

МОДЕЛИРОВАНИЕ ДИНАМИЧЕСКОГО ВЗАИМОДЕЙСТВИЯ РЕЧЕВОГО СИГНАЛА С ПАМЯТЬЮ В СИСТЕМАХ РАСПОЗНАВАНИЯ РЕЧИ

Г.Н.Вандакурова, Р.Я.Гумецкий, Л.А.Мелень

Кафедра биофизики и математических методов в биологии
Львовский государственный университет, Львов, СССР, 290000

АННОТАЦИЯ

Распознавание речевого сообщения рассматривается как целенаправленный динамический процесс взаимодействия ассоциативно организованной памяти с речевым сигналом, позволяющий генерировать гипотезы о каждом из элементов сообщения, которые проверяются на соответствие воспринимаемым акустическим образам. При этом максимально используется информация многоуровневой памяти распознающей системы, а ее рецептивные возможности задействуются на глубину, достаточную для однозначной смысловой интерпретации речевого сообщения.

ВВЕДЕНИЕ

В естественной речи многие из номинальных элементов сообщения не могут быть непосредственно восприняты, так как они зачастую недостаточно четко реализованы или даже редуцированы в произнесенной фразе. Решения о них могут появляться при восприятии речи только ретроспективно, после того, как сообщение становится осмысленно понятным. Поэтому в плане выполнения функции восприятия речи принципиально важна особенность распознающей системы заключается в способности текущего предсказания возможных в данной ситуации элементов сообщения, которое уточняется до непротиворечивости (соответствия) воспринимаемому речевому сигналу [1].

Методологическая основа рассматрива-

емого подхода к моделированию восприятия речи основывается на предположении, что многоуровневый процесс распознавания речевых сообщений заключается не столько в процедурах таксономии и классификации конкретных акустических образов, сколько в целенаправленной динамической актуализации ими языковой памяти распознающей системы, что в конечном итоге позволяет однозначно интерпретировать сообщение на основе переданных в речевом сигнале смыслозначительных признаков, достаточных для осуществления этой цели. Наряду с выявлением таких информативных признаков, в этом процессе основополагающую роль играют механизмы дедукции, включая активный поиск аргументов для оценки соответствия воспринимаемых акустических образов предполагаемым на каждом шаге анализа элементам сообщения [2].

С этой точки зрения процедуры активного взаимодействия между памятью, организованной с учетом лингвистических и иных ассоциаций, и специфическими механизмами обработки речевого сигнала не могут игнорироваться и должны быть введены в общую структуру модели распознавания речи. Возможные алгоритмы взаимодействия речевого сигнала с памятью распознающей системы исследовались в настоящей работе.

ОРГАНИЗАЦИЯ МОДЕЛИ

Описанная ниже модель основывается на дедуктивном подходе к распознаванию речи [1,4]. В соответствии с его основными по-

ложениями, параллельно с процессом обработки воспринимаемого речевого сигнала текущим образом генерируются обоснованные в данной ситуации гипотезы о распознаваемом речевом образе, которые верифицируются путем сопоставления предполагаемой фонемной структуры с ее акустической реализацией в произнесенном сообщении. При этом моделируются три принципиально различных вида перцептивной активности, определенные как:

1/ мыслительная активность (выдвижение словесно сформулированных гипотез в соответствии с лексиконом и закономерностями используемого для речевой коммуникации языка), управляемая результатами распознавания сообщения;

2/ слуховая активность (обработка и представление речевого сигнала в виде динамики обнаруживаемых в нем фонетических качеств, подготавливающие его к сопоставлению с фонемными матрицами слов, выдаваемых на данном отрезке сообщения алгоритмом предсказания);

3/ активность принятия решения (верификация предсказываемых фонемных последовательностей по результатам фонетической интерпретации речевого сигнала), в результате которой определяется максимально правдоподобная фонемограмма высказывания [1].

В распознающей системе на используемый словарь всевозможных слов, находящихся в ее памяти (лексикон), априорно действует или текущим образом накладывается целый ряд ограничивающих факторов, которые постоянно задают и перестраивают подмножества слов-гипотез, актуальных на каждом шаге анализа конкретного сообщения. Ограничения, которые позволяют упорядочить (при обучении) и целесообразно задействовать память при распознавании речи, могут задаваться как изнутри самой системы, так и из внешних источников. В первом случае имеет место воздействие на процесс восприятия речи внутренней структуры используемого проблемно-ориентированного языка (лексики, синтаксиса, семантики). Внешние источники информа-

ции имеют место, когда известна ситуация речевого общения (тематическая, контекстуальная, в т.ч. и текущие результаты распознавания сообщения).

Таким образом, на основе взаимодействия априорной информации языковой памяти с текущей информацией воспринимаемого сигнала формируется результат распознавания пословно предсказываемого сообщения как последовательность слов, наилучшим образом соответствующая фонетической структуре высказывания. Сама мера подобия речевого сигнала эталонам верифицируемых элементов речи (слов) может быть оценена с помощью различных решающих правил и мер близости.

АЛГОРИТМИЧЕСКИЕ РЕШЕНИЯ

В моделируемом алгоритме распознавания речевого сообщения максимально используется априорная информация, заданная в памяти воспринимающей системы, а рецептивные возможности системы задействуются при анализе определенного участка речевого сигнала в той мере, в какой это необходимо для однозначного декодирования соответствующего лексического элемента сообщения.

На каждом шаге распознавания текущее подмножество актуальных лексических гипотез из заданного словаря определяется на основе пересечения подмножеств слов, которые:

1/ задаются априорно вводимыми прагматическими и лингвистическими ограничениями на высших уровнях системы;

2/ определяются текущим образом, исходя из лингвистических ограничений, на основе решений о предшествующих отрезках речи;

3/ определяются по результатам оценки интегральных надежно регистрируемых характеристик очередного участка речи на низших ступенях его анализа.

Для реализации последнего в модель введена динамическая организация памяти по свойствам речевых сигналов, которые не являются традиционными свойствами лингвистической категоризации. В этом плане испыты-

валась ассоциативная группировка слов (осуществляемая параллельно с эталонизацией словаря) по близости их интегральных акустических характеристик, типичный пример которой рассмотрен в работе [3].

На этапе обучения системы для разбиения заданного словаря на подмножества близких по акустическому образу слов могут использоваться процедуры кластеризации обучающих реализаций слов по различным - структурным, просодическим, маркирующим и другим надежно регистрируемым характеристикам их первичного описания. Обнаружение соответствующих характеристик в акустической реализации распознаваемого слова существенно уменьшает область поиска решения. Алгоритм группового распознавания слов, который испытывался в составе описываемой модели, рассмотрен в работе [3].

В результате наложения всех указанных выше ограничивающих факторов на выходе памяти формируется осмысленное, акустически и лингвистически обоснованное предсказание для каждого лексического (следовательно и фонемного) элемента конкретного сообщения, неоднозначность которого (если она имеет место) разрешается оценкой соответствия эталонных описаний слов-конкурентов фонетической структуре воспринимаемого речевого сигнала. Таким образом, составом актуальной части лексической памяти на каждом шаге восприятия сообщения постоянно управляют как первые ступени, так и конечный результат текущего анализа очередного отрезка речи.

Алгоритм принятия решения об элементах речевого сообщения состоит в том, что поступающие из памяти эталоны актуальных слов-гипотез в нужный момент сопоставляются с очередным отрезком речевого сигнала. Существенно, что при этом происходит предварительная унификация описаний речевого сигнала и эталонов с использованием для этой цели многомерного пространства надежно различимых фонетических качеств, достаточных для сегментного представления речи.

Последняя в одном из вариантов модели проводилась на основе двоичного структурного описания речи, предложенного в работе [5] и позволяющего проводить нелинейное во времени сопоставление без использования процедур ДП-согласования.

ЭКСПЕРИМЕНТАЛЬНАЯ РЕАЛИЗАЦИЯ МОДЕЛИ

В диалоговом режиме с использованием ограниченных сменных словарей моделировались типичные задачи распознавания фраз-команд в автоматизированных системах управления (речевое управление вычислительной машиной, речевой запрос в ИПС и другие ситуации). Аппаратурно-программные средства реализованных экспериментальных систем включали:

а/ устройство выделения и ввода речевых признаков в ЭВМ (УВРП);

б/ базовый комплект устройств мини-ЭВМ (УВК СМ-4);

в/ программное обеспечение речевого управления (ПОРУ).

Исходным дискретным описанием речевого сигнала, вводимым в ЭВМ, являлась последовательность выдаваемых УВРП каждые 20 мс векторов качественных и количественных речевых признаков, отражающих акустические характеристики источника, способа и места образования звуков в речевом тракте человека. Вторичное структурное описание речи определялось наличием-отсутствием в речевом сигнале конкретной последовательности сегментов с различным фонетическим качеством из потенциально возможной последовательности сегментов всех различимых типов. Переход от временного описания к структурному осуществлялся с одновременным определением наличия пауз - границ слов.

Алгоритмы, разработанные для речевого обращения к ЭВМ, были реализованы в ПОРУ в виде комплекса: 1/ универсальных управляющих программ, которые обеспечивают работу системы в режимах обучения (эталонизации, группировки произносимых слов исполь-

зуемого словаря) и распознавания речи (с выполнением воспринятых речевых команд), и 2/ ряда общих подпрограмм (ввода речи, формирования структурного описания, вывода текущих результатов и др.). Связи между группами слов, которые могут встретиться в определенных местах фраз или ассоциируются по акустическому подобию, представлялись в памяти в виде графов, задаваемых списочными структурами.

В режиме распознавания программно реализованы и опробованы:

а/ структурный метод распознавания речи на основе бинарного признакового описания речевых сигналов и хеминговой меры их близости к эталонам слов из памяти системы;

б/ метод распознавания сообщений по полному признаковому описанию речевых сигналов с ДП-согласованием в качестве меры их сходства с эталонами.

Исследовалась работоспособность распознающих моделей при объеме экспериментальных словарей до 100 слов. В условиях шумов до 75 дБ оценена надежность распознавания речевых команд различных операторов, которая при заданных ограничениях составляет 90±3% для лиц, имеющих опыт работы с системой.

ВЫВОДЫ

Результаты исследования рассмотренных алгоритмов взаимодействия параллельных потоков актуальной информации из памяти с потоком текущей информации речевого сигнала, полученные в частных моделях, подтверждают эффективность описанного подхода к распознаванию речевых сообщений формализованных проблемно-ориентированных языков. В теоретическом аспекте - показана возможность реализации разработанных алгоритмов в виде целостной иерархической модели восприятия речи.

Следует особо подчеркнуть значение целенаправленной активности в работе всех уровней и систем, принимающих участие в

процессе восприятия речи. Это означает, что и все задействованные в распознающей модели блоки неизбежно должны находиться под воздействием управляющих факторов, наиболее мощным из которых является фактор языка, направляющий восприятие речи. Учет всего многообразия факторов, как внутренних, так и внешних, в том числе и самой речи, которые управляют корректностью и адекватностью упреждающей активности распознающей системы, представляется нам как феномен ее интеллектуализации. Именно в таком плане следует понимать значение предварительной организации памяти распознающей системы и ее динамического взаимодействия с распознаваемым сигналом, вне которого осмысленное восприятие, т.е. понимание речи автоматом представляется невозможным.

ЛИТЕРАТУРА

- [1] Derkach M. Deductive approach to automatic recognition of russian spoken sentences // Proc. ICASSP, 1980. - pp. 1041-1044.
- [2] Динамические спектры речевых сигналов / М.Ф.Деркач, Р.Я.Гумецкий, Б.М.Гура, М.Е.Чабан; под ред. М.Ф.Деркача. - Львов: Вища школа. Изд-во при Львов. ун-те, 1983. - 168 с.
- [3] Гумецкий Р.Я., Вандакурова Г.Н., и др. Двухступенчатый алгоритм распознавания слов, использующий информацию о совокупности и последовательности акуст.-фонет. признаков в речевом сигнале // Вест. Львов. ун-та, сер. биол., 1983. С. 105
- [4] Вандакурова Г.Н., Гумецкий Р.Я., Мелень Л.А. К разработке систем распознавания речевых команд из ограниченных сменных словарей // Тез. докл. Всесоюз. шк.-семинара АРСО-13. Новосибирск, 1984. - С. 126-127.
- [5] Гумецкий Р.Я., Гура Б.М., Мишин Л.Н. Использование двоичного структурного описания речевого сигнала в задачах распознавания речи // Там же. - С. 136-137.