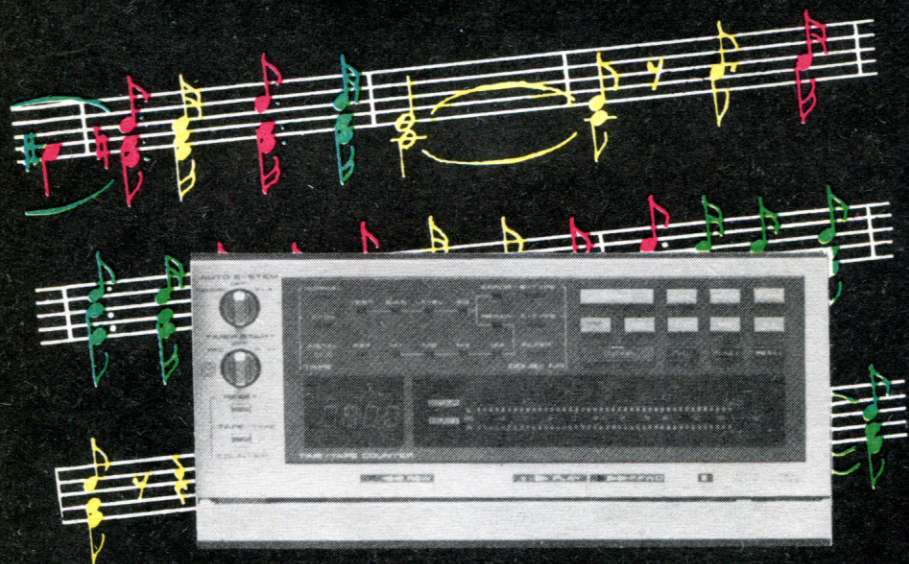


БИБЛИОТЕЧКА
ВЫЧИСЛИТЕЛЬНАЯ
ТЕХНИКА
И ЕЕ ПРИМЕНЕНИЕ



*Издательство
"Триумф"*

РЕЧЬ И ЭВМ



ВЫПУСК **8**

БИБЛИОТЕЧКА .
ВЫЧИСЛИТЕЛЬНАЯ
ТЕХНИКА
И ЕЕ ПРИМЕНЕНИЕ

В помощь
лектору

РЕЧЬ И ЭВМ

ВЫПУСК **8**

Под общей редакцией
члена-корреспондента АН СССР
В.И.Сифорова

OldPC.ru

5 0 5 1

музей компьютеров

Издательство "Знание"
Москва 1987

ББК 32.973

Р 46

Составитель — Свириденко В. А., доктор технических наук,
профессор.

Р 46 Речь и ЭВМ: Сб. статей/Под ред. В. И. Сифорова. — М.: Знание, 1987. — 64 с. — (В помощь лектору. Библиотечка «Вычислительная техника и ее применение». Вып. 8).

20 к.

25 000 экз.

Общение с вычислительной машиной на естественном для человека языке — одна из актуальных проблем в разработке ЭВМ нового поколения. В брошюре рассказано о решении задач синтеза (генерации речевых образов) и анализа (распознавания их) — основных систем современной речевой технологии.

Брошюра рассчитана на широкий круг читателей — лекторов, производственников, студентов.

Р 240500000—152
073(02)—87

ББК 32.973

© Издательство «Знание», 1987 г.

Речевой диалог с ЭВМ В. И. Сифоров	4
Компьютерный анализ речи А. С. Вайсман	6
Компьютерный синтез речи В. Н. Сорокин	27
Речевой ввод-вывод для персональной ЭВМ В. А. Свириденко	46
Диалоговый вычислительный комплекс «Электроника НЦ 8020-4» В. С. Кокорин, А. А. Попов	49
Интегральные схемы для средств вычислительной техники Н. В. Воробьев	52

Речевой диалог с ЭВМ

В. И. Сифоров

Функционирование экономики, а также управление структурами технически развитого общества во все большей степени начинают зависеть от возможностей получения, обработки и передачи информации. Объем информационных потоков и сложность информации все время возрастают, поэтому способность общества к дальнейшему развитию определяется не только технологией добычи и переработки сырьевых ресурсов, но и технологией обработки информации. Создание и быстрое расширение возможностей ЭВМ, обладающих способностью не только к вычислениям, но и к логическим действиям, привели к появлению новых средств в обработке информации, в принципе решающих проблему «информационного кризиса». На повестке дня стоит разработка ЭВМ пятого поколения, обладающих элементами искусственного интеллекта, т. е. способных решать такие задачи, которые раньше были под силу только человеку. Эти ЭВМ смогут взять на себя решение многих рутинных задач по обработке информации, они станут неисчерпаемыми источниками сведений, накопленных человечеством, окажутся неоценимыми помощниками и советчиками в решении технических и научных задач.

Одним из существенных признаков искусственного интеллекта является способность к речевому общению с человеком. Поэтому во многих странах ведутся интенсивные исследования с целью разработки систем автоматического распознавания и синтеза речи. Призванная для обслуживания сложных ЭВМ, речевая технология сама требует вычислительных средств высокой производительности для обработки речи. Таким образом, развитие новых типов ЭВМ не только ставит задачу создания систем распознавания и синтеза речи, но и предоставляет необходимые средства для решения этой задачи.

Имеются две ситуации в использовании речевых систем: в одной из них речевой диалог можно обеспечить различными способами, а в другой связь с ЭВМ может быть осуществлена только посредством речи. Последняя ситуация возникает, когда источник информации или объект управления (база знаний, искусственный интеллект, информационно-справочная система, подвижный робот) удален от человека и связь происходит по телефону или радиоканалу. Кроме того, общение с ЭВМ в таких условиях, когда все внимание сосредоточено на объекте наблюдения, возможно только с помощью речи.

Но и в тех случаях, когда речь не является единственным средством связи с ЭВМ, даже современные, но еще довольно несовершенные системы распознавания и синтеза речи повышают производительность труда от 20 до 300%. Широкое распространение персональных ЭВМ возможно только при их доступности людям, не владеющим специальными навыками программирования. Такие самопрограммирующиеся ЭВМ способны создавать алгоритмы по качественному, словесному описанию задачи, а также нуждаются в системах распознавания — синтеза речи, являющихся необходимым элементом комфортности общения с ЭВМ.

Речевые системы найдут применение практически во всех областях человеческой деятельности: промышленности, делопроизводстве, связи, военной технике, обучении. Особую область составляет диагностика заболеваний по характеристикам голоса и создание речевых протезов — читающих, говорящих и распознающих устройств для инвалидов.

Системы, позволяющие вести разговор с ЭВМ на естественном языке, изменят наше общество, возможно, в не меньшей степени, чем это сделало широкое распространение телефона.

Компьютерный анализ речи

А. С. Вайсман

Звуки мертвых,
Музыку я разъял, как труп. Поверил
Я алгеброй гармонию.

А. С. Пушкин. Моцарт и Сальери

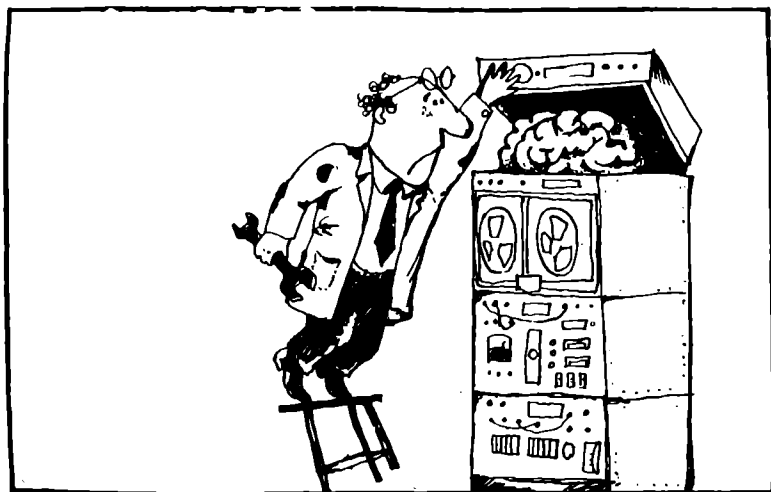
О распознающих и «говорящих» машинах люди мечтают с глубокой древности. В течение веков было проведено колоссальное количество исследований и создан целый ряд интересных устройств. Так, сто лет назад русская газета «Берег» сообщила: «Среди замечательных изобретений года можно выделить одно новшество, которое привлекает всех своей оригинальностью. Это изобретение — зоофон, воспроизводящий крик тридцати семи различных животных и между прочим голос человека». По существу, старинная шарманка была не только «играющей», но и «певающей» машиной. Характерно, что в ее конструкции роль своеобразных пневматических фильтров играли... перфокарты.

Но все это были машины и устройства, так или иначе воспроизводившие речь без ее анализа. А для того чтобы подавать команды голосом, нужно знать не только ее звучание, но и то, как она образуется. В эпоху НТР, когда ЭВМ проникают буквально повсюду — с их помощью ежесекундно перерабатываются огромные массивы информации, проблема автоматического распознавания человеческой речи с каждым днем становится все насущнее. В наши дни от машин требуется, чтобы они были «умными». Это не только красивая метафора — особенностью современного производственного процесса является его растущая интеллектуальная насыщенность. А вся действующая «армия» счетно-решающих устройств требует общения с нею, иначе она не сможет работать, — ведь армией нужно управлять. Да, но на каком языке отдавать команды? Естественно, на «машинном».

Сейчас работа программиста и оператора ЭВМ

очень сложна. С машиной необходимо разговаривать исключительно на специальных языках программирования. К тому же написанную на бумаге программу нужно еще перевести на машинный язык и перенести на машинный носитель — перфокарту, перфоленгу, магнитную ленту, магнитный диск. Долго, хлопотно, дорого. Таким образом, ввод информации «с голоса» — тот путь, который кардинально приблизит к нам пока еще очень отдаленную ЭВМ.

К сожалению, мы до сих пор не знаем, за счет чего (как это ни парадоксально звучит) люди, говорящие на одном языке, прекрасно понимают друг друга, несмотря на различия в дикции, произношении, интонации. Очевидно, распознавание осуществляется за счет того, что в речевом сигнале есть какие-то постоянные составляющие — определенные физические параметры и их соотношения, — присущие любому диктору (мы не берем в расчет больших дефектами речи) и независящие от интонации при произношении одного и того же текста. Ученые называют это **фонетическим кодом**. Наш мозг, располагая 10 млрд. нервных клеток с многочисленными связями между собой, с легкостью нащупывает и дешифрует этот код. Мы так же до сих пор не знаем, как в мозгу осуществляется процесс распознавания. Именно здесь причина того, что при сравнительно боль-



ших успехах в создании синтезаторов речи мы весьма мало достигли в создании анализаторов речи.

Необходимо подчеркнуть, что для нас неизмеримо большую ценность представляют не «говорящие», а «понимающие» машины. Особенно важно, чтобы машина правильно реагировала на наши речевые команды, а в какой форме она будет «говорить» с нами, имеет меньшее значение. Мы ждем от нее «внимания и повиновения».

Но, как мы уже отмечали, любой диктор говорит со своей интонацией, со своим произношением. Помимо более или менее постоянных голосовых данных, интонация сильно зависит от состояния, от эмоции в момент произнесения. По нашей речи легко определить, что говорится и кто и как говорит.

Но, несмотря на огромные различия в произношении, все люди, говорящие на одном языке, прекрасно понимают друг друга. Чем можно объяснить это удивительное явление? Именно наличием такого фонетического кода, такого соотношения физических компонентов, которое легко выявляется на слух у всех дикторов.

Длительное эволюционное развитие глубоко замуровало смысловой элемент, фонетический код нашей речи. По отношению ко всему речевому сигналу он играет роль ядра в атоме или клетке. Можно сказать, что фонетический код — это тот характер и тот минимум информации, который необходим для распознавания речевого сигнала машиной, в какой бы форме ни производилась эта операция, и вне зависимости от того, насколько она совпадает с распознаванием речи человеком. Ведь было бы большой ошибкой считать, что они могут полностью совпадать. Распознавание машиной базируется на принципе распознавания человеком, но детали могут совсем не совпадать.

Итак, в конечном итоге мы должны заложить в память машины математическое описание речевого сигнала — другого языка она просто не понимает.

Известный швейцарский лингвист Ф. де Соссюр вывел положение о том, что «...мы прежде всего различаем внутри общего феномена, каким является речевая деятельность, два фактора: язык и речь. Язык для нас — это речевая деятельность минус сама речь. Он есть совокупность лингвистических навыков, позволяю-

щих отдельному человеку понимать других и быть ими понятым» [1]. В реальной деятельности это, разумеется, не простая арифметическая сумма, а достаточно сложная интегральная система.

Итак, мы можем сравнить семантическую часть речевого сигнала (язык) с передаваемой информацией (сообщением), а речь представить как различные способы передачи, например как передачу одной и той же информации на разных волнах или по разным каналам. В совокупности они образуют модулированный сигнал — речевую деятельность. Необходимо сразу оговориться — под семантикой речевого сигнала мы подразумеваем не какой-то психический смысл, какой-то психический образ, а лишь тот набор физических параметров, с помощью которого мы выделяем какой-либо отрезок речи из всех остальных. Например, слово «стол» в смысле фонетического кода есть всего лишь набор четырех фонем (с-т-о-л), не имеющих ничего общего с абстрактным или конкретным понятием сооружения для работы, еды и т. д.

Сравним кодирование информации в речи с кодированием в радиопередачах. При этом мы имеем в виду обычные, общепринятые коды связи. Возьмем всем известную систему кодов связи — азбуку Морзе. Основу ее составляет комбинация двух элементов — точки и тире. Точка — короткий сигнал, тире — длинный. Как видим, даже в одной из самых простых систем кодирования определение передаваемой буквы зависит от длительности составляющих элементов, интервалов между ними в пределах одной буквы, которые должны быть весьма короткими, и интервалов гораздо более длительных между буквами.

Ученые называют это временным кодированием. Здесь необходимо внести стилистическую коррективу. Под временным кодированием часто подразумевают более или менее автономную определенную ритмическую структуру сигнала или его значительную часть. Это очень ярко видно при игре на африканских барабанах — тамтамах. В них соблюдается более или менее равномерное чередование ударов разной силы. Аналогичную роль играют и длительности элементов и интервалов между ними во многих типах сигналов у различных животных.

Отметим: в структуре таких кодов нет частоты, а

есть ритмика, т. е. временные интервалы. Передача индивидуальна у каждого радиста (так называемый почерк, по которому радисты сразу могут узнать друг друга) и колеблется в определенных пределах. Однако структура букв дает возможность практически безошибочно отделять их одну от другой и распознавать.

По аналогии с радиосвязью можно сказать, что речь — это сведения о «передатчике» — дикторе, и о «режиме работы» — психическом состоянии диктора: спокойном, возбужденном и т. д. Соответственно можно говорить о двух речевых кодах: индивидуальном коде диктора — **диктоскопическом** и коде его состояния в момент произнесения речевого сигнала, определяемом законами психолингвистики, — **диктофоническом**. Итак, имеем три компонента, практически сливающихся в одном речевом сигнале.

Таким образом, для создания универсального устройства распознавания речевых сигналов необходимо обнаружить фонетический код, т. е. тот объем и ту форму информации, которой будет достаточно для создания эталона в памяти ЭВМ, чтобы выделить семантическую, «языковую» часть из всего речевого потока речевой информации. Поскольку фонетический код является универсальным в смысле доступности его любому слушателю, владеющему соответствующим языком, мы можем утверждать, что именно в фонетическом коде используется наиболее простая форма кодирования. В общем виде таким является временное кодирование, что чрезвычайно важно. Язык, т. е. фонетический код, мы можем представить себе как цифровой код, как информацию, записанную в цифровой форме, а речь, т. е. диктоскопический и диктофонический коды, — как аналоговые.

Говоря о кодах речевого сигнала, нужно отметить, что каждый из нас был участником своеобразного эксперимента. Это речь, которую мы слышим, когда разговаривают за стенкой в помещении. Мы часто различаем, кто и как говорит, но не различаем, что или о чем говорится. При этом мы можем определить в общей форме и голоса незнакомых людей, речь которых мы никогда ранее не слышали, пол, с определенной точностью возраст, состояние и манеру говорить того человека, речь которого мы слышим. Стенка «съедает» се-

мантическую часть речевого сигнала, т. е. является тем естественным фильтром, который пропускает лишь собственно речь (интонацию, индивидуальные особенности диктора, мелодические особенности речи), но задерживает семантику сигнала, его смысл, то, что говорится. Это объясняется тем, что информацию о фазах сигнала несут его «пики», по которым и ведется анализ его ритмики. Инерционность стенки как фильтра не позволяет передать наиболее кратковременные и резкие переходы давления по его величине и направлению. С учетом различных модификаций можно сказать, что в известном смысле фильтр-стенка пропускает мелодику сигнала, но не пропускает его ритмику. И здесь возникает вопрос о соотношении между разборчивостью сигнала и его распознаваемостью.

Кстати, стенка-фильтр является своеобразным демодулятором, выделяющим собственно «речь» в речевом сигнале и поглощающим его семантику — «язык». В известном смысле можно сказать, что фильтр-стенка пропускает большую часть сведений о форме импульсов, но не пропускает сведения об их положении на временной оси.

Анализатор, который необходим для прямого ввода речевых команд в ЭВМ, должен решать противоположную задачу: выделить и передать семантику сигнала и «поглотить» информацию о речи диктора.

Итак, задача автоматического распознавания речи, если иметь в виду универсальное, работающее с голоса любого диктора устройство, сводится к отделению стандартных, устойчивых частей нашей речи от индивидуальных, от интонации.

Уже долгие годы исследования в области анализа речи, попытки найти фонетический код речевого сигнала базируются на исследовании так называемого амплитудно-частотного спектра. Однако длительные усилия лучших ученых и конструкторов не привели к значительному результату.

Есть, разумеется, и другие мнения. Так, крупнейший советский специалист в области информации и связи академик А. А. Харкевич много лет назад предлагал взять в качестве основного компонента для автоматического распознавания речи изменение звукового давления по времени.

Согласно гипотезе, выдвинутой видным советским

специалистом в области техники связи профессором А. А. Пироговым, нет однозначного соответствия между физическим качеством звуков речи и амплитудно-частотным спектром. Различия между фонемами следует искать в характерных для каждой из них изменениях спектра, а не в самом спектре. Гипотеза Пирогова в известном смысле исходит из динамического, а не статического подхода к анализу спектра. Но подтверждения она не нашла.

Автор этой статьи предполагает, что изменения частоты колебаний в процессе речи характеризуют не смысловую ее часть, а произношение, дикцию. Амплитудно-частотный спектр не определяет совсем или определяет косвенно смысловую часть речевого сигнала.

Другой важнейший вопрос, который стоит перед исследователями и конструкторами, — вопрос о той единице речевого сигнала, которую необходимо взять за основу для распознавания. Наиболее привлекательно было бы взять первичную единицу — фонему (условно ее можно назвать звуком-буквой). Например, слова «дам» и «там» различаются фонемами «д» и «т», а специфическим признаком такого отличия является «глухость» (звонкость) звука. Так вот, говоря о положении фонемы в слитной речи, де Соссюр отмечал, что она, с одной стороны, всегда как-то связана с другими фонемами, а с другой — представляет собой автономную единицу. В русском языке 41 фонема, которые объединяясь, дают все многообразие звучащей речи.

Возникает вопрос: а можно ли дать фонеме как автономной единице адекватное математическое описание, с помощью которого мы могли бы выделить ее из всего речевого потока вне зависимости от характера текста и произношения диктора? После этого мы могли бы дать возможность ЭВМ четко выделять фонетический код и читать слова и целые фразы пофонемно. Свести распознавание речевых сигналов к анализу нескольких десятков фонем — вот перспективное решение труднейшей и актуальнейшей задачи.

Но для этого нужно вначале перевести фонему из акустического вида в какой-нибудь иной, сделать ее пригодной для приборного анализа. В свое время исследователи так и пытались поступить. Но, как мы уже говорили, успеха достичь не удалось. Сразу отметим,

что послоговый и пословный анализы также пока не дали кардинальных результатов.

Наиболее экономичный способ нахождения фонетического кода речевого сигнала — прочесть слово пофонебно, как мы читаем письменный текст побуквенно.

Очень характерно, что развитие письменности наглядно отражает как процесс эволюции речеобразования, так и осмысление этого процесса самим человеком. Вначале были целые блоки фраз — человек мог целое большое понятие выразить сразу. Отражение этого процесса — рисованные символы. Затем появились пиктография и иероглифы. В одном из древнейших центров цивилизации — Шумере имели уже слоговое, иероглифическое письмо, выделение неких отдельных единиц.

Попутно отметим одну чисто техническую сторону шумерской письменности. В ней использовался эффект света и тени. Подобная клинопись сохранилась и на территории СССР, в Армении. В наши дни этот метод находит все большее применение в очень перспективной системе сбора, обработки и передачи информации — фазовой рельефографии.

Вероятно, запись речевой информации, а возможно, и вообще запись информации в нашем мозгу происходит в форме фазовой рельефографии. Только вместо светового луча действуют электрические или даже механические колебания.

В письменности другого центра мировой цивилизации и культуры — Библие — слоги членились на звуки, и это давало возможность создавать с помощью ограниченного числа письменных символов бесконечное число сообщений.

Очень характерно, что форма букв древнееврейского алфавита условно отображала конфигурацию речевого тракта при произнесении звуков, соответствующих этим буквам.

Как отмечает член-корреспондент АН СССР С. С. Лавров, «долгий путь развития естественных языков привел к тому, что в большинстве из них в устной речи используется несколько десятков звуков, а в письменной — несколько десятков букв и других знаков. По-видимому, это количество оптимально для человека». Свести анализ и распознавание речевых сигналов к

нескольким десяткам фонем — это было бы оптимальным решением.

Итак, в теоретическом плане для решения вопроса о нахождении фонетического кода речевого сигнала представляется необходимым:

1) отказаться от спектрального анализа и перейти к анализу изменения звукового давления по времени;

2) основной единицей распознавания считать первичную фонетическую единицу-фонему (звук-букву);

3) исходить из того, что фонетический код характеризуется тем, что импульсы, подобные по форме, занимают определенное положение на временной оси, т. е. подобны по фазе.

В настоящее время существуют два основных типа распознающих машин: 1) со сравнительно большим словарем, но с обязательным обучением, подстройкой ЭВМ под диктора и 2) с маленьким словарем, но работающих без предварительного обучения.

Системы распознавания речи, работающие с произвольным диктором, строить гораздо труднее, чем системы, в которых происходит настройка на диктора, — это объясняется тем, что на характеристики звуков речи накладываются параметры индивидуальной манеры произнесения звуков каждым диктором.

Системы распознавания речи, работающие на голос произвольного диктора — носителя нормы произношения данного языка (русского, английского, японского, американского диалекта и т. д.), которые ориентированы на использование ограниченного словаря, строятся различными способами.

Во-первых, в некоторых промышленных системах используют для небольших рабочих словарей большой набор эталонов для каждого слова (система фирмы Dialog Systems, США). Распознавание осуществляется сравнением неизвестного слова со всеми эталонами. Тот эталон, который более всего походит на неизвестное слово, и определяет результат распознавания. В США собиралась статистика произнесений по всем штатам различными возрастными категориями дикторов. Так что на одно слово (а словарь составил всего 20 слов — названия цифр и некоторые служебные слова) выходило 500 эталонных произнесений, каждое из которых характеризовалось своим амплитудно-частотным

динамическим временем. Это — первый путь построения многодикторских систем распознавания речи.

Распознавание основано на том, что неизвестное слово, например seven (семь), будет более похоже на один из эталонов этого слова или на один из эталонов, характеризующих обобщенный эталон этого слова по нескольким дикторам со сходной манерой произношения, чем на слово «восемь» (eight) или какое-то другое.

Во-вторых, использовались параметры первичного описания, которые в отличие от динамического спектра описывают речевой сигнал так, что таблица этих признаков для каждого слова в существенно меньшей степени зависит от индивидуальной манеры произнесения, чем, скажем, спектрограммы. Такие признаки получаются со специальной аналоговой аппаратуры, которая их выделяет, или с помощью цифровых, алгоритмических методов [2].

Третий способ построения систем, работающих на голос произвольного диктора, заключается в некоторой относительно небольшой подстройке алгоритмов распознавания под голос после произнесения новым диктором некоторых парольных, обучающих фраз, не включающих все слова словаря. В США в 1985 г. была разработана система автоматического распознавания 5000 слов из словаря деловой переписки фирмы IBM [3]. Чтобы пользоваться этой системой распознавания, оператор должен в течение 5—20 мин ее обучать, произнося в микрофон по слову несколько предложений, которые заранее известны машине. Раз они известны, то не составляет труда рассегментировать, разбить каждое слово на отрезки, соответствующие отдельным фонемам, которые тоже известны, и на этих фонемах, находящихся в определенной последовательности, измерить их параметры — средние спектры и другие энергетические и частотные характеристики звуков.

В дальнейшем измеренные таким образом характеристики используются во вновь конструируемых эталонах слов и в алгоритмах распознавания (как опорные характеристики эталонов), которые, в свою очередь, используются при распознавании, — с ними сравниваются неизвестные слова. Нужно отметить, что в этой американской системе применяются два рабочих словаря: один большой — 5000 слов и один маленький (но трудный для распознавания), который включает 78

слов — цифры, названия букв и знаков препинания. Этот второй словарь нужен для надиктовывания любого слова, не входящего в большой словарь, по буквам, как мы обычно говорим.

Если, например, надо ввести в ЭВМ (распознать) какую-то аббревиатуру или фамилию, которых, как правило, нет в общем словаре, тогда диктор говорит «произношу по буквам», и система настраивается на побуквенную диктовку (распознавание нового слова).

Очевидно, машины с очень большим, может быть, с неограниченным словарем, но с подстройкой под диктора будут созданы в ближайшее время. Как сообщил в своей беседе с членом-корреспондентом АН СССР Н. Т. Федоренко видный японский писатель Кобо Абэ, «мы накапуне создания и выпуска в Японии новых компьютерных устройств, которые будут записывать с голоса диктующего и выдавать письменный текст. Правда, лишь с голоса определенного человека. Словом, машина, как это ни фантастично, будет признавать только голос своего хозяина» [4].

Итак, наряду с персональным компьютером в Японии, очевидно, появится персональный диктограф — устройство, печатающее прямо с голоса. Это будет доведение до кондиции распознающего устройства одного из двух основных направлений в создании распознающей техники.

В нашей стране задаче создания устройств, работающих по речевым командам, придается огромное значение. В план фундаментальных исследований АН СССР включена программа «Исследование принципов реализации акустического диалога «человек — ЭВМ».

..Один из коллективов, решающих задачу автоматического распознавания и синтеза речи, — Вычислительный центр АН СССР. Соответствующими исследованиями в нем руководит кандидат технических наук В. Н. Трунин-Донской. Он рассказывает, что, как и за рубежом, при разработке отечественных систем делается упор на такие, которые работают без предварительной настройки на голос диктора, в основу которых должны быть положены принципы фонетической обработки речевого сигнала.

Уже лет 20 назад мы демонстрировали систему, работающую от голоса произвольного диктора. Она распознавала 58 слов с точностью 93%. Из этих слов со-

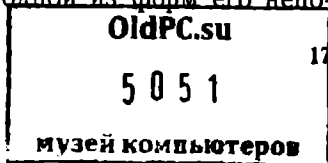
ставлялись предложения-задания на вычисление математических выражений. Любой человек, настроив систему на свой голос, последовательностью слов давал задание на вычисление различных значений математических функций, иногда достаточно сложных. Подобный речевой ввод будет, по-видимому, в сверхмощных ЭВМ пятого поколения, которыми станут пользоваться не только профессиональные программисты, но и просто специалисты самых различных областей и даже неспециалисты — ведь формой общения с машиной станет естественная речь.

Сейчас в ВЦ АН СССР разрабатывается система на микро-ЭВМ, которая должна работать от голоса произвольного диктора и иметь возможность быстрой смены словаря (и рабочего языка) в зависимости от требований потребителя. Конечно, словарный состав такого рода систем пока невелик — примерно 40—50 слов. Точность распознавания — около 95% — для ряда применений вполне удовлетворительная.

Мы подходим также к проблеме создания системы понимания слитной речи, к изучению ее характеристик... И пусть только для решения производственных задач, но такие устройства уже существуют. Одно из них — ИКАР. О том, как оно работает и где может использоваться, я попросил рассказать одного из разработчиков кандидата технических наук С. Б. Аврина.

— Сергей Борисович, такой «провокационный» вопрос: все время говорят, что ЭВМ стала понимать нашу речь, а вот есть ли уже реально действующие устройства и если есть, то каково их использование? Если же это только научно-техническая экзотика, то когда же появятся такие устройства, которые можно будет использовать в производстве и управлении?

— Сейчас ГлавУКС Мосгорисполкома приступает к обучению большой группы своих сотрудников работе с ЭВМ. В курс входит и обучение работе с ИКАРОм. После прохождения курса обучения инженерно-технический персонал будет на «ты» с современной техникой и в том числе с устройством ввода речи. Впоследствии эти лица смогут без каких-либо затруднений работать как с ИКАРОм, так и с любым другим устройством, понимающим речь человека. Таким образом, можно считать подобное использование ИКАРа одной из форм его непо-



средственного использования в сфере производства и управления.

— А почему надо учиться работе с устройством речевого ввода? По телефону, скажем, я разговаривать умею безо всяких курсов.

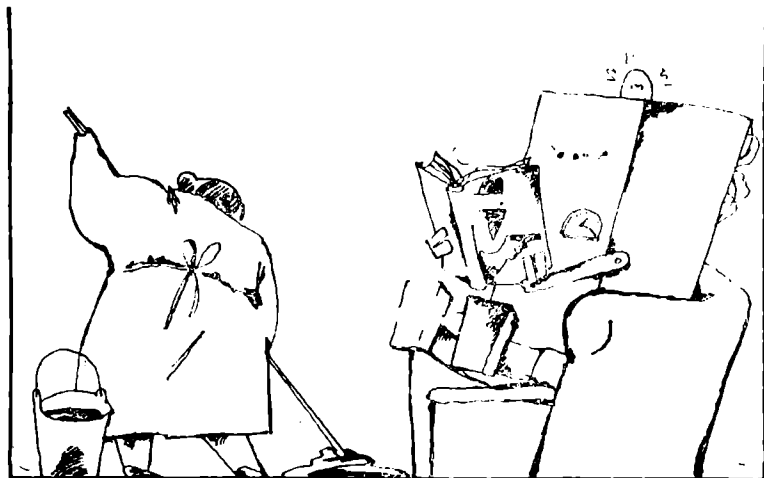
— Это вам только кажется. Ведь вы должны знать, что надо набрать семь цифр, а не какое-либо другое их число. Далее, вы ведь набираете их не произвольно, а строго в определенном порядке. Вы отличаете длинный гудок от короткого, знаете, что каждый из них означает. Так что определенное обучение вы прошли, хотя и не на специальных курсах.

При работе с речевым устройством пока надо в какой-то мере приспособливаться к нему.

Дело в том, что человек за тысячи лет эволюции научился «выжимать» из получаемого речевого сигнала необходимую информацию, нивелируя особенности произношения собеседника. Машина же пока этого не умеет. Поэтому, работая с ней, необходимо стараться произносить слова всегда более или менее одинаково, не слишком варьруя интонацию.

— Как конкретно выглядит работа с устройством?

— Представим себе, что один из соответствующих работников задает через ИКАР вопрос ЭВМ: «Выполнение плана строительно-монтажных работ за второй квартал». Машина, поняв это, ищет соответствующие



данные и пишет их на экране дисплея. Лишь бы слова запроса были заложены в память машины, связанной с устройством ввода.

— А сколько может быть таких слов?

— До 200. В рамках этого количества их можно заменять, т. е. «стирать» одни и записывать другие.

— А как происходит распознавание речевого сообщения в устройстве?

— Сразу отметим, что процесс распознавания состоит из двух частей: обучения системы и собственно распознавания. При обучении каждое слово из словаря нужно произнести один раз. Характеристики произнесенных слов будут использоваться как эталоны для сравнения с ними при распознавании. Эти характеристики получают так: из микрофона сигнал поступает в спектральный анализатор, представляющий собой набор стандартных для радиопромышленности фильтров звуковых частот. На выходах этих фильтров получаем сигналы, которые, как и входной речевой, развиваются во времени. Таким образом, каждое слово представляется трехмерной «картинкой» в координатах амплитуда — частота — время. Далее на этой «картинке» находятся характерные для этого слова точки (признаки), которые и запоминаются в качестве эталона данного слова.

— Ну а каков механизм собственно распознавания?

— При распознавании для произнесенного слова, так же как и при обучении, находятся характерные признаки, набор которых сравнивается со всеми эталонными наборами, после чего выбирается наиболее похожий из них. Слово, набор признаков которого выбран, и является результатом распознавания. Для контроля правильности распознавания оно пишется на индикаторной панели ИКАРа. Эталоны могут использоваться многократно, для чего они должны храниться во внешней памяти ЭВМ, с которой работает ИКАР. Поскольку устройство понимает речь только конкретного человека, на речь которого оно в данный момент настроено, то человек перед началом работы должен «назвать» себя. ИКАР позволяет работать с ним в разное время 256 различным людям, каждый из которых имеет словарь из 200 слов, или меньшему числу людей при пропорциональном увеличении числа словарей у каждого из них.

— А как часто ошибается устройство?

— Если человек выполняет правила произнесения, о которых я говорил, то редко. Достаточно сказать, что оно правильно понимает 95 из 100 слов даже при достаточно высоком уровне шума (80 дБ). А у человека, который работает с ИКАРОм регулярно, надежность распознавания может быть еще выше.

— Какова возможность применения ИКАРа в сферах производства и управления?

— Устройство можно применять в системах автоматизации проектирования (речевое управление вводом и коррекцией графических и других данных), в АСУ (ввод данных при контроле качества изделий, при складских и почтовых операциях и т. п.), в качестве входного звена информационно-поисковых систем, для непосредственного ввода и коррекции программ работы ЭВМ, программирования станков с ЧПУ, управления различными приборами и механизмами для инвалидов и т. п.

К настоящему времени пользователи еще только начали осваивать в эксплуатации первые образцы ИКАРов в системах автоматизации проектирования печатных плат, тренажерных и информационно-поисковых системах и для программирования станков с ЧПУ, поэтому суммы реального экономического эффекта от их применения пока не известны.

Однако имеющиеся зарубежные данные по использованию устройств, аналогичных по своим параметрам ИКАРу, включая стоимость, показывают, что рост производительности труда составляет при сортировочных операциях 50%, при программировании станков с ЧПУ — 95, при контроле качества изделий — 30, при перевозке грузов и их инвентаризации — 90, в задачах автоматизации проектирования — от 8 до 25% [5, 6, 7].

Институтом кибернетики АН УССР им. В. М. Глушкова и Минским отделом ЦНИИ связи создана система речевого диалога «Речь-1».

Я попросил показать ее действие на одной из выставок.

— Продиктуйте, пожалуйста, текст.

— Пусть машина скажет: «Москва — один из красивейших городов мира».

Текст набирается на клавиатуре, и через краткое время очень разборчиво и довольно «приятным мужским голосом» система произносит его.

В этот момент к экспонату подошел корреспондент радио. Набран текст, и он записывает на магнитофон: «Здравствуйте, дорогие товарищи радиослушатели. Вас приветствует система «Речь-1».

Но система не только «говорит», т. е. осуществляет синтез речи. В ней объединены функции распознавания и синтеза. Это комплексная система.

...Произношу в микрофон целый ряд слов, причем не только на русском, но и на английском и немецком языках. Они были предварительно записаны мной на бумаге, и против каждого я поставил его номер по порядку, произносил их тоже в этом порядке.

Система «слушала» мои слова и «обучалась» (о технике процесса мы расскажем дальше. Однако сразу отметим, что распознавание состоит из двух процессов: обучения системы и собственного распознавания).

Затем я услышал ее «голос».

— Готова распознавать. Говорите!

Произношу те же слова, но уже в другом порядке. Система все равно распознавала их: на индикаторном табло появлялась цифра — номер слова по порядку записи.

Почти все слова распознаны правильно. Но было и несколько сбоев. Как только я произвольно изменил произношение, машина тут же «поставила меня на место». Она совершенно четко «человеческим голосом» сообщила, что не понимает. Мне пришлось следить за тем, чтобы оба раза — и при обучении, и при распознавании — была одинаковая интонация. Тогда номер произнесенного слова высветился правильно.

Вкратце механизм действия системы при распознавании речи таков. Мы произносим в микрофон отдельные слова (команды). Из микрофона сигнал поступает в специальный анализатор. В нем производится вычисление значений признаков, характеризующих звуки речи.

Это каскад фильтров, «процеживающих» характеристики речевого сигнала; он включает в себя 16 каналов. Это стандартные для радиопромышленности фильтры звуковой частоты. Затем происходит сравнение характеристик. Оно очень просто: анализируются последовательно два соседних канала: там, где напряжение выше, — «1», там, где ниже, — «0». При этом применяют-ся типовые микросхемы.

Сигнал из аналоговой формы преобразуется в цифровую. По каналу связи уже передается не сам сигнал, а как бы сведения о его составляющих, при этом резко возрастает помехоустойчивость и разборчивость передаваемого сигнала.

В памяти машины формируется как бы обезличенный образ слова, некое цифровое его описание. При обучении может сформироваться до 200 таких образов.

Итак, процесс обучения сводится к преобразованию речевого сигнала из аналоговой в цифровую форму, состоящую из «1» и «0». Этот образ слов можно увидеть на экране в виде таблицы. Количество строк в ней зависит от длины произнесенного слова. Количество же столбцов всегда постоянно — 16. Оно определяется разрядностью ЭВМ. В целом процесс обучения основан на спектральном анализе речевого образа и наперед заданных правилах формирования эталона.

Ну а каков механизм собственно распознавания?

В процессе произнесения слова (одного из тех, что были произнесены при обучении) в машине формируется точно такого же типа описание его (в цифровой форме), как и сформированный при обучении эталон, производится сравнение описания произнесенного слова с эталоном. Номер эталона, наиболее похожего на описание произнесенного слова, выдается на индикаторную панель в виде цифры.

Но в пределах 200 слов мы можем записать в память машины специальную программу. Тогда она сможет не только распознавать соответствующие слова — команды (из числа записанных в ее память), но и выдавать их на экран дисплея или печатать на бумаге электропечатающей машинкой типа «Консул» и одновременно произносить синтезатором.

Система может распознавать в один прием 200 любых слов на любом из индоевропейских языков. Например, можно сказать в микрофон 50 слов на русском, 50 — на немецком и 100 — на чешском и т. п. По произнесении все их можно стереть и записать 200 любых других слов.

Для решения данной задачи была признана наиболее подходящей очень высокого уровня универсальная и дешевая отечественная ЭВМ — «Электроника-60». В ее память заложена специальная программа. Попутно отметим, что в Институте кибернетики им. М. В. Глушко-

ва на базе машины БЭСМ действует устройство, распознающее до 2000 команд.

При синтезе речи в зависимости от текста, вводимого с клавиатуры дисплея, перфоленты, магнитной ленты, магнитного диска, программа микро-ЭВМ вычисляет параметры фонем. Они заложены в ее память. Каждые 10 мс в синтезатор в кодах поступают просчитанные параметры фонем. В нем находится ЦАП, преобразующий сигнал из цифровой в аналоговую форму (в уровне электрического напряжения). Аналоговые сигналы управляют специальными формантными фильтрами (частотой и амплитудой). Эти фильтры соответствуют частотам речевого сигнала (речевого тракта). В результате на выходе мы слышим искусственную речь. Хотя синтез имеет неограниченный словарь, он может осуществляться лишь на двух языках — русском и украинском. Для каждого языка заложена специальная программа.

Необходимо подчеркнуть, что по некоторым параметрам система превосходит зарубежные аналоги. Это прежде всего использование в ней серийно выпускаемой универсальной ЭВМ. За рубежом в этих целях применяются вычислительные машины специального назначения. Большое достоинство и в том, что в память машины введены параметры всех фонем русского языка и их взаимозависимости между собой.

Но, естественно, возникает вопрос: почему, имея в памяти эти характеристики, система распознает только до 200 слов (в каждом цикле) и делает это, распознавая слово лишь целиком?

К тому же система очень чувствительна к однородности произношения при обучении и распознавании.

Потому что при распознавании, при анализе мы имеем дело с конкретным человеческим голосом, который у всех людей сугубо индивидуален, подобно индивидуальности почерка. Создатели устройства исходили из того, что наиболее полно информация о речевом сигнале сосредоточена в слове, а не в отдельной фонеме, поэтому они ввели пословное, а не фонемное распознавание. В синтезаторе же мы имеем дело с обезличенной информацией как при ее вводе (нажим клавиатуры), так и на выводе — машинный синтезированный голос.

В наши дни чрезвычайно актуальна проблема прямого речевого ввода информации в ЭВМ и получения ре-

чевого ответа от нее. «Речь-1» позволяет вести оперативный диалог человека с машиной в наиболее привычной и удобной для нас форме — голосом. У нее высокая надежность распознавания — 95%, она малогабаритна и легка. Сфера применения достаточно широка. Она может использоваться как диалоговое устройство в АСУ, в системах автоматизированного проектирования, системах связи, информационно-справочных системах, робототехнике и др. Использование системы повышает производительность труда человека и эффективность машины [8].

Мы редко придаем значение тому, как говорим, с какой интонацией. Но существуют профессии, для которых точное и тонкое владение своим языком, умение пользоваться интонационными оттенками речи являются показателями профессиональной подготовленности. Это — актеры, дикторы, певцы и переводчики. Обученные интонации — один из важнейших этапов в их специальной подготовке.

На основе учета значимости этого фактора базировался получивший широкое применение метод югославского профессора Петера Груберины. Обучение проводилось с помощью сконструированного им специального аппарата. Он пропускал только те частоты, которые не типичны для родного языка. Этим устранялась важная помеха при изучении иностранного языка — когда его звуки слышатся так, как они произносятся в родном. Несмотря на всю привлекательность и эффективность метода, он имел существенный недостаток — процесс обучения был чисто акустическим, без воздействия зрительных образов.

Для облегчения и ускорения изучения интонационных особенностей речи, в первую очередь иностранной, в Московском педагогическом институте иностранных языков им. М. Тореза создан специальный прибор «Обучающий комплекс-3» — ОК-3 [9].

Вот что говорит о нем заведующий лабораторией экспериментальной фонетики института А. П. Беликов.

— Прежде всего разрешите рассказать интересную, несколько курьезную историю, произошедшую с великим русским певцом Ф. И. Шаляпиным в начале его артистической карьеры. Когда он плыл на пароходе, к нему подошел пассажир-англичанин и завел разговор. Шаляпин знал тогда всего одно английское слово yes (да).

И он вставлял его в ходе «беседы» настолько выразительно, с такими интонациями, что у англичанина сложилась полная иллюзия знания языка его отзывчивым и приятным «собеседником». При прощании он долго и благодарно пожимал ему руку. Вот что значит интонация в речи.

Классический вариант обучения интонации предусматривает в общем случае диалог между преподавателем и обучаемым. Преподаватель, произнося правильный интонационный вариант, задает обучаемому своеобразный уровень, достижение которого и является конкретной его целью. Преподаватель выполняет в этой схеме роль эталона и роль корректирующего звена.

С помощью ОК-3 можно осуществить иную схему процесса обучения. В ней в процессе обучения интонации включается зрительный образ в виде соответствующей диаграммы на экране телевизионного устройства. Обучаемый не только слышит, но и как бы видит свою интонацию. Это позволяет преподавателю и обучаемому, кроме естественного диалога на слух вести его на языке зрительных образов.

Из речевого сигнала выделяется физический параметр, характеризующий интонацию — частоту колебаний голосовых связок. Графики изменения этой частоты на протяжении произносимой фразы (интонационный контур) запоминаются в электронной памяти и демонстрируются на экране монитора. График может стираться и записываться вновь. Можно записать параллельно и сравнивать интонационный контур учителя и обучаемого. На экране монитора одновременно с интонационными контурами могут демонстрироваться в виде стандартного телевизионного сигнала изображения графиков, нарисованных фломастером на листе бумаги, или методические программы, инструктирующие обучаемого. Это достигается с помощью портативной видеокамеры, которая может передать изображение с доски или с листа бумаги на основной экран. Включение в процесс обучения зрительного канала информации позволяет быстрее и эффективнее достичь цели обучения: безакцентного владения интонацией иностранного языка.

ОК-3 — многофункциональное (многоцелевое) электронное устройство. Оно может использоваться как индивидуально для тренажа, для развития своего голосо-

вого аппарата, так и при обучении интонации иностранного языка. Техническая реализация такой схемы защищена двумя авторскими свидетельствами.

В заключение следует отметить, что авторы прибора награждены медалями ВДНХ СССР; ОК-3 успешно демонстрировался на шести международных выставках. Эта разработка удостоена Золотой медали Лейпцигской ярмарки.

ЛИТЕРАТУРА

1. Соссюр Ф. М. Курс общей лингвистики. — М., 1933. — С. 86.
2. Цемель Г. И. Оpozнaвание речевых сигналов. — М.: Наука, 1971.
3. ТИИЭР. — 1985. — № 11.
4. Литературная газета. — 1985. — 20 ноября. — № 47 (5061).
5. Авторское свидетельство № 1 040 512 БИ № 33, 07.09.83.
6. Авторское свидетельство № 1 076 938 БИ № 8, 28.02.84.
7. Авторское свидетельство № 1 094 049 БИ № 19, 23.05.84.
8. Авторское свидетельство № 1 159 059 БИ № 20, 30.05.85.
9. Авторское свидетельство № 824 267.

Компьютерный синтез речи

В. Н. Сорокин

В технически развитом обществе сложность и скорость производственных процессов и процессов управления экономикой пришли в противоречие со способностью человека получать, обрабатывать и передавать информацию. Имеющиеся где-либо сведения становится все труднее получать. Например, считается, что если научное исследование стоит не более 100 тыс. долл., то часто его дешевле выполнить заново, чем разыскать нужные результаты. Поэтому доступ к информационным ресурсам становится столь же важным, как и доступ к сырьевым ресурсам, а увеличение скорости обработки информации необходимо для обеспечения устойчивости экономических систем и повышения производительности труда.

Однако трудности, созданные быстрым развитием техники, могут быть решены средствами, которые эта техника предлагает в виде электронных цифровых машин, способных не только управлять производственными процессами, но и обрабатывать информацию, необходимую для человека. В Японии, США, Франции и других странах созданы национальные программы по разработке ЭВМ пятого поколения, обладающих элементами искусственного интеллекта. Эти машины смогут читать тексты, распознавать зрительные образы, произносить и понимать человеческую речь.

Возможности речевого общения с системами, располагающими искусственным интеллектом, придает большее значение, поскольку в ряде случаев связь человека с машиной должна осуществляться на расстоянии, например по телефонному каналу. Речь является для человека наиболее удобным и естественным способом обмена информацией. При таком способе общения человек делает меньше ошибок, меньше устает, быстрее реагирует, а скорость обмена информацией выше, чем при других способах — визуальном, тактильном, то-

нально-звуковом. Повышение роли человека как элемента производственных и экономических процессов привело к созданию и интенсивному развитию новой научно-технической области — речевой технологии или речевой информатики.

Речевая технология занимается разработкой систем автоматического распознавания и синтеза речи, распознавания (идентификации) и подтверждения (верификации) диктора, диагностики заболеваний по речевому сигналу, передачи и засекречивания речи по телефонным каналам. В основе речевой технологии лежат фундаментальные исследования свойств речевого сигнала, процессов речеобразования и восприятия, связи между речью и мышлением. Техническую базу речевой технологии создают ЭВМ, производительность которых непрерывно повышается, а стоимость и габариты уменьшаются.

Синтез речи является важным элементом речевой технологии, который обеспечивает получение информации в речевой форме от различных источников. Применение синтеза речи целесообразно в тех случаях, когда зрительный канал не работает (в темноте), либо по каким-то причинам не может быть использован, либо создает трудности в обработке информации. Если же связь осуществляется на расстоянии, как в информационно-справочных системах, автоматизированных системах управления, при общении с банком знаний или искусственным интеллектом, то синтез речи является основным способом передачи информации для человека.

ПРИМЕНЕНИЯ СИНТЕЗА РЕЧИ

Во многих странах мира быстро увеличиваются объемы выпуска разнообразных синтезаторов и накапливается опыт их практического применения. Имеющиеся сведения, с одной стороны, подтверждают экономическую и техническую эффективность синтезаторов речи, с другой — выявляют необходимость в дальнейшем улучшении их качества с тем, чтобы удовлетворить запросы всех потребителей. Приводимые ниже примеры существующих и потенциальных приложений систем синтеза речи свидетельствуют о большой экономической и социальной значимости этой области речевой технологии.

На протяжении нескольких десятилетий предпринимаются усилия по созданию читающих машин для слепых. Имеются опытные образцы, но широкому использованию читающих машин препятствует сложность и дороговизна оптико-механического блока, предназначенного для чтения текста с различными шрифтами. Тем не менее в США уже давно выпускается читающее устройство такого рода. Переход полиграфии на набор с помощью ЭВМ, при котором сначала создается копия книги в цифровом коде на магнитном носителе — ленте, устраняет необходимость в оптико-механическом блоке. Речь теперь может синтезироваться непосредственно по тексту, представленному в цифровом виде. Другая возможность состоит в чтении текста диктором, записи речи в сжатом виде, а затем в синтезе по параметрам. Этот способ уже реализуется для выпуска речевых курсов обучения.

В результате некоторых болезней, приводящих к расстройству управления движениями, люди лишаются возможности нормальной артикуляции, речь становится неразборчивой. Обслуживание таких людей часто требует специального персонала и стоит очень дорого, поэтому весьма актуально получение для них возможности сообщать о своих потребностях. Во Франции испытывается устройство, в котором больной фиксирует свой взгляд на одной из ячеек прямоугольного табло, находящегося перед ним. На каждой ячейке написано слово, и устройство, следящее за положением глаз больного, фиксирует координаты слова на табло, а синтезатор речи генерирует фразу, состоящую из слов, последовательно выбираемых больным.

В Калифорнийском университете было успешно испытано устройство предварительного диагноза шизофрении с помощью интеллектуальной системы, задающей пациенту вопросы в речевой форме, на которые он отвечал через клавиатуру алфавитного дисплея. Оказалось, что такой безличный опрос способствует уменьшению напряжения пациента и облегчает постановку предварительного диагноза, с которым пациент отправляется к специалисту.

Оценивая перспективы применения синтезаторов речи в сфере помощи инвалидам и больным, наряду с довольно высокой стоимостью этих устройств следует принимать во внимание и другие факторы. Поскольку чис-

ло нуждающихся в синтезаторах насчитывается сотнями тысяч, то при таком массовом производстве их цена должна понизиться, тем более что микропроцессоры и электронная память — один из немногих видов продукции, стоимость которой постоянно падает.

Другая гуманитарная область применения синтезаторов речи относится к обучению языку. На начальных этапах школьного обучения ребенку необходимо установить соответствие между произношением и написанием слов, и в этом ему может помочь синтезатор. В конце 70-х годов в США было выпущено устройство «Speak and spell» («Говори и пиши»), которое по заданной программе произносило слова и оценивало правильность их буквенного представления на алфавитной клавиатуре. Это устройство приобрело огромную популярность, тем более что стоимость его была вполне приемлемой — около 30 долл. Затем появились различные варианты обучающих устройств. В перспективе вполне возможен перевод некоторой части процесса обучения в диалоговый режим с участием синтезаторов речи.

Использование синтезаторов наиболее эффективно в диалоговых системах с автоматическим распознаванием речи. В ближайшем будущем ЭВМ пятого и последующих поколений окажутся в состоянии вести речевой диалог в заданной предметной области, выступая как советчики. Перевод на диалоговый режим с использованием синтеза и распознавания речи систем автоматизированного управления (АСУ) и информационно-справочных систем (в том числе адресных и телефонных) обещает большой экономический эффект вследствие круглосуточной готовности и исключения человека из операций поиска информации.

В новых поколениях низколетающих вертолетов и самолетов, при операциях на орбите информационная нагрузка на пилотов столь велика, что становится совершенно необходимой выдача информации о характеристиках полета и обстановке на борту с помощью синтезаторов речи. На бомбардировщиках ВВС США Б-52 установлена система речевого оповещения об аварийной ситуации. Синтезаторы речи используются в США и в тренажерах пилотов истребителей.

В США успешно прошла испытания система выдачи команд монтажнику сложной электронной аппаратуры с помощью синтезатора. Это позволило повысить произ-

водительность труда и сократить число ошибок за счет исключения отвлечения внимания для чтения инструкций на алфавитно-цифровом печатающем устройстве, связанном с ЭВМ, контролирующей процесс монтажа.

Многие технические устройства, в том числе и бытовые, стали настолько сложны, что для их эксплуатации необходимо знать иногда весьма труднопони­маемые инструкции. Это особенно затруднительно в ситуациях, требующих быстрого вмешательства в работу устройства. Перевод инструкций в речевую форму облегчает пользование такими устройствами. Например, в Японии выпускается швейная машинка с программным управлением, выдающая инструкции в речевой форме. Выпускаются также электронные калькуляторы, сообщающие результат действий через синтезатор. Речевой вывод в контрольных и измерительных приборах является не только удобным качеством, но и может предотвратить серьезные последствия, сообщая, например, об изменениях критически важных параметров технологических процессов, загазованности шахт, показателях пациента во время хирургической операции или в реанимационном отделении.

В токийском метро с конца 70-х годов объявления об остановках поручены синтезатору речи. Аналогичные информаторы устанавливаются в лифтах, в гостиницах, аэропортах и магазинах для объявлений и аварийного оповещения.

Синтезаторы речи могут стать средством творчества, если их использовать при создании радиоспектаклей и озвучивании сказочных персонажей.

В ближайшем будущем появятся роботы, которым потребуется общаться не только с человеком, но и друг с другом. Эти роботы могут быть электрически или программно несовместимы или находиться на расстоянии. Поэтому и здесь применение синтезаторов речи может оказаться вполне уместным. Наконец, синтезаторы речи являются мощными научными инструментами для исследования процессов речеобразования и свойств речевого сигнала, для изучения певческого голоса и патологии речи.

Из вышеприведенного краткого обзора следует, что область применения синтезаторов речи очень широка, а экономические эффекты и социальные последствия их использования могут быть весьма велики.

ИСТОРИЯ СИНТЕЗА РЕЧИ

Разработка синтезаторов речи тесно связана с исследованиями процессов речеобразования и восприятия, а также грамматической структуры языка. В некотором смысле развитие методов синтеза повторяет развитие речи и письменности.

В настоящее время считается, что вплоть до неандертальцев речь не возникала. Это связывают с особенностями строения ротовой полости и структуры мозга, в котором отсутствовали лобные доли и те зоны височной и теменной областей, которые связаны с функционированием второй сигнальной системы. Первые речевые сообщения представляли собой монологичные звукосочетания, неразложимые на элементы. Этими звукосочетаниями пользовались для приманивания и отпугивания животных («зверинные слова»), для обмена простейшей информацией, причем полагают, что раньше всех появились глаголы в повелительном наклонении.

По-видимому, осознание структуры речи сказалось и на развитии письменности. Примерно 6 тыс. лет назад в Шумере появилась иероглифическая письменность, в которой каждый иероглиф обозначал слово или понятие. Около 4 тыс. лет назад в Индии возникло слоговое письмо, а еще через тысячу лет в финикийском городе Библе впервые было создано фонетическое письмо, в котором каждому согласному звуку соответствовал отдельный знак, а гласные вообще никак не обозначались. В V—VI вв. до н. э. в Греции появилось «звукое» письмо, где обозначались также и гласные звуки. Такое изменение письменности отражает объективные свойства речи. Слово является наименьшим семантическим элементом, слог — наименьший элемент программы управления артикуляционными движениями, фонема — наименьший смысловоразличительный звуковой элемент речи. Интересно, что написание букв древнееврейского алфавита напоминало форму речевого тракта при артикуляции соответствующих звуков.

Речь оказывает большое влияние на психические процессы человека. В гипнозе с помощью слова можно изменять состав крови и биохимии организма, причем это воздействие проявляется и при прослушивании голоса гипнотизера, записанного на магнитную ленту. Поэтому в древности возникла своеобразная «наука», за-

нимающаяся речевыми заклинаниями, причем верили, что с помощью заклинаний можно воздействовать и на неодушевленную природу. Врачи древнего мира придавали большое значение слову как лечебному фактору. В эпоху Возрождения предпринимаются первые попытки исследования свойств голосообразования. Этому уделял внимание и Леонардо да Винчи. С развитием механики пытаются создать искусственный голос.

В 1779 г. немецкий ученый Кратценштейн, находившийся на службе в Санкт-Петербургской академии наук, представил свое сообщение в ответ на объявленный академией конкурс на лучшее объяснение процесса генерирования гласных звуков. Кратценштейн изготовил органные трубы различной формы, звучание которых имитировало гласные звуки. Другие исследователи с этой целью использовали зубчатые колеса с разным числом зубцов, возбуждавшие при вращении колебания прижатой к зубцам пластины. В конце XVIII в. венгерский инженер и изобретатель Ф. В. Кемпелен предпринял попытку механического моделирования речевого тракта. Его синтезатор имел мехи, создающие поток воздуха, и эластичную трубку, форма которой менялась руками для получения речеподобных звуков. Эта же идея механического моделирования была реализована в электромеханических синтезаторах Белла, Пэйджета, Риша. Такой синтезатор действует в Институте математики Сибирского отделения АН СССР.

Все эти устройства были лишь забавными игрушками и не имели практического значения. Только появление электронно-цифровой техники сделало реальным создание синтезаторов речи.

МЕТОДЫ СИНТЕЗА РЕЧИ

Методы синтеза речи зависят от практической задачи и условий использования. Сам термин «синтез» используется в различных смыслах, обозначая целый ряд методов, начиная от воспроизведения записанного речевого сигнала до генерации речи по заданному тексту. Иногда только последний метод признают «синтезом» речи, однако, как мы увидим, нет строгой границы между этим методом и различными способами восстановления речевого сигнала по записанным параметрам.

Все задачи, в которых требуется вывод речевого сиг-

нала, можно грубо разделить на два класса. Первый класс — это задачи, в которых число сообщений невелико и они не изменяются или изменяются редко. К ним относятся, например, системы типа «говорящие» часы, объявление остановок и этажей, считывание показаний приборов. Второй класс — это задачи с заранее не известным текстом, как, например, в «читающих» машинах для слепых или «говорящих» устройствах для немых. Между этими классами имеется целый ряд промежуточных задач.

Параметрический синтез. В задачах с фиксированными сообщениями не требуется синтез речи в строгом значении этого термина, а достаточно обеспечить воспроизведение записанного речевого сигнала. В простейшем случае речевой сигнал записывается в аналоговой форме на магнитную ленту или другой носитель и по заданному коду сообщения осуществляется его воспроизведение. Электромеханические устройства, однако, весьма дороги в обслуживании и недолговечны. Если же записать речевой сигнал в дискретной форме в электронную память, то длина сообщения зависит от емкости памяти и скорости передачи информации для выбранного способа записи.

Оценим скорость передачи информации для некоторых способов. Высококачественное воспроизведение речи достигается при частотной полосе 5 кГц, а полоса телефонного канала (до 3,4 кГц) обеспечивает лишь удовлетворительное качество. Для записи речевого сигнала в электронную память его нужно дискретизировать по времени и амплитуде. Минимальная частота дискретизации по времени, согласно теореме Котельникова, должна быть равна удвоенной частоте верхней границы спектра сигнала, т. е. 6,8—10 кГц. Квантование по амплитуде на 2^{10} — 2^{12} уровней обеспечивает необходимый динамический диапазон. Таким образом, скорость передачи речевого сигнала при сохранении его качества составляет от 68 тыс. бит в секунду при удовлетворительном качестве и до 120 тыс. бит при высоком качестве. При этом для памяти, например, в 64 Кбайт возможна запись речевого сигнала длительностью 4—8 с, т. е. трех—шести слов.

В таком способе, называемом импульсно-кодовой модуляцией (ИКМ), используются минимальные сведения о свойствах речевого сигнала и соответственно до-

стигается наименьшее сжатие информации. При этом требуются наименее сложные устройства для анализа и воспроизведения сигнала (для восстановления аналоговой формы требуется лишь пропустить последовательность импульсов через фильтр низкой частоты с частотой среза, равной половине частоты дискретизации). Поэтому характеристики ИКМ применяются в качестве опорных при сравнении различных методов записи речевого сигнала. Как видно, при ИКМ требуется большая память даже для сравнительно коротких сообщений. Вследствие этого было создано большое разнообразие методов, позволяющих уменьшить скорость передачи речевого сигнала.

Имеется несколько разновидностей ИКМ. Например, если использовать нелинейное квантование амплитуды сигнала таким образом, чтобы кванты увеличивались по мере увеличения амплитуды, или подвергнуть амплитуду нелинейному преобразованию логарифмического типа при сохранении равномерного квантования, то можно уменьшить число бит с 10—12 до 7, т. е. использовать для записи только 128 уровней сигнала (при соответствующем обратном преобразовании во время вывода сигнала).

Речевой сигнал не обязательно дискретизировать по времени с постоянным шагом. Следующий отсчет можно брать лишь тогда, когда амплитуда сигнала изменится на некоторую заданную величину. Такой способ называется адаптивной ИКМ. Еще большего выигрыша можно достигнуть, если записывать не амплитуду отсчета, а лишь приращение этой амплитуды относительно предыдущего отсчета. Такой способ называется адаптивно-разностной ИКМ. Эти приемы позволяют понизить скорость передачи речевого сигнала в несколько раз. Естественно, при этом усложняются блоки анализа и восстановления сигнала.

К разновидности ИКМ относится и группа методов, называемых дельта-модуляцией. При дельта-модуляции определяется не амплитуда сигнала в тот или иной момент времени, а знак приращения сигнала относительно предыдущего отсчета. Положительному приращению соответствует импульс (1 бит), а отрицательному — отсутствие импульса, или импульс противоположной полярности. При восстановлении речевого сигнала соответствующие этим импульсам кванты приращения сум-

мируются во времени. Информация о каждом отсчете при дельта-модуляции минимальна — 1 бит, но частота отсчетов должна быть значительно увеличена по сравнению с частотой Котельникова для обеспечения достаточной плавности изменения отсчетов сигнала. С математической точки зрения дельта-модуляция соответствует дифференцированию сигнала, а процесс восстановления — интегрированию. Величина порога при этом определяет максимальную скорость изменения в сигнале, которая, как известно, связана с шириной его спектра. Выигрыш в памяти для записи сигнала происходит вследствие подавления операцией дифференцирования низкочастотных компонент сигнала, которые в речи имеют наибольшую амплитуду. В дельта-модуляции также используются адаптивные и разностные методы. В итоге скорость передачи информации для речевого сигнала снижается до 20—40 Кбит/с, что сравнимо со скоростью передачи для ИКМ, однако устройства кодирования — восстановления сигнала оказываются проще.

Значительное снижение скорости передачи — до 9 Кбит/с и даже до 2 Кбит/с — достигается с использованием группы методов с линейным предсказанием речевого сигнала. Метод линейного предсказания состоит в вычислении значения сигнала в некоторый момент времени как линейной комбинации (взвешенной суммы) значений сигнала в нескольких предыдущих отсчетах. Предсказанное таким образом значение сигнала сравнивается с действительным, и разность этих значений квантуется и запоминается. Число отсчетов, по которым производится предсказание (т. е. число коэффициентов взвешивания), находится в диапазоне 10—14. Одновременно определяется период следования импульсов возбуждения (если звук образован с участием голосового источника), а также запоминается индикатор наличия шумового возбуждения для шипящих звуков. Коэффициенты линейного предсказания оказываются связанными с параметрами передаточной функции речевого тракта, поэтому синтез речевого сигнала сравнительно прост, но, конечно, он гораздо сложнее, чем восстановление сигнала по ИКМ или дельта-модуляции. В то же время анализ, т. е. вычисление коэффициентов линейного предсказания, значительно сложнее синтеза, поскольку приходится решать систему линейных уравнений высокого порядка (равного числу коэффициентов).

На основе линейного предсказания можно записывать весьма длинные сообщения. Например, в Японии созданы микросхемы, в каждую из которых помещаются данные, с помощью которых можно в течение 100 с воспроизводить речь хорошего качества, причем число таких микросхем в синтезаторе можно наращивать до 10. Описанные принципы кодирования речевого сигнала используются во множестве методов и устройств. Однако задачей этих методов изначально являлось сжатие полосы речевого сигнала при передаче по телефонному каналу. В этом случае важна не только степень сжатия, но и устойчивость передаваемых параметров к шумам и искажениям каналов связи. Эти факторы не влияют на выбор метода в задачах запоминания и восстановления речи, поэтому основными параметрами методов являются скорость передачи при заданном качестве речи, а также сложность анализа и синтеза. Сложность анализа, в свою очередь, важна лишь тогда, когда речевые сообщения не зафиксированы, а время от времени изменяются, для чего необходимо производить новую запись речевых сигналов.

Факторы стоимости и габаритов устройств электронной памяти влияют лишь до известного предела, когда сам метод синтеза фиксированных сообщений теряет смысл. Например, с появлением метода лазерной записи и считывания информации на одном диске размером около 15 см помещается от нескольких сотен до нескольких тысяч гигабайт, причем выборка информации осуществляется достаточно быстро для целей синтеза речи. Это сильно расширяет область применения параметрического синтеза, однако было бы неправильно думать, что это решает проблему синтеза произвольных сообщений.

Действительно, на такие диски можно записать огромное число слов и, вызывая их по очереди, компоновать (компилировать) из них фразы. Но тут вступает в действие новый фактор, влияющий на оценку качества синтеза. Известно, что в зависимости от длины и типа фразы и места слова во фразе изменяются длительность элементов этого слова и контур частоты основного тона, поэтому параметры заранее записанных слов не будут соответствовать требуемым, и субъективно оцениваемое качество и разборчивость таких фраз ухудшаются. Конечно, в определенных случаях

такое ухудшение будет вполне терпимо, но, очевидно, имеются и такие задачи, где оно неприемлемо. Запись же всех возможных фраз принципиально неосуществима. Таким образом, задача синтеза произвольного текста с высоким качеством речевого сигнала в принципе не может быть решена средствами параметрического синтеза речи.

В заключение этого раздела отметим, что параметрический синтез наряду с ограничениями на число возможных сообщений обладает большим преимуществом в виде высокой натуральности и разборчивости, а также возможности записи сообщений на любом языке, любом диалекте и от любых дикторов. Не случайно пилоты самолетов предпочитают для оповещения именно синтезаторы параметрического типа.

Компиляционный синтез. Стремление к расширению возможностей синтеза привело к попыткам использования в качестве элементов речи не слов, а отдельных звуков или слогов, вырезанных из естественной речи с применением какого-либо из описанных методов сжатия речевого сигнала. В этом случае, располагая конечным алфавитом элементов, можно было бы синтезировать произвольные сообщения. Многие исследователи, однако, стали жертвой нынешней алфавитной письменности, в которой каждая буква якобы соответствует одному звуку. В действительности же физические параметры звуков, обозначаемых одним и тем же фонетическим символом, сильно зависят от окружающих звуков. Поэтому использование отрезка речевого сигнала, воспринимаемого как некоторый звук в одном контексте, в другом окружении неприемлемо, и сигнал, составленный из таких элементов, звучит как неречевой.

Более успешным оказался такой подход, когда в качестве элементов для синтеза выбираются участки речи, где взаимное влияние звуков проявляется сильнее всего. Такими элементами обычно служат сочетания согласный — гласный. Для японского языка, например, такой метод вполне адекватен, поскольку в нем используются только такие сочетания. Доля слогов согласный — гласный велика в итальянском, французском, английском и русском языках, так что этот метод синтеза развивается для всех этих языков. Основная проблема в этом методе та же, что и при пословном синтезе.

Длительность, частота основного тона и другие важ-

ные параметры слогов изменяются в зависимости от окружения, места в слове, ударной или безударной позиции, места во фразе и типа фразы (восклицание, вопрос, повествование и т. д.). Для создания естественно звучащей речи необходима разработка правил изменения параметров элементов синтеза в зависимости от синтезируемого сообщения, а это чрезвычайно трудная задача. Поэтому компиляционный синтез находится как бы на полпути между параметрическим синтезом целых сообщений и синтезом произвольных сообщений по заранее не заданному тексту, т. е. синтезом речи по правилам. По-видимому, качество компиляционного синтеза никогда не достигнет качества параметрического синтеза, хотя этот вид синтеза может найти определенное применение в ограниченном круге задач.

Синтез по правилам. Когда говорят о синтезе речи по правилам, то имеют в виду генерацию речевых сигналов, соответствующих данному, заранее не известному тексту. При этом удобство алфавитного письма для чтения сменяется большими затруднениями для формирования команд управления синтезатором, поскольку фонема — буква является символом высокого уровня абстракции, малосвязанным с физическими характеристиками звуков речи. Кроме того, в тексте почти полностью отсутствует информация о том, на каком слове во фразе находится логическое ударение, на каком слове — словесное ударение, как должны меняться частота основного тона и другие параметры речевого сигнала. Точки, запятые, восклицательные и вопросительные знаки несут довольно мало информации об этих так называемых просодических параметрах. Вместе с тем несоблюдение правил, которым подчиняются просодические параметры, значительно ухудшает качество речи. Поэтому необходимо разработать систему алгоритмов для перевода письменного текста в управляющие команды.

Сложность этих алгоритмов зависит от соответствия между письменной и звуковой речью. Например, для японского, финского и белорусского языков эти правила проще, чем для русского, а для английского языка сложнее, чем для русского. Это приводит к тому, что в английском языке для правильного преобразования текста в одном из лучших синтезаторов записано около 12 тыс. буквосочетаний, тогда как в японском языке

почти все преобразования осуществляются по правилам.

Для русского языка характерно оглушение звонкого согласного в конце слова («дуб» произносится, как «дуп») и перед глухими согласными («подполье» произносится, как «потполье»). Глухие же согласные, наоборот, озвончаются перед звонкими согласными (слово «отбор» произносится, как «адбор»). В сочетаниях гласных иногда выпадают средние звуки («сердце» произносится, как «серце»). В безударной позиции вместо шести гласных А, О, У, И, Ы, Э остается лишь три типа: А-образные, И-образные и У-образные. Лингвисты составляют правила преобразования от букв к фонемам, однако, как мы помним из школьного опыта, в грамматике русского языка имеется множество исключений, доставлявших нам неприятности. К тому же положение ударения в слове иногда изменяется в разных словоформах («доска» — «доски»), причем правила сложны и непоследовательны. Это приводит к необходимости не только формирования правил, но и хранения в памяти синтезатора большого числа слов, представляющих исключения, или характеризующих группу слов, подчиняющихся одним и тем же правилам. Положение осложняется тем, что нормативный литературный язык сильно отличается в произношении от разговорного, причем существует ряд стилей, зависящих от предмета и условий разговора.

Определение логического ударения во фразе, которое характеризуется более четким и длительным произношением ударного слова, представляет собой нерешенную пока задачу семантического анализа текста, хотя в проблемно-ориентированных языках (например, в служебной документации) трудности семантического анализа могут быть меньшими. Для отдельного предложения зачастую вообще нельзя определить логическое ударение. Так, фраза «Петя пришел домой» может быть произнесена с ударением на каждом из трех слов в зависимости от намерений говорящего, тогда как, будучи представленной в буквенном виде и в отрыве от контекста, эта фраза не дает информации для определения логического ударения. Такие же явления, как «подтекст», пока даже не рассматриваются создателями синтезаторов.

Следует также отметить разницу между озвучиванием какого-либо текста, например чтением книги, и

речевым выводом сообщения, источник которого осведомлен о его смысле, как, например, в диалоге с АСУ или искусственным интеллектом. В этом случае источник сообщения может сам разметить некоторые просодические характеристики сообщения.

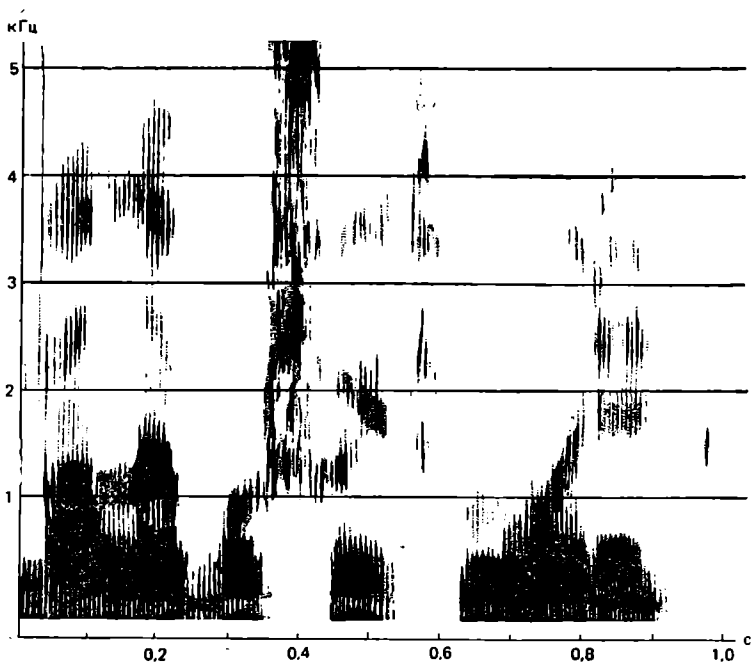
Формантный синтез. Синтез речи по правилам возможен лишь на основе какой-либо модели речеобразования. Простейшая модель, описывающая резонансные частоты речевого тракта и источники возбуждения акустических колебаний, положена в основу формантного синтезатора. В середине прошлого века Гельмгольц обнаружил максимум энергии в спектре гласных звуков и назвал его формантом. Позднее были найдены и другие максимумы, связанные с резонансами речевого тракта. Термин «форманта» поменял свой род — из мужского стал женским, и его стали применять для обозначения как резонансов, так и максимумов в спектре (которые не всегда совпадают). Для правильного восприятия звуков речи достаточно 2—3 формант, но для надлежащего качества требуется 4—5 формант.

В настоящее время практически единственным видом синтеза по правилам, обеспечивающим приемлемое качество звука, является именно формантный синтез. Его схема весьма проста: четыре или пять резонансных контуров соединяются последовательно или параллельно, причем в процессе синтеза управлению подвергаются только частоты (иногда и амплитуды) первых 2—3 формант. В соответствии со звонкостью или шумностью звука применяется либо голосовой источник возбуждения с периодической последовательностью импульсов, либо равномерный случайный шум, либо тот и другой источник (для звонких шумных звуков типа Ж, З, В). Для шумных звуков иногда используется подавление сигнала в определенных участках частотного диапазона, а чаще вводят дополнительный резонансный контур. Дополнительный контур с фиксированной частотой также важен при образовании носовых звуков М и Н. Разнообразные формантные синтезаторы, точнее, их акустический блок, уже выпускаются серийно многими фирмами за рубежом в виде единой микросхемы.

Качество звучания синтетической речи зависит не только от схемы генерации звука, но и от алгоритмов управления частотами формант. Сведения о траекториях формантных частот и местах расположения шумных

сегментов получают путем анализа так называемой видимой речи, которая образуется путем фиксации на электрочувствительной бумаге, огибающей на выходах гребенки полосовых фильтров. Степень почернения бумаги грубо отображает амплитуду сигнала в той или иной частотной полосе и дает возможность следить за изменением частотных характеристик звуков речи. Пример видимой речи показан на рисунке, где изображена запись фразы «Волна бушует в море», произнесенной мужским голосом.

Число правил, управляющих траекториями формант в зависимости от окружения звуков, весьма велико — в английском языке оно более 200, но и такое количество не дает возможности детально описать все имеющиеся случаи взаимодействия звуков. В формантном синтезаторе, разработанном в Минском отделе МОНИИС под руководством Б. М. Лобанова, таких правил гораздо меньше, поскольку была предпринята попытка классифицировать сами правила в зависимости от места артикуляции звуков речи.



Качество речи даже лучших формантных синтезаторов, однако, еще весьма далеко от качества естественной речи. Так, сравнительное испытание нескольких синтезаторов, проведенное в Университете штата Индиана (США), показало, что наилучший синтезатор Deck Talk имеет ошибку в восприятии слов в несколько раз большую, чем у естественной речи, причем даже незначительный шум в канале связи резко ухудшает разборчивость синтетической речи. Если во время прослушивания синтезатора человеку необходимо выполнять какую-либо работу, то не только ухудшается разборчивость, но и падает эффективность самой работы. Это связано с тем, что в синтетической речи нарушены соотношения между важными параметрами, а избыточность, необходимая для надежного восприятия при помехах, практически отсутствует. В результате мозг затрачивает больше времени на обработку синтетической речи, сигнал дольше находится в оперативной памяти, отнимая ресурсы, необходимые для выполнения работы. Замечен также эффект полного выпадения фраз при восприятии синтетической речи, связанный с тем, что из-за накапливающейся задержки в обработке сигнала в определенный момент оперативная память полностью занята и вновь пришедший сигнал в нее просто не помещается.

Таким образом, формантный синтез речи, хотя и обладает преимуществом неограниченного числа сообщений, может быть использован лишь в условиях хорошей акустической обстановки, при малой загрузке мозга слушателя и сильной заинтересованности в понимании синтетической речи. К числу недостатков формантного синтеза относятся большие трудности в создании индивидуальности голоса — даже проблема синтеза женского голоса пока решается неудовлетворительно.

Артикуляторный синтез. Существенное улучшение качества синтетической речи и приближение ее к качеству естественной речи возможны лишь при условии моделирования всех процессов речеобразования: акустических колебаний, движений артикулярных органов и процессов в периферической системе управления артикуляцией. Взаимодействия между этими процессами настолько сложны, что наблюдение лишь за поведением формантных частот, да еще в таком грубом представ-

к качеству естественной речи, которую приходится платить при моделировании процессов речеобразования. Надо сказать, однако, что уже существующая электронная технология вполне может обеспечить такую производительность.

Сравнивая свойства различных видов синтеза речи, можно предположить, что в будущем останутся лишь два типа синтезаторов — параметрический (для ограниченных сообщений) и артикуляторный (для произвольных сообщений).

ЭТИКЕТ ОБЩЕНИЯ С СИНТЕЗАТОРАМИ

Синтезаторы речи вскоре станут таким же обычным устройством, как телефон или телевизор. Однако поскольку синтезаторы предназначены для общения с людьми, то психологические факторы, влияющие на готовность принять или отвергнуть новую технику, в данном случае являются определяющими. Специальное исследование, проведенное в США, показало, что в половине случаев неудачное применение систем речевой технологии обуславливалось не их низким качеством, а психологической неподготовленностью людей, консерватизмом и плохим согласованием устройств речевой технологии с техническими системами, в которые они включены. Поэтому необходимо уделять внимание не только качеству и техническим характеристикам синтезатора, но обеспечивать надлежащий этикет общения, т. е. время и форму, в которой представляется речевое сообщение. Синтезатор, говорящий невпопад или пользующийся труднопонимаемыми речевыми оборотами, скорее всего, будет выключен. Иными словами, прежде чем наделить компьютер даром речи, нужно подумать о том, что же он должен будет говорить.

ЛИТЕРАТУРА

- Рабинер Л. Р., Шафер Р. В. Цифровая обработка речевых сигналов. — М.: Радио и связь, 1981.
Кейтер Д. Компьютеры — синтезаторы речи. — М.: Мир, 1985.
Сорокин В. Н. Теория речеобразования, — М.: Радио и связь, 1985.



Речевой ввод-вывод для персональной ЭВМ

В. А. Свириденко

Развитие процесса компьютеризации и повсеместное распространение персональных ЭВМ делают актуальной разработку аппаратных и программных средств, облегчающих взаимодействие пользователя со своим персональным компьютером. Такие средства представляют интерес и для подготовленных пользователей, и для людей, не специализирующихся в вычислительной технике.

Наиболее естествен вводом информации в голосовой форме. В этом случае через микрофон осуществляется ввод команд и данных (речевое управление или речевое программирование), и машина адекватно реагирует на них. При выводе результаты обработки могут быть представлены синтезированными речевыми фразами и одновременно (при необходимости) в визуальной форме на экране дисплея. Это позволяет значительно разгрузить визуальный канал восприятия, освободить руки, до-

пускает перемещение человека при общении с машиной, обеспечивает простоту обучения работе на машине и увеличивает скорость ввода данных.

Что же выполняет такие интерфейсные функции в подобной человеко-машинной системе? Ответ таков: средства распознавания слуховых образов и средства синтеза речевых сообщений по тексту.

Первые коммерческие «речевые изделия» были созданы японской фирмой НЕК в 1978 г. С начала 80-х годов на Западе начался бум в области речевой технологии. В различные приборы бытового, промышленного и военного назначения (часы, калькуляторы, игрушки, видеоигры, различные компьютеры, системы управления технологическими процессами, кабины боевых самолетов и т. п.) стали встраивать модули распознавания речевых команд и синтеза речи. Сдерживающим фактором в приложениях речевых устройств яв-

лялась лишь относительно высокая стоимость. Однако по мере развития и все большего использования интегральной технологии экономические затраты на речевые устройства имеют устойчивую тенденцию к снижению. Так, по данным фирмы Logica (Великобритания), с 1974 по 1984 г. стоимость систем распознавания речевых команд (РК) упала на четыре порядка и составляет в настоящее время десятки долларов.

В системах речевой технологии наиболее сложный — блок классификации речевых команд, к которому предъявляются высокие требования по быстродействию (более 10^7 арифметико-логических операций в секунду в зависимости от объема словаря РК), не может быть реализован полностью аппаратно, ему требуется программная поддержка. Именно такой блок (модуль) разработан фирмой IBM в 1979 г.

В 1981 г. фирма Interstate выпустила в продажу модуль VRT101, встраиваемый в интеллектуальный терминал на основе микро-ЭВМ, который стал главным компонентом ряда приборов с речевым вводом. Через год был создан набор из двух БИС VRC 100, включающий анализатор спектра ASA-16, выполняющий функции блока сопряжения и блока цифровой обработки и содержащий перепрограммируемое ПЗУ, на базе которого строится классификатор РК. Дальнейшим развитием этих инженерных решений стал модуль VRT300, совместимый с персональным компьютером фирмы IBM.

Японская фирма NEC создала набор из трех БИС для распознавания РК. В этот набор входят блок сопряжения MC 4780 (преобразующий речевой сигнал в цифровую фор-

му), специальный микропроцессор распознавания РК D7761D и микропроцессор D7762G (выполняющий функции управления), который может также обеспечивать интерфейс с главной ЭВМ.

Все эти разработки речевых модулей ориентированы на предварительное обучение диктором, который должен произнести каждую речевую команду от одного до трех раз. После преобразования речевого сигнала в цифровую форму и выделения параметров формируется его эталон, который заносится в память. Обычно словарь эталонов хранится в накопителе на дисках, откуда переписывается в ОЗУ непосредственно перед работой пользователя на ЭВМ. Очевидно, что с увеличением объема словаря возрастают затраты емкости ЗУ и число операций на классификацию РК, так как требуется сравнение вводимой речевой команды с каждым из эталонов. Процедура вынесения решения сводится к выбору наиболее похожего эталона. При этом следует считаться с возможными ошибками. Подобный подход к распознаванию РК, называемый дикторозависимым распознаванием дискретной речи, достаточно отработан, и современные устройства ввода РК в ЭВМ обеспечивают надежность в 90—99% при словаре 100—200 слов или словосочетаний. Этого вполне достаточно для многих применений.

Более интересно распознавание слитной речи, исключение зависимости от диктора, а также работа в условиях акустических помех. Совмещение всех этих возможностей в одном модуле слишком сложно, хотя развитие речевой технологии приведет в будущем к таким системам. Современные

же системы распознавания слуховых образов строятся на достаточно традиционных концепциях теории распознавания образов, хотя и имеются определенные успехи по распознаванию слитной речи. Но эксплуатационно надежных таких систем еще нет.

Следует отметить, что уже разработан одноплатный модуль РВ для ПЭВМ фирмы IBM, способный слитно распознавать произносимые команды из словаря десяти цифр (это так называемое распознавание составной речи). Алгоритмы распознавания фактически базируются на алгоритмах классификации РК и микропроцессорах с высокой производительностью.

Для ввода данных в речевой форме в структуру ЭВМ включается речевой синтезатор. Однако пока качество разборчивой речи еще недостаточно для их широкого применения.

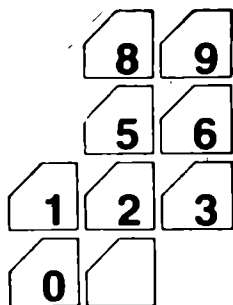
Отечественная промышленность выпускает системы речевого диалога «Речь-1», «Марс-2» и др., которые способны выполнять роль устройств ввода-вывода. Следующей задачей является разработка «речевых» БИС и одноплатных модулей, встраиваемых в ЭВМ. Другие направления исследований связаны с разработкой эффективных алгоритмов блока цифровой обработки и распознавания слуховых образов.

Одно из решений в этом направлении позволяет создать средства ввода-вывода речевой информации в машину — это многофункциональный речевой терминал (МРТ). С его помощью в общем случае, помимо распознавания речи, можно осуществить кодирование речевых сигналов с низкой информационной скоростью, верификацию дикто-

ра и взаимодействие с другими терминалами или с ЭВМ через сеть.

Частный вариант такого терминала разработан под руководством автора (отмечен серебряной медалью ВДНХ СССР). Он реализует ввод РК в диалоговый вычислительный комплекс ДВК-2М. «Речевая плата», встраиваемая в микро-ЭВМ, выполняет роль модуля первичного анализа речевых сигналов и выделяет информативные параметры РК. В зависимости от возможностей подсистемы распознавания речи может производиться простая обработка речевых сигналов, рассчитанная на фонетически сбалансированный словарь объемом до 100 РК, или сложная, предполагающая фильтрацию акустических помех и оценку спектральных параметров речевых сигналов.

Модуль распознавания реализован программными средствами. Для уменьшения затрат процессорного времени на классификацию РК применяется специальная процедура, называемая локальным динамическим программированием, которая производит сравнение введенного образа РК с эталоном не в целом, а последовательно по частям. Разумный их выбор более чем на порядок сокращает число операций, оставляя практически без изменений точность вынесения правильного решения. Дикторозависимый ввод РК со словарем в 150 слов, требующий однократного предварительного обучения, характеризуется надежностью не хуже 97%. После введения последовательности команд и данных микро-ЭВМ ДВК-2М может производить обработку запроса пользователя или выполнять другие программы.



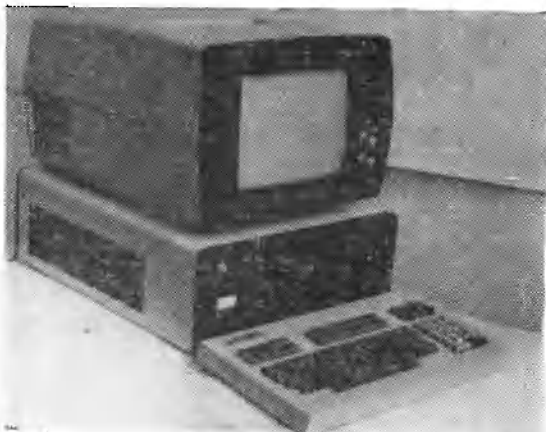
НОВАЯ ЭВМ

Диалоговый
вычислительный комплекс
«Электроника НЦ 8020 - 4»

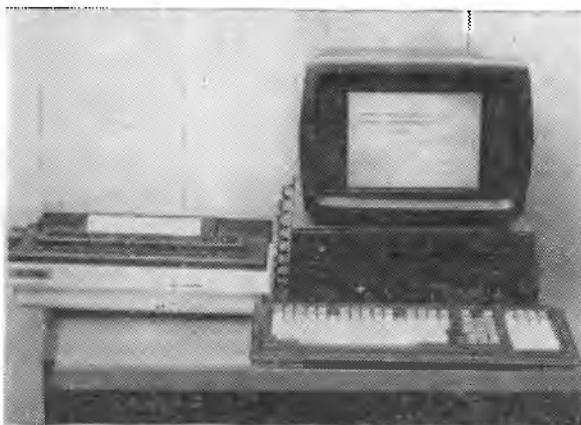
В. С. Кокорин, А. А. Попов

Диалоговый вычислительный комплекс «Электроника НЦ 8020/4» (ДВК-4) представляет собой одну из последних моделей отечественных профессиональных персональных ЭВМ (ПЭВМ). От ранее разработанных ПЭВМ, таких, как ДВК-1-, 2-, 3-й, модель ДВК-4 отличается главным образом значительно увеличенным объемом оперативной и внешней памяти и возможностью вывода цветной графической информации на экран дисплея. Новые свойства диалогового комплекса существенно расширили его оперативные возможности и улучшили восприятие человеком с дисплея графической информации.

По способу взаимодействия с пользователем ДВК-4 относится к классу диалоговых, по представлению ресурсов в распоряжение пользователя — к классу персональных, по ориентации на область применения — к классу профессиональных. Благодаря сочетанию в ДВК-4 больших функциональных возможностей наряду с хорошими эксплуатационными данными (настольное исполнение, малая потребляемая мощность, относитель-



ДВК-3



ДВК-4

ная простота эксплуатации) комплекс можно использовать для выполнения функций:

- секретаря-машинистки;
- редактора текстовой и графической информации;
- интеллектуального центра систем тестирования и регулировки сложнейшей электронной аппаратуры;

— банка знаний и элемента искусственного интеллекта информационно-справочных систем;

— высокоинтеллектуального средства отображения процессов, протекающих в графической форме в реальном масштабе времени (многоцветные графики, диаграммы и т. п.);

— служить элементом простейших САПР;

— выступать в роли персонального наставника школьника, студента, специалиста.

По архитектуре, системе команд и программному обеспечению ДВК-4 совместим с ДВК-1-, 2- и 3, с ЭВМ типа «Электроника-60», «Электроника-79», «Электроника 100/25» и мини-ЭВМ типа СМ 3 и СМ 4. Далее расскажем об основных технических характеристиках ДВК-4.

Быстродействие процессора при выполнении коротких операций — 700—900 тыс. операций в секунду. Информационная емкость ОЗУ — от 256 Кбайт до 1 Мбайт в зависимости от модификации ДВК. Адресное пространство — до 4 Мбайт. Информационная емкость внешней памяти (на гибких и жестких магнитных дисках) — от 6 до 12 Мбайт в зависимости от модификации. Количество одновременно воспроизводимых цветов на экране дисплея — любые 16 из палитры на 64 цвета. Формат графической информации (число строк и столбцов) — 240×400. Формат алфавитно-цифровой информации — 24 строки по 80 символов. Имеется возможность подключения печатающего устройства и графопостроителя. Габариты — 440×540×470.

марки ТИПЫ характеристики



продолжение выпуска

« ЭЛЕМЕНТАРНАЯ
БАЗА ЭВМ »

6

Интегральные схемы для средств вычислительной техники

Н. В. Воробьев

Современный этап развития полупроводниковой электроники характеризуется созданием широкой номенклатуры и массовым выпуском интегральных микросхем с быстро растущей степенью интеграции и непрерывно совершенствующихся дискретных полупроводниковых приборов. Интегральные схемы занимают доминирующее положение в вычислительной технике и системах обработки информации.

Темпы развития электронной промышленности все время ускоряются. Ни одна промышленность не развивалась такими быстрыми темпами, как электроника. Это прежде всего обусловлено достижениями технологии в полупроводниковой области. Технология часто является движущей силой в реализации новых разработок — от схем до архитектурных решений вычислительных систем.

Выдающимся технологическим достижением после изобретения сплавного транзистора был планарный процесс создания транзистора, основанный на диффузии. Использование диффузии сразу создало преимущества по сравнению с разработанными ранее процессами изготовления транзисторов. Например, из-за того что примеси проникали в полупроводник на глубину, которую можно было точно предсказать, зная температуру, давление паров и время воздействия, толщину базы можно было задавать в 10 раз точнее, чем в прежних методах. Другое существенное преимущество за-

ключалось в том, что, используя диффузию, можно было наладить групповое производство транзисторов.

Созрели условия для появления новой технологии, такой, которая бы позволила приспособиться к растущей сложности схем путем исключения соединений между их дискретными элементами. Поэтому изобретение интегральных схем стало насущным требованием. Из полупроводниковых материалов уже умели изготавливать резисторы, конденсаторы, транзисторы и диоды. Для изготовления резисторов использовали омические свойства тела полупроводника, а для создания конденсаторов — смещенные в обратном направлении *p-n*-переходы. Стало возможным изготовить все это для одной схемы в одном куске кремния. Такой прибор был разработан в начале 1959 г., и широкой общественности была представлена «твердая схема» в марте 1960 г. После этого разработка и изготовление интегральных схем начали развиваться быстрыми темпами. С конца 1961 г. интегральные схемы в больших количествах поступили в продажу. Логические схемы, известные в настоящее время как стандартные ТТЛ ИС, появились в результате работ, проведенных в нескольких местах, но основа их заложена в 1961 г. сотрудниками небольшой фирмы из Калвер-Сити (штат Колорадо, США), называвшейся Pacific Semiconductors Inc.

В 1962 г. на основе планарного технологического процесса был создан новый тип полевых транзисторов, работающих на принципах использования только основных носителей, — полевой транзистор МДП-типа с изолированным затвором. Внедрение планарного процесса для группового изготовления схем на основе транзисторов МДП-типа обеспечило развитие нового схемотехнологического направления — МДП ИС, которые, так же как и биполярные ИС, стали использоваться прежде всего для построения логических схем и запоминающих устройств.

С момента появления МДП ИС началось параллельное развитие двух главных схемотехнологических направлений в области цифровых ИС — биполярных технологий и МДП-технологий, носящих характер соперничества, которое продолжается и, вероятно, будет продолжаться в будущем, поскольку преимущества биполярных и МДП ИС до конца еще не реализованы, а недостатки постоянно устраняются. Внутри этих двух больших категорий в зависимости от схем, материалов и особенностей обработки существует много различных технологических процессов. Ниже будут кратко рассмотрены схемы на базе промышленно освоенных технологий.

БИПОЛЯРНЫЕ ТЕХНОЛОГИИ

Три основные биполярные технологии представлены транзисторно-транзисторной логикой (ТТЛ), эмиттерно-связанной логикой (ЭСЛ) и интегральной инжекционной логикой (И²Л).

ТТЛ-схемы представляют собой самые массовые логические схемы с момента выпуска первого семейства ТТЛ-схем в 1964 г. по настоящее время.

Принципиальная схема базового элемента широко распространенной серии К155 приведена на рис. 1, а, а функциональное обозначение — на рис. 1, б. Данный элемент выполняет функцию логического произведения с инверсией (И-НЕ). Схема содержит три основные части (на рисунке выделены пунктиром). Диоды VD1—

VD4 не выполняют функционального назначения, а служат для ограничения импульсов напряжения помехи отрицательной полярности, возникающих в линиях связи между элементами. Фазорасщепительный каскад управляет третьей основной частью схемы — выходным каскадом, способным отдать в нагрузку и принять от нагрузки большие токи. Резистор R5 защищает элемент от воздействия короткого замыкания выхода на общую шину. К сожалению, ТТЛ-элементы не защищены от кратковременного замыкания выхода на шину источника питания. Так, если на всех входах ТТЛ-элемента действуют уровни логической единицы, то в этом случае выходит из строя транзистор VT5.

В ТТЛ-схемах используются биполярные транзисторы в режиме насыщения, поэтому их динамические параметры существенно зависят от накопления и рассеивания избыточных зарядов в базах транзисторов. В настоящее время разработано большое количество модификаций ТТЛ-элементов. Модификация, как правило, преследует одну или несколько целей, среди которых можно выделить увеличение нагрузочной способности, уменьшение задержек распространения сигналов, увеличение уровней токов нагрузки, создание возможности объединения выходов элементов (монтажная логика), уменьшение мощности потребления, получение специальных элементов (например, И, индикации, контроля, И-ИЛИ-НЕ, преобразователей уровней и т. д.), увеличение порогового напряжения и логического перепада, получение парафазных выходов и т. п.

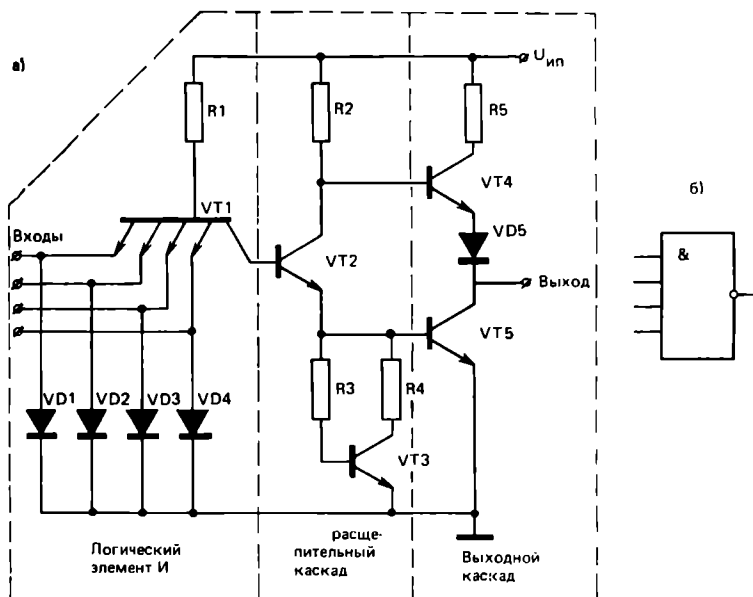


Рис. 1. Базовый ТТЛ-элемент:
а — принципиальная схема; *б* — функциональное обозначение.
 Полярность включения диодов VD1—VD4 на рисунке должна быть противоположной.

Наиболее интересна модификация ТТЛ-элемента с использованием транзисторов Шоттки и диодов Шоттки. Структура транзистора Шоттки и его обозначения приведены на рис. 2. Введение диодов Шоттки исключает накопление избыточных базовых зарядов, увеличивающих время выключения транзистора, и обеспечивает стабильность времени переключения в диапазоне рабочих температур. Такая модификация при той же мощности рассеяния элемента позволяет улучшить его динамические параметры в 5—10 раз либо при тех же динамических параметрах снизить в 5—10 раз мощность потребления.

Среди общих достоинств ТТЛ-элементов можно выделить: схемно-технологическую отработанность и как следствие высокий процент выхода годных схем, низкую стоимость, широкий функциональный набор логических элементов, высокую помехоустойчивость, большую нагрузочную способность при приемлемых динамических параметрах. Модификации ТТЛ-элементов стали схемотехнической основой многих БИС и микропроцессорных БИС.

К недостаткам ТТЛ-элементов можно отнести: наличие скачков тока по цепям питания и земли при переключении ТТЛ-элементов, что, в свою очередь, накладывает ограничение на максимальную длительность фронтов входных сигналов (для серии К155 эта величина составляет 150 нс); даже кратковременное замыкание выходов на шину питания может привести к выходу из строя ТТЛ-элемента; со схемами, показанными на рис. 1, *а*, нельзя реализовать монтажную логику (для этих целей разработаны модификации ТТЛ-элементов с открытым коллектором и с тремя состояниями выхода).

Популярность и распространенность ТТЛ-элементов в настоящее время таковы, что логические схемы любой степени интеграции стали, по существу, стандартными для цифровой техники.

ЭСЛ-схемы представляют другое широко распространенное семейство биполярной технологии. Определяющим свойством этих схем является устранение возможности насыщенного режима работы биполярных транзисторов, что позволяет получить задержки распространения сигналов в диапазоне долей — единиц наносекунд. В настоящее время ЭСЛ-схемы являются самыми быстродействующими. Опыт проектирования аппаратуры показывает, что примененные ЭСЛ-схем рационально для построения быстродействующих радиоэлектронных устройств, в частности ЭВМ высокого быстродействия, и менее эффективно при разработке радиоэлектронных устройств малого и среднего быстродействия.

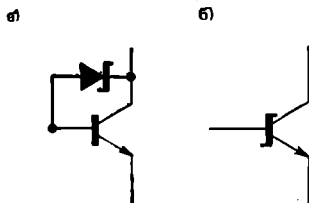


Рис. 2. Структура транзистора Шоттки *а* и его обозначение *б*

Принципиальная схема базового элемента широко распространенной серии 500 приведена на рис. 3, а, а функциональное обозначение — на рис. 3, б. Данный элемент выполняет функцию логического сложения и логического сложения с инверсией (ИЛИ, ИЛИ-НЕ). Наличие парафазных выходов существенно увеличивает функциональные возможности ЭСЛ-схем. Схема содержит четыре основные части: 1. Логическая часть схемы. Количество логических входов определяется количеством транзисторов в данной группе. Резисторы R1 и R2 служат для надежного заперения неиспользуемых входных транзисторов, они имеют сопротивление около 50 кОм, причем требования по точности этого номинала весьма слабы. 2. Переключатель тока. Величина тока задается резистором R4. Резисторы R3 и R5 являются определителями тока, т. е. указывают, в какой цепи течет ток. 3. Источник опорного напряжения ($-1,2$ В), которое подается на базу транзистора VT3. Величина опорного напряжения составляет полусумму уровней логического нуля ($-1,6$ В) и логической единицы ($-0,8$ В). Диоды VD1, VD2 обеспечивают температурную стабилизацию величины переключаемого тока. 4. Выходные эмиттерные повторители работают совместно с внешними нагрузочными резисторами. Основное назначение эмиттерных повторителей — в смещении выходных напряжений для обеспечения совместности входных и выходных уровней напряжений.

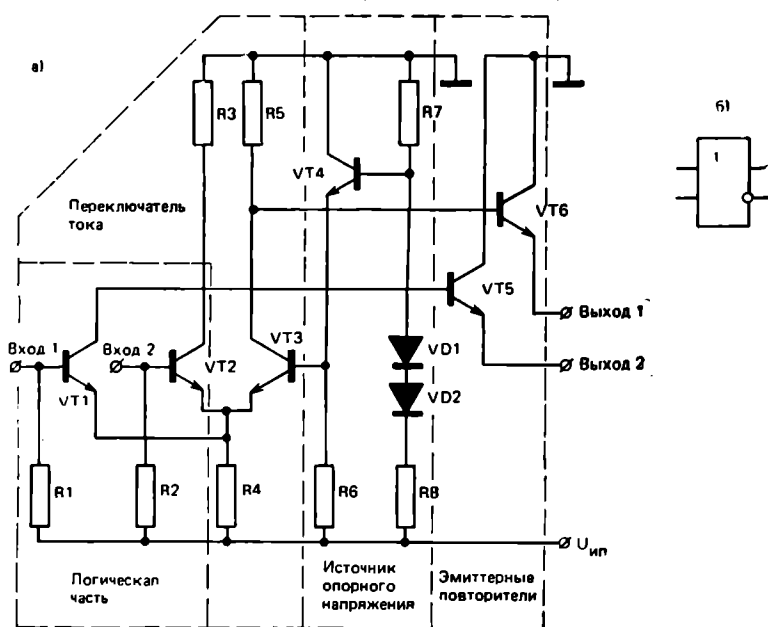


Рис. 3. Базовый ЭСЛ-элемент:

а — принципиальная схема; б — функциональное обозначение

Для достижения еще больших логических возможностей ЭСЛ-элементов, в частности для реализации «монтажного ИЛИ» по выходам, выводы эмиттеров транзисторов VT5 и VT6 делают изолированными, что позволяет объединять несколько эмиттеров различных логических элементов и подключать к ним только один внешний нагрузочный резистор. Нагрузочные резисторы выполняются как набор резисторов в тех же корпусах, что и логические элементы.

Высокое быстродействие ЭСЛ-элементов обеспечивается тремя основными факторами:

- активным режимом работы транзисторов в обоих логических состояниях элемента, благодаря чему устраняются этапы накопления и рассасывания избыточных зарядов;

- использованием низкого значения логического перепада, благодаря чему уменьшается время заряда и разряда собственных емкостных составляющих схемы;

- использованием на выходах элемента эмиттерных повторителей, обеспечивающих значительные токи для перезаряда емкостной составляющей нагрузки.

Микросхемы на основе ЭСЛ-схем имеют ряд достоинств, среди которых все достоинства рассмотренных выше ТТЛ-схем, кроме того:

- высокое быстродействие при средней потребляемой мощности или сверхвысокое быстродействие при большой потребляемой мощности;

- высокая стабильность динамических параметров при изменении рабочей температуры и напряжения питания;

- способность работать на низкоомные согласованные линии связи;

- возможность реализации монтажной логики;

- применение двух- и трехуровневого переключения тока для еще большего расширения логических возможностей ЭСЛ-элемента;

- возможность конструктивного снижения уровня помех в системах, выполненных на ЭСЛ-элементах.

К недостаткам ЭСЛ-элементов можно отнести:

- высокую потребляемую мощность при субнаносекундных задержках;

- более узкий температурный диапазон работы по сравнению с ТТЛ-элементами;

- относительно большую площадь, занимаемую вентилями на кристалле ($>10^4$ мкм²);

- проблематично применение ЭСЛ-элементов в сверхбольших интегральных схемах СБИС, хотя они с успехом используются в БИС.

И²Л-схемы являются третьим широко распространенным схемотехническим направлением биполярной технологии. Прежде всего необходимо отметить, что если ТТЛ- и ЭСЛ-элементы были разработаны для реализации интегральных схем, то И²Л-схемы — для создания БИС и СБИС на основе биполярных транзисторов. Схемы малой и средней степени интеграции на И²Л-элементах не выпускаются.

Принципы, на которых основано «конструктивное» и схемное отличие И²Л-элементов от других биполярных схем, заключаются в использовании совмещения электрически связанных однородных областей полупроводника в одном кристалле. В этих схемах тра-

диционный способ питания цепей базы и коллектора транзисторов через резисторы заменен непосредственным введением избыточных подвижных носителей заряда в базу переключаемых транзисторов.

Разработанные вначале как логические элементы для БИС и СБИС И²Л-элементы сейчас широко применяются для создания БИС памяти, микропроцессорных наборов, базовых кристаллов, разнообразных аналоговых устройств.

Базовый И²Л-элемент содержит *p-n-p*-транзистор, который генерирует постоянный ток (через внешний резистор), поступающий на базу многоколлекторного переключающего *n-p-n*-транзистора (рис. 4, а). Логический вход, «0» или «1», соответствует закороченной или разомкнутой цепи в базе транзистора VT2. Базовый вентиль представляет собой инвертор, поскольку если цепь базы закорочена, то транзистор находится в выключенном состоянии, обеспечивая на выходе разомкнутое состояние цепи, а если цепь базы будет разомкнута, ток включит транзистор, закорачивая тем самым его выходы на землю. С помощью трех вентилях-инверторов, как показано на рис. 4, б, можно организовать логический элемент ИЛИ.

Для изготовления подобных структур можно использовать технологический процесс с четырьмя фотошаблонами, эквивалентный по сложности базовой технологии схем МДП-типа. Это рекорд в биполярной технологии.

Стремление к увеличению быстродействия таких схем и их логической гибкости привело к разработке большого числа интегральных структур и разнообразных схемотехнических вариантов базового ключа. Одним из важнейших направлений развития инжекционных схем стало использование в них диодов Шоттки. При этом достигается увеличение быстродействия благодаря ограничению степени насыщения ключевого *n-p-n* транзистора (шунтирование перехода база — коллектор), использованию в качестве перехода база — коллектор диода Шоттки (транзистор с металлическим коллектором), уменьшению логических перепадов в схеме (последовательное включение диодов Шоттки в цепь «коллектор предыдущего *n-p-n* транзистора — база последующего»).

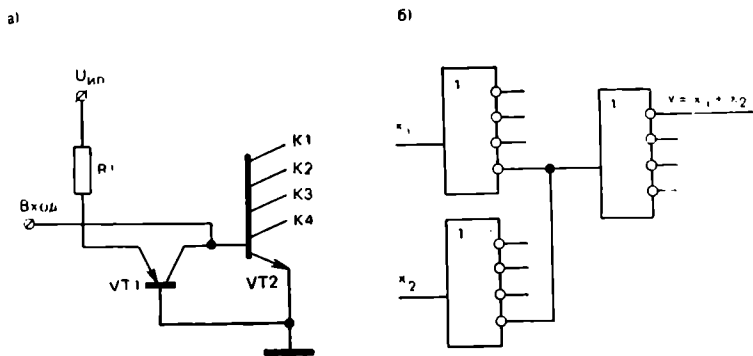


Рис. 4. Базовый И²Л-элемент:

а — принципиальная схема; б — реализация функции ИЛИ

С использованием диодов Шоттки возможно увеличение функциональной логической гибкости — объединение входов с помощью диодов Шоттки (логическая функция И).

К достоинствам И²Л-элементов можно отнести:

— возможность работы в широком диапазоне токов (10^{-9} — 10^{-2} А);

— возможность сохранения логического состояния И²Л-схем переводом их в режим микротоков, когда они не работают в предельном частотном режиме или вообще должны находиться в нерабочем состоянии;

— простоту разветвления сигнала на выходе за счет добавления коллекторов в ключевом транзисторе;

— чрезвычайно низкую мощность рассеяния и высокую плотность компоновки (малая площадь на кристалле $3 \cdot 10^2$ мкм²), что делает ее идеальной для СБИС;

— возможность создания инжекционно-полевых структур, для чего заменяют ключевой транзистор на полевой с вертикальным каналом (в этом направлении ведутся большие исследования).

Основной недостаток И²Л-элемента: площади коллекторов меньше площади эмиттера, т. е. транзистор работает в инверсном включении. Это в значительной степени определяет технологию изготовления И²Л-элементов и создает основные трудности при их реализации. К недостаткам интегральных схем инжекционно-полевой логики следует отнести прежде всего технологические трудности получения и воспроизведения геометрических размеров канала при массовом производстве схем. Несмотря на большие успехи в области разработки БИС и СБИС на И²Л-элементах, разработчикам и технологам предстоит решить еще много сложных задач.

Основные требования, предъявляемые к элементной базе БИС и СБИС, — высокая плотность элементов на кристалле, малая мощность рассеяния и технологичность структур при достаточно высоком выходе годных и низкой их стоимости.

Для повышения плотности компоновки совершенствуют полупроводниковую технологию. В частности, уменьшают линейные размеры компонентов, увеличивают площадь кристалла до такой степени, когда производство базовых кристаллов остается еще экономически целесообразным, а также увеличивают функциональные возможности базового элемента, применяют схемотехнику, обеспечивающую большую плотность компоновки в интегральном исполнении, упрощают схему, в частности, исключением буферных усилителей на входе или выходе логических элементов.

Снижение мощности рассеяния кристаллов матричных БИС и СБИС возможно при уменьшении напряжения питания, уменьшении логического перепада, совершенствовании добротности полупроводниковых компонентов, применении многоярусных схем на переключателях тока и др.

К наиболее распространенным разновидностям и модификациям элементов логических биполярных схем можно отнести: низкочастотную эмиттерно-связанную логику (НУ ЭСЛ), эмиттерно-функциональную логику (ЭФЛ), базосвязанную логику (БСЛ), интегральную инжекционную логику (И²Л), пороговую логику (НПЛ), многоуровневую переключательную логику (МУПЛ), эмиттерно-эмиттерную логику (ЭЭЛ, Э²Л), разнообразные модификации ТТЛ-элементов: маломощные ТТЛ (МТТЛ), ТТЛ-схемы с диодами Шоттки (ТТЛШ), маломощные ТТЛШ (МТТЛШ), ТТЛ-элементы с повышенным порогом переключения (Т³Л) и др.

Оценивая современное положение в области технологии биполярных схем, необходимо отметить, что она переживает подлинное возрождение в связи с переходом в область субмикронных размеров. Об этом убедительно свидетельствует возобновление работы в США конференции по биполярным схемам и технологиям их изготовления после примерно 20-летнего перерыва.

МОП-ТЕХНОЛОГИИ (*p*-МОП, *n*-МОП и КМОП)

Преимущества МОП ИС были известны с самого начала: процесс их изготовления значительно проще, чем для биполярных ИС, так как число необходимых технологических операций уменьшалось более чем в 2 раза; они потребляли гораздо меньшую мощность и, следовательно, допускали более высокий уровень интеграции, чем биполярные приборы, и наконец, их изготовление обходилось дешевле. Эти схемы, однако, не были лишены и недостатков: их производству мешали дефекты в окисле; они были чрезвычайно чувствительны к статическим зарядам — небольшое перенапряжение могло пробить тонкий окисел и мгновенно разрушить МОП-транзистор. Кроме того, рабочие напряжения МОП ИС были значительно выше, чем рабочие напряжения серий логических биполярных интегральных схем, выпускавшихся в то время; МОП ИС обладали значительно меньшим быстродействием, чем биполярные схемы. К настоящему времени большинство из этих недостатков устранено.

Канал МОП-транзистора представляет собой легированный кремний *p*- или *n*-типа, что определяет тем самым *p*-МОП и *n*-МОП-технологии. Несмотря на простоту изготовления, *p*-МОП-технология была повсеместно вытеснена *n*-МОП-технологией вследствие более высокого быстродействия. Изготовители МОП-приборов обнаружили, что можно существенно повысить быстродействие, если заменить *p*-канальные структуры структурами с каналом *n*-типа. Этот переход был сложным, так как прецизионное легирование в случае *n*-МОП-структур осуществить гораздо сложнее, чем при изготовлении *p*-канальных структур. Однако повышение быстродействия было очевидным: носители в канале *n*-типа (электроны) движутся быстрее, чем носители в *p*-канале (дырки).

В третьей основной МОП-технологии используются как *n*-, так и *p*-МОП-транзисторы. Она называется комплементарной МОП-технологией (КМОП). Комплементарные структуры оказались удивительными. От скромного начала в качестве медленно действующих структур с металлическими затворами и малой плотностью упаковки элементов, использовавшихся в ИС наручных часов, они прошли через этапы уменьшения размеров элементов, перехода на изоляющую окисел, замены металлических затворов на кремниевые. Это позволило увеличить быстродействие и плотность упаковки элементов, так что КМОП-структуры смогли соперничать с *n*-МОП-струк-

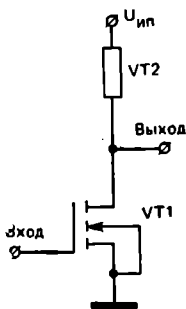


Рис. 5. Схема инвертора на *n*-МОП-транзисторах

турами, рассеивая при этом в режиме покоя порядка микроватт.

Основное преимущество всех МОП-технологий — относительная простота производственных процессов и высокая плотность компоновки. Поэтому МОП-технология применялись почти во всех без исключения приложениях БИС, обеспечивая тем самым выпуск чрезвычайно недорогих изделий (например, микропроцессоров и больших кристаллов памяти).

Базовый *n*-МОП-вентиль представляет собой инвертор (*p*-МОП-вентиль имеет идентичную структуру при противоположной полярности источника питания). Этот вентиль состоит из двух транзисторов (рис. 5) — ключевого VT1 и нагрузочного VT2. Нагрузочный транзистор всегда включен и ведет себя как постоянный резистор с большим сопротивлением, поэтому на рис. 5 он обобщенно представлен в виде резистора. В открытом состоянии ключевой транзистор имеет сопротивление в 20 раз меньше, чем нагрузочный. Когда транзистор VT1 включен ($U_{вх} = U_{п.п.}$), напряжение $U_{вых}$ близко к потенциалу земли ($U_{вых} \approx 0$). Когда транзистор VT1 выключен ($U_{вх} \approx 0$), $U_{вых}$ приблизительно равно потенциалу источника питания $U_{п.п.}$.

Если параллельно соединить два ключевых транзистора, то получим логический элемент ИЛИ-НЕ (рис. 6, а), если же ключевые транзисторы соединить последовательно, то получим логический элемент И-НЕ (рис. 6, б). Поскольку в схеме И-НЕ уровень логического нуля определяется суммарным сопротивлением последовательно включенных ключевых транзисторов, необходимо обеспечить меньшее их сопротивление, чем в схемах инвертора или ИЛИ-НЕ. Это достигается увеличением ширины канала, т. е. увеличением геометрических размеров структуры логического вентиля. Следовательно, схемы ИЛИ-НЕ более предпочтительны. Поскольку транзистор VT2 не переключается, а работает только как пассивный резистор, в *p*- и *n*-канальных схемах имеется асимметрия в длительности спада и подъема на выходе. Считают, что спад — это активный процесс, а подъем — пассивный. Различие в длительности может быть до порядка.

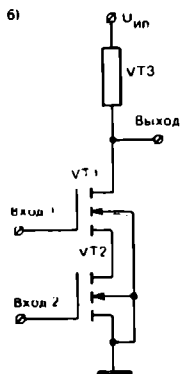
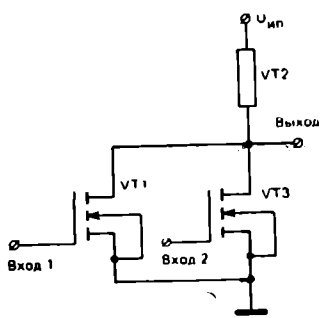


Рис. 6. Схема логических элементов на *n*-МОП-транзисторах:
а — ИЛИ-НЕ; б — И-НЕ

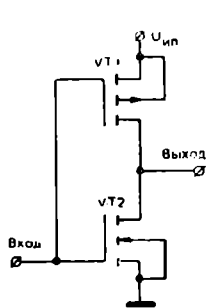


Рис. 7. Схема инвертора на КМОП-структуре

Поскольку в p -канальных МОП-схемах полярности тока и напряжения противоположны полярности в n -канальных МОП-схемах, то они могут применяться как комплементарные пары для формирования базового инверторного вентиля (рис. 7). Такое свойство позволяет формировать почти симметричный инвертор (существуют некоторые отличия в подвижности носителей), в котором включается только один транзистор. С помощью двух комплементарных пар транзисторов могут быть получены логические элементы И-НЕ и ИЛИ-НЕ. В элементе И-НЕ, изображенном на рис. 8, транзисторы p -типа включены параллельно, транзисторы n -типа — последовательно. Элемент ИЛИ-НЕ создается из последовательно включенных транзисторов p -типа и параллельно включенных транзисторов n -типа.

Замечательное свойство КМОП-схем выражено в том, что потребление тока возможно только во время переключения, когда происходит заряд и разряд емкостной составляющей нагрузки логических элементов. В этом довольно просто удостовериться, если вспомнить, что в комплементарной паре один транзистор всегда выключен. Мощность рассеяния растет с ростом частоты тактирования. Это дает огромное преимущество при изготовлении схем памяти, причем существующие n -канальные МОП-схемы по мощности рассеяния приближаются к тепловому барьеру (примерно 1—2 Вт на корпус).

Как бы ни были велики нынешние достижения электронной техники и технологии, в настоящий момент (80-е годы) они представляют собой всего лишь основу для дальнейшего огромного роста быстродействия, плотности упаковки и сложности логических, запоминающих, микропроцессорных и других схем.

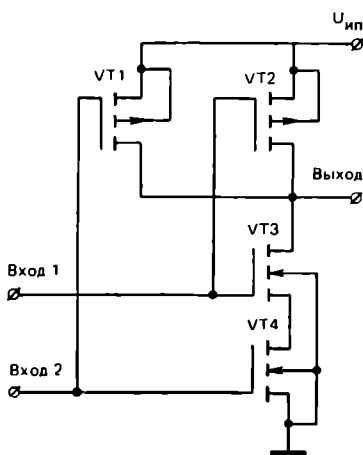


Рис. 8. Схема логического элемента И-НЕ на КМОП-структуре

НОМЕНКЛАТУРА ИС

Номенклатуру современных интегральных схем удобно рассматривать, выделив следующие их основные классы: цифровые, аналоговые, пленочные и оптоэлектронные микросхемы.

Цифровые микросхемы включают в себя логические комбинационные и последовательностные схемы, запоминающие ИС, микросхемы микропроцессорных комплектов, микросхемы интерфейсов, базовые матричные кристаллы. Все перечисленные цифровые микросхемы являются наиболее массовыми и употребительными. Число наименований схем различного назначения и различной степени интеграции даже в составе одной серии элементов приближается к сотне. Следует отметить всевозрастающий уровень выпуска микросхем памяти. Если в 1986 г. они составили приблизительно 50% объема всех ИС, то к 2000 г. их доля увеличится до более 75%.

Аналоговые микросхемы включают в себя операционные усилители, коммутаторы и ключи, компараторы напряжений, цифроаналоговые и аналого-цифровые преобразователи (ЦАП и АЦП), вторичные источники электропитания, микросхемы для радиосвязи, звуковоспроизведения, телевидения и видеозаписи, микросхемы для широкого технического применения или многофункциональные и т. п.

Аналоговые микросхемы широко используются в устройствах первичной обработки информации, источниках питания, периферийных устройствах ЦВМ и т. п.

Пленочные микросхемы включают в себя различно организованные матрицы прецизионных резисторов, которые могут применяться в качестве последовательных делителей напряжения, масштабирующих устройств, согласователей уровней, двоично-кодированных матриц для ЦАП и АЦП, просто наборов резисторов и т. д.

Оптоэлектронные микросхемы содержат ключи и коммутаторы цифровых и аналоговых сигналов, выполненные в виде двух разновидностей пар: излучатель — приемник; излучатель — приемник с усилителем мощности.

Оптоэлектронные микросхемы являются идеальным средством гальванической развязки для широкого круга применения.

В заключение отметим, что сегодня (80-е годы) используется лишь 5% тех реальных возможностей, которые микроэлектроника в действительности способна предложить промышленности. Предполагается, что через 20 лет будут реализованы все эти возможности на 100%.

ЛИТЕРАТУРА

Преснухин Л. Н., Воробьев Н. В., Шишкевич А. А. Расчет элементов цифровых устройств: Учеб. пособие /Под ред. Л. Н. Преснухина. — М.: Высшая школа, 1982.

Баринов В. В., Орликовский А. А. Сверхбыстродействующие элементы кремниевых цифровых БИС: /Учеб. пособие по курсу «Микросхемотехника». — М.: РИО, МИЭТ, 1981.

Электроника: прошлое, настоящее, будущее. — М.: Мир, 1980 (специальный выпуск журнала «Электроника», 1980. — № 9).

Бернард Конрад Коул. Активное возрождение биполярной технологии //Электроника. — 1986. — № 18. — С. 48—51.

ТЕМА
следующего выпуска

9

СИСТЕМЫ АВТОМАТИЗАЦИИ

Под редакцией академика АН УССР
В. И. Скурихина

Научно-популярное издание

РЕЧЬ И ЭВМ

Библиотечка «Вычислительная техника и ее применение»

Выпуск 8

Главный отраслевой редактор *Л. А. Ерлыкин*

Редактор *Б. М. Васильев*

Мл. редактор *Л. В. Дегтярева*

Оформление художника *И. А. Емельяновой*

Худож. редактор *П. Л. Храмцов*

Техн. редактор *Т. В. Луговская*

Корректор *В. В. Каночкина*

Сдано в набор 28.04.87. Подписано к печати 14.10.87. Т-13853. Формат бумаги 84×108¹/₃₂. Бумага тип. № 1. Гарнитура литературная. Печать высокая. Усл. печ. л. 3,36. Усл. кр.-отг. 3,78. Уч.-изд. л. 3,58. Тираж 25 000 экз. Заказ 1635. Цена 20 коп.

Издательство «Знание». 101835, ГСП, Москва, Центр, проезд Серова, д. 4. Индекс заказа 874811.

170000, г. Калинин, Студенческий пер., 28. Областная типография.

20 коп.

БИБЛИОТЕЧКА

**ВЫЧИСЛИТЕЛЬНАЯ
ТЕХНИКА
И ЕЕ ПРИМЕНЕНИЕ**



ВЫПУСК

8

**Авторы
выпуска**

СИФОРОВ Владимир Иванович — член-корреспондент АН СССР, директор Института проблем передачи информации, председатель научно-методического совета отделения радиоэлектроники, оптики и связи общества «Знание».

ВАЙСМАН Анатолий Савельевич — журналист, сотрудник журнала «Изобретатель и рационализатор», автор около 150 научно-популярных статей, предпочтение в которых отдает тематике автоматического распознавания образов.

ВОРОБЬЕВ Николай Васильевич — доцент, ученый секретарь специальных советов. Научный интерес — применение вычислительной техники для автоматической обработки результатов измерений и контроля состояния технологического оборудования.

КОКОРИН Владимир Сергеевич, ПОПОВ Александр Анатольевич — кандидаты технических наук, специализируются в области разработки средств вычислительной техники.

СВИРИДЕНКО Владимир Александрович — доктор технических наук, профессор, автор ряда статей, монографий и изобретений по обработке и передаче данных и речевой информации.

СОРОКИН Виктор Николаевич — доктор физико-математических наук, автор более 60 публикаций, 6 авторских свидетельств.

**РЕЧЬ И
ЭВМ**