

СТРУКТУРА СИСТЕМ СИНТЕЗА И РАСПОЗНАВАНИЯ РЕЧИ

Р.В. Мещеряков

Томский государственный университет систем управления и радиоэлектроники

E-mail: office@security.tomsk.ru

Рассматриваются общие подходы к системам синтеза и распознавания речи. Предлагаются схемы процесса обработки информации в них как в двух изолированных системах, решающих прямую или обратную задачу. Показывается объединение двух подсистем синтеза и распознавания речи в единую диалоговую систему с новым качеством.

Ключевые слова:

Система, анализ речи, распознавание речи, обработка информации.

Key words:

System, speech analysis, speech recognition, information processing.

Введение

Реализация различных речевых систем решает вопросы создания алгоритмического и программного обеспечения обработки информации при оптимальном использовании ресурсов вычислительных систем. При этом целесообразно создание моделей систем обработки информации и их оценка.

На каждую из моделей систем обработки информации накладываются свои ограничения. Это обусловлено тем, что из поступающей информации выделяются характеристические элементы различных уровней. Далее происходит порождение элементов нового типа и их конфигурации, т. е. из нескольких элементов образуются конфигурации одного элемента высокого уровня или наоборот — происходит разбиение элемента высокого уровня на мелкие элементы более низкого уровня [1]. Остаются открытыми вопросы хранения в системе информации по правилам регулярности конфигураций и функционирования системы на различных уровнях. Рассмотрим обобщенные структуры систем для наиболее распространенных задач синтеза и распознавания речи, которые характеризуются схемами обработки информации в них.

Обобщенная модель синтеза речи

Внешние и внутренние данные, используемые для формирования речевого сигнала, отражены на рис. 1. В блоки левой части выделены все сведения, на основе которых строятся конфигурации и определяются признаки получаемых данных. В средней части идет информация по этапу преобразования и результат его выполнения (выделяются курсивом и через дефис приведен признак результата). Блоки правой части являются правилами языка и входят непосредственно в блоки левой и центральной части в виде таблиц, правил, условий регулярности, алгоритмов обработки. Данная схема может быть использована в качестве методологических основ систем синтеза речи.

Так как данная модель описывает прогнозируемые параметры модели генерации речевого сигнала,

в приведенной схеме отсутствуют некоторые обратные связи. Они будут добавлены на этапе непосредственного управления легкими и речеобразующим трактом в процессе генерации речевого сигнала.

На основе схемы, рис. 1, и полученных практических результатов можно дать рекомендации по организации баз знаний для языковых систем [2]:

1. Любая речевая система должна быть предварительно рассмотрена для выявления наиболее информативных блоков, влияющих на результат. По результатам анализа должны быть созданы критерии, разграничивающие информацию по уровням значимости. В частности, необходимо использовать критерий эффективности передаваемой информации [3].
2. Всю информацию, необходимую для преобразования, целесообразно разделить на базовые сведения (таблицы и словари) и правила, в соответствии с которыми производится преобразование.
3. Для неформализованной информации, которая не может быть представлена в виде правил, используются словари. Размер словарей ограничивается требуемым качеством и допустимым количеством объектов.
4. Для информации, однозначно определяемой объектом, целесообразно использовать таблицы, в соответствии с которыми по требуемому объекту определяется его признак. Например, таблица длительностей.
5. Общие параметры речевого сигнала должны основываться на физиологических параметрах речеобразующей системы человека. Это позволяет настроить систему синтеза речи на диктора и получить естественную речь.
6. Введение критериев эффективности необходимо как на различных промежуточных этапах, так и на результирующем. Их задание может быть как в явном виде, так и в неявном виде в виде ограничений. Таким образом, получаем систему управления процессом синтеза речи.

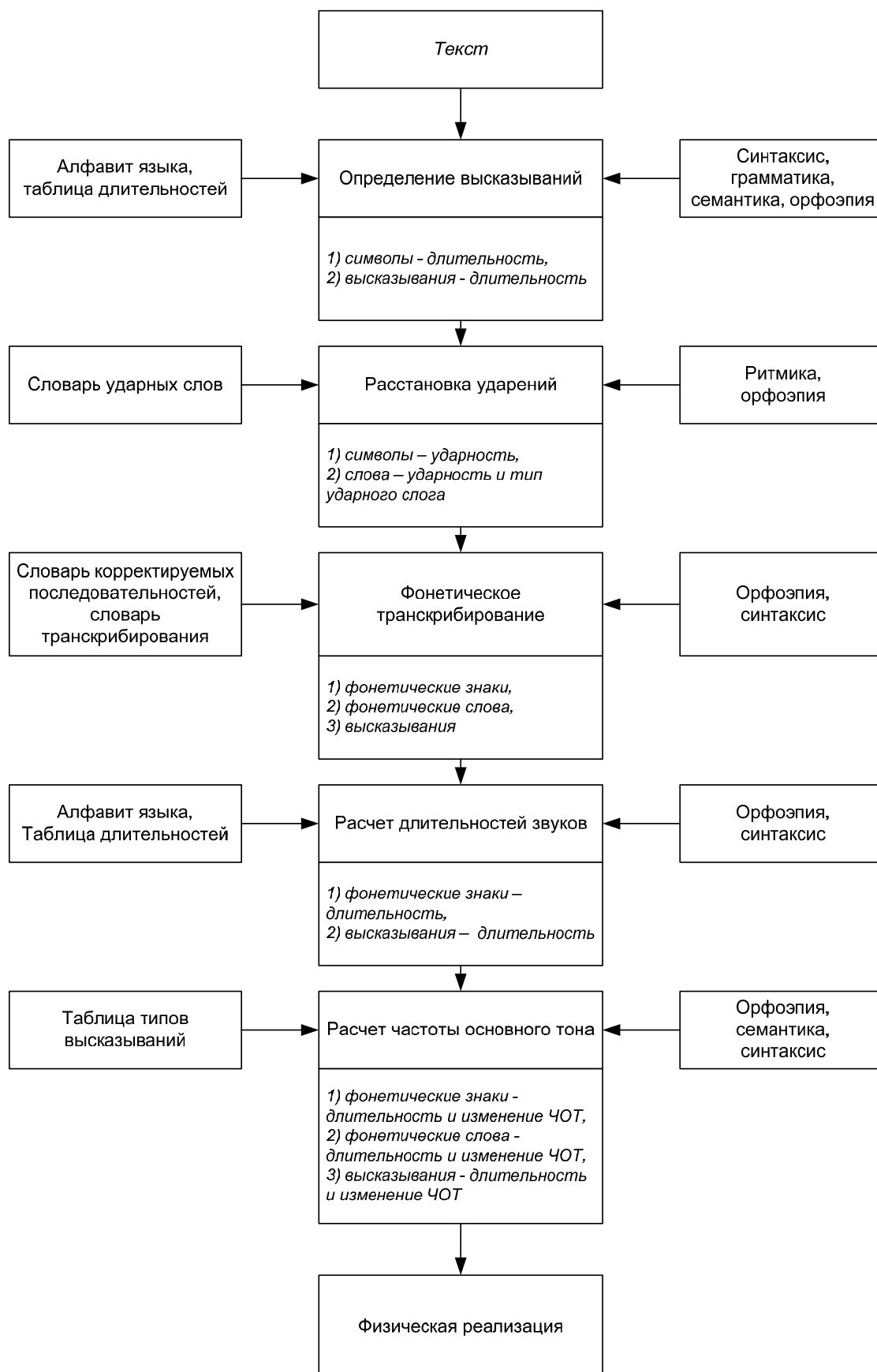


Рис. 1. Схема построения модели управления

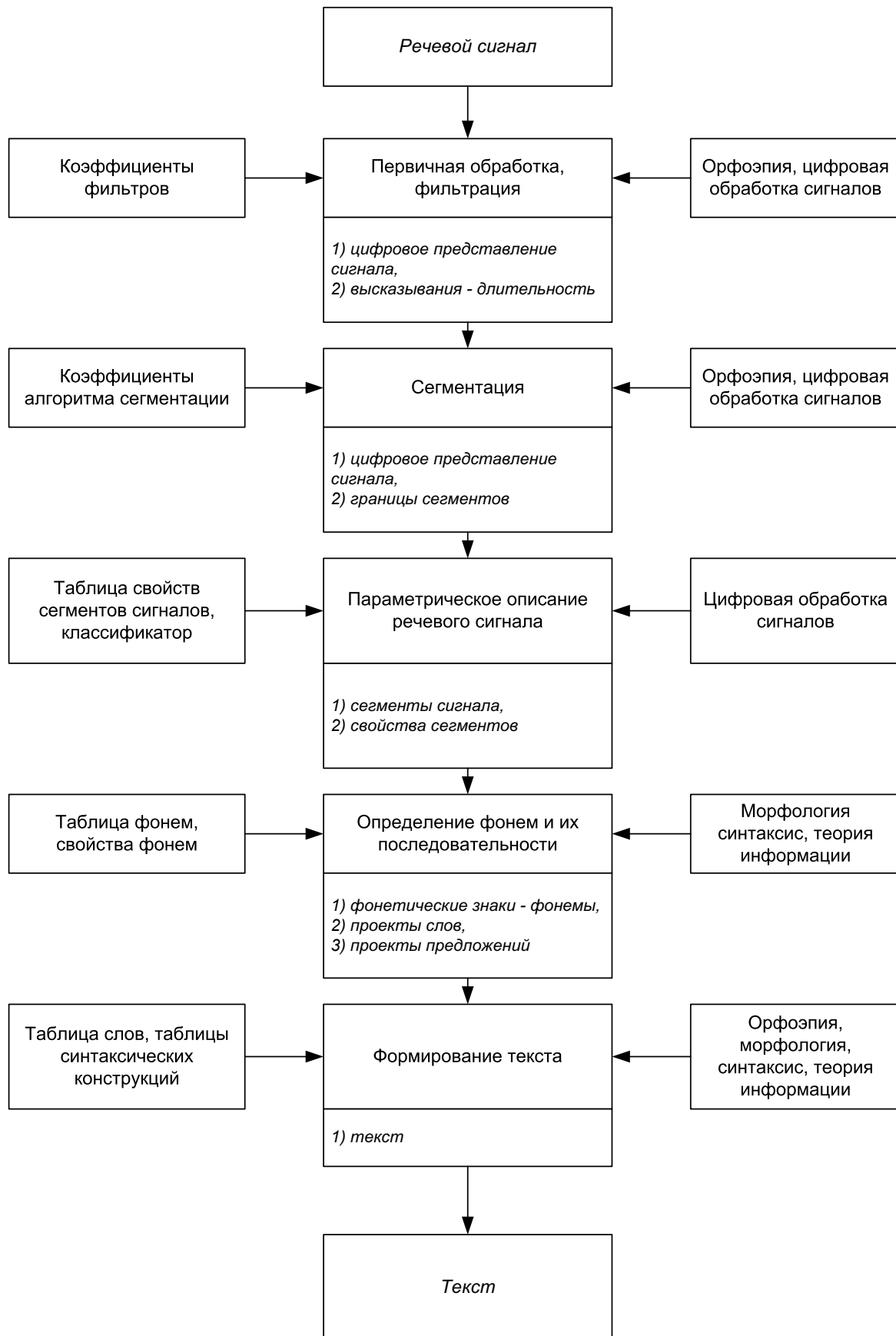


Рис. 2. Схема распознавания речевого сигнала для получения печатного текста

По результатам исследования блоков генерации речевого сигнала были выделены наиболее информативные. В них вся необходимая информация для формирования речевого сигнала была собрана в виде правил. По требуемым входным данным для реализации правил сформированы таблицы, основой которых являлись физиологические и информационные сведения, приведенных в [4–9]. Результаты работы алгоритма синтеза показали адекватность созданной модели данным, приведенным в работах [8, 10, 11].

Обобщенная модель распознавания речи

Рассмотрение процесса анализа речевого сигнала показало необходимость введения обратных связей [8]. Очевидно, обобщенная модель распознавания речевого сигнала подобна синтезу речи с той лишь разницей, что движение информации идет в обратном направлении.

Для реализации базовых алгоритмов необходимо сформировать прогнозируемое получение смысла и цели принимаемого речевого сигнала. Таким образом, необходимо отобразить схему получения смысла речевого сообщения.

На рис. 2 отражены все внешние и внутренние данные, используемые для формирования просодических характеристик. В блоки левой части вы-

делены все сведения, на основе которых строятся конфигурации и определяются признаки. В средней имеются две составляющие: этап преобразования и выделенный курсивом результат его выполнения. Блоки правой части являются правилами языка и входят непосредственно в блоки левой и центральной части в виде таблиц, правил, условий регулярности, алгоритмов обработки. Данная схема может быть использована в качестве методологических основ систем синтеза речи.

Приведенная модель описывает получение печатного текста, однако в этой схеме отсутствует собственно получение семантики и прагматики сообщения и диалога в частности. При этом целесообразно рассматривать полную систему диалога при человеко-машинном взаимодействии.

Обобщенная модель речевого диалога

Приведенные выше схемы используют только прямой канал преобразования, показанные непрерывными линиями на рис. 3. Прерывистыми стрелками показаны предполагаемые для наполнения связи для согласования формирования и разбора сообщения.

Для повышения эффективности работы системы преобразования информации предложена следующая система синтеза и распознавания речи, рис. 4.

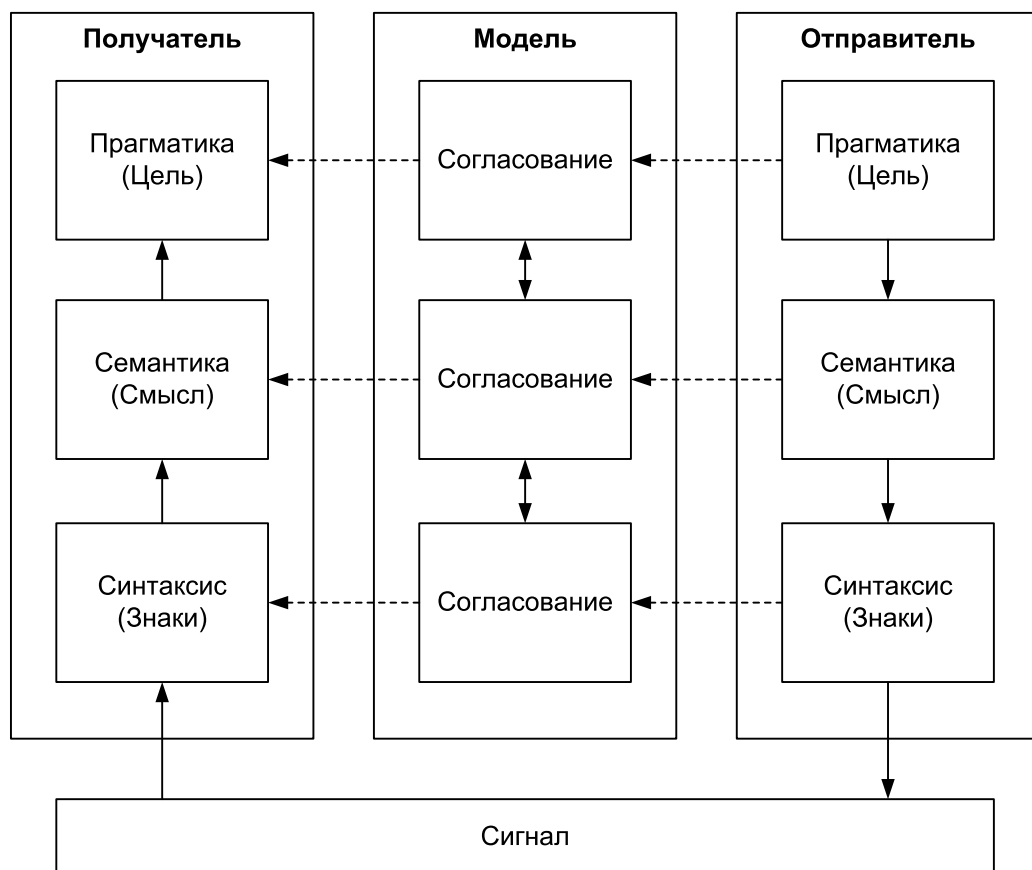


Рис. 3. Упрощенная схема преобразования информации

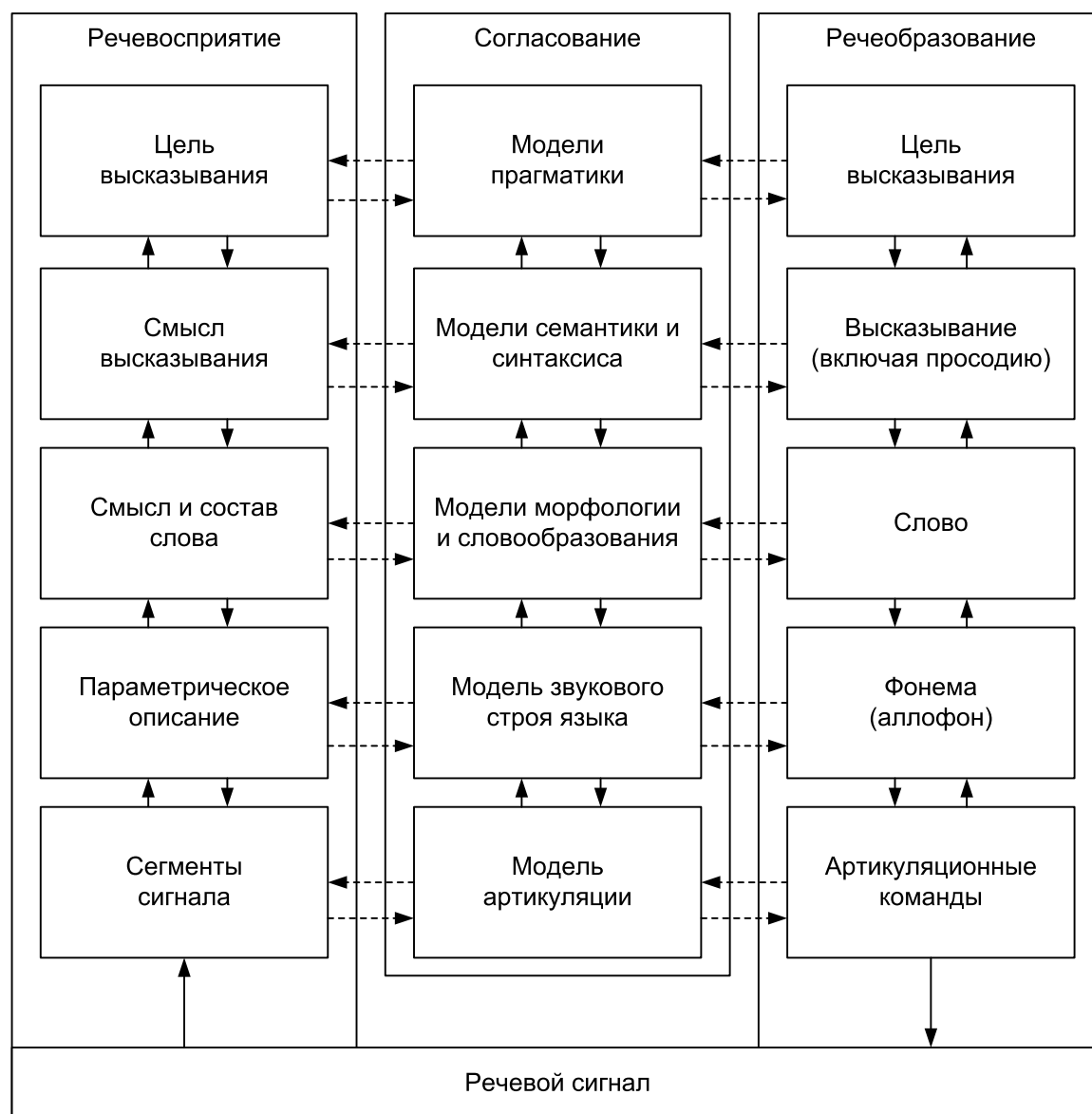


Рис. 4. Обобщенная схема речеобразования и речевосприятия

Таким образом, предложенная обобщенная схема учитывает различные априорные знания: языковые (фонетика, лексика, синтаксис, семантика, просодия и т. д.) и неязыковые, т. е. знания предметной области диалога. Внешними входными и выходными данными речевых систем являются: смысловое пространство слов и фраз заданного языка и предметной области, речевой сигнал, а также для систем синтеза — параметры речеобразующей системы.

Отметим, что восприятие речи человеком состоит из нескольких этапов обработки поступающей информации. Эти этапы можно представить в виде иерархии уровней (рис. 4). Каждый из этих уровней имеет свой набор данных и правил, обеспечивающих обработку информации. Соответственно, для решения задач, стоящих перед верхними уровнями (распознавание фраз, текста), необходимо решение задач нижних уровней.

При речеобразовании, наоборот, для решения задач нижнего уровня необходимо решение задач верхних уровней. Взаимодействие систем речевосприятия и речеобразования обеспечивается за счет обратных связей между этими системами, которые существуют на каждом уровне иерархии. Например, при синтезе речи требуется непрерывно подстраиваться к изменяемым характеристикам разборчивости речевого сигнала. Для этого вводятся обратные связи с использованием системы восприятия речевого сигнала, на основе которых корректируется синтез [12].

Каждый выделенный объект системы определяется своим набором сведений о языке, правилах преобразований и связей с другими уровнями. В схеме модели речевой системы отражены непосредственные взаимосвязи только двух уровней: вышестоящего и нижестоящего, в реальности же

существует больше взаимосвязей. Так, на вышестоящих уровнях большое значение имеют знания о модели конкретного языка, то есть о звуковом строе, морфологии, синтаксисе, а также сведения о текущей предметной области диалога. На нижестоящих уровнях эти знания теряют свое значение.

На нижних уровнях основное значение приобретают знания о речевом сигнале, которые являются общими для любой модели языка. Эта независимость от различий в языках обусловлена одинаковыми принципами строения речеобразующего тракта и слуховой системы разных людей.

Получаем, что для решения задач распознавания на нижнем уровне необходимо использовать сведения по структуре и работе слуховой системы человека, закономерностям образования различных звуков, фонетическому составу различных языков (для выявления общих классов фонем) и параметрическому описанию речевых сигналов. Предлагаемые сведения наполняют модели, рис. 1–4.

Например, после транскрибирования текста символы-буквы не используются. Раздельные базы данных используются лишь на этапах, ограничивающих их применение. Так, на этапе оценки длительности используется около 5 % всей информации, расстановки ударений – 20 % и т. д.

На рассматриваемом этапе обработки текста используются, в основном, закономерности, присутствующие в языке. Параметры диктора (объем легких, параметры участков речеобразующего тракта, пол, параметры артикуляционных органов, темп речи и т. п.) учитываются только на уровне средних значений и ограничений. Очевидно, что учитываемые особенности можно представить и использовать в виде баз данных:

- 1) длительностей звуков (количество звуков определяется языком, их конкретные значения – диктором);

- 2) слов, для которых расставляем ударения (количество слов определяется языком (для покрытия 70 % текстов), но возможны отличия для конкретного диктора или говорящего);
- 3) транскрибирования (определяется полностью языком, хотя могут быть некоторые вариации для говорящих и конкретного диктора);
- 4) мелодик типовых высказываний (формируется для языка и учитывает индивидуальные особенности диктора (параметры легких, постоянные времени, ограничения по частоте основного тона, средние значения)).

Разработанные модели и схемы синтеза и распознавания речи реализованы в виде программного обеспечения. Наибольшую эффективность они показали при реализации реабилитационного комплекса по восстановлению речи после резекции гортани [14]. Учет характеристик процесса голосообразования позволил повысить качество исследования голоса пациентов и сократить время реабилитационного периода по сравнению с работой логопеда в 2...3 раза.

Выводы

Рассмотрены системы синтеза и распознавания речи с позиций теории иерархических многоуровневых сложных систем. Предложены алгоритмы обработки речевой информации в виде двух изолированных систем, что позволяет решать как прямую, так и обратную задачу синтеза и распознавания речевого сигнала, отличающихся от аналогов разделением на информацию, зависящую от диктора и от языка. Предложено объединение двух подсистем синтеза и распознавания речи в единую диалоговую систему с новым качеством. Показаны перспективы реализации алгоритмов синтеза и распознавания речи в виде единого программно-аппаратного комплекса, использующего диалоговую структуру.

СПИСОК ЛИТЕРАТУРЫ

1. Гренандер У. Лекции по теории образов: Т. 1. Синтез образов / под ред. Ю.И. Журавлева; пер. с англ. – М.: Мир, 1979. – 383 с.
2. Мещеряков Р.В., Бондаренко В.П., Организация баз знаний в системе синтеза речи // Теория и практика речевых исследований (АРСО-99). Матер. конф. – М.: МГУ им. М.В. Ломоносова, 1999. – С. 37–38.
3. Флейшман Б.С. Элементы теории потенциальной эффективности сложных систем. – М.: Советское радио, 1971. – 223 с.
4. Златоустова Л.В., Потапова Р.К., Трунин-Донской В.Н. Общая и прикладная фонетика. – М.: Изд-во МГУ, 1986. – 304 с.
5. Математическая лингвистика. Сб. переводов / под ред. Ю.А. Шрейдера, И.И. Ревзина, Д.Г. Лахути, В.К. Финна. – М.: Мир, 1964. – 144 с.
6. Филичева Т.Б., Чевелева Н.А., Чиркина Г.В. Основы логопедии. – М.: Просвещение, 1989. – 223 с.
7. Потапова Р.К. Речь: коммуникация, информация, кибернетика. – М.: Радио и связь, 1997. – 528 с.
8. Сорокин В.Н. Синтез речи. – М.: Наука, 1992. – 392 с.
9. Трунин-Донской В.Н. Автоматический синтез звучащего текста // Звучащий текст. – М.: Институт научной информации по общественным наукам, 1983. – С. 218–250.
10. Потапова Р.К. Речевое управление роботом. – М.: Радио и связь, 1989. – 246 с.
11. Дмитриев Л.Б., Теляева Л.М., Таптапова С.Л., Ермакова И.И. Фонология и фонопедия. – М.: Медицина, 1990. – 272 с.
12. Бондаренко В.П., Мещеряков Р.В. Диалог как основа построения речевых систем // Кибернетика и системный анализ. – 2008. – № 2. – С. 30–41.
13. Бондаренко В.П., Квасов А.Н., Конев А.А., Мещеряков Р.В., Чойнзонов Е.Л., Чижевская С.Ю. Программные средства комплекса исследования речевого сигнала при злокачественных заболеваниях гортани // Медицинская техника. – 2009. – № 4. – С. 33–37.

Поступила 29.10.2009 г.