Об исследованиях проблемы речевых технологий

Н.Г. Загоруйко,
доктор технических наук

Эта работа была опубликована в институтском сборнике трудов более 10 лет назад [1] в самый острый момент перестроечного разрушения. Прошли годы тяжёлого восстановления, и вновь настали времена, не самые лучшие для финансирования науки. Так что минорный тон текста введения, пожалуй, не требует существенных изменений и далее приводится в прежнем виде.

Та часть статьи, которая касается истории исследований речи в нашей стране в период существования АРСО, по понятным причинам и не должна меняться. Что касается раздела с описанием прикладных систем, то он, конечно, устарел, но, к сожалению, не настолько сильно, как этого хотелось бы. Многие высказанные идеи либо не реализованы, либо реализованы не полностью. Так что список возможных сценариев использования речевых систем остаётся достаточно актуальным. Отдельные части этого раздела можно было бы осовременить, но я не считаю вправе делать это: в годы перестройки наш коллектив речевиков Института математики Сибирского отделения переключился на другие задачи распознавания образов, и сейчас я не считаю себя достаточно компетентным в области современных речевых исследований и разработок. Я надеюсь, что речевикам «со стажем» мои записки напомнят незабываемые времена расцвета движения АРСО, а молодым исследователям речи, может быть, будет интересно узнать одну из страниц истории своей науки. По указанным выше причинам я решил ничего не менять в старой статье. Далее следует её текст.

Сокращение бюджетных средств, выделяемых на научные исследования, сказывается в первую очередь на таких фундаментальных научных направлениях, развитие которых требует координации усилий специалистов нескольких научных дисциплин и потребность в прикладных результатах от которых ещё недостаточно осознаётся потенциальными потребителями. Отсутствие централизованных средств не позволяет организовать серьёзную координацию работы коллективов разного профиля и разной ведомственной принадлежности, а ориентация потенциальных инвесторов на сиюминутную прибыль не позволяет финансировать разработки для рынка отдалённого будущего.

По этим причинам в числе серьёзно пострадавших находится и проблема исследований речевой коммуникации между человеком и техническими системами. Среди современных научно-технических проблем трудно найти проблему такой же сложности. Для создания машин, свободно понимающих человеческую речь, требуется, во-первых, воспроизвести техническими средствами механизмы работы слуховой системы человека. Для этого физикам, биофизикам, физиологам и психоакустикам нужно понять хотя бы, как работает слуховая мембрана, которая обеспечивает необъяснимо тонкую спектральную избирательность, причём с таким малым временем реакции, которое не согласуется с известным соотношением неопределённости.

Далее, психоакустикам совместно с фонетистами и лингвистами следует построить модели преобразований потока нервных импульсов от периферии слуховой системы в последовательность фонетически значимых признаков, фонем, слогов, слов и фраз речи, которая звучит в исполнении разных дикторов, с разной громкостью, в разном темпе и в различных акустических условиях.

Затем лингвистам вместе со специалистами в области искусственного интеллекта нужно описать процесс восприятия речи на уровне синтаксиса, семантики и прагматики.

Далее, математикам и электронщикам на базе указанных выше знаний предстоит построить программно-аппаратные комплексы для автоматического восприятия речевых сообщений без ограничений на объём словаря, индивидуальный тембр голоса и дикцию, эмоциональное состояние диктора, акустические помехи и т.д.

Наконец, системотехникам, эргономистам и психологам нужно разработать такую технологию использования речевых систем, чтобы это использование было экономически оправданным и психологически комфортным (дружественным) для пользователя.

СССР и Россия располагали высококвалифицированными научными коллективами, работавшими во всех указанных направлениях. Существовала система координации работ и оперативного обмена свежими научными результатами. Сейчас всё это практически разрушено. Значительная часть речевых коллективов распалась, оперативные связи между оставшимися отсутствуют. Фундаментальные исследования фактически прекращены, продолжаются лишь разработки ограниченных по своим характеристикам технических систем для конкретных заказчиков на базе ранее полученных научных результатов. Можно лишь надеяться, что слабый ручеёк речевых разработок не прервётся совсем и настанут времена, когда наряду с другими фундаментальными проблемами начнёт возрождаться и проблема речевого взаимодействия человека с машиной.

Краткая история проблемы АРСО

Приятно осознавать, что первая серьёзная попытка построить систему автоматического распознавания речи была предпринята в нашей стране. В 1941–1942 гг. в блокадном Ленинграде Лев Леонидович Мясников завершил свою докторскую диссертацию по работе, связанной с системой распознавания изолировано произносимых звуков — всех гласных и некоторых согласных. (Этот впечатляющий факт позволяет надеяться, что интерес к речевой коммуникации поможет проблеме АРСО пережить и нынешние трудные времена.) Работы Мясникова Л.Л. на 10 лет опередили первые зарубежные работы по распознаванию речи.

В послевоенные годы речевые исследования велись, в основном, в интересах уплотнения каналов связи. При этом нужно выявить наиболее информативные характеристики речевого сигнала, экономно закодировать их на входе и по кодам восстановить качественную речь на выходе линии связи. Естественно, возникал вопрос: нельзя ли по этим же характеристикам распознавать звуки, слова, фразы? Если бы это удалось, то по линиям связи можно бы передавать только телеграфные коды распознанных фонем. На этой основе в Ленинградском НИИ дальней связи начал работать сильный научный коллектив под руководством Л.А.Варшавского и В.С.Федоровича. В этом же коллективе или в тесном сотрудничестве с ним начинали свои исследования многие будущие лидеры советских речевиков: Л.А. Чистович, Л.В. Бондарко, В.И. Галунов, М.Ф. Деркач и др.

Резкий подъём интереса к проблеме распознавания и синтеза речи как у нас, так и за рубежом отмечается в начале 60-х годов в связи с развитием вычислительной техники. Компьютерные фирмы осознали, что в перспективе машина должны включать в число средств общения с человеком и наиболее удобный для него вид — речевой диалог. Решить эту проблему было очень заманчиво и казалось не очень трудно. Иллюзия лёгкости решения проблемы была у всех, кто был мало знаком с ней. Один известный кибернетик, не занимавшийся речевыми сигналами, после доклада о скромных результатах исследований сказал: «То, что наши речевики не могут научить машину распознавать речь, — это недоразумение. Человек пользуется речью миллионы лет, а они не могут понять, как это делается. Надо будет поговорить со знакомыми немецкими акустиками, они наверняка это могут делать».

Под воздействием социального заказа и иллюзии лёгкости проблемы над речевыми сигналами начали работать тысячи научных коллективов во многих странах мира. В СССР таких коллективов в 70-е годы насчитывалось более 150. По мере углубления в работу обнаруживались всё новые трудности, но одновременно проблема открывалась всё новыми и новыми интереснейшими гранями, в результате чего речевые коллективы не прекращали своей работы и стали появляться важные фундаментальные научные результаты. До конца 60-х годов, когда основные результаты касались речевой психоакустики и фонетики, советская научная школа прочно занимала лидирующие позиции в мире.

Многие советские речевики начинали знакомство с проблемой по книге Сапожкова М.А. «Речевой сигнал в кибернетике и связи» (1963 г.).

Выдающийся вклад в становление серьёзных речевых исследований в нашей стране, кроме отмеченного выше НИИ дальней связи, внесли учёные Института физиологии АН СССР им. академика Павлова (Чистович Л.А., Кожевников В.А., Люблинская В.В. и др.). Их коллективная монография «Речь. Восприятие и речеобразование» (1968 г.) была настольной книгой всех, кто изучал физиологию и психоакустику речи.

Проблемы физиологической акустики были глубоко представлены в работах Акустического института РАН (Дубровский Н.А., Бибиков Н.Г.).

Фонетические и лингвистические проблемы речевой коммуникации плодотворно изучались коллективами Ленинградского, Московского и Одесского университетов, Московского института иностранных языков (Бондарко Л.В., Вербицкая Л.А., Златоустова Л.В., Бровченко Т.А., Нушикян Э.А., Потапова Р. К., Торсуева И.Г. и др.).

Исследования особенностей индивидуальной дикции интенсивно велись в НИИ дальней связи (Галунов В.И., Бром Н.С., Коваль С.Л. и др.) и в Тбилисском институте автоматики и электроники (Какауридзе Г.А., Ромишвили Г.С., Тушишвили М.А., Сердюков В.Д. и др.).

Инженерно-математические разработки речевых систем велись во многих организациях. Наиболее многочисленные коллективы в этом направлении работали в Новосибирском институте математики и Новосибирском университете (Загоруйко Н.Г., Волошин Г.Я., Величко В.М., Кельманов А.В., Тарабунов И. и др.), Киевском институте кибернетики (Винцюк Т.К., Людовик Е.К., Богино В.И. и др.), Минском институте технической кибернетики (Лобанов Б.М., Дегтярев Н.П., Панченко Б.В.), Московском вычислительном центре АН СССР (Трунин-Донской В.Н., Чучупал В.Я. и др.) и Московском институте проблем передачи информации (Турбович И.Т., Цеммель Г.И., Сорокин В.В., Книппер А.В. и др.).

Большой вклад в разработки речевых проблем внесли сотрудники Московского института связи (Пирогов А.А., Прохоров Ю.Н., Акинфиев Н.Н. и др.), МВТУ им. Баумана (Жигулевцев Ю.Н., Плотников Ю.Н. и др.), ЦНИИПИАС (Фролов Г.Д. и др.), ЦНИИ электронного машиностроения (Петров Г.М., Копейкин А.Б.), Ленинградского НПО «Аврора» (Петров А.Н., Туркин В.Н.), Пензенского НИИ электросвязи (Голубцов С.В., Белявский В.Н.), Ижевского политехнического института (Гитлин В.Б., Сметанин А.М. и др.), Львовского университета (Деркач М.Ф., Гумецкий Р.Я. и др.), Вильнюсского политехнического института (Кемешис П.П., Рудженис А.И.), Таллинского института технической кибернетики (Кюннап Е.Ю., Рохтла М., Отт А. и др.), Ереванского политехнического института (Григорян А.А., Закарян А.Б. и др.) и многих других организаций.

Координация работ осуществлялась по официальным каналам — вначале через Секцию речи Комиссии по акустике при Президиуме АН СССР (председатель Галунов В.И.), а затем через Совет по распознаванию и синтезу речи при Президиуме АН СССР (председатель Журавлев Ю.И.).

Однако более важную роль играл неофициальный орган — Всесоюзная школа-семинар по проблеме Автоматического распознавания слуховых образов (АРСО), которая собиралась ежегодно, а затем раз в два года в период с 1965 по 1992 годы. Идея АРСО родилась во время узкого рабочего совещания в 1963 году в Новосибирске, в котором участвовали Бондарко Л.В., Волошин Г.Я., Голубцов С.В., Загоруйко Н.Г., Кожевников В.А. и Чистович Л.А. Участники совещания, представляющие коллективы математиков, инженеров, лингвистов и физиологов, пришли к выводу, что обсуждение проблемы речевой коммуникации в таком составе было исключительно полезным для понимания каждым из участников проблемы в целом. Было решено организовать школу-семинар, на которой в «школьном» разделе ведущие специалисты разных аспектов проблемы делали бы обзорные доклады и читали учебные лекции, а в «семинарском» разделе участники делали бы сообщения о результатах своих последних исследований. Организацию АРСО брали на себя различные центры речевых исследований.

Автору, которому довелось быть председателем Программного комитета и открывать заседания всех семнадцати школ-семинаров, доставляет большое удовольствие напомнить города и годы их проведения: АРСО–1 (Новосибирск, 1965), АРСО–2 (Тракай, 1966), АРСО–3 (Новосибирск, 1967), АРСО–4 (Киев–Канев, 1968), АРСО–5 (Сухуми, 1969), АРСО–6 (Таллин, 1971), АРСО–7 (Алма-Ата, 1973), АРСО–8 (Львов, 1974), АРСО–9 (Минск, 1976), АРСО–10 (Тбилиси, 1978), АРСО–11 (Ереван, 1980), АРСО–12 (Одесса, 1982), АРСО–13 (Новосибирск, 1984), АРСО–14 (Каунас, 1986), АРСО–15 (Таллин, 1989), АРСО–16 (Суздаль, 1991), АРСО–17 (Ижевск, 1992).

АРСО была центральным событием в жизни речевых коллективов, регулярным смотром их последних достижений. В отличие от больших конференций с разноплановой тематикой, на которые каждый раз собирались новые люди, АРСО сохраняла высокую стабильность состава основных участников, чему способствовал практически постоянный состав программного комитета. «АРСОшники» хорошо знали друг друга, что исключало возможность недобросовестной рекламы сомнительных результатов. Как разработчики, так и присутствовавшие обычно на АРСО заказчики речевых систем имели возможность по докладам и дискуссиям составлять объективные суждения об уровне работ и возможностях того или иного коллектива. По итогам, АРСО речевики могли корректировать направление своих исследований. По существу АРСО играла важную роль неформального, но эффективно работающего Всесоюзного центра по координации исследований в области речевых технологий. Остаётся только сожалеть, что нынешние обстоятельства разрушили это уникальное научно-организационное явление.

Характерная черта проблемы АРСО, необычная для многих других проблем, состоит в том, что по мере проникновения в неё обнаруживаются всё новые слои нерешённых вопросов возрастающей сложности. На каждом этапе проблема становилась более трудной, чем представлялось раньше. Сейчас достаточно очевидно, что полное решение проблемы АРСО по времени совпадёт с решением проблемы создания искусственного интеллекта, не уступающего человеческому.

Очень показательны прогнозы, которые давали советские специалисты-речевики относительно возможных сроков решения разных проблем распознавания речи. Опрос экспертов был проведён нами трижды: в 1967, 1977 и 1988 гг. Один из вопросов касался сроков решения следующих четырёх задач: распознавание с надёжностью 98% речи любого диктора без подстройки на его голос при полном стиле произношения изолированных команд в обычном помещении при объёмах словаря 20, 200, 2000 слов и слитной речи на базе словаря в 2000 слов. Среди экспертов были оптимисты, считавшие, что системы на 20 и 200 слов уже практически имелись ещё в 1967 году, были и пессимисты, откладывавшие решение этих задач на начало XXI века. Результаты этих опросов в виде математического ожидания оценок года, в котором, по мнению экспертов, будет решена соответствующая проблема, приведены в таблице:

Год опроса	20 слов	200 слов	2000 слов	Слитная речь
1967	1969	1971	1977	—
1977	1980	1984	1988	1994
1988	1993	2000	2008	2029

Как видно из таблицы, первоначальный оптимизм сменился осторожностью, а затем и явным пессимизмом. Но прошло 7 лет со времени последнего опроса — и снова видно, что оценки,

казавшиеся пессимистичными, на самом деле были оптимистичными. Систем, которые надёжно работали бы с любым диктором «с улицы», причём не в лабораторных, а в нормальных производственных условиях, хотя бы со словарём в 20 слов, пока нет ни у нас, ни у японцев, ни у американцев. Имеются лабораторные системы, хорошо распознающие большие словари и даже слитно произносимые фразы, однако преодолеть барьер инвариантности к диктору, акустическим условиям на рабочем месте, особенностям входного электроакустического тракта пока не удалось. Исследования в этом направлении (теперь уже практически без нас) ведутся во всём мире, причём расходы на них с каждым днём увеличиваются. В качестве примера можно привести фирму Dragon Systems UK Ltd, в которой 130 сотрудников работают над развитием перспективной системы Dragon Dictate, распознающей речь на базе словаря из 60 000 слов (пока не в реальном масштабе времени). Помимо собственных средств, эта компания получила грант от правительства США на сумму 7 миллионов долларов. Так что не следует терять надежды на то, что полезные для практического применения речевые системы будут созданы в обозримом будущем. И теперь становятся более актуальными, чем раньше, проблемы эргономического и психологического характера. Как разработчикам, так и потенциальным пользователям важно знать как можно большее число возможных областей применения систем распознавания и синтеза речи.

Ниже описываются некоторые из возможных сценариев применения речевых систем, находящихся на разных стадиях реализации. Некоторые системы представлены действующими опытными образцами, некоторые находятся в разработке, но значительная часть систем ещё не разрабатывалась.

Сценарии применения речевых систем

Чтобы оценить важность речевой коммуникации в жизни людей, достаточно представить себе глухонемое человечество. В нормальной ситуации речевая связь, в отличие от визуальной или тактильной, может использоваться вне прямой видимости, в темноте, на определённом удалении от терминала, в условиях, когда руки заняты другим делом, и т.д. Как показали американские исследования поведения пилотов современного самолёта, плотно насыщенного приборами и другими средствами отображения ситуации, среднее время реакции лётчика на звуковые сигналы на разных режимах полёта находится в пределах от 2,8 до 3,0 сек., а на световые сигналы — от 7,1 до 128,3 сек. Отмечен случай, когда пилот в течение почти 60 сек. не замечал светового сигнала, сообщавшего о пожаре двигателя.

Другими исследованиями показано, что ввод информации в машину через речевой сигнал занимает в несколько раз меньше времени, чем при использовании клавиатуры. Число ошибок оператора при управлении машиной с помощью устных команд также значительно меньше, чем при использовании кнопок, особенно если число этих кнопок (команд) увеличивается. Все эти факты подтверждают большие перспективы применения средств речевой технологии для общения человека с техническими системами.

Варианты применения речевых систем будем излагать в порядке возрастания объёма распознаваемого словаря W. При этом будем иметь в виду так называемый «коэффициент ветвления» K, указывающий на то, сколько слов нужно распознать на каждом отдельном этапе речевого диалога или управления.

Описывая варианты системы, будем указывать, возможна ли подстройка под диктора или нет. При этом будут иметься в виду разные способы подстройки под диктора: централизованный, автономный и дискетный. При централизованном способе средства подстройки не входят в комплект речевого устройства, а находятся в пункте продаж или обслуживания речевых систем, и каждый потенциальный пользователь на этом пункте наговаривает свой обучающий материал, который заносится в память устройства. Если у одного автомата пользователей несколько, то в начале сеанса речевой связи данный пользователь должен сообщить автомату своё имя (или номер), по которому автомат вызовет в рабочее поле его эталоны.

При автономном способе средства подстройки входят в состав устройства, что позволяет настраивать и перестраивать систему под любого диктора в любое время.

Возможен и комбинированный вариант («дискетный»), при котором каждый диктор записывает в пункте обслуживания свои эталоны на дискету, которую вводит в автомат в начале работы с ним. Можно организовать «заочное» обучение: пользователь высылает в пункт обслуживания запись обучающего материала в аналоговой форме на обычном магнитофоне, а центр из этой записи готовит машинные эталоны и высылает пользователю дискету с эталонами.

Перейдём к описанию вариантов применения речевых систем. Начнём с систем, для управления которыми на каждом этапе требуется распознавать только одно слово (K=1).

1. **Система аварийного останова.** Бывают ситуации, когда оператор должен быстро остановить станок или процесс и при этом не имеет возможности воспользоваться обычными устройствами управления (кнопкой выключателя, рычагом, ключом и т.д.). Так бывает, например, когда станок захватил халат или волосы оператора или когда между оператором и аппаратурой возникли непреодолимые препятствия. В этих условиях было бы очень полезно иметь на станке микрофон и систему, которая распознаёт всего одну устную команду «Стоп!». Решение задачи осложняют следующие обстоятельства: система не должна реагировать на другие сигналы, в том числе на другие слова или на громкие механические звуки производственного цеха. Кроме того, она не должна требовать подстройки под диктора и должна быть компактной и дешёвой. При удачном решении задачи возможный тираж исчисляется сотнями тысяч устройств в год.

2. **Управление станком.** В обычной обстановке было бы полезно иметь возможность запускать и останавливать станок или процесс с помощью двух команд типа «Старт» и «Стоп». Это освободило бы руки оператора для других действий. На каждом этапе управления система должны распознавать только одно слово: если станок стоит, то она должна ждать только команду «Старт», а если работает, то только команду «Стоп». Здесь также требуется устойчивость системы против других сигналов, дешевизна и компактность, а также инвариантность к диктору.

3. **Доступ к базам данных и банковским счетам («строгий вахтёр»).** Вначале пользователь вставляет в автомат свою карточку с информацией на магнитном носителе, по которой автомат определяет владельца этой карточки. Затем пользователь должен произнести в микрофон свой пароль. Система должна распознать, тот ли пароль произнёс пользователь, и, если да, по характеристикам голоса определить, является ли данный человек владельцем данной карточки. При положительном результате открывается доступ к счёту или проход на охраняемую территорию. Можно допустить две или три неудачные попытки, после чего система может прекратить общение или включить сигнал тревоги.

Общий словарь равен числу паролей, но в каждом сеансе работы распознаётся одно известное системе слово. Здесь нужно обеспечить возможность настройки системы на индивидуальный

голос каждого пользователя. Требуется высокая надёжность системы. Больших ограничений на её габариты и стоимость нет.

4. **Обучение устной речи.** Компьютер показывает ученику предмет и просит назвать его на изучаемом языке. Ученик произносит слово и компьютер определяет, нужное ли слово произнесено или нет. В случае ошибки,компьютер может через запись или синтезатор воспроизвести правильное звучание этого слова. Общий словарь для распознавания и синтеза зависит от объёма материала, изучаемого на данном уровне. В каждый момент система имеет дело только с одним словом или словосочетанием. Для того, чтобы система нашла применение в учебных заведениях, она должна быть дешёвой и инвариантной к диктору.

5. **Исправление дефектов речи.** Компьютер показывает ученику одно слово из имеющихся в памяти, просит его произнести и сравнивает качество произнесения с эталоном. При больших отклонениях от эталона (появление звука «Л» вместо «Р», звука «С» вместо «Ш», ударения не на том слоге и т.д.) система указывает ученику на ошибку и воспроизводит по его желанию правильное звучание этого слова. Устройство должно быть инвариантным к диктору и недорогим.

6. **Обучение речи глухих.** Компьютер показывает предмет или слово и просит произнести его. Если произнесено не то слово или произнесение сильно отличается от эталона, показываются динамические спектрограммы эталонного произнесения и произнесения ученика. Кроме того, показывается динамическое изображение лица человека, произносящего это слово. Здесь, как и в предыдущем случае, система должна распознать без подстройки под диктора одно слово и оценить степень отличия полученного произнесения с эталоном. Но подсказка должны быть не звуковой, а зрительной.

7. **Узнавание диктора.** В следственной практике встречается так называемая задача «верификации» диктора. Имеется исследуемая запись речи неизвестного человека и требуется определить, является ли эта запись речью данного конкретного человека. Для сравнения делается контрольная запись одного или нескольких слов в произнесении этого человека, а система должна ответить, принадлежат ли исследуемая и контрольная записи одному и тому же диктору или нет. Задача в такой постановке практически совпадает с описанной выше задачей 3 («Строгий вахтёр»).

Более сложная задача («Идентификация диктора») выглядит так: имеется исследуемая запись разговора группы неизвестных лиц и контрольная запись речи нескольких конкретных лиц (подозреваемых), каждый из которых должен произнести одно или несколько слов, содержащихся в групповой записи. Требуется определить, участвовали ли подозреваемые в данном групповом разговоре и, если да, кому из них принадлежит какая часть исследуемой записи. Больших ограничений на стоимость системы не налагается, но требуется высокая надёжность решения.

8. **Управление краном.** Под крышей задымлённого цеха движется кран, в кабине которого крановщица выполняет роль автомата, распознающего 8–10 слов (вперёд, назад, влево, вправо, вира, майна, прямо, тише, стоп и т.д.), отдаваемых стропальщикам. Удобная микрофонная гарнитура с радиопередатчиком у стропальщика в сочетании с приёмником и распознающим устройством на борту крана позволили бы освободить человека от выполнения примитивных функций в опасных для здоровья условиях. Направленный микрофон подводится ко рту

стропальщика только на время отдачи команд крану, что помогает уменьшить влияние шума и не лишает стропальщика возможности свободного устного общения с окружающими в другие моменты времени. Требуется решить вопрос борьбы с помехами в радиоканале, вызываемыми электросваркой и другими производственными процессами. Возможна подстройка системы под диктора в автономном или дискетном режиме.

9. **Управление коляской.** Для управления инвалидной коляской было бы достаточно распознавать 8 команд одного диктора (вперёд, назад, налево, направо, кругом, стоп, начало, конец). Такое управление освободило бы руки инвалида во время движения и было бы полезно для людей с поражёнными ногами и руками. Устройство должно быть компактным и недорогим. Желательна подстройка под диктора (централизованная или дискетная).

10. **Управление микроскопом.** Микрохирург наблюдает за операционным полем через окуляр микроскопа, и ему приходится отрывать руки от операции, чтобы изменить поле зрения или фокусировку микроскопа. Было бы гораздо удобнее, если бы он мог делать это, произнося в микрофон небольшой набор команд (выше, ниже, влево, вправо, вперёд, назад, стоп и пр.). Возможен любой способ подстройки под диктора. Жёстких ограничений на стоимость и габариты устройства нет.

11. **Речевой блокнот.** При пользовании обычным блокнотом требуется найти нужную страницу, прочитать нужную запись, зачеркнуть устаревшую информацию или сделать новую запись. «Речевой блокнот» содержит два блока: блок управления и блок работы. На этапе управления пользователь должен иметь возможность с помощью устных команд выбирать нужную «страницу» и нужную «строку» на ней и задавать режим работы: запись, считывание или стирание. Эти функции выполняет блок управления, представляющий собой устройство, распознающее 50–60 слов при KJ10. Таким словарём можно обеспечить выбор страниц типа «телефоны», «совещания», «встречи», «заказы» и т.д., а также выбор нужной части страницы: адрес или телефон по фамилии, перечень запланированных мероприятий по дате и пр.

Блок работы позволяет производить запись в выбранную строку устного сообщения произвольного содержания и длины (желательно в экономном закодированном виде), считывание и звуковое воспроизведение записи, содержащейся в выбранной строке, а также стирание сигнала в выбранной строке. Здесь предстоит скомбинировать два направления речевых технологий: автоматическое распознавание управляющих команд и сжатие речевых сигналов с последующим их разборчивым и качественным воспроизведением. Желательно предусмотреть возможность работы с блокнотом по телефону. Естественна подстройка под диктора, возможно сочетание с системой «Строгий вахтёр».

12. **Управление роботом.** Можно организовать диалог с роботом так, чтобы при общем словаре в 100–150 слов на каждом шаге диалога требовалось распознать не более 10 слов. Требования к надёжности распознавания зависят от того, какую функцию выполняет робот. Возможна подстройка под диктора.

13. **Игральные автоматы.** Голосовое управление фигурами, участвующими в компьютерной игре (люди, машины, предметы и т.д.), позволили бы оживить некоторые игры или создать новые виды игр. Система должна распознавать 10–20 слов, быть дикторонезависимой, компактной и дешёвой.

14. **Управление бытовой аппаратурой.** Словарь из 20 слов был бы достаточным, чтобы управлять режимами работы радио- или телеаппаратуры (включать и выключать её, менять номер канала, громкость звука и т.д.). Очень высокой надёжности не требуется. Подстройка под диктора нежелательна, стоимость не должна быть высокой.

15. **Телефонный номеронабиратель.** Десять цифр и десять команд типа «дом», «офис», «Петров» и т.д., распознаваемых автоматом, находящимся на телефонном узле связи, позволили бы быстро соединяться с наиболее частыми абонентами и упростить периферийную аппаратуру телефонной сети (телефонная трубка вместо телефонного аппарата с номеронабирателем). Автомат должен распознавать цифровые и кодовые команды и знать, какие номера набирать по кодам, принятым с данного аппарата. Подстройка под диктора неприемлема, нужны средства компенсации помех в телефонном канале.

16. **Автомобильный радиотелефон.** Водителю нужно соединиться по радиотелефону с нужным абонентом и при этом не отрывать глаза и руки от управления автомобилем. Бортовое распознающее устройство позволяет водителю набирать номер, произнося в микрофон, как и в предыдущем случае, составляющие его цифры или кодовые слова. Словарь — 20 слов. Очень высокой надёжности не требуется, возможен режим переспроса при неуверенном распознавании. Приемлема централизованная или дискетная подстройка под диктора. Сложности состоят в наличии шумов в салоне движущегося автомобиля и нестабильном расстоянии между микрофоном и диктором.

17. **Банковские операции по телефону.** Работа включает в себя два этапа: этап доступа к счёту и этап управления счётом.

На первом этапе клиент с помощью номеронабирателя сообщает компьютеру свой условный номер. Компьютер вызывает в рабочее поле имеющиеся в его базе данных речевые эталоны данного клиента и просит произнести свою фамилию и пароль. Компьютер проверяет правильность этих слов, по особенностям голоса удостоверяется, действительно ли говорит хозяин данного счёта, и, если да, открывает доступ к управлению счётом. На этом этапе словарь состоит из двухтрёх слов при K=1.

На втором этапе в словарь входят десять цифр и команды типа: «перевести», «списать», «зачитать» и т.д. Общий словарь порядка 20 слов при K=10, с подстройкой под диктора.

Документирование работы со счётом ведётся с помощью записи сеанса связи в архив системы. Трудность задачи состоит в очень высоких требованиях к надёжности работы, особенно на первом этапе, и в помехах телефонного канала.

18. **Управление локомотивом.** Машинист обязан наблюдать за дорогой и светофором и устно фиксировать состояние светофора.

Аналогичное применение система может найти и в работе диспетчера морского порта. Здесь достаточно словаря в 200–300 речевых единиц с K=30. Учитывая, что ситуация здесь меняется не так быстро, как в воздухе, допустимы режимы «эхо» и переспроса.

19. **Справки о компьютерной сети.** Задавая в микрофон вопросы типа: «загрузка пятой машины», «количество отказов сети», «сколько времени занял абонент номер двадцать семь» и т.д., можно услышать из динамика ответы в устной форме. При большом общем словаре (W<1000) можно организовать диалог при K<30. Возможна автономная подстройка под диктора. Очень высокой надёжности не требуется, возможен режим переспроса.

20. Управление кораблём или самолётом. Оператор (т.е. капитан надводного, подводного или воздушного корабля) в процессе управления должен наблюдать за окружающей обстановкой и следить за состоянием управляемых систем по многочисленным приборам, лампочкам и сигнальным табло. Деятельность глаз и рук оператора насыщена выполнением своих функций до предела, в то время как речевой канал почти не используется. Эффективность управления сильно возросла бы, если бы оператор мог, не отрывая глаз от наблюдения за обстановкой, дать, например, команду «горючее» и услышать устное сообщение о количестве оставшегося топлива. Помимо информационных функций речевой канал можно эффективно использовать и непосредственно в управлении, отдавая команды типа «убрать закрылки», «вправо на 30 градусов» и т.д.

Набор устных команд в 100–150 слов можно разделить на подсловари, каждый из которых предназначен для своего этапа движения (взлёт, посадка, всплытие и т.д.) и может состоять из 20–30 слов. Возможна и даже необходима автономная или дискетная подстройка под диктора. Акустическая обстановка характеризуется высоким уровнем шумов или резонансами в объёме гермошлема пилота. Дополнительная трудность — переменное эмоциональное состояние оператора, а для лётчиков ещё и большие физические перегрузки (до 7 G) во время воздушных манёвров.

Повышение эффективности корабля путём добавления речевого канала в контур управления настолько значительно, что на создание таких речевых систем тратится сейчас много средств и усилий во многих странах, в частности, в рамках проектов НАТО.

21. Обучение авиадиспетчеров. Курсант — будущий авиадиспетчер — наблюдает на экране дисплея ситуацию в воздушном пространстве в районе аэропорта и ведёт радиопереговоры с пилотами, роль которых играет инструктор, сидящий в соседней комнате. Курсант отдаёт команды управления типа «Борт 82645, займите четвёртый эшелон», «Борт 7650, заходите на посадку» и т.д. Инструктор контролирует правильность и своевременность команд и, если всё в порядке, задаёт на дисплей новую учебную ситуацию, а если реакция курсанта неправильна, инструктор указывает на допущенную ошибку и продолжает сеанс связи.

Автоматизированная система должна включать в себя блок распознавания команд курсанта и блок синтеза сообщений пилотов и инструктора. Кроме того, в системе должен иметься генератор учебных ситуаций, а также блок анализа и оценки реакции курсанта и блок выбора корректирующей реплики инструктора.

Характеристики речевой системы таковы. Словарь содержит примерно 1000 речевых единиц в виде отдельных слов («понял», «повторите», «выполняйте» и т.д.), а также коротких словосочетаний, произносимых слитно («взлёт разрешаю», «восемьдесят два», «девятьсот сорок семь» и т.д.). Облегчающий момент состоит в том, что для каждого типа занятий (управление в ближайшей зоне, управление в дальней зоне и т.д.) используется свой подсловарь. Самая трудная часть словаря — цифровые номера бортов. Но в каждой учебной ситуации участвует не более 30 известных ей номеров, так что,коэффициент ветвления в системе не превышает 30. Дискетная подстройка под диктора. Другим облегчающим фактором является стандартная конструкция фраз, которой диспетчер должен неукоснительно придерживаться. Трудность — шум в учебной аудитории, где рядом сидят и отдают команды другие курсанты.

22. Ассистент диспетчера. Система должна использоваться в условиях работы реального диспетчерского пункта, вести протокол переговоров диспетчера с пилотами и по возможности предотвращать грубые ошибки диспетчера, чтобы он не разрешил посадку самолёта на занятую полосу, не направил бы два самолёта по пересекающимся маршрутам в одном и том же эшелоне и т.д. Если нецифровая часть будет вызывать большие трудности, то возможен

ввод с клавиатуры бортового номера каждого самолёта, появляющегося в поле наблюдения, после чего система будет работать с эталонами только этих номеров, по ходу работы исключая те номера, связь с которыми прекращается и которые содержатся во фразе диспетчера «Борт такой-то, конец связи», так что коэффициент ветвления можно ограничить числом 40–50. Подстройка под диктора вполне приемлема (автономная или дискетная). Система не должна раздражать диспетчера частыми переспросами, так что требования к надёжности достаточно высоки.

23. **Ввод данных из журнала наблюдений.** В биологии, ботанике, геологии и т.д. накоплены большие массивы экспериментальных данных, записанных в своё время в полевые журналы. В некоторых организациях (например, во Всероссийском институте растениеводства им. Н.И. Вавилова) имеются сотни тысяч таких журналов о наблюдениях за особенностями выращивания и развития десятков тысяч сортов растений. Ввод данных из журналов в ЭВМ с помощью оптических сканеров затрудняется тем, что записи делались в полевых условиях, разными почерками, чернилами и карандашами и т.п. Ручной ввод с клавиатуры потребовал бы огромных трудозатрат. Оператор должен поочерёдно смотреть в журнал, на клавиатуру и экран компьютера, что делает процесс ввода утомительным, медленным и сопровождается большим числом опечаток.

Ситуация радикально меняется, если оператор, не отрывая рук и глаз от журнала, диктует в микрофон содержание очередной журнальной записи, контролируя правильность ввода с помощью «эха», синтезирующего слова, воспринятые машиной. Как показали наши исследования, по сравнению с клавиатурой скорость ввода при этом возрастает в 6–8 раз при заметно меньшем числе ошибок [2]. Словарь порядка 250 слов при K=40. Есть возможность дополнительного контроля правильности ввода с использованием семантических и прагматических ограничений (например, дата цветения не может быть более ранней, чем дата всхода растения).

24. **Ведение протокола наблюдений.** Не составляет труда для оператора записать на магнитофон сообщение о том, что он видит, не отрываясь от наблюдений. Но иногда требуется не просто запись наблюдений для дальнейшего анализа, но и немедленная реакция на наблюдаемую ситуацию, при этом для принятия решения требуется помощь информационной или экспертной системы. Следовательно, нужно, чтобы система понимала содержание устного сообщения и вырабатывала бы адекватную реакцию (синтезировала устную подсказку, включала бы требуемый исполнительный механизм и т.д.).

Для многих применений достаточно распознавать 50–100 слов, при K=20. В литературе описывались эксперименты с системами, в которых контролёр замеряет размеры детали и описывает их состояние — цвет, качество поверхности, — а автомат маркирует деталь или отправляет её на переделку. Оператор на почтовом конвейере читает код города на посылке, переворачивая её, если требуется для прочтения кода, а конвейер отправляет посылку в нужный отсек накопителя. Артиллерийский наблюдатель корректирует по радио огонь с помощью команд типа «недолёт», «левее 20» и т.д., а система автоматической наводки управляет орудием. Пилот космического корабля описывает видимые признаки предметов, летающих в космосе, а система пытается определить, что это за предмет или деталью какого изделия он является.

25. Ввод банковских платежей. В каждом банке имеются сотрудники, занятые вводом информации, записанной на стандартных бланках, например, платёжных поручений. Нужно ввести наименования и номера счетов плательщика и получателя и информацию о назначении платежа. Было бы удобно иметь возможность сочетать ручной и устный ввод такой информации. Например, по желанию оператора вводить номера счетов либо через микрофон, либо через клавиатуру; стандартные назначения платежей (а таких порядка 80) вводить с помощью ключевых слов типа «налог с прибыли», «соцстрах» и пр. Такие возможности сделали бы работу оператора более комфортной и менее трудоёмкой. Нецифровой словарь системы от 50 до 1000 слов (если включать наименования клиентов) при К от 10 до 200. Приемлема дискетная подстройка под диктора.

26. Справки по телефону. Абонент речевой справочной сети может запросить по телефону информацию о других абонентах данной сети, о репертуаре кинотеатров, о номере канала и времени спортивных телерепортажей, о расписании поездов, о времени работы магазинов и т.д. Система вначале запрашивает, какой из 10–20 видов справок интересует абонента. Затем спрашивает, например, вид транспорта (самолёты, поезда, пароходы и т.п.). Потом определяется направление движения и наименование пункта прибытия и т.д. После уточнения всей входной информации система синтезирует устную справку.

Диалог, построенный по иерархическому принципу, позволяет при общем словаре в несколько тысяч слов на каждом этапе распознавать не более 20 слов. Реплики абонента могут содержать неизвестные системе слова, что требует распознавания заданного словаря в произвольном речевом потоке. Переспросы допустимы, подстройка под диктора неприемлема. Дополнительные трудности состоят в шуме и искажениях сигнала в телефонной линии.

Если терминал системы установлен, например, в большом офисе или в гостинице, то блок распознавания и синтеза может входить в состав этого терминала и телефонная линия из тракта исключается. При этом терминал можно объединить с дисплеем для высвечивания табличной информации.

27. Система бронирования. Абонент в телефонном диалоге сообщает системе о своём желании забронировать билет на самолёт, поезд, автобус или пароход. Система ведёт диалог и из кратких ответов абонента («самолёт», «в Москву», «апрель», «десятого», «утром» и т.д.) получает всю необходимую информацию, после чего сообщает абоненту о том, что билет такого-то типа может быть забронирован и что его можно выкупить в такой-то кассе, в такое-то время и за такую-то цену. Абонент подтверждает своё согласие с данным предложением, система бронирует билет и сообщает абоненту номер брони.

Технические требования к речевой части такой системы точно такие же, как и для системы «Справки по телефону».

28. Речевой путеводитель. Абонент в диалоге, которым управляет система, сообщает ей место своего нахождения и то место, где ему хочется быть: «на пересечении улицы Академической и Морского проспекта», «в пункт обмена валюты», «доллары», «на карбованцы», «городским транспортом». После этого система сообщает, что требуемый обмен валюты делается в таком-то банке, адрес которого такой-то, работает он в такие-то часы, обменный курс такой-то, а проехать к банку можно на таком-то автобусе до такой-то остановки и пр.

Данная система может быть сочленена с монитором, на котором изображается план города и курсором показывается кратчайший путь к требуемому месту. От двух предыдущих систем эта система отличается только информационными базами, с которыми работает речевой блок.

29. Речевой интерфейс с экспертной системой. В начале сеанса работы экспертная система идентифицирует пользователя по голосу и определяет круг доступных для него функций. Если это рядовой пользователь, то ему не разрешается вносить изменения в базу данных и знаний и можно пользоваться только определённой её частью.

Затем пользователь делает запрос или сообщает системе какую-то информацию, система формирует ответную реакцию и синтезирует устный ответ или задаёт очередной вопрос.

Общий словарь системы порядка 500 слов при K=40. Возможна дискетная или автономная подстройка под диктора, допускается режим переспросов. Если связь с системой осуществляется по телефону, то добавляются проблемы телефонных шумов и искажений.

30. Фразовый вокодер. Пользователь говорит системе фразу с описанием наблюдаемой ситуации. Например, брокер сообщает своему банку о том, что акции фирмы «Русич» поднялись на 13%. Система распознаёт ключевые слова, по которым определяет, к какому из возможных типов ситуаций относится данная. Затем во фразе выделяются и распознаются слова, характеризующие её индивидуальные особенности (параметры). После этого система выбирает из памяти и передаёт по линии связи только номер типа ситуации и кодовые номера, соответствующие значениям наблюдаемых параметров. Достигается максимальное сжатие сигнала и закрытие передаваемой информации от несанкционированного доступа.

Размер словаря определяется числом типов ситуаций, интересующих пользователя в данном рабочем сеансе, и количеством различаемых значений их параметров. При общем словаре в несколько сотен фраз можно пользоваться подсловарями из нескольких десятков фраз. Приемлемы подстройка под диктора и режим переспроса.

31. Пульт спортивного комментатора. Комментатору было бы желательно в определённые моменты работы быстро получить справку о том или ином спортсмене или спортивном событии. Для этого нужно сообщить системе парольное слово (например, «система справка»), означающее переход от режима репортажа к справочному режиму. Система задаёт комментатору вопросы в той последовательности, при которой можно за минимальное число шагов добраться до нужных данных, и озвучивает ему данные. Затем комментатор сообщает об окончании справочного режима («конец справки») и переходит к продолжению репортажа.

Диалог можно организовать так, чтобы K был не больше 30. Подстройка под диктора и режим переспросов возможны. Шум трибун можно минимизировать направленным микрофоном.

32. Диктовка деловых писем. Значительная часть деловых писем имеет стандартную форму, что позволяет вызвать на экране монитора письмо нужного вида и ввести по запросу системы только специфическую информацию: кому, от кого, сколько и т.д. Частую информацию можно вводить в устной форме, а уникальную — через клавиатуру. Можно ограничиться словарём в несколько сотен слов при K=40–50. Возможны переспросы и подстройка под диктора.

33. Диктовка отчёта. Ситуация аналогична предыдущей и представляет собой одну из простейших реализаций гипертекстовой технологии работы по подготовке документов. На экране высвечивается форма типового отчёта, содержащего типовые фразы с пробелами для указания конкретных данных. Курсор указывает на очередной пробел, и пользователь произносит нужное слово или словосочетание. При K=40–50 общий словарь состоит из нескольких сотен слов — названий цифр, городов, объектов, месяцев и т.п. Допустимы подстройка под диктора и переспросы. Имеются такого рода системы, например, для отчёта о состоянии лёгких по рентгеновскому снимку. На одном мониторе показывается снимок, а на другом — форма отчёта. Врач вставляет в стандартную фразу типа «затемнение в районе ___ ребра», например, слово «четвёртого», и курсор переходит к следующей неполной фразе.

34. Психофизиологические тесты. Оператор (пилот или космонавт) произносит некоторые слова, часто используемые в радиопереговорах (например, «понял», «хорошо» и т.д.), а система должна по характеристикам голоса определить его состояние — в хорошей ли рабочей форме он находится или устал, спокоен или перевозбуждён, испуган или подавлен. Это служит основанием для принятия решения о том, можно ли оператору в данный момент поручить выполнение некоторой ответственной операции или нет.

Важен такой анализ и при анализе катастроф, чтобы понять, в какой момент оператор почувствовал реальную опасность ситуации. Приемлема подстройка под диктора. Количество ключевых слов — 2–3, число различаемых состояний — до 10. Возможны большие помехи в тракте связи.

35. Автоматическая стенография. Система должна в реальном времени распознавать и записывать в буквенной форме слова непрерывной речи на базе большого словаря (10–20 тысяч слов). Подстройка под диктора допустима, переспросы неприемлемы. Очень высокие требования к надёжности распознавания не предъявляются.

36. Фонемный вокодер. Слитная речь без ограничений на словарь перекодируется в фонемы, и коды фонем передаются по линии связи. Количество ошибок распознавания фонем не должно мешать человеку понимать речь, синтезированную на выходном конце связи по принятым фонемным кодам. В некоторых случаях подстройка под диктора возможна.

37. Синхронный перевод. Эту задачу можно отнести к задачам предельной трудности. Система должна не только распознавать, но и понимать смысл услышанного, чтобы генерировать и синтезировать соответствующую фразу на другом или других языках. Словарь 10–20 тысяч слов, непрерывная речь. Предварительная подстройка под диктора нежелательна. В некоторых хорошо подготовленных условиях возможна дискетная подстройка (например, участники международной конференции, выступающие с докладами, привозят записи с обучающим речевым материалом, из которого накануне доклада формируются речевые эталоны для данного диктора).

38. Монтаж и испытание схем. Синтезатор в нужной последовательности даёт команды, которые должен выполнить монтажник или контролёр сложной схемы: «жёлтым проводом соединить контакт два третьего ряда с контактом девять тринадцатого ряда» или «между точкой А2 и «землёй» должно быть 27 вольт» и пр. Требуется хорошая разборчивость речи, высоких требований к натуральности звучания не предъявляется. Процесс монтажа или контроля заметно ускоряется, сокращается число ошибок в работе.

39. Речевые информаторы. Система считывает показания метеорологических приборов, синтезирует фразы с информацией о параметрах погоды и передаёт это сообщение по радио для лётного и наземного персонала аэропорта: «ветер северный, пятнадцать метров в секунду, облачность 200 метров».

Другой вариант — информаторы на вокзалах, передающие повторяющуюся или текущую информацию: «Комната матери и ребёнка находится на втором этаже, камеры хранения в левом крыле вокзала», «Поезд номер один Москва — Владивосток отправляется с первого пути через три минуты. Будьте осторожны» и пр. Требуется высокая разборчивость и хорошее качество синтезированной речи.

40. **Речевые оповещатели.** Речевая система является частью автоматической системы контроля или наблюдения. Если измеряемые параметры выходят за определённые пределы, то система синтезирует устную фразу с сообщением об этом факте. Например, «Утечка сернистого газа в узле номер восемь», «Шасси не убраны», «Открылся багажник» и пр. В ситуации, требующей принятия немедленных мер, система указывает последовательность нужных действий: «Пожар на пятом участке. Включить сигнал пожарной тревоги. Выключить вентиляцию. Вызвать аварийную бригаду...».

Опыт показывает, что в системах, предупреждающих об опасности, более желателен женский голос при высокой разборчивости и натуральности его звучания.

41. **Звуковая клавиатура.** Слепым и слабовидящим пользователям компьютеров было бы удобно слышать звуковые сигналы, соответствующие каждой клавише, и иметь возможность прослушать содержание информации на экране дисплея. Высоких требований к качеству звука не предъявляется.

42. **Читающие автоматы.** Система считывает текст, распознаёт буквы, слова и знаки препинания и синтезирует устные фразы соответствующего содержания. Это позволило бы слепым людям читать обычные книги и пр. Чтение может быть продолжительным во времени, поэтому желательно наряду с хорошей разборчивостью иметь и высокую натуральность синтезированной речи.

Литература

1. *Загоруйко Н.Г.* АРСО и речевые технологии // Сборник трудов ИМ СО РАН, «Вычислительные системы», Вып. 153, Новосибирск, 1995. С. 3–31.

2. Zagoruiko N.G., Tambovtsev Yu.A. Aspects of Human Performance in Intensive Speech Task // Int. Journal on Man-Machine Studies. V. 16, 1982, Academic Press, London.

Загоруйко Николай Григорьевич —

зав. отделом Информатики Института математики им. С.Л. Соболева доктор технических наук, профессор, академик Международной академии информатизации. Профессор Н.Г. Загоруйко — один из ведущих отечественных специалистов по машинным методам обнаружения эмпирических закономерностей. Его работы в области анализа данных и распознавания образов имеют мировую известность. Более 30 лет ведет активную педагогическую деятельность, читая курсы лекций по искусственному интеллекту и анализу данных в Новосибирском государственном университете, ряде зарубежных университетов. Участвует в международных и всероссийских научных конференциях, является членом советов по защитам диссертаций. Среди его учеников — 5 докторов и более 30 кандидатов наук. Н.Г. Загоруйко — автор более 190 научных работ, в том числе 11 монографий.