АЛГОРИТМЫ И ПРОГРАММНЫЕ СРЕДСТВА ОБРАБОТКИ ИНФОРМАЦИИ

ПРОЕКТИРОВАНИЕ И РАЗРАБОТКА ПРОГРАММНОГО КОМПЛЕКСА СИНТЕЗА КАЗАХСКОЙ РЕЧИ ПО ТЕКСТУ

Мусабаев Р.Р.
Институт проблем информатики и управления МОН РК,
Алматы, Казахстан e-mail: rmusab@gmail.com

Введение

В данной статье рассматривается вопрос программной реализации системы синтеза казахской речи [1-3]. Разработанный комплекс программ является инструментарием, с помощью которого имеется возможность проводить научные исследования в области изучения интонационной структуры различных языков. С применением данной системы процесс исследования казахской интонации возможно осуществлять на качественно новом уровне – с применением технологии речевого синтеза. При этом разработанная система может выступить не только в качестве инструментария исследователя, но и в роли платформы для разработки систем синтеза более высокого уровня, использующих интонационную модель языка.

1 Структура информационной системы

Фонетический уровень звукового синтеза – это базовый фундамент, на котором строится любая комплексная система компилятивного синтеза речи. На рисунке 1 представлена общая схема комплексного взаимодействия фонетического уровня (серый цвет) с другими подсистемами синтеза речи.

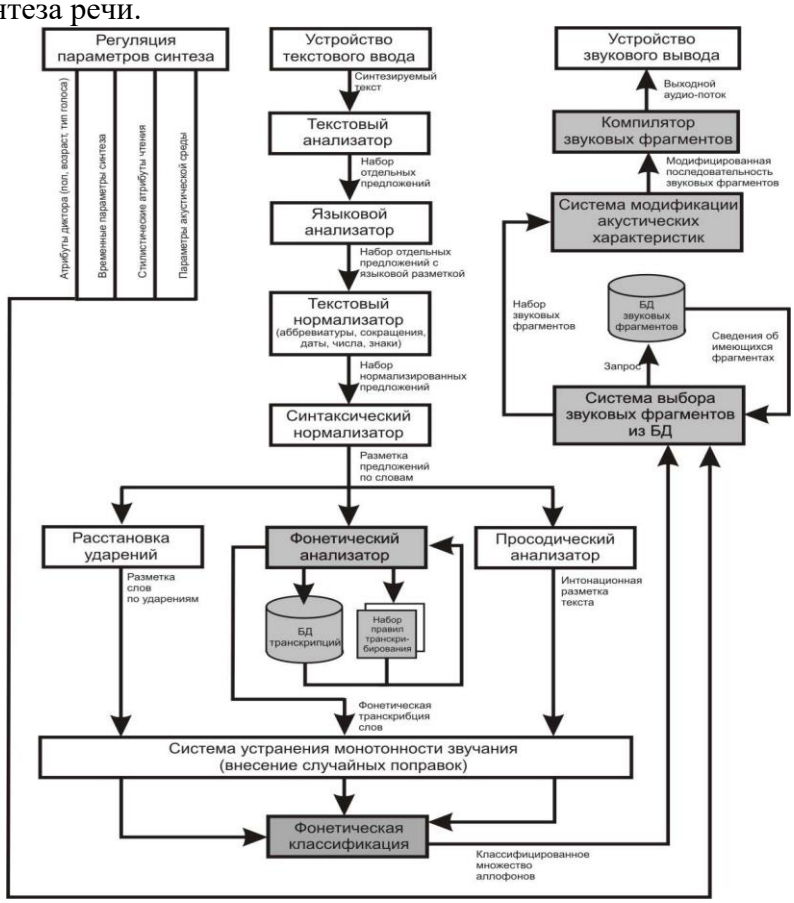

Рисунок 1. Общая схема комплексного взаимодействия фонетического уровня (серый цвет) с другими подсистемами синтеза речи

Система компилятивного синтеза речи – это многоуровневая комплексная система, состоящая из множества подсистем. Все подсистемы в рамках единой системы решают общую задачу получения синтезированного аудио-потока. Каждая из подсистем решает свою собственную задачу и имеет набор входных и выходных параметров. Выходные параметры подсистем более высокого уровня передаются на входные параметры подсистем более низкого уровня.

Можно сказать, что фонетический уровень звукового синтеза является ядром системы. От его реализации во многом зависит качество всей системы в целом. Если его реализация не удовлетворяет необходимым условиям качества, то это может свести к минимуму все качественные показатели, полученные на более высоких уровнях. Этот уровень также является наиболее трудоёмким при реализации в технологическом плане. Необходимо учитывать специфику и правила акустического представления отдельных фонем, их взаимное влияние друг на друга, учитывать эффекты коартикуляции и наложения нескольких звуков, правила изменчивости звукового представления фонем при различных просодических условиях. Также необходимо иметь функциональную зависимость частоты основного тона, энергии и длительности фонетических единиц от текущей просодической схемы. При звуковом синтезе фонем необходимо учитывать особенности психоакустики, которая описывает нелинейные свойства человеческого слуха. Концепции психоакустики используются при регулировании таких параметров как:

1. Громкость речевого сигнала.
2. Высота звука.
3. Тембр.
4. Продолжительность звучания.

2 Разработанный программный инструментарий

При практической реализации системы речевого синтеза имеется множество трудоёмких задач. К ним относятся:

1. Запись исходного речевого материала.
2. Первичная обработка записанного речевого материала.
3. Выделение речевых макрофрагментов (предложения, фразы, слова, фонемы и др.).
4. Выделение периодических микросегментов вокализированных участков речевого сигнала согласно изменению частоты основного тона.
5. Нормализация речевого сигнала.
6. Сохранение результатов выделений в отдельные файлы.

Так если первые две задачи могут быть решены с применением стандартных инструментальных средств, то процесс решения остальных задач необходимо максимально автоматизировать так как он требует значительных затрат времени и усилий.

Так для целей создания дифонного синтезатора речи разработано специализированное ПО WavTranscriber, которое в значительной степени облегчило процесс выделения и классификации дифонов. ПО построено по универсальному принципу - его можно с лёгкостью перестроить на работу с любым языком. Основным назначением данного ПО является формирование классифицированной БД дифонов, которая впоследствии будет использована при построении универсального речевого синтезатора.

Программа для формирования дифонной базы языка позволяет выполнять следующие функции:

1. Выделение дифона из звукозаписи натуральной произношения выбранного слова.
2. Помещение выделенного дифона в БД.

3. Автоматическая классификация выделенного дифона в зависимости от его фонемного окружения, положения в слове, длительности собственного звучания, собственной амплитуды и частоты.

4. Возможность ручной классификации с помощью регулирования уровня значимости внутри множества дифонов подпадающих под определения одного класса.

5. Имеются различные режимы сортировки и навигации по сформированной базе дифонов.

6. При выделении нового дифона автоматически отображается список всех доступных звукозаписей в которых данный дифон должен присутствовать.

7. Выделение, воспроизведение и сохранение во внешний файл любого фрагмента звукозаписи речи. Данная функция особенно важна при опытном определении положения и границ дифона внутри звуковой записи слова.

8. Осуществление автоматического контроля качества базы данных звукозаписей естественной речи (контроль обрезки звуковой волны по максимальному уровню, контроль по средней продолжительности звукозаписи приходящейся на одну фонему).

Таким образом, сформированная база данных в автоматическом режиме проклассифицирована по 1500 дифонам по следующим признакам:

1. Расположение дифона в слове (начало, середина, конец, особняком).

2. Предыдущая фонема (отсутствует, гласная, согласная).

3. Последующая фонема (отсутствует, гласная, согласная).

4. Длительность звучания (в виде отклонения от средней длительности).

5. Амплитудная и частичная классификация начального участка дифона.

6. Амплитудная и частичная классификация конечного участка дифона.

В приложении задействованы следующие БД:

1. БД звукозаписей из 118 965 продиктованных слов в формате WAV и MP3. Звукозаписи продиктованы носителями языка, имеют частоту дискретизации 22050 Гц и разрядность сигнала в 16 бит. Эти данные являются источником для формирования БД дифонов.

2. БД фонетических транскрипций. Содержит 129 468 транскрипции.

3. БД фонетического состава языка.

4. БД всех возможных полуфонемных комбинаций – дифонов, со статистической информацией, которая используется для расчета приоритета в обработке. В первую очередь выделяются дифоны, которые имеют наибольшую частоту вхождения в языке.

В программе можно просматривать форму звуковых колебаний, выделять и проигрывать фрагменты, сохранять выделенные фрагменты в БД. Классификация дифонов выполняется автоматически: программа выдаёт звукозапись, указывает дифон, который необходимо выделить и после выделения сохраняет его со всей классификационной информацией в БД. Для наибольшего удобства применяется цветовая подсветка различных состояний и классификаций.

В качестве системы управления базами данных (СУБД) используется MS SQL Server 2005. Доступ к данным осуществляется с помощью языка структурированных запросов (SQL) с применением технологии dbExpress. Само приложение написано на языке Object Pascal в среде Delphi 7.

Вид главной формы приложения WavTranscriber показан на рисунке 2.

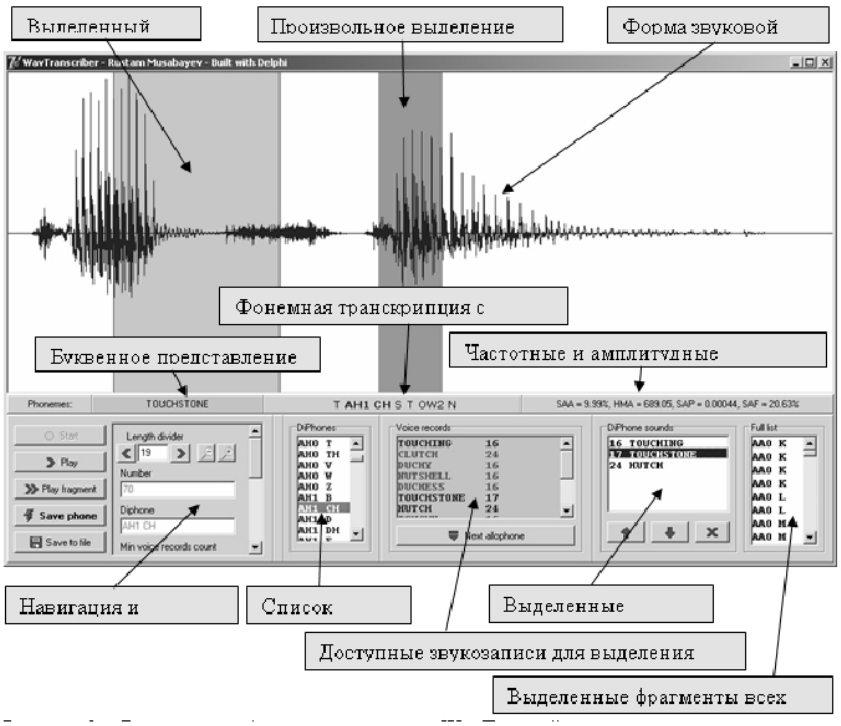

Рисунок 2 – Вид главной формы приложения WavTranscriber

Необходимость в использовании данной программы возникает только на этапе разработки синтезатора речи. Также её можно использовать для устранения выявленных дефектов синтеза – можно заменить имеющийся менее качественный дифон вновь выделенным и более качественным по произношению.

Заключение

Все основные исследования и разработки проводились на примере казахского языка. Для целей автоматизации синтеза была исследована его фонетико-акустическая структура, проведена полная классификация фонетического состава, получена статистическая информация и сформированы различные по назначению и содержанию лингвистические базы данных, которые в последствии могут быть использованы для проведения отдельных научных исследований в области казахского языкознания.

Для качественного решения поставленной задачи разработан унифицированный язык фонетического представления, который позволяет задавать и описывать разнообразие фонетических и интонационных форм речи. Все исходные данные модели описываются с помощью унифицированного языкового представления, что позволяет осуществлять гибкое межсистемное взаимодействие.

Нерешённым остаётся вопрос построения интонационной модели казахского языка и последующий процесс её алгоритмизации для построения полнофункционального синтезатора казахской речи по тексту.

Список литературы

1. Амиргалиев Е.Н., Мусабаев Р.Р. Методы анализа и проектирования системы синтеза искусственной речи // Таврический Вестник Информатики и Математики Таврического Национального Университета. Украина, – 2008. №1. – С.51–59.
2. Амиргалиев Е.Н., Мусабаев Р.Р. Методы обработки сигналов в системе синтеза речи. / Труды Института вычислительной математики и мат. геофизики СО РАН. 2009.– С.14–22.
3. Амиргалиев Е.Н., Мусабаев Р.Р. Один метод модуляции речевого сигнала по амплитуде и его применение в системах синтеза и клонирования речи // Вычислительные технологии ИВТ СО РАН. 2010, № 1. – С. 25–29.