

**ДИАЛОГ КАК ОСНОВА ПОСТРОЕНИЯ РЕЧЕВЫХ СИСТЕМ**

**Ключевые слова:** *диалог, структура, иерархия, качество.*

**ВВЕДЕНИЕ**

К интеллектуальным системам относятся такие речевые системы, как системы синтеза и распознавания речи, диалоговые системы, системы машинного перевода и др. Они должны, во-первых, понимать речь, несмотря на возможные помехи, нечеткость дикции, отклонения от нормативного синтаксиса, эллиптические конструкции и т.п., и, во-вторых, обеспечивать качественную речь, имеющую высокую разборчивость и естественность речевого сигнала. Человек интегрально использует как языковые знания (фонетика, лексика, синтаксис, семантика, просодия и т.д.), так и неязыковые (предметная область диалога). Основной психоакустической характеристикой речи является ее разборчивость, т.е. степень правильного восприятия звуков, слов и смысла. Максимальная разборчивость, характерная для восприятия связной речи, это фразовая разборчивость [1]. Если человек воспринимает изолированные слова, процент разборчивости оказывается меньше. Он еще больше снижается при восприятии изолированных фонетических элементов речи типа слогов.

Большинство подходов основано на последовательном выполнении распознавания слов и синтаксического разбора. Опыт лингвистов показывает, что человек вначале постигает смысл фразы и только затем приступает непосредственно к задаче понимания речи, которая часто рассматривается как процедура, не зависящая от этапа распознавания. Использование подобных подходов приводит к слабым результатам. Из теории речевосприятия [1] становится очевидным, что распознавание и понимание — это две тесно связанные процедуры, которые интегрированы в процесс диалога.

**ОСНОВНЫЕ ПОЛОЖЕНИЯ КОНЦЕПТУАЛЬНОЙ МОДЕЛИ СИСТЕМ РЕЧЕВОГО ДИАЛОГА**

Внешними входными и выходными данными систем речевого диалога являются смысловое пространство слов и фраз заданного языка и предметной области, речевой сигнал, а также параметры речеобразующей системы (для канала синтеза). Речевая система имеет следующие иерархические уровни: прагматический, семантический, синтаксический, фонетический, физический. Каждый уровень имеет свой набор данных и правил, обеспечивающих обработку информации. Соответственно для достижения целей обработки информации на верхних уровнях при восприятии речи необходимо решение задач нижних уровней и наоборот. Подобное возникает в системах синтеза речи — для формирования и произнесения высказывания необходимо решение задач верхних уровней, а также в других речевых системах. Эти два процесса настолько взаимосвязаны, что при решении прямой задачи требуется решение обратной задачи.

Таким образом, основой диалоговой системы может служить многоуровневая иерархическая модель [2], предполагающая:

— полное описание языка как иерархию тесно интегрированных событий с определением соответствующих правил преобразования информации на соответствующих уровнях;

— учет метаязыковых знаний — предметной области диалога, т.е. прогноз развития событий, как особенность восприятия речи.

В результате основные положения при построении модели диалоговой системы можно свести к следующим:

— модели мира субъектов, ведущих диалог, пересекаются, т.е. для них частично общие знания о мире, определяемые конкретной предметной областью диалога [3];

— субъекты всегда (с учетом предыстории диалога) с определенной вероятностью предсказывают речевую реакцию оппонента.

Следовательно, эффективная процедура восприятия и синтеза речи должна включать «полные» знания о языке, параметры внешней среды, прогноз реакции оппонента.

Эти три положения позволяют ограничить область допустимых решений в каждый текущий момент диалога практически на всех уровнях иерархии и повысить надежность восприятия речи и качество ее синтеза.

### КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ ДИАЛОГА

Исходя из требований, изложенных в предыдущем разделе, упрощенную модель системы диалога представим схемой (рис. 1), где стрелками указаны направления движения потоков информации. Движение потоков снизу вверх характеризует канал восприятия (распознавания), а сверху вниз — канал синтеза. Каждый выделенный объект определяется своим набором сведений о языке, правилах преобразований и связях с другими уровнями. В схеме отражены непосредственные взаимосвязи только двух уровней: вышестоящего и нижестоящего. В действительности существует больше взаимосвязей. Так, на вышестоящих уровнях большее значение имеют языковые знания, в частности сведения о текущей предметной области. Отметим, что на нижестоящих уровнях эти знания о языке теряют свое значение.

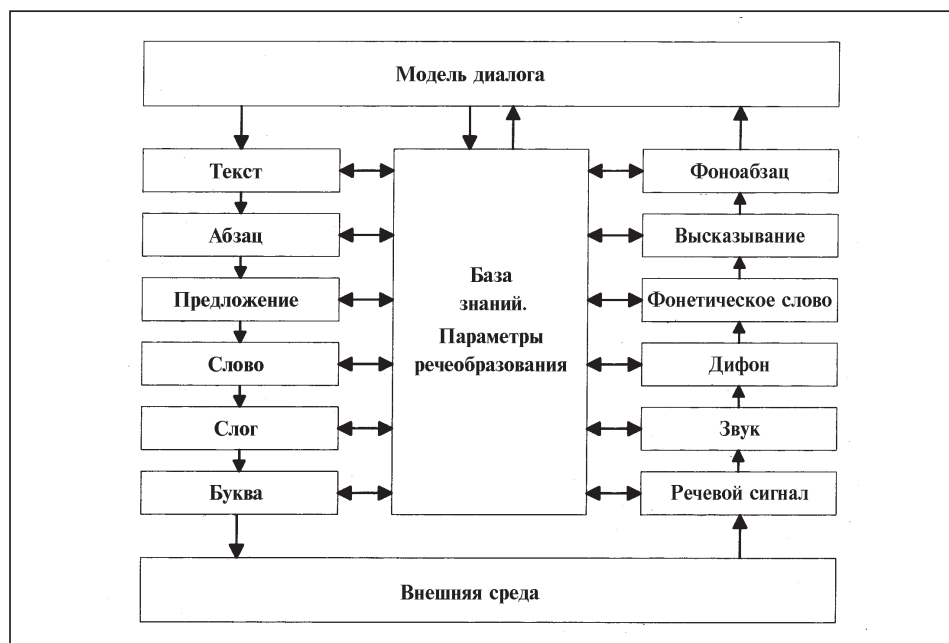


Рис. 1. Упрощенная модель системы диалога

Рассмотрим взаимное влияние верхних и нижних уровней. Очевидно, что смысл предложения накладывает ограничения на состав слов, содержащихся в предложении, а также на их синтаксические связи. Бесспорно, что и синтаксические связи слов в предложении определяют смысл высказывания и его целевую функцию. Известно [4], что буквы в тексте имеют различную вероятность появления, а вероятность появления следующей буквы зависит от предшествующих букв. Однако уже при трех и более буквенных сочетаниях вероятность появления буквы снижается. Аналогичные рассуждения имеют место и для слов в предложениях. Таким образом, текущий прогноз на уровне букв или слов позволяет снизить области принятия решений при распознавании.

Представляет определенный интерес влияние результатов прогноза развития диалога на уровне модели мира, например на лексическую базу области допустимых решений на уровне фонетических слов. Известно, что для бытового общения достаточно 100–200 наиболее часто употребляемых слов. Естественно, что любая другая специальная предметная область потребует включения своих терминов и понятий, т.е. новых слов, что приведет к расширению лексической базы. Оказывается, что общая лексическая база увеличивается незначительно. С этой целью на основе алгоритмов морфологического и синтаксического анализа, рассмотренных в работе [5], была проведена оценка лексической базы для таких предметных областей, как компьютерная техника, радиоэлектроника. Оказалось, что наиболее часто встречаемые слова в этих областях увеличивают лексическую базу на 30–50%. Отметим, что если ведется текущий прогноз диалога, то область допустимых решений даже на этой лексической базе будет значительно ограничена, что должно привести к повышению надежности распознавания на уровне слов.

Приведенная схема диалоговой системы позволяет провести предварительную классификацию речевых систем:

- системы восприятия речи, реализующие полностью канал обработки от речевого сигнала до модели мира;
- системы полного синтеза речи, реализующие канал от модели мира до речевого сигнала;
- системы распознавания слов и фраз;
- системы синтеза по печатному тексту и др.

В настоящее время в той или иной степени реализованы системы распознавания слов и фраз, а также системы синтеза по печатному тексту, сопоставительный анализ которых приведен в [6].

Отметим, что системы, использующие только ограниченное количество уровней, например только фонемный, имеют невысокую надежность распознавания. При использовании информации по слоговому распознаванию надежность повышается. Аналогичное происходит при синтезе речи: при непосредственном преобразовании буквы в звук качество и разборчивость речевого сигнала будут невысокими. При учете влияния более высоких уровней качество синтезированной речи повышается.

Существуют взаимосвязи не только уровней одного канала, но и каналов между собой. Например, при синтезе речи по артикуляторной модели [7] требуется постоянно подстраиваться по изменяемым характеристикам разборчивости сигнала. Для этого вводятся обратные связи с использованием канала распознавания речевого сигнала, на основе которых корректируется синтез.

Таким образом, на основании вышесказанного сделаем заключение, что существующие обратные связи определяются уровнем иерархии в системе обработки информации и степенью необходимости языковых знаний.

Необходимость обратных связей может быть показана на примере системы речеобразования. Человек, произнося речь, корректирует ее через обратные свя-

зи: во-первых, через слуховую систему: по воздуху, по скелету; во-вторых, с учетом местоположения органов речеобразования; в третьих, фиксируя реакцию собеседника по зрительным каналам. В соответствии с получаемой информацией человек вносит коррективы в произносимую речь.

#### СТРУКТУРА УРОВНЕЙ ДИАЛОГА

Диалог предполагает взаимодействие, как минимум, двух подсистем, структуры которых в общем случае подобны и модели мира пересекаются. Каждая подсистема может быть представлена в виде

$$A_i = (A_{ic}, A_{in}), \quad (1)$$

где  $A_{ic}$ ,  $A_{in}$  — структура и поведение подсистемы. Диалоговая система в этом случае будет представлена как объединение подсистем (1)

$$A = \bigcup_i A_i = \bigcup_i (A_{ic}, A_{in}). \quad (2)$$

Каждая подсистема в процессе диалога будет стремиться реализовать свою цель  $A_{in}$ , поэтому можно считать, что структура и поведение систем  $A_i$  могут не совпадать. Реализация частной цели  $A_{in}$  предполагает соответствующую структуру и поведение подсистемы  $A_i$ , которые можно (в некотором смысле) представить в виде оптимального обмена [8]. Ясно, что это требует определения, с одной стороны, функционала качества, с другой — области допустимых решений. Наиболее трудно определить функционал качества для диалога в целом, однако его можно декомпозировать по уровням иерархии, представив в виде иерархии качеств; следовательно, должна возникнуть иерархия областей допустимых решений.

Основную цель определяет прагматическая составляющая диалога, которая на

более высоком уровне иерархии определяется глобальным смыслом диалога. На более низких уровнях глобальный смысл представлен своими частями (конкретным смыслом).

На рис. 2 изображены две подсистемы:  $A_1$  и  $A_2$ , участвующие в диалоге через внешнюю среду (ВС). Уровни  $K_{11}$  и  $K_{21}$  подобны, так как реализуют подобные функции: формирование и передачу сообщений. Уровни также являются подсистемами. Уровни  $K_{12}$  и  $K_{22}$ , реализующие функции приемника, также подобны. Таким образом, функции субъектов диалога должны быть подобны. Можно предположить, что при формировании сообщений используется часть информации по каналу восприятия. Таким

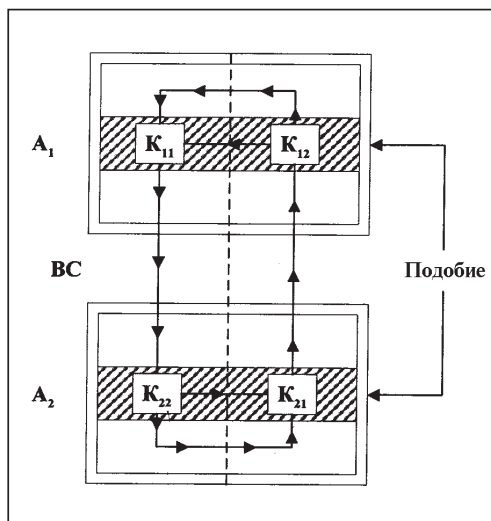


Рис. 2. Взаимодействие подсистем при диалоге

образом, источник сообщения учитывает возможности приемника сообщения. В данном случае можно сказать, что при формировании и передаче сообщения приемник стремится достичь не только цели оптимального кодирования, но и достаточного уровня надежности распознавания. Для этого используются имеющиеся в его системе сведения о возможностях приемника; некоторые его параметры он может использовать, располагая структурой и поведением своего восприятия.

Исходя из этого можно предположить, что на каждом уровне имеются свои определения и понятия, например модели грамматики, фонетики и т.п. Эти модели по аналогии с моделями мира должны пересекаться для различных подсистем, участвующих в диалоге.

Будем считать, что ресурсы системы, например сложность, ограничены. В частности, при генерации высказывания у человека ограничено количество воздуха в легких, быстрдействие артикуляторных органов и т.п. Это приводит к необходимости передать сообщение с достаточной степенью надежности восприятия с ограничением на ресурсы, что позволяет говорить об оптимальном обмене в смысле Флейшмана [8].

В свою очередь, подсистемы генерации и распознавания следует рассматривать как иерархически построенные в соответствии с рис. 1. Поскольку каждый уровень иерархии диалоговой системы выделяется как самостоятельный, необходимо найти целевую функцию его возникновения и определить, каким образом проводится его оптимизация при диалоге. Это рассматривается как достижение некоторого качества системы при заданных ограничениях.

Использование данного подхода предполагает учет иерархии качеств сложной системы. Рассматриваемый тип систем относится к классу решающих систем, которые должны обладать следующей иерархией [8]:

— *O*-качество — система, в которой присутствует обратная связь; автоматическая система; система, обладающая свойствами устойчивости, точности и др.;

— *R*-качество — надежная система, обладающая свойством структурной устойчивости;

— *I*-качество — помехоустойчивая система, включающая свойство кодирования и декодирования;

— *C*-качество — управляемая система.

Роль качества изменяется с ростом уровня иерархии диалоговой системы. Если на уровне сигнала, звука, дифтонга основную роль играют *O*-, *K*-, *I*-качества, то на уровнях высказывания, фоноабзаца существенное значение, кроме этих качеств, приобретает *C*-качество.

#### **ОБОБЩЕННАЯ СТРУКТУРА ПРЕДСТАВЛЕНИЯ ИНФОРМАЦИИ В ДИАЛОГОВОЙ СИСТЕМЕ**

Диалог двух подсистем ведется элементарными сигналами через внешнюю среду, которая вносит искажения и помехи в эти сигналы. Ограничения на сложность системы в целом приводит к ограничению сложности уровней. Следовательно, в ограниченный промежуток времени уровень может обработать конечный объем информации. Для этого требуется использовать иерархическое представление, при котором объем информации, требуемый для описания входных и выходных сигналов в единицу времени, сокращается при движении по каналу распознавания и увеличивается при движении по обратному каналу синтеза. Рассматриваемым свойством, связанным с необходимостью представления информации в сложной системе с разной степенью абстрагирования, обладают все сложные иерархические системы, что является следствием ограничений на структурную сложность любых систем (1). Данное свойство можно рассматривать как особое *M*-качество сложных иерархических систем.

Любой уровень иерархии характеризуется парой преобразований, которые описывают прямой и обратный каналы взаимодействия в данной системе:

$$\begin{aligned} s: X &\rightarrow Y, \\ s^* : Y^* \times Z &\rightarrow X^*, \end{aligned} \quad (3)$$

где  $X$  и  $Y$  — входное и выходное множества системы снизу, т.е. для канала восприятия;  $Y^* \times Z$  и  $X^*$  — вход и выход системы сверху, т.е. для канала синтеза;  $Z$  — множество трансформаций, задаваемых вышестоящими уровнями.

Объемы описания  $\mu(X)$  и  $\mu(Y)$  должны находиться в соотношении

$$\mu(X) > \mu(Y). \quad (4)$$

Это возможно, если на множестве  $X$  устанавливается отношение эквивалентности, т.е. разбиение на классы  $X_Y \in X$ . Следовательно, описание  $Y$ , а возможно, и  $X_Y$ , должно содержать как минимум две компоненты: имя класса и перечисление или правило генерации элементов  $x$ , принадлежащих данному классу  $X_Y$ . Но если элементы  $x$  каким-либо образом сгенерированы, то они могут храниться на соответствующем уровне иерархии в виде некоторой базы знаний (модели), позволяющей установить соответствие между именем класса и его содержанием. Тогда  $Z$  будет характеризовать допустимые преобразования подмножества  $X_Y$ , не выходя из данного класса. Если множество  $Z$  гомеоморфно единичному отрезку, то отображение  $s^*$  становится гомотопией, что равносильно разбиению входного множества на гомотопические классы эквивалентности. Тогда при

$$z \in [0, 1] \text{ и } z = 0 \quad (5)$$

должно выполняться условие

$$\delta(X, X^*) \leq \delta_0, \quad (6)$$

где  $\delta$  — погрешность синтезированного сигнала  $X^*$  по описанию множества  $X$  в случае отсутствия его трансформаций;  $\delta_0$  — заданная допустимая погрешность.

Наиболее ярко отображение (3) прослеживается на таких понятиях, как фонема и аллофон. С учетом этого в описание элемента  $Y$  необходимо включить дополнительную составляющую, характеризующую ее состояния, а также составляющую, характеризующую положение  $X_Y$  в множестве  $X$ . Это приводит к форме описания элемента  $Y$  (называемого образующей), предложенного в работе [9]:

$$O(Y) = \{N_Y, P_Y, S_Y\}. \quad (7)$$

Здесь  $N_Y$  — имя класса,  $P_Y$  — признаки,  $S_Y$  — связи  $X_Y$  с другими классами эквивалентности множества  $X$ .

Представление входного множества в виде отображения (3), удовлетворяющего условиям (4)–(6), на уровне сигналов не вызывает сомнений. В частности, оценка (7) может быть получена путем вычисления расстояния. Задача резко усложняется на более высоких уровнях, когда процедура оценки (6) становится неочевидной.

В рассмотренной постановке каждый уровень характеризуется некоторым переменным количеством элементов (7), зависящим от конкретного речевого сообщения. Элемент, как отмечалось ранее, может иметь связи с другими элементами выделенного уровня иерархии. В отдельных случаях структуры конфигурации элементов обретают свойство регулярности [9], что позволяет качественно переходить на более высокий уровень иерархии, принимая за элемент регулярную структуру нижестоящего уровня. Аналогичное происходит и в отношении обратного канала: элемент разбивается на несколько элементов низкого уровня, каждый из которых имеет наименование, признаки и связи. Однако ключевая информация может иметь большое значение на всех уровнях диалоговой системы. Примером такой информации является частота основного тона речевого сигнала, которая требуется на разных уровнях и сохраняется в признаках элементов-образующих.



Предположим, что при использовании  $M$ -качества количество объектов, одновременно обрабатываемое каждым уровнем, приблизительно одинаково [10]. Однако объекты имеют различную природу и различные свойства. Очевидно, что представление информации на физическом уровне отличается от ее описания на уровне слов. В частности, информация на уровне сигналов представлена в шкалах интервалов, а на уровне слов — в шкалах наименования и порядка.

При использовании подхода, изложенного в работе [9], несколько регулярных структур текущего уровня (точнее, одного класса) со своими связями и признаками преобразуются в образующие более высокого уровня или в регулярные структуры более низкого уровня. Возникают случаи, когда для получения другой регулярной структуры требуется использование более одного уровня иерархии диалога. Примером может служить влияние уровня предложений на фонетический состав звука (произнесение в начале, в конце предложения). Для перехода от одного уровня к другому недостаточно только определить правила выделения образующих каждого уровня. Необходимо задать алгебру на множестве выделенных элементов каждого уровня.

Для выделения элементарных объектов физического уровня (уровня звуков) наиболее часто используют предварительную сегментацию на квазиоднородные участки звукового сигнала с последующей их классификацией. Будем считать, что множество разнородных сегментов определяется как  $S = \{s_i\}$ ,  $i = 1, 2, \dots, I$ , где  $I$  — конечно. Для каждого выделенного объекта можно определить свое тело кластера  $K_i$ . С использованием какой-либо меры близости [11] в данном пространстве можно и нужно определить образ, в который попадает искомым сегмент речевого сигнала. Очевидно, что на данном уровне может происходить сравнение в шкалах интервалов и в качестве меры близости использоваться расстояние.

Очевидно, что на множестве сегментов  $s_i$  можно определить отношение эквивалентности  $R = \sim$ . Два сегмента,  $s_1$  и  $s_2$ , будут эквивалентными, если принадлежат одному кластеру/классу в пространстве состояний, т.е.  $s_1 \in K$  и  $s_2 \in K$ . Следовательно, все сегменты, расположенные в одном кластере, принадлежат одному и тому же классу эквивалентности  $[K_{\sim}]$ . Заметим, что по определению отношение эквивалентности будет рефлексивным, транзитивным, симметричным. При этом признаки  $P_1$  и  $P_2$  будут определять различие объектов  $s_1$  и  $s_2$ , попадающих в один класс эквивалентности. Таким образом, отношение эквивалентности позволяет отобразить на множество наименований  $M_N = \{m_i\}$  классы эквивалентности:  $m: S \rightarrow M_N$ , где  $m_i$  — имя класса эквивалентности  $[K_{\sim i}]$ . Можно определить свойства образующих, описывающих характеристики пространства состояний, т.е. признаки.

Следующий уровень описания речевого сигнала — параметрический, его входом является множество элементов  $M_N = \{m_i\}$ . Введем отображение на множестве:  $r: M_N \rightarrow R_N$ , где  $r$  — бинарное отношение связности элементов  $M_N$ . Данные последовательности могут образовывать регулярные конфигурации [9], ограниченные фонемой, дифтонгом, фонетическим словом.

От длительности цепочки зависит необходимость учета связи не только двух последовательно стоящих образующих, но и всех находящихся в данной конфигурации, т.е. арность образующих увеличивается от конфигураций малых длительностей к конфигурациям большой длительности. Построение конфигурации будет происходить как действие над свойствами образующих. Можно определить следующую функцию  $p(m_1, m_2): M_N^2 \rightarrow [0..1]$ , где  $p$  — вероятность (частотность) сочетаемости объектов  $m_1$  и  $m_2$  между собой; точнее, за объектом  $m_1$  должен следовать объект  $m_2$ . Значение 0

соответствует несочетаемости объектов. Значение 1 соответствует обязательной сочетаемости объектов между собой, т.е. всегда за объектом  $m_1$  следует объект  $m_2$ .

#### О-КАЧЕСТВО ДИАЛОГОВЫХ СИСТЕМ

Речевая система в общем случае рассматривается как диалоговая иерархическая многоуровневая система и представляется цепочкой следующих отображений:

$$\begin{array}{cccccccccccc}
 \varphi_n & & \varphi_3 & & \varphi_2 & & \varphi_1 & & \psi_1' & & \psi_2' & & \psi_3' & & \psi_n' \\
 \rightarrow & \dots & \rightarrow & F_2 & \rightarrow & F_1 & \rightarrow & S & \rightarrow & G_1' & \rightarrow & G_2' & \rightarrow & \dots & \rightarrow \\
 M_1 & & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & & M_2. \\
 \subset & \dots & \subset & G_2 & \subset & G_1 & \subset & S^* & \subset & F_1' & \subset & F_2' & \subset & \dots & \subset \\
 \psi_n & & \psi_3 & & \psi_2 & & \psi_1 & & \varphi_1' & & \varphi_2' & & \varphi_3' & & \varphi_n'
 \end{array} \quad (8)$$

Здесь  $F_i, F_i', G_i, G_i'$  — описание высказываний на разных уровнях иерархии в системах речеобразования и речевосприятия;  $\varphi_i', \varphi_i, \psi_i', \psi_i$  — отображение описаний;  $S, S^*$  — речевые сигналы;  $M_1, M_2$  — модели мира участников диалога. Стрелки между описаниями  $F_i$  и  $G_i$  характеризуют взаимосвязи между ними. Реально этими стрелками представлены соответствующие модели преобразования.

Существуют различные типы связей в речевой системе, например:

$$G_1 \rightarrow F_1 \xrightarrow{\varphi_1} S \rightarrow S^* \xrightarrow{\psi_1} G_1; \quad (9)$$

$$G_2 \rightarrow F_2 \xrightarrow{\varphi_2} F_1 \xrightarrow{\varphi_1} S \rightarrow S^* \xrightarrow{\psi_1} G_1 \xrightarrow{\psi_2} G_2. \quad (10)$$

Кроме того, возможны биологические обратные связи, включающие участников диалога:

$$G_2 \rightarrow \dots \xrightarrow{\varphi_1} S \xrightarrow{\psi_1} \dots \xrightarrow{\psi_2} G_2' \rightarrow F_2' \xrightarrow{\varphi_2'} \dots \xrightarrow{\varphi_1'} S^* \xrightarrow{\psi_1'} \dots \rightarrow G_2. \quad (11)$$

Приведенные структуры определяются структурой естественного языка. В частности, связи вида (5) реально существуют [7]. Более того, они замыкаются на очень низком уровне, т.е. на уровне образования звонких звуков речи [12]. Поэтому для нижних уровней иерархии диалоговой системы необходимо обеспечить устойчивость и точность управления. Соотношение (6) можно определить как условие обратимости цепочки преобразований (8) при восприятии и синтезе речевого сигнала и, возможно, речи.

#### R- И I-КАЧЕСТВА ДИАЛОГОВОЙ СИСТЕМЫ

Эти два качества обеспечивают достоверность передаваемой информации. Ясно, что для участников диалога должны во многом совпадать соответствующие модели, описанные в терминах образующих (7).

В настоящее время уровень надежности технических систем, т.е. R-качество, находится на весьма высоком уровне. Поэтому для обеспечения достоверности передаваемой информации необходимо обеспечить высокую степень помехоустойчивости, т.е. I-качество диалоговой системы. Обеспечение I-качества связано, во-первых, с выбором вида передаваемых сигналов, обеспечивающих наименьшую вероятность ошибки при приеме, и, во-вторых, — с выбором помехоустойчивого способа кодирования. Решение первой части задачи определяется особенностями слуховой системы восприятия, и, по-видимому, речевой сигнал настроен с учетом этих особенностей, которые во многом определили кодовую структуру речи [7].



Решение второй части задачи связано в классическом понимании помехоустойчивого кодирования [13] с выбором основания кода и метода избыточного кодирования и декодирования. До определенного уровня иерархии эта задача может рассматриваться как задача выбора множества образующих [9], точнее их имен и связей. Формально задача сводится к выбору основания кода и конечной алгебры. Неоднозначность решения этой задачи заключается в том, что существует множество различных языков, имеющих различный фонемный состав, звуковой строй и т.п., однако они имеют много общего на уровне речевого сигнала.

Анализ отображений (3) и описания (7) показывает, что множество трансформаций  $Z$  и признаки  $P_Y$  должны иметь соответствие между собой, возможно, быть связанными однозначным или взаимно однозначным отображением. При этом признаки  $P_Y$  могут играть и основную роль при формировании регулярных конфигураций. Это позволяет предположить, что при формировании отображения (3) существенное значение имеют требования помехоустойчивости, вернее, надежности принятия решения при разбиении множества  $X$  на гомотопические классы. По-видимому, должно быть некоторое соотношение между числом классов и объемом описания признаков  $\mu(P_Y)$ . В частности, известны разбиения звуков по месту и способу образования. Рассмотрим три группы классификации звуков:

I группа: гласные; сонанты; шумные звонкие; шумные глухие;

II группа: гласные; сонанты; шумные: смычные звонкие; смычные глухие, щелевые звонкие, щелевые глухие;

III группа: гласные; сонанты: смычные, щелевые, дрожащие; шумные: щелевые звонкие; щелевые глухие, взрывные звонкие, взрывные глухие, аффрикаты.

Для проведения эксперимента после ручного транскрибирования текста проводилась замена звука соответствующим ему классом. Затем решалась обратная задача — по полученной записи

находились повторяющиеся последовательности сочетаний, соответствующие определенным словам. При этом учитывалось не только правильное определение слова, но и нахождение однокоренного слова. Полученные соотношения для раздельной и слитной речи приведены в табл. 1 (соотношения однокоренных слов указаны в скобках).

**Таблица 1**

Речь	Соотношение между объемом описания признака и общим объемом текста для групп, %		
	I	II	III
Раздельная	49 (66)	54 (72)	69 (92)
Слитная	18 (30)	27 (47)	44 (74)

Таким образом, из полученных результатов следует, что при увеличении количества образующих, которые характеризуют описания признаков, повышается надежность определения конфигурации и соответственно образующих более высокого уровня.

#### **С-КАЧЕСТВО ДИАЛОГОВОЙ СИСТЕМЫ**

Управляемость диалоговой системы (рис. 3) может быть обеспечена выполнением целевых функций на уровне прагматики и семантики. Цель диалога [14] — информирование участников диалога, передача знаний и достижение консенсуса.

Ясно, что цель диалога определяется на прагматическом уровне, который проецируется на уровень семантики в виде определенных областей (см. рис. 3). Поскольку в процессе диалога цель достигается путем изоморфного отождествления цепочек образующих нижнего уровня в образующие высокого уровня, то каждой точке

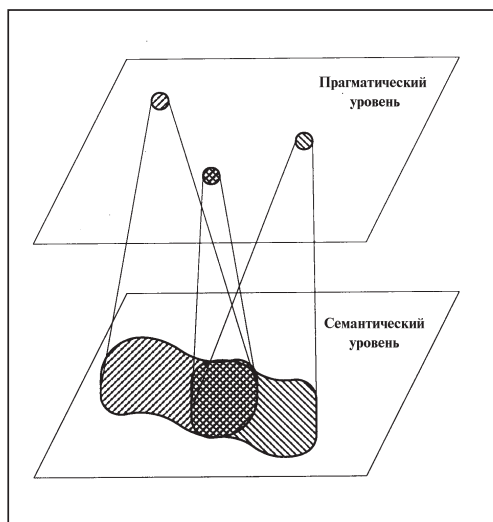


Рис. 3. Проекция целей участников диалога

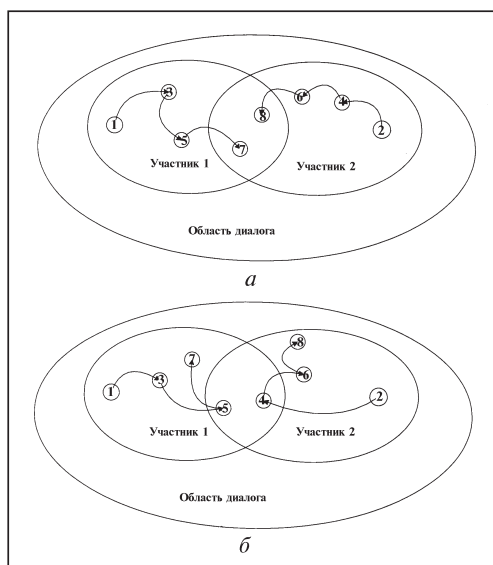


Рис. 4. Траектории продвижения диалога между участниками

Тем не менее, в некоторых случаях собеседники могут считать, что говорят об одном и том же, однако в действительности их предметные области различаются (рис. 4, б).

Если перейти на уровень целей, то также возможны различные ситуации. Например, собеседники могут не стремиться к взаимопониманию или их цели диалога расходятся. При этом тренды в семантическом пространстве диалога могут находиться в одной области и цель может быть достигнута.

Пример реализации переходов при стационарной прагматической компоненте, например при проведении экзамена, представлен на рис. 5. Здесь участник 1 генерирует сообщения Z1, Z3, Z5, а участник 2 — ответы Z4, Z6, Z8. В состоянии 7 и 8 оба участника переходят в общую область прагматики, для которых состояния 7 и 8 эквивалентны.

прагматического уровня будет соответствовать целая область семантического уровня. Здесь возникает достаточно сложная задача определения эквивалентности на множестве высказываний семантического уровня. Суть заключается в том, что эти отношения эквивалентности для различных участников диалога могут не совпадать. Такое несовпадение возможно и на более низких уровнях диалога, например на уровне синтаксиса и морфологии. Это возникает в случае, когда слова и термины воспринимаются участниками диалога по-разному (см. рис. 3).

Таким образом, цель диалога достигается путем изоморфного отождествления конфигураций образующих нижнего уровня в образующие высоких уровней. При этом траектории продвижения диалога (или преобразования конфигураций) будут управляемыми при условии наличия целевой функции.

Необходимо отметить, что при диалоге стороны преследуют свои цели. При этом они могут совпадать или различаться. С другой стороны, участники диалога имеют различные понятия, которые могут не пересекаться. Рассмотрим вариант, когда оба участника обладают близкими знаниями и их диалог направлен на достижение единой цели. На рис. 4, а отображены тренды движения в семантическом пространстве диалога, фиксирующие траекторию продвижения диалога собеседников. Как видно, участники диалога постепенно приходят в область, которая является общей для них.

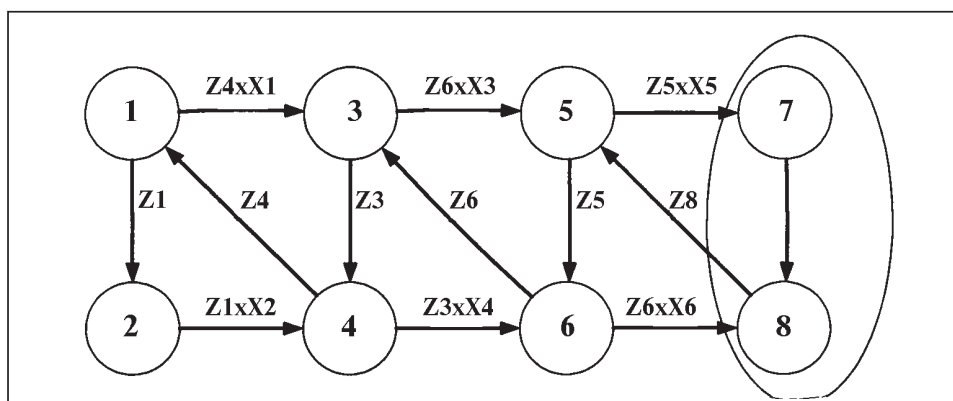


Рис. 5. Граф перехода при стационарной прагматической компоненте

### ЗАКЛЮЧЕНИЕ

Проведенный анализ концептуальной модели диалоговой системы позволяет сделать выводы, что построение качественной системы речевосприятия невозможно без реализации системы синтеза и наоборот. При этом необходимо учитывать такие явления, как прогноз развития событий не только на каждом уровне иерархии (цепочки слов, фонем, букв, слов и т.п.), но и между уровнями. В частности, знание предметной области диалога позволяет ограничить лексическую базу при принятии решения на уровне слов. Рассмотренная концептуальная модель позволяет целенаправленно формулировать требования по распознаванию и синтезу речи на разных уровнях иерархии диалоговой системы.

Анализ концептуальной модели показывает, что между системами речевосприятия и речеобразования имеется глубокая связь: общая база знаний; тесное взаимодействие их при диалоге. Общая база знаний предполагает, что при распознавании речевого сигнала на уровне звука должно формироваться параметрическое описание, соответствующее моделям движений артикуляторных органов речеобразующей системы. Иными словами, описание звуков в речевом потоке во многом должно быть близким к описанию состояния артикуляторов, иначе обратная связь при распознавании и речевосприятии может оказаться неэффективной.

Следует отметить, что *M*-качество системы ставит весьма актуальную задачу, связанную с выбором структуры образующих различных уровней иерархии диалоговой системы. Фактически решение этой задачи должно определить объемы описания имени, признаков и связей образующих [7], т.е. определить отношение эквивалентности на множестве элементов различного уровня иерархии. Увеличение объема описания имени обуславливает снижение объема описания признаков, но это может привести, во-первых, к снижению надежности распознавания и, во-вторых, к разрушению структуры связей, т.е. к разрушению регулярности в цепочках образующих. Определенную основу для решения этой задачи дают известные данные по фонетике, морфологии, грамматике и других языков.

Введение понятия *M*-качества иерархических систем позволяет использовать математические методы поиска оптимума, так как критерии оценки качества имеют также иерархический характер. При этом не только сокращается область значений, но и появляется возможность прогнозировать начальную точку поиска оптимального значения.

#### СПИСОК ЛИТЕРАТУРЫ

1. Быков Ю.С. Теория разборчивости в линиях связи. — М.: Оборониздат, 1954. — 196 с.
2. Мессарович М., Мако Д., Такахара И. Теория иерархических многоуровневых систем / Пер. с англ. под ред. И.Ф. Шахнова. — М.: Мир, 1973. — 311 с.
3. Искусственный интеллект. В 3-х кн. Кн.2: Модели и методы: Справочник / Под ред. Д.А. Поспелова. — М.: Радио и связь, 1990. — 304 с.
4. Шеннон К. Работы по теории информации и кибернетике. — М.: Изд-во иностр. лит., 1963. — 489 с.
5. Белоногов Г.Г., Новоселов А.П. Автоматизация процессов накопления, поиска и обобщения информации. — М.: Наука, 1979. — 257 с.
6. Бондаренко В.П., Коцубинский В.П., Мещеряков Р.В. Иерархические структуры распознавания и синтеза речи // Интеллектуальные автоматизированные системы проектирования, управления и обучения. — Томск: Изд-во НИИ АЭМ, 2000. — С. 115–125.
7. Сорокин В.Н. Теория речеобразования. — М.: Радио и связь, 1985. — 312 с.
8. Флейшман Б.С. Элементы теории потенциальной эффективности сложных систем. — М.: Сов. радио, 1971. — 223 с.
9. Гренандер У. Лекции по теории образов. Синтез образов / Под ред. Ю. Журавлева; пер. с англ. — М.: Мир, 1979. — 383 с.
10. Клацки Р. Память человека: Структуры и процессы / Под ред. Е. Соколова; пер. с англ. — М.: Мир, 1978. — 319 с.
11. Ту Дж., Гонсалес Р. Принципы распознавания образов. — М.: Мир, 1978. — 412 с.
12. Реклейтис Г., Рейвиндран А., Рэгсдел К. Оптимизация в технике: В 2-х кн. Кн. 1, пер. с англ. — М.: Мир, 1986. — 350 с. @REFERENCE = 13. Галлагер Р. Теория информации и надежная связь. — М.: Сов. радио, 1974. — 720 с.
14. Бондаренко В.П., Коцубинский В.П., Мещеряков Р.В. Синтез речевого сигнала по печатному тексту // Автоматическое и автоматизированное управление сложными системами / Под. ред. В.П. Тарасенко. — Томск: Изд-во Том. ун-та, 1998. — С. 204–217.

*Поступила 07.09.2005*