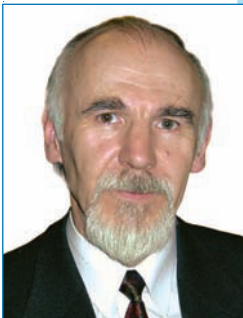




Исследования методов речевого диалогового управления в МВТУ – МГТУ им. Н.Э. Баумана

В.А. Суханов,
кандидат технических наук,



Ю.Н. Жигулёвцев,
кандидат технических наук

Памяти Вадима Николаевича Плотникова посвящается

Разработки в области речевых технологий на кафедре «Системы автоматического управления» МВТУ им. Н.Э. Баумана, основанной и возглавлявшейся В.В. Солодовниковым, начались в 1968 году по заказу ЦКБЭМ (ныне РКК «Энергия»). Целью работы было создание систем речевого управления пилотируемыми космическими аппаратами (ПКА). Это было время реабилитации кибернетики, а поскольку теория автоматического управления является одним из разделов этой науки, привлечение кафедры, готовящей специалистов по управлению летательными аппаратами, для решения поставленной задачи было вполне логичным, тем более, что многие видные специалисты в области управления, например Ю. Ту, Я.З. Цыпкин и др., внесли заметный вклад и в развитие теории распознавания образов. Работы проводились в лаборатории моделирования ЦКБЭМ, начальником которой был Е.И. Харченко. Лаборатория относилась к отделу под руководством О.Н. Воропаева, входившему в состав комплекса, возглавлявшегося Б.Е. Чертоком.

Сформировал и возглавил группу речевиков на кафедре В.Н. Плотников. Вначале в неё входили авторы настоящей статьи, а также Б.А. Бурак и Г.О. Жгутов, в последующем в работах одновременно принимали участие до 10 человек, включая инженеров, научных сотрудников, аспирантов и студентов. Аспирантуру по речевому направлению закончили А.А. Харламов, В.А. Кононенко, А.В. Белинский, М.А. Абдуллаев, И.В. Воронов, В.И. Спорыш, М.Н. Поклад, В.А. Гусев, Д.А. Кушнир. В силу изначальной заданности прикладной задачи собственно речевые исследования носили подчинённый характер, служа средством достижения практических результатов по управлению. Поэтому наряду с разработками распознающих устройств большое внимание уделялось исследованию вопросов организации речевого управления, включая роль и место оператора в таких системах, учёт особенностей предметной области и т.п. С другой стороны, прикладная сфера предъявляет предельно жёсткие требования к надёжности во всех смыслах — как аппаратной, так и распознавательной, а также к конструктивным характеристикам аппаратных средств (устойчивость к воздействиям, ограничения весов и габаритов и т.п.). В связи с этим большое внимание уделялось разработке специализиро-

ванных технических средств на самой современной элементной базе. Например, как только в 1972-м году появились опытные образцы отечественных интегральных операционных усилителей, они были использованы для реализации блока выделения информативных параметров (БВИП) в виде печатной платы размером 120x180 мм. Блок состоял из двух фильтров четвёртого порядка, выделявших полосы первых двух формант, двух выпрямителей, двух усилителей-ограничителей с формирователями импульсов и пяти сглаживающих фильтров шестого порядка. Другой пример — применение первой отечественной однокристалльной микро-ЭВМ 1801BE1 и одноплатной микро-ЭВМ «Электроника НЦ-8001» на её основе в 1981 году. В 1994 году были приобретены университетские комплекты отладочных средств на основе двух семейств цифровых сигнальных процессоров фирмы Analog Devices с фиксированной и плавающей точкой.

Первые результаты по речевому управлению движущимся объектом были получены в 1970 г. с использованием аналоговой вычислительной машины ЛМУ-1. Распознавание осуществлялось по частоте переходов через ноль на основе линейных решающих правил. Комбинация знаков сигналов на выходах двух линейных пороговых элементов определяла одно из четырёх распознаваемых слов и управляла двумя электродвигателями гусеничной тележки, обрабатывая команды «вперёд-назад», «направо-налево». Конечно, этот макет носил скорее «пиаровский» характер и был, по существу, воплощением дипломных проектов, однако позволил убедиться в возможности решения поставленной задачи.

Отработка алгоритмов распознавания речи проводилась на аналого-цифровом вычислительном комплексе на основе ЭВМ М-222 с блоком аналого-цифрового преобразования УП-1 и упомянутым выше БВИП. В последующем был организован обмен данными М-222 с БЭСМ-6, что позволило существенно расширить возможности исследований и перейти к экспериментам по речевому управлению. Основой для этих экспериментов послужил стенд полнатурного моделирования, созданный в ЦКБЭМ первоначально для тренировок по посадке пилотируемого спускаемого аппарата на лунную поверхность. После прекращения лунной программы он был переоборудован для отработки алгоритмов управления сближением и стыковкой с орбитальной станцией, а также моделирования скафандра с автономной двигательной установкой — прототипа применяемых в настоящее время на космической станции.

Визуальная имитация движения осуществлялась на экране телевизионного монитора, изображение на который передавалось с телекамеры, перемещающейся по шести степеням свободы (три линейных перемещения вдоль осей в прямоугольной системе координат и три вращения вокруг этих осей). Управление телекамерой выполнялось следящими системами по сигналам управления, которые формировались моделирующим комплексом, воспроизводящим модель движения объекта. Перед телекамерой помещался макет внешней среды, в результате при перемещении телекамеры на экране монитора имитировалось движение объекта. Сигналы управления движением вводились в модель объекта на основании распознавания речевых команд, заменяющих в данном случае команды от органов ручного управления. Одновременно результат распознавания отображался на транспарантах блока отображения информации, встроенного в телевизионный монитор. Таким образом, осуществлялся двойной визуальный контроль выдачи команд — на табло и по отображаемому на экране движению объекта. Начиная с 1979 года и до начала 90-х годов на этом стенде проведено несколько серий экспериментов по речевому управлению движением различных ПКА, доказавших возможность, а в некоторых случаях необходимость применения речевых технологий для управления движущимися объектами. Вначале в качестве распознающей подсистемы использовался программный комплекс на М-222 [1]. В дальнейшем были реализованы макеты системы речевого диалогового управления, получившей имя «Сириус» (в частности, по созвучию с аббревиатурой РИУС — речевая информационно-управляющая система) на основе применения микро-ЭВМ «Электроника

НЦ-8001», «Электроника-60» и ДВК. Разработаны были и новые блоки выделения признаков и алгоритмы распознавания, позволившие системе «Сириус» принять участие в сравнительных испытаниях отечественных систем распознавания речи, проводившихся в 1984 году на ВДНХ [3].

Результаты исследований и разработок докладывались на заседании секции управления движением и навигации АН СССР в 1985 году, на котором присутствовали представители КБ имени Арт. Ив. Микояна. В результате было положено начало циклу разработок средств речевого управления для перспективных высокоманевренных самолётов, к участию в которых был привлечён ещё ряд научных коллективов, в том числе под руководством Н.Г. Загоруйко, а также предприятий авиационной отрасли, в частности НИИ Авиационного оборудования. На полунатурных стендах названных предприятий проводились исследования методов речевого диалога пилотов с бортовым оборудованием, при этом в подсистеме синтеза речи применялся синтезатор «Фонемафон-3».

К началу 90-х годов речь шла о переходе к опытно-конструкторской разработке системы речевого управления для автономного скафандра. К сожалению, этим планам не суждено было сбыться в связи с известными процессами. В течение целого ряда лет речевые исследования ничем не поддерживались и продолжались в академическом плане преимущественно на основе научно-педагогических интересов и частично утилитарных разработок типа телефонного робота-информатора, системы аварийно-пожарного оповещения для электропоездов, говорящих приборов для слепых и т.п.

Опыт и результаты, полученные авторами при проведении исследований различных аспектов задач анализа, распознавания и синтеза речевых сообщений/сигналов, нашли отражение в публикациях [1,2] и позволили определить наиболее перспективные направления и новые задачи исследования речевых технологий. Была сформулирована проблема организации диалога в процессе речевого общения с техническими средствами — речевого диалогового управления (РДУ). Ориентиром послужило то, как подобный диалог осуществляется в «идеальной диалоговой системе» — общении двух или более операторов, выполняющих общую задачу. А именно: человек в процессе общения с другим человеком может распознавать и понимать речевые сообщения только в том случае, если они общаются на одном (известном им) языке и используют понятия (о теме диалога, о предметной области), известные им обоим, в акустической среде, допускающей ведение диалога. Таким образом, в систему, реализующую РДУ (СРДУ), «закладывается» определённая совокупность знаний, представленных в виде модели языка, на котором ведётся общение (язык описывается как с точки зрения восприятия и понимания, так и синтеза речевых сообщений), модели предметной области (о которой ведётся диалог в процесс управления), и модели акустической среды, в которой осуществляется диалог. Можно считать, что термин РДУ служит обобщением термина СПР (система понимания речи), предполагая не только восприятие смысла сказанного, но и формирование адекватной ответной реакции в рамках соответствующей предметной области.

Учитывается, что модель предметной области для конкретных устройств и систем, используемых для управления, ограничена совокупностью функций, выполняемых ими. Отсюда следует, что и словарь, используемый для формирования речевых сообщений, также ограничен. Аналогично выявляются ограничения

на стиль и темп речи. Введенные в СРДУ знания о предметной области позволяют использовать более сложные команды-фразы (учёт прагматики), а также формировать сценарии диалогов, классифицировать входную информацию по нескольким типам, соответственно разбивать на группы словарь. В итоге — возможность повысить надёжность распознавания до приемлемого для технических систем уровня.

В качестве центральной задачи, решение которой определяет потенциальные возможности и практическую ценность СРДУ, выделена разработка внутреннего языка представления разнородных знаний и механизма оперирования ими. Именно это определяет способность СРДУ использовать в нужный момент необходимые (релевантные) знания, является показателем интеллектуальности СРДУ.

Предлагаемый подход к построению СРДУ базируется на С-технологии, или технологии систем, основанных на знаниях, которая характеризуется тем, что знания системы должны быть выражены в явном виде, к ним необходимо обеспечить прямой доступ и возможности их постепенного (в том числе и во время работы системы) наращивания и модификации [4, 5]. Определённая совокупность знаний «приобретается» в процессе обучения СРДУ.

База знаний СРДУ строится как сетевая структура, представляющая собой интеграцию источников знаний, двудольного графа (сети Петри с управляемыми переходами) и дискретной динамической системы. Особенно важно это для представления знаний о предметной области, так как сети Петри описывают не только сам процесс, но и управление этим процессом [5, 6].

Компонентам сети сопоставлены элементарные источники знаний, в частности, отдельные продукционные правила (обобщенно — источники знаний — ИЗ). В отдельных подсетях располагаются совокупности ИЗ, представляющие собой поуровневое описание языка, предметной области и акустической среды. Исключительно важным свойством такой модели является однотипность организации взаимодействия ИЗ как на её отдельных уровнях, так и между уровнями. Этим обеспечивается возможность применения высших уровней для формирования адекватной реакции СРДУ на целенаправленное речевое или другое поведение.

Общность подхода обеспечивает возможности моделирования разнообразных процессов, обладающих асинхронностью, параллелизмом, недетерминированностью и т. п. Именно эти свойства характеризуют процессы речевого взаимодействия и должны быть присущи перспективным системам РДУ. Модульный принцип формирования базы знаний позволяет добавлять, удалять или изменять отдельные ИЗ независимо от других, что обеспечивает возможность перехода к автоматизированной процедуре проектирования. Отдельные ИЗ связаны между собой только через поток данных, которые они обрабатывают. Представление ИЗ в формате «Что делать и когда» делает их естественными с точки зрения здравого смысла.

Следует подчеркнуть перспективность подобного подхода и для построения «обычных» систем распознавания речи, поскольку повышения их характеристик до уровня, приемлемого для практического применения, можно добиться лишь за счёт максимально полного учёта всех знаний о речи и речевом взаимодействии. Результаты последних лет в области речевых технологий всё с большей очевидностью указывают на ограниченность подходов, связанных исключительно со статистическим моделированием процессов речевого общения. Этот экстенсивный путь, поддерживаемый бурным ростом производительности (быстродействия и объёмов памяти) вычислительных систем, позволил добиться значительных результатов в улучшении характеристик речевых систем, однако в последнее время всё ча-

ще высказывается мнение о необходимости продолжения исследований и разработок, ориентированных на концепции искусственного интеллекта [7, 8].

В плане фундаментальных исследований проблем речевого общения основной и всеобъемлющей задачей является построение базы знаний о речи, интегрирующей совокупность декларативных и процедурных источников знаний, опирающихся на представительную, многодикторную, многостилевую и т.д. базу речевых данных. Очевидно, что решение такой задачи не под силу одному отдельно взятому научному коллективу. Только объединение усилий всего научного сообщества в рамках масштабной исследовательской программы с поддержкой государственных и коммерческих структур поможет осуществить прорыв в области речевых технологий. Хотелось бы, чтобы появление профильного отечественного журнала внесло свой вклад в решение этой проблемы.

Литература:

1. Плотников В. Н., Суханов В. А., Жигулёвцев Ю. Н. Речевой диалог в системах управления. — М.: Машиностроение, 1988. — 244 с., ил.
2. Плотников В. Н., Белинский А. В., В. А. Суханов, Жигулёвцев Ю. Н. Цифровые анализаторы спектра. — М.: Радио и связь, 1990. — 184 с.
3. Потапова Р.К. Речевое управление роботом. — М.: Радио и связь, 1989. — 248 с.
4. Искусственный интеллект: В 3 кн. Кн.1. Системы общения и экспертные системы: Справочник / Под ред. Э.В. Попова. — М.: Радио и связь, 1990.
5. Плотников В.Н., Суханов В.А. Системы, основанные на знаниях. — М.: Изд-во МГТУ, 1995. — 88 с., ил.
6. Котов В.Е. Сети Петри. — М.: Наука, 1984. —160 с.
7. Potapova R.K. The Knowledge Based Speech-Input Expert System for Russian // Proc. of Intern. conf. «Speech and Computer» (SPECOM'99). Moscow, 1999.
8. Chin-Hui Lee. Fundamentals and Technical Challenges in Automatic Speech Recognition. — The XIIIth International Conference «Speech and Computer» SPECOM'2007, 15-18 October 2007, Moscow, Russia, p. 25.

Суханов Владимир Александрович,

кандидат технических наук, старший научный сотрудник, доцент МГТУ им. Н.Э.Баумана. Окончил в 1968 г. МГТУ им. Н.Э.Баумана по специальности «Системы автоматического управления». Область научных интересов: Системы, основанные на знаниях (системы речевого диалогового управления, экспертные системы). Автор более 70 научных публикаций, 6 авторских свидетельств на изобретения, соавтор 3 монографий.

Жигулёвцев Юрий Николаевич,

кандидат технических наук, старший научный сотрудник, доцент МГТУ им. Н.Э.Баумана. Окончил в 1969 г. МГТУ им. Н.Э.Баумана по специальности «Системы автоматического управления». Область научных интересов: Методы и средства построения систем речевого взаимодействия. Автор более 80 научных публикаций, 6 авторских свидетельств на изобретения, соавтор 2 монографий.