

ФОРМИРОВАНИЕ ПЕРЕХОДОВ В ИЕРАРХИИ ШКАЛ РЕЧЕВЫХ СИСТЕМ

к.т.н., доцент Мещеряков Р.В.,
д.т.н., проф. Бондаренко В.П.

Речевые системы (системы синтеза и распознавания речи, диалоговые системы, системы машинного перевода и др.) относятся к интеллектуальным системам. Они должны, во-первых, понимать речь, несмотря на возможные помехи, неточности 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, т.е. знания предметной области диалога. Основной психоакустической характеристикой речи является ее разборчивость, т.е. степень правильного восприятия звуков, слов и смысла речи. Максимальная разборчивость характерна для восприятия связной речи — фразовая разборчивость. Если человек воспринимает изолированные слова, процент разборчивости оказывается меньше. Она еще больше снижается при восприятии изолированных фонетических элементов речи типа слогов.

Однако, большинство подходов основано на последовательном выполнении распознавания слов и синтаксического разбора. Но опыт лингвистов показывает, что человек вначале схватывает смысл фразы и только затем приступает собственно к задаче понимания речи, которая часто рассматривается как процедура, независимая от этапа распознавания. Эта точка зрения естественно приводит к слабым результатам. Из психологии речевосприятия становится очевидным, что распознавание и понимание — это две тесно связанные процедуры, которые интегрированы в процесс диалога.

Внешними входными и выходными данными систем речевого диалога являются: смысловое пространство слов и фраз заданного языка и предметной области; речевой сигнал; а для канала синтеза — такие параметры речеобразующей системы. Речевой сигнал является связующим звеном, с помощью которого передается информация в системе речевого диалога.

Общепринято представлять речевую систему в виде иерархии уровней: прагматический; семантический; синтаксический; фонетический; физический (см. [1-3]). 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 при синтезе речи — для формирования и произнесения высказывания необходимо решение задач верхних уровней. Эти два процесса (синтез и распознавание речи) настолько сильно взаимосвязаны, что при решении прямой задачи требуется решение обратной.

Таким образом, основой диалоговой системы может служить многоуровневая иерархическая модель, что предполагает:

- 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 и другие внешние сведения.

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

- модели мира субъектов, ведущих диалог, пересекаются, т.е. для них являются частично общими знания о мире, определяемые конкретной предметной областью диалога [4];
- субъекты всегда, по 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 позволят ограничить область допустимых решений в каждый текущий момент диалога практически на всех уровнях иерархии и, как следует ожидать, повысить надежность восприятия речи и ее синтеза.

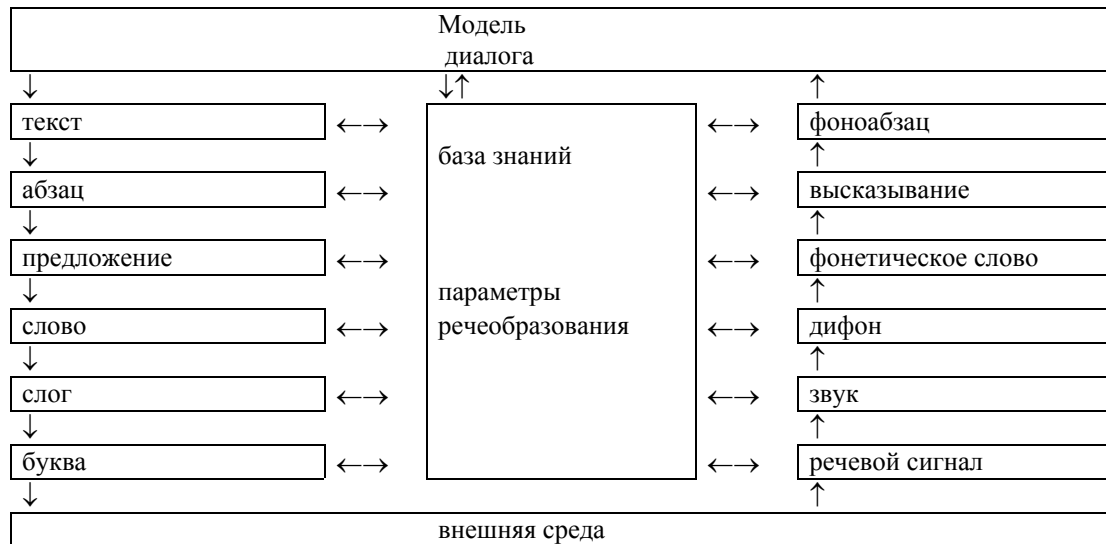


Рисунок 1. Иерархическая структура диалога.

Известно [1-3], что речевая система имеет иерархическую структуру, представленную на различных уровнях (см. рис.1). Каждый уровень речевой системы имеет свои собственные элементарные единицы, которые могут сочетать определенным образом. Особую сложность представляют собой согласование переходов с уровня на уровень.

Для перехода от одного уровня к другому недостаточно просто определить правила выделения образующих каждого уровня. Необходимо, кроме того, определить правила соединения образующих в регулярные конфигурации [4]. Ясно, что если имеется конфигурация

$$K_i = \{O_1^i, O_2^i, \dots, O_n^i\}, \quad (1)$$

которая получена, например, при использовании правила L_i при объединении образующих O_j^i в регулярную конфигурацию, то она может быть описана

$$K_i = \langle N_{k_i}, n, L_i \rangle. \quad (2)$$

где N_{k_i} – наименование конфигурации;

n – количество образующих в конфигурации;

L_i - правила сочетания образующих в конфигурации.

Ясно, что (2) есть компактная форма записи (1). Таким образом при описании уровней необходимо определение образующих и правил, т.е. получаем алгебру.

Для выделения элементарных объектов физического уровня (уровня звуков) чаще всего используют предварительное выделение квазиоднородных участков звукового сигнала. Будем считать, что множество квазиоднородных участков $S = \{s_i\}$. Для выделения из сигнала элементарных объектов можно использовать пространство состояний критериев качеств сложной системы. Для каждого объекта можно определить свое тело кластера K_i . Используя какую-либо меру близости [4,5] в данном пространстве можно и нужно определить образ, в который попадает искомый квазиоднородный участок речевого сигнала. Очевидно, что на данном уровне

сравнения происходят в определенных измерительных шкалах, в данном случае в шкалах интервалов.

Ясно, что измерительная шкала, в которой будут представлены образующие определенного уровня, в существенной мере определяют возможные операции. С другой стороны, шкалы могут быть построены на основе множества операций [5].

Определенной сложностью является определение отношений в этих шкалах на заданном пространстве. С одной стороны имеем выраженные в числовом выражении параметры сигнала (частота, амплитуда, мощность, временные характеристики), с другой стороны для кластеризации необходимо определение правил отнесения к какому-либо классу квазиоднородных участков.

Очевидно, можно определить отношение эквивалентности $R \sim$ двух квазиоднородных участков. Два квазиоднородных участка s_1 и s_2 будут эквивалентными, если будут принадлежать одному кластеру/классу в пространстве состояний, т.е. $s_1 \in K$ и $s_2 \in K$. Следовательно, все квазиоднородные участки, расположенные в одном кластере принадлежат одному и тому же классу эквивалентности $[K_i]$. Таким образом происходит классификация квазиоднородных участков. Можно заметить, что по определению отношение эквивалентности будет рефлексивным, транзитивным, симметричным.

Действия на множестве квазиоднородных участков выделить не представляется возможным, да и нет необходимости дополнительных преобразований речевого сигнала.

Используя отношение эквивалентности можно сделать отображение на множество наименований $M_N = \{m_i\}$ классов эквивалентности, $m: S \rightarrow M_N$, m_i — наименование кластера класса эквивалентности $[K_i]$. У образующих наименований кластеров можно определить свойства, описывающих характеристики пространства состояний.

Определив классы эквивалентности переходим к следующему уровню описания речевого сигнала: параметрическому. В этом случае получаем множество поименованных элементов $M_N = \{m_i\}$, соответствующим классам эквивалентности. Введем следующее отображение на данном множестве: $g: M_N \rightarrow R_N$, g — отношение связности двух элементов M_N , т.е. возможности идти в последовательной записи речевого сигнала. Данные последовательности являются регулярными конфигурациями в образах [3]. Конфигурации могут быть ограничены фонемой, дифоном, фонетическим словом. В зависимости от длительности последовательности необходимо учитывать связи не только двух последовательно стоящих объекта, но и все находящиеся 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 будет являться вероятностью (частотностью) сочетаемости объектов m_1 и m_2 между собой, точнее за объектом m_1 должен идти объект m_2 . Значение 0 соответствует значению их не сочетаемости. Значение 1 — объекты должны обязательно сочетаться между собой, т.е. всегда, когда встречается объект m_1 за ним идет объект m_2 .

Таким образом, при определении регулярности той или иной конфигурации может быть использовано два подхода:

- использование комбинации образующих (1) в качестве регулярности, что наблюдается на верхних уровнях иерархии речевых систем.
- использование описания правил регулярности конфигураций образующих (2), присутствующей в большей степени на нижних уровнях иерархии.

Проведенные исследования позволяют предположить достаточно строгое описание речевых систем, позволяющих формировать правила синтеза и распознавания на основе формальных моделей языка.

ЛИТЕРАТУРА

1. Потапова Р.К. Речь: информация коммуникация, кибернетика.
2. Бондаренко В.П., Мещеряков Р.В., Коцубинский В.П. Иерархические структуры распознавания и синтеза речи. // Сборник: Интеллектуальные автоматизированные системы проектирования, управления и обучения. - Томск: STT, 2000. - с.115-125.
3. Бондаренко В.П., Мещеряков Р.В., Коцубинский В.П. АРСО
4. Гренандер У. Лекции по теории образов. Синтез образов. / под. Ред. Ю.Журавлева; пер. с англ. - М.: Мир, 1979. - 383 с.
5. Маркус С. Теоретико-множественные модели языков. - М.:Наука, 1970, - 332с.