle _ \rangle, \quad F_{| \# |} = L_i \equiv \langle \# \rangle, \quad F_{| + |} = L_i \equiv \langle + \rangle, \quad F_{| = |} = L_i \equiv \langle = \rangle.$$

4.2. Функции генерации гласных фонем

В русской речи существует шесть гласных фонем: /u/, /e/, /o/, /a/, /y/, /i/.

Генерация фонемы /u/ происходит, если текущий символ входного орфографического текста – «У» или «Ю»; генерация фонемы /e/, если текущий символ – «Э» или «Е», в соответствии с формулами

$$F_{/u/} = (L_i \equiv \langle \text{У} \rangle) \vee (L_i \equiv \langle \text{Ю} \rangle); \quad (1)$$

$$F_{/e/} = (L_i \equiv \langle \text{Э} \rangle) \vee (L_i \equiv \langle \text{Е} \rangle). \quad (2)$$

Примеры: У+ТКА – u,+,t,k,a; ЮЛА+ – j',u,l,a,+; Э+ТОТ – e,+,t,a,t; Е+ЛБ – j',e,+,l'.

Фонема /o/ генерируется в случаях, когда текущий символ – буква «Ё» или же текущий символ – буква «О», за которой следует символ ударения, в соответствии с формулой

$$F_{/o/} = (L_i \equiv \langle \text{Ё} \rangle) \vee ((L_i \equiv \langle \text{О} \rangle) \wedge (L_{i+1} \in M_3)). \quad (3)$$

Примеры: ЛЁ+ТЧИК – $l', \underline{o}, +, ch', ch', i, k$; О+СЕНЬ – $\underline{o}, +, s', e, n'$.

Генерация фонемы /a/ происходит всегда, когда текущий символ – буква «А» или «Я», а также в случаях, когда текущий символ – буква «О» при условии, что за ней не следует символ ударения:

$$F_{/a/} = (L_i \equiv \langle \text{А} \rangle) \vee (L_i \equiv \langle \text{Я} \rangle) \vee ((L_i \equiv \langle \text{О} \rangle) \wedge (L_{i+1} \notin M_3)). \quad (4)$$

Пример: АНАНА+С – $\underline{a}, n, \underline{a}, n, \underline{a}, +, s$; Я+БЛОКА – $j', \underline{a}, +, b, l, \underline{a}, k, \underline{a}$; МЯ+ГКИЙ – $m', \underline{a}, +, h, k', i, j$; МОЛОКО – $m, \underline{a}, l, \underline{a}, k, o, +$.

Фонема /y/ генерируется, если выполняется одно из трех условий:

1. Текущий символ – буква «Ы».
2. Текущий символ – буква «И», а предшествующий символ – одна из букв «Ц», «Ш», «Ж».
3. Текущий символ – буква «И», находящаяся в начале слова, а предшествующее слово заканчивается согласной буквой.

Если же текущий символ – буква «И», но не выполняется ни одно из вышеперечисленных условий, генерируется фонема /i/. Эти правила реализованы в следующих функциях генерации:

$$F_{/y/} = (L_i \equiv \langle \text{Ы} \rangle) \vee ((L_i \equiv \langle \text{И} \rangle) \wedge F_1); \quad (5)$$

$$F_{/i/} = (L_i \equiv \langle \text{И} \rangle) \wedge \neg F_1, \quad (6)$$

где $F_1 = (L_{i-1} \in M_{18}) \vee ((L_{i-1} \in M_1 \cup M_5) \wedge (L_{i-2} \in M_8))$.

Примеры: СЫ+ТЫЙ – $s, \underline{y}, +, t, \underline{y}, j'$; ШИ+ЛО – $s, h, \underline{y}, +, l, a$; ПРИНЕ+С_ИГРУШКУ – $p, r', i, n', o, s, \underline{y}, g, r, u, sh, k, u$.

4.3. Функции генерации непарных по глухости/звонкости согласных фонем

Фонемы /sh'/, /ch'/, /c/, /h/, /h'/, /j'/, /m/, /m'/, /n/, /n'/, /l/, /l'/, /r/ и /r'/ не являются парными по глухости/звонкости, т. е. для глухих фонем /sh'/, /ch'/, /c/, /h/, /h'/ не существует соответствующих парных звонких, а для звонких /j'/, /m/, /m'/, /n/, /n'/, /l/, /l'/, /r/, /r'/ – соответствующих парных глухих.

Фонема /sh'/ генерируется в случае, когда текущая буква – «Ш» и в случаях, когда текущая буква – «С», «Ш» или «Ж», а последующая – «Ч». Фонема /ch'/ генерируется только в случае, когда текущая буква – «Ч» при условии, что предыдущая буква – не «С», «Ш» или «Ж». Если же текущая буква – «Ч», а последующая – одна из букв «С», «Ш» или «Ж», ни одна из фонем не генерируется. Описанные особенности учтены в формулах (7), (8):

$$F_{/sh'/} = (L_i \equiv \langle \text{Ш} \rangle) \vee ((L_i \in M_{19}) \wedge (L_{i+1} \equiv \langle \text{Ч} \rangle)); \quad (7)$$

$$F_{/ch'/} = (L_i \equiv \langle \text{Ч} \rangle) \wedge (L_{i-1} \notin M_{19}). \quad (8)$$

Примеры: ШЕКА+ – $sh', e, k, a, +$; СЧА+СТЬЕ – $sh', a, +, s', t', j', e$; ПЕРЕБЕ+ЖЧИК – $p', e, r', e, b', e, +, sh', i, k$; ЧАСЫ+ – $ch', a, s, y, +$.

В функции генерации фонемы /c/ учитывается, что, во-первых, и буква «Т», и буква «С» генерируют фонему /c/ в сочетаниях «ТС», «ТЬС», во-вторых, буква «Т» генерирует фонему /c/ в сочетании «ТЦ» и, в-третьих, буква «Д» генерирует фонему /c/ в сочетаниях «ДС» и «ДЦ», кроме «РДЦ». Функция генерации /c/ описывается формулой

$$\begin{aligned} F_{/c/} = & (L_i \equiv \langle \text{Ц} \rangle) \vee (((L_i \equiv \langle \text{Д} \rangle) \vee (L_i \equiv \langle \text{Т} \rangle)) \wedge (L_{i+1} \equiv \langle \text{С} \rangle)) \vee \\ & \vee ((L_i \equiv \langle \text{Д} \rangle) \wedge (L_{i-1} \neq \langle \text{Р} \rangle) \wedge (L_{i+1} \equiv \langle \text{Ц} \rangle)) \vee ((L_i \equiv \langle \text{Т} \rangle) \wedge (L_{i+1} \equiv \langle \text{Ц} \rangle)) \vee \\ & \vee ((L_i \equiv \langle \text{Т} \rangle) \wedge (L_{i+1} \equiv \langle \text{Б} \rangle) \wedge (L_{i+2} \equiv \langle \text{С} \rangle)) \vee \\ & \vee ((L_i \equiv \langle \text{С} \rangle) \wedge ((L_{i-1} \equiv \langle \text{Т} \rangle) \vee ((L_{i-1} \equiv \langle \text{Б} \rangle) \wedge (L_{i-2} \equiv \langle \text{Т} \rangle)))). \end{aligned} \quad (9)$$

Примеры: ЦА+ПЛЯ – $\underline{c}, a, +, p, l', a$; СПЕЦСВЯ+ЗЬ – $s, p', e, \underline{c}, s, v', a, +, s'$; ПЕРЕВОЛНО-
ВА+ТЬСЯ – $p', e, r', e, v, a, l, n, a, v, a, +, \underline{c}, \underline{c}, a$; РУЧА+ЕТСЯ – $r, u, ch', a, +, j', e, \underline{c}, \underline{c}, a$; БЛЮ+ЩЕ –
 $b, l', u', +, \underline{c}, \underline{c}, e$; ОЦА+ – $a, \underline{c}, \underline{c}, a, +$.

При генерации фонем /h/ и /h'/ учитываются следующие особенности:

– буква «Г» в сочетании «ГК», за которым следует любой символ, кроме букв «Е», «Ё», «Ю», «Я», «И», «Ь», генерирует фонему /h/;

– буква «Г» в сочетании «ГК», за которым следует одна из букв «Е», «Ё», «Ю», «Я», «И», «Ь», генерирует фонему /h'/.

Для того чтобы определить, является ли k-й символ входного текста одной из букв «Е», «Ё», «Ю», «Я», «И», «Ь», используется вспомогательная функция $F_2(k)$:

$$F_2(k) = (L_k \in M_4 \cup M_7),$$

причем в функциях генерации /h/ и /h'/ индекс k принимает значения i+1 и i+2:

$$F_{/h/} = ((L_i \equiv \langle X \rangle) \wedge \neg F_2(i+1)) \vee ((L_i \equiv \langle \Gamma \rangle) \wedge (L_{i+1} \equiv \langle X \rangle) \wedge \neg F_2(i+2)); \quad (10)$$

$$F_{/h'/} = ((L_i \equiv \langle X \rangle) \wedge F_2(i+1)) \vee ((L_i \equiv \langle \Gamma \rangle) \wedge (L_{i+1} \equiv \langle X \rangle) \wedge F_2(i+2)). \quad (11)$$

Примеры: ХИ+ТРЫЙ – $\underline{h}', i, +, t, r, y, j'$; ТРЁ=ХХВО+СТКА – $t, r', o, =, \underline{h}, \underline{h}, v, o, +, s, t, k, a$;
ЛЁ+ГКИЙ – $l', o, +, \underline{h}', k', i, j'$; ЛЕГКО+ – $l, e, \underline{h}, k, o, +$.

Функция генерации фонемы /j'/ более сложная, чем предыдущие: /j'/ генерируется не только в случае, когда текущая буква – «Й», но и когда текущая буква – «О» ударная или «И», находящиеся после буквы «Ь», а также в случаях, когда текущая буква – «Я», «Е», «Ю» или «Ё», находящаяся на границе слова, после мягкого или твердого знаков, после гласной или после знака ударения, в соответствии с формулой

$$F_{/j'/} = (L_i \equiv \langle \text{Й} \rangle) \vee ((L_i \equiv \langle \text{И} \rangle) \wedge (L_{i-1} \in M_4)) \vee ((L_i \equiv \langle \text{О} \rangle) \wedge (L_{i-1} \in M_4) \wedge (L_{i+1} \in M_3)) \vee \quad (12) \\ \vee (((L_i \equiv \langle \text{Я} \rangle) \vee (L_i \equiv \langle \text{Е} \rangle) \vee (L_i \equiv \langle \text{Ю} \rangle) \vee (L_i \equiv \langle \text{Ё} \rangle)) \wedge F_3),$$

где $F_3 = (L_{i-1} \in M_1 \cup M_2 \cup M_3 \cup M_4 \cup M_5 \cup M_6 \cup M_7)$.

Примеры: ЯМА+ЙКА – $\underline{j}', a, m, a, +, \underline{j}', k, a$; ВОРОБЬИ+ – $v, a, r, a, b', \underline{j}', i, +$; БУЛЬО+Н –
 $b, u, l', \underline{j}', o, +, n$; КОНЬЯ+К – $k, a, n', \underline{j}', a, +, k$; ОБЪЁ+М – $a, b, \underline{j}', o, +, m$.

Функции генерации фонем /m/, /m'/, /n/, /n'/, /l/, /l'/, /r/, /r'/ учитывают эффект ассимиляции по твердости/мягкости, заключающийся в смягчении фонемы, находящейся перед мягкой, одинаковой по способу образования. Эффект ассимиляции по твердости/мягкости имеет место внутри орфографического слова, но не на стыках слов (см. рисунок).

Функции генерации фонем /m/ и /m'/ учитывают, что смягчение фонемы /m/ происходит перед губными /p'/, /f'/, /b'/, /m'/, /v'/, что отражено в соответствующих формулах:

$$F_{/m/} = (L_i \equiv \langle M \rangle) \wedge \neg F_2(i+1) \wedge ((L_{i+1} \notin M_{11} \cup M_{12}) \vee \neg F_2(i+2)); \quad (13)$$

$$F_{/m'/} = (L_i \equiv \langle M \rangle) \wedge (F_2(i+1)) \vee ((L_{i+1} \in M_{11} \cup M_{12}) \wedge F_2(i+2)). \quad (14)$$

Примеры: МРА+МОР – $\underline{m}, r, a, +, m, a, r$; ИЮ+НЬСКИМ_ВЕ+ТРОМ – $i, j', u, +, n', s, k', i, \underline{m}, _v', e, +$
 $+, t, r, a, m$; ТРИУМВИРА+Т – $t, r', i, u, \underline{m}', v', i, r, a, +, t$; НАЪТУ+МБЕ – $n, a, t, u, +, \underline{m}', b', e$; МЯ+ЧИК –
 $\underline{m}', a, +, ch', i, k$.

Функции генерации фонем /n/, /n'/ учитывают, что смягчение /n/ происходит перед переднеязычными /t'/, /s'/, /z'/, /d'/, /n'/, /l'/, /ch'/:

$$F_{/n/} = (L_i \equiv \langle N \rangle) \wedge \neg F_2(i+1) \wedge ((L_{i+1} \notin M_{13} \cup M_{14}) \vee \neg F_2(i+2)); \quad (15)$$

$$F_{/n'/} = (L_i \equiv \langle N \rangle) \wedge (F_2(i+1)) \vee ((L_{i+1} \in M_{13} \cup M_{14}) \wedge F_2(i+2)). \quad (16)$$

Примеры: НО+ША – $\underline{n}, o, +, sh, a$; ЗАКО+НЧИТЬ – $z, a, k, o, +, \underline{n}', ch', i, t'$; ВА+СИН – $v, a, +, s', i, \underline{n}, _ch', e, m, a, d, a, +, n$; БА+НТИК – $b, a, +, \underline{n}', t', i, k$.

Функции генерации фонем /l/, /l'/, /r/, /r'/ учитывают следующие особенности:

– смягчение фонем /l/ и /r/ осуществляется, соответственно, перед фонемами /l'/ и /r'/;

– буква «Л» в сочетании «ЛНЦ» не генерирует ни одну фонему –

и выражаются формулами

$$F_{ll} = (L_i \equiv \langle \text{Л} \rangle) \wedge ((L_{i+1} \neq \langle \text{Н} \rangle) \vee (L_{i+2} \neq \langle \text{Ц} \rangle)) \wedge \neg F_2(i+1) \wedge ((L_{i+1} \neq \langle \text{Л} \rangle) \vee \neg F_2(i+2)); \quad (17)$$

$$F_{ll'} = (L_i \equiv \langle \text{Л} \rangle) \wedge (F_2(i+1) \vee ((L_{i+1} \equiv \langle \text{Л} \rangle) \wedge F_2(i+2))); \quad (18)$$

$$F_{rr} = (L_i \equiv \langle \text{Р} \rangle) \wedge \neg F_2(i+1) \wedge ((L_{i+1} \neq \langle \text{Р} \rangle) \vee \neg F_2(i+2)); \quad (19)$$

$$F_{rr'} = (L_i \equiv \langle \text{Р} \rangle) \wedge (F_2(i+1) \vee ((L_{i+1} \equiv \langle \text{Р} \rangle) \wedge F_2(i+2))). \quad (20)$$

Примеры: СОЛНЕЧНЫЙ – $s, o, +, \underline{l}, n', e, ch', n, y, j'$; СО+ЛНЦЕ – $s, o, +, \underline{n}, c, e$; Э+ЛЛИПС – $e, +, \underline{l}', l', i, p, s$; ПО=ЛБЛИСТА+ – $p, o, =, \underline{l}, l', i, s, t, a, +$; КРОМЕ+ШНАЯ – $\underline{k}, r, a, m', e, sh, n, a, j', a$; КОР – k, o, r ; РИ+ДА – $k, a, r', r', i, +, d, a$.

4.4. Функции генерации парных по глухости/звонкости согласных фонем

По критерию глухости/звонкости 22 согласные фонемы русской речи объединяются в 11 пар: /b/ - /p/, /b'/ - /p'/, /v/ - /f/, /v'/ - /f'/, /d/ - /t/, /d'/ - /t'/, /g/ - /k/, /g'/ - /k'/, /z/ - /s/, /z'/ - /s'/, /zh/ - /sh/.

При формировании функций генерации парных по глухости/звонкости учитывается описанный выше эффект ассимиляции по твердости/мягкости, который в данном случае выражается следующими правилами:

– фонемы /b/, /p/, /v/ и /f/ смягчаются перед фонемами /b'/, /v'/, /m'/, /p'/, /f'/;

– фонемы /d/, /t/, /z/ и /s/ смягчаются перед фонемами /n'/, /d'/, /z'/, /l'/, /t'/, /s'/;

– фонемы /g/ и /k/ смягчаются перед фонемами /g'/, /k'/, /h'/;

– фонемы /zh/ и /sh/, не имеющие парных мягких, не смягчаются.

Кроме того, в соответствующих функциях генерации отражен эффект ассимиляции фонем по глухости/звонкости, который проявляется в оглушении парных звонких фонем и озвончении парных глухих фонем в некоторых позициях. Эффект ассимиляции по глухости/звонкости наблюдается не только внутри фонетического слова, но и на стыках слов (см. рисунок). Проведенные исследования [18] позволили сформулировать следующие внутрисловные и межсловные критерии оглушения/озвончения фонем:

– в позициях внутри фонетического слова все парные звонкие переходят в глухие перед глухой фонемой, парные глухие /p/, /p'/, /t/, /t'/, /k/, /k'/, /s/, /s'/, /sh/ переходят в звонкие перед любой парной звонкой, кроме /v/, /v'/, парные глухие /f/, /f'/ переходят в звонкие перед любой парной звонкой;

– в позициях на стыке слов все парные звонкие переходят в глухие лишь в случае, когда следующее слово начинается с парной глухой, а парные глухие (как и во внутрисловных позициях) переходят в звонкие, когда следующее слово начинается с парной звонкой фонемы.

Для отражения внутрисловного эффекта ассимиляции по глухости/звонкости используются вспомогательные функции $F_4(k)$ и $F_5(k)$, при этом функция F_4 принимает значение «истина», если фонема, генерируемая по k -му символу входного текста, глухая, а функция F_5 принимает значение «истина», если фонема, генерируемая по k -му символу входного текста, парная звонкая.

Для отражения межсловного эффекта ассимиляции по глухости/звонкости используются вспомогательные функции $F_6(k)$ и $F_7(k)$, причем $F_6(k)$ принимает значение «истина», если k -й символ входного текста является разделителем между словами, $F_7(k)$ принимает значение «истина», если k -я фонема является парной глухой.

На основе функций $F_4 - F_7$ определены вспомогательные функции $F_8 - F_{11}$, проверяющие условия, общие для генерации различных парных звонких фонем и учитывающие как внутрисловные, так и межсловные правила ассимиляции по глухости/звонкости:

$$F_8(k) = \neg F_4(k) \vee (F_6(k) \wedge F_5(k+1));$$

$$F_9(k) = F_6(k) \wedge \neg F_5(k+1);$$

$$F_{10}(k) = (L_k \in M_4) \wedge (\neg F_4(k+1) \vee (F_6(k+1) \wedge F_5(k+2)));$$

$$F_{11}(k) = (L_k \in M_4) \wedge (F_4(k+1) \vee (F_6(k+1) \wedge \neg F_5(k+2))).$$

С учетом описанных эффектов функции генерации фонем /b/, /p/, /b'/ и /p'/ (формулы (21)–(24)) определяются следующим образом:

$$F_{/b/} = ((L_i \equiv \langle \text{Б} \rangle) \wedge F_8(i+1)) \vee ((L_i \equiv \langle \text{П} \rangle) \wedge F_5(i+1) \wedge ((L_{i+1} \neq \langle \text{Б} \rangle) \vee \neg F_2(i+2))); \quad (21)$$

$$F_{/p/} = ((L_i \equiv \langle \text{Б} \rangle) \wedge ((L_{i+1} \in M_2) \vee (F_4(i+1) \wedge ((L_{i+1} \notin M_{11}) \vee \neg F_2(i+2)) \vee F_9(i+1)))) \vee ((L_i \equiv \langle \text{П} \rangle) \wedge \neg F_2(i+1) \wedge ((L_{i+1} \notin M_{11} \cup \{\langle \text{М} \rangle\}) \vee \neg F_2(i+2)) \wedge \neg F_5(i+1)); \quad (22)$$

$$F_{/b'/} = ((L_i \equiv \langle \text{Б} \rangle) \wedge ((L_{i+1} \in M_7) \vee ((L_{i+1} \in M_{12}) \wedge F_2(i+2)) \vee F_{10}(i+1))) \vee ((L_i \equiv \langle \text{П} \rangle) \wedge (L_{i+1} \equiv \langle \text{Б} \rangle) \wedge F_2(i+2)); \quad (23)$$

$$F_{/p'/} = ((L_i \equiv \langle \text{Б} \rangle) \wedge (((L_{i+1} \in M_{11}) \wedge F_2(i+2)) \vee F_{11}(i+1))) \vee ((L_i \equiv \langle \text{П} \rangle) \wedge (F_2(i+2) \vee ((L_{i+1} \in M_{11} \cup \{\langle \text{М} \rangle\}) \wedge F_2(i+2)))); \quad (24)$$

Примеры: СТОЛБНЯ+К – s,t,a,l,b,n',a,+,k; ДУ+Б ЗЕЛЕ+НЬЙ – d,u,+,b,_,z',e,l',o,+,n,y,j'; КРЕПДЬШИ+Н – k,r',e,b,d,y,sh,y,+,n; ОБМЯ+К – a,b',m',a,+,k; СЕ+РП ЗАБЛЕСТЕ+Л – s',e,+,r,b,_,z,a,b,l',e,s',t',e,+,l; НАБПНЕ+ – n,a,p',n',e,+.

Генерация фонем /v/, /f/, /v'/ и /f'/ происходит по описанным выше правилам, учитывающим внутрисловную ассимиляцию по твердости/мягкости, а также внутрисловную и межсловную ассимиляцию по глухости/звонкости. Соответствующие формулы отражают тот факт, что буква «Г» в сочетаниях «ОГО», «ЕГО» в конце слова генерирует фонему /v/.

Примеры: ГРА+Ф ВИ+КТОР – g,r,a,v,_,v',i,+,k,t,a,r; ЗА+ВТРА – z,a,+,f,t,r,a; СА+МОГО – s,a,+,m,o,v,o; ВВЕ+З – v',v',o,+,s; ГОТО+ВЬСЯ – g,a,t,o,+,f',s',a; АФГАНИСТА+Н – a,v,g,a,n',i,s,t,a,+,n.

Функции генерации фонем /d/, /t/, /d'/ и /t'/ учитывают следующие особенности:

– буква Т в сочетаниях «СТН», «СТЛ», «НТГ», «ТЦ», «ТС», «ТЬС» не генерирует ни одну из фонем /d/, /t/, /d'/, /t'/;

– буква Д в сочетаниях «ЗДН», «НДШ», «НДТ», «ГДТ», «РДЧ», «ДЦ», «ДС» не генерирует ни одну из фонем /d/, /t/, /d'/, /t'/.

Примеры: ПОДУ+Л – p,a,+,d,u,l; ЧА+СТНЬЙ – ch',a,+,s,n,y,j'; МЛА+ДШЕ – m,l,a,+,t,sh,e; ДНЯ+МИ – d',n',a,+,m',i; ОТДИРА+ТЬ – a,d',d',i,r,a,t'; ПОДТИ+П – p,a,t',t',i,+,p; ЯГДТА+Ш – j',a,g,t,a,+,sh.

Функции генерации фонем /g/, /k/, /g'/ и /k'/ отражают следующие особенности:

– буква Г в сочетании «ГК» не генерирует ни фонему /k/, ни фонему /k'/;

– буква Г в сочетаниях «ОГО», «ЕГО» в конце слова не генерирует фонему /g/.

Примеры: ВОКЗА+Л – v,a,g,z,a,+,l; СТОКГО+ЛЬМ – s,t,a,g,g,o,+,l',m; БЕ+ГСТВО – b',e,+,k,s,t,v,a; ТРИ+ГПЕР – t,r',i,+,g',g',e,r; АКСИО+МА – a,k,s',i,o,+,m,a; ХИ+МИКГЕ+НА – h',i,+,m',i,g',_,g',e,+,n,a.

Функции генерации фонем /z/, /s/, /z'/ и /s'/ учитывают следующие особенности:

– буква З в сочетании «ЗЖ» не генерирует фонему /z/;

– буква С в сочетаниях «СЧ», «СШ», а также в сочетаниях «ТС», «ТЬС» не генерирует ни одну из фонем /z'/, /s'/, /z/, /s/.

Примеры: АГРЕССИ+ВНОСТЬ – a,g,r',e,s',s',i,+,v,n,a,s',t'; СБА+ВИТЬ – z,b,a,+,v',i,t'; БЕЗПОЩА+ДНО – b',e,s,p,a,sh',a,+,d,n,a; СДЕ+ЛАТЬ – z',d',e,+,l,a,t'; КОСБА+ – k,a,+,z',b,a; ВЕЗТИ+ – v',e,s',t',i,+.

Функции генерации фонем /zh/, /sh/ учитывают следующие особенности:

– буква З в сочетании «ЗЖ» генерирует фонему /zh/;

– буквы Ж и Ш в сочетаниях «ЖЧ», «ШЧ» соответственно не генерируют ни фонему /zh/, ни фонему /sh/;

– буква С в сочетании «СШ» генерирует фонему /sh/.

Примеры: ВЪЕЗЖА+ТЬ – $v, j', e, zh, zh, a, +, t'$; БРО+ШЬ ЖЕ+НИ – $b, r, o, +, zh, zh, e, n', i$; МОЛОДЕ+ЖЬ – $m, a, l, a, d', o, +, sh$; РАСШИ+Б – $r, a, sh, sh, y, +, p$.

4.5. Принятие окончательных фонемных решений

Функции генерации фонем описаны таким образом, что возможны два варианта результатов, поступающих на блок принятия решения:

– только одна из 42 функций F_{ϕ_j} принимает значение «истина», при этом генерируется фонема ϕ_j ;

– две из 42 функций принимают значение «истина», причем одна из них – $F_{j'}$, другая – одна из функций $F_{/el}, F_{/ol}, F_{/ul}, F_{/al}, F_{/il}$. В этом случае генерируется последовательность двух фонем, первая из которых – $/j'$.

Заключение

Алгоритм преобразования буква-фонема, представленный в работе, реализован в виде динамически подключаемой библиотеки для 32-битных операционных систем семейства Windows и используется в составе многоязыковой системы синтеза речи по тексту на славянских языках [20], созданной в рамках проекта «Разработка многоголосовой и многоязыковой системы синтеза и распознавания речи (языки: белорусский, польский, русский)» в соответствии с грантом INTAS № 04-77-7404.

Список литературы

1. Лобанов, Б.М. Синтез речи по тексту / Б.М. Лобанов // Четвертая Международная летняя школа-семинар по искусственному интеллекту: сб. науч. тр. – Минск: Изд-во БГУ, 2000. – С. 57–76.
2. Bell Laboratories Russian text-to-speech System / E. Pavlova [et al.] // Proc. of the Int. conf. EUROSPEECH. – Rhodes-Greece, 1997. – P. 2451–2454.
3. Arora, S. Using Syllable as a Major Unit for Developing an Efficient Concatenative Hindi Speech Synthesiser / S. Arora, K. Arora, S.S. Agrawal // Proc. of the Int. conf. SPECOM. – Greece, 2005. – P. 675–679.
4. Development of a Kiswahili Text to Speech System / M. Gakuru [et al.] // Proc. of the Int. conf. INTERSPEECH. – Portugal, 2005. – P. 1482–1485.
5. Tucker, R. Supporting the Creation of TTS for Local Language Voice Information Systems / R. Tucker, K. Shalovov // Proc. of the Int. conf. INTERSPEECH. – Portugal, 2005. – P. 453–457.
6. A Comparison of Letter-to-Sound Conversion Techniques for English text-to-speech Synthesis / R. I. Damper [et al.] // Proc. of the Institute of Acoustics. – № 20(6). – 1999. – P. 21–30.
7. Allen, J. From Text to Speech: The MITalk System / J. Allen, S. Hunnicutt, D. Klatt. – Cambridge: Cambridge University Press, 1987.
8. Comparison of Two Tree-Structured Approaches for Grapheme-to-Phoneme Conversion // R. Andersen [et al.] // Proc. of the Int. conf. ICSLP. – USA, 1996. – P. 1808–1811.
9. Sejnowski, T.J. Parallel Networks that Learn to Pronounce English Text / T.J. Sejnowski, C.R. Rosenberg // Complex systems. – Vol. 1. – 1987. – P. 145–168.
10. Yvon, F. Grapheme-to-Phoneme Conversion Using Multiple Unbounded Overlapping Chunks / F. Yvon // Proc. of the Int. conf. NeMLaP. – Turkey, 1996. – P. 218–228.
11. Dermatas, E. A Language-Independent Probabilistic Model for Automatic Conversion Between Graphemic and Phonemic Transcription of Words / E. Dermatas, G. Kokkinakis // Proc. of the Int. conf. EUROSPEECH. – Hungary, 1999. – P. 2071–2074.
12. Taylor, P. Hidden Markov Models for Grapheme to Phoneme Conversion / P. Taylor // Proc. of the Int. conf. INTERSPEECH. – Portugal, 2005. – P. 1973–1976.
13. Алгоритм преобразования русских орфографических текстов в фонемную запись // Л.В. Златоустова [и др.]. – М.: МГУ, 1970.

14. Лобанов, Б.М. Преобразователь графема-фонема для синтеза речи по орфографическому тексту / Б.М. Лобанов, Б.В. Панченко // Тр. Междунар. конф. «Автоматическое распознавание слуховых образов АРСО-8». Ч. 4. – Львов, 1974. – С. 15–18.
15. Зиновьева, Н.В. Прикладные системы с использованием фонетических знаний / Н.В. Зиновьева, О.Ф. Кривнова // Проблемы фонетики: сб. статей. – М.: Прометей, 1993. – С. 288–300.
16. Захаров, Л.М. Акустическая вариативность звуковых единиц в русской речи / Л.М. Захаров // Язык и речь: проблемы и решения: сб. науч. тр. – М.: Макс-Пресс, 2004. – С. 240–269.
17. Русская разговорная речь / Под ред. Е.А. Земской. – М.: Наука, 1973. – 485 с.
18. Лобанов, Б.М. Внутрисловные и межсловные правила обработки текста для полного и разговорного стилей речи / Б.М. Лобанов, Л.И. Цирульник // Функциональные стили звучащей речи: сб. науч. тр. – М.: Макс-Пресс, 2006. – С. 21–30.
19. Русская грамматика. Т. I. Фонетика, фонология, ударение, интонация, словообразование, морфология. – М.: Наука, 1982. – С. 14–95.
20. Lobanov, B. Development of multi-voice and multi-language TTS synthesizer (languages: Belarussian, Polish, Russian) / B. Lobanov, L. Tsirulnik // Proc. of the Int. conf. SPECOM. – Russia, 2006. – P. 274–283.

Поступила 23.06.06

*Объединенный институт проблем
информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: Liliya_Tsirulnik@ssrslab.com*

L.I. Tsirulnik

PHONEMIC SEQUENCE GENERATION ALGORITHM BY ORTHOGRAPHIC TEXT IN TTS-SYNTHESIS SYSTEM

The paper deals with the description of peculiarities of letter-to-phoneme conversion for Russian language. The logical functions of phonemes generation by orthographic text are presented. The algorithm described, in contrast to existing ones, embraces phonetic phenomena that take place not only within words, but also at word boundaries. The software of letter-to-phoneme conversion is implemented in multilingual TTS-synthesis system for Slavonic languages.