К формированию акустической базы синтезатора татарской речи[1]

Т.И. Ибрагимов (Tavzich.Ibragimov@ksu.ru)
Казанский государственный университет, филологический факультет

Ф.И. Салимов (Farid.Salimov@ksu.ru),
Д.Ш. Сулейманов (dvdt@telecet.ru),
Р.Р. Хусаинов (rust@ksu.ru)
Казанский государственный университет, факультет ВМК

Рассматривается задача построения акустической базы для синтезатора татарской речи. Рассмотрены лингвистические аспекты построения такой базы.

Разрабатываемый синтезатор является конкатенативным, в котором в качестве исходной единицы используется дифон. Акустическая база синтезатора включает три типа дифонов – начальный, срединный и конечный.

В синтезаторах описываемого типа начальные и конечные дифоны, как правило, представляют половинки первой и последней фонем слова с включением переходных участков от пробела к фонеме, а также от фонемы к пробелу соответственно[1]. В разрабатывемом синтезаторе татарской речи начальные дифоны состоят из сочетания первой в слове фонемы и половины следующей фонемы; срединные дифоны представляют сочетание пограничных половинок соседствующих фонем, конечные дифоны состоят из сочетания последней в слове фонемы и половины предшествующей фонемы. Таким образом, начальные и конечные дифоны в синтезаторе татарской речи соответствуют полуслогам. Так же, как и полуслоги, они могут содержать кластер согласных. Следует отметить, что для фонетической системы татарского языка стечение согласных не характерно (встречаются лишь следующие сочетания /йт/, /рт/, /лт/, /нт/, /йк/, /йм/).

Включение в акустическую базу дифонов в виде полуслога несколько увеличивает объем базы. Но оно, как нам представляется, оправдано. Дело в том, что для татарского языка не характерно словесное ударение. В осутствие словесного ударения в формировании ритмической организации речи важную роль играет разбиение предложения на ритмические группы. Последние могут состоять из одной или из объединения двух-пяти лексем, произносимых как одно слово. Таким образом, начальные и конечные дифоны (полуслоги) в разговорной речи обозначают не просто границу слова, а границу ритмической группы. Более того, они в значительной степени формируют просодию ритмической группы. Естественно, что своеобразия начального и конечнего слогов полнее отражаются в оцифрованных полуслогах, чем в оцифрованных половинках фонем.

Создание акустической базы происходило следующим образом.

1. Была построена таблица сочетаемости фонем в фонетически транскрибированном тексте с учетом пробела между словами.

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

2. В целях унификации условий озвучивания сочетаний фонем и, по мере возможности, исключения влияния фразового ударения, все фонемосочетания произносились в псевдофразе. Псевдофраза состояла из трех ритмических групп. Начальная и конечная ритмические группы представляли осмысленные словосочетания, средняя ритмическая группа состояла из одного трех-четырехсложного бессмысленного слова. Однако по звуковому составу и структуре это слово полностью соответствовало системе татарского языка. За редким исключением все дифоны озвучивались в следующей псевдофразе "Безнен'[2] авылда (*ш*) / …….. / тик торып булмый", в которой косая черта указывает границы первой и второй ритмических групп, многоточия заменялись трех– или четырехсложным бессмысленным словом, содержащим необходимое для выделения данного дифона сочетание фонем. Фонемосочетание, из которого выделялся дифон, располагалось либо в начале, либо в середине, либо в конце бессмысленного слова в зависимости от позиции, в которой дифон использовался при синтезе речи. Приведенная в скобках фонема /ш/ присоединялась к слову "авылда" в тех случаях, когда трех - четырехсложное бессмысленное слово начиналось на гласную фонему. Добавление .шипящей согласной /ш/ позволяет точнее определить границу между первой и второй ритмическими группами.

Помимо указанных типов дифонов, в акустическую базу записывались целые слова, состоящие из одной или двух фонем. В этом случае ритмическая группа, состоящая из одного бессмысленного слова, заменялась на сочетание слов, включающее записываемое в базу слово, союз *ha'м*, а также местоимение *син*.

Следует отметить, что к созданию акустической базы на основе дифонов мы пришли не сразу. В первой версии синтезатора татарской речи в качестве исходной единицы использовались слоги [2,3]. Попытка построения слогового синтезатора было предпринята на том основании, что слог, по определению, наименьшая произносительная единица, а число слогов в татарском языке сравнительно невелико – около 2500 единиц. Однако практически сразу же выяснилось, что число слогов, подсчитанных на базе изолированных слов, несколько меньше числа используемых в речи слогов. На стыках слов, входящих в одну ритмическую группу, происходит перераспределения слогов, появляются новые слоги, число которых трудно установить. Кроме того, как оказалось, слог, будучи единицей просодической организации речи, несет в себе особенности ритмико-интонационного построения той фразы, в которой он озвучивается. Это означает, что использование данного слога в других фразах может привести к ухудшению качества синтезированной речи.

Переход к созданию акустической базы синтезатора на дифонах потребовал проведение дополнительных исследований. В частности, необходимо было выяснить - могут ли дифоны (пограничная область сочетания согласных), выделенные в псевдословах твердого произношения, использоваться в озвучиваниии слов мягкого произношения. Экспериментальные исследования показали, что в фонетических структурах типа *ат + тк +ка*, *эш´+ш´с´+с´е+ез´* палатальность/непалатальность согласных, не граничащих с гласными (в приведенных примерах дифоны *тк* и *ш´с´´*), не различаются слуховой системой. Иными словами, дифоны *тк* и *т´к´* или, например, *шс* и *ш´с´*, выделенные в парах как то *атка – этка'* и *ашсыз – эшсез,* являются взаимозаменяемыми. Исключение составляют сочетания согласных, содержащих в своем составе сонорную фонему /л/. Дифоны, представляющие сочетания *сс* (согласный + согласный), одним из которых является звук [л], были включены в акустическую базу синтезатора в двух вариантах - палатализованном и непалатализованном.

В свете результатов проведенных исследований вызывает некоторое сомнение целесообразность включения в состав фонем татарского языка твердых согласных /q/ и /g//. [4] Как показали эксперименты, твердость/мягкость фонем /к/ и /г/ определяются гласными, точнее, переходными участками от гласной к согласной и от согласной к гласной фонеме. Если же фонемы /q/ и /g/ вводятся в звуковую систему языка в целях правильного произношеня букв *к* и *г* в составе заимстований, то для достижения такой цели необходимо включить в состав фонем все гласные русского языка.

Создание акустической базы синтезатора на дифонах вызвало также необходимость проведения исследований по определению значимости слоговой границы в фонетическом или фонологическом отношениях. Дело в том, что в татарском языке слоги выделяются довольно четко. Исследования,

[2] В примерах, где встречаются татарские буквы будет использоваться следующая нотация: а',о',у',ж',н',h (со знаком апострофа в правом верхнем углу). Символ а' произносится как [a] в слове [cat], символ о' произносится как [ir] в слове [bird], символ у' произносится как [w] в слове [down], символ ж' произносится как [sio] в слове [adhesion] или как [зж] в слове [дозжек], символ н' произносится как [ng] в слове [sing], символ h произносится как [h] в слове [hat].

проведенные в рамках теории информации и математической статистики, показали, что фонемы, входящие в слог, проявляют более жесткую организацию, чем неслоговые сочетания. Было выявлено [5], что в изолированно произнесенных словах длительность слогоконечной фонемы несколько больше длительности неслогоконечной. Все это вызвало необходимость предварительного изучения проблемы – находят ли указанные особенности формирования слогов и слоговой границы отражение в акустических характеристиках дифонов, а, следовательно, в качестве синтезированной речи. Другими словами, необходимо было выяснить - можно ли дифоны *ал* и *аш*, выделенные, допустим, в словах *алан* и *кашык* (1), использовать при синтезе слов *алма*, *кашлы* (2) и т.д. (Слова *алан* и *кашык* представляют сочетание слогов *а+лан* и *ка+шык*, а слова *алма* и *кашлы* – сочетание слогов *ал+ма* и *каш+лы*) Эксперименты по аудированию синтезированной речи, в которой дифоны, включающие границы слогов, использовались как при озвучивании слов типа (1), так и при озвучивании слов типа (2), не обнаружили значимых различий. Следовательно, одноименные (состоящие из пограничных половинок одних и тех же фонем) дифоны, включающие границу слогов и не включающие границу слогов, являются взаимозаменяемыми.

Создана элементная база оцифровок дифонов для мужского голоса. При создании дифонной базы были выбраны следующие параметры квантования сигнала: частота дискретизации -22025Гц, разрядность выборки - 16 бит. На данный момент словарь содержит порядка 1000 различных дифонов. По нашим оценкам это составляет 90% всех дифонов. Как упоминалось в начале статьи, дифонная база была построена на базе фонематического транскрибированного текста. В [6] отмечалось, что в речи слова объединяются в ритмические группы, и что при этом происходит перераспределение слогов и появляются новые слоги, и, следовательно, новые фонемосочетания. Пополнение акустической базы в настоящее время происходит, в основном, за счет неучтенных фонемосочетаний. Продолжаются работы по выявлению некачественно записанных дифонов и по перезаписи их.

Литература

1. Meelis Mihkla, Arvo Eek, Einar Meister. Diphone synthesis of Estonian. //Труды Международного семинара Диалог'99 по компьютерной лингвистике и ее приложениям. т.2. - Таруса: 1999.
2. Ибрагимов Т.И., Рахматуллина Г.А., Салимов Ф.И., Хусаинов Р.Р. Компилятивный синтез татарской речи, Тр. Международного семинара «Диалог-98» по компьютерной лингвистике и ее приложениям, т.2 – Казань 1998 с. 472-477.
3. Ибрагимов Т.И., Рахматуллина Г.А., Салимов Ф.И., Хусаинов Р.Р. Синтез татарской речи методом конкатенации слогов, Материалы конференции «Теория и практика речевых исследований» (АРСО-99) 1999 с. 23-24.
4. Татарская грамматика. т.1. - Казань: 1993, 584с.
5. Т.И.Ибрагимов. Изучение образования слогов и структуры их сочетаний в татарском литературном языке. Автореф. дисс. на соиск. уч. степ. канд. филол. наук. Казань – 1970.
6. 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 "Сэлэт", 2000, с. 80-86.