OSTIS-2011
(Open Semantic Technologies for Intelligent Systems)

УДК 004.934.1

МЕТОДЫ СЕМАНТИЧЕСКОГО АНАЛИЗА ДЛЯ ПОСТРОЕНИЯ ГОЛОСОВЫХ ИНТЕРФЕЙСОВ: СИНТЕЗ РЕЧИ

Б.М.Лобанов (*lobanov@newman.bas-net.by*)

Объединенный институт проблем информатики, Минск, Республика Беларусь

В статье предлагается подход к синтезу речи на основе многоуровневых семантических сетей, позволяющий сделать технологию синтеза речи открытой и легко модифицируемой.

1. Структура синтезатора речи по тексту

Описываемая компьютерная модель синтеза речи базируется на результатах многолетних исследований по созданию лингво-акустических основ синтеза речи по тексту [1]. В модели аккумулированы теоретико-экспериментальные сведения о специфике лингвистической обработки текстов, фонетической и просодической структуре русской речи, артикуляторно-акустических явлений процесса речеобразования. Структура системы синтеза речи по тексту представлена на рисунке 1.

Отличительной особенностью описываемой модели, нашедшей отражение в её названии – «мультиволновый синтез», является использование в качестве элементов компиляции речи отрезков естественной речевой волны, соотносимой с элементами различной фонетической длинны: аллофонами, диаллофонами и аллослогами.

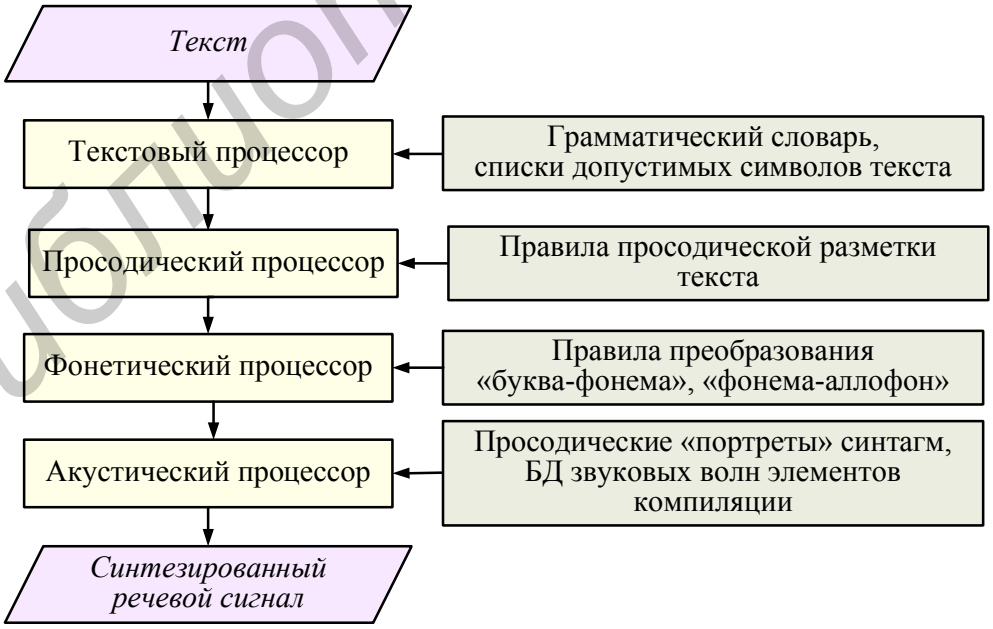

Рисунок 1 - Структура системы синтеза речи по тексту

Синтез устной речи по тексту осуществляется на основе лексико-грамматического анализа входного текста путём моделирования процессов речеобразования с учётом правил произношения звуков и интонирования, свойственных данному языку. Орфографический текст документа (книги, статьи, веб-страницы и т.п.) поступает на вход синтезатора и далее подвергается последовательной обработке рядом специализированных процессоров в соответствии с общей структурой синтезатора речи по тексту, представленной на рис. 4.1. Синтезатор включает четыре основных модуля: текстовый процессор, просодический процессор, фонетический процессор и акустический процессор. Каждый из этих модулей поддерживается наборами соответствующих БД и правил. Рассмотрим основные функции этих модулей.

2. Текстовый процессор

Текстовый процессор (рисунок 2) включает два основных блока, которые поддерживаются соответствующими базами данных, словарями и правилами. Он выполняет предварительную обработку входного текста, а также морфологическую и акцентную маркировку слов текста.

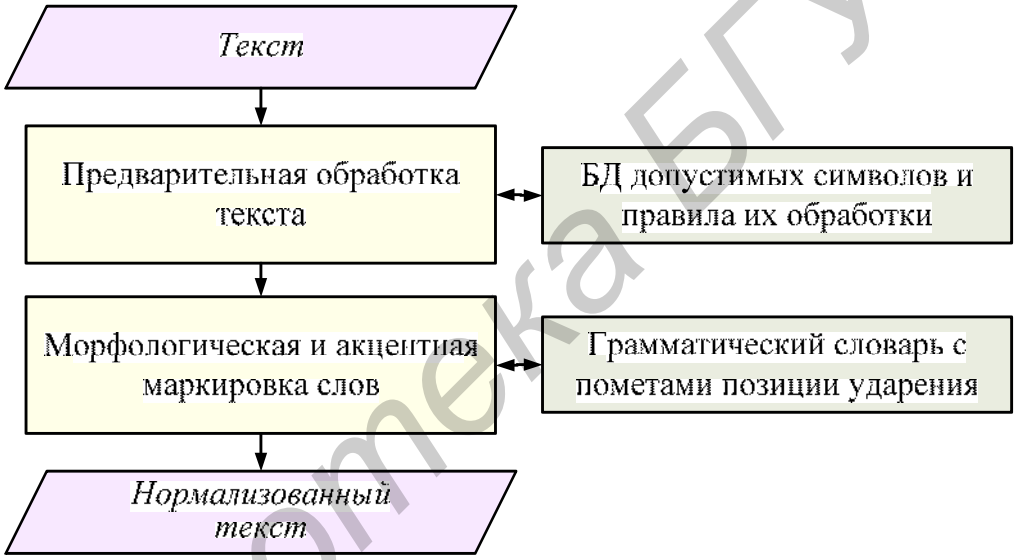

Рисунок 2 - Структура текстового процессора

На вход системы синтеза речи могут поступать тексты, взятые из разных источников и зачастую содержащие графические объекты, ссылки, числовые значения, формулы, а также другие объекты и символы, непригодные для синтеза речи. Основной задачей первого блока - блока предварительной обработки - является нормализация текста, т.е. приведение его к такому виду, когда текст состоит из последовательности слов русского языка. Следующий блок – блок морфо-фонетической маркировки – осуществляют маркировку каждого слова входного текста, необходимую для адекватного синтеза звуков и интонации речи. Для такой маркировки используется грамматический словарь, в котором каждое слово снабжено пометой позиции ударения.

Блок предварительной обработки включает в себя целый ряд этапов, среди которых можно выделить этапы очистки текста, дешифровки чисел, дешифровки аббревиатур, дешифровки иностранных слов и корректировки “ё” (рисунок 3). Очистка текста осуществляется с целью удаления из входного текста графических объектов, ссылок, различных маркеров и других неинформативных для синтеза речи символов. Для реализации этой задачи необходимо иметь БД допустимых символов и объектов, содержащую русские и латинские буквы, знаки пунктуации, цифры, математические символы, а также специальные символы, такие как «@», «^»и т.д. Вообще, в данной БД должны содержаться только те символы, которые могут быть «озвучены» синтезатором речи. Например, если в базе данных содержатся римские цифры или

сложные математические символы, такие как «Σ», «∫», то на последующих этапах обработки текста должны быть блоки, преобразующие последовательности этих символов в слова.

Примечательно, что данный блок с точки зрения разработчика не представляет ни особой трудности, ни особого интереса, и в большинстве случаев при разработке систем синтеза речи реализуется в последнюю очередь. Для пользователей же системы синтеза речи этот блок, напротив, очень важен, поскольку от алгоритмов его работы зависит полнота «озвучивания» входного текста.

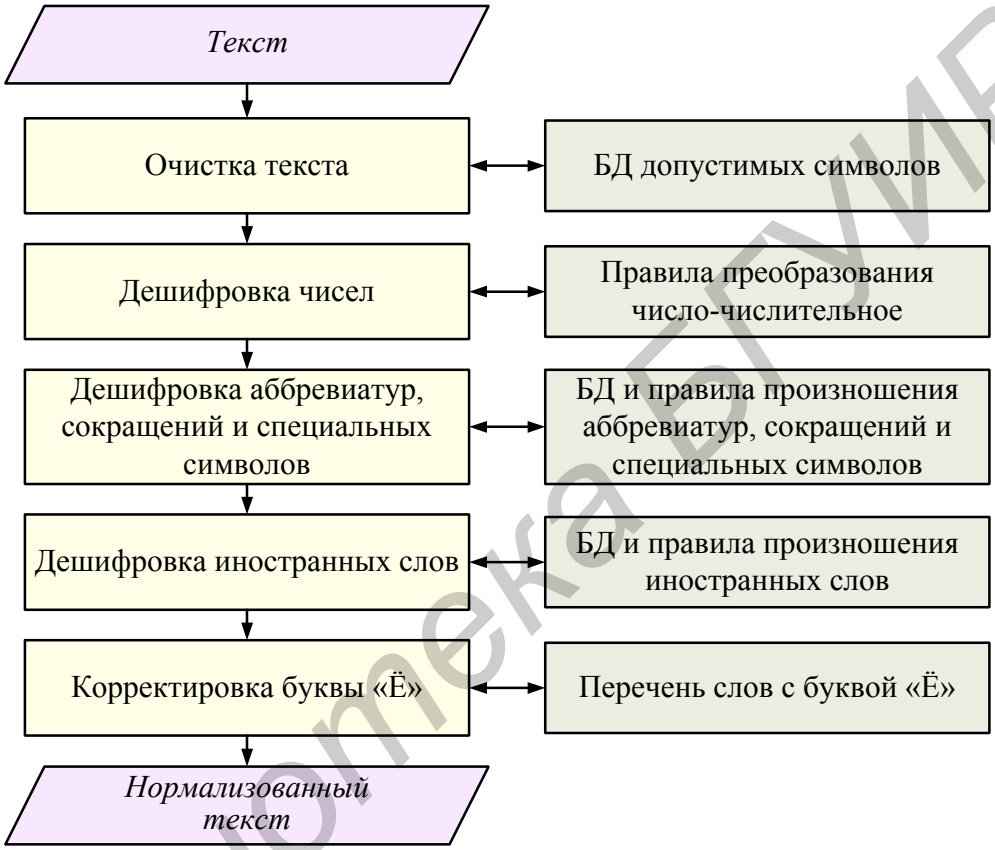

Рисунок 3 - Структура блока предварительной обработки текста

Задача блока дешифровки чисел это преобразовать числа, встретившиеся в тексте, в числительные. При этом необходимо учитывать, что числа, встретившиеся в тексте, могут обозначать целые, десятичные и дробные количественные числительные, порядковые числительные (которые могут быть записаны как арабскими, так и римскими цифрами), дату, время, номера телефонов и т.д. Для корректного преобразования чисел необходимо использовать правила преобразования число – числительное, учитывающие не только число, но и окружающие его слова, сокращения, которые позволяют определить характеристики числа.

Кроме того, необходимо учитывать, что знаки «.» и «,» могут использоваться как для разделения разрядов в целых числах, так и для отделения целой части от дробной. Например, в записи числа 53,45 запятая отделяет целую часть от дробной, а в записи 378,812,547 служит для разделения разрядов.

При синтезе речи необходимо учитывать, что правила чтения аббревиатур, сокращений и специальных символов отличаются от соответствующих правил для слов русского языка. Для решения этой задачи необходимо преобразовать аббревиатуры, сокращения и специальные символы в слова, для которых применимы стандартные правила, используемые на этапах

фонетической и просодической обработки текста. При дешифровке необходимо учитывать следующие факторы:

1. Аббревиатуры в текстах не всегда пишутся заглавными буквами. Это характерно в первую очередь для текстов электронных писем, блогов и других текстов, полученных из различных интернет-ресурсов.
2. Некоторые аббревиатуры и сокращения могут расшифровываться по-разному в зависимости от предметной области, от контекста, например «г.» может означать «город» или «год», «т.» - «товарищ» или «тонн».
3. Некоторые аббревиатуры читаются не в соответствии со стандартными правилами дешифровки, например «США» по правилам расшифровывается как «эс-ше-а», тем не менее общепринятое произношение – «сэ-ше-а».
4. Специальные символы могут преобразовываться по-разному, например «%» – «процент», «процента» или «процентов», «$» - «доллар», «доллара», «долларов».

Для решения этих задач необходимо использовать базу данных и правила произношения аббревиатур, сокращений и специальных символов. Содержащийся в базе данных перечень аббревиатур русского языка позволит обнаружить в тексте аббревиатуру даже в случае, если она записана прописными символами. Перечень сокращений и варианты их расшифровки, а также анализ контекста сокращения позволят корректно преобразовать сокращение в слово.

Аббревиатуры произносятся, как правило, по буквам, например «КГБ» – «ка-гэ-бэ», «ФРГ» – «эф-эр-гэ», при этом каждый слог является ударным. Однако наиболее употребительные аббревиатуры, а также аббревиатуры, содержащие большое количество гласных, произносятся, как правило, в одно слово, например, «ЮНЕСКО» – «юнэ́ско». Это должно учитываться правилами произношения аббревиатур и сокращений. Правила произношения специальных символов для корректного преобразования должны учитывать контекст символа.

В текстах на русском языке могут встречаться интернет-адреса, адреса электронной почты, названия организаций, записанные латинскими символами. Для преобразования таких слов в последовательность русских букв, читаемых по общим правилам, используется блок дешифровки иностранных слов. Этот блок использует БД и правила дешифровки латинских символов. В БД должны содержаться наиболее употребительные иностранные слова, а также их эквиваленты на русском языке, например «Microsoft» – «ма́йкросо́фт», «www» – «три да́блъю». Кроме того, правила дешифровки латинских символов должны содержать русские эквиваленты каждой латинской букве. Тогда в случае, если встретившееся в тексте слово, записанное латинскими буквами, не будет найдено в БД, каждая буква будет преобразована по соответствующим правилам.

Проблема расстановки точек над «ё» - это, пожалуй, проблема только русского языка. Интересно, что человек при чтении текста не задумывается, как правильно прочитать слово, с буквой «ё» или «е», используя для коррекции свои знания о языке. Если же при синтезе вместо, например, слова «ёлка» прозвучит «елка» или вместо «весёлый» – «веселый», такая неточность будет сразу же замечена пользователем. В подавляющем большинстве случаев для корректировки буквы «ё» достаточно лексической информации, а именно БД, содержащей наиболее полный перечень слов с буквой «ё» в русском языке. Тогда в каждом слове текста, содержащем одну или несколько букв «е», каждая из них последовательно заменяется на «ё» и осуществляется поиск соответствующего слова в БД. Однако в некоторых случаях такой информации недостаточно, например, как корректно прочитать слово «все»: «Все в машине?» или же «Всё в машине?». Очевидно, что в этом случае необходимо использовать не только лексический и синтаксический, но и семантический и прагматический анализ.

3. Просодический процессор

Синтез речи по тексту предполагает наличие автоматической процедуры формирования текущих контуров мелодии, силы звука, фонемной длительности и длительности пауз на основе анализа определенных свойств входного текста и его просодической разметки. Просодическая разметка текста заключается в его членении на синтагмы, разметке синтагм на акцентные единицы и маркировке интонационного типа синтагм в соответствии с определёнными правилами.

Под синтагмой понимается самостоятельная в интонационном смысле часть предложения или всё предложение. Установка границ синтагм влияет на передачу интонационных характеристик при синтезе речи, а также на передачу смыслового содержания. При разбиении текста на синтагмы важно не поставить границу синтагмы там, где она может нарушить смысловое восприятие речи (или передачу смыслового содержания текста), например, между предметом и его признаком. Для установки границ синтагм используются определённые правила синтагматического членения, базирующиеся на пунктуационном, морфологическом и синтаксическом анализе текста, а также на статистическом анализе синтагматического членения в естественной речи.

Синтагмы в речи отделяются, как правило, паузами. Паузы принимают участие в передаче определённых синтаксических и смысловых отношений. Кроме того, временные интервалы, создаваемые паузами, позволяют слушателю производить лингвистическую обработку текста, запоминать её результаты и строить смысловую структуру, необходимую для восприятия текста. В естественной речи различают грамматические паузы, отделяющие друг от друга интонационно-оформленные части фразы, выделительные паузы и паузы хезитации (неуверенности). Граница синтагмы может быть промаркирована не только физическим перерывом в речевом сигнале, но и резкой сменой высоты тона и (или) других просодических характеристик, которые воспринимаются как нарушение плавного течения речи.

Важно заметить, что процесс синтагматического членения должен удовлетворять решению двух основных задач: установить границы синтагм в тех местах, где они обязательно должны присутствовать, и не устанавливать границу синтагмы там, где она может нарушить смысловое восприятие речи [2].

Структура просодического процессора представлена на рисунке 4.

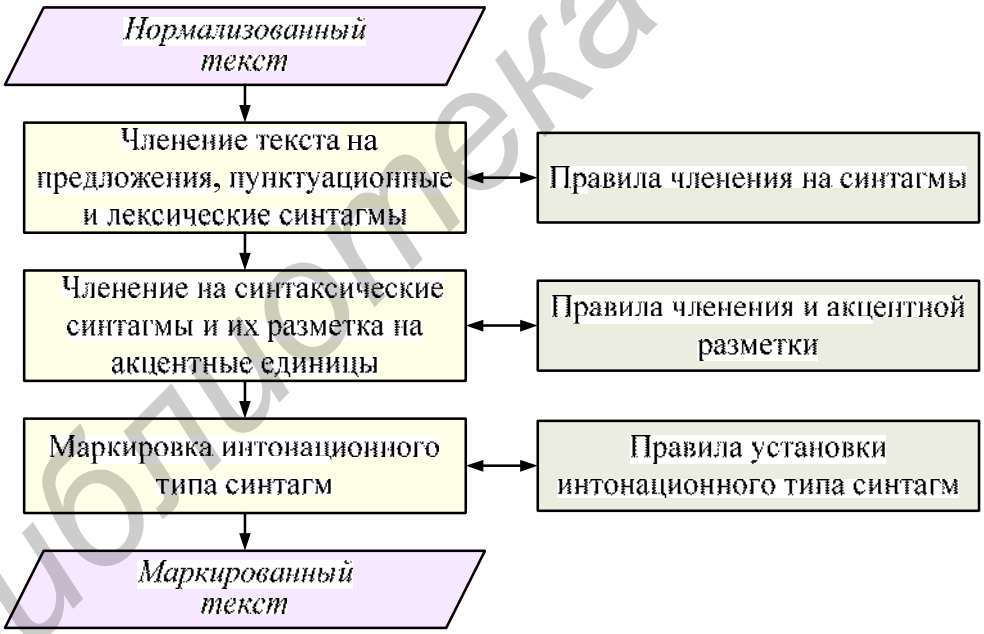

Рисунок 4 - Структура просодического процессора

4. Фонетический процессор

Задачей фонетического процессора является преобразование орфографического текста в последовательность аллофонов, которая используется на этапе акустической обработки при синтезе речевого сигнала.

В фонетическом процессоре заложены правила преобразования орфографического текста в последовательность фонем (преобразование буква-фонема) и правила преобразования

последовательности фонем в аллофонную последовательность (преобразование фонема-аллофон). Общая структура фонетического процессора представлена на рисунке 5.

5. Акустический процессор

Общая структура акустического процессора представлена на рисунке 6. Задачей первого блока акустического процессора является преобразование просодически размеченной последовательности аллофонов синтагмы в последовательность их звуковых волн со значениями ЧОТ – F0, амплитуды – A и длительности – T, задаваемыми БД просодических портретов. Во втором блоке осуществляется синтез речевого сигнала путём выбора из БД звуковых волн мультифонов (единичных аллофонов, диаллофонов, аллослогов), соответствующих входному аллофонному тексту, и их конкатенации (соединения).

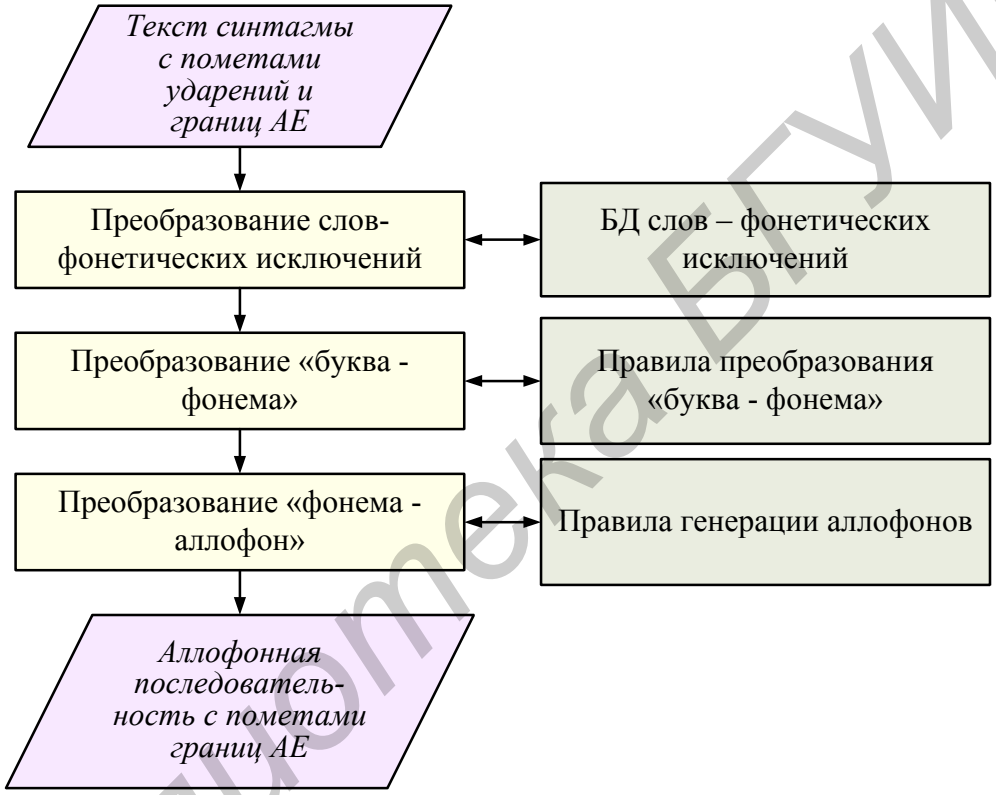

Рисунок 5 - Структура фонетического процессора

6. Семантическая система синтеза речи по тексту

Основной сложностью при создании описанных выше процессоров обработки текста для систем синтеза русской речи является значительная гибкость языка. Практически каждый из представленных выше этапов сопряжен с определенными сложностями, связанных с тем, что в русском языке не существует строгих правил как построения фраз, так и произношения. В результате этого большоое количество фраз допускает разночтение, разное произношение и интонация, устранить которое может только анализ всего предложения, а иногда и всего текста.

Человек при чтении текстов на основе предшествующего содержания и своего опыта легко восстанавливает пропущенные в тексте слова, определяет ударения, части речи слов и т.д. Однако автоматическая система на настоящем этапе развития сделать это не в состоянии, поскольку как уже было упомянуто выше, в русском языке нет строгих правил построения текста, и поэтому жестко алгоритмизировать этапы обработки текста не удается.

Использование базы данных, в которой содержатся различные варианты употребления слов и словосочетаний так же не всегда решают проблему, поскольку ни одна база данных не способна охватить все богатство языка. Кроме этого, в некоторых случаях правильность произношения определяется только на основе анализа и понимания смысла синтезируемого текста.

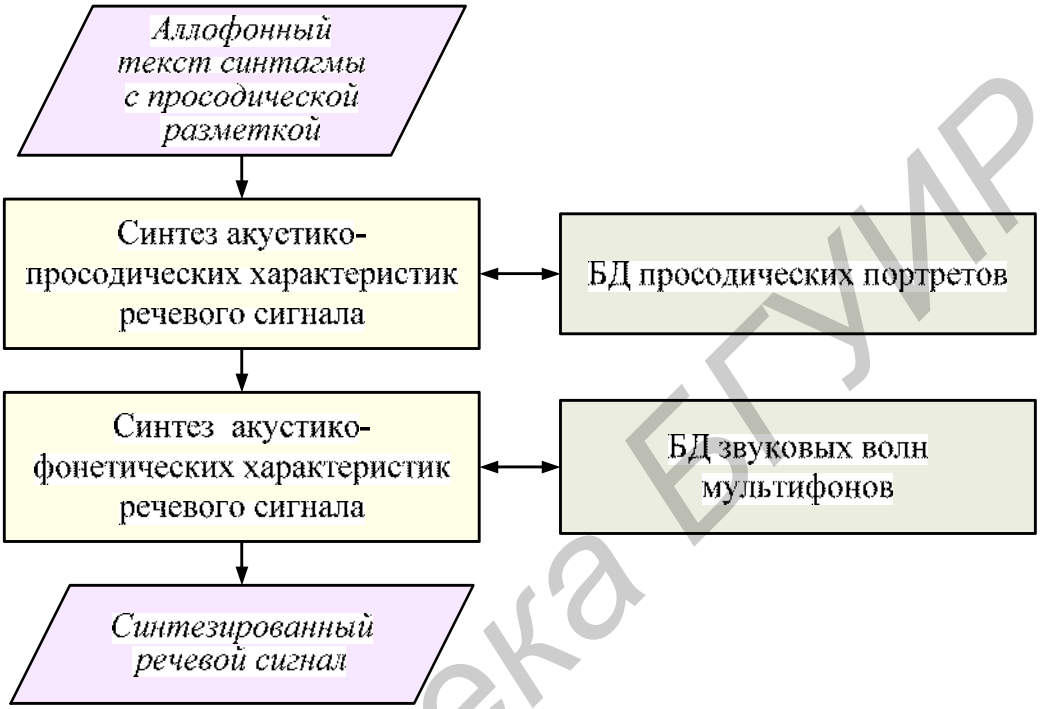

Рисунок 6 - Структура акустического процессора

Для решения такого рода проблем и создания высококачественной системы синтеза русской речи предлагается следующий подход на основе многоуровневого семантического анализа (рисунок 7).

На каждом этапе обработки текста в рамках текстового, просодического, фонетического и акустического процессоров происходит обнаружение так называемых “коллизий”, неоднозначности в применении правила обработки. Алгоритм обнаружения данных “коллизий” в общем случае может носить сложный характер, в простейшем случае будем предполагать, что коллизия образуется тогда, когда в базе данных содержатся слова и выражения, имеющие одинаковое написание, однако разное произношение.

При возникновении коллизии блок коррекции последовательно опрашивает семантические сети разного уровня, начиная с самого низкого (уровень слов), если коллизия не устранена на некотором уровне, то она передается в семантическую сеть более высокого уровня. Семантические сети должны быть созданы таким образом, чтобы устранять коллизии разного типа, от коллизий связанных с расстановкой ударений до коллизий смысловых, разрешение которых позволит выбрать правильный просодический контур и выделить в произношении некоторое слово.

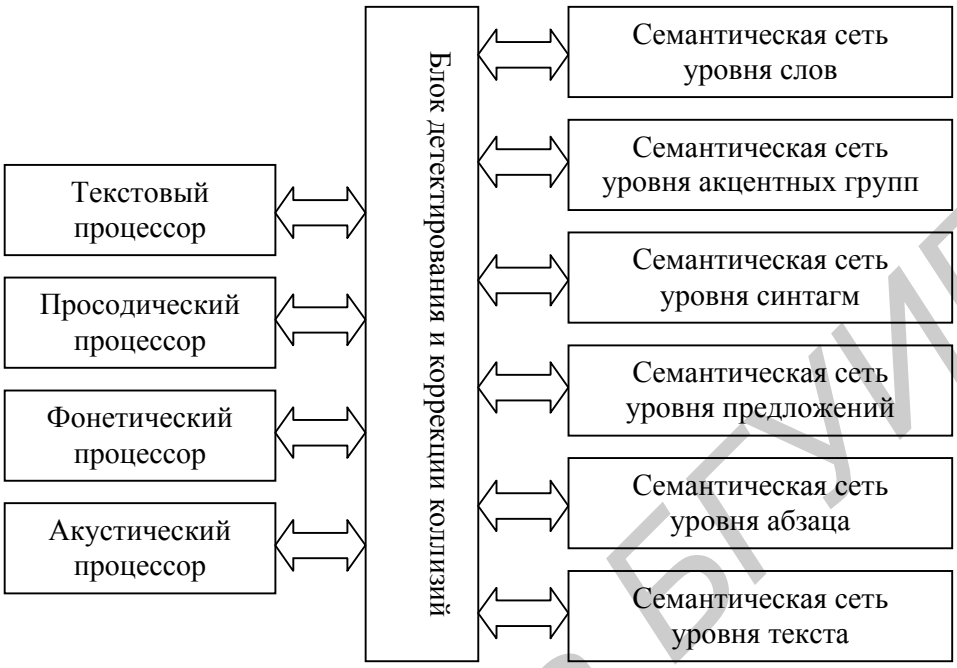

Рисунок 7 - Многоуровневый семантический анализ для синтеза речи

Совершенствование систем синтеза речи возможно только при помощи использования всесторонней и всеобъемлющей информации о языке, которая может быть органично внедрена в систему на основе предложенного выше подхода. Путем совершенствования и усложнения семантических сетей разного уровня возможно улучшение качества синтезированной речи, не изменяя при этом общую структуру и основные модули синтезатора речи, что позволяет привлекать к совершенствованию голосовых интерфейсов специалистов, не знакомых с обработкой сигналов и основными алгоритмами синтеза. Предложенный подход, основанный на детектировании и устранении коллизий на основе многоуровневых семантических сетей, позволяет сделать технологию синтеза речи открытой и легко модифицируемой.

Библиографический список

1. Лобанов Б.М., Цирульник Л.И. Компьютерный синтез и клонирование речи, Белорусская наука, С.344, 2008.

2. Лобанов Б..М., Цирульник Л.И., Сизонов О.Г Алгоритм интонационной разметки повествовательных предложений для синтеза речи по тексту // Тр. Международной конференции «Диалог'2008» "Компьютерная лингвистика и интеллектуальные технологии" – М.: Наука, 2008. – 7 с.