

Синтезатор татарской речи: фонетический эллипсис и изменения речевых звуков на границе слов и слогов¹

Т.И. Ибрагимов

Казанский государственный университет, филологический факультет
Tavzich.Ibragimov@ksu.ru

Ф.И. Салимов

Казанский государственный университет, факультет ВМК
Farid.Salimov@ksu.ru

Д.Ш. Сулейманов

Казанский государственный университет, факультет ВМК
dvdt@telecet.ru

Р.Р. Хусаинов

Казанский государственный университет, факультет ВМК
rust@ksu.ru

Ключевые слова: синтез речи, фонетическая транскрипция, ритмические группы, речевой звук, фонетический эллипсис

Рассматривается проблема повышения качества синтезированной татарской речи. Приводятся данные о позиционных и контекстных изменениях фонем, в том числе данные о частичной редукции и полном выпадении кратких гласных в потоке речи.

В отличие от синтезаторов, разрабатываемых для различных европейских языков, алгоритм преобразования графем в речевые звуки синтезатора татарской речи состоит из двух частей. На первом этапе текст преобразуется в фонематическую запись, на втором - полученный результат переводится в последовательность речевых звуков.

Двуступенчатость алгоритма перевода графем в речевые звуки обусловлена несовершенством орфографии татарского языка, а именно, существованием правила - писать заимствования так, как в оригинале. В результате действия данного правила доля варваризмов (неадаптированных заимствований [1]) в лексике современного татарского языка составляет 25 - 30% , а в словарях эта цифра достигает до 46%. Наличие такого количества не ассимилированных в языке заимствований нарушает систему языка, сильно затрудняет его алгоритмическое описание и компьютерную обработку. Правила, лежащие в основе алгоритма фонематической записи приведены в [6].

В статье рассматривается вторая часть алгоритма преобразования графем в последовательность речевых звуков - фонетическая транскрипция фонематически транскрибированного текста. Преобразование проводится после семантико-синтаксического анализа текста, т.е. после объединения слов входного предложения в ритмические группы (РГ) или, иными словами, выделения в нем границ речевых тактов.

Таким образом, в работе рассматривается не просто фонетическая транскрипция, а моделирование ритмической группы. Перевод графем в звуки осуществляется с учетом фонетических явлений, вызванных слитным произнесением входящих в РГ слов.

¹Работа выполнена при финансовой поддержке гранта РФФИ №03-07-90285

По определению [2, 3] РГ представляет собой фонетическое единство. Данное свойство РГ проявляется как на сегментном, так и суперсегментном уровнях ее организации. В силу присутствия интонационной компоненты ритмические группы, находящиеся в начальной, серединной и конечной позициях предложения, отличаются друг от друга. В целях упрощения задачи, в данной статье интонационная составляющая просодии РГ была исключена из рассмотрения, что позволило сосредоточиться на моделировании ее ритмической организации.

РГ в татарском языке может состоять из одного слова или из сочетаний от двух до пяти слов. Слова, входящие в РГ, произносятся слитно, как одно слово. Последнее имеет большое значение для синтеза качественной речи, в силу того, что в РГ длительности начального и конечного слогов, если даже они по составу сегментных единиц полностью совпадают, не одинаковы. Так, в РГ $//[без_1]не(нг)^2 килене[без_2]//$, произнесенной в составе предложения «Узган атнада // безне(нг) киленебез // Арчага барып кайтты» («На прошлой неделе наша сноха съездила в Арск»), длительность слога $[без_1]$ равна 155 мсек, а длительность слога $[без_2]$ составила 289 мсек. Если бы слова «безне(нг)» и «киленебез» не были объединены в одну РГ, то они бы произносились как отдельные ритмические группы, что негативно отразилось бы на качестве синтезированной речи.

Фонетическая транскрипция, точнее моделирование ритмической организации РГ, предполагает в нашем случае расстановку слогов по длительностям, а также репрезентацию фонетических явлений, в частности, изменение речевых звуков на границе объединяемых в РГ слов.

Как было отмечено, длительность слога к концу РГ возрастает. Надо полагать, что увеличение длительности слогов происходит постепенно. Так, в бессмысленной РГ, представляющей сочетание слогов //та - та - та - та - та//, произнесенной в середине предложения, длительности распределились в следующем порядке //259 мсек, 282 мсек, 296 мсек, 303 мсек, 319 мсек.

Но поскольку в реальной речи словосочетания из одинаковых слогов не встречаются и, кроме того, ритмико-интонационная структура речевых тактов формируется наложением различного рода факторов, при моделировании ритмической структуры все слова, составляющие ритмическую группу, склеиваются в одно слово, в результате чего слоги, за исключением слогов начальной и конечной позиций, считаются серединными. Этот факт учитывался при построении элементной базы синтезатора: дифоны серединной позиции вырезались из бессмысленной трехсложной конструкции, в которой слог с нужным дифоном находился во втором слоге от начала слова.

Более существенным в плане формирования ритмики речевого такта было выявление случаев редукции и полного выпадения гласных (фонетический эллипсис), главным образом гласных $[o]$, $[o']$, $[ы]$, $[e (э)]$.

В фонетике татарского языка указанные гласные называются краткими. Более того, при фонологическом описании звуков языка признак краткости употребляется в качестве дифференциального признака данных гласных [4]. Последнее утверждение, на наш взгляд, является не совсем корректным, поскольку краткость их не носит абсолютный характер. В позициях начала и конца слова, в полностью закрытых слогах типа CVC их длительность сопоставима с длительностью остальных (некратких) гласных той же позиции (С - консонант/согласный, V - вокал/гласный).

Своеобразие гласных $[o]$, $[o']$, $[ы]$, $[e (э)]$ проявляется в случаях, когда они следуют в составе слогов CV. Исследования с использованием программ COOL Edit и Goldwave выявили зависимость редукции данных гласных от следующих факторов:

² Здесь и далее для обозначения ряда специфических татарских фонем используются следующие обозначения: a' – соответствует звуку $[a]$ слова «сядь», y' – соответствует звуку $[y]$ слова «тюк», o' – соответствует звуку $[o]$ слова «сёмги», $(нг)$ - соответствует звуку $[ng]$ английского слова «thing»

- 1) Редукция кратких гласных, в том числе и фонетический эллипсис³, наблюдается в слогах типа CV.
- 2) Гласные [o], [o'], [ы], [e (э)] в большей степени подвергаются редукции и эллипсису, когда они располагаются ближе к началу слова.
- 3) Редукция и эллипсис кратких гласных в значительной степени зависят от контекста.
- 4) Редукция кратких гласных зависит от позиции их в слове и в РГ.
- 5) Явления редукции и эллипсиса зависят от длины слова

Указанные зависимости были учтены при моделировании ритмической организации речевого такта. Были приняты две градации степени редукции гласного - редукция на 50% и редукция полная, фиксировались явления фонетического эллипсиса. При этом мы руководствовались принципом разумной достаточности в приближении к естественной речи, а также сведением к минимуму внесения каких бы то ни было изменений в исходные единицы конкатенации.

Фонетическая транскрипция, или моделирование структуры РГ, состоит из двух частей. В первой части алгоритм работает на уровне отдельных слов РГ, слова анализируются на наличие слогов CV с кратким гласным и осуществляется их транскрипция. Во второй части исследуются пограничные области объединяющихся слов РГ и формируются стыки согласно правилам составленного алгоритма

В целях более полного представления изложенных положений по формированию ритмической структуры речевого такта приведем часть алгоритма фонетической транскрипции.

Элементарной единицей просодии является слог. Поэтому естественно предположить, что алгоритм моделирования ритмической структуры речевого такта оперирует сегментированными на слоги словами вида $u = q_1q_2 \dots q_k$, где u - слово, q_1, q_2, \dots, q_k , соответственно 1-й, 2-й, ..., k - й (последний) слог слова. Ниже описывается фрагмент алгоритма фонетического транскрибиатора, который начинается с проверки j-го слога, $j=1, 2, \dots, k$, на слог типа CV.

1. проверить слог q_j на тип CV (2, 100)
 2. проверить V_j на гласные [o], [ль], [ы], [e (э)] (3, 100)
 3. проверить $j = 1$ (20, 4)
 4. проверить $j = k$ (100, 5)
 5. проверить $j = k - 1$ (6, 9)
 6. проверить $C_{k,1}$ на согласные [p], [л] (7, 8)
 7. присвоить $V_j = 0,5V_j$ (Примеры: утырам, кайттылар)
 8. сохранить V_j без изменения (Примеры: итегем, киттегез)
 9. <переход к анализу слова, в котором CV с кратким гласным не находится ни на первом, ни на предпоследнем и ни на последнем местах>
- * *
*
* *
20. проверить $k - j = 1$ (21, 50)
 21. проверить $q_2 = C_{21}V_2C_{22}$ (22, 40)

³ Выражение «полная редукция» означает сокращение длительности гласного до нуля с сохранением переходных участков. Термин «фонетический эллипсис» обозначает выпадение гласного в контексте CVC, т.е. замену CVC на CC.

22. проверить $C_1 = \{T\}$ (23,30)
 23. проверить $C_{21} = \{R\}$ (24,25)
 24. удалить V_1 (Примеры: кырык , телим)
 25. проверить $C_{21} = \{S\}$ (26,27)
 26. выполнить $V_1 \rightarrow \emptyset$ (Примеры: пычак, тышау)
 27. <переход к рассмотрению случая, когда согласные как анализируемого слога, так и слога, следующего за ним, относятся к классу смычных>

* *
 *

40. проверить C_{21} на согласные [p], [л] (41, 42)
 41. присвоить $V_1 = 0,5V_1$ (Примеры: тере, теле)
 42. сохранить V_j без изменения (Примеры: кеше, тыга)
 * *
 *
 50. проверить $k - j = 2$ (<переход к анализу трехсложных слов >)
 * *
 *
 100. <переход к анализу следующего слова предложения>

Использованные обозначения: графемы, заключенные в квадратные скобки, соответствуют речевым звукам; значения прописных букв в фигурных скобках: $\{T\}$ - класс смычных, $\{S\}$ - класс спирантов, $\{R\}$ - класс сонорных звуков; запись "присвоить $V_j = 0,5V_j$ " означает "сократить длительность находящегося в j - й позиции гласного в два раза"; запись "удалить V_1 " означает заменить TVR на TR (случай фонетического эллипсиса); запись "выполнить $V_1 \rightarrow \emptyset$ " означает "удалить гласную из слога, стоящего в позиции начала слова, сохранив при этом его переходные участки"; цифры в круглых скобках: первая - соответствует ответу "да" и указывает, какую команду на следующем шаге следует выполнить, вторая - соответствует ответу "нет" и указывает, куда в этом случае следует переходить.

Приведем краткий комментарий к отдельным правилам.

Из правила 4 следует, что, когда слог CV с кратким гласным находится в конце слова, то редукция не наблюдается. В сочетании двух CV, образованных краткими гласными, гласный предпоследнего слога сокращается в два раза только в том случае, если слог, следующий за предпоследним начинается на согласные /p/ или /л/ (Правила 5, 6, 7).

Правила с 20 по 27 дают некоторое представление о зависимости редукции от позиции слога CV в слове, а также от контекста. Отметим, что, если слог с кратким гласным употребляется в начале трех- и более сложных слов, то, как правило, краткая гласная выпадает, т.е. в этих случаях наблюдается фонетический эллипсис.

Вторая часть алгоритма фонетической транскрипции относится к исследованию и формированию пограничной области входящих в РГ слов.

Объединение слов в РГ в одно целое приводит к многочисленным фонетическим изменениям пограничных фонем. Частично они были рассмотрены в [5]. Остановимся более подробно на явлениях перераспределения слогов в сочетаниях, входящих в одну ритмическую группу слов.

Перераспределение слогов в РГ происходит в следующих случаях, когда:

- 1) первое слово в сочетании заканчивается на закрытый слог, а второе слово начинается на неприкрытый закрытый слог (китап алдым, со'т эчтек).
- 2) первое слово в сочетании заканчивается на закрытый слог, а второе слово начинается на неприкрытый открытый слог (ит ашый, укып утыра).
- 3) первое слово в сочетании заканчивается на прикрытый открытый слог, а второе слово начинается на закрытый неприкрытый слог (алма алдым, киттеме инде).
- 4) первое слово в сочетании заканчивается на прикрытый открытый слог, а второе слово начинается на неприкрытый открытый слог (карама а'ле, ярма ашы).

В случаях 1 и 2 перераспределение происходит регулярно. Для звуковой системы татарского языка (впрочем, как для системы многих других языков) не характерно употребление слогов типа V и VC в середине и в конце слова. Поэтому при слитном произнесении слов u_1 и u_2 , когда u_1 заканчивается на закрытый, а u_2 начинается на открытый слог, обязательно происходит перераспределение слогов, т. е. наблюдается картина

$$\dots C_{11}V_1C_{12} + V_2 \dots \rightarrow \dots C_{11}V_1 + C_{12} + V_2 \dots$$

$\dots C_{11}V_1C_{12} + V_2C_{21} \dots \rightarrow \dots C_{11}V_1 + C_{12}V_2C_{21} \dots$. При этом, если первое слово в сочетании заканчивается на глухие [п] и [к], то они переходят в свои звонкие пары.

В случаях 3 и 4 характер фонетических изменений, в том числе перераспределение слогов, зависят от нескольких факторов.

Известно, что произнесение сочетания гласных, как и произнесение сочетания согласных, является неудобным для речевых органов. Сказанное целиком относится и к звуковой системе татарского языка, о чем свидетельствует отсутствие в ней истинных дифтонгов. С другой стороны, назначение речи - в передаче значений - и, говорящий вынужден произносить (обозначить) неудобные для артикуляции сочетания гласных, чтобы не нанести ущерба передачи смысла слов.

Так, в случае, когда u_1 заканчивается на гласную (V_1), а u_2 начинается с гласной (V_2), то конечная гласная слова u_1 редуцируется. Степень редукции в значительной степени зависит от употребительности данного сочетания слов в языке. Полная редукция, эллипсис, а также перераспределение слогов наблюдаются в том случае, когда объединяющиеся слова являются устойчивыми словосочетаниями, образованными либо сочетанием основного и вспомогательного глаголов (бара алмадым), либо сочетанием знаменательного и служебного слов (со'йла' а'ле).

Редукция в сочетаниях u_1 и u_2 , где u_1 заканчивается, а u_2 начинается на гласный, зависит от качества гласного V_1 . В тех случаях, когда u_1 заканчивается на краткие гласные [ы] или [е], а u_2 является служебным словом, то при их слитном произнесении наблюдается сокращение кратких гласных до нуля (кила'ме ика'н — кила'мика'н, кайттымы а'лла' — кайттыма'лла').

Редуцирование V_1 (а не V_2), возможно, вызвано тем, что гласные абсолютного начала произносятся с некоторым приступом, фиксирующим начало гласного и обеспечивающим его полное произнесение.

Во всех остальных случаях слитного произнесения слов u_1 и u_2 , заканчивающихся и начинающихся на гласные звуки, гласная V_1 значительно редуцируется. Редукция, по нашим оценкам, составляет в среднем 50% от длительности гласного в данной позиции. Понятно, что при этом перераспределения слогов словоформ u_1 и u_2 не происходит.

Выявленные фонетические изменения (перераспределение слогов, сокращение длительности V_1 до нуля и на 50%, фонетический эллипсис) были внесены в алгоритм

фонетической транскрипции (моделирования ритмической группы) и, по отзывам auditors, улучшили качества синтезируемой речи.

Приведенный анализ построения ритмической группы способствует раскрытию принципов, лежащих в основе звуковой системы татарского языка.

The tatar Speech Synthesizer: Phonetical Ellipsis and Sound alteration at the ends of Syllables.

T.I. Ibragimov, F.I.Salimov, D.S.Suleymanov, R.R.Khusainov

Key words: Speech Synthesizer, phonetic transcription, phonetical ellipsis

Upgrading of quality of the tatar speech being synthesized is the concern of the present paper. Data for the position and context alternations of phonemes are given, among them the data for the partial reduction and the complete disappearance of the short vowels in the speech flow

Литература

- 1) Л. Блумфильд. Язык. М.: Прогресс, 1968.
- 2) Щерба. Фонетика французского языка. М.: Иностранная литература, 1957.
- 3) Виллер М.А. О зависимости французской ритмической группы от стиля речи. // Сб. "Звуковой строй языка". М.: Наука, 1979 - С.44 - 51.
- 4) Татарская грамматика. Казань, Татиздат, 1993, Т. 1
- 5) Ибрагимов Т.И., Хусаинов Р.Р. Синтезатор татарской речи: формирование ритмических групп в речевом потоке. // Сб. научных трудов Казанской школы по компьютерной и когнитивной лингвистике ТЕЛ - 2000. Казань: Изд - во "Селет", 2000, С. 80 - 86.
- 6) Ибрагимов Т.И., Салимов Ф.И., Хусаинов Р.Р. Синтезатор татарской речи: Вопросы транскрипции заимствований и планирования языка // Труды международного семинара Диалог 2002 Т. 2. М.: Наука, 2002, С.228-233.