

1558522

На правах рукописи

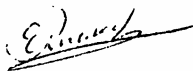
Ермаков Александр Евгеньевич

**Разработка и исследование методов обработки знаний
в системах распознавания речи
на основе
иерархических структур из динамических нейронных сетей**

Специальность 05.13.16:

Применение вычислительной техники,
математического моделирования и
математических методов в научных исследованиях

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук



Москва 1998

Работа выполнена в Московском государственном техническом университете им. Н.Э. Баумана

Научный руководитель:

кандидат технических наук,
профессор Плотников В.Н.

кандидат технических наук,
старший научный сотрудник
Харламов А.А.

Официальные оппоненты:

доктор технических наук,
профессор Женило В.Р.

кандидат технических наук,
ведущий научный сотрудник
Жигулевцев Ю.Н.

Ведущая организация:

ВЦ РАН

1 сн9 1998г. в 16.40 часов на заседании
15.12 в Московском государственном
Баумана по адресу: 107005, Москва, 2-я

иться в библиотеке университета.

претя 1998г.

1558522

Яковлев А.В.



1558522

Ермаков А.Е. Разработка и

ЦАЯ ХАРАКТЕРИСТИКА РАБОТЫ

НА ДОМ
НЕ ВЫДАЕТСЯ

~~задачи~~ проблемы. В связи с успехами в развитии вычислительной техники и новых информационных технологий в последнее десятилетие определилась тенденция к нарастанию сложности систем управления, а также всех видов человеко-машинных систем. Возрастающая потребность в автоматизации всех сфер производственной и организационно-управленческой деятельности и необходимость решения новых задач, традиционно считавшихся прерогативой человека, привели к выделению самостоятельного класса человеко-машинных систем, называемых интеллектуальными системами.

Актуальными задачами, связанными с развитием интеллектуальных систем, являются автоматическое извлечение и представление знаний в универсальной форме, в качестве которой следует рассматривать естественно-языковую форму, а также организация эффективного взаимодействия между человеком и системой на естественном языке. Только создание систем понимания естественного языка (распознавания речи и анализа текстов) позволит автоматизировать процесс накопления знаний интеллектуальной системой и устранить разрыв между высокой скоростью обработки информации в ЭВМ и низкой эффективностью обмена информацией с человеком.

Несмотря на то, что исследования в области создания систем распознавания речи (СРР) ведутся более 20 лет, современные СРР обладают рядом ограничений. К наиболее существенным из них следует отнести ориентированность на изолированно произносимые слова и на ограниченный тезаурус предметной области (в наиболее мощных системах - около 20 тысяч слов), а также сильную подверженность влиянию шумов и вариаций стиля произнесения диктора. Существующие системы понимания речи обладают словарем порядка тысячи слов и поддерживают ограниченные форматы языковых конструкций. Вместе с тем, на фоне появления зарубежных СРР на рынке, отчетливо проявляется отсутствие отечественных систем распознавания речи, что отчасти вызвано сложностью русского языка.

Вычислительные и алгоритмические трудности, встающие на пути создания систем распознавания и понимания слитной речи в больших словарях, требуют для своего разрешения новой аппаратной базы, отличной от традиционных последовательных ЭВМ, а также соответствующих однородных алгоритмов, ориентированных на автоматическую обработку знаний. В связи с этим перспективным является применение однородных вычислительных сред с массовым параллелизмом - нейронных сетей, способных к самообучению.

Цель работы. Целью диссертационной работы является разработка методов автоматического представления речевых и языковых знаний всех

МГТУ
им. Н.Э. Баумана
БИБЛИОТЕКА

уровней в системах распознавания и понимания речи на основе однородных структур с массовым параллелизмом - нейронных сетей, а также механизмов анализа (расознавания и понимания) речевых сообщений на их основе.

Основные задачи исследования.

1. Провести анализ существующих методов представления знаний всех типов, а также алгоритмов обработки информации, применяемых в СРР с целью понимания принципиальных проблем, возникающих при их использовании;

2. Провести анализ существующих нейросетевых парадигм и их применений к задаче распознавания речи с целью выбора адекватной структуры нейронной сети для достижения поставленной цели - автоматического однородного представления знаний в СРР;

3. Модифицировать структуру выбранной нейронной сети для представления различных типов знаний в СРР на основании анализа основных особенностей речевых сигналов и представлений лингвистики о структуре языка;

4. Разработать подход к анализу содержания языковых сообщений (пониманию речи) на основе выбранной нейросетевой структуры;

5. Провести экспериментальное исследование разработанных нейросетевых алгоритмов моделированием на традиционной последовательной ЭВМ.

Научная новизна работы. Основной научный результат заключается в разработке методов автоматического формирования представления речевых и языковых знаний и анализа речевых сообщений в системах распознавания и понимания речи на основе иерархических структур из динамических нейронных сетей:

1. Использована парадигма иерархических структур (ИС) из динамических нейронных сетей на основе нейроподобных элементов с временной суммацией входных сигналов - динамических ассоциативных запоминающих устройств (ДАЗУ) для обработки речевой и тестовой информации;

2. Для управления ИС ДАЗУ в режиме распознавания использован механизм, реализующий динамическую подстройку чувствительности нейроподобных элементов к искажениям информации на входе на основе ее предсказания, называемый механизмом внимания;

3. Разработан метод формирования эталонов речевых единиц в ДАЗУ, реализующий покрытие областей отображения речевых сигналов в сигнальное пространство множеством связанных гиперсфер;

4. Разработан метод автоматического формирования представления языковых знаний семантического уровня на основе статистического анализа множества текстов в ИС ДАЗУ, учитывающего взаимосвязи языковых элементов в рамках структурных единиц текста;

5. Предложен подход к анализу содержания языкового сообщения как связного текста с опорой на семантическое представление на основе ИС ДАЗУ, управляемой механизмом внимания;

6. Предложена структура системы анализа (распознавания и понимания) связной речи с опорой на семантическое представление, включающей в свой состав речевую и языковую подсистемы на основе двух ИС ДАЗУ, управляемых единым механизмом внимания.

Реализация результатов. Разработанные в диссертации методы использованы:

- в МГТУ им. Н.Э. Баумана в комплексе обучающих программ для студентов по обработке звуковой и текстовой информации в динамических нейронных сетях;

- в МГТУ им. Н.Э. Баумана в аппаратно-программной системе распознавания речевых команд, в рамках программы «Интеллектуальные системы»;

- в ГНТЦ «ГИНТЕХ» в системе распознавания речи, в рамках проекта «Разработка концепции и инженерно-технических решений по созданию перспективных ЭВМ на базе нейротехнологий. Система для диктовки деловой корреспонденции голосом - фонетическая пишущая машинка», поддержанного Министерством науки и технологий РФ;

- в Мэрии г. Москвы в системе автоматического смыслового анализа текстовой информации, в рамках проекта «Разработка технологии представления языковых знаний на основе иерархических структур из динамических нейронных сетей и разработка аппаратно-программного комплекса автоматической обработки текстов», поддержанного Министерством науки и технологий РФ;

- в издательстве «Humtec Publisher» в системе автоматического смыслового анализа текстовой информации.

Практическая ценность. Разработанные в диссертации методы ориентированы на следующие применения: 1) построение систем распознавания и понимания связной речи, 2) формирование баз знаний на основе автоматического анализа текстовой информации, 3) использование в других областях: для автоматического анализа информации с выделением иерархии составляющих элементов ее структуры, для управления процессом обработки информации при распознавании образов сложной структуры на основе нейронных сетей.

Апробация работы. Основные положения работы докладывались: на втором международном симпозиуме «Интеллектуальные системы» (С.-Петербург, 1996г.); на конференциях «Нейрокомпьютеры и их применения» НКП-96 и НКП-97 (Москва, 1996 и 1997гг.); на научном семинаре кафедры прикладной лингвистики в Московском Государственном Лингвистическом

Университете (1996г.); на научных семинарах по программе «Интеллектуальные системы» в МГТУ им. Н.Э. Баумана, (1996 и 1997гг.).

Материалы диссертационной работы использованы автором в курсах лекций «Технические средства обработки звуковой информации» и «Введение в кибернетику», читаемых в МГТУ им. Н.Э. Баумана и Российском Университете Дружбы Народов.

Разработанные в рамках диссертационной работы системы демонстрировались:

- на выставках конференций «Нейрокомпьютеры и их применение» (ИКП-96 и НКП-97) (Москва, МинОборонПром, 1996, 1997 гг.);
- на Лейпцигской инновационной ярмарке (Лейпциг, Германия, 1996г.);
- на выставке «Телеком-97» (Москва, ВВЦ, 1997г.);
- на выставке-ярмарке в области вычислительной техники, информатики и коммуникаций «SIMO-97» (Мадрид, Испания, 1997г.);

Публикации. По теме диссертации автором опубликовано 8 печатных работ, общим объемом 2,7 печатных листа.

Объем работы. Диссертация состоит из введения, пяти глав и заключения, изложенных на 130 страницах, содержит 13 рисунков, включает 125 наименований отечественной и зарубежной литературы, и четыре приложения, всего 157 страниц.

СОДЕРЖАНИЕ РАБОТЫ

Первая глава диссертации посвящена аналитическому обзору литературы по методам обработки информации в СРР, а также теории нейронных сетей и опыту их применения для распознавания речи.

Рассматривается общая структура СРР, функции всех модулей и алгоритмы их функционирования.

Ввиду высокой вариативности речевого сигнала распознавание речи требует привлечения всей доступной информации, позволяющей разделить перцептивно близкие образы. Такая информация относится к собственно речевой (звуковой), а также к языковой сторонам речи. Первая представляет знания об особенностях акустических реализаций отдельных единиц речи и позволяет выделить из сигнала информативные признаки для классификации его фрагментов. Вторая, языковая, охватывает все знания как о формальной стороне языка - грамматике, так и о смысловой - семантике, и позволяет сформировать ограничения на сочетаемость речевых единиц.

Проведенный анализ методов, применяемых при распознавании речи, позволяет сформулировать основные проблемы, встающие при разработке СРР:

1) сложность совмещения разнородных способов представления речевых и языковых знаний и алгоритмов оперирования с ними на разных уровнях СРР;

2) сложность реализации в реальном масштабе времени вычислительно емких алгоритмов распознавания на последовательных однопроцессорных и даже мультипроцессорных ЭВМ;

3) невозможность автоматического формирования представления знаний высшего языкового уровня - семантического;

4) ориентированность существующих методов анализа языковых сообщений на формальный разбор грамматических конструкций

Результаты обзора позволяют сформировать следующие предпосылки для разработки нового подхода к обработке знаний в СРР:

- в качестве универсальной формы, позволяющей реализовать однородное представление как речевых, так и языковых знаний, следует избрать представление в форме сетей состояний и переходов;

- ввиду того, что существующие сетевые реализации не решают задачу автоматического формирования представления языковых знаний и при распознавании требуют применения вычислительно емких алгоритмов, необходима реализация сетевого представления на нейронных сетях, обладающих способностью к самообучению и позволяющих достичь высоких скоростей обработки информации в случае аппаратной реализации.

Анализ нейросетевых парадигм показывает, что традиционные типы нейронных сетей (многослойные персептроны и рекуррентные сети) не предназначены для формирования модели структурированных знаний.

Адекватной решаемым задачам была признана нетрадиционная парадигма нейронной сети - динамическое ассоциативное запоминающее устройство (ДАЗУ) - динамическая нейронная сеть на основе нейроподобных элементов с временной суммацией входных сигналов, которая позволила автоматически формировать однородное представление речевых и языковых знаний на всех уровнях СРР в форме сетей состояний и переходов.

Во *второй главе* излагается теория ДАЗУ и иерархических структур (ИС) на их основе. Основные принципы обработки информации в ИС ДАЗУ были описаны в работе [Харламов А.А. Нейроподобные элементы с временной суммацией входного сигнала и блоки ассоциативной памяти на основе этих элементов. // Вопросы кибернетики. Устройства и системы. / Под ред. Н.Н. Евтихьева. - М.: МИРЭА, 1983. - С. 57-68.] для случая двоичных последовательностей. Автором диссертационной работы модифицирована структура нейроподобного элемента и проведено дальнейшее развитие теории ДАЗУ для решения поставленных задач, а также использован механизм внимания для управления ИС ДАЗУ в режиме распознавания.

Основным элементом ДАЗУ является нейроподобный элемент (НЭ) с временной суммацией входных сигналов, реализующий следующие функции:

- обучение - запоминание n -членного фрагмента входной последовательности $a^i = (a_{i-p+1}, \dots, a_{i-1}, a_i)$ (n -граммы), интерпретируемого как координаты точки или центра гиперсферы в n -мерном сигнальном пространстве: $w^i_i = a^i_i$.

- избирательная адресация - подача сигнала y_i обученным НЭ w^i_i при появлении на входе n -граммы, близкой к запомненной: $y_i \neq 0$, если $D(w^i_i, y^i_i) \leq \epsilon$.

- подсчет частоты встречаемости z запомненной n -граммы w^i_i или близких к ней во входных последовательностях;

- пороговое преобразование, реализующее забывание статистически недостоверной n -граммы.

Чувствительность НЭ к отклонению входной n -граммы от ранее запомненной определяется радиусом расфокусировки НЭ r - радиусом гиперсферы СП.

ДАЗУ представляет собой множество НЭ W^A , имеющих общий вход и общий выход, и моделирует n -мерное сигнальное пространство. Выход ДАЗУ замыкается на дополнительный общий вход всех НЭ через механизм управления, называемый механизмом внимания (МВ), что реализует обратную связь. Общим входом всех НЭ является регистр сдвига (РС), сканирующий входную последовательность окном длины n элементов (рис.1).

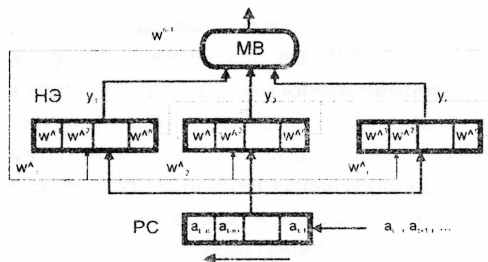


Рис. 1

Отображение F , реализуемое РС, преобразует входную последовательность $A=(a_i)$ в последовательность точек сигнального пространства: $F(A)=(a^i_i)=A^A$.

При обучении ДАЗУ запоминает во множестве НЭ W^A точки образа A^A входной последовательности. При этом НЭ выделяются динамически:

$$w^i_{p+1} = a^i_i, \text{ если } \exists w^i_i : w^i_i = a^i_i, i=1..p,$$

где p - количество НЭ, уже задействованных к моменту времени t .

Основное свойство отображения F таково, что на множестве НЭ в ДАЗУ определяется функция автоассоциации, которая задает для каждого n -членного фрагмента входной последовательности a^i_i , представленного в НЭ w^i_i , его возможные продолжения a^i_{i+1} , представленные в НЭ $\{w^i_{i+1}\}$, и реализуется за счет обратной связи:

$$F_{ac}(w^{\wedge}_i) = \{w^{\wedge}_j: w^{\wedge p+1}_j = w^{\wedge p}_i, p=1..n-1\},$$

где $w^{\wedge p}_j$ - p -ая компонента $w^{\wedge}_j$. Множество НЭ $\{w^{\wedge}_i\}$, представляет в СП траекторию автоассоциативно связанных точек.

ДАЗУ реализует представление входных последовательностей в форме сети состояний, представленных n -граммами входных последовательностей в соответствующих НЭ, и направленных переходов между ними, соответствующих автоассоциативным связям. Представление информации в ДАЗУ обладает свойством компактности - общие участки последовательностей представляются в общих участках траекторий. При этом каждой входной последовательности A_k , встретившейся при обучении ДАЗУ, соответствует траектория в сети $W^{\wedge}_k$.

Свойство ДАЗУ отображать повторяющиеся фрагменты входных последовательностей в общий участок траектории СП позволяет проводить анализ структуры входной информации с выделением словаря повторяющихся элементов. Применение порогового преобразования N к обученному ДАЗУ сохраняет только n -граммы с частотой встречаемости z не ниже порога h - участки траекторий, соответствующие составляющим элементам множества входных последовательностей: $N(W^{\wedge}) = \{w^{\wedge}_i: z(w^{\wedge}_i) \geq h\}$.

Обученное ДАЗУ способно к распознаванию входных последовательностей. Для обеспечения распознавания, устойчивого к искажениям, НЭ могут быть расфокусированы с радиусом r , моделируя гиперсферы СП $\{w^{\wedge}_{(r)}\}$. Траектории ДАЗУ в этом случае представляют собой гипертрубки, образованные автоассоциативно связанными гиперсферами. Распознавание осуществляется ассоциативным вложением входной траектории $A^{\wedge} = F(A)$ в область гипертрубки $W^{\wedge}_{(r)}$, и ему соответствует непрерывная последовательность откликов НЭ $\{y_{i(t)}\}$.

Для управления распознаванием в ДАЗУ используется механизм внимания (МВ), реализующий функции, подобные концентрации и рассредоточению внимания. Работа МВ основана на использовании ранее распознанной информации для предсказания следующей и подстройки параметров НЭ под наилучшее распознавание ожидаемой. Для этого в ДАЗУ вводится варьирование во времени радиусов расфокусировки r НЭ зависимости от состояния процесса распознавания:

$$r_i(t+1) = R(\{r_j(t)\}, \{y_j(t)\}, i),$$

где R - функция управления вниманием, определяемая в работе. При наличии распознавания в одном из НЭ МВ увеличивает радиусы расфокусировки на ассоциативно связанных с ним НЭ и уменьшает на остальных, увеличивая вероятность вложения в ожидаемые продолжения траектории в случае искажения входной информации, что соответствует концентрации внимания. При отсутствии распознавания во всех НЭ МВ увеличивает все радиусы, помогая вложиться в ближайший участок траектории, что соответствует рассредоточению внимания.

Наибольшей эффективности при обучении и распознавании позволяет достичь использование ИС ДАЗУ, где ДАЗУ нижних уровней работают как фильтры, пропуская нераспознанную информацию на вход более высоких уровней.

$$Y_t(W^{\wedge}, A) = (y_t), \text{ где } y_t = \begin{cases} y_j, & \text{если } \exists j: D(w^{\wedge}_j, a^{\wedge}_t) \leq r_j \\ a^{\wedge}_t & \text{в противном случае,} \end{cases}$$

К-уровневая ИС ДАЗУ автоматически осуществляет анализ структуры входной информации A с формированием иерархии взаимосвязанных словарей $\{W^{\wedge 1}, W^{\wedge 2}, \dots, W^{\wedge k}\}$ составляющих ее элементов различной степени сложности:

$$S^k(A) = HF(Y_t(W^{\wedge k-1}, \dots, W^{\wedge 1}, A)) = W^{\wedge k}$$

Иерархия словарей представляет собой многократно вложенный суперграф, в котором элементы словарей более высоких уровней являются комбинациями элементов более низких уровней, а между элементами одного уровня определяются отношения автоассоциативной связанности.

Простейший пример представления информации в ИС ДАЗУ при $n=3$ и $k=3$ для случая символьных последовательностей, соответствующих словам «нейрон» и «корона», приведен на рисунке 2.

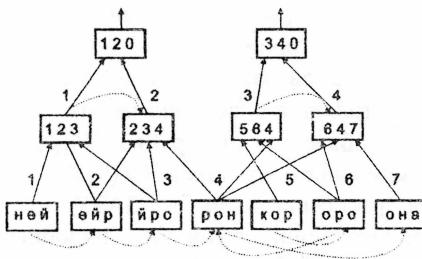


Рис. 2

При распознавании в ИС ДАЗУ управление вниманием принимает на себя ДАЗУ наиболее высокого уровня из числа тех, на которых происходит распознавание. Таким образом реализуется глобальный МВ, использующий информацию высших уровней для адекватной подстройки радиусов расфокусировки НЭ на более низких уровнях ИС и обеспечивающий наибольший запас устойчивости к искажениям входной информации.

В заключении главы качественно описывается динамический процесс распознавания в ИС ДАЗУ под управлением глобального МВ, сопровождающийся переключением внимания по отдельным НЭ в ДАЗУ и по уровням ДАЗУ в ИС.

В третьей главе диссертации разрабатывается метод представления речевых знаний для формирования эталонов речевых единиц и алгоритм их

распознавания на основе ИС ДАЗУ в речевой подсистеме системы распознавания речи.

В начале главы рассматриваются основные особенности речевых сигналов и предлагается способ описания эталона речевой единицы в сигнальном пространстве (СП) в форме гипертрубки переменного радиуса, траектории которой соответствуют отдельным акустическим реализациям речевой единицы. Разработанный метод представления эталона основан на покрытии области гипертрубки множеством связанных гиперсфер в СП ДАЗУ (рис.3).

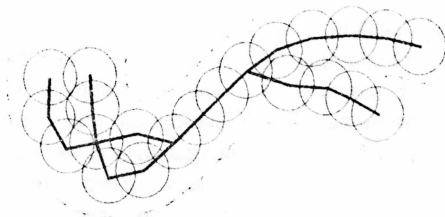


Рис.3.

Отображение F общего вида, введенное во второй главе для произвольных входных последовательностей, конкретизируется с учетом особенностей обрабатываемой в ДАЗУ информации - последовательностей векторов параметров (ВП) предобработки речевого сигнала, извлекаемых с дискретным шагом на коротких фрагментах сигнала. Определяется метрика:

$$D = D_n(\hat{a}^j, \hat{w}^j) = \max \{ d_j; d_j = d(\hat{a}_i^j, \hat{w}_i^j), j=1..n \},$$

где $\hat{a}^j$ и $\hat{w}^j$ - j -ые компоненты соответствующих n -грамм, представляющие отдельные ВП. При обучении запоминание входных n -грамм в ДАЗУ реализуется динамически с игнорированием близких:

$$\hat{w}_{p+1}^i = \hat{a}_i^i, \text{ если } \exists \hat{w}_i^i: D_n(\hat{w}_i^i, \hat{a}_i^i) \leq r, i=1..p,$$

где p - количество НЭ, уже задействованных к моменту времени t .

Расширяется определение функции автоассоциации, учитывающей радиусы расфокусировки НЭ:

$$F_{ac}(\hat{w}_{(r)}^i) = \{ \hat{w}_{(r)}^i; D_{n-1}[(\hat{w}_1^i, \dots, \hat{w}_i^i), (\hat{w}_j^i, \dots, \hat{w}_j^{n-1})] \leq r_i \}.$$

Рассматривается топология такого СП и показывается, что при отображении последовательностей ВП в гипертрубки СП ДАЗУ нормализуется длительность стационарных участков сигнала, а форма гипертрубки отражает особенности изменения сигнала во времени. Весь эталон речевой единицы представляется в виде графа, образованного в СП пересекающимися гипертрубками - последовательностями автоассоциативно связанных гиперсфер (рис.3). Сформированный эталон представляет собой компактное описание множества реализаций речевой единицы, использованных при обучении, а также всех реализаций, удаленных от них не далее, чем на радиус расфокусировки НЭ. Варьирование размерности СП

(длины регистра сдвига НЭ) и радиуса расфокусировки позволяет изменять форму графа-эталона, определяя количество гиперсфер и автоассоциативных связей между ними.

С целью оптимизации покрытия гиперсферами области эталона в СП при обучении ДАЗУ применяется итеративный алгоритм самоорганизующихся карт Кохонена. Координаты центров гиперсфер сдвигаются к максимумам плотности распределения образов обучающих реализаций в СП:

$$w_i(t+1) = w_i(t) + \varepsilon(a_i - w_i(t)), \text{ если } D_n(a_i, w_i) = \min D_n(a_i, w_j), (\varepsilon < 1).$$

Эталон каждой речевой единицы формируется при обучении в отдельном ДАЗУ. Распознавание на множестве обученных ДАЗУ происходит как ассоциативное вложение траектории, соответствующей входной последовательности ВП сигнала, в область гипертрубки графа-эталона в одном из ДАЗУ. При этом все ДАЗУ работают как единое СП, в котором области эталонов могут частично перекрываться на перцептивно близких участках речевых единиц. Применение механизма внимания, динамически регулирующего радиусы гипертрубок при распознавании, позволяет уменьшить вероятность неоднозначного вложения входной траектории в несколько перекрывающихся гипертрубок и выделить только одну из них.

В завершении главы описывается структурное представление эталонов речевых единиц в многоуровневой ИС ДАЗУ. При этом эталоны более сложных (составных) единиц речи формируются в ДАЗУ на более высоких уровнях ИС как комбинация более простых, представленных в ДАЗУ более низких уровней, соединяемых переходными участками.

Четвертая глава диссертации посвящена разработке метода представления языковых знаний и подхода к анализу речевых сообщений в системах распознавания и понимания связной речи.

В начале главы кратко анализируются особенности задач распознавания и понимания речи, а также недостатки существующих формальных методов, используемых для анализа языковых сообщений. Формируется единый подход к решению обеих задач с опорой на семантическое представление, позволяющий расширить функции систем понимания речи и требующий целостного анализа речевого сообщения как связного текста.

Далее на основе представлений лингвистики описывается целостная организация языковых знаний и механизм анализа связного текста, что иллюстрируется рисунком 4.

В структуре языковой семантики выделяются две взаимосвязанные составляющие целостных значений языковых единиц - грамматические и лексические значения, соответственно слева и справа на рисунке. Лексические компоненты значений представляются корневыми морфемами и их цепочками, а грамматические - флективными морфемами на уровне поверхностного синтаксиса, и классами флективных морфем, эквивалентных с точки зрения передачи основных аспектов значения, а также их цепочками,

на уровне семантического синтаксиса. Объединяясь на верхнем уровне, лексические и грамматические значения образуют семантическую сеть. Элементы семантической сети соответствуют лексическим значениям, связи между элементами отражают смысловые взаимосвязи, а грамматические значения определяют характер связей.

Анализ языкового сообщения, представляющего связный текст, реализуется как динамический процесс его отображения на семантическую сеть, в ходе которого лексические и грамматические компоненты значений языковых единиц актуализируют соответствующие элементы и связи в семантической сети. В ходе этого реализуется глубинное синтаксическое членение текста на сверхфразовые единства (СФЕ) с формированием тематической иерархии, характеризующей смысловую структуру сообщения в целом.

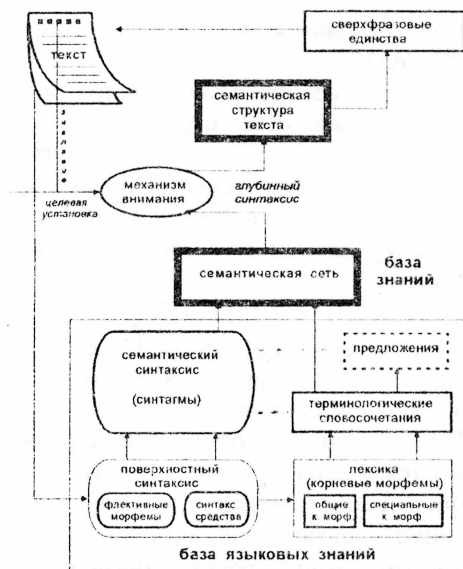


Рис.4.

В основу разработанного метода представления языковых знаний положена идея о том, что все составляющие элементы схемы (рис. 4) - корневые морфемы и их устойчивые сочетания, соответствующие базовым понятиям предметной области, а также флексии и устойчивые конструкции семантического синтаксиса, характеризующие типы отношений между объектами, выражаемыми понятиями, могут быть выделены на основе статистического анализа множества текстов. Связи между понятиями семантической сети могут быть установлены по их совместной встречаемости в рамках структурных единиц текста, в простейшем случае - предложений.

Организовать статистический анализ с выделением повторяющихся элементов текста различной длины и их взаимосвязей позволяет ИС ДАЗУ. В качестве обрабатываемой в ДАЗУ информации выступают последовательности символов (букв). Механизмы обработки в ИС ДАЗУ позволяют отобразить входные тексты в многоуровневый суперграф, аналогично представленному на рисунке 2, и выделить пороговым преобразованием как повторяющиеся элементы всех уровней, так и их связи. При этом словари морфем формируются на основе анализа цепочек букв, с последующей коррекцией экспертом, словари словосочетаний - на основе цепочек корневых морфем, а словари конструкций семантического синтаксиса - на основе цепочек классов, на которые разделяются флексивные морфемы в соответствии с категориями семантического синтаксиса. Семантическая сеть формируется в отдельном ДАЗУ на основе 2-грамм, представляющих пары ассоциативно связанных понятий из лексической ИС ДАЗУ, снабженные гетероассоциативными ссылками к элементам грамматической ИС ДАЗУ.

Каждому элементу семантической сети (понятию) ставится в соответствие численная величина, отражающая вероятность появления соответствующего понятия в текстах предметной области - вес элемента w^*_i . Каждая связь между парой элементов характеризуется величиной, отражающей условную вероятность появления второго в смысловой связи с первым - весом связи w_{ij} . Для оценки весов элементов и связей предлагается итерационный алгоритм:

$$w^*_i(t+1) = f(\sum w_{ji}(t)w^*_j(t)), \quad w_{ij}(t+1) = f(\sum w_{ip}(t)w^*_p(t)w_{pj}(t))$$

$$w^*_i(0) = \ln z_i; \quad w_{ij}(0) = z_{ij}/z_j, \quad f(s) = \sigma(s) = 1/(1+e^{-ks}),$$

где z_i - частота встречаемости понятия в тексте, z_{ij} - частота совместной встречаемости понятий в предложениях текста.

На основе сформированной в соответствии с рисунком 4 базы знаний в ИС ДАЗУ предлагается подход к анализу содержания языкового сообщения как иерархии СФЕ. Грамматическая и лексическая ИС ДАЗУ работают параллельно в режиме распознавания, активируя распознанные элементы в семантической сети. Повышение уровня активации элемента семантической сети происходит также при распознавании ассоциативно связанных с ним элементов, с учетом весов связей, что позволяет учитывать смысловой контекст:

$$x_i(t) = x_i(t-1) + x_{ивх}(t) + \sum w_j w_{ji} x_{jac}(t) - c_{jторм}.$$

Превышение уровнем активации x_i чувствительности НЭ r_i позволяет говорить об отнесении участка сообщения к соответствующей данному элементу теме и выделении СФЕ. Семантическая сеть работает под управлением МВ, варьирующего чувствительности НЭ $r_i(t)$ или постоянные торможения $c_{jторм}(t)$, что повышает устойчивость к отклонениям содержания текста от темы.

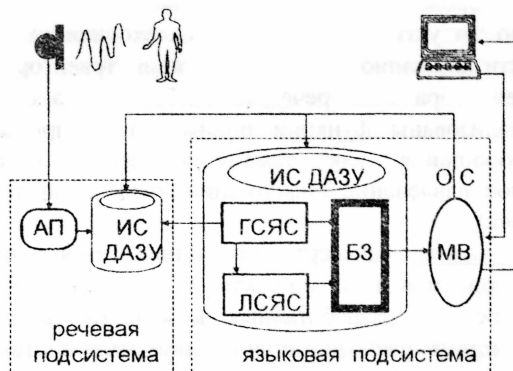


Рис.5.

Целостная структура системы распознавания и понимания связной речи (рис. 5) включает в свой состав две ИС ДАЗУ - речевую и языковую (текстовую), работающие под управлением глобального МВ. Первая ИС ДАЗУ реализует распознавание отдельных речевых единиц в соответствии с алгоритмами, разработанными в главе 3; а вторая осуществляет подстройку параметров ДАЗУ речевой ИС на основе знаний грамматики и семантики предметной области.

Основные элементы системы на рисунке 5: АП - акустический процессор, ГСЯС - грамматическая составляющая языковой семантики, ЛСЯС - лексическая составляющая языковой семантики, БЗ - база знаний (семантическая сеть), МВ - механизм внимания, ОС - обратная связь.

Распознавание осуществляется как соотнесение входной последовательности распознанных речевых единиц с ближайшей из траекторий, представленных в семантической сети. Каждой траектории соответствует множество последовательностей языковых единиц, представленных на более низких уровнях ИС своими лексическими и грамматическими составляющими. Слабо искаженные участки входного сообщения, распознаваемые безошибочно в ДАЗУ речевой ИС, проходя через лексическую и грамматическую ИС, актуализируют соответствующие элементы и связи в семантической сети, что позволяет локализовать внимание только на допустимых продолжениях траектории в семантической сети и соответствующих им лексических и грамматических конструкциях на нижних уровнях речевой ИС. Знание допустимых продолжений сообщения позволяет локализовать внимание на эталонах соответствующих речевых единиц в ДАЗУ речевой ИС, увеличивая вероятность отображения последующего участка сообщения в допустимые элементы языковой ИС, и так далее. Работа двух ИС из ДАЗУ с МВ происходит как разворачивание одной из записанных в семантической сети траекторий в ответ на речевую

сигнал, при котором записанная информация корректирует входную, направляя ее по допустимой траектории, а входная информация снимает неопределенности записанной в местах ветвлений траектории.

В процессе отображения речевого сообщения на семантическую сеть могут быть реализованы функции понимания - определение структуры содержания сообщения как тема-рематической иерархии СФЕ и целенаправленное извлечение информации, относящейся к заданной теме (смысловой поиск).

В пятой главе диссертации описывается экспериментальное исследование алгоритмов ДАЗУ и практическая реализация двух прикладных систем на их основе, разработанных в качестве речевой и языковой подсистем для системы распознавания связной речи: системы распознавания отдельных речевых единиц (слогов и слов) и системы автоматического анализа текстов с формированием языковых баз знаний.

На основе разработанных в третьей главе алгоритмов реализована программная модель одноуровневого ДАЗУ с использованием языка C++ для ПЭВМ IBM PC, а также средства визуализации эталонов в СП ДАЗУ. Для получения векторов параметров на этапе предобработки речевого сигнала использован метод перцептивного линейного предсказания, учитывающий как особенности психоакустического восприятия человека, так и свойства механизма порождения речи. Эксперименты по оценке параметров ДАЗУ проведены на базе 30 первых ударных слогов из частотного словаря, произносимых одним диктором. В ходе предварительных экспериментов оценено влияние параметров ДАЗУ (радиусов расфокусировки НЭ, размерности СП и количества итераций при обучении) на форму траекторий в СП, количество НЭ в эталоне и точность распознавания. При выбранных оптимальных параметрах среднее число НЭ на эталон составило 22, а точность распознавания 91%. При этом применение локального МВ повысило точность распознавания на 4%.

Система распознавания речи реализована в двух вариантах: а) в виде подсистемы распознавания речевых команд для речевого управления на основе цифрового процессора обработки сигналов ADSP-2181, и б) в виде системы ввода текстовой информации с голоса в ПЭВМ на основе ЦП Pentium со звуковой картой Sound Blaster под ОС Windows-95. После обучения на 100 слов, произносимых изолированно одним диктором, точность распознавания составила 95%.

Во второй части главы описывается система автоматического формирования баз знаний на основе анализа корпуса текстов, разработанная в качестве языковой подсистемы для системы распознавания связной речи, и, кроме того, имеющая самостоятельное применение для смыслового анализа и структурирования текстовой информации. Программная модель ИС ДАЗУ и алгоритмы, разработанные в главе 4 диссертации, реализованы на ПЭВМ IBM PC с использованием языка C++ для ОС Windows-95.

В результате автоматического анализа корпуса текстов системой формируется база языковых знаний в соответствии со схемой, приведенной на рисунке 4 (за исключением уровня семантического синтаксиса). Семантическая сеть представляется выделенными из текстов значимыми понятиями (словами и словосочетаниями) с их ассоциативными взаимосвязями. Понятия и их взаимосвязи ранжируются по весам, а также снабжаются отсылками в соответствующие места исходных текстов. Сформированная семантическая сеть может быть использована в качестве верхнего уровня в системах распознавания и понимания речи.

В приложениях приводятся таблицы и рисунки, иллюстрирующие результаты экспериментов с моделью ДАЗУ, а также 3 документа, подтверждающие внедрения результатов диссертационной работы.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

1. Использована наиболее адекватная задаче распознавания речи нейросетевая структура - иерархическая структура (ИС) из динамических ассоциативных запоминающих устройств (ДАЗУ). ИС ДАЗУ способна автоматически организовать структурное иерархическое представление информации. Развита теория ИС ДАЗУ и модифицирована структура нейроподобного элемента;

2. Для управления ИС ДАЗУ в режиме распознавания использован механизм внимания, реализующий динамическую подстройку параметров ДАЗУ на основе предсказания ожидаемой информации на входе;

3. Разработан метод формирования представления эталонов речевых единиц в ДАЗУ на основе покрытия областей отображения речевых сигналов в сигнальное пространство множеством связанных гиперсфер;

4. Проведен анализ лингвистических представлений о семантической организации языка и текста, на основе которого определена структура представления языковых знаний в системах распознавания и понимания речи, а также принципы анализа связного текста;

5. Разработан метод автоматического формирования представления языковых знаний семантического уровня на основе статистического анализа множества текстов предметной области в ИС ДАЗУ, учитывающего веса взаимосвязей элементов в рамках структурных единиц текста;

6. Предложен подход к анализу содержания языкового сообщения как связного текста с опорой на семантическое представление на основе ИС ДАЗУ, управляемой механизмом внимания;

7. Предложена целостная структура системы анализа (расознавания и понимания) связной речи, включающей в свой состав речевую и языковую подсистемы, на основе двух ИС ДАЗУ, управляемых глобальным механизмом внимания;

8. На основе разработанных методов реализованы две прикладные системы в качестве речевой и языковой подсистем для системы распознавания связной речи:

- система распознавания отдельных речевых единиц;
- система автоматического анализа текстов с формированием базы знаний.

Основное содержание диссертации изложено в следующих работах:

1. Харламов А.А., Ермаков А.Е. Динамическая нейронная сеть для программного представления эталонов речевых событий // Проблемы нейрокибернетики. Материалы XI Международной конференции по нейрокибернетике. - Ростов-на-Дону, 1995. - С. 272-273.

2. Харламов А.А., Ермаков А.Е., Рябых Д.А. Иерархические структуры из динамических нейронных сетей для распознавания речевых сигналов и представления языковых знаний. // Интеллектуальные системы: Труды второго международного симпозиума. - М., 1996. - Т.2. - С. 143-146.

3. Система распознавания речи на основе динамических нейронных сетей: Научно-технический отчет по теме «Интеллектуальные системы» / МГТУ им. Н.Э. Баумана. Рук. темы Пупков К.А., исп. Ермаков А.Е. ГР N 01.970003063, инв N 02.970002081. - М., 1997. - С. 200-217.

4. Автоматический анализ текста на основе иерархических структур из динамических нейронных сетей: Научно-технический отчет по теме «Интеллектуальные системы» / МГТУ им. Н.Э. Баумана. Рук. темы Пупков К.А., исп. Ермаков А.Е., Харламов А.А. ГР N 01.970003063, инв N 01.970002081. - М., 1997. - С. 218-234.

5. Харламов А.А., Ермаков А.Е. Распознавание речи на основе динамических нейронных сетей, управляемых механизмом внимания // Информационные технологии. -1997. - N 9. - С. 22-27.

6. Харламов А.А., Ермаков А.Е. Динамическая нейронная сеть для распознавания речевых сигналов // Вестник МГТУ им. Н.Э. Баумана. - 1998. - N 1, с.18-23

7. Харламов А.А., Ермаков А.Е., Кузнецов Д.М. TextAnalyst - комплексный нейросетевой анализатор текстовой информации // Вестник МГТУ им. Н.Э. Баумана. - 1998. - N 1, С. 32-36

8. Харламов А.А., Ермаков А.Е., Кузнецов Д.М. Технология обработки текстовой информации с опорой на семантическое представление на основе иерархических структур из динамических нейронных сетей, управляемых механизмом внимания // Информационные технологии. -1998. -N2. - С.26-32.