

Компьютеры — синтезаторы речи

**ELECTRONICALLY SPEAKING:
COMPUTER SPEECH
GENERATION**

by
John P. Cater

Indianapolis, Howard W. Sams & Co., Inc., 1983

В мире науки и техники

Дж. Кейтер

Компьютеры— синтезаторы речи

Перевод с английского

Э. Я. Пастрона

под редакцией

канд. техн. наук В. А. Усика



МОСКВА «МИР» 1985

ББК 32.973.2

К 33

УДК 681.3

Кейтер Дж.

К33 Компьютеры — синтезаторы речи: Пер. с англ./
/Под ред. В. А. Усика.— М.: Мир, 1985.—237с.,
ил. (В мире науки и техники)

Книга американского инженера Джона Кейтера рассказывает о теоретических исследованиях и практических разработках в одной из самых передовых областей современной электронной техники — технике синтеза речи. Автор приводит также конкретные схемы электронных блоков, используемых в реальных синтезаторах речи.

Книга адресована широкому кругу читателей, интересующихся достижениями современной техники; особенно полезна она будет радиолюбителям.

К $\frac{2405000000-514}{041(01)-85}$ 181 — 86, ч. 1

ББК 32.973.2
6Ф7.3

*Редакция научно-популярной
и научно-фантастической литературы*

Copyright © 1983 by Howard W. Sams and Co., Inc.

© перевод на русский язык, «Мир», 1985

ПРЕДИСЛОВИЕ РЕДАКТОРА ПЕРЕВОДА

Природа наделила человека уникальной способностью общения с себе подобными — даром речи. Тысячелетиями люди мечтали поделиться этим чудесным даром с окружающими их объектами живого и неживого мира. Однако долгое время результаты таких попыток оказывались ничтожными: за исключением попугаев, механически заучивающих десяток-другой слов и выражений, животные и предметы разговаривали с людьми только в сказках. Лишь в конце XIX в., постигнув физиологический механизм действия голосового тракта, люди начали конструировать «говорящие» машины, но соответствующие технические возможности были тогда предельно ограничены. И даже в наш век электроники первые устройства, генерирующие искусственную речь, — синтезаторы речи — были настолько громоздки и дороги, что не могли выйти из стен лабораторий в сферу широкого применения.

Сейчас положение радикально изменилось. Электронная технология, вооруженная математическими моделями голосового аппарата, поднялась на уровень, который позволяет производить в массовых количествах весьма дешевые и компактные синтезаторы речи. И, возможно, недалек тот день, когда человек окажется в окружении предметов, наделенных способностью говорить. Уже сейчас появились разнообразные говорящие устройства: компьютеры и калькуляторы, бытовые и измерительные приборы, игрушки и электронные игры, читающие машины.

К сожалению, информация, касающаяся этой новой бурно развивающейся области техники, весьма скудна и в основном рассеяна в специальной литературе, малодоступной широкому кругу читателей. В связи с этим перевод на русский язык книги американского инженера Джона Кейтера «Компьютеры — синтезаторы речи»

представляется нам весьма своевременным. Эта книга, безусловно, будет полезна не только специалистам, занимающимся разработкой или использованием синтезаторов речи, но и всем, кто с интересом следит за достижениями современной электронной техники. Книга, несомненно, привлечет внимание и подготовленных радиолюбителей, которые, пользуясь приведенными в ней идеями и конкретными схемами, смогут на доступной им элементной базе приступить к захватывающим экспериментам по синтезу речи.

В нашей стране широкие возможности для использования синтезаторов речи открывает осуществляемая ныне программа по внедрению микропроцессорной техники и персональных компьютеров. Советские ученые и инженеры давно и плодотворно занимаются проблемами синтеза речи. Так, в минском отделении Центрального научно-исследовательского института связи под руководством Б. М. Лобанова созданы и внедрены в промышленное производство синтезаторы речи Фонемофон-К и Фонемофон-Т. Как следует из названия, эти синтезаторы речи работают по фонемному принципу (с ним читатель подробно ознакомится в книге Дж. Кейтера) и потому обладают неограниченным словарем. В качестве периферийного устройства синтезатор Фонемофон-К можно подключать к любой ЭВМ. Единственное, что требуется от оператора, — это проставить правильные ударения в словах текста, отображаемого на экран дисплея. Фонемофон-Т имеет аналогичную конструкцию, но дополнительно оснащен схемой подключения к обычной телефонной сети. Благодаря этому его можно широко использовать в телефонной службе и коммунальном хозяйстве. Так, он сам звонит телефонным абонентам и сообщает размер оплаты за междугородные переговоры или напоминает забывчивым жильцам о необходимости своевременно внести квартплату. В Институте кибернетики Академии наук Эстонской ССР разработан формантный синтезатор речи, который также выпускается серийно. В настоящее время упомянутые синтезаторы переводятся на более совершенную элементную базу с использованием микропроцессоров, что позволит значительно уменьшить их размеры и стоимость.

В заключение следует сказать несколько слов о работе над переводом книги Дж. Кейтера. Используемые

ВСТУПЛЕНИЕ

По мере совершенствования компьютеров, которые человек моделирует по своему образу и подобию, они «интеллектуально» все больше приближаются к человеку. В лабораториях разных стран с головокружительной быстротой проводятся исследования в таких областях, как искусственный интеллект, аудиовизуальный вывод информации, и моделирование других способов восприятия информации человеком. В этой книге мы расскажем об искусственной, или компьютерной, речи и о том, как она используется в современных больших и малых компьютерах.

Цель книги — познакомить читателя с методами и принципами генерации искусственной речи, не упуская при этом из виду более практические, но тем не менее важные вопросы применения синтезаторов речи в современном обществе, которое с интересом и пониманием относится к развитию вычислительной техники. В большинстве опубликованных работ по компьютерной речи рассматривается в основном техническая база компьютерных синтезаторов речи и ничего не говорится о «подводных камнях» и возможностях их обхода при конкретных применениях. Часто нужные рекомендации даются лишь после того, как система уже спроектирована, разработана и поступила в практическое пользование. Во многих случаях эти рекомендации становятся доступными для разработчиков слишком поздно, чтобы, учтя их, внести необходимые изменения в изделия, — в результате вся идея оказывается обреченной на провал.

Сначала мы вкратце познакомимся с основами человеческой (естественной) речи, рассматривая их с точки зрения физиологии и биомеханики, а также дадим нечто вроде краткого курса лингвистики, которая, собственно, и является основным инструментом нашего исследова-

автором термины из области радиоэлектроники, физиологии и лингвистики в основном совпадают с отечественными. Что же касается техники синтеза речи, то здесь некоторые новые понятия, возможно, еще не нашли четкого определения в нашей терминологии, но их разъяснение в Словаре терминов облегчает понимание материала. Кроме того, необходимо отметить следующее: автор касается американских разработок в области синтеза речи, и естественно, что в книге имеется ряд фонетических примеров на английском языке, которые целесообразно было сохранить, поскольку с ними неразрывно связан весь текстовый и иллюстративный материал. Список литературы, приведенный автором, включает исключительно англоязычные источники, большинство которых малодоступно советскому читателю. Мы дополнили этот список отечественными и переведенными на русский язык зарубежными публикациями (см. разд. «Дополнительная литература»), которые помогут заинтересованному читателю глубже разобраться в затронутых в книге вопросах.

В. Усик

дования. После этого мы займемся методами генерации искусственной речи и расскажем о выпускаемых промышленностью изделиях, которые могут служить для построения синтезаторов речи на базе малых компьютеров.

Первая глава нашей книги служит введением в область компьютерного генерирования речи; она рассчитана в первую очередь на новичка, желающего получить хотя бы общее представление о положении дел в данной области. Знакомство с основными принципами компьютерного синтеза речи поможет читателю лучше понять современное состояние знаний в этой области и технику, используемую в новых приложениях синтезированной речи.

Во второй главе рассмотрен механизм речи человека. Эта глава касается преимущественно биологических вопросов, но понимание основных принципов механизма речи необходимо, чтобы в достаточной мере разобраться в содержании последующих глав. Поскольку синтезатор речи функционально должен моделировать человеческую речь, беглое знакомство с биомеханикой естественной речи позволит читателю осознать всю сложность искусственного воспроизведения тех же самых звуков из последовательности единиц и нулей.

Третья глава книги в каком-то смысле напоминает инструкцию по эксплуатации органов речи человека или компьютера. Описанный в предыдущей главе механизм речи останется бесполезным, если не знать ряда правил, используемых при разговоре. Изложение здесь ведется на примерах из английского языка. К сожалению, область лингвистики (фонетика), описываемая в этой главе, сродни тому предмету, который вызывал у нас раздражение в школе. Однако она имеет непосредственное отношение к компьютеру—синтезатору речи и поэтому заслуживает внимания! Это один из наиболее важных аспектов компьютерной речи. Чтобы правильно передать наши знания речи компьютеру, необходимо усвоить основные положения, которые изложены здесь. Если же вы пропустите эту главу, то вряд ли созданный вами компьютер заговорит как диктор в программе новостей. Синтезатор речи, разработанный без учета основ лингвистики, возможно, будет говорить, но как Буратино или утенок Дональд. Иначе говоря, эта глава знакомит с «программными средства-

ми», необходимыми для получения речи от говорящего механизма. Без усвоения данного материала вы, по существу, обречены на попытки программировать компьютер без программного обеспечения.

В четвертой главе мы, отклонившись от технической направленности нашей книги, постараемся дать общее представление о некоторых довольно важных этических правилах, которые следует учитывать при реализации говорящего компьютера. Если вы взялись за чтение этой книги с намерением создать собственную компьютерную систему, способную разговаривать с людьми, то вам, естественно, надо помнить о необходимости соблюдения определенных правил речевого этикета, которыми, к сожалению, часто пренебрегают как люди, так и компьютеры. При желании вы можете отложить чтение этой главы до той поры, когда ваш синтезатор речи перестанет быть новшеством. Но без нее не обойтись, если вы захотите заставить компьютер замолчать. Эта глава не бесполезна: в ней изложены кое-какие соображения здравого смысла, которые следует учитывать при конструировании говорящих компьютеров — от говорящих устройств, используемых для охраны жилых домов, до «болтливых» игровых автоматов.

В пятой главе дается краткий исторический обзор исследований в области синтезирования речи. Эта глава весьма важна, ибо, как говорит пословица, не знающий прошлого обречен повторять его ошибки. Кроме того, небезынтересно узнать всю историю этих исследований и насколько усовершенствовались за 100 лет методы обучения машин разговору.

Глава шестая вводит нас в круг технических проблем современных систем синтеза речи. Рассмотрены и детально описаны все имеющиеся методы синтеза речи. Сначала описывается способ синтеза речи путем преобразования речевого сигнала в цифровой код. Далее рассматриваются методы аналогового формантного синтеза, линейное предиктивное кодирование (ЛПК) речи для систем с ограниченным и неограниченным словарями и методы частичной автокорреляции. В заключение читатель познакомится с перспективными методами, которые пока еще находятся в стадии исследований, но, вполне вероятно, найдут применение в будущем, когда это позволит развитие техники.

Седьмая глава представляет собой обзор выпускаемых в продажу синтезаторов речи, которые могут применяться в качестве периферийных устройств малых компьютеров. Изделия классифицированы по используемым в них методам синтеза речи. Краткое описание принципа действия каждого изделия помогает выбрать то, которое лучше удовлетворяет конкретным условиям применений.

Восьмая глава знакомит с широким диапазоном применений синтезаторов речи, начиная с устройств, которые уже поступили в серийное производство, и кончая теми, которые, быть может, впервые будут реализованы в вашем будущем персональном компьютере. Приводится также некоторая информация о возможных перспективных применениях синтезаторов речи, основой для которых послужат нынешние исследования. И независимо от того, на каком из синтезаторов речи вы остановите выбор, он непременно найдет подходящее применение.

В заключение приведен список литературы, который поможет читателю лучше разобраться в интересующих его вопросах. Следует, однако, иметь в виду, что многие из указанных в этом списке работ довольно трудны и рассчитаны на серьезное изучение.

Наконец, приводится словарь терминов, где читатель найдет разъяснение незнакомых ему терминов. Этот словарь может оказаться полезным и при чтении другой технической литературы, относящейся к области компьютерного синтеза речи.

Джон Кейтер

Моей дорогой жене Джей, которая терпеливо ждала, пока я писал эту книгу

Глава I

ОСНОВЫ КОМПЬЮТЕРНОГО СИНТЕЗА РЕЧИ

Написать введение к книге по электронному синтезу речи — задача, в какой-то мере подобная той, с которой пришлось столкнуться Христофору Колумбу, когда он приступил к описанию своего путешествия в Новый Свет в 1492 г. Углубившись в чтение этой книги, вы совершите фантастическое путешествие в глубь нашего организма и мозга — путешествие по «земле», где по соседству уживаются научный факт и фантастика. Всего лишь несколько быстротечных лет назад на страницах научно-фантастических произведений возникли образы мощных компьютерных систем и роботов с речевым выходом. Теперь — спустя время меньшее, чем жизнь одного поколения, — наука и техника включили этот мир в свои владения. Методы компьютерного синтеза речи не только вышли за стены лабораторий, но и стали использоваться в сочетании с персональными компьютерами.

Нашу книгу, рассказывающую о компьютерном синтезе речи, можно рассматривать как своего рода научное путешествие внутрь человека, ибо впервые наука предприняла попытку скопировать физиологический процесс. Чтобы заставить электронные машины говорить, ученые, занимавшиеся проблемами синтеза речи, не могли просто ограничиться решением уравнений с помощью логарифмических линеек и калькуляторов — им пришлось разбираться в биомеханике речи человека, а затем находить электронные эквиваленты для моделирования голосового тракта. Можете ли вы вообразить себе трудности, с которыми пришлось бы столкнуться первым специалистам по компьютерам, если бы они попытались моделировать цифровой компьютер из серого вещества мозга? Если бы такая попытка была осуществлена и увенчалась успехом, то знакомые нам сегодня компьютеры были бы по своей сути химическими, а не электронными системами.

Разумеется, вполне возможно попытаться моделировать электронные конструкции на основе живых аналогов, однако следует учитывать, что сам механизм речи человека во много раз проще, чем механизмы процессов, протекающих в мозге. В порядке некоторой предварительной подготовки мы познакомимся сначала с некоторыми физиологическими сторонами процесса речи и связанными с ними мыслительными процессами — вместе они и образуют то, что мы называем речевым общением. Затем, остановившись кратко на истории исследований в области говорящих машин, мы покажем сегодняшнее состояние и перспективы техники синтеза речи. Одна глава нашей книги адресована тем, кто, не желая вникать в механизмы компьютерного синтеза речи, заинтересован лишь в обзоре современной «говорящей продукции».

Чтобы участвовать в нашем путешествии, не требуется особой научной подготовки. Хотя читатель, возможно, и знает кое-что о компьютерах, эта книга одновременно предназначена и для новичка, совершенно незнакомого с синтезом речи, и для профессионала, работающего в этой области. Чтобы облегчить понимание, автор при изложении основных понятий пользуется простыми терминами, а затем постепенно переходит к более сложным вопросам. Завершив свою «одиссею» в область компьютерного синтеза речи, читатель далее познакомится с новыми перспективными разработками.

В чем суть вопроса?

Много тысячелетий назад, на заре человеческого общества, когда человек только начинал произносить первые звуки, он воспринимал речевое общение как некую часть жизни. Сейчас мы каждый день используем речь для связи или передачи сообщений. К счастью, сам процесс генерации речи требует от нас минимальных умственных усилий. В противном случае нам, возможно, пришлось бы при разговоре постоянно делать паузы. И поскольку устное общение — одно из наиболее эффективных средств, имеющихся у нас для выражения своих мыслей и чувств, мы пользуемся им весьма широко. В самом деле, не будет преувеличением сказать, что каждый из нас произносит ежедневно в среднем 40 тыс.

слов. Можете ли вы себе представить, как бы это выглядело в письменной форме? Запись разговоров только одного человека на протяжении всей его жизни могла бы составить целые библиотеки. Итак, совершенно ясно, что устная речь представляет собой одну из наиболее важных и полезных для нас форм общения. Напомним, однако, что в большинстве случаев устная связь не оставляет после себя никаких «следов», кроме эха и того, что откладывается из услышанного в нашей памяти. Напротив, письменное сообщение считается долговременным отпечатком мыслительных процессов.

С появлением в XX в. электронных вычислительных машин человек стал сознавать, что он создал устройство, способное разумно реагировать на его команды. В соответствии с имевшимися техническими средствами для вывода информации из компьютера сначала использовали мигающие лампочки; позднее перешли на печатные сообщения. Хотя максимальная скорость печати у первых компьютеров достигала 60 слов в минуту, совершенствование продолжалось. Сегодня в нашем распоряжении лазерные печатающие устройства (принтеры), производительность которых достигает 600 страниц в минуту. Вообразите себе компьютер, выдающий информацию со скоростью 10 страниц в секунду! Всего за сутки работы такой компьютер может дать почти миллион страниц информации.

Параллельно с разработкой компьютеров ученые трудились над созданием электронных моделей голосового тракта, способных искусственно генерировать голос человека. Занимаясь этими проблемами, специалисты сначала не ориентировались на компьютеры, а предполагали использование устройств, управляемых человеком. Но по мере становления названных параллельно развивающихся технологий они взаимно обогащали друг друга. Компьютер начал говорить. Первые несколько слов, с трудом произнесенных компьютером, озаменовали еще одно проявление эгоцентризма человека. Действительно, поначалу компьютер моделировали по образу и подобию человека. С тех пор, однако, мы поняли, что способность компьютера к непосредственному взаимодействию с человеком путем речевого вывода и ввода информации весьма удобна и естественна для нас. В последние годы эта область привлекла к себе внимание электронной промышленности,

и началось ее довольно быстрое развитие, хотя пока она еще пребывает в младенческом возрасте.

Описываемые в этой книге основные принципы, методы и технические средства, используемые для компьютерного синтеза речи, характеризуют уровень, достигнутый на 80-е годы. Однако, поскольку способности человека понимать и преодолевать трудности, по-видимому, удваиваются каждые пять лет, весьма вероятно, что с высоты 90-х годов наши достижения покажутся устаревшими. Но ведь и ребенок, обучаясь ходить, сначала ползает и лишь потом обретает способность бегать. Сейчас по своим возможностям в области генерации искусственной речи мы пребываем в состоянии где-то между ползанием и ходьбой. Существует три основных технологически различных подхода к проблеме синтеза речи:

1. Метод кодирования-восстановления формы сигналов.

2. Аналоговый метод синтеза формантных частот.

3. Цифровое моделирование голосового тракта.

Первый из этих методов представляет собой один из самых основных и элементарных подходов к созданию говорящего компьютера. По существу, компьютер в данном случае просто служит устройством для записи речи. Но для хранения информации в нем используется не магнитофонная лента или грампластинка, а цифровое запоминающее устройство. Фразы и слова записываются отдельно и воспроизводятся в нужный момент по командам, поступающим от соответствующей программы. Возможность «сказать» слово, которое не было заранее введено в память, здесь отсутствует. Для реализации «говорящего» выхода в подобных системах требуется очень мало дополнительных аппаратных средств. У этого метода есть, к сожалению, очень серьезный недостаток — для хранения речевых сигналов в их непосредственной форме нужна память значительного объема. Хотя полученная таким образом речь по качеству и может приближаться к весьма хорошей звукозаписи, требования к объему памяти часто оказываются неприемлемыми.

Тем не менее, когда необходимый словарь не слишком велик, т. е. не превышает 10—15 слов или фраз, данный метод вполне пригоден и весьма недорог. Действие этого метода иллюстрирует рис. 1.1. Его приме-

нение вполне естественно там, где нужен ограниченный и неизменный словарь, например при создании говорящего приборного щитка для автомобиля. Поскольку необходимые речевые сообщения здесь полностью определяются особенностями работы автомобиля, для их хранения достаточно памяти ограниченного объема. Таким образом, автомобильный приборный щиток дополняется довольно компактным и недорогим устройством речевого вывода.

Второй метод синтеза речи — метод синтеза формантных частот. Синтезатор, построенный в соответ-

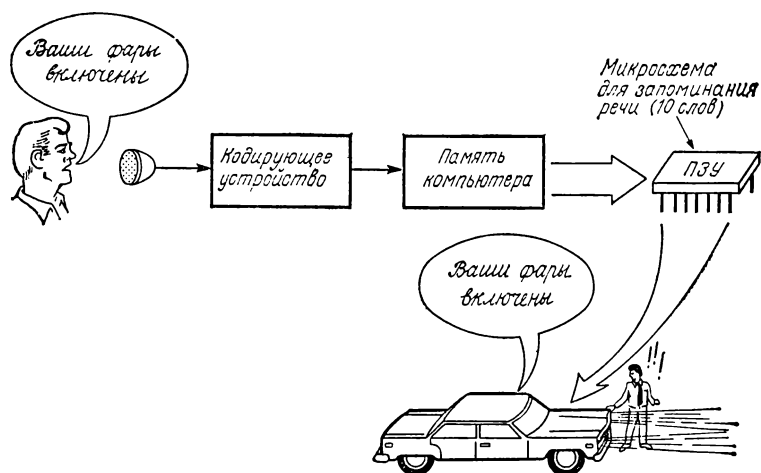


Рис. 1.1. Синтез речи с предварительным кодированием.

ствии с этим методом, принципиально отличен от описанного выше и, как правило, имеет неестественное звучание, что объясняется особенностью источника речи. Формантный синтезатор говорит голосом робота, ибо получаемая на его выходе речь не имеет в своей основе естественной человеческой речи, как это было в предыдущем методе. Другими словами, речь, генерируемая формантным синтезатором, действительно создается в компьютере. В основу формантного синтезатора положены принципы акустического моделирования голосового тракта человека. Более детально они будут рассмотрены в дальнейшем. Для генерирования формантных частотных полос, которыми характеризуется чело-

веческая речь, используются полосовые фильтры. Суммарный выходной сигнал формантных фильтров достаточно близко соответствует частотному спектру речи человека, и наш слух интерпретирует его как речевое сообщение.

Преимущество формантного метода синтеза — в его универсальности, вытекающей из присущей этому методу возможности иметь неограниченный словарь. Поскольку речь создается в этом методе из отдельно генерируемых звуков, правильно расставив звуки, можно произнести любое слово. Эта универсальность, однако, достается не бесплатно — за нее приходится расплачиваться ухудшением разборчивости речи. Во многих случаях понять, что говорит формантный синтезатор, нелегко. Кроме того, дополнительные трудности при реализации неограниченного словаря создает множество имеющихся в английском языке исключений из правил написания и произношения слов.

Наиболее распространенный способ возбуждения синтезатора формантных частот состоит в использовании многочисленных поддающихся идентификации звуков речи, называемых фонемами. Образуя последовательность фонем, подобную той, что существует в естественной речи, фонемный синтезатор может довольно хорошо копировать нашу речь. Общая схема устройства, осуществляющего такой процесс, показана на рис. 1.2. Само собой разумеется, что при столь упрощенном представлении формантного синтеза опускаются трудности реальной генерации речи по этому методу. Но основные особенности типичного формантного синтезатора на этой схеме все же отражены.

Преимущество формантного синтезатора с фонемным возбуждением с точки зрения его использования в персональном компьютере заключается в том, что любое слово, поддающееся фонетическому описанию, таким синтезатором может быть произнесено. Выходной словарь синтезатора, подключенного в качестве периферийного устройства к персональному компьютеру, не зависит от набора слов, помещенных изготовителем в память. Возможно, следующее соображение поначалу покажется вам незначительным, но задумайтесь на мгновение, могли ли бы, например, изготовители ввести в заранее составляемый для синтезатора словарь вашу фамилию? Вряд ли! Этот пример как раз и иллюстри-

рует универсальность формантного синтезатора. Такая возможность есть у всех типов синтезаторов с фонемным возбуждением [существуют также фонемные синтезаторы с линейным предиктивным кодированием (ЛПК)], а роскошь неограниченного словаря никакому другому методу синтеза речи не доступна.

Формантный метод синтеза речи точнее всего можно описать как цифровое моделирование голосового трак-

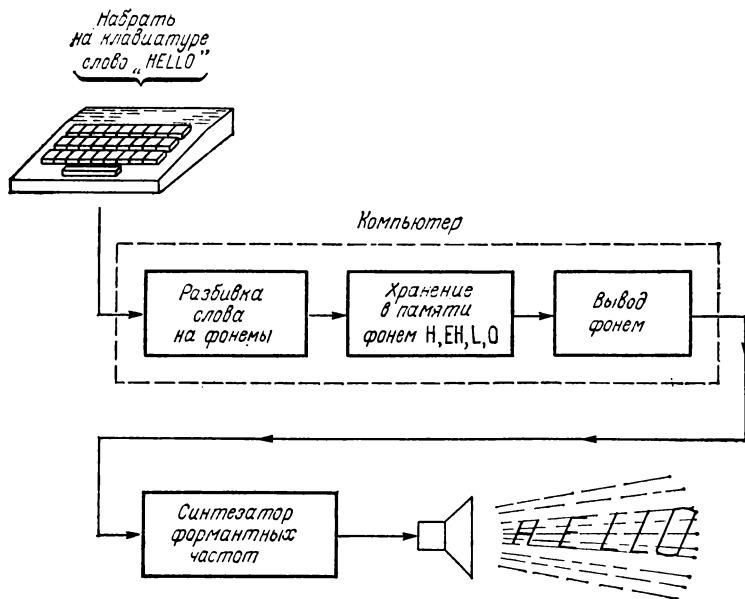


Рис. 1.2. Формантный синтез речи.

та человека. Наиболее распространенная реализация этого метода известна под названием линейного предиктивного кодирования (ЛПК) речи. Есть еще и другие способы реализации, весьма похожие на ЛПК, — это частичная автокорреляция (паркор) и параметрическое кодирование сигналов речи. Все эти способы насыщены математическими выкладками, поскольку в них используют математические модели голосового тракта. Голосовой же тракт человека — довольно сложная акустическая система, и уравнения, с помощью которых описывается его работа, также весьма сложны.

Преимущества синтезаторов третьего типа обуслов-

лены простотой их реализации в виде цифровых интегральных микросхем, вытекающей отсюда меньшей себестоимостью производства и меньшей эквивалентной скоростью передачи информации. Словарь в синтезаторах этого типа, как и в первом методе синтеза речи, обычно создается с участием говорящего человека. Но в данном случае в память не записываются непосредственно слова и фразы, а производится выделение частотных и голосовых параметров речи. Такой подход позволяет значительно уменьшить объем памяти, необходимый для получения речевого вывода. В этом и заключается одно из главных достоинств данного метода

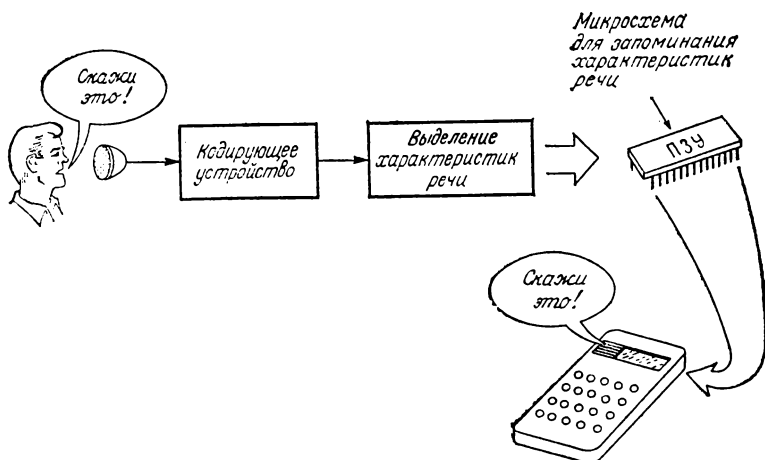


Рис. 1.3. Синтез речи с цифровым моделированием голосового тракта.

синтеза речи. Примером преимуществ этого метода может служить выпущенная фирмой «Тексас инструментом» обучающая детская игра «Говори и пиши по буквам». Это сравнительно недорогое устройство синтеза речи по методу ЛПК обладает словарным запасом примерно в 200 слов. Компьютер, управляющий синтезатором, размещен внутри устройства и выполнен на базе 4-разрядного микропроцессора. После этой обучающей игрушки в последующие годы было создано много других аналогичных говорящих устройств малого размера. Хотя говорящие игрушки не были абсолютной новинкой (куклы и телефоны со встроенными в них мини-

турными проигрывателями появились задолго до них), введение в игрушки синтезаторов значительно улучшило разборчивость речи и позволило существенно повысить механическую надежность изделий.

Принцип действия устройств, основанных на цифровом моделировании речевого тракта человека, поясняется на рис. 1.3. Внешне этот метод весьма напоминает синтез речи путем кодирования формы сигнала, однако он существенно превосходит последний длительностью речевого сообщения при одинаковых объемах памяти. Различия в качестве речи очень малы и едва заметны для среднего слушателя. Благодаря эффективному использованию памяти снижается стоимость устройства в расчете на одно записанное слово. Поэтому расширение спроса на говорящие устройства со все более объемным словарем будет, несомненно, способствовать распространению синтезаторов речи, основанных на цифровом моделировании голосового тракта.

Что будут говорить синтезаторы?

Диапазон применения синтезаторов речи сегодня поистине безграничен. Прежде всего наибольшую пользу речевой вывод может принести тем, кто плохо видит или полностью потерял зрение. Таким людям говорящая машина дает возможность пользоваться новыми читающими устройствами, в которых применение кода Брайля исключается. И значение говорящего читающего устройства здесь трудно переоценить. Метод оптического распознавания знаков в сочетании с синтезаторами речи, имеющими неограниченный словарь, позволит слепым знакомиться с литературой без посторонней помощи. Пока создано лишь первое поколение таких читающих машин, но они уже становятся большим благом для слепых.

Еще одно важное назначение синтезаторов речи — помощь тем, кто по каким-либо причинам не может говорить. В прежние времена единственным способом общения для таких людей были либо разговор жестами, либо письменная форма общения. Говорящая машина с клавиатурой для генерации речи позволит им говорить искусственным голосом почти так же легко, как пзобразать слова жестами. Пока высокая стоимость таких машин не позволяет выпускать их серийно, но расшире-

ние производства синтезаторов речи должно, по-видимому, привести к снижению цен на эти устройства.

Если теперь от этих весьма нужных применений синтезаторов речи обратиться к другим, на первый взгляд менее существенным, то и здесь мы увидим много интересного. Например, синтезатор речи можно встретить в электронном оснащении кабины пилота на современном авиалайнере; он служит здесь для предупреждения пилотов о неполадках, грозящих опасностью. Если вы обратите внимание на количество приборов, за которыми приходится следить пилоту авиалайлера, то поймете, что при таком обилии шкал летчик вполне может пропустить момент, когда зажжется маленькая лампочка, сигнализирующая, например, о возникновении пожара в двигателе. Поэтому, чтобы уменьшить информационные перегрузки пилотов во время полета, чрезвычайные сообщения такого типа, как «пожар в двигателе», «опасная высота полета» или «нарушения в механических узлах», переданы в функции автоматически срабатывающему синтезатору речевых сообщений. Его механический голос немедленно обращает внимание экипажа на возникшую опасность. Это один из тех синтезаторов, которые лучше вообще никогда не слышать!

Еще одно весьма полезное (но относящееся к числу неосновных) применение синтезаторов речи — это телефонные автоответчики. Поскольку телефонные линии используются главным образом для речевой связи (за исключением случаев, когда на телефонную линию включен цифровой терминал), синтезатор речи может прекрасно обеспечивать связь между человеком и компьютером. Сейчас уже существуют машины, которые сочетают в себе кнопочный телефонный аппарат и компьютерный синтезатор речи, автоматически передающий биржевые курсы акций, сальдо банковского баланса и данные о кредитоспособности. В дальнейшем на смену кнопочному набору наверняка придут устройства, распознающие речь, и тогда, возможно, нелегко будет установить, разговариваем мы с человеком или с машиной. Вообразите весь ужас телефонного разговора с компьютером, который не только говорит, но и слушает.

Продолжая обзор областей применения синтезаторов речи, мы без труда можем назвать новые их приложения, ибо синтезатор речи позволяет дополнить и расши-

рить основные функции многих современных устройств и приборов. Но, несмотря на то что бытовые электроприборы и устройства получают определенные преимущества от приобретенной ими способности «говорить», синтезаторы речи устанавливаются в них прежде всего потому, что сейчас они являются новинкой. Одно из главных достоинств говорящего бытового прибора состоит, например, в том, что словесные сообщения и инструкции могут быть в таком случае выданы без применения электронно-лучевого индикатора или печатных текстов. А как и следовало ожидать, стоимость синтезаторов речи значительно ниже, чем соответствующих средств визуального вывода информации.

Однако независимо от того, как синтезаторы речи могут использоваться в промышленных изделиях, для вас, вероятно, самым интересным будет использование их с персональным компьютером. Если вас занимает вопрос покупки или изготовления собственными силами синтезатора речи для компьютера, то применение ему, несомненно, найдется. Некоторые рекомендации в этом плане приведены в специальной главе нашей книги, посвященной применениям синтезаторов речи. В числе рассмотренных там возможных применений — говорящие компьютерные игры, говорящие системы охраны жилья и даже электронный дверной замок, отпирающийся голосом. Если вы действительно увлечетесь намерением дополнить свой персональный компьютер синтезатором речи, то в один прекрасный момент обнаружите, что у вас почти не стало свободного времени, хотя раньше его было достаточно. Оказывается, учить персональный компьютер говорить — одно из самых приятных и увлекательных занятий, которыми вам когда-либо приходилось заниматься. И не удивительно, что, с интересом выполняя эту работу, вы попутно узнаете немало нового.

Первой остановкой в нашем путешествии в мир синтеза речи будет знакомство с голосовым трактом человека. Если же вам очень хочется побыстрее прочесть об использовании электронных методов синтеза речи, то обратитесь к гл. 6. Однако не следует игнорировать и гл. 2—5, ибо они рисуют широкую картину пути, который прошла в своем развитии современная техника синтеза речи. Для полного понимания работы любой системы синтеза речи прочесть эти главы необходимо.

Глава 2

КАК МЫ ГОВОРИМ?

Скажите: *Hello*

Прежде чем приступить к рассмотрению биомеханических аспектов речи, проведем простой эксперимент. Откройте рот и скажите: *hello* (хэллоу). Неправда ли, это было совсем легко? По существу, мы совсем не задумываемся над тем, как идет процесс, называемый речью.

Далее мы рассмотрим, какие процессы происходят, когда мы мысленно формулируем слово *hello* и передаем его (через центральную нервную систему) в виде команд различным мышцам для генерации речи. Мы узнаем также, как человек при помощи слуха постоянно контролирует свою речь, сравнивая звучание произносимого слова с речевой памятью. На первый взгляд может показаться, что для произнесения столь простого слова, как *hello*, не требуется сколько-нибудь заметного участия биологических и биохимических процессов организма. Если бы это было так, то и генерирование искусственной речи было бы достаточно простым делом. По мере углубления в процесс генерации всего одного слова человеческой речи мы начнем понимать сложности, присущие этому процессу, который справедливо называют «даром речи». И тогда нам станут более понятны трудности, с которыми приходится сталкиваться при создании искусственной речи.

Работа центральной нервной системы

Основой для произнесенного слова *hello* была мысль, возникшая в мозге. В нашем случае эту мысль пробудило напечатанное слово *hello*. При этом в мозг по зрительному нерву поступили нервные импульсы. Затем мозг, быстро осуществив поиск в нашем речевом словаре, нашел там слово *hello* и выдал соответствующие параметры речи. На рис. 2.1 в упрощенном виде показана схема физиологического процесса речи с визуальным

входным звеном. Эта схема дает представление о входных и выходных каналах, через которые мозг управляет органами нашей речи. Перейдем к более подробному их рассмотрению.

Визуальные сигналы поступают из глаза в мозг по зрительному нерву. Одна из наиболее интересных особенностей этого пути состоит в том, что он непосредственно связывает глаз с головным мозгом — минуя спинной мозг, через который в головной мозг поступают

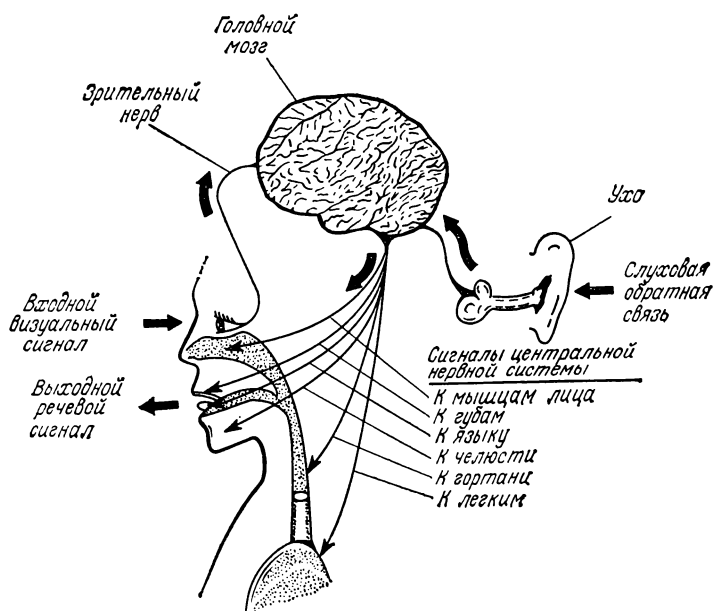


Рис. 2.1. Схематическое изображение процесса генерации речи.

другие сигналы. Тем самым не только предотвращается какая-либо путаница визуальных сигналов с сигналами от других органов тела, но и обеспечивается почти мгновенная зрительная реакция. В противном случае — из-за большого времени распространения нервных импульсов через центральную нервную систему — время нашей реакции на визуальные сигналы намного бы возросло.

Поскольку нас интересует работа мозга и нервной системы, необходимо иметь хотя бы общее представление

о «проводниках», по которым сигналы поступают к различным периферийным устройствам нашего организма или от них. У человека роль этих «проводников» выполняют нервные волокна. Естественно спросить: можно ли рассматривать живые организмы как состоящие из «проводников»? Разумеется, процесс электрической проводимости представляет собой явление более сложное, чем просто направленный переход электронов от одного атома к другому. Рассматривая процесс на уровне атомов, легче понять, каким образом пучок нервных клеток, называемых нейронами, может за малые доли секунды передавать по телу электрохимические импульсы.

Крошечные нейроны (у человека их насчитывается приблизительно 15 млрд.) представляют собой микроскопические массы протоплазмы, заключенные в сравнительно неплотные оболочки, или мембраны. Ради простоты будем считать протоплазму основным компонентом живой клетки. И поскольку в нашу задачу не входит изложение основ биологии, допустим, что именно нейроны и выполняют функции проводников, по которым сигналы передаются от мозга и к нему. Процесс, посредством которого сигнал проходит через живые клетки, напоминает электрохимический процесс, происходящий в аккумуляторе. Иначе говоря, если, скажем, никель и кадмий вместе с их окислами образуют аккумулятор, то такая система способна удерживать электрический заряд. Подобным же образом нейроны пропускают электрические заряды от головного мозга к различным нервным окончаниям.

На конце каждого пучка нервных волокон имеются клетки «ввода-вывода», называемые соответственно рецепторными и эффекторными. Зрительные и слуховые рецепторы действуют как телекамера и микрофон, осуществляя ввод визуальных и звуковых сигналов. Эти чувствительные датчики посылают свои сигналы по нейронам в головной мозг, где деятельность нервной системы продолжается в форме поисков в памяти нужной информации. Когда нужная информация, хранящаяся в клетках головного мозга, найдена, производится ее выборка для последующего использования в распознавательных и двигательных центрах мозга. Эту последовательность действий нам будет легче понять, если мы вспомним, что в нашем эксперименте произнесению сло-

ва *hello* предшествовало его прочтение. Вероятно, это было обычное приветствие, которое произносят при встрече люди. Оно, возможно, вызвало также чувство дружелюбия и благосклонности. Когда эти мысли — почти бессознательно — появились у нас, мгновенно начался процесс выборки хранящейся в памяти информации, связанной с управлением двигательными центрами и нужной для того, чтобы сказать слово *hello*. Как мы увидим дальше в этой главе, для управления примерно 50 мышцами, которые участвуют в процессе произнесения этого слова (содержащего приблизительно пять основных фонетических звуков), требуется по меньшей мере 250 обращений к хранящемуся в памяти слову.

Сложность процесса еще более возрастает, если учесть, что для перехода от одного звука к другому требуется плавное изменение работы мышц лица. Наконец, последнее взаимодействие, которое нужно рассмотреть при анализе нашего эксперимента, — это звуковая обратная связь, обеспечиваемая нашим органом слуха, т. е. ухом. Следовательно, чтобы мы могли произнести слово точно таким, как оно зафиксировано в памяти, головной мозг должен непрерывно контролировать работу наших голосовых связок и те видоизменения, которые происходят со звуками во рту и в горле.

Функции уха

Прочитав название этого параграфа, вы, возможно, захотите задать вопрос: «Какое отношение имеет ухо к процессу речи?» И если вы действительно сомневаетесь в том, что слух играет первостепенную роль в правильной и разборчивой речи, позвольте предложить вам еще один простой эксперимент, который, возможно, развеет ваши сомнения. Возьмите магнитофон с микрофоном и, прочитав вслух какое-нибудь предложение из этой или другой книги, запишите себя на магнитофон. Теперь прочтите и запишите на магнитофон то же самое предложение, но при условии, что ваш слух будет искусственно ухудшен. Это можно сделать несколькими способами. Самый примитивный из них — это просто заткнуть уши пальцами или, надев на себя стереонаушники, включить громкость, чтобы они заглушали вашу речь. Другой способ — попросить кого-либо го-

ворить близко от вашего уха в то время, когда вы делаете запись. Ни один из этих способов, конечно, не имитирует полной потери слуха, поскольку в обоих случаях сохраняется обратная связь посредством костной проводимости. Но, сравнив записи, сделанные при разных условиях слышимости, вы, безусловно, заметите небольшие различия в речи, вызванные отсутствием слуховой обратной связи, которая обычно имеет место в процессе речи.

Самый потрясающий эффект, иллюстрирующий первостепенное значение слуховой обратной связи для генерации правильной речи, достигается, если использовать линию задержки звука, или эхо-камеру, между микрофоном и наушниками. Если вам не приходилось ранее испытывать этот эффект, скажем на какой-нибудь научной выставке или при выступлении через мощную усилительную систему в большой аудитории, то вы обнаружите, что задержка обратной связи (вашей речи) вызывает еще худшие последствия, чем вообще отсутствие обратной связи. Речь, получаемую при задержке слуховой обратной связи, точнее всего характеризует ее отрывистость, которая почти граничит с заиканием.

Описанный эксперимент можно провести и без собственного участия, наблюдая за речью людей с дефектами слуха. В их речи можно заметить некоторые общие особенности. Прежде всего речь таких людей отличается более высокой (по сравнению с нормой) громкостью, иногда она даже переходит в крик. При этом они говорят громче нормального не для того, чтобы вы поступали таким же образом, а для того, чтобы повысить уровень собственной слуховой обратной связи. Еще одна особенность речи людей с пониженным слухом состоит в том, что звуки, издаваемые с участием губ и зубов (фрикативные и взрывные звуки), находятся у них в ином соотношении с звонкими звуками (например, гласными), чем у людей с нормальным слухом. Этот эффект обусловлен тем, что костная проводимость действует как фильтр нижних частот, обеспечивая слуховую обратную связь для звонких звуков, содержащих низкочастотные компоненты. Эти колебания попадают в органы слуха с большей амплитудой, чем более высокочастотные фрикативные и взрывные звуки, к которым в английском языке относятся звуки *F*, *P*, *T*, *K*, *S*, *Z* и *CH*. Если слух ослаблен равномерно в некоторой по-

лосе частот, то в речи сохранится нормальное соотношение высокочастотных и низкочастотных звуков, но громкость ее будет несколько выше, чем обычно. При более значительной потере слуха обратная связь на высоких частотах, как уже отмечалось выше, исчезает и возникает диспропорция между звуками речи высокой и низкой частоты. К рассмотрению звуков речи и их частотных составляющих мы еще вернемся в гл. 3.

Итак, убедившись теперь, что уши действительно играют очень важную роль в устной речи человека, познакомимся вкратце с возможностями нашего органа слуха. Наше ухо — изумительный инструмент, выполняющий функции микрофона для головного мозга и обладающий динамическим диапазоном (т. е. диапазоном слышимых амплитуд звуков), который превосходит возможности почти любой электронной системы.

Частотный диапазон звуков, воспринимаемых ухом (при нормальном слухе) простирается приблизительно от 15 до 20 000 Гц. Правда, с возрастом этот диапазон несколько сужается за счет высокочастотной части. Если у детей младшего возраста высокочастотная граница может достигать 30 000 Гц, то вследствие отвердевания с возрастом барабанной перепонки и связанных с ней частей внутреннего уха верхняя частотная граница диапазона слышимости смещается в область частот 10 000—15 000 Гц. К счастью, этот возрастной процесс не сказывается на генерации речи, так как самая высокая частотная компонента речи, участвующая в слуховой обратной связи, лежит примерно в области 6000—8000 Гц.

Теории слуха

В разные времена выдвигались различные теории, призванные объяснить, каким образом ухо преобразует звуковые волны в сигналы, поступающие в кору головного мозга. Первая теория слуха была предложена Германом Гельмгольцем, выдающимся немецким ученым XIX в.; согласно этой теории, в ухе имеется ряд резонансных нитей, которые, подобно струнам арфы, колеблются в такт с приходящими звуковыми сигналами. Это было равносильно допущению, что ухо содержит множество настроенных фильтров, каждый из которых связан со своим акустическим преобразователем и

посылает сигналы в головной мозг. Эта теория получила название *резонансной теории слуха*.

Вторая теория, сформулированная в начале XX в., известна под названием *телефонной теории*. Она предполагает, что барабанная перепонка и другие части уха преобразуют звуки в электрические импульсы, подобно тому как микрофон преобразует звуковые волны в электрические сигналы. Далее предполагалось, что сигналы от барабанной перепонки передаются в головной мозг по нервам, как по проводам, связывающим микрофон с

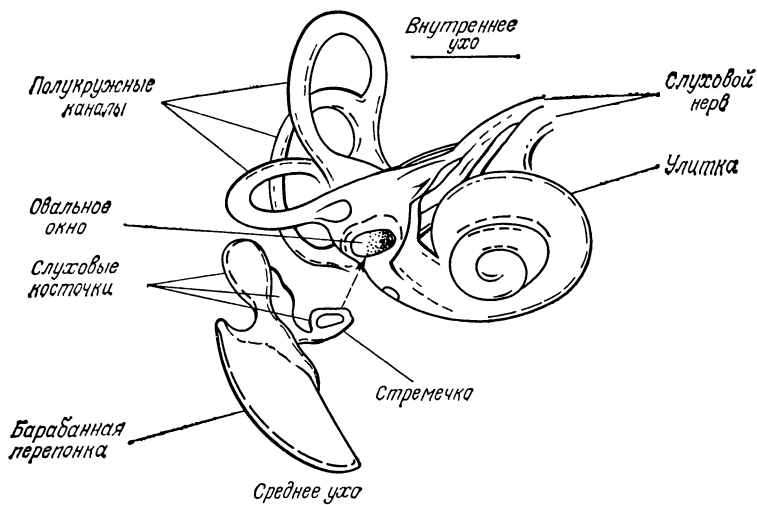


Рис. 2.2. Строение среднего и внутреннего уха человека.

усилителем. Таким образом, вся обработка сигналов, согласно этой теории, осуществляется в центральной нервной системе. Название «телефонная», естественно, объясняется тем, что в основу теории положен принцип действия систем микрофон — телефонный приемник (наушник), что в начале XX в. выглядело весьма логично и современно.

Наиболее точной следует, по-видимому, считать третью теорию. Ее называют *теорией места*. Это современное представление о слуховом механизме базируется на способности уха различать частоты, как это предусматривает резонансная теория, но без настроенных резонаторов. Различение частот здесь связывается с опреде-

ленным участком, или местом колебаний, в полукружных каналах или улитке уха (рис. 2.2). В зависимости от частоты сигналы, воспринятые рецепторными нервными клетками внутреннего уха, передаются затем по определенным слуховым нервам в виде параллельно-последовательных комбинаций в кору головного мозга, где и происходит их распознавание и анализ.

Согласно этой теории, можно упрощенно считать, что восприятие высоты тона зависит от того, какие именно нервные волокна параллельно передают импульсы в головной мозг. Восприятие же громкости зависит от числа импульсов, посылаемых по каждому из частотозависимых нервных волокон. Таким образом, мы слышим, воспринимая амплитуду сигналов нескольких частот одновременно; при этом информация поступает о каждой отдельной частоте, а не об общем спектральном уровне. Иначе говоря, наша слуховая система во многом похожа на спектральный анализатор с высоким разрешением, позволяющий определять амплитуды частот в звуковом диапазоне. Это представление не только согласуется с физическим строением уха, но и объясняет способность человека различать отклонения на 3 Гц в чистых звуковых тонах в области частот 500—1000 Гц (т. е. изменение частоты менее чем на 1 %!).

Механизм вывода речи у человека

Как возникает речь?

До сих пор мы пытались дать читателю некоторые основные сведения о том, как наш организм воспринимает и генерирует сигналы для управления речевым выводом, или речью. После того как кора головного мозга приняла сигнал генерации речи, проанализировала его и сформулировала командные коды для речевого тракта, начинается непосредственно процесс генерации слышимой речи. Это, очевидно, самое важное звено в цепи получения речи. Правда, возможна «речь» и без участия этого звена, т. е. без произнесения слов. Например, мы можем прочесть слово *hello* про себя, «сказать» и «услышать» его мысленно, не приводя в действие органы речи. При этом фактически мы приводим механизм речи в готовность, но не подаем выходных команд на многочисленные органы голосового тракта человека.

Решив же произнести слово вслух, мы отдаем большому числу наших органов команду выполнить одновременно разные действия при минимальном сознательном участии с нашей стороны. Сложность этой почти подсознательно выполняемой задачи необходимо понять, чтобы представить себе всю трудность генерации синтезируемой компьютером речи. Проследив за работой голосового тракта — он начинается в легких и заканчивается ртом, — легче будет понять, почему были выбраны те или иные методы синтеза речи. Возможно, вы сами сможете предложить какие-либо новые методы синтеза речи.

От легких до губ

Участвующие в генерации речи органы человека образуют канал сложной формы, идущий от легких к кончикам губ. Этот канал длиной около 30 см вместе с остальными органами речи изображен схематически на рис. 2.3. Чтобы разобраться в процессе образования речи от начала до конца, надо проследовать по этой схеме от легких по трахее до гортани, пройти глотку и наконец добраться до выхода через нос или рот. Хотя вся полость имеет длину около 30 см, голос действительно начинает обретать свою форму лишь на последних примерно 15 см. В конце концов после модификации и артикулирования в носовой полости, глотке и во рту звуки обретают форму определенного осмысленного слова. Но если вы не проучились два первых курса в медицинском институте, вам все же нелегко будет понять из приведенной схемы, как в голосовом тракте создается речь. Однако, сопоставив эту схему с моделью речевого механизма, показанной на рис. 2.4, можно связать различные функции голосового тракта с соответствующими речевыми эффектами. На участке ниже рта между обеими схемами существует однозначное соответствие, так как звук здесь очень мало модифицируется. Когда же звук или воздух под воздействием легких и голосовых связок проходит через глотку в рот, аналогия становится достаточно сложной. В этом легко убедиться, если попытаться сказать *ah*, сохраняя неизменными открытое положение рта, челюсти, языка, губ и зубов. Попробуйте теперь заметно изменить характеристики вашего голоса, не приводя в движение упомянутые голосовые

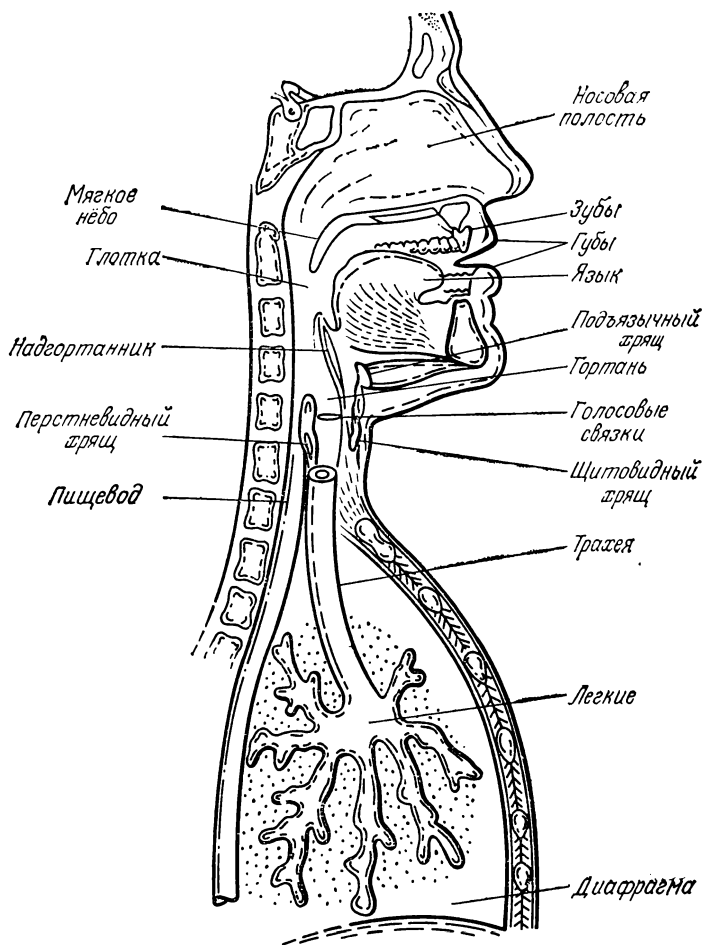


Рис. 2.3. Органы, участвующие в процессе генерации речи.

«фильтры» рта. Если вы поступили честно и оставили все без движения, то единственно, что вы сможете изменить, это высоту основного тона произносимого вами звука или его громкость (амплитуду). Внимательно выполняя этот эксперимент, заметим, что можно произвести еще одно изменение в голосовом тракте, не нарушая условий опыта. Оно достигается полным расслаблением мышц голосовых связок, расположенных ниже глотки, и выражается в прекращении колебаний связок

и превращении речи с голосом в простое прохождение воздуха, не сопровождающееся голосом. Такой вид речи, как и речь с голосом, так же важен для нормального разговора, как и звук, создаваемый колебаниями голосовых связок. Далее в этой главе мы познакомимся с белым шумом, возникающим при простом прохождении воздуха из легких через голосовой тракт. Но прежде попытаемся разобраться в работе голосового тракта и найти электронные аналогии каждому виду взаимодействий, ибо наша цель — понять, как можно с помощью электронных схем синтезировать речь.

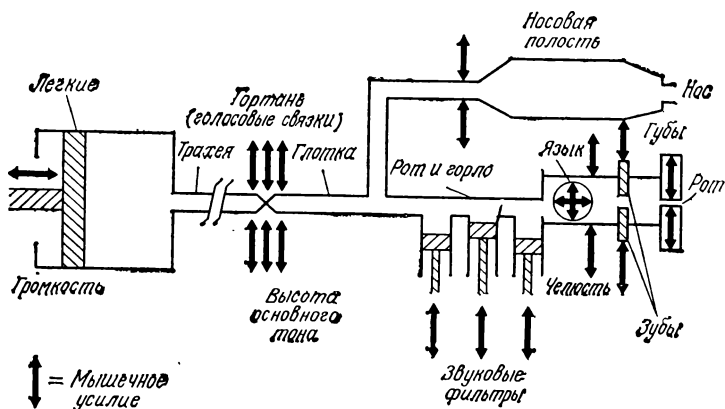


Рис. 2.4. Механическая модель голосового тракта,

Первое, что следует сделать для генерации речи, — это выдохнуть воздух из легких, в результате чего создается воздушный поток через голосовой тракт. Такой процесс происходит всегда, когда мы дышим через рот, но никакого звука при этом не слышно, пока в работу не включаются голосовые органы. Как только мы решили, что выдох должен стать речью, мы начинаем формировать звук. Это можно сделать, либо заставив колебаться голосовые связки (гортань), либо ограничивая прохождение воздуха языком, зубами и губами. В результате получаются звонкие, например гласные, звуки или соответственно фрикативные типа *S* или *F*.

Допустим теперь, что мы хотим произнести слово из гласных звуков. Подходящее слово, не требующее ни фрикативных, ни согласных звуков, — это восклицание

ow. Рассмотрим последовательно, что происходит, когда мы произносим это слово.

Легкие начинают выпускать воздух в объеме достаточно, чтобы возникли произвольные колебания голосовых связок. Если поток воздуха из легких недостаточен для возбуждения колебаний голосовых связок, то слово произносится шепотом. Иначе говоря, голосовые связки начинают колебаться при некотором пороговом значении воздушного потока из легких. Мы подсознательно «задаем» нужное количество выдыхаемого воздуха, прежде чем начать говорить, и тем самым определяем громкость (амплитуду) речи. Если же попытаться

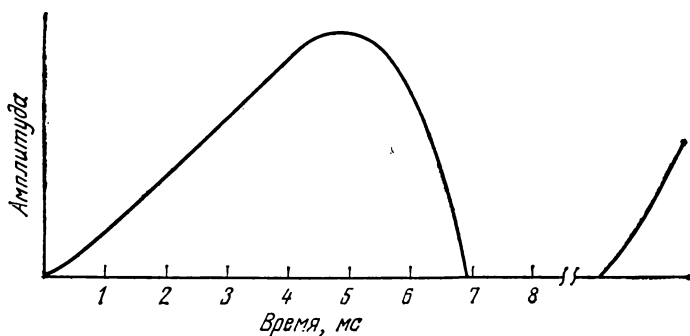


Рис. 2.5. Форма импульса на выходе голосовой щели.

сказать слово очень тихо, то нетрудно заметить, что это можно сделать, повысив высоту основного тона голоса. Но если пытаться говорить с прежней громкостью, существенно понизив тональность, то придется значительно увеличить количество воздуха, пропускаемого через голосовые связки.

Форма волны звуковых колебаний, создаваемых голосовыми связками, подобна той, что показана на рис. 2.5. Именно такую форму волны мы увидели бы на экране осциллографа с достаточным усилением, если бы нам удалось поставить микрофон в горле непосредственно на выходе из гортани. Спектральная характеристика того же сигнала, поданного на анализатор спектра или систему частотного анализа, показана на рис. 2.6. Для тех, кто незнаком с анализаторами спектра, поясним, что эта кривая характеризует распределение частотных составляющих импульса на выходе голосовой щели

(см. рис. 2.5) в полосе звуковых частот. Как можно видеть, максимум этого распределения смещен в область нижних частот: 200—300 Гц.

Когда в процессе речи голосовые связки натягиваются, частота основной составляющей (или составляющих) спектра повышается, что выражается в повышении высоты основной тональности голоса. Частота основного тона мужского голоса лежит в пределах 130—146 Гц и имеет среднее значение около 141 Гц. Частота основного тона женского голоса находится в области 188—295 Гц; ее среднее значение близко к 233 Гц. В особых случаях, когда речь содержит сильные уда-

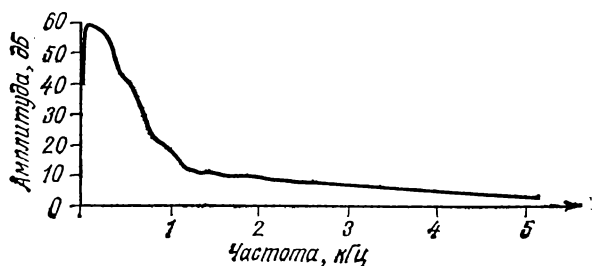


Рис. 2.6. Частотный спектр импульса на выходе голосовой щели.

рения, частота основного тона на выходе голосовой щели может достигать 480 Гц.

Не следует забывать, что рассматриваемые нами частоты — это генерируемые вследствие колебаний голосовых связок основные гармоники, а не обертоны, возникающие при резонансах в горле, носовой и ротовой полостях. Другими словами, форма волны, или импульс, показанный на рис. 2.5, повторяется с интервалом, соответствующим основным гармоникам.

Голосовые фильтры

Мы рассмотрели функции голосовой системы человека, которая создает основной звуковой тон речи посредством колебаний голосовых связок. До этого момента электронный синтез голоса сравнительно прост, так как упрощенную кривую колебаний голосовых связок можно записать в электронной памяти, а затем воспроизводить для получения основного звукового тона речи.

Наиболее сложные речевые структуры формируются в голосовом тракте выше голосовых связок. Рот, нос и губы производят фильтрацию гармоник, или обертонов, создавая частотные спектры, соответствующие известным нам звукам, которые мы называем нормальными речевыми сигналами. Электронным приборам непросто справиться с такой задачей. В самом деле, вернувшись к рис. 2.4, можно увидеть, что во рту и горле есть много подвижных, регулируемых, элементов. К тому же на этой упрощенной схеме представлены далеко не все звуковые фильтры, имеющиеся в ротовой полости.

В предыдущих параграфах этой главы мы познакомились с процессом формирования естественной речи и узнали, как головной мозг генерирует сигналы, управляющие речью. В конечном счете эти сигналы приводят в движение мышцы лица, от которых зависит звуковая фильтрация. Таким образом, для создания электрической модели голосового тракта, позволяющего получать искусственную речь, близкую к естественной, компьютер должен управлять таким же числом фильтров. Если учесть, что на лице, во рту и горле существует около 50 мышц, движением которых можно изменить импульс на выходе гортани, то трудно даже представить, насколько сложна задача моделирования 50 фильтров с помощью компьютера. Мало кого привлекла бы попытка создать систему из 50 фильтров; еще меньше, видимо, найдется энтузиастов заняться программированием управляемого изменения таких фильтров в процессе разговора. В этом основная трудность синтеза речи. Как же довести такой сложный механизм до достаточно простой формы, с которой можно работать?

Прежде всего посмотрим, что происходит, когда в звуковой фильтр (электронный или акустический), называемый также резонатором, вводится колебание, по форме соответствующее импульсу на выходе голосовой щели. Реакция резонатора на этот импульс возбуждения может быть такой, как показано на рис. 2.7. Хотя импульсы возбуждения следуют с небольшой частотой повторения, порядка 100—300 Гц (верхний график), выходной сигнал резонирующего звукового фильтра получается подобным «звону» колокола после удара по нему, причем частота «звона» в четыре-пять раз превышает частоту исходного возбуждающего сигнала. «Звон» звукового фильтра (или резонатора) служит

характеристикой резонирующего устройства. Вспомним, например, как дети часто дуют поперек горлышка открытой бутылки с лимонадом, создавая при этом звук определенного тона. Если из бутылки отпить немного лимонада, то частота звука становится ниже. В этом случае изменяется длина резонансной полости в бутылке, которая и определяет резонансную частоту звукового фильтра. Примерно то же самое происходит в полости рта, когда там возбуждаются обертоны

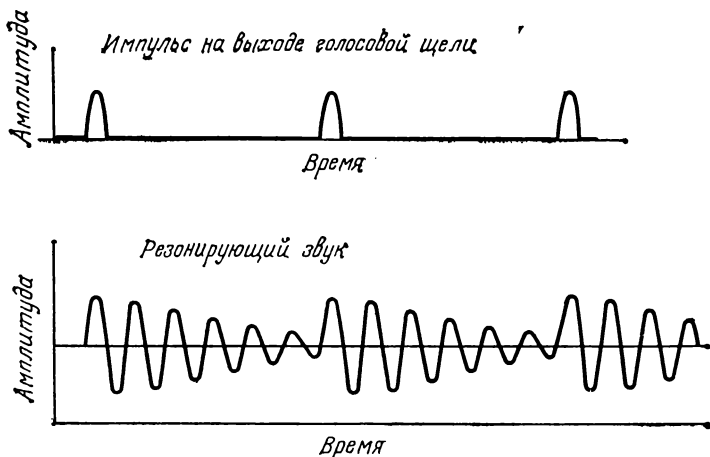


Рис. 2.7. Процесс возбуждения звукового резонатора.

импульса, образовавшегося на выходе голосовой щели. Хотя мы ограничиваем наше рассмотрение одним частотным резонатором, в голосовом тракте действует множество резонаторов-фильтров, и каждый из них характеризуется своей резонансной частотой (зависящей от формы резонатора), которая быстро меняется в процессе речи.

Различные резонансные частоты полостей голосового тракта называют *формантными* частотами. Для удовлетворительного синтеза речи обычно нужны три-четыре формантные частоты. Они лежат в диапазоне от 200 (первая форманта мужского голоса) до 2000 Гц (третья форманта женского голоса). Точным расположением формантных частот в звуковом спектре и определяется звук, который мы интерпретируем как речь. Процесс речи заметно усложняется вследствие того,

что все формантные частоты присутствуют в речи одновременно и непрерывно перемещаются вверх-вниз по частотному спектру в соответствии с особенностями произносимого слова. Поэтому, слушая говорящего человека, вы слышите звук не какой-либо одной частоты, а множество обертонов, которые образуются при фильт-

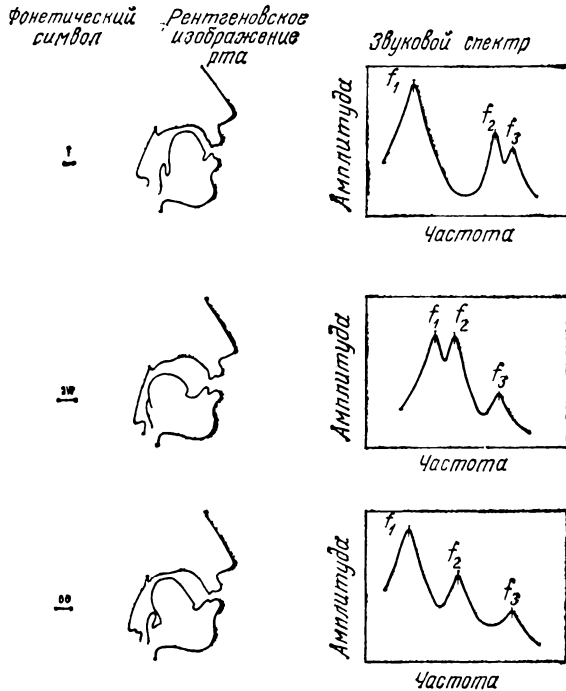


Рис. 2.8. Форманты в спектре речи.

рации импульсов, формируемых на выходе голосовой щели.

На рис. 2.8 приведены примеры частотных спектров речи. В каждом из спектров содержание формантных частот различно, так как произносятся разные гласные звуки. Максимумы спектров на частотах f_1 , f_2 , и f_3 соответствуют резонансам голосового тракта, или формантам, образующимся в процессе нормальной речи.

Теперь, исходя из спектров, показанных на рис. 2.8, посмотрим, как будет меняться со временем спектр некоего набора звуков, вроде *i-aw-oo*. Результирующий

спектр будет чем-то похож на ряд кинокадров, идущих на рисунке в последовательности сверху вниз. При переходе от звука *i* к звуку *aw* форманты смещаются по частоте так, как показано на втором «кадре». А при произнесении звука *oo* форманты становятся такими, как на третьем «кадре». Для визуального представления непрерывных спектров звуков речи разработаны специальный метод и устройство. На выходе такого устройства фактически получается видимая речь, которую часто называют «отпечатком голоса» или *спектрограммой*. Рассмотрим, что представляет собой спектрограмму.

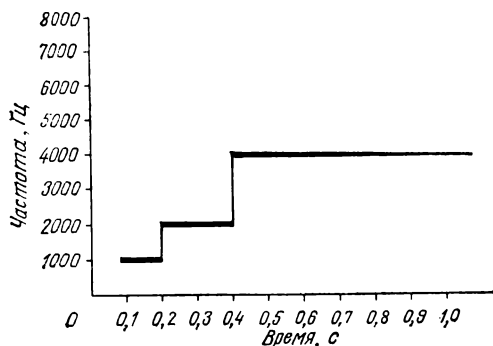


Рис. 2.9. Спектрограмма тональных частот.

ма, поскольку нам придется встречаться с таким представлением речи в дальнейшем.

Начнем с рис. 2.9, где приведено несколько моментальных снимков частотных спектров (они соединены между собой вертикальными линиями), расположенных последовательно вдоль оси времени. С увеличением амплитуды колебания данной конкретной частоты горизонтальная линия, соответствующая этой частоте, становится более черной. Другими словами, чем больше амплитуда колебания (громкость звука) какой-либо частоты, тем темнее соответствующая ей линия в каждый момент времени. Так, в момент времени 0,1 с на анализатор спектра подается тональный сигнал частотой 1000 Гц. Затем частота сигнала скачком возрастает до 2000 Гц, сохраняя ту же самую амплитуду в продолжение еще 0,2 с, после чего (в момент времени 0,4 с) снова увеличивается до 4000 Гц и начинает умень-

шаться по амплитуде на протяжении оставшегося (до окончания секунды) интервала. С помощью этой довольно упрощенной спектрограммы можно следить за изменением амплитуды и частоты входного сигнала, получая двумерное представление входного сигнала в зависимости от времени. Разумеется, если мы подадим на вход анализирующей системы две (или больше) частоты, то в процессе записи получится две или больше черных линий, соответствующих частотам и амплитудам

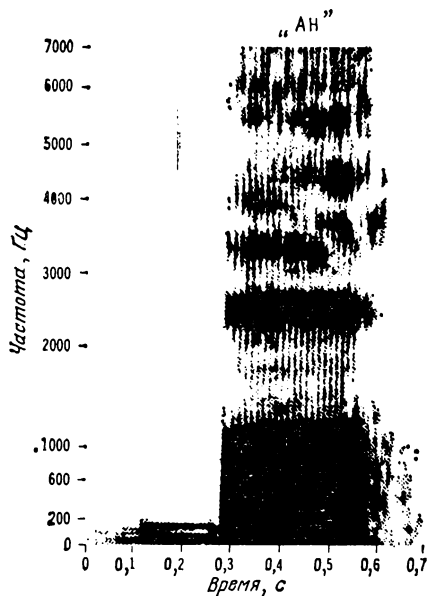


Рис. 2.10. Видеограмма звука *ah* (как в слове *father*).

звуковых сигналов. Таков механизм получения записей видимой речи, или спектрограмм, с которыми мы еще неоднократно встретимся.

Чтобы лучше представить, как выглядит типичная спектрограмма, на рис. 2.10 показана видеограмма звука *ah* в слове *father*. Заметим, что каждая из темных полос — это формантная частота, присутствующая на протяжении всего времени произнесения данного звука. Предположим далее, что мы намерены сконструировать три полосовых фильтра, управляемых напряжением, и допустим, что их характеристики должны изменяться со

временем так, чтобы они соответствовали ходу темных полос на спектрограмме. Если на входы фильтров подавать низкочастотные импульсы, подобные тем, что формируются на выходе голосовой щели, то выходной сигнал такой системы фильтров будет весьма похож на речь, записанную на спектрограмме.

Эта система фильтров в принципе представляет собой простой синтезатор речи, основанный на формантном синтезе, который используется во многих системах фонетического синтеза. Более подробно мы рассмотрим этот метод в главе, посвященной методам синтеза речи.

Резонансы со свистом и шипением

Последняя группа звуков, которой мы коснемся в данной главе, — это глухие звуки типа *s*, *t*, *f*, *p*, *sh*, *ch* и *k*, образующиеся при участии зубов, губ и языка. Эти звуки (о них мы будем говорить также в следующей главе), необходимые для понимания речи человека, возникают исключительно в результате резонансов при ограничении потока воздуха из легких. Если, например, мы ограничим выход воздуха из полости рта губами и зубами, то получим звук *f*. Если ограничить выход воздуха языком и зубами, то образуется звук *th*. Главное различие между этими двумя звуками — в силе резонанса, который возбуждается белым шумом, генерируемым ограниченным воздушным потоком. (Белый шум в данном случае определяется как звук с равномерным распределением спектра частот по звуковому диапазону.)

Поскольку источник шума при произнесении этих звуков находится у самого выхода ротовой полости, различные резонансные полости в горле и задней части рта не оказывают существенного влияния на звук, создаваемый губами или зубами. Следовательно, идея формантных частот для этих звуков (называемых фрикативными или взрывными) фактически не применима, так как большинство резонаторов в процессе формирования звуков не участвует. Более простой путь к пониманию сути дела — рассмотреть спектрограмму фрикативного звука *s* в слове *seal* (рис. 2.11). Этот звук, создаваемый губами и зубами, очень слабо фильтруется в ротовой полости и в горле; поэтому обнаружить резонансы на его спектрограмме очень трудно. Но имеющиеся резонансы характеризуют информационное

содержание этого звука и позволяют распознавать его в речи.

В данной главе мы познакомились, хотя и весьма кратко, с механизмом генерации речи у человека. Начав с описания голосового тракта человека, мы закончили наш рассказ упоминанием о синтезе речи с использованием формантных частот. Мы попытались пояснить,

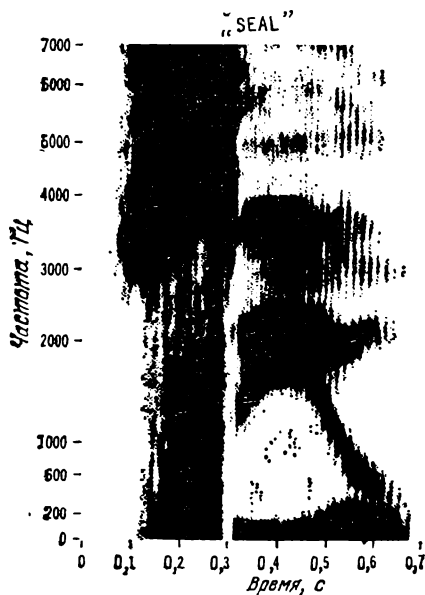


Рис. 2.11. Видеограмма, показывающая спектр звука *s* (как в слове *seal*).

как головной мозг управляет процессом речи — от момента возникновения возбуждающего сигнала в мозге до формирования звуковой речи. Внимательно прочитав эту главу, вы получите общее представление о механизме генерации речи человека. А это создает основу для понимания проблем компьютерного синтеза речи.

До сих пор нам было не важно, какой язык формируется из звуков, генерируемых голосовым трактом. Переходя к следующей главе, мы должны оговориться, что при знакомстве с другим важным разделом науки о речи — лингвистикой мы будем пользоваться английским языком.

Глава 3

НЕМНОГО О ЛИНГВИСТИКЕ

Теперь подошло время познакомить читателя с основными идеями и принципами генерации искусственной речи. Однако это почти равносильно словам: «Вот вам компьютер — напишите для него программу исчисления подоходного налога». Задача, казалось бы, поставлена вполне конкретная, но мы не определили языка, на котором должна быть написана программа. И мы потерпели бы неудачу, если бы приступили к составлению программы, не зная языка компьютера. Точно так же мы встретились бы с большими трудностями, если бы попытались исследовать и анализировать различные методы синтеза речи, не познакомившись с основами языка, на котором должен «разговаривать» компьютер. И хотя представление о сложностях синтеза речи можно составить, не зная основ лингвистики, это представление оказалось бы малополезным при решении вопроса о применении синтезатора речи. Поэтому мы начнем с краткого знакомства с некоторыми основными разделами лингвистики — фонологией и фонетикой.

Прежде всего определим, что такое лингвистика. Лингвистика — это наука об общих законах строения и функционирования человеческого языка. И если бы у нас была возможность изучать этот предмет четыре-пять лет, то тогда мы действительно отдали бы ему должное. Но мы лишь кратко коснемся некоторых областей лингвистики — наиболее существенных для понимания методов синтеза речи, а остальные разделы этой науки оставим для тех, кто глубоко интересуется исследованиями в области языка.

Знакомые с содержанием этой главы, вы, безусловно, заметите, что основное внимание в ней уделяется фонетическому анализу речи. Возможно, некоторые читатели усомнятся в необходимости изучения фонем, так как они не интересуются фонемным синтезом речи.

Пытаясь разрешить их сомнения, приведем такой пример. Допустим, когда-нибудь в будущем вы обзаведетесь собственным синтезатором речи, умеющим говорить. Однако некоторые слова у него звучат не вполне правильно. Если вы обратитесь за помощью к специалисту в области языка и словообразования, то он, вероятно, скажет вам, что синтезатор неточно произносит некоторые фонемы. Но если вы не имеете никакого представления даже о подобных терминах, то ответ специалиста будет для вас абсолютно непонятным. Прочитав эту главу, вы вряд ли станете лингвистом, но будете в состоянии с пониманием дела говорить об особенностях речи вашего синтезатора, пользуясь при этом принятыми научными терминами. Итак, фонемы — это основные звуковые единицы речи. Познакомимся с ними.

Скажите: *aaa*

Каждый помнит из детства, как врач, нажимая ложечкой на язык, просил нас произнести: *aaa*. Смысл этого упражнения заключался в том, что оно позволяло врачу поглубже заглянуть в горло говорящего. Произнося этот звук, мы создавали один из основных элементов языка — *фонему*. Данная конкретная фонема была выбрана как универсальный тест для проверки горла потому, что при ее произнесении мышцы, управляющие резонаторами в горле и ротовой полости, расслабляются, давая наиболее широкий угол обзора. Далее мы узнаем, что увеличение этих резонансных полостей приводит к уменьшению основных резонансных, или формантных, частот.

Поскольку речь пойдет о фонетическом составе типичной английской речи (которую можно слышать в США), необходимо выбрать некий стандарт в качестве основы для звуков. В данной главе мы будем анализировать звуки общеамериканского диалекта.

Один из возможных способов анализа основных звуков речи и их связи с нашим разговорным языком заключается в том, чтобы установить сходство между разговорным языком, с одной стороны, и языком компьютера — с другой. (Между ними имеется довольно большое сходство, хотя, конечно, в первом случае разговор ведется с людьми, а во втором — с компьютерами.) Возьмем для примера базовую команду, записанную

машинным кодом. Если мы сможем установить связь между таким кодом и базовым звуком речи, то в таком случае мы сможем расположить ряд звуков, или фонем, в правильном порядке, создав таким образом слова. Точно так же можно, взяв логическую последовательность машинных кодов, создать компьютерные подпрограммы. Разумеется, эти подпрограммы могут составлять часть компьютерного языка более высокого уровня (в зависимости от их структуры), например, таких языков, как Бейсик, Паскаль, Фортран и т. д. Подобным же образом соответствующие звуки речи, или фонемы, можно упорядочить в различные языки — английский, немецкий, испанский и т. д. Итак, звук речи можно рассматривать как эквивалент самой элементарной команды, записанной машинным кодом; тогда слово будет соответствовать «строке» программы, записанной на компьютерном языке более высокого уровня. Если теперь разместить структурную группу программных строк вместе так, чтобы образовалась законченная программа, то мы получим компьютерное «предложение». Итак, между компьютерным и разговорным языками действительно существует аналогия. Понимание элементов каждого из них поможет пользователю синтезатора стать знатоком как компьютерного, так и разговорного языков.

Кроме фонем в структуре английского языка имеются и другие элементы речи, которые помогают определять произносимые слова. Эти элементы — вариации произносимых фонем (их число составляет около 40); они называются *аллофонами*. Если проанализировать одну и ту же фонему, встречающуюся в 30 различных словах, то вполне может оказаться, что имеется 30 различных вариантов произнесения данной конкретной фонемы. Каждый из этих вариантов — аллофон. Поэтому число аллофонов значительно превышает число основных фонем (40), соответствующих принятому стандарту произношения каждого звука. Вариации аллофонов зависят не только от конкретных фонем и произносимого слова, но и от положения фонемы в данном слове.

Еще одним элементом нормально артикулируемой речи является *дифтонг*. Хотя дифтонг нельзя представлять определяемым символом, он характеризует звук, который получается при «переключении» голосового

тракта от произнесения одной фонемы к произнесению другой. Типичным примером может служить последовательное произнесение двух фонем гласных звуков, таких, как [a:] и [i:]. Таким образом, дифтонг характеризуется изменением звучания при переходе от одной фонемы к другой и, следовательно, может находиться в произносимом слове лишь между двумя фонемами.

Если дифтонг образуется сочетанием двух аллофонных фонем, то, естественно, он звучит по-разному. Итак, в разговорном английском языке имеется множество разных звуков, которые и обеспечивают нормальное звучание человеческой речи. При описании произношения какого-либо слова необходимо учитывать не только все звуки, но и соответствующие дифтонги и аллофоны.

Такое обилие фонетических и аллофонических звуков может охладить пыл энтузиастов, которые предпринимают попытки синтезировать речь при помощи компьютера. Однако они не должны отчаиваться: в конце туннеля все же виден свет. Даже используя ограниченную группу комбинаций фонем, аллофонов и дифтонгов, можно «сконструировать» вполне распознаваемую речь. Хотя синтезированные звуки могут несколько отличаться от звуков естественной речи, смысловое различие их незначительно. Обычно компьютерная речь напоминает «речь с иностранным акцентом», но большинство людей легко понимают ее.

Гласные звуки

Большинство гласных звуков в общеамериканском диалекте произносятся при относительно фиксированном положении мышц рта и голосового тракта, причем многие звуки произносятся с открытым ртом. При произнесении этих звуков носовая полость играет очень малую роль или вообще не играет никакой.

По существу, в общеамериканском диалекте английского языка насчитывается 12 гласных звуков (рис. 3.1).

Чтобы проиллюстрировать положения языка относительно губ и зубов при произнесении каждого из этих звуков, на рисунке показаны (приблизительно) профильные разрезы артикуляционного аппарата речи.

Отметим, в частности, что, когда язык образует горбик в задней части открытого рта, получающийся звук близок к звуку [a:], как в слове *father*. Проведем простой эксперимент. Откройте рот и скажите *a-a-a* [a:]; затем, удерживая челюсть рукой в неизменном положении и оставляя без изменения все артикуляционные органы рта, кроме языка, медленно перемещайте язык в переднюю часть рта так, чтобы он коснулся нижних зубов. Заметьте, что произносимый вами звук изме-

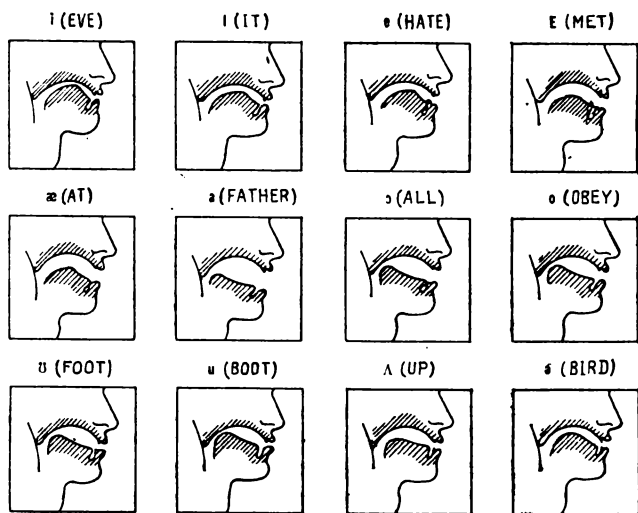


Рис. 3.1. Фонемы гласных звуков и соответствующие положения элементов голосового тракта,

нился теперь на [i:], как в слове *eve*. При произнесении большинства гласных звуков положение и форма языка отличаются, тогда как положение элементов голосового тракта в основном остается фиксированным. Если, продолжив наш эксперимент, сложить теперь губы так, как при насвистывании, то нетрудно заметить, что звук [a:] переходит в звук [u:], как в слове *foot*. Если вам захочется продолжить эксперимент с артикуляцией гласных, то, руководствуясь рис. 3.1 для выбора положения языка и губ, проследите, насколько произнесение фонем гласных сходно с тем, что изображено на рисунке.

Другие звуки: согласные

Остальные произносимые звуки английского языка — это согласные. Если при произнесении описанных в предыдущем параграфе гласных звуков положение артикуляционных элементов рта остается в основном неизменным, то произнесение согласных требует большей подвижности этих элементов. Есть, правда, несколько согласных, для произнесения которых не требуется движения элементов голосового тракта, — это фрикативные звуки; они весьма близки к гласным, но обычно гораздо сложнее по своему частотному спектру, так как в их формировании участвует значительно больше фильтрующих элементов голосового тракта.

В английской разговорной речи можно выделить пять типов согласных звуков: 1) взрывные; 2) фрикативные; 3) носовые; 4) промежуточные и 5) полугласные звуки.

Произнесение звуков первого типа — *взрывных* — сопровождается довольно активным движением элементов голосового тракта в сочетании с полной паузой в речи при произнесении фонемы. Хотя пауза может длиться всего лишь малую долю секунды, без нее речь отличается от нормальной.

В английском языке имеется шесть основных взрывных согласных. Три из них — звонкие и произносятся с колеблющимися голосовыми связками. Три — глухие, и при их формировании голосовые связки не работают. Звонкие взрывные согласные: *b*, как в слове *butter*; *d*, как в слове *dog*; *g*, как в слове *green*.

Произносятся слова, содержащие такие звонкие взрывные согласные, очень медленно, нетрудно заметить, что в речи возникает весьма короткая пауза, за которой следует характерное для взрывного звука резкое высвобождение воздуха из легких. Произнесение трех глухих взрывных звуков имеет ту же самую особенность, но колебаний голосовых связок при этом не происходит. Глухие взрывные согласные: *p*, как в слове *paper*; *t*, как в слове *tomato*; *k*, как в слове *kick*.

По звукообразованию последняя группа похожа на фрикативные звуки, которые мы сейчас и рассмотрим.

Наиболее характерная черта *фрикативных* фонем — отсутствие специфических формантных частот. Причину этого легко понять, взглянув на девять звуков, ко-

торые считаются ффрикативными. Как и взрывные звуки, они также делятся на звонкие и глухие. При произнесении звонких ффрикативных звуков (их четыре) голосовые связки должны колебаться, генерируя резонансное шипение (свист): *v*, как в слове *valve*; *th*, как в слове *there*; *z*, как в слове *zebra*; *zh*, как в слове *azure*.

Остальные пять ффрикативных (глухих) согласных звуков также содержат резонансное шипение, но при их произнесении голосовые связки не колеблются: *f*, как в слове *fanfare*; *th*, как в слове *think*; *s*, как в слове *sound*; *sh*, как в слове *sugar*; *h*, как в слове *help*.

Каждая из приведенных ффрикативных фонем содержит в себе резонансное шипение, что существенно для получения некоторых наиболее высокочастотных звуков речи. Поскольку эти фонемы «статические», голосовой тракт при их произнесении остается неизменным.

Согласные третьего типа называют *носовыми* фонемами, ибо при их произнесении звук излучается в основном через нос. Если у вас простуда и заложило нос, то эти звуки, как вы знаете, трудно произнести правильно. В самом деле, если, зажав нос пальцами, попытаться произнести эти звуки, то легко убедиться, насколько сильно они зависят от фильтрации в носовой полости. Итак, сделайте глубокий вдох, зажмите нос и попытайтесь произнести носовые согласные звуки: *m*, как в слове *mike*; *n*, как в слове *noon*; *ng*, как в слове *ring*.

Теперь вы, надеюсь, понимаете, почему у этих звуков такое название. А то, что вы произнесли с зажатым носом, больше похоже на слова *bike*, *dude* и *reek*. Иначе говоря, носовая фильтрация и ее связь с речевым выходом абсолютно необходимы для произнесения носовых фонем.

Фонемы согласных двух последних типов — *промежуточных* и *полугласных* — образуют очень малую группу в общем числе звуков английского языка. Эти фонемы очень схожи с гласными звуками и произносятся с участием голоса и при очень малой назализации. Промежуточные звуки, как можно представить себе по их названию, — это динамические звуки, для произнесения которых требуется участие артикуляционных элементов рта: *y*, как в слове *yo-yo*; *w*, как в слове *winter*.

Полугласные звуки, похожие по своей природе на

гласные, относятся к фрикативным звукам, и их произношение долгое. Полугласных звуков всего два; им сопоставляются следующие фонемы: *r*, как в слове *run*; *l*, как в слове *list*.

Этим завершается перечень основных фонем разговорного английского языка, необходимых для получения распознаваемой речи. У этих фонем существуют, конечно, вариации, которые называют *дифтонгами*, если они образуются при последовательном произнесении двух гласных фонем. При последовательном произнесении двух согласных звуков аналогичным образом получаются *аффрикаты*.

Вперед и выше

До сих пор мы занимались наименьшими звуковыми элементами английского разговорного языка — фонемами. Очевидно, однако, что из-за физических различий и эмоциональной окраски, которую придает речи говорящий, речь нельзя описать полностью одними лишь фонемами. Звучание каждой фонемы существенно зависит также от окружающих ее фонем. Когда мы начинаем группировать фонемы, создавая осмысленные фразы, мы произносим слова. Лингвистически состав слова можно анализировать в *морфемах*, *морфах* и *алломорфах*. Определение этих единиц речи довольно сложно; оно зависит не только от звуков, из которых состоит слово, но и от смыслового значения слова. Поскольку изучение смысловых значений слова и соответствующих разделов грамматики не входит в круг наших интересов, связанных с созданием синтезированной речи, мы просто примем к сведению, что эти элементы присутствуют в произносимых словах. Читатель, желающий глубже познакомиться с этими вопросами, может обратиться к списку литературы, приведенному в конце книги.

После того как фонемы выстроены в ряд, образуя более длинные и имеющие смысл последовательности звуков (называемые словами) можно выявить некоторые количественные характеристики этих последовательностей. Хотя, как уже отмечалось выше, правильное произнесение каждой фонемы в большой степени зависит от ее положения в слове, некоторые звуковые особенности фонем остаются неизменными, если рассматривать речь среднего человека. Эти особенности можно

использовать при проектировании компьютерной речи и ее сравнении с «идеальной» человеческой речью.

Возьмем, к примеру, частоту употребления наиболее распространенных слов в разговорном английском языке. Имея перечень таких слов, необходимо позаботиться, чтобы синтезатор речи по крайней мере эти слова произносил точно, ибо они встречаются чаще других. Такой список, составленный много лет назад Годфри Дьюн из Гарвардского университета, приведен в табл. 3.1. Слова расположены здесь в порядке уменьшения частоты их появления, так что слова, употребляемые чаще всего, находятся сверху списка. При этом для каждого слова указана (в процентах) относительная частота его употребления. Например, из этой таблицы видно, что слово *the* встречается чуть чаще 7 раз на каждые 100 слов. Следующее наиболее употребляемое слово *of* встречается почти 4 раза на 100 слов и т. д. Приведенная таблица может найти разнообразные применения. Так, следует заставить синтезатор речи «поупражняться» в произнесении приведенных здесь слов, чтобы убедиться, что они произносятся правильно. Данная таблица незаменима также при составлении минимального словаря для системы синтеза речи (если предусматривается хранение такого словаря в памяти системы).

Двигаясь дальше, мы можем также выявить звуки, наиболее часто встречающиеся в нашей речи (табл. 3.2). Эта таблица найдет себе применение не только при работе с системами фонетического синтеза, но и в тех случаях, когда требуется обеспечить правильное произношение звуков в любых системах синтеза речи. Каждый звук (фонема) в этой таблице дается с указанием относительной частоты его употребления; приводится также слово, характерное для произношения данного звука. Эта таблица дает нам информацию о наиболее употребляемых фонемах. Руководствуясь ею, мы можем решить, какие звуки или фонемы следует опустить в упрощенной системе синтеза речи. Иными словами, если наша цель — создать элементарную фонетическую систему синтеза речи, то, вероятно, можно опустить такие звуки, как *zh* (*azure*) и дифтонг *oi* (*boil*), поскольку частота их употребления очень мала. Такие же фонемы, как *i* (*tip*), используемые наиболее часто, необходимо включить. При создании и анализе работы синтезатора речи полезна также информация об относительной фонетиче-

**Таблица 3.1. Слова, наиболее часто встречающиеся
в английской речи**

№ п/п	Слово	Час- тота появ- ления, %	№ п/п	Слово	Час- тота появ- ления, %	№ п/п	Слово	Час- тота появ- ления, %
1	the	7,31	36	one	0,36	71	into	0,16
2	of	3,99	37	our	0,33	72	men	0,16
3	and	3,28	38	an	0,33	73	must	0,16
4	to	2,92	39	been	0,32	74	people	0,16
5	a	2,12	40	no	0,32	75	said	0,16
6	in	2,11	41	their	0,31	76	may	0,16
7	that	1,34	42	there	0,30	77	man	0,15
8	it	1,21	43	were	0,30	78	about	0,15
9	is	1,21	44	so	0,30	79	over	0,15
10	I	1,15	45	my	0,29	80	some	0,15
11	for	1,03	46	if	0,26	81	these	0,15
12	be	0,84	47	me	0,25	82	two	0,14
13	was	0,83	48	what	0,25	83	very	0,14
14	as	0,78	49	would	0,25	84	before	0,13
15	you	0,77	50	who	0,24	85	great	0,13
16	with	0,72	51	when	0,23	86	could	0,13
17	he	0,68	52	him	0,23	87	such	0,13
18	on	0,64	53	them	0,22	88	first	0,13
19	have	0,61	54	her	0,22	89	upon	0,12
20	by	0,60	55	am	0,21	90	every	0,12
21	not	0,58	56	your	0,21	91	how	0,12
22	at	0,58	57	any	0,21	92	come	0,12
23	this	0,57	58	more	0,21	93	us	0,12
24	are	0,54	59	now	0,21	94	shall	0,12
25	we	0,52	60	its	0,20	95	should	0,11
26	his	0,51	61	time	0,20	96	then	0,11
27	but	0,50	62	up	0,20	97	like	0,11
28	they	0,47	63	do	0,20	98	will	0,11
29	all	0,46	64	out	0,20	99	little	0,11
30	or	0,45	65	can	0,19	100	say	0,111
31	which	0,45	66	than	0,19			
32	will	0,44	67	only	0,18			
33	from	0,43	68	she	0,18			
34	had	0,41	69	made	0,17			
35	has	0,39	70	other	0,16			

Таблица 3.2. Частота появления звуков речи

№ п/п	Звук	Как в слове	Частота появления, %
1	i	tip	7,94
2	n	navy	7,24
3	t	tot	7,13
4	r	rare	6,88
5	uh	tun	5,02
6	s	sis	4,55
7	d	dad	4,31
8	ae	tap	4,17
9	ee	feet	3,89
10	l	lilly	3,74
11	eh	ten	3,44
12	th	then	3,43
13	ah	top	3,33
14	z	zebra	2,97
15	m	mama	2,78
16	k	kick	2,71
17	a	tape	2,35
18	v	vivacious	2,28
19	w	wine	2,08
20	p	paper	2,04
21	f	fluffy	1,84
22	h	hat	1,81
23	b	bob	1,81
24	oh	tone	1,63
25	oo	tool	1,60
26	i	bike	1,59
27	aw	talk	1,26
28	ng	sing	0,96
29	sh	sugar	0,82
30	g	go	0,74
31	u	took	0,69
32	y	yoyo	0,60
33	ou	our	0,59 (дифтонг)
34	ch	chalk	0,52
35	j	judge	0,44
36	th	thick	0,37
37	ew	few	0,31 (дифтонг)
38	oi	boil	0,09 (дифтонг)
39	zh	azure	0,05

Таблица 3.3. Относительная мощность звуков речи

№ п/п	Звук	Как в слове	Относительная мощность	№ п/п	Звук	Как в слове	Относительная мощность
1	aw	talk	680	17	ch	church	42
2	ah	top	600	18	n	Nancy	36
3	uh	ton	510	19	j	judge	23
4	ae	tap	490	20	zh	azure	20
5	oh	tone	470	21	z	zoo	16
6	u	took	460	22	s	sis	16
7	a	tape	370	23	t	tot	15
8	eh	ten	350	24	g	go	15
9	oo	tool	310	25	k	cook	13
10	i	tip	260	26	v	vote	12
11	ee	peek	220	27	th	that	11
12	r	rare	210	28	b	bob	7
13	l	lilly	100	29	d	dad	7
14	sh	sugar	80	30	p	paper	6
15	ng	sing	73	31	f	fluffy	5
16	m	mama	52	32	th	thick	1

ской мощности (или громкости), которая характеризует каждую фонему, произнесенную (средним) человеком.

Из табл. 3.3 мы получаем представление об относительной мощности наиболее широко используемых фонем. Эта информация помогает нам приблизить речь компьютерного синтезатора к человеческой. В табл. 3.3 каждая фонема приводится с характерным словом и с указанием относительной громкости (по отношению к другим фонемам в таблице). У самой мощной фонемы *aw* (*talk*) громкость в 680 раз больше, чем у фонемы *th* (*think*). Если эту разность уровней выразить в децибелах (дБ), то она составит 28 дБ. Заметим, что по громкости первыми стоят гласные звуки, за ними идут промежуточные и носовые. Остальную часть таблицы занимают фонемы фрикативных согласных, которые произносятся более мягко. Конечно, значения громкости, приведенные в табл. 3.3, не учитывают слоговых ударений, которые могут увеличивать или уменьшать соотношения громкостей некоторых фонем. Приведенные

здесь громкости соответствуют некоторой усредненной группе фонем, произносимых при спокойном разговоре.

Эти три таблицы могут оказаться полезными при проведении испытаний и проверке систем синтеза речи. Разумеется, конечные результаты испытаний подобных устройств зависят от субъективной оценки слушателя. Ибо независимо от того, насколько искусственная речь с технической точки зрения соответствует естественной речи человека, оценку качества компьютерной речи дает все же человек. Наши же таблицы могут помочь приблизить параметры систем синтеза речи к количественным характеристикам обычного человеческого разговора. Если нужно оценить лингвистические способности системы синтеза речи, то следует подготовить серию контрольных слов, позволяющих проверить звучание фонем, используемых в речи. Чтобы лучше проверить качество речи синтезатора, можно пригласить какого-нибудь «свежего» человека, который ранее не слышал, как говорит система. Пусть тогда компьютер произнесет несколько слов из своего словаря, а вы оцените процент понимания слушателем каждого слова. Если беспристрастный слушатель неправильно воспринимает отдельные слова и группы слов, это, безусловно, говорит о недостаточном качестве воспроизведения синтезатором определенной фонемы или группы фонем. Следующий параграф поможет вам оценить имеющуюся в вашем распоряжении систему синтеза речи, если у вас только есть желание заниматься такой проверкой и имеется возможность программирования всего речевого словаря синтезатора.

Итак, предположим, что вам удалось создать устройство, которое говорит. Если вы намерены быть единственным слушателем своего детища, то, возможно, вполне удовлетворитесь имеющимися у него разговорными способностями. Но если вы хотите, чтобы речь вашей системы понимали и «неподготовленные» слушатели, то вам придется позаботиться, чтобы синтезатор имел правильное, общепринятое произношение. Ниже предлагается список слов, который можно использовать для развития артикуляции синтезатора. Эти слова можно запрограммировать в память, чтобы синтезатор мог произносить их последовательно или в случайном порядке, давая возможность проверить правильность его произношения. Если неподготовленный слушатель смо-

жет правильно опознать каждое из перечисленных слов, то это будет означать, что система имеет очень хорошее (точное) произношение. Неправильное понимание слушателем слов свидетельствует о наличии слабых мест в конкретной фонетической области и о необходимости внести изменения в программу или отрегулировать уровень громкости при генерации данного конкретного звука. Для каждой фонемы подобрана группа контрольных слов. Проверьте, сможете ли вы идентифицировать на слух каждую фонему, использованную в приведенных группах слов.

1. Saw, Horse, Horn, Ball, Talk
2. Yard, Clock, Top, Block, Star, Arm
3. Gloves, Rug, Truck, Tub, Button, Ton
4. Tap, Hat, Can, Black, Grass, Basket
5. Tone, Boat, Coat, Snow, Stove, Comb
6. Book, Cook, Foot, Look, Took
7. Tape, Cake, Grapes, Table, Lady, Tail
8. Ten, Bed, Dress, Red, Steps, Feather, Sled
9. Tool, Blue, Moon, Tooth, Shoe
10. Tip, Chicken, Fish, Pillow, Pig
11. Peek, Cheese, Meet, Sleep, Trees, Green, Feet
12. Radio, Rake, Barrel, Car, Tire, Rabbit, Red
13. Ladder, Lease, Leg, Letter, Ball, Bottle, Look
14. Sheep, Shelf, Dish, Fish, Brush, Push, Shoulder, Shake
15. Finger, Sing, Swinging, Ring, Tongue, Blanket
16. Move, Music, Memory, Most, More, Meek, Mimic, Movie
17. Chair, Cheese, Chicken, Watch, Catch, Matches, Teacher, Speech, Church
18. Nasal, Know, Knife, Candle, Woman, Nancy, Spoon, Man
19. Juice, Engine, Orange, Soldier, Bridge, Joke, Jump
20. Glacier, Azure, Measure, Television
21. Music, Zoo, Roses, Ears, Nose, Zebra, Scissors
22. Seven, See, Saw, Sleep, Spoon, Basket, Glasses, Face
23. Table, Tire, Butter, Tot, Letter, White
24. Gloves, Grass, Gun, Golf, Digging, Wagon, Rug, Flag
25. Crack, Pocket, Black, Clock, Cook, Fake
26. Vase, Violet, Vivacious, Cover, Drive, River, Stove
27. Thimble, Three, Thin, Thick, Mouth, Teeth
28. Bed, Boat, Rabbit, Ribbon, Umbrella, Table, Bob
29. Dog, Drink, Indian, Radio, Dud, Bed, Wood
30. Paper, Pencil, Airplane, Apple, Pop, Cap, Rope, Sleep
31. Feather, Finger, Fire, Fluffi, Elephant, Laugh, Roof, Knife
32. These, Those, Brother, Then, Father, Feather, Loathe

Каждая из 32 основных фонем, представленных в списке, необходима для адекватного понимания речи, генерируемой синтезатором. Если у синтезатора обнаруживаются какие-либо дефекты в произношении одной или нескольких групп слов, то недостаток его

речи можно характеризовать конкретными фонемами, содержащимися в неправильно произнесенных или неверно понимаемых группах слов. Сведения, изложенные в предыдущих параграфах данной главы, должны позволить вам идентифицировать каждую из специфических фонем в приведенном выше списке слов. Но не огорчайтесь, если у вас возникнут трудности с какой-либо группой слов, — ответы даны ниже:

№ п/п	Фонема	Слово	№ п/п	Фонема	Слово
1.	aw	talk	17.	ch	church
2.	ah	top	18.	n	Nancy
3.	uh	ton	19.	j	joke
4.	ae	tap	20.	zh	azure
5.	oh	tone	21.	z	zoo
6.	u	took	22.	s	see
7.	a	tape	23.	t	tot
8.	eh	ten	24.	g	golf
9.	oo	tool	25.	k	cook
10.	i	tip	26.	v	vivacious
11.	ee	peek	27.	th	thick
12.	r	rake	28.	b	bob
13.	l	look	29.	d	dud
14.	sh	shake	30.	p	pop
15.	ng	ring	31.	f	fluffy
16.	m	move	32.	th	then

Описанный тест пригоден для любого синтезатора речи и вообще для любого говорящего, если требуется проверить правильность произношения различных фонем или поупражняться с целью улучшения их произношения. Этот тест можно также использовать для сравнительных оценок качества речи.

Последнее слово о лингвистике

Лингвистика и техника синтеза речи очень тесно переплетены друг с другом. Правда, если в ваши намерения не входят программирование звуков или генерирование связной речи из фонем, то изложенные выше сведения по фонетике, возможно, помогут вам лишь при исследовании других особенностей синтезированной речи. Если же, напротив, вы занимаетесь именно фонетическим (формантным) синтезатором, который строит

речь из отдельных звуков, то вам необходимо иметь представление о фонетическом запасе вашей системы. Тогда можно позаботиться и о том, чтобы система произносила звуки так, как вы хотите. Надо иметь, однако, в виду, что посторонний слушатель может воспринимать звуки совершенно иначе, чем вы.

При создании системы синтеза речи, рассчитанной на посторонних слушателей, необходимо учесть еще одно важное обстоятельство. Дело в том, что в рамках североамериканского варианта английской речи существует очень большое число различных *диалектов*. Диалект — под которым понимается разновидность языка, обусловленная географическими или социальными признаками говорящих на нем людей, — значительно отличается от «стандартного» языка использованием некоторых специфических грамматических форм и особенностями произношения. Когда создатель синтезатора приступает к программированию речи компьютера, он бессознательно пользуется для этого диалектом, на котором говорит сам. И если кто-либо в другой части страны услышит речь синтезатора, то он, возможно, сумеет определить, откуда родом человек, составивший программу говорящего компьютера.

Лучший способ избежать программирования на своем диалекте — это внимательно слушать передачи национального телевидения. Такие передачи, как обзор новостей, могут служить образцом звучания для компьютера. И лишь когда система обретет «дикторскую» речь, ее смогут легко понимать люди из любой части страны.

Наконец, есть еще один момент, на который следует обратить внимание при рассмотрении лингвистических аспектов: программируя фразы для компьютера, следует учитывать зависимость некоторых грамматических форм от географии места, где будет использоваться говорящая система. Например, твердую сердцевинку плода вишни в разных частях страны могут называть по-разному: косточкой, сердцевинкой или семечком. В определенной географической области каждое из этих слов имеет один и тот же смысл, но в другой части страны его значение может оказаться совсем другим. Касаясь таких «нюансов» лингвистики, я вовсе не хочу сказать, что речь компьютера обязательно должна быть грамматически идеальной. Я хочу лишь показать, что

на большой территории, где говорят на одном языке, в речи существуют различия. И чтобы речь одинаково хорошо понимали во всех частях страны, она должна звучать по возможности «более нейтрально». Именно такую речь мы чаще всего слышим в телевизионных передачах новостей.

Подводя итоги, можно сказать, что глубокое понимание и использование элементов лингвистики послужит надежной основой при составлении довольно полного и хорошо понимаемого словаря для системы синтеза речи. Чем больше времени и внимания вы будете уделять вопросам лингвистики при обучении компьютера навыкам речи, тем разборчивее и понятнее он будет говорить. Естественно, нельзя забывать о грамматических конструкциях и структуре предложения. Но тех, кто интересуется вопросами, связанными со структурой предложений, мы отсылаем к списку литературы, приведенному в конце книги. Внимательность при составлении предложений, безусловно, необходима, ибо в противном случае посторонний слушатель может иногда не так понять смысл фразы, которую по вашей воле прознесет компьютер. Чтобы убедить вас в возможности подобных недоразумений, предлагаю вам для размышления два следующих предложения:

Time flies like an arrow. (Время летит, как стрела.)

Fruit flies like a banana. (Плодовые мушки любят бананы.)

Глава 4

ЭТИКА «ПОВЕДЕНИЯ»

КОМПЬЮТЕРА — СИНТЕЗАТОРА РЕЧИ

В двух предшествующих главах мы рассказали, как формируется речь человека, начиная с базовых звуков и кончая образованием английских слов и предложений. Теперь мы располагаем довольно прочной основой для перехода к компьютерной речи. Однако между двумя, казалось бы, сходными понятиями — человеческой и компьютерной речью — лежит некая «серая зона», мостом через которую служит интеллект. Я, разумеется, имею в виду интеллект не читателей, а тех, кто говорит. В этой главе мы вкратце остановимся на участии разума в процессе речи и на том, как это сказывается на различиях между речью естественного и искусственного происхождения.

Хотя, возможно, некоторые посчитают эту главу экстраординарной, излагаемая в ней информация исключительно важна для практических применений компьютерных синтезаторов речи. И от того, будет ли она принята во внимание при использовании готового устройства, может зависеть успех или неудача его в реальных условиях. Разговор здесь пойдет о восприятии человеком речи, генерируемой компьютером.

Произнося ту или иную фразу, мы (почти подсознательно) учитываем множество внешних факторов, определяющих, как и когда говорится фраза. Компьютер, естественно, сделать этого не может, если только он не обучен этому специально или не обладает искусственным интеллектом. В настоящей главе вы узнаете, как позаботиться о том, чтобы высказывания компьютера были своевременными и «по существу».

Правила компьютерной речи

Для начала рассмотрим некоторые основные правила «этикета», которым должен следовать говорящий компьютер. Далее мы более подробно познакомимся с этими

правилами и необходимостью их соблюдения при непосредственных применениях компьютерных синтезаторов речи. Предлагаю вам поразмышлять над соображениями автора по поводу компьютерного речевого этикета.

1. Если компьютер может сказать что-нибудь не вовремя, он это сделает.

2. Если вы будете неоднократно демонстрировать говорящий компьютер одним и тем же людям, то каждый раз они будут надеяться услышать речь лучшего качества.

3. В то время как человека можно попросить помолчать, к запрограммированному компьютерному синтезатору речи бесполезно обращаться с подобной просьбой.

4. Компьютеры, прерывающие говорящего, как и люди, поступают бестактно. «Идеально воспитанный» компьютер должен вступать в разговор, лишь когда в помещении тишина.

5. Короткая, неожиданно сказанная компьютером фраза потеряется в шуме обычного разговора.

6. Если компьютер намерен сообщить что-то важное, он должен заранее оповестить об этом сигналом или каким-либо звуком, привлекающим внимание слушателей.

7. Говорящий компьютер похож на ребенка: он знает, как говорить, но не знает когда.

8. Компьютер, говорящий слишком много, будет выключен.

9. Компьютер, говорящий слишком мало, останется без внимания.

10. Слушатель должен быть готов услышать искусственный голос компьютера, если, конечно, последний не обладает натуральным голосом.

Эти правила особенно важны, когда вы только начинаете «выводить» ваш компьютер в общество. Конечно, первые сказанные им слова могут вызвать у вас «головокружение от успеха», но помните, что новизна компьютерной речи скоро потеряет свой блеск. И вам придется начать применять изложенные здесь правила компьютерного речевого этикета для «воспитания» вашей действующей системы. Поступая так, вы разовьете у говорящего компьютера некоторые зачатки искусственного интеллекта, что сделает его более прият-

ным для окружающих. В противном случае может оказаться, что в один прекрасный день вы обнаружите громкоговоритель своего компьютера в ближайшем мусорном ящике.

Следуйте правилам

Даже понимая общие идеи изложенных выше правил, вы, возможно, затрудняетесь применять их к конкретной говорящей системе. Здесь и может проявиться ваша творческая жилка. Большинство из перечисленных правил пока находится на самой ранней стадии развития, если иметь в виду их применение к компьютерным синтезаторам речи. Следуя этим правилам, важно все время помнить, что вы, возможно, необъективны в оценке вашего синтезатора речи. Вы с любовью обучили его первым словам и постепенно свыклись с его необычными способностями. Но ведь его речь предназначена и для других ушей. А другие слушатели вполне могут воспринять вашу систему как нечто сверхъестественное, поскольку это — КОМПЬЮТЕР.

У тех, кто впервые слышит говорящий компьютер, обычно возникает желание спросить: «Что он сказал?» В таком случае вам, подавив свою гордость, придется самому спокойно повторить фразу, которую не понял слушатель. И в ответ на это вам, как правило, скажут: «Да, кажется, я именно это и слышал». Неловкое положение часто возникает из-за того, что компьютер вмешивается в разговор людей. Чтобы разрешить эту проблему, компьютер должен перед «выступлением» подать тональный сигнал, звук колокольчика или даже звук, имитирующий прочистку компьютерного «горла». На первый взгляд такое поведение может показаться бестактным, но если вы обратите внимание на обычное разговорное общение людей, то заметите, что любой человек, прежде чем начать говорить, старается каким-то образом привлечь к себе внимание. Обычно это может быть жест или взгляд, показывающий, что речь начинается. Почему же компьютер не должен соблюдать подобные правила? То, о чем я здесь говорю, фактически представляет собой установление связи (нечто вроде рукопожатия) между двумя людьми перед началом разговора. Поскольку предполагается, что компьютер обладает каким-то интеллектом, естественно ожидать от

него соблюдения тех же правил. Существуют и другие способы, которые в зависимости от обстоятельств могут использоваться для «анонсирования» выступления компьютера. Например, если человек внимательно следит за компьютером в ожидании его речи, то перед началом фразы может автоматически включаться сигнальная лампочка. Той же цели может служить, например, тональный звуковой сигнал или звон колокольчика. Не следует, однако, начинать анонсированное сообщение сразу после предупредительного сигнала, так как вполне вероятно, что говорящему в данный момент человеку потребуется две-три секунды, чтобы закончить фразу. Поэтому желательно соблюдать такое хорошо проверенное на опыте правило: речевому сообщению компьютера должны предшествовать предупреждение и вслед за ним пауза продолжительностью две-три секунды. Если вы предполагаете пользоваться речевым выходом компьютерной системы в большом помещении или же в таком месте, откуда компьютер непосредственно не виден, то для предупреждения слушателей о предстоящем сообщении следует использовать тональный сигнал или какое-либо другое звуковое оповещение.

От соблюдения данного правила в большой степени зависит успех компьютерной речевой системы. Если же вы откажетесь от устройства предварительного оповещения, то, по всей вероятности, сообщение компьютера не будет воспринято на фоне разговора окружающих его людей. Для достижения той же цели пригоден еще один способ: повторение фразы несколько раз. Такое «поведение» говорящего компьютера, может, однако, раздражающе действовать на слушателей, если только в его сообщении не содержится чрезвычайно важной и экстренной информации. Например, если компьютер сообщает о пожаре или о взломе жилого либо служебного помещения, то сообщению не только должны предшествовать предупреждающие и привлекающие внимание сигналы, но и само оно должно повторяться очень громко до тех пор, пока человек-оператор не прореагирует на сообщение. Так мы подошли к следующему вопросу: какие же имеются в нашем распоряжении средства взаимодействия с компьютером, пока он говорит фразу или после нее? На данном этапе исследований ответ на этот вопрос, к сожалению, однозначен: никаких. Далее мы рассмотрим несколько примеров использования обрат-

ной связи с системой речевого вывода, которая позволит сделать поведение этой системы более вежливым и и обходительным, как это свойственно человеку.

Можно ли мне сказать это?

Описанные в предыдущем параграфе попытки установить связь компьютера с пользователем носят исключительно пассивный характер, ибо они ограничиваются предупреждением о предстоящем сообщении. Чтобы действи-

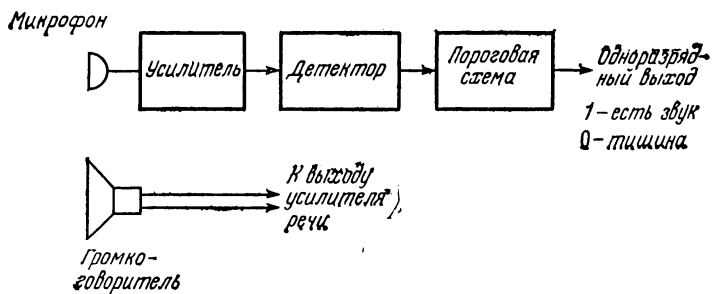


Рис. 4.1. Простейшая схема устройства акустического взаимодействия с компьютером.

тельно обрести способность к речевому взаимодействию с оператором, система должна реагировать на поступающие в нее входные сигналы. Если придерживаться концепции обмена звуковой информацией, то эта обратная связь также должна быть звуковой. Отсюда, однако, не следует, что компьютер должен непременно понимать смысл слов или обладать способностью распознавания речи. Хотя такой воспринимающий звуковые сообщения вход был бы идеальным и позволил бы вам беседовать с компьютером, для осуществления акустического взаимодействия он необязателен. Итак, для обретения некоторого «интеллекта» компьютеру потребуется микрофонный вход.

Простейшая схема звукового входного устройства, которое можно подключить к компьютеру, — это комбинация микрофона с усилителем и несложной цепочки из амплитудного детектора и пороговой схемы. Такое входное устройство «предупредит» компьютер, что в данный момент в данном помещении нет тишины и поэтому приготовленную к выводу фразу следует временно записать в память. На рис. 4.1 показана блок-схема та-

кого простейшего, но вместе с тем действенного устройства звукового входа. Уровень интеллекта этого устройства при взаимодействии с человеком-слушателем определяется программным обеспечением устройства. Как показано на рисунке, микрофон следует установить рядом с громкоговорителем компьютера, а схему, содержащую усилитель, детектор и пороговое устройство, отрегулировать так, чтобы при тишине в помещении на его выходе был логический ноль. Если же на расстоянии слышимости выходного громкоговорителя про-

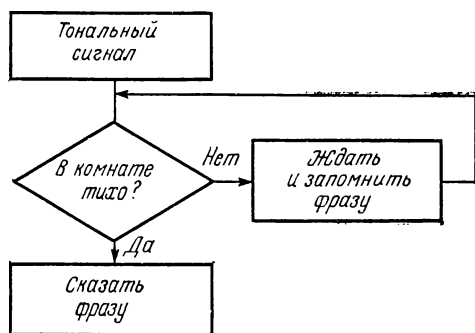


Рис. 4.2. Блок-схема алгоритма «вежливого» говорящего компьютера.

исходит разговор, то порог надо отрегулировать так, чтобы на входе компьютера была логическая единица. Обработывая эту входную информацию в соответствии с заложенной программой, компьютер определяет наиболее подходящий момент для своего высказывания. Структурная схема типичной программы, позволяющей значительно повысить интеллект компьютерной речевой системы, показана на рис. 4.2.

Перед началом вывода речевой информации прежде всего подается тональный сигнал оповещения или речевой сигнал прерывания, предупреждающий о том, что скоро компьютер сделает свое сообщение. Сразу же после этого компьютер делает паузу приблизительно в 1 с, чтобы при помощи схемы звукового взаимодействия определить, наступила ли в помещении тишина. Если разговор продолжается, то приготовленная фраза записывается в память и периодические проверки на наличие тишины производятся до тех пор, пока уровень шу-

мов не станет ниже порогового. Выявив такой момент, компьютер может начать говорить с определенной уверенностью, что речевого столкновения не произойдет. Программная реализация такого режима выдачи сообщений без прерывания других говорящих очень проста, а достижимые при этом результаты могут быть весьма впечатляющими. Несколько более сложная программа



позволит даже проверять уровень шумов в помещении между определенными словами, произнесенными компьютером, и если разговор возобновился, то компьютер может начать свое сообщение снова, когда представится подходящий момент. Для слушателя компьютер, ведущий себя подобным образом, будет похож на говорящего человека, которого прервали и который замолчал, ожидая момента, когда сможет повторно высказать свою мысль.

Иными словами, главное в описанной системе — имитировать поведение говорящего человека, проявляющего определенную степень уважения к тем, кто на-

ходится рядом с ним. На рис. 4.3 показана принципиальная схема, позволяющая реализовать описанный режим работы с акустическим взаимодействием. Эта схема собрана на трех обычных операционных усилителях интегрального типа. С их помощью производится усиление сигнала с выхода микрофона, детектирование

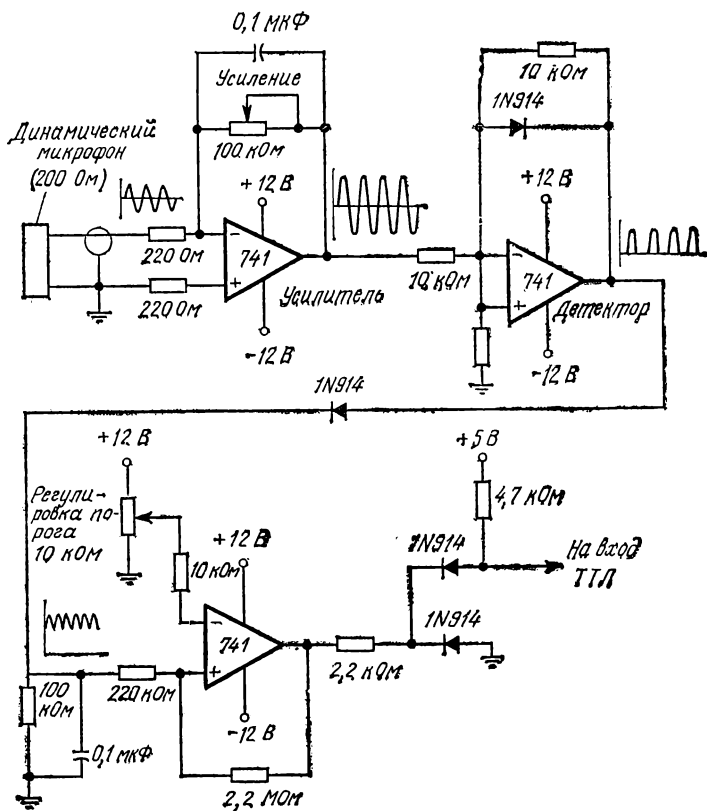


Рис. 4.3. Принципиальная схема микрофонного входа.

ние и пороговое ограничение, позволяющее определить уровень шумов в помещении. Предусмотрены две регулировки: усиления сигнала, получаемого с выхода микрофона, и порогового уровня окружающих шумов. Практически такая схема звукового взаимодействия может сопрягаться всего лишь с одним входом говорящего компьютера — это сигнальная линия, позволяю-

шая определять, тихо или шумно в помещении в данный момент. Если ручку регулировки порога установить приблизительно в среднем положении, то усиление нужно отрегулировать таким образом, чтобы человек, говорящий в помещении спокойным голосом, создавал на входе компьютера сигнал, соответствующий логической единице, т. е. примерно напряжению 5 В. Когда же человек прекращает говорить, уровень сигнала на входе компьютера должен быстро падать до 0 В. Работа данной схемы мало зависит от разброса параметров компонентов схемы и от способа монтажа. При сборке схемы следует лишь обязательно предусмотреть экранирование микрофонного провода и соединений его с входным усилителем, чтобы избежать сетевых наводок и других помех. Собрав схему и подключив ее на вход компьютера, вы получите говорящий компьютер, ведущий себя весьма разумно и нуждающийся в исключительно малых программных средствах для реализации блок-схемы алгоритма (рис. 4.2).

Повышение уровня интеллекта

Разумеется, входное устройство компьютера, способное не только просто реагировать на наличие звуков в помещении, но и распознавать смысл сказанных вами слов, относится к устройствам более высокого класса. Описанию различных конструкций этих приборов — распознавателей речи можно было бы посвятить отдельную книгу такого же объема, как наша. Но не будем погружаться в мир этих сложных устройств и способов их реализации, а предположим просто, что они уже имеются и их можно установить на входе говорящего компьютера. Когда эта колоссальная задача по созданию «разумного» речевого ввода и такого же вывода будет выполнена, область применения использующего их компьютера станет поистине беспредельной. Так, если у вас есть распознаватель речи, способный понимать, скажем, 16 слов, то вы можете вводить в него такие слова, как: да, нет, замолчи, что?, повтори еще раз, громче, тише и т. д. И когда вы начнете программировать речевой вывод с учетом разумного ввода, то, возможно, вам больше не захочется возвращаться к старой клавиатуре. На деле, однако, вам все еще придется вводить слова, которым система пока не обучена.

Занимаясь расширением речевого словаря компьютера и повышением его интеллекта, можно составить программу речевого обучения, которая позволит вводить в компьютер слово через клавиатуру, а затем заставит компьютер методом проб и ошибок учиться правильно произносить это слово, подчиняясь речевым командам человека. Конечно, такая форма взаимодействия с речевым вводом и выводом применима главным образом к системам фонетического синтеза, которые будут описаны ниже. Можно использовать системы и других типов, но при наличии у них в памяти достаточно большого речевого словаря. Результаты речевого взаимодействия с компьютером в процессе его обучения речи могут оказаться впечатляющими как для вас, так и для достаточно подготовленных посторонних наблюдателей. Предлагаю вам поразмышлять над тем, как лучше реализовать подобную речевую систему, обладающую определенным интеллектом. А что касается компьютеров, способности которых отвечали бы поставленным целям, то они, несомненно, появятся в ближайшие пять — десять лет. Аппаратные средства существуют уже сейчас, недостает лишь программного обеспечения.

Управление выводом речи

Даже если вы решили не создавать у говорящего компьютера речевого ввода, вам обязательно нужно предусмотреть какие-то средства воздействия на его речевой вывод. Следует помнить, что речевой вывод информации отличается от вывода с помощью принтера, который не требует к себе внимания, работает автоматически и записи которого можно внимательно прочесть позднее в любое время. Речь же должна быть услышана и понята в самом процессе ее вывода. На деле программа, генерирующая много речевых сообщений, может вызвать раздражение, если только не предусмотреть какого-либо способа понизить громкость речи или совсем выключить речевой вывод. Если вам приходилось бывать в залах электронных игровых автоматов и слышать их бессмысленное бормотание, то вы, конечно, поймете, что обязательно нужно иметь возможность управления речевым выводом. Если же вы еще сомневаетесь в этом, постоитесь полчаса у одного из таких автоматов.

О двух органах управления синтезатора речи, который можно было бы считать «хорошо воспитанным», уже говорилось ранее. Регулировка громкости, входящая в этот обязательный минимум, позволяет оператору по желанию изменять громкость речи. Важно также иметь простой выключатель громкоговорителя. В ряде случаев той же цели служит выключатель питания синтезатора речи. Такой выключатель позволяет мгновенно отключать выход синтезатора, не меняя установленный



уровень громкости. Располагая этими двумя органами управления, оператор может выключить синтезатор или регулировать его громкость от шепота до крика.

Свобода речи?

Создавая говорящие машины, способные привлекать наше внимание своими высказываниями и советами, мы сами становимся объектами воздействия со стороны техники, лишая тем самым себя тишины. Можно привести неплохой пример того, как развитие и распространение техники влияет на окружающую нас среду. Обратите внимание на шум в театре во время спектакля. Вы услышите различный звон и «бип-бипы» часов, кальку-

ляторов и других электронных «штучек», отсчитывающих время. Вообразите, что было бы, если бы все эти часы и калькуляторы стали еще и объявлять время. Наверняка начнется мешанина всех этих сообщений, произносимых говорящими устройствами различными голосами и с разными акцентами. А если еще начнут говорить привычные нам пишущие машинки, будильники, кухонные плиты и камины, холодильники и ав-



томобили, то мы постоянно будем окружены странными голосами, от которых некуда скрыться. Поэтому нужно тщательно учесть возможные последствия добавления речевого вывода таким изделиям или устройствам. Помните, что речевой вывод компьютера полезен, пока соблюдаются определенные правила. Речевой вывод нужен далеко не для каждого применения, и во многих случаях отсутствие визуального вывода информации — большой недостаток. Поскольку, создав говорящий компьютер, мы будем иметь дело с такой реакцией компьютера, от которой нельзя просто отвернуться, следует более ответственно подойти к выбору формы и содержания произносимого компьютером речевого сообщения.

Глава 5

НЕМНОГО ИСТОРИИ

Прежде чем углубиться в технику электронного синтеза речи, читателю, возможно, будет интересно хотя бы вкратце познакомиться с ранними попытками человека реализовать искусственным путем удивительную способность к речи, которой природа наделила лишь людей. В эпоху Древней Греции и Рима дар речи почитался как божественный. Стремясь убедить свою паству в подлинности божественных идолов и статуй богов, жрецы всячески пытались сделать их говорящими. Используемые для этого методы были, разумеется, весьма примитивны, но еще более примитивным было представление о механизме речи: в ход пускались переговорные трубы и рупоры, а также сооружались ловко замаскированные туннели, через которые можно было добраться до рта статуй. Вообразите, что было бы, если бы этих древних «иллюзионистов» вооружить современной техникой. Сколь сильное впечатление произвели бы на несведущих древних людей «говорящие» идолы и боги!

Шли века, но интерес людей к волшебному дару речи не пропал. Предпринимались различные попытки, чаще всего весьма безуспешные, создавать звуки с помощью машин. Некоторые из первых документированных попыток такого рода относятся к эпохе Возрождения. В прежние времена исследования в области создания искусственной речи основывались большей частью на качественных представлениях о механизме речи, а основополагающим принципам речи должного внимания не уделялось. В эпоху Возрождения человек начал пылливо, всеми доступными ему средствами изучать жизненные и физиологические функции организма. Одной из старых проблем, вновь оказавшихся на переднем плане, было исследование физиологии человека и количественное описание механизма речи. Приступив к изучению физиологии речи, ученые начали строить

практические модели говорящих машин, и некоторые из них действительно работали. Если учесть состояние науки в конце XVIII в., то можно считать, что при создании моделей говорящих машин было проявлено немало творческой смекалки.

Говорящие механизмы

Если смотреть с позиций современной технологии, то до наступления эры электроники успехи ученых в области создания говорящих устройств сдерживались отсутствием транзисторов и интегральных схем. Им приходилось полагаться лишь на механические устройства. Подобно аналитической машине (компьютеру) Чарльза Беббиджа, собранной только из механических деталей и занимавшей целую комнату, первые говорящие машины были больше похожи на музыкальные инструменты, чем на сложные системы, которые можно было бы ожидать. Тем не менее все эти механизмы по-своему интересны, и попытки создания механических речевых систем до сих пор продолжают параллельно с развитием методов электронного синтеза речи.

Одна из первых вошедших в историю попыток механической генерации речи была предпринята примерно в 1770 г. в Петербургской академии наук. Сотрудник Санкт-Петербургского университета Кратценштейн успешно решил поставленную задачу — объяснить физиологические различия речи при произнесении пяти гласных звуков и был удостоен ежегодной премии за моделирование и построение ряда акустических резонаторов, имитирующих голосовой тракт человека. Построенное им говорящее устройство, примитивное по сегодняшним меркам, действительно создавало гласные звуки при помощи вибрирующих язычков, возбуждаемых воздушным потоком. Меняя акустические резонаторы и вручную выбирая формантные частоты, при помощи этого устройства можно было, хотя и ограниченно, генерировать речь.

Одновременно с Кратценштейном и независимо от него Вольфганг фон Кемпелен работал над более сложной машиной для генерации связной речи. Он ставил перед собой более сложную задачу, нежели воспроизведение гласных звуков: его целью было получение фраз и слов. Так же как и в устройстве Кратценштейна, в

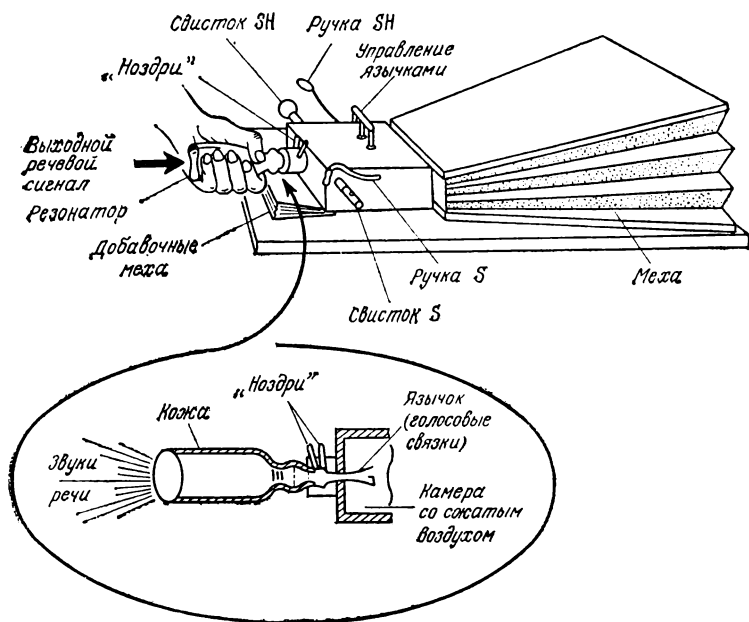


Рис. 5.1. Говорящие меха модели Кемпелена, построенные Уитстоном.

машине Кемпелена имелись меха, которые создавали воздушный поток для возбуждения вибрирующих язычков, игравших роль голосовых связок. Резонаторы ротовой полости имитировались гибкой трубкой из кожи, которая действием руки создавала звуки речи. У этого необычного механического устройства (рис. 5.1) имелись также отверстия, имитирующие носовые резонансные полости, и две ручки, при помощи которых можно было управлять свистками, создававшими фрикативные звуки (человек создает такие звуки посредством зубов и губ). Говорящее устройство было сделано весьма остроумно, но для работы на нем оператору требовалось немало умения и опыта. При «игре» на этом инструменте приходилось выполнять такие же примерно действия, как при игре на самых больших органах. Достоинством удивления, что инженерное искусство того времени позволило создать столь сложную говорящую машину, с помощью которой оператор мог действительно создавать синтезированную речь. Оператор, сидевший сбоку

от говорящей машины, правой рукой приводил в движение меха, одновременно управляя «носовыми» отверстиями и плунжером язычка.левой рукой оператор манипулировал кожаной трубкой, посредством которой осуществлялась артикуляция звонких звуков, приходящих от вибрирующего язычка. Как утверждал фон Кемпелен, его машина производила 19 согласных звуков, которые были вполне различимы. О том, насколько по-разному эти звуки могли быть поняты неподготовленными людьми, мы уже говорили в гл. 4.

Более усовершенствованный вариант «говорящих мехов» фон Кемпелена был позднее построен Чарлзом Уитстоном (создателем известного моста Уитстона); однако до XIX в. исследования в этой области были крайне малочисленны. Машину Кемпелена, реализованную Уитстоном, довелось увидеть молодому экспериментатору из Эдинбурга (Шотландия) Александру Грейаму Беллу. Впечатление, которое произвело на него это говорящее механическое устройство, не только направило его интересы на научное исследование речи, но, вероятно, во многом определило и все дальнейшее развитие науки и техники. Несколько лет спустя, в феврале 1876 г., Белл получил американский патент № 174 465 на принцип и способ реализации устройства, получившего название «телефон».

Увидев «говорящие меха» Уитстона, Александр Белл решил собрать под руководством отца собственную аналогичную модель. С помощью брата Мелвилла Александр построил из дерева, резины и ткани действующую физическую модель голосового тракта человека. Корпус его говорящей машины имел форму головы, а различные органы речи — губы, щеки, язык и голосовые связки — были сделаны из резины и мягкого фетра. Движущимися частями этой модели головы можно было управлять посредством рычагов, а голосовые связки имитировались резиновой мембраной со щелью, через которую проходил воздух. Хотя модель выглядела, по всей видимости, весьма устрашающе, напоминая монстра, Белл утверждал, что она может произносить гласные, носовые звуки и даже (при достаточном опыте оператора) простые связные фразы. Играя впоследствии на говорящей механической голове, Белл вспоминал о тех днях, когда он только начал разбираться в физиологии и механике речи.

Интерес А. Белла к механизму речи во многом был вызван имевшейся у его отца школой речи, и довольно скоро он понял перспективность работ в этой области. Все эти механизмы потенциально были готовы «заговорить», но время для них еще не подошло. Вскоре Белл научил своего терьера Ская, садясь на задние лапы, издавать продолжительное рычание, означавшее требование пищи (которую ему и давали как поощрение). В это время Белл осторожно манипулировал с горлом и ртом собаки, создавая звуки, на которые только был способен ограниченный по своим возможностям голосовой тракт собаки. В результате этих усилий Беллу удалось заставить собаку «произносить» пять звуков, состоящих из гласных *ah* и *oo*, морфем *ta* и *ga* и дифтонга *oi*. Рассказывают, что с помощью этого весьма ограниченного «словаря» и нескольких фунтов пищи Белл сумел заставить своего любимца произносить фразу «How are you, Grandmama?» (Как поживаете, бабушка?) Говорили также, что иногда пес с блеском в глазах делал попытки самостоятельно сказать эту фразу, но у него не получалось ничего, кроме «гррр». Старания юного Белла научить своего пса «говорить» хотя и не увенчались успехом, тем не менее послужили основой для его первых манипуляций с механической говорящей головой.

Двадцатое столетие

В нашем веке сохранился интерес к синтезу речи, основанному на тех механических методах, которые разрабатывались в прошлом. Однако с зарождением электроники были предприняты попытки построить электрические аналоги речевой системы. Так, в 1922 г. Дж. Стюарт сообщил в журнале *Nature Magazine* о построенной им электрической модели голосового тракта. Эта довольно примитивная схема (рис. 5.2) включала электрический зуммер для моделирования голосовых связок и пару индуктивно-емкостных резонаторов для моделирования резонансов горла и ротовой полости. Модель Стюарта, способная в лучшем случае генерировать две формантные частоты, давала речь, состоящую главным образом из гласных звуков. Изменением емкостей, индуктивностей и сопротивлений компонентов схемы можно было плавно переходить от одного глас-

ного звука к другому, получая таким образом некоторые дифтонги.

Сложные электрические компоненты в то время были сравнительно малодоступны, и поэтому продолжались поиски в области механических систем синтеза речи. В 1937 г. Р. Риш продемонстрировал свое механическое говорящее устройство, которое, как и другие механические устройства такого назначения, больше напоминало музыкальный инструмент. Это устройство, в своей звуковой части повторявшее голосовой тракт человека и выполненное в основном из резины и металла, содержало также клавиши, подобные тем, какие мы привыкли видеть на музыкальных трубах (рис. 5.3). Опытный опе-

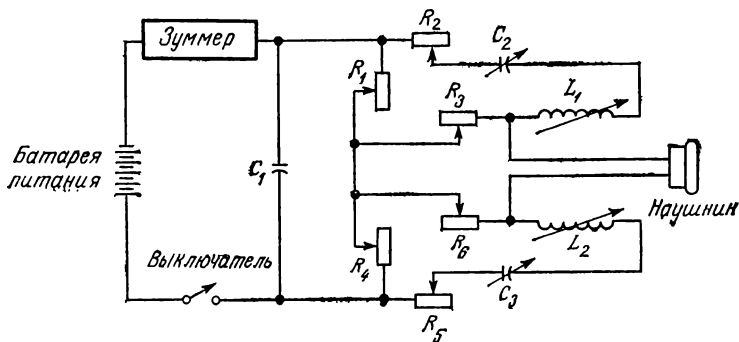


Рис. 5.2. Электрическая модель голосового тракта, построенная Стюартом.

ратор извлекал из этого говорящего устройства речь довольно хорошего качества. Управляя (двумя руками) 10 клавишами, оператор мог осуществлять артикуляцию звуков. Имелась возможность управлять движением почти каждой подвижной части голосового тракта человека. В качестве примера наиболее артикулированной речи этого устройства обычно приводили слово *cigarette* (сигарета).

Почти одновременно Риш работал также над созданием одного из первых электрических синтезаторов речи, способных создавать связную речь. Как и его механическое говорящее устройство, электрическая говорящая система Риша, построенная им с помощью Дадли и Уоткинса, демонстрировалась на международных выставках в Нью-Йорке в 1939 г. и в Сан-Франциско

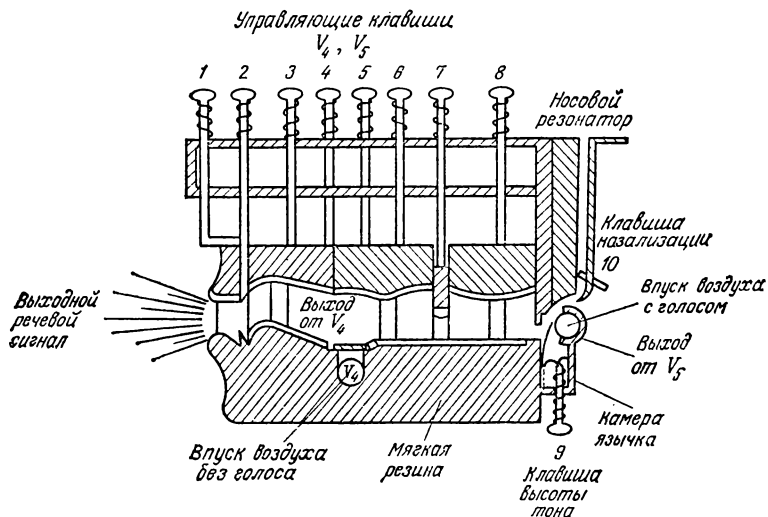


Рис. 5.3. Механическое говорящее устройство Риша.

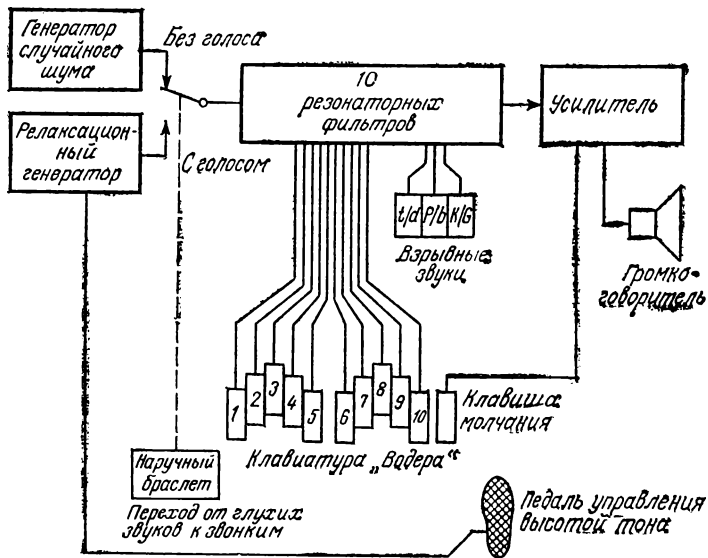


Рис. 5.4. Синтезатор речи «Водер», демонстрировавшийся на международных выставках в Нью-Йорке (1939) и Сан-Франциско (1940).

в 1940 г. Схема этой машины, получившей название «Водер», приведена на рис. 5.4. Она содержит электрические генераторы для моделирования звонкой и глухой речи, за которыми следуют 10 включенных параллельно полосовых фильтров, охватывающих весь спектр звуковых частот. Управление каждым из этих фильтров осуществляется нажатием на клавиши (типа клавишей фортепьяно), которые связаны с фильтрами через переменные резисторы (потенциометры). Для перехода от звонкой речи к глухой оператор пользуется наручным браслетом, а высота тона изменяется ножной педалью. Были предусмотрены также клавиши для генерации взрывных звуков и пауз.

Посетители международных выставок видели оператора, который играл на говорящей машине, подобно опытному органисту или пианисту. Подготовка оператора занимала не менее года. Получаемая речь, как сообщалось, была весьма высококачественной.

По своей конструкции «Водер» в принципе очень похож на современные системы фонетического синтеза речи. Клавиатура управления в последних заменена, однако, выходами (портами) компьютера, благодаря чему управление этими системами существенно упростилось. После «Водера» дальнейшие исследования в области синтеза речи были связаны с использованием компьютеров и в основном велись по трем направлениям: восстановленная речь, фонетически синтезированная речь и математически синтезированная речь.

Поскольку это уже техника синтеза речи сегодняшнего дня, мы и перейдем теперь к ее описанию.

Глава 6

СОВРЕМЕННЫЕ МЕТОДЫ СИНТЕЗА РЕЧИ

Итак, мы подошли к главной теме нашей книги. Познакомившись в предыдущих главах с основами генерации и синтеза речи, займемся теперь современными электронными методами синтеза речи. Эта глава вследствие сложности изложенного в ней материала — одна из самых больших в нашей книге. Но, изучив ее, читатель получит достаточно полное представление о методах синтеза речи и средствах, которыми они реализуются. Не исключено, что некоторые, пока еще не вполне ясные и менее известные методы генерации искусственной речи не нашли здесь отражения, но основные компьютерные методы синтеза, существующие на сегодняшний день, рассмотрены весьма подробно.

Существуют три основных метода генерации речи; перечислим их в порядке возрастания сложности:

1. Кодирование сигналов речи с последующим прямым восстановлением речи.
2. Фонетический синтез речи с помощью аналоговых формантных синтезаторов.
3. Математическое восстановление речи в частотной области (например, синтезаторы с линейным предиктивным кодированием, ЛПК).

Чтобы лучше понять возможности и ограничения каждого из этих методов, рассмотрим их более подробно. Познакомившись с методами синтеза речи, читатель сможет ориентироваться в особенностях различных синтезаторов. И если ваше внимание привлечет какой-либо конкретный способ, то возможно, что с него и начнется ваш путь к занятиям синтезом речи. Данная глава поможет также найти наиболее эффективный способ вывода речевой информации из компьютера. Итак, мы вступаем в одну из самых быстро развивающихся областей современной электронной техники — область синтеза речи.

Триединство методов синтеза речи

Как мы уже говорили, существуют три основных метода синтеза речи. А почему именно три? Почему, скажем, не два, четыре, восемь или десять?

На этот вопрос вряд ли можно ответить просто и однозначно. Пытаясь найти ответ, мы можем сравнить методы синтеза речи, например, с видами транспортных средств. Так, говоря о транспортных средствах, можно назвать три основных типа легковых автомобилей: обычный легковой автомобиль, спортивный автомобиль и вездеход. Все эти автомобили выполняют одну и ту же задачу — перемещение человека в пространстве. Основные же различия между ними связаны с тем, как они это делают, каковы их стоимость и скорость. Каждое из этих транспортных средств обладает своими достоинствами и недостатками, но все они всегда дают один и тот же конечный результат: вы выезжаете из пункта *А* и приезжаете в пункт *Б*. Аналогия с методами синтеза речи может показаться здесь слишком далекой, тем не менее каждый из трех методов генерации искусственной речи служит одной и той же цели — выводу информации из компьютера в виде речи. Способы же, которыми достигается эта цель (кодирование-восстановление формы сигналов, фонетический формантный синтез или математическое восстановление речи), различаются скоростью выполнения синтеза речи, стоимостью аппаратуры и быстродействием компьютера. И каждая из таких возможных реализаций синтезатора обеспечивает различные уровни разборчивости и качества синтезированной речи. Иначе говоря, каждому из трех методов синтеза речи присущи свои плюсы и минусы: один дает речь с лучшей артикуляцией, но его реализация обходится гораздо дороже; другой достаточно дешев, но его речевые способности весьма ограничены. И мы должны учитывать это, когда останавливаем свой выбор на том или ином методе.

«Тахометр» синтеза речи

На рис. 6.1 изображена диаграмма, позволяющая визуально сравнить три названных типа систем. Этот «тахометр» синтеза речи показывает, как соотносятся скорости работы компьютера, необходимые для генераторов

искусственной речи каждой из трех основных категорий. Сами по себе эти скорости для большинства современных персональных компьютеров вполне реальны, но скорость выборки информации (выражаемая в битах в секунду, бит/с) в неявном виде связана непосредственно с объемом памяти, необходимым для хранения речи. И если ввести некий относительный показатель, позволяющий сравнивать три названных метода синтеза речи по необходимому для каждого из них объему памяти,

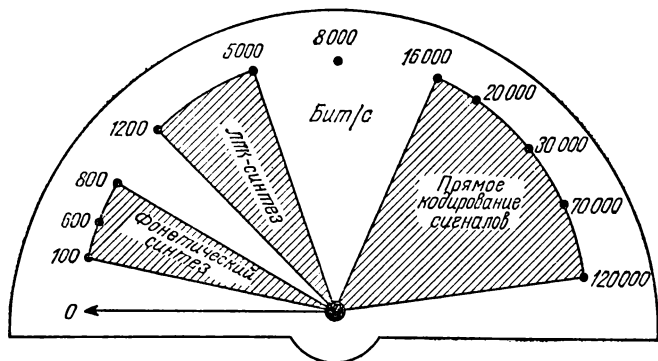


Рис. 6.1. «Тахометр» синтеза речи.

то идея подобного «тахометра» приобретает более глубокий смысл.

Предположим, например, что у базового персонального компьютера байт памяти содержит восемь бит (у большинства компьютеров это именно так). Поскольку на «тахометре» речи показаны скорости в бит/с, простым делением скорости на восемь можно найти объем памяти, требуемый для каждой секунды речи. Допустим теперь, что мы хотим сравнить методы синтеза речи по эффективности хранения такого слова, как *hello*. Время, затрачиваемое на произнесение этого слова, зависит от манеры речи каждого человека, но приблизительно его можно считать равным 0,3 с. Как видно из табл. 6.1, где три основных метода синтеза речи сопоставляются по объему памяти, необходимой для произнесения данного слова, их эффективность в этом отношении очень различна. На рис. 6.1 для каждого из методов показана область скоростей, в которой данный конкретный метод может использоваться. Нижняя

Таблица 6.1. Сравнение методов синтеза речи

Метод синтеза	Скорость, бит/с	Объем памяти для хранения слова HELLO, байт	Полное количество речи, которое может хранить компьютер с памятью 48К
Фонетический	100—800	4—30	1 ч 4 мин—8 мин
Линейное предиктивное кодирование	1200—5000	45—188	5 мин 20 с—1 мин 17 с
Непосредственное кодирование	16 000—120 000	600—4500	24 с—3,2 с

граничная скорость выборки в каждом случае соответствует относительно низкому качеству речи и связанному с ним ухудшению разборчивости. Верхняя граничная скорость в каждом методе соответствует системе с очень хорошей артикуляцией, качество речи которой приближается к естественной речи человека. Объем памяти, требуемый для произнесения слова *hello*, меняется в пределах от 4 байт (сравнительно низкокачественный фонетический метод синтеза; см. третью колонку таблицы) до 4500 байт (метод непосредственного кодирования сигналов с высоким качеством речи). Разумеется, нижние граничные значения скоростей для каждого метода синтеза выбраны так, чтобы уровень разборчивости был примерно одинаков, а верхние граничные скорости соответствуют в каждом случае почти идеальному качеству речи.

Запоминание и хранение речи

В табл. 6.1 для нас наиболее важна последняя колонка, где указана полная длительность речевого сообщения, которое можно хранить в типичном персональном компьютере с памятью 48К байт ($K=1024$). Максимальная длительность (1 ч 4 мин), как видим, соответствует фонетическому методу и позволяет хранить не только довольно большой словарь, но и другие программы, используемые для работы говорящей системы. А на другом конце интервала длительностей речевого сообщения находится наименее эффективный метод синтеза речи — непосредственное кодирование сигналов речи, при работе с которым в памяти компьютера емкостью 48К

можно хранить речевое сообщение длительностью лишь 3,2 с.

Не следует забывать, что эти требования к памяти относятся только к речи. А в памяти должен также содержаться словарь, который желательно использовать. Так, если вы хотите иметь словарь из 100 слов, предназначенный для составления произвольных предложений и фраз, то, учитывая, что для произнесения каждого словарного слова потребуется приблизительно 0,5 с, для хранения словаря потребуется объем памяти, соответствующий длительности сообщения по меньшей мере 50 с.

Обратившись снова к табл. 6.1, мы увидим, что при использовании метода непосредственного кодирования наш словарь может содержать максимум 48 слов (24 с), и при этом совсем не остается места для другого программирования фраз речи. Проблему можно решить, если хранить речь непосредственно в виде фраз без возможности произвольной выборки слов, но с возможностью выборки каждой из записанных фраз в нужные моменты. Это позволит ограничиться гораздо меньшим словарем, но вместе с тем выделить часть памяти для хранения программ управления речью.

Приведенные результаты сравнения методов речевого вывода нельзя рассматривать как основу для вынесения окончательного решения относительно выбора того или иного метода. Эти данные служат лишь иллюстрацией возможностей и ограничений каждого из них. Выбор предстоит сделать вам самим. В гл. 7 даны характеристики многих выпускаемых в настоящее время систем с указанием их приблизительных цен. Знакомясь с этими системами, вспомните о «тахометре» синтеза речи, определяющем требования к объему памяти систем, основанных на различных методах синтеза речи.

Искусственная (синтезированная) речь

Сравнивая три метода генерации искусственной речи, важно прежде всего определить понятие «искусственная (синтезированная) речь». Существуют по крайней мере два основных критерия, по которым можно судить, что

речь, воспроизводимая компьютером, в самом деле является искусственной.

Во-первых, если компьютер генерирует действительно «синтетическую» речь, то произносимые им слова не должны быть словами, предварительно сказанными человеком и затем введенными в память компьютера. Наиболее уместна здесь аналогия с «магнитофоном», роль ленты в котором выполняет цифровая или электронная память. Действительно, если слова или фразы, которые произносит компьютер, были предварительно «наговорены» человеком (лично вами или кем-то на предприятии, изготовившем интегральные схемы), то в таком случае речь компьютера следует считать восстановленной. В речевом «тахометре» представлены два метода получения восстановленной речи: метод непосредственного кодирования сигналов и метод ЛПК-синтеза. Эти методы существенно различаются по способам реализации, но в обоих из них обычно требуется предварительно наговоренный словарь, который в дальнейшем используется при восстановлении речи. Таким образом, единственным методом получения действительно синтезированной речи оказывается метод фонетического синтеза.

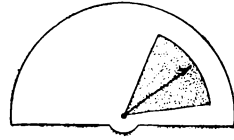
Разумеется, синтезированная при этом речь по своему звучанию менее похожа на речь человека, чем в двух других случаях. По существу, на современном уровне развития техники речь, приемлемую для нашего слуха, дают только методы с восстановлением речи. Настоящая же «синтетическая» речь фонемного синтезатора по своему звучанию неестественна. Другими словами, если вы хотите иметь действительно синтетическую речь, то должны быть готовы к тому, что речь вашего компьютера будет звучать «механически». Если же вы предпочтете использовать один из методов восстановления речи, то звучание вашей системы будет приятным для слуха и близким к нормальному произношению человека.

Второй критерий, определяющий искусственность речи, связан с основной теоремой из области сбора информации — теоремой выборки или отсчетов. Много лет назад Генри Найквист из научно-исследовательской фирмы «Белл телефон лабораторис» установил, что для адекватной выборки и запоминания типичного аналогового сигнала частота выборки должна по меньшей

мере вдвое превышать наивысшую частотную компоненту в обрабатываемом сигнале¹. Эту теорему можно проиллюстрировать следующим примером. Если нужно делать выборки речевого сигнала, наивысшие частотные компоненты которого имеют частоту около 4 кГц, то частота выборок должна быть не менее 8 кГц. Кроме того, если мы хотим получить сравнительно понятную восстановленную речь, при каждой выборке нужно запоминать амплитуду колебаний, что потребует по меньшей мере 4 бит.

Таким образом, умножив 4 бит на частоту выборок 8 кГц, мы получим битовую частоту выборок, равную приблизительно 32 000 бит/с. Поскольку обычно речь отличается некоторой избыточностью (информации) и минимально вразумительную речь можно получить, когда частоты наивысших частотных компонентов речи составляют лишь 2 кГц, нижний предел битовой частоты выборок должен составлять 16 000 бит, что и показано на диаграмме «тахометра» речи (рис. 6.1). Если же мы какими-то средствами восстанавливаем речь такого же качества при частотах выборки ниже расчетного значения (16 000 бит/с), то получаем речь, которая перед этим не дискретизировалась. Отсюда второе возможное определение искусственной речи, согласно которому методы фонетического и ЛПК-синтеза дают действительно искусственную речь. По существу, любую систему речевого вывода, работающую со скоростью менее 16 000 бит/с, можно считать создающей искусственную речь. Но независимо от принятой терминологии каждый из названных методов синтеза речи со всеми присущими ему достоинствами и недостатками предназначен для получения речевого вывода информации из компьютера. С этих позиций приведенные выше доводы становятся несущественными. Целесообразно выбрать тот из методов, который больше подходит для данного компьютера и лучше соответствует возможностям пользователя. Рассмотрим теперь более подробно способы реализации каждого из трех основных методов синтеза речи.

¹ В отечественной литературе данный критерий известен как теорема Котельникова. — *Прим. ред.*



Метод непосредственного кодирования-восстановления

Этот простейший метод генерации речи более других похож на цифровую запись. Выше мы ввели понятие выборки аналоговых сигналов с целью последующего восстановления речи. Правильное представление о механизме выборки очень важно для понимания методов кодирования-восстановления речи. Чтобы пояснить, как происходит выборка аналоговых сигналов, приведем такой пример. Рассмотрим выборку визуального (зрительного) сигнала. Здесь все происходит так же, как и в случае звукового аналогового сигнала, с той лишь разницей, что теперь мы имеем дело с визуальным сигналом.

Прежде всего выберем «сцену», на которой разворачиваются действия, — скажем, экран телевизора или проходящую поблизости скоростную магистраль с мчащимися по ней автомобилями. Теперь посмотрите на эту сцену и закройте глаза. Если далее вы будете быстро (с частотой примерно раз в секунду, т. е. 1 Гц) открывать и закрывать глаза, подобно тому как это делает обтюратор кинокамеры, то без труда сможете наблюдать за происходящим на сцене. Продолжайте мигать, открывая глаза на несколько секунд, и вы убедитесь, что вам действительно удастся следить за событиями, протекающими перед вашим взором. То, что вы делали, соответствует визуальной выборке движения (аналогового сигнала). Если же какое-то действие произойдет «мгновенно» — пока глаза были закрыты, хотя мигание происходит с частотой 1 Гц, — то вы этого просто не заметите. Уменьшим теперь частоту мигания примерно до одного раза в десять секунд, т. е., закрыв глаза, вы считаете до 10, затем на мгновение открываете их и снова закрываете на 10 с и т. д. Повторив этот опыт в течение одной-двух минут, вы поймете, что теперь вам не удастся пронаблюдать всю последовательность собы-

тий, происходящих на «сцене». По существу, вы уменьшили частоту выборки ниже предельной величины, определяемой критерием Найквиста, соблюдение которого необходимо для адекватного наблюдения и описания аналогового сигнала. При мигании с более высокой частотой у вас не должно возникать никаких трудностей при наблюдении за сценой и описании происходящих на ней действий. Рассмотренный пример полностью аналогичен процессу выборки слышимой речи, с тем лишь исключением, что в нашем примере выборка производится визуальным путем, а в случае речи — электронным. Различие между этими двумя случаями очень мало. В процессе преобразования звуков речи микрофоном генерируется электрический сигнал, который содержит как медленно, так и весьма быстро меняющиеся компоненты. Если бы мы стали делать выборки из этого сигнала с относительно низкой частотой, то самое большое, что нам удалось бы установить, — это то, что в сигнале действительно существуют медленно меняющиеся компоненты. Быстро меняющиеся компоненты сигнала ускользнули бы от нас при таком методе исследования. Если же мы начнем осуществлять выборку электрических сигналов микрофона со все возрастающей частотой, то будем обнаруживать все более высокочастотные компоненты сигнала. В целях иллюстрации на рис. 6.2 сравниваются выборки, осуществляемые с разными частотами.

На самом верхнем графике рисунка показан аналоговый сигнал, из которого должны быть получены выборки. По своему виду он весьма похож на сложные колебания, с которыми приходится иметь дело при обработке речи. Нетрудно заметить, что в составе этого сигнала имеются как медленно, так и быстро меняющиеся составляющие. Пользуясь терминами спектрального анализа, можно сказать, что медленно меняющиеся составляющие представляют собой низкочастотные компоненты, а быстро меняющиеся (центральная часть графика) — более высокочастотные компоненты. Если из этого сигнала, общая длительность которого равна 5 мс, производить выборку с частотой одна выборка в миллисекунду, то мы получим амплитудно-модулированный импульсный сигнал (рис. 6.2, *a*), который дает сравнительно мало информации об исходном колебании. Если попытаться восстановить форму первоначального

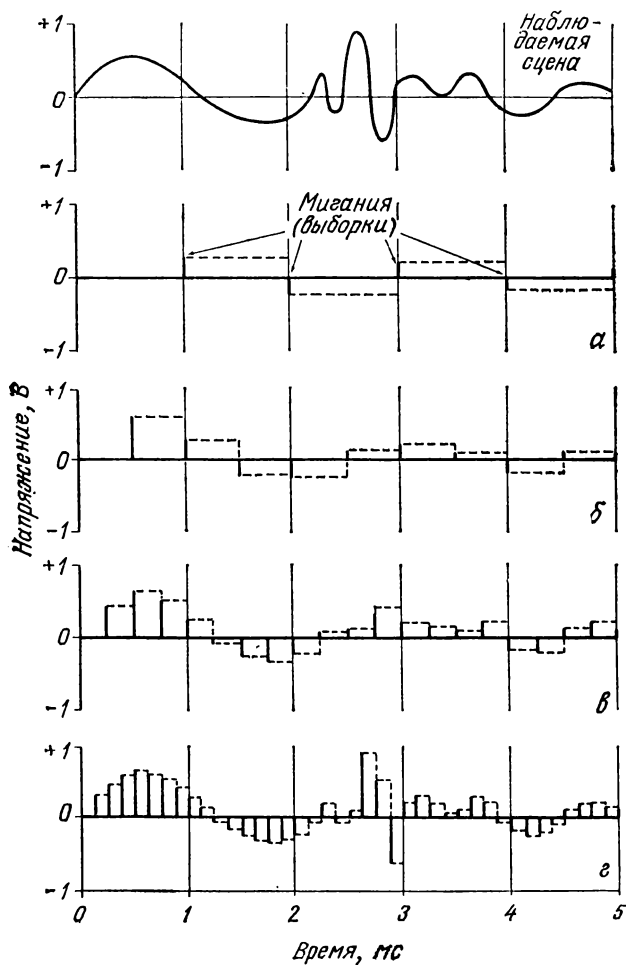


Рис. 6.2. Увеличение частоты выборки сигнала.

сигнала по этим пяти импульсам, то это позволит нам судить лишь о самых низкочастотных компонентах. Удвоив частоту выборок, мы получим сигнал (рис. 6.2, б), который дает несколько больше информации об исходном сигнале. Еще и еще удваивая частоту выборок (рис. 6.2, в, г), мы получим амплитудно-модулированные импульсные сигналы, более точно воспроизводящие исходный сигнал. Процесс с удвоением частоты выборки

можно проводить многократно — до тех пор, пока полученная форма сигнала не станет настолько точно соответствовать исходному колебанию, что дальнейшее повышение частоты выборки ничего более не добавит к ней. Поскольку наивысшая частотная составляющая в исходном (входном аналоговом) сигнале уместается во временном интервале несколько менее 0,5 мс, для адекватной характеристики сигнала мы должны, согласно теореме Найквиста, производить выборки с частотой четыре раза в миллисекунду. Этот случай представлен на рис. 6.2, в. В самом деле, показанный здесь сигнал содержит в себе большую часть высокочастотных выбросов исходного сигнала. Для большей точности воспроизведения сигнала частоту выборок можно еще удвоить, доведя до восьми раз в миллисекунду (рис. 6.2, г). В этом случае первоначальный сигнал отобра-

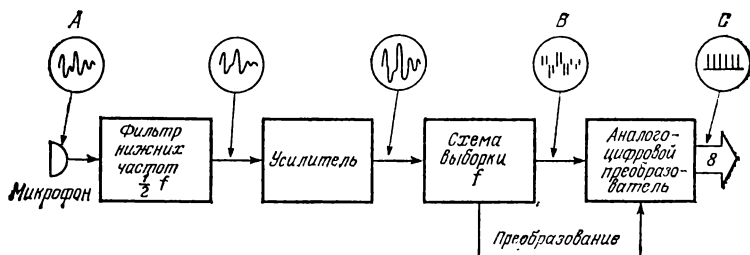


Рис. 6.3. Система выборки речевого сигнала.

жается амплитудно-модулированным импульсным сигналом почти без искажения.

Процесс выборки аналогового сигнала, как было показано, дает нам последовательность импульсов, высота которых соответствует амплитуде сигнала в момент выборки, а длительность — ширине импульсов выборки. Информацию такого вида весьма трудно вводить в память компьютера. Для этого нужно сначала преобразовать изменения амплитуды импульсов к цифровому виду. Блок-схема, показанная на рис. 6.3, иллюстрирует, как это можно сделать.

В системе выборки (дискретизации) речи, представленной на рис. 6.3, сигналы речи, поступающие с микрофона, прежде всего пропускаются через фильтр нижних частот, что позволяет удалить из них составляющие с частотой, превышающей половину частоты выборки.

Это нужно, чтобы не допустить проникновения посторонних сигналов, возникающих, когда выборка производится на частоте, менее чем в два раза превышающей частоты сигнала (вспомните теорему Найквиста). Затем сигнал подается на усилитель, где его уровень повышается настолько, чтобы стало возможным подать его на вход собственно схемы выборки. Здесь-то и выполняется процесс выборки, проанализированный на рис. 6.2. На выходе схемы выборки получается последовательность импульсов, промодулированных по амплитуде в соответствии с величиной аналогового сигнала

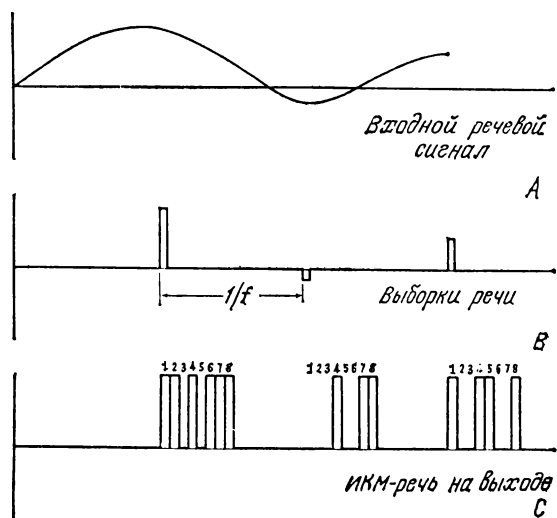


Рис. 6.4. Преобразование амплитудно-модулированного сигнала в импульсно-кодовый.

в каждый из моментов выборки. Наконец, последний блок системы представляет собой аналого-цифровой преобразователь (АЦП). Заметим, что этот блок имеет одну входную линию, на которую поступают выборки аналогового сигнала, и восемь выходных линий. Существуют и более сложные АЦП с 16 или 18 битами (разрядами) на выходе; однако мы ограничимся величиной выходного сигнала 8 бит. Этот 8-битовый выходной сигнал АЦП используется для ввода данных в компьютер. В АЦП аналоговый входной сигнал преобразуется в цифровую форму и в виде сигнала импульсно-кодовой

модуляции (ИКМ) поступает с выхода АЦП в память компьютера.

Чтобы проиллюстрировать процесс импульсно-кодовой модуляции, на рис. 6.4 показаны более детально формы сигналов, обозначенных на рис. 6.3 буквами *A*, *B* и *C*. Мы рассмотрели, как осуществляется преобразование аналогового речевого сигнала в форму, пригодную для ввода в память компьютера. При этом сначала производится выборка, а лишь затем преобразование аналогового сигнала в ИКМ-сигнал. Важная особенность сигнала *C*, представленного на рис. 6.4, состоит в том, что импульсы этой последовательности постоянны по амплитуде и, следовательно, пригодны для ввода в память компьютера. Каждый из 8 бит ИКМ-сигнала речи может подаваться на последовательный вход компьютера для записи в его память. Вместе с тем эти 8 параллельных бит могут передаваться через параллельный вход компьютера и записываться непосредственно байт за байтом. Данные, поступившие на вход компьютера, вводятся в его память в виде цифровой записи амплитуд исходного речевого сигнала. Речевой информацией, хранящейся в памяти, можно манипулировать, используя для более плотного размещения информации программные средства. Подобные манипуляции, которые будут рассмотрены ниже, позволяют снизить требования, предъявляемые к памяти.

Итак, мы знаем, как осуществляются выборка и запись аналоговой речи в память нашего компьютера. Теперь можно попытаться создать очень простой синтезатор речи с непосредственным кодированием. Примерная схема такого синтезатора приведена на рис. 6.5. В верхней части рисунка мы видим уже знакомую нам цепочку блоков, осуществляющих выборку речевого сигнала после микрофона (см. рис. 6.4). При частоте выборки 6 кГц эта схема дает на выходе 6000 8-битовых байтов в секунду. Ранее мы отметили, что для произнесения слова *hello* требуется примерно 0,3 с. При частоте выборки 6000 Гц число выборок для этого слова составит около 2000 байт. Сюда следует также включить короткие паузы в начале и в конце слова. Прямоугольник (обведен штриховой линией) в центре рис. 6.5 — это компьютер с 8-битовым параллельным портом ввода и 8-битовым параллельным портом вывода. Оба порта, конечно, могут быть реализованы в виде одного, и тогда

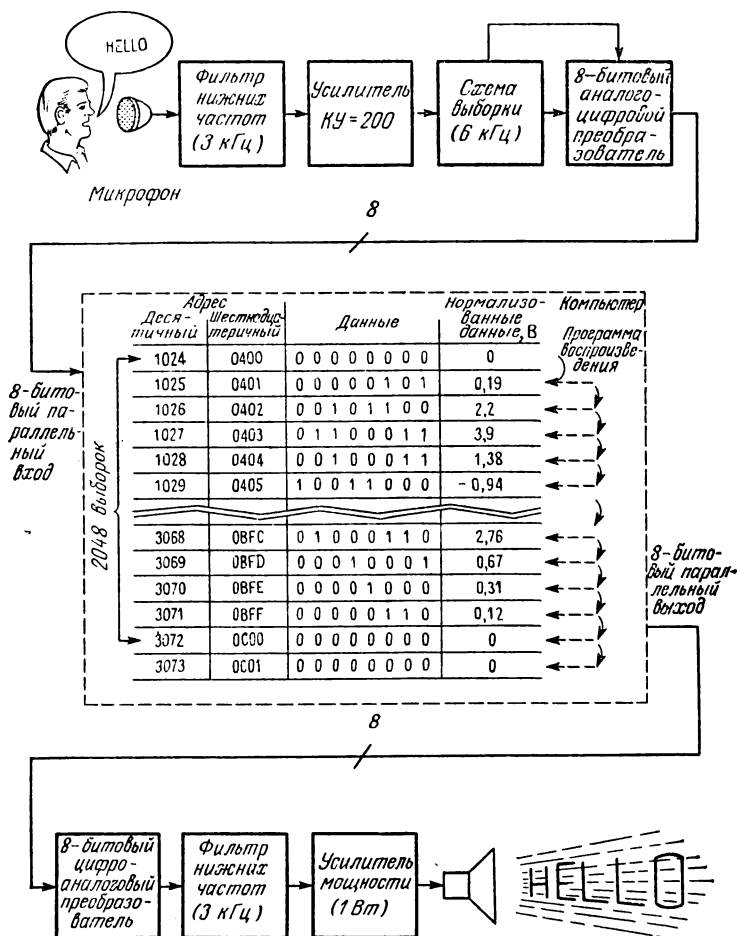


Рис. 6.5. Упрощенная схема синтезатора речи с непосредственным кодированием.

речь воспроизводится через тот же самый двунаправленный порт, через который она была сначала записана.

В процессе сбора данных компьютером его резидентная программа переносит информацию из 8-разрядного АЦП и последовательно записывает ее в память. После произнесения слова процесс выборки заканчивается, а в памяти оказывается записанным цифровое представление слова *hello*. Как видно из рисунка, эта информация

занимает адресное пространство от ячейки 1024 до примерно 3073. В данном случае информация представлена в двоичном коде при полном диапазоне изменений напряжения ± 5 В. Колонка «нормализованные данные» содержит величины напряжений, эквивалентных записанным входным сигналам. Эта нормализованная информация в вольтах не имеет существенного значения для компьютера и показана здесь только для того, чтобы проиллюстрировать эквивалентность аналого-цифрового процесса выборки. В этой точке схемы синтезатора речи мы располагаем записанной в цифровом виде речью, и эту запись можно сохранять на диске или касете для последующего воспроизведения.

Воспроизводящая часть схемы синтезатора представлена внизу рис. 6.5. Программа воспроизведения речи, хранящаяся в памяти компьютера, — это простая индексирующая программа, которая последовательными шагами просматривает записанную ранее информацию и выводит ее побайтно на 8-разрядный цифро-аналоговый преобразователь (ЦАП). Если ЦАП совместим с входным АЦП, то нормализация данных и их преобразование в напряжения не вызывает беспокойства. Преобразователи следят друг за другом и дают на выходе тот же самый сигнал напряжения, какой был подвергнут выборке на входе. Фактически это означает, что если у обоих преобразователей напряжения полный размах по напряжению составляет ± 5 В, то выходной сигнал будет в точности соответствовать сигналу, поступившему на вход синтезатора.

Выход 8-разрядного ЦАП подсоединен к фильтру нижних частот (3 кГц), который отфильтровывает нежелательные высокочастотные компоненты, возникающие при восстановлении сигнала. На рис. 6.2, г им соответствуют прямоугольные верхушки импульсов. Затем выходной сигнал фильтра усиливается усилителем мощности и поступает на громкоговоритель. Речь, которую мы услышим из громкоговорителя, должна в точности повторять исходный речевой сигнал на входе синтезатора, за исключением небольших потерь в высокочастотной части спектра речевого сигнала, вызванных фильтром нижних частот. Иначе говоря, полученная речь по своему звучанию соответствует голосу говорящего и имеет такие же модуляцию и тональность, как и входной речевой сигнал.

В схеме синтезатора речи (рис. 6.5) не отражены лишь способ запоминания речи, используемый в компьютере при приеме сигнала, и программа воспроизведения в процессе синтеза речи. Вследствие довольно высокой частоты выборки (6 кГц) у компьютера остается примерно 17 мкс для приема и записи в память каждого байта информации. Такое требование к быстродействию компьютера исключает возможность применения программ записи и воспроизведения речи, составленных на языке Бейсик или каком-либо другом языке высокого уровня. Поэтому приемлем лишь один способ записи и

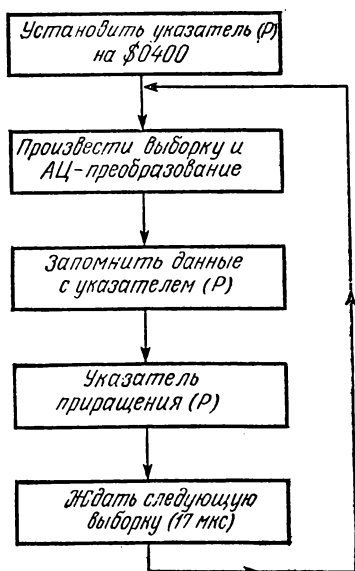


Рис. 6.6. Структурная схема возможной программы записи речи.

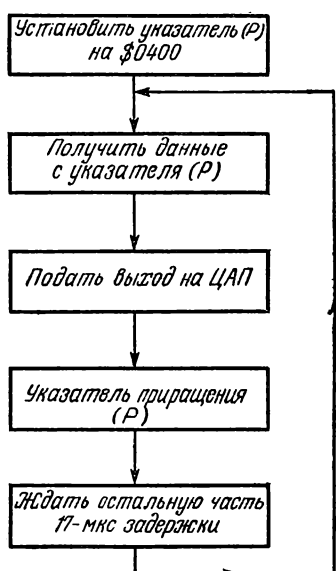


Рис. 6.7. Структурная схема возможной программы воспроизведения речи.

воспроизведения синтезированной речи — программы, составленные на ассемблере. Насколько прост процесс кодирования речи, показывает блок-схема одной из возможных программ записи речи, приведенная на рис. 6.6. После того как речь введена в память, можно запускать программу воспроизведения первоначально закодированной речи. Простоту этой программы иллюстрирует рис. 6.7, где дана блок-схема программы на ас-

семблере, являющаяся обратной по отношению к блок-схеме рис. 6.6. Единственное важное сходство между этими двумя программами состоит в том, что программа вывода речи имеет ту же самую эквивалентную частоту выборки, что и программа ввода. Другими словами, если выборка входного сигнала производится со скоростью 6000 раз в секунду, то, естественно, желательно, чтобы закодированная информация поступала на ЦАП также со скоростью 6000 раз в секунду. Если эти скорости различны, то звучание речи на выходе синтезатора будет отличаться от исходного — подобный эффект получается, если пластинку, записанную со скоростью вращения 45 об/мин, проигрывать со скоростью вращения $33\frac{1}{3}$ об/мин или наоборот.

Заменяя последний блок на рис. 6.7 таким, который будет уменьшать или увеличивать время задержки, можно получить на выходе компьютера довольно необычные звуковые эффекты. Если вообще исключить этот блок, то цикл выполнения программы займет менее 17 мкс и по своему звучанию речь на выходе будет напоминать механический голос Буратино. А если задержку в данном блоке сделать больше 17 мкс, то речь станет похожей на ворчание. При таком экспериментировании с блоком задержки фактически происходит сжатие и растяжение во времени поступающей на вход речи. При этом создается впечатление, что нормальный голос стал звучать быстрее или медленнее, сохраняя при этом все первоначально содержащиеся в нем частоты. В действительности же, увеличивая время задержки воспроизведения с 17 до 34 мкс, мы вдвое понижаем все частотные составляющие голоса по сравнению с их исходными значениями. Поэтому, если при испытании такого простого синтезатора с кодированием и восстановлением речи получается речь искаженной тональности, то следует обратиться к блоку задержки (рис. 6.7) и, изменяя задержку, добиться правильного звучания. Запись и воспроизведение речи и других звуков с помощью этой простой системы доставит вам немало удовольствия. При творческом подходе можно попытаться объединить программу речевого вывода с игровыми и другими программами, записанными на языке Бейсик (напомним, однако, что программа воспроизведения речи должна быть записана на ассемблере).

Дельта-модуляция

Рассмотренный нами пример преобразования речи в цифровой код с последующим декодированием — отнюдь не единственный возможный способ реализации данного метода синтеза речи. На практике используются другие способы, позволяющие более эффективно преобразовывать речь в цифровой код и хранить сигналы. Один из таких способов, получивший довольно широкое распространение в телефонии и в некоторых из существующих микросхем синтезаторов, называется *дельта-модуляцией*. Главное отличие этого способа от описан-

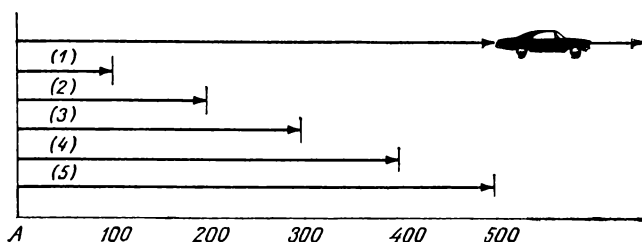


Рис. 6.8. Пример взятия абсолютных отсчетов.

ного ранее (см. рис. 6.5) заключается в том, что он основан на относительных изменениях амплитуды, а не на ее абсолютных величинах. Чтобы пояснить способ дельта-модуляции, обратимся снова к уже использованной аналогии, иллюстрирующей процесс выборки на примере визуальных сигналов наблюдения за автомобилем, движущимся по дороге. Пользуясь рассмотренным ранее методом параллельного кодирования, мы можем описать движение автомобиля из пункта *A* в пункт *B* следующим образом (по выборкам):

1. Автомобиль находится в 100 м от пункта *A*.
2. Автомобиль находится в 200 м от пункта *A*.
3. Автомобиль находится в 300 м от пункта *A*.
4. Автомобиль находится в 400 м от пункта *A*.
5. Автомобиль находится в 500 м от пункта *A* и т. д.

Таким образом, в любой момент мы можем сказать, где находится автомобиль относительно начальной точки его движения. Этот способ отсчета абсолютных значений меняющейся величины показан на рис. 6.8. Если попытаться теперь представить движение автомобиля

во времени в виде аналогового сигнала, то для точности аналогии достаточно заменить в отсчетах метры на вольты.

Посмотрим теперь, как меняется процесс взятия выборок при использовании дельта-модуляции, когда мы интересуемся лишь относительными изменениями, получаемыми в каждой конкретной выборке по сравнению с предыдущей. Следуя такому подходу, мы теряем информацию об абсолютных величинах сигнала при каждой выборке, но при последовательных выборках это не имеет значения. Исключая же избыточную информацию, мы увеличиваем возможности хранения информа-

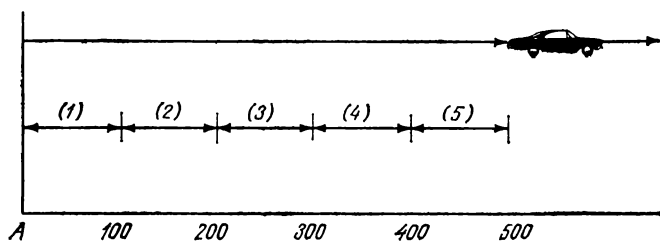


Рис. 6.9. Кодирование методом дельта-модуляции.

ции о наблюдаемом аналоговом сигнале. При использовании дельта-модуляции движение автомобиля можно представить следующими отсчетами:

1. Автомобиль прошел 100 м от пункта А.
2. Автомобиль прошел 100 м от точки предыдущей выборки.
3. Автомобиль прошел 100 м от точки предыдущей выборки.
4. Автомобиль прошел 100 м от точки предыдущей выборки.
5. Автомобиль прошел 100 м от точки предыдущей выборки и т. д.

Если бы мы захотели продолжить наблюдение за автомобилем на протяжении сотен километров, то при первом способе отсчета (рис. 6.8) нам пришлось бы хранить в памяти числа, скажем, до 150 000 м и более. Каждый из отсчетов (выборок) давал бы нам точное значение расстояния, на которое автомобиль удался от пункта А. Но с каждой выборкой величина, которую приходится хранить в памяти, возрастает. А как видно из рис. 6.9, при кодировании с дельта-мо-

дуляцией то же самое перемещение автомобиля описывается через разности (дельта) расстояний между каждыми соседними моментами выборки. Следовательно, даже при расстояниях в сотни километров наибольшее число, которое нужно хранить в памяти для описания поездки, составляет 100 м (отсчитываемые от точки последней выборки). Естественно, что для запоминания числа 100 в памяти требуется гораздо меньше места, чем для запоминания числа 150 000. Таким образом, на этом примере ясно, насколько должна увеличиваться

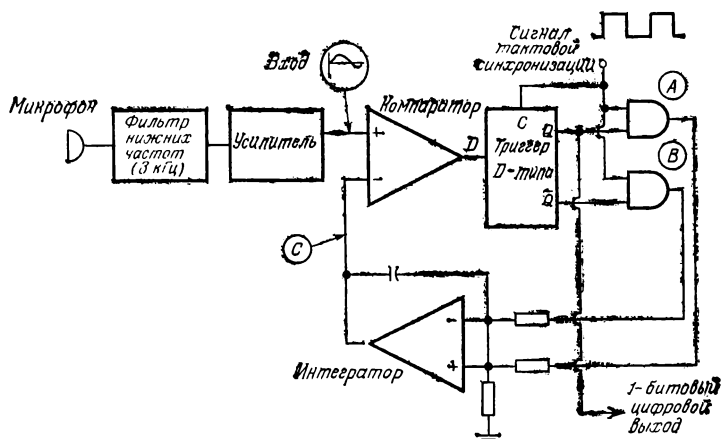


Рис. 6.10. Схема дельта-модулятора.

эффективность цифрового кодирования и запоминания речи при использовании дельта-модуляции.

Описанный процесс дельта-модуляции часто называют кодированием последовательных приращений, так как в нем кодируется лишь изменение сигнала между соседними выборками. Но как с помощью электроники реализовать кодирование речевого сигнала, поступающего с микрофона? Одна из возможных схем цифрового дельта-модулятора представлена на рис. 6.10. Сигнал речи, принятый микрофоном, фильтруется и усиливается так же, как в предыдущих случаях. Затем сигнал подвергается дельта-модуляции; для этого используется простая схема, состоящая из аналогового компаратора и контура обратной связи, содержащего аналоговый интегратор. В контуре обратной связи име-

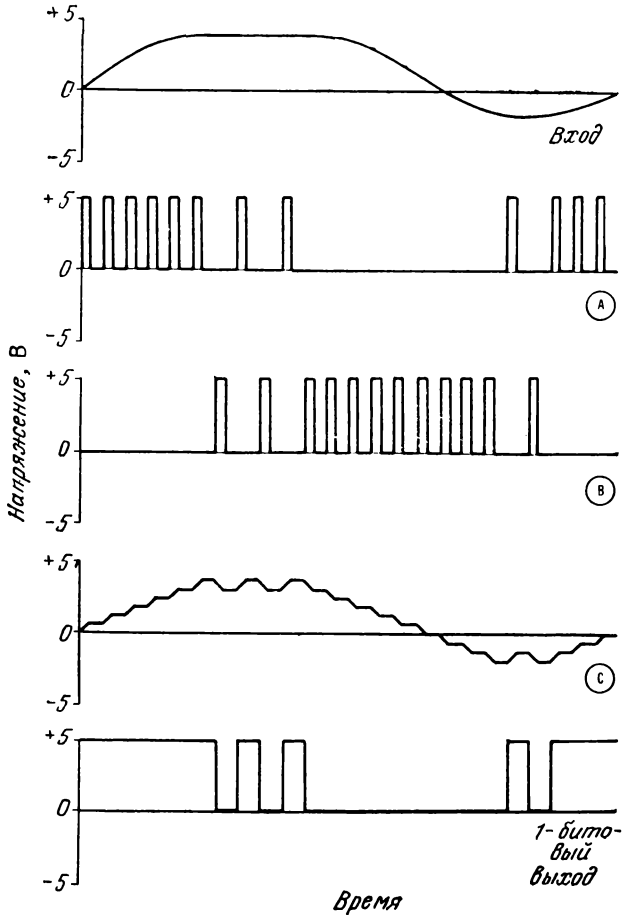


Рис. 6.11. Сигналы при кодировании методом дельта-модуляции.

ется также триггер D-типа, синхронизирующий дельта-компоненту сигнала (изменение наклона) с тактовой частотой аналого-цифрового преобразования. Сигналы, с которыми имеют дело при кодировании методом дельта-модуляции, показаны на рис. 6.11. Отметим, что входной сигнал (верхний график) представляет собой колебание, где участки нарастания и спада чередуются с участками относительного постоянства (стабильности) сигнала. Стробированные выходные сигна-

лы триггера D-типа (рис. 6.10) представлены на рис. 6.11 графиками *A* и *B* — это импульсы неизменной амплитуды 5В, которые поступают на интегратор, создавая на его выходе сигнал *C*. Изменения сигнала отсчитываются в каждый из периодов выборки *T* (график *A*). Если входной сигнал продолжает нарастать, то о его увеличении свидетельствуют импульсы в точке *A* схемы. Импульсы же в точке *B* указывают на уменьшение

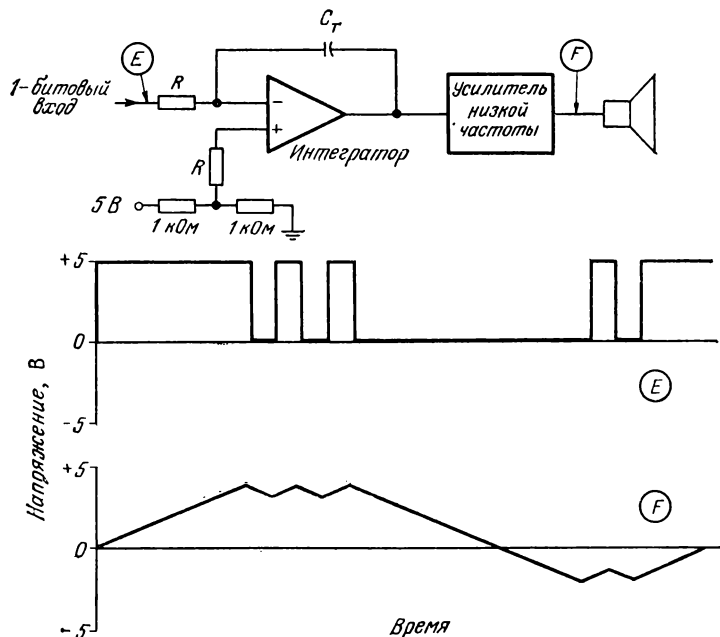


Рис. 6.12. Схема воспроизведения речи, закодированной методом дельта-модуляции.

входного сигнала. В те интервалы времени, когда величина входного сигнала остается неизменной, дельта-модулятор начинает генерировать поочередно импульсы *A* и *B*. Выходной сигнал в точке *C* — это результат сравнения сигнала обратной связи с сигналом на входе интегратора.

На нижнем графике рис. 6.11 показан цифровой сигнал, который эта схема подает на компьютер. Этот двухуровневый сигнал записывается в память ком-

пьютера с той же самой частотой синхронизации, какая использовалась в схеме кодирования с дельта-модуляцией (рис. 6.10).

При кодировании с дельта-модуляцией для получения адекватного воспроизведения речи достаточно даже еще более простой схемы. На рис. 6.12 показан простой интегратор, весьма похожий на тот, что использовался в схеме кодирования. На вход E интегратора от компьютера поступает двухуровневый логический сигнал, следующий с частотой, равной частоте выборок. Постоянную времени интегратора RC_T выбирают так, чтобы она соответствовала продолжительности интегрирования в схеме выборок. Если на вход схемы воспроизведения речи поступает логическая единица, то интегратор дает кривую постоянного наклона одного знака. Когда же на вход поступает логический нуль, наклон кривой выходного напряжения интегратора меняет знак. В результате получается сигнал такого типа, как показано на графике F . Этот сигнал есть не что иное, как воспроизведение исходного входного сигнала, закодированного с использованием дельта-модуляции.

Как видим, полученная форма сигнала довольно хорошо соответствует исходной, отличаясь лишь небольшими искажениями. Степень искажений, создаваемых в процессе дельта-модуляции, обычно определяется одним из видов искажений при кодировании — так называемой перегрузкой по крутизне. Если крутизна входного сигнала меняется быстрее, чем могут проследить интеграторы, то выходной сигнал просто «не успевает» за изменениями входного. Эту типичную для процесса линейной дельта-модуляции трудность можно преодолеть, применив некую модификацию этого вида модуляции, называемую адаптивной импульсно-кодовой дельта-модуляцией. Этот модифицированный метод использован в некоторых микросхемах синтезаторов речи, в том числе в системе синтеза речи Диджитокер фирмы «Нэшнл семикондактор».

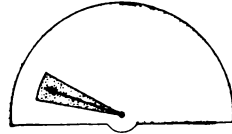
Система на основе адаптивной импульсно-кодовой дельта-модуляции (АИКДМ) весьма сходна с аналогичной линейной системой; единственное различие между ними заключается в том, что система АИКДМ хранит в своей памяти не только направление изменения крутизны с момента последней выборки, но и величину

этого изменения с того же момента. Это позволяет сохранить все преимущества предыдущей системы и одновременно исключить проблему перегрузки по крутизне. Проще говоря, вместо использования единственного бита, указывающего, является ли сигнал нарастающим или спадающим, можно брать четыре бита, которые будут характеризовать как нарастающий, так и спадающий сигнал одним из восьми возможных уровней в зависимости от конкретных условий изменения. Это дает возможность более точно описывать изменение сигнала между выборками, а также снимает проблему перегрузки по крутизне, с которой приходится сталкиваться при линейной дельта-модуляции.

Помимо описанных используются и другие методы синтеза речи путем кодирования сигнала; однако созданные на их основе системы по существу представляют собой лишь модификации описанных. Синтез речи по методу кодирования-восстановления речевого сигнала считается одним из простейших по способам реализации; он несложен также и для понимания. Как уже указывалось, этот метод сходен с цифровой записью. Способы запоминания речи могут быть, однако, различными, как было показано при рассмотрении «тахометра» речи; чтобы добиться высокой точности воспроизведения речи, необходима очень большая скорость выборки.

Из всех методов цифрового кодирования речевого сигнала наиболее эффективен метод АИКДМ. Представляя изменения входного сигнала от выборки к выборке значениями его наклона и амплитуды, удается сократить число необходимых битов памяти по каждой выборке. А как известно, основная проблема при разработке синтезаторов речи сводится к тому, как сократить объем памяти, необходимый для хранения одной произнесенной фразы.

Принципиально иной метод синтеза речи воплощен в синтезаторе с фонемным возбуждением, в котором используется синтез формантных частот. Внимательно познакомившись с действием этой системы, мы обнаружим значительные различия в названных методах электронного синтеза речи.



Аналоговый синтез формантных частот речи

Описанный метод кодирования-восстановления речевого сигнала можно назвать «фотографированием» речи. В то же время синтез речи по методу аналоговых формант, или формантных частот, лучше характеризовать как живописное изображение того же речевого рисунка. Кодированный речевой сигнал должен существовать еще до того, как его произнес синтезатор, точно так же, как сами по себе существуют в природе пейзаж или предметы, которые мы хотим сфотографировать. В отличие от этого метод синтеза формантных частот представляет собой способ выражения той же речи искусственным путем, аналогично тому, как художник рисует наблюдаемую картину. Особенности каждого из этих двух методов синтеза речи в значительной мере совпадают с особенностями соответствующего изобразительного средства: фотография дает точное воспроизведение оригинала, тогда как живописное полотно, хотя и достаточно близко к оригиналу, никогда не совпадет с ним точно. Однако в зависимости от поставленной цели каждый из методов с известным приближением описывает исходную обстановку или речь.

Существует много методов реализации формантного синтеза речи. Но основные функциональные операции для генерации речи при разных способах формантного синтеза в принципе одинаковы. Все они основываются на детальном знании фонем и фонетическом расчленении речи (см. гл. 3). Познакомившись немного подробнее с количественными характеристиками фонетической речи, мы сможем заложить основы для понимания сущности формантного метода синтеза речи. Обратившись к описаниям спектрограмм речи, показанным на рис. 2.10 и 2.11 (гл. 2), вспомним, что темные частотные полосы визуализированной речи называются формантными частотами. Тот факт, что у нас имеется возможность визуальными и электронными способами распознавать в человеческой речи эти меняющиеся частоты, поз-

Таблица 6.2. Типичные формантные частоты для ряда фонем гласных звуков

Фонема	Как в слове	F_1	F_2	F_3
ee	feet	250	2300	3000
i	hid	375	2150	2800
eh	head	550	1950	2600
ae	had	700	1800	2550
ah	tot	775	1100	2500
aw	talk	575	900	2450
u	took	425	1000	2400
oo	tool	275	850	2400

воляет сформулировать метод синтеза искусственной речи, основанный на одновременной генерации необходимого числа формантных частот.

Если теперь посмотреть на табл. 6.2 и вспомнить приведенное в гл. 3 описание фонем гласных звуков, то мы обнаружим в этой таблице фонемы некоторых знакомых нам звуков, но здесь они охарактеризованы более подробно. Для каждого звука в таблице даны три основные формантные частоты, которые наблюдаются в спектрограммах соответствующих фонем, произносимых «средним» мужским голосом. Эти частоты, F_1 , F_2 и F_3 , можно различать визуально на спектрограмме каждой произносимой гласной. Поскольку каждая из этих гласных является «статической», частоты остаются стабильными на протяжении всего времени их произнесения. Теперь, призвав на помощь немного воображения, нетрудно представить электронную схему (рис. 6.13) с тремя параллельными полосовыми фильтрами (их средние частоты равны формантным частотам F_1 , F_2 , F_3), которая возбуждается задающим генератором с выходным сигналом, аналогичным импульсу, формируемому в голосовой щели. Как ни проста эта схема, она может служить основой фонемного синтезатора гласных при условии, что формантные частоты F_1 , F_2 и F_3 регулируются в пределах, указанных в табл. 6.2.

Средняя частота каждого из параллельных полосовых фильтров регулируется таким образом, чтобы в процессе речи синтезатора достигалось совпадение с эквивалентными резонансами полостей рта и горла. Резуль-

тирующая речь, получаемая суммированием выходов этих трех фильтров, весьма близка к речи человека. В спектрограмме выходного сигнала этой схемы содержатся три формантные частоты, идентичные формантным частотам в спектрограмме речи человека, произносящего те же гласные. Если в схеме на рис. 6.13 четыре ручки регулировки заменить компьютерным управлением высотой основного тона голоса и тремя формант-

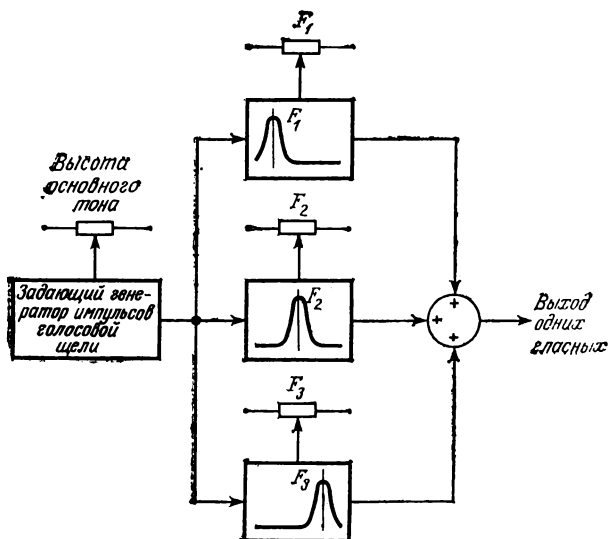


Рис. 6.13. Базовая схема формантного синтезатора речи.

ными частотами, то можно программировать компьютер, который по нашим командам будет произносить требуемые гласные. То, что мы описали, и есть основа системы синтеза речи по методу формантных частот.

Единственная трудность в создании такого синтезатора формантных частот связана теперь с тем, что его речевые способности строго ограничены — он может генерировать только гласные звуки. Хотя построенная таким образом схема весьма интересна, поскольку наглядно демонстрирует метод электронной генерации фонетической речи, при работе с ней мы очень быстро исчерпаем весь запас слов и фраз, которые она способна произносить. Чтобы расширить диапазон возможностей формантного синтезатора, необходимо ввести источник

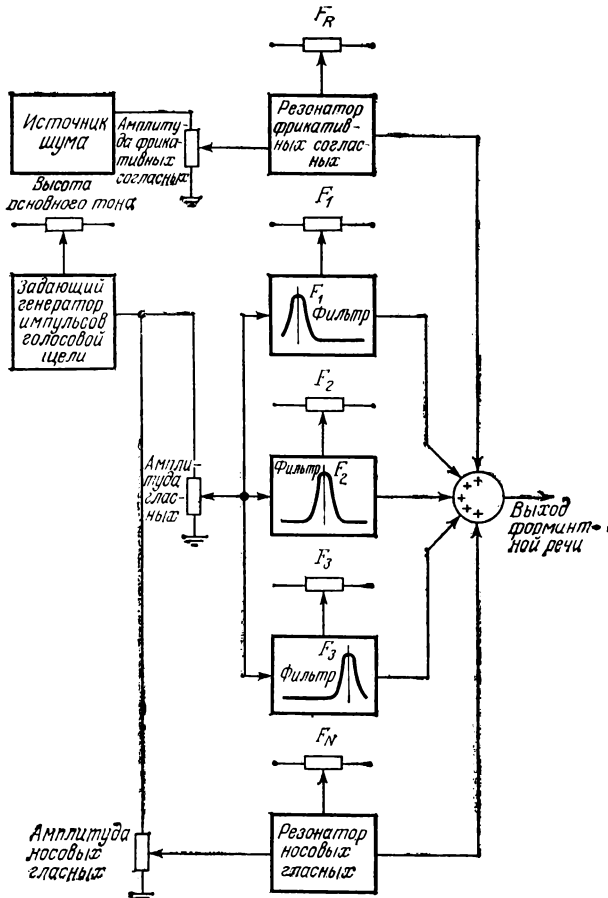


Рис. 6.14. Полная схема формантного синтезатора речи.

шума для формирования фриктивных и взрывных согласных, а также аналог носовой резонансной полости, имитирующий носовые согласные. Блок-схема этого расширенного формантного синтезатора приведена на рис. 6.14.

Когда мы введем в новую схему синтезатора весь набор средств, необходимых для полного синтеза речи, она, конечно, усложнится по сравнению с предшествующей схемой, но не слишком сильно. Заметим, что вместо прежних четырех органов управления и регулиров-

ки теперь стало девять. Три из них служат для управления амплитудами фрикативных, гласных и носовых звуков, один — для регулирования высоты тона, а пять остальных — для регулирования частот различных резонансов. Если построить систему, которая имела бы пять ручек управления на передней панели, и если бы ваши руки обладали исключительной подвижностью, то посредством этой системы вы могли бы воспроизводить речь, подобно тому как это делали творцы первых «говорящих» машин. К счастью, мы располагаем таким преимуществом, как возможность компьютеризации управления органами регулировки, так что компьютер окажет нам неоценимую помощь в генерации синтезированной речи. Заменяв каждый из девяти регулировочных потенциометров в схеме, изображенной на рис. 6.14, соединением с параллельным компьютерным портом, мы можем приступить к программированию синтезатора. Позже в процессе синтеза можно производить все необходимые регулировки со скоростью, которая требуется, чтобы добиться приемлемого приближения к нормальной речи человека.

Теперь предположим, что мы в состоянии надлежащим образом синтезировать изменения формантной частоты, необходимые для воспроизведения дифтонгов и «нестатических» фонем, путем перенастройки управления в полной схеме синтезатора с частотой 100 Гц; тогда, учитывая, что каждое управляющее слово состоит из 8-битового байта, мы можем управлять синтезатором при параллельной передаче управляющих слов со скоростью 900 байт/с. Но это еще не скорость выдачи используемых нами фонем, а всего лишь скорость передачи данных в самом синтезаторе, необходимая для текущих регулировок фильтра, высоты голосового тона и амплитуд при воспроизведении каждой отдельной фонемы. Данная информация, как правило, хранится в специальной управляющей таблице в памяти компьютера отдельно для каждой фонемы и ее вариаций — аллофонов. Она вызывается в виде последовательности, которая определяется входной последовательностью фонем в речевом выходе управляющей программы. Например, по каждой фонеме, которая вводится в компьютер, компьютерное управление синтезатора может потребовать до 30 полных перенастроек системы фильтров. Это означает следующее: чтобы компьютер

выдавал схеме управления формантного синтезатора по 900 байт в секунду, в программу опроса справочной таблицы, которая управляет синтезатором, требуется вводить всего по 30 байт в секунду. Это соответствует тому, что выдача данных конечному пользователю производится со скоростью примерно 240 бит/с.

В то время как указанные выше значения скорости перенастройки фильтров могут меняться от системы к системе, зависимость между числом воспроизводимых фонем и размерами управляющей таблицы остается практически неизменной. Естественно, чем больше число обращений к справочной таблице по каждой фонеме, тем большей плавностью будет отличаться синтетическая речь и тем ближе она будет по звучанию к естественной артикуляции человеческой речи. Однако не следует надеяться, что речь подобной системы когда-нибудь будет звучать так, что ее не отличишь от человеческой. В ее синтезированной речи всегда прослушивается голос робота, что, согласно представлениям большинства людей, и должно быть характерно для компьютерной речи. Если вы захотите продемонстрировать подобную систему посторонним людям, то эффект будет весьма впечатляющим, но им, возможно, будет нелегко понять, что именно говорит система. Однако совершенно исключено, чтобы кто-нибудь мог принять синтезированную речь формантного синтезатора с фонемным возбуждением за предварительно наговоренную или восстановленную речь. Выход подобной системы всегда будет компьютерной речью.

Теперь посмотрим, как компьютер мог бы управлять типичным формантным синтезатором речи с фонемным возбуждением. Одна из возможных схем управления такой говорящей системы приведена на рис. 6.15. Слева на этой схеме изображен компьютер. Чтобы управлять такой системой, компьютер должен обладать следующими основными средствами:

1. Набором 8-битовых выходных портов для реализации различных функций управления.

2. Одним 8-битовым входным портом, служащим для уведомления компьютера о готовности синтезатора к приему следующего кадра данных для текущей перенастройки синтезатора.

Если проследить ход различных управляющих линий, соединяющих на этой схеме различные блоки, то в

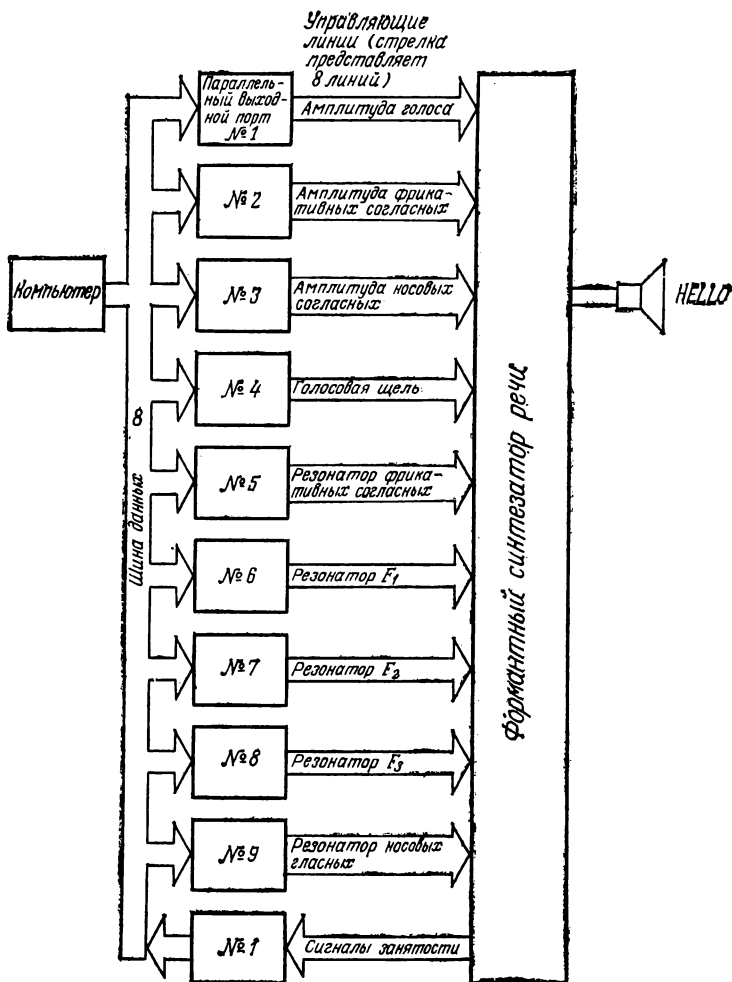


Рис. 6.15. Схема формантного синтезатора речи с компьютерным управлением.

них можно узнать те же средства управления формантным синтезатором, которые представлены на схеме рис. 6.14. Сигналы занятости, поступающие по управляющим линиям (показаны в нижней части рис. 6.15), сообщают компьютеру, что формантный синтезатор речи закончил произносить очередной кадр данных. Термин «кадр данных» выбран не случайно: он говорит

о том, что управление синтезатором через отдельные порты производится не в произвольном порядке. Напротив, компьютер обновляет управляющие данные на всех девяти параллельных портах одновременно в соответствии с подлежащей воспроизведению фонемой или ее частью, затем ожидает окончания произнесения этого кадра (типичное время ожидания составляет 33 мс), после чего возобновляет цикл перенастройки управления для завершения произнесения фонемы.

На рис. 6.16 приведена функциональная блок-схема компьютерной программы для управления типичным формантным синтезатором. Если предположить, что синтезатор должен произнести слово *hello*, то программа в первую очередь должна расчленить вводимое слово на составляющие его фонемы. Ради простоты допустим, что в первом приближении для произнесения слова *hello* нужно воспроизвести всего четыре фонемы: *h*, *eh*, *l* и *o*. Если попытаться произнести это слово с помощью такой цепочки фонем, то оно будет звучать очень грубо и, вполне возможно, на слух окажется недостаточно понятным. Большинство фонемных синтезаторов располагает возможностью воспроизведения такого разнообразия фонем и их аллофонов для генерации речи, что для более точного произнесения слова *hello* могут использовать семь-восемь фонемных звуков.

Подойдя далее к той части блок-схемы программы, которая связана с процессом вывода речи, мы заметим, что компьютер обращается к управляющей таблице по каждой фонеме, которую необходимо произнести. Если фонема является «статической», то фонемная управляющая таблица (имеющая ширину 9 байт и длину, равную 30 кадрам обновляющей информации) может содержать один и тот же обновляющий кадр, который повторяется 30 раз. Фонемам «нестатического» типа, или дифтонгам, будет соответствовать управляющая таблица, которая содержит набор меняющихся кадров обновляющейся информации, что связано с динамическими свойствами произносимого звука. Когда компьютер получает доступ к каждому из этих обновляющих кадров и выводит их на выходные порты, сопряженные с синтезатором речи, система начинает говорить. Когда же синтезатор речи завершает работу с очередным обновляющим кадром по произносимой фонеме, он сигнализирует компьютеру, что произнесение закон-

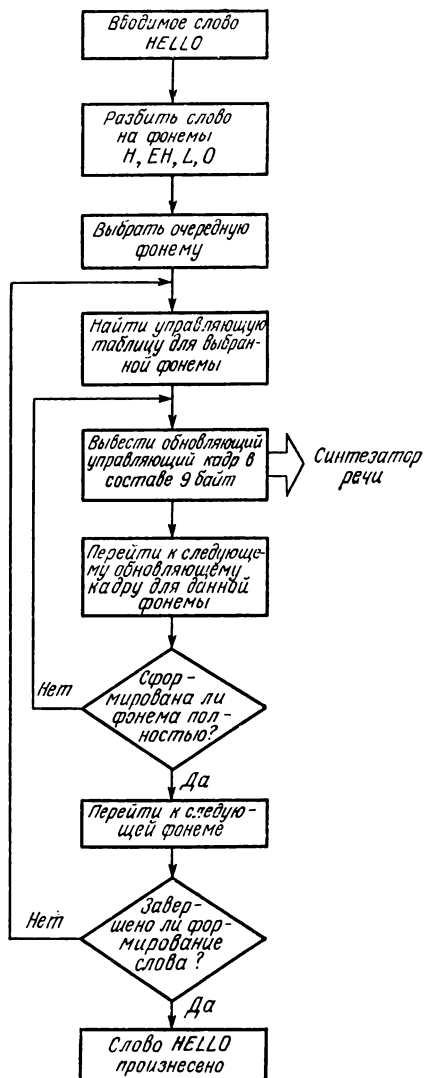


Рис. 6.16. Блок-схема программы для управления формантным синтезатором речи.

чено, после чего компьютер обращается к следующей позиции управляющей таблицы, чтобы получить очередной обновляющий кадр для перенастройки синтезатора.

Завершив воспроизведение очередной фонемы, компьютерная программа переходит к следующей фонеме произносимого слова и повторяет описанный цикл. Процесс продолжается до тех пор, пока не будут произнесены все фонемы расчлененного слова. После этого система оказывается готовой к приему следующего слова из произносимой фразы.

На этом, собственно, и можно закончить описание синтезатора, работающего по методу синтеза формантных частот. Довольно просто, не так ли? Или все же весьма сложно? Если при изучении блок-схемы программы вас поставит в тупик блок «Найти управляющую таблицу для выбранной фонемы», то не особенно огорчайтесь. Изготовители подобных систем формантного синтеза достаточно предусмотрительны и поставляют говорящие периферийные устройства, снабженные фонемными управляющими таблицами, которые являются резидентными средствами, т. е. размещены в постоянной памяти. В некоторых устройствах, например в формантном синтезаторе Тайп'н ток¹, функция расчленения на фонемы даже включена в состав речевой системы. Сказанное означает, что для правильного произнесения слова достаточно просто ввести его в систему. Все логические операции, указанные на блок-схеме рис. 6.16, выполняются в самом периферийном устройстве для синтеза речи. Эта особенность характерна также и для ряда фонемных и аллофонных синтезаторов, построенных по принципу линейного предиктивного кодирования (ЛПК), которые описаны в следующем разделе. Потоки информации в фонетическом ЛПК-синтезаторе обычно имеют такой же вид, и его отличие от аналогового формантного синтезатора состоит лишь в том, что здесь используются не аналоговые, а цифровые управляемые речевые фильтры.

Благодаря особенностям своей конструкции синтезаторы с фонемным и аллофонным возбуждением обладают способностью к воспроизведению неограниченного словаря. Поскольку синтезаторы этого типа произносят последовательности звуков, которые составляют естественную речь человека, все, что необходимо сделать, чтобы синтезатор воспроизвел слово, — это пра-

¹ «Тайп'н ток» — торговая марка отделения «Вотракс» компании «Федерал скрю уоркс».

вильно составить цепочку из образующих слово фонем. Число фонем, используемых в фонемных синтезаторах, различно в разных конструкциях. Некоторые довольно грубые синтезаторы обходятся всего 32 фонемами, тогда как в других в распоряжение пользователей, синтезирующих слова, предоставляются сотни фонем и аллофонов. Совершенно очевидно, что, чем большее разнообразие звуков речи доступно пользователю, тем лучше будут звучать составленные из фонем слова и выражения. Однако увеличение набора фонем повышает требования к скорости передачи данных для воспроизведения речи и необходимому объему памяти. Это также приводит к увеличению нагрузки на компьютер, если он должен производить расчленение слов на отдельные фонемы.

В приведенном примере слово *hello* произносилось как последовательность четырех фонем. Для артикуляции дифтонгов и отражения динамических изменений речи в ходе произнесения подобного слова высококачественному аллофонному синтезатору может потребоваться от десяти до пятнадцати звуков речи. Задача нахождения четырех основных звуков слова *hello* может оказаться довольно простой, однако попробуйте найти 14—15 аллофонов, необходимых для воспроизведения этого слова в более сложных системах! Подобная задача почти неразрешима для более длинных слов. Для получения окончательно отработанной речи придется провести несколько часов за клавиатурой, действуя методом проб и ошибок.

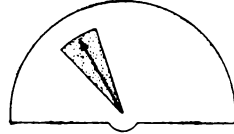
И наконец, последнее замечание относительно формантных синтезаторов речи с фонемным возбуждением. При большем числе доступных пользователю звуков речи лучшего звучания окончательно отработанной речи можно добиться, лишь затратив достаточно много времени на правильную сборку слов из фонем и аллофонов. Если вас удовлетворяет звучание слова, составленного из меньшего числа фонем, то в таком случае вам, по существу, не нужны сотни вариаций фонем. Добавляя к слову, составленному из основных фонем, дополнительные звуки речи, вы лишь делаете речевой выход синтезатора более понятным все большему кругу нетренированных слушателей. Это не следует упускать из виду.

Конечно, если вы решите приобрести систему

речевого синтеза с фонемным возбуждением, наделенную способностью преобразовывать текст в речь, то вам придется столкнуться с другим ограничением: вы будете связаны словарем, который заложен в таблицу расчленения слов, встроенную в синтезатор. Естественно, можно воспроизводить слова, не входящие в эту таблицу, однако для этого придется самостоятельно заниматься фонетической композицией. Опыт показывает, что достаточно синтезатору, снабженному преобразователем текста в речь, произнести одно-два предложения, как система сталкивается со словом, которое она не способна правильно расчлениить на фонемы,— тогда в синтезируемой речи появляются «марсианские» слова. Этот сбой даже нельзя отнести к недостаткам говорящей системы, ибо он скорее обусловлен чудовищной непоследовательностью правил, действующих в английской речи.

Метод аналогового синтеза формантных частот речи довольно специфичен. Если прослушать речь синтезаторов других типов и сравнить ее с речевым выходом синтезатора формантного типа, то их без труда можно различить. Звуки речи, генерируемые формантным синтезатором, отличаются значительной мягкостью и даже отдают некоторой слащавостью. Первые системы, в которых применялись аналоговые формантные фильтры, произносили слова так, как говорит человек с набитым ртом.

По мере совершенствования динамического изменения управляющих параметров речь формантных синтезаторов становилась более разборчивой. Этот метод синтеза речи весьма перспективен, и его использование, безусловно, будет продолжаться. Главное преимущество его — очень низкая скорость передачи данных, управляющих речевым выходом; подобная скорость характерна и для синтеза речи по методу ЛПК, о котором мы будем говорить в следующем разделе. Эти два метода синтеза речи обладают и другими сходными чертами. Знакомясь с методами синтеза речи на основе ЛПК и с тем, как цифровая фильтрация влияет на конечный речевой выход, не забывайте о схеме, представленной на рис. 6.14,— она позволит обнаружить удивительное сходство этих методов.



Синтез речи по методу линейного предиктивного кодирования (ЛПК)

Линейное предиктивное кодирование речи — один из наиболее стремительно развивающихся методов синтеза речи. Вместе с тем этот метод наиболее труден для понимания. Причина состоит в том, что регенерация речи при его использовании выполняется математически с использованием уравнений преобразования закодированной речи в спектры ее исходных частот. Однако существует определенное и большое сходство между этим методом и методом кодирования-восстановления речевого сигнала. Так, было бы очень трудно построить ЛПК-речь, начав с нуля. Как правило, генерация параметров для синтеза речи методом ЛПК производится человеком, который наговаривает нужные слова. Сходство здесь, по существу, так велико, что если мы подробно изучим механизм процесса кодирования и восстановления речи по методу ЛПК, то в скрытой форме обнаружим в нем весь процесс кодирования речевого сигнала.

На рис. 6.17 показана упрощенная схема генерации речи по методу ЛПК. Слева на рисунке изображен микрофон, после которого речевой сигнал, произнесенный человеком, пройдя через аналого-цифровой преобразователь, принимает вид выборки дискретизированной речи. В этой части системы речевой сигнал представлен непосредственно в цифровом коде, и его можно хранить в памяти, а затем восстанавливать так, как было описано в предшествующих разделах, рассказывающих о восстановлении речи из закодированного речевого сигнала. Однако если выборки речевого сигнала направляются в компьютер, который производит цифровой анализ речи с целью определения ее спектрального состава, то мы уже не располагаем последовательностью цифровых байтов, непосредственно соответствующей исходному речевому сигналу. Результирующие выходные кадры данных, поступающие из системы ана-

лизи речи, дают нам такую информацию, как высота основного тона голоса, характеристики формантных частот, а также данные об амплитуде и интонации. Массив информации на выходе цифрового процессора речи, таким образом, весьма впечатляет своими размерами; однако, не располагая всеми этими сведениями, невозможно в конечном счете осуществить ЛПК-кодирова-

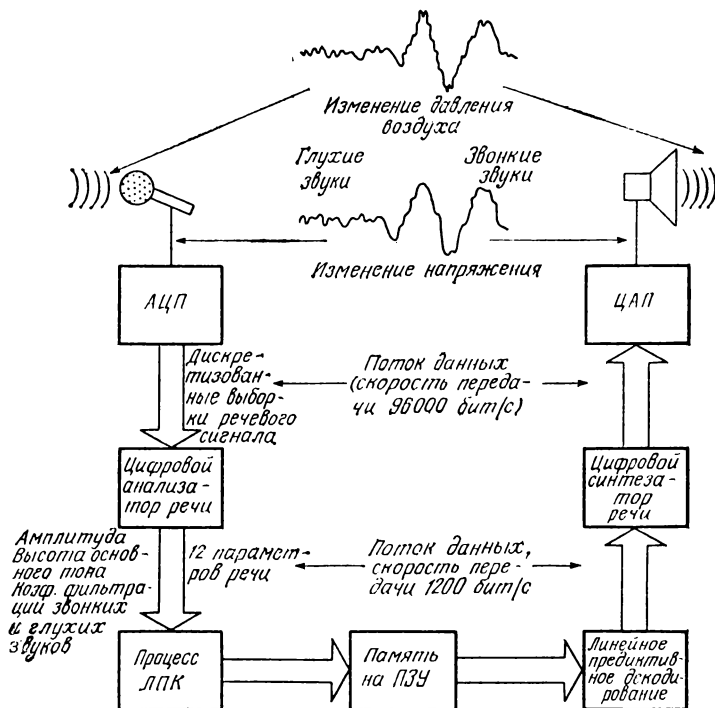


Рис. 6.17. Принцип линейного предиктивного кодирования.

ние. Блок «Линейное предиктивное декодирование», реализованный на базе компьютера для обработки речи, преобразует спектральный состав речи в ЛПК-параметры. Основной принцип, положенный в основу линейного предиктивного кодирования, сводится к тому, что поступающие выборки речевого сигнала могут рассматриваться как линейные комбинации прошлых выборок речевого сигнала. Таким образом, блок кодирования превращается в устройство краткосрочного прогнози-

рования, т. е. своеобразного «предсказателя» последующего речевого сигнала. Выход этого предсказателя управляет параметрами и определяет числовые коэффициенты, которые используются для линейной предиктивной генерации речи.

Параметры управления, формируемые в процессе линейного предиктивного кодирования, в обычном случае состоят из информации об амплитуде и высоте тона, которая сопровождается выдачей управляющего сигнала, указывающего, является ли звук звонким или глухим. Наиболее важными выходными данными в процессе ЛПК считаются коэффициенты для устройства прогнозирования — предикторные коэффициенты, которые в конечном счете используются в процессе восстановления речи цифровым фильтром. Число подобных коэффициентов меняется в зависимости от частоты дискретизации — частоты получения выборок и требуемого качества восстановленной речи. В современных коммерческих изделиях число предикторных коэффициентов, как правило, равно 10—12.

Выходные данные с линейного предиктивного кодирующего устройства (кодера) имеют форму цифровой информации, которая содержится в постоянном запоминающем устройстве (ПЗУ) и служит для регенерации речи, формируемой системой ЛПК-синтеза. Далее в этом разделе мы рассмотрим типичный кадр ЛПК-данных и расскажем, какую роль играют различные коэффициенты. После того как ЛПК-параметры записаны в ПЗУ синтезатора, можно приступить к синтезу речи, используя процесс линейного предиктивного декодирования. Соответствующие операции выполняются в блоках, которые на рис. 6.17 расположены справа. Иногда все эти блоки изготавливаются в виде одной интегральной схемы, что позволяет получать речь, синтезированную ЛПК-методом, при минимальном объеме необходимых электронных схем. Структура интегральной схемы ЛПК-синтеза, выпускаемой фирмой «Тексас инструментс» (этой компании первой удалось построить схемы ЛПК-синтеза речи в виде микросхемы), показана на рис. 6.18.

Эта схема синтезатора, отражающая особенности внутреннего устройства системы TMC-0281 (TMS 5100) фирмы «Тексас инструментс», показывает, как проходит информационный поток в полноструктурной систе-

ме восстановления искусственной речи по методу ЛПК. На рисунке логический поток данных следует из верхнего правого угла в левый, затем вниз (через штриховую линию), после чего поступает на выход (в правом нижнем углу схемы). Информация выводится в форме аналогового речевого сигнала. Часть схемы, изображенная выше штриховой линии, относится к системам цифровой обработки данных, которые приводят в действие механизм ЛПК-синтеза. Ниже штриховой линии пред-

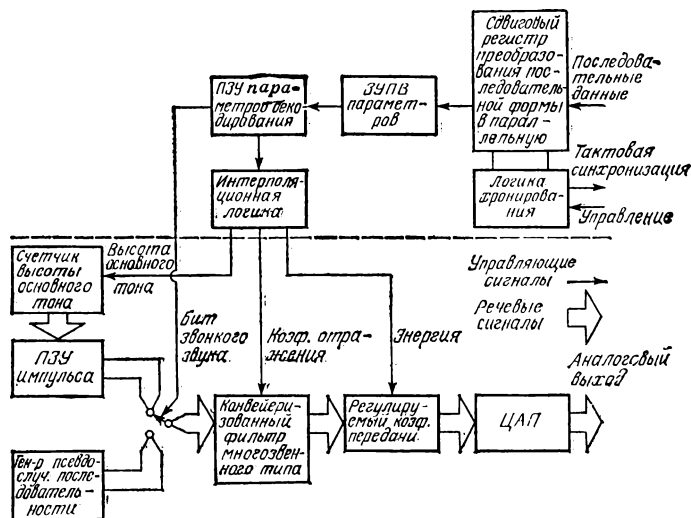


Рис. 6.18. Структурная схема ЛПК-синтезатора.

ставлены схемные компоненты, обеспечивающие прямой синтез речи с помощью расчетных ЛПК-коэффициентов.

Некоторые блоки на этой схеме не являются универсальными: здесь показан лишь один из возможных вариантов реализации ЛПК-синтеза, спроектированный в отделе говорящих изделий фирмы «Тексас инструментс». Однако блоки, приведенные в нижней части рис. 6.18, — действительно универсальные; они служат неременными компонентами практически всех речевых ЛПК-синтезаторов. Сказанное относится к генератору высоты основного тона (вспомните об импulsive для голосовой щели), генератору псевдослучайного шума для получения фрикативных звуков и к системе циф-

ровой фильтрации, служащей для модификации звонких и глухих звуков с целью их превращения в речь. Последний блок в системе ЛПК-генерации речи — это цифро-аналоговый преобразователь, который преобразует речь, записанную в форме единиц и нулей, в аналоговый сигнал, используемый для возбуждения громкоговорителя.

Как и у аналогового формантного синтезатора, возбуждение ЛПК-синтезатора производится периодически, причем частота поступления обновляющего управляющего кадра должна быть достаточно высокой, чтобы следить за динамикой изменения голосового тракта человека. В зависимости от изготовителя интегральной схемы и числа используемых предикторных коэффициентов (коэффициентов отражения) типичное значение частоты повторения кадров лежит в пределах 40—100 раз в секунду. Иными словами, в схемы ЛПК-синтеза каждые 10 или 25 мс вводится совершенно новый набор речевых данных, который используется для генерации речевого сигнала в остальную часть периода того же управляющего кадра. Исследовав более детально структуру кадра ЛПК-данных и то, как этот кадр связан со структурной схемой, изображенной на рис. 6.18, мы лучше почувствуем всю сложность такого процесса, как ЛПК-синтез. На рис. 6.19 дано наглядное описание ЛПК-данных, которые используются в ЛПК-методе, принятом фирмой «Тексас инструментс». В верхней части рисунка приведен кадр данных, относящийся к звонкому звуку. Каждая ячейка кадра представляет собой элемент информации, необходимой для формирования синтетического речевого выхода методом ЛПК. Первый элемент, «Энергия», служит для непрерывного управления амплитудой произносимой речи. Следующий элемент, «Повторить кадр», использует такое свойство речи, как медленное изменение характеристик голосового тракта при воспроизведении «статических» звуков. Если, например, во время произнесения какой-то фразы звук *ah* должен воспроизводиться в течение более длительного времени, чем продолжительность одного управляющего кадра, то используется бит (разряд) «Повторить кадр», который приказывает синтезатору повторить десять предыдущих коэффициентов фильтра ($K_1 — K_{10}$). Введение такого разряда повторения в кадр управляющих ЛПК-данных позволяет

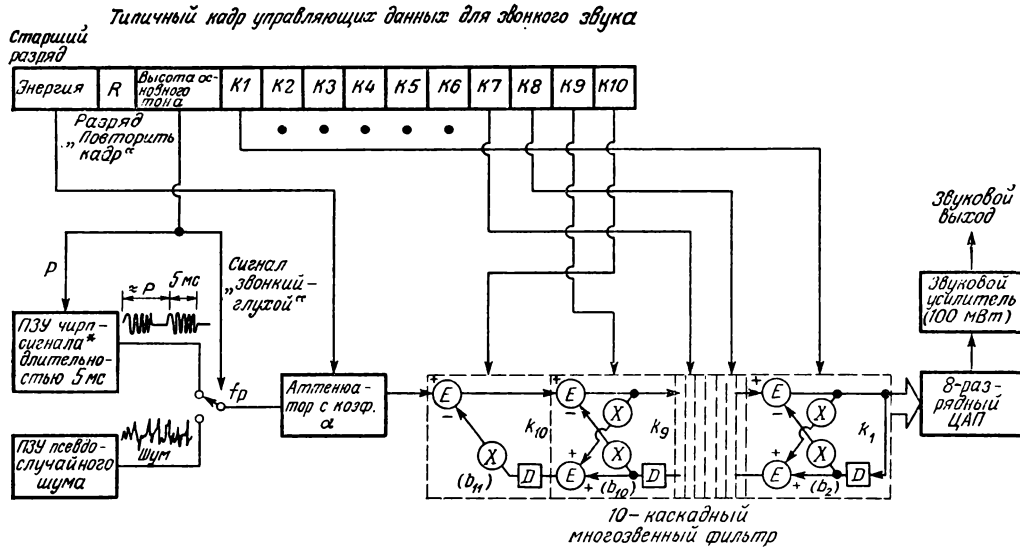


Рис. 6.19. Наглядное представление кадра управляющих ЛПК-данных фирмы «Тексас инструментс». (* Импульс с линейной частотной модуляцией, ЛЧМ.)

сократить почти на 80% число элементов в кадре для «статических» звуковых данных. Следующий элемент в ЛПК-кадре — управляющее слово «Высота основного тона», которое регулирует не только высоту тона (или период P) синтезатора импульса голосовой щели, но и определяет, является ли воспроизводимый звук звонким или глухим. Если высота тона произносимой фразы равна нулю, то, согласно определению, это соответствует глухому звуку и тогда переключатель «звонкий — глухой» подключает к входу цифрового многозвенного фильтра источник псевдослучайного шума.

Десять последних элементов кадра ЛПК-данных соответствуют предикторным коэффициентам, которые используются в цифровом многозвенном фильтре для регенерации исходной речи. Каждый из этих 10 коэффициентов представляет собой численную величину, в соответствии с которой в многозвенном фильтре устанавливается коэффициент передачи в цепи специальной обратной связи. Для восстановления речи такое число коэффициентов используется не всегда (при воспроизведении глухих звуков оно меньше); это — максимальное количество коэффициентов для кадра управляющих ЛПК-данных в системе, разработанной фирмой «Тексас инструментс».

Различные типы кадров управляющих данных для ЛПК-синтезаторов речи проектируются с таким расчетом, чтобы сократить потребность в памяти и уменьшить предельную скорость передачи данных при синтезе речи. В синтезаторах, изготовляемых фирмой «Тексас инструментс», используются пять базовых форматов для кадров цифровой речевой информации (рис. 6.20). В синтезаторе TMS5100 максимальное число разрядов в полном управляющем кадре для звонкого звука равно 49. Кадр такого формата используется для формирования любого звука, который естественным образом воспроизводится вибрацией голосовых связок. Фрикативные и взрывные согласные генерируются с помощью управляющего кадра для глухих звуков; одновременно управляющее слово «Высота основного тона» устанавливается равным нулю. Обратите внимание на то, что число коэффициентов для фильтра снижается до четырех, поскольку шипящие звуки генерируются в части голосового тракта, расположенной выше, и, следо-

вательно, в меньшей степени фильтруются в голосовом тракте. Длина управляющего кадра для глухих звуков равна 28 разрядам, что обеспечивает весьма эффективный речевой синтез, особенно если учесть, что таким путем генерируются высокочастотные звуки. Следующий формат кадра, приведенного на рис. 6.20, — формат «Повторить кадр». Как уже говорилось, такой кадр используется для повторения любого из предшествующих кадров при сохранении всех коэффициентов

	Старший разряд энергии	Повторить	Высота тона	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	Длина
Кадр для звукового звука	XXXX	0	XXXXX											49 бит
Кадр для глухого звука (Высота основного тона = 0)	XXXX	0	00000											28 бит
Повторяемый кадр	XXXX	1	XXXXX											10 бит
Кадр с нулевой энергией	0000													4 бит
Кадр с кодом останова	1111													4 бит
Число разрядов	4	1	5	5	5	4	4	4	4	4	3	3	3	

X — безразличное состояние

Рис. 6.20. Форматы ЛПК-данных, используемые фирмой «Тексас инструментс».

K1—K10 неизменными. Уменьшение длины управляющего кадра до 10 разрядов значительно повышает эффективность воспроизведения звуков «статического» типа. Два последних формата на рис. 6.20 — это форматы нулевой энергии и кода останова. Формат нулевой энергии есть не что иное, как «произносимое» молчание. При его воспроизведении выход синтезатора, по существу, выключается — с целью воспроизведения пауз в речи. Формат останова дает синтезатору команду прекратить говорить и вернуться в состояние готовности выполнять следующие приказы управляющего компьютера. Хотя описанные кадры управляющих данных специфичны именно для систем синтеза речи, выпускаемых фирмой «Тексас инструментс», им присущи характерные особенности и других систем, например ЛПК-синтезаторов, выпускаемых фирмами «Дженерал инструментс»

и «Телесенсорн», а также японских синтезаторов, работающих по принципу частотной корреляции (Паркор).

Описание синтеза речи по методу линейного предиктивного кодирования мы вынуждены были дать на примере соответствующих аппаратных реализаций, хотя сам этот метод — чисто математический. Если бы мы попытались изложить процесс линейного предиктивного кодирования речи более строго, то 20—30 страниц этой книги нам пришлось бы заполнить одними уравнениями. Тем, кто желает ознакомиться с этими уравнениями, мы рекомендуем обратиться к книгам по ЛПК-синтезу речи, указанным в списке литературы.

Процесс, который происходит в ЛПК-синтезаторе при выводе речи, трудно смоделировать посредством любого из существующих персональных компьютеров. Во время синтеза речи каждый предикторный коэффициент, используемый в цифровом фильтре, обновляется с частотой подачи управляющих кадров примерно 50 кадров в секунду. При каждом очередном обновлении, или перенастройке, фильтра (производимом каждые 20 мс), схема ЛПК-синтезатора выполняет 4000 умножений 10-разрядных чисел на 14-разрядные и 4000 сложений 14-разрядных чисел. Попробуйте-ка повторить все это на персональном компьютере!

Системы, используемые для выполнения вычислений со столь огромной скоростью, представляют собой довольно сложные схемы реализации математических алгоритмов умножения и деления. Если же производить эквивалентные вычисления с помощью программы на Бейсике средствами персонального компьютера, то частота поступления кадров управления речью составит, вероятнее всего, не 50 кадр/с, а лишь 1 кадр в 5—10 мин. Следовательно, решающее значение для успешного практического синтеза по ЛПК-методу имеет его аппаратная реализация. Каждый изготовитель располагает своими собственными методами выполнения вычислений, связанных с ЛПК-синтезом, а также способами изготовления соответствующих интегральных микросхем. Независимо от того, как это в конечном счете достигается, в будущем метод ЛПК синтеза речи, несомненно, станет одним из самых распространенных. Говорящие изделия, построенные по этому методу, появились сравнительно недавно и быстро завоевывают

успех у потребителей. Линейное предиктивное кодирование обеспечивает синтез речи, которая близка по качеству к естественной человеческой речи, при относительно низкой скорости передачи данных. Стоимость этих изделий также сравнительно невысока, что объясняется простотой их воплощения в виде интегральных схем. Все это позволяет создавать довольно дешевые синтезаторы.

Естественно, что здесь (как и в случае прямого кодирования-восстановления речевого сигнала) остается нерешенной одна весьма важная проблема. Необходимо иметь в виду, что речевой выход компьютера, работающего с такими системами, ограничен предварительно наговоренным словарем. Чтобы несколько ослабить это ограничение, некоторые изобретательные изготовители ввели в синтезаторы фонетические и аллофонические звуки, записанные в прямом ЛПК-коде. Это предоставляет пользователю неограниченный словарь, который он сам должен строить путем последовательного фонетического соединения ЛПК-звуков речи. По принципу использования эти системы весьма сходны с синтезаторами формантных частот с фонемным возбуждением, описанными в предыдущем разделе. Сочетание ЛПК-метода с фонемным синтезом дает исключительно мощную систему синтеза речи. Однако речь, генерируемая такой системой, произносится голосом робота, что обусловлено объединением отдельных фонем в цепочки. Вместе с тем качество такой речи не хуже, чем у любого синтезатора с фонемным возбуждением. Естественно, пользователь всегда должен идти на компромисс, выбирая между неограниченным словарем, который будет воспроизводиться голосом робота, и предоставляемым изготовителем ограниченным словарем, который будет звучать, как естественная речь человека. Конечно, при наличии средств можно попросить изготовителя ЛПК-схем закодировать в ЛПК-речь весь желательный для пользователя словарь. Так часто поступают крупные фирмы, которые затем выпускают миллионы изделий с синтезированной речью и одинаковым словарем, компенсируя таким образом затраты по кодированию заказного неограниченного словаря.

В ближайшем будущем устройства синтеза речи по методу ЛПК начнут в ограниченных количествах поступать в продажу. По мере повышения эффективности

их производства стоимость изделий будет понижаться, что позволит энтузиастам-любителям генерировать собственный словарь, наговаривая слова своим голосом.

Другие методы

Хотя существующие теперь методы синтеза речи в основном относятся к трем описанным выше типам, не исключено, что в будущем могут появиться и другие способы. Системы завтрашнего дня, которые находятся на пределе возможностей данной области техники, пока еще не вышли из стен лабораторий. Большинство из таких «экзотических» синтезаторов речи действует по принципу математического моделирования голосового тракта человека. Внешне они очень напоминают ЛПК-синтезаторы, отличаясь от последних характером вычислений, которые используются для получения речевого выхода. Существуют возможности для воспроизведения речи на основе прямого фурье-синтеза, синтеза с помощью функций Уолша и методов, использующих корреляцию и автокорреляцию сигналов (Паркор).

Теоретически большинство этих методов весьма сходны с методом ЛПК-синтеза речи. Различия между ними связаны в основном с их аппаратной реализацией. Если какой-то из этих методов превзойдет по своей эффективности метод ЛПК, то это произойдет потому, что его аппаратная реализация окажется менее сложной. Только время покажет, какая из систем синтеза речи завоюет будущее. Учитывая непрерывное возрастание сложности при переходе от каждого предшествующего метода синтеза к последующему, можно сказать, что с позиций современного технического уровня мы просто не в состоянии предсказать, каким будет следующее поколение синтезаторов речи.

Глава 7

СЕРИЙНЫЕ СИСТЕМЫ СИНТЕЗА РЕЧИ

Автор берет на себя смелость предполагать, что, ознакомившись с содержанием предыдущих глав, читатель в известной степени стал «специалистом» в области генерации синтезированной речи и технических средств ее реализации. Теперь целесообразно рассмотреть выпускаемые серийно синтезаторы речи. Перечень изделий, который мы здесь рассмотрим, очевидно, не отличается полнотой, поскольку все время создаются новые говорящие устройства. Наиболее удачный метод синтеза речи скорее всего определится так же, как это было в свое время с микропроцессорами. На заре производства микропроцессоров появилось большое число изделий, из которых собирались самые различные микрокомпьютеры. По мере изготовления новых, все более совершенных систем многие ранее выпущенные постепенно исчезали, не выдержав конкуренции. Аналогичная картина, вероятно, будет наблюдаться и на рынке говорящих устройств. По этой причине некоторых изделий, упоминаемых в этой книге, возможно, сейчас уже нет в продаже — их заменили другие, ибо, как говорится, «свято место пусто не бывает». По-видимому, верх в этой борьбе одержат наиболее эффективные и обладающие наилучшим звучанием системы синтеза речи, которые завоевывают широкое признание у потребителей.

Наш обзор выпускаемых в настоящее время устройств синтезированной речи основан главным образом на данных, полученных от изготовителей. Если вы столкнетесь с тем, что описание той или иной системы слишком кратко, то это значит, что либо ее изготовитель не захотел раскрывать «секреты фирмы», либо же во время публикации настоящей книги не было достаточно подробных сведений о данной системе.

Мы будем рассматривать эти изделия в последовательности, отвечающей используемым в них методам

синтеза речи. К первой группе синтезаторов относят устройства, созданные по методу кодирования-восстановления речевого сигнала. Следующая группа говорящих изделий включает аналоговые формантные синтезаторы. Эти приборы чаще всего называют фонемными синтезаторами; они обеспечивают возможность генерации речи с неограниченным словарем. К последней (конечно, не по важности) группе изделий относятся устройства с механизмом речевого синтеза на основе линейного предиктивного кодирования (ЛПК).

Некоторые из изделий, описанных в этой главе, представляют собой одиночные интегральные схемы. Они интересны прежде всего тем, что их можно своими силами ввести в персональную компьютерную систему. Если же вы интересуетесь конечным результатом, то вам, вероятно, более придется по вкусу говорящие периферийные устройства, которые можно приобрести в готовом виде.

Системы кодирования-восстановления речевого сигнала

К первой группе рассматриваемых нами говорящих устройств относятся устройства, использующие метод цифровой записи и воспроизведения, который в предыдущей главе был назван методом кодирования-восстановления речевого сигнала. Некоторые из упоминаемых здесь систем специально предназначены для совместного использования с такими серийными персональными компьютерами, как Эппл фирмы «Эппл компьютер корпорейшн», TRS-80 отделения «Рейдио шэк» фирмы «Тенди корпорейшн» и т. д. Другие говорящие устройства спроектированы как автономные приборы или же рассчитаны на применение с существующими ЭВМ в качестве компьютерных периферийных устройств. Все эти системы, конечно, выполняют свою функцию — синтез речи, но качество генерируемой речи у них различно, и, чем оно выше, тем дороже стоит система.

Сравнительные характеристики синтезаторов речи с кодированием-восстановлением речевого сигнала, рассматриваемых в данном разделе, приведены в табл. 7.1. Перечисленные в таблице синтезаторы имелись в продаже, когда печаталась наша книга, так что указанные технические характеристики каждого устройства от-

Таблица 7.1. Характеристики некоторых типов синтезаторов, выпущенных в 1982 г. (устройства с прямым кодированием-восстановлением речевого сигнала)

Название и фирма-изготовитель	Тип синтеза	Размеры, мм	Речевые возможности	Тип сопряженного компьютера
LISA фирмы «Сентиграм»	Параметрическое кодирование речевого сигнала	317×290×106	Предварительно записанный в память словарь	Любая машина с последовательным выводом через интерфейс RS-232
Чипток (TRS-80)	ЧМ кодирование-восстановление речевого сигнала	Чисто программное средство	Хранимый в памяти словарь пользователя	TRS-80, H-8
Компьюкордер фирмы «Компьютокер консалтантс»	Прямое кодирование речевого сигнала	12,7×254 (печатная плата)	Хранимый в памяти словарь пользователя	S-100, IEEE-696
Майкромаус фирмы «Майкроминт»	Микросхема устройства Диджитокер фирмы «Нэшнл семикондактор»	177×176×50	Покупной словарь на ПЗУ	Эпл, TRS-80, S-100, H-8
Диджитокер фирмы «Нэшнл семикондактор»	ИКДМ-кодирование речевого сигнала	Различные	Покупной словарь на ПЗУ	Автономные платы и микросхемы любого типа
Сериес III фирмы «Телесенсори сиш системс»	Кодирование речевого сигнала с помощью микросхемы заказного контроллера ПЗУ	101×114 (плата)	Покупной словарь на ПЗУ	Любой параллельный ТТЛ-порт
Войстек фирмы «Конинокс Ви-Ай-Оу»	Прямое кодирование речевого сигнала	127×152×32	Слова, наговоренные пользователем. Осуществляет также распознавание речи.	Эпл, TRS-80, ПЕТ, AIM-65

ражают действительное положение дел. Как видим, в таблице представлены следующие характеристики перечисленных в ней синтезаторов: метод используемого в них синтеза; примерные размеры систем; особенности речевых возможностей системы; тип компьютера, с которым может использоваться синтезатор.

Система LISA фирмы «Сентиграм корпорейшн»

Система LISA фирмы «Сентиграм корпорейшн» представляет собой автономный терминал, который служит компьютерным периферийным устройством. Он рас-

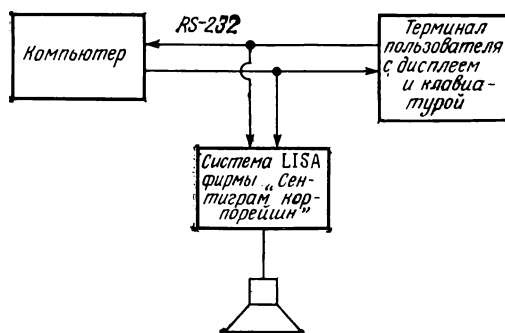


Рис. 7.1. Синтезатор речи LISA.

считан на включение между телетайпом и терминалом (на ЭЛТ) пользователя, с одной стороны, и ведущим компьютером — с другой. Эта система, по всей видимости, принадлежит к наиболее дорогим серийным устройствам с кодированием речевого сигнала (если к этому классу действительно можно отнести параметрическое кодирование речевого сигнала, применяемое фирмой «Сентиграм корпорейшн»). Поскольку фирма строго хранит секрет метода параметрического кодирования речевого сигнала и информация по нему весьма незначительна, располагая этими скудными сведениями, мы рискнули отнести его к первой группе рассматриваемых методов синтеза. Структурная схема возможного подключения терминала с речевым выходом LISA к типичной компьютерной системе приведена на рис. 7.1. Одной из особенностей устройства является то, что оно может быть включено между клавишным

пультом и компьютером и в таком случае работает как прозрачное (не требующее участия пользователя) периферийное устройство, способное генерировать речь.

Говорящий терминал системы LISA изображен на рис. 7.2. В процессе работы это периферийное устройство выделяет из информационного потока адресованные ему речевые команды и речевые данные, не оказывая никакого влияния на обмен обычными компьютерными командами, происходящий между клавиш-

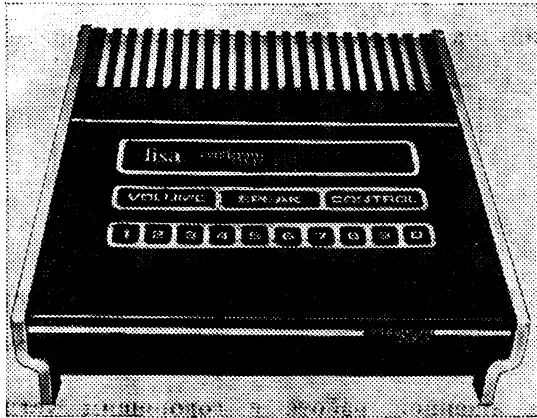


Рис. 7.2. Внешний вид терминала с речевым выходом LISA, (С разрешения фирмы «Сентиграм корп.».)

ным пультом и компьютером. Словарь системы LISA опирается на набор хранимых в памяти слов, предложений или наговоренного текста любой длины. В качестве памяти служит дисковый накопитель ведущего компьютера, откуда хранимый текст считывается для воспроизведения нужных выражений. При работе по такой схеме система LISA предоставляет компьютеру, к которому она подключена, неограниченные в словарном отношении речевые возможности.

«Встроенный» словарь для дискового накопителя фирма «Сентиграм корпорейшн» изготавливает в виде дискового файла. Затем он переносится из компьютера в систему LISA с помощью прикладной программы. И наконец, команды на произнесение требуемых слов из словаря выдаются системе LISA операционной

системой пользователя. Словарь в дисковом файле может быть построен на основе предварительно подготовленного текста, который пользователь направляет фирме «Сентиграм» для кодирования (за соответствующую плату), или же на основе вашей собственной библиотеки наговоренных слов: ее вы можете самостоятельно составить с помощью генератора библиотеки наговоренных слов, выпускаемого фирмой «Сентиграм». Этот (довольно дорогой) прибор дает пользователю возможность строить свои собственные словари с той же легкостью, с какой производится цифровая запись на магнитную ленту. Поток данных в компьютере при

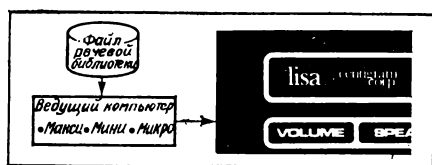


Рис. 7.3. Схема включения системы LISA (С разрешения фирмы «Сентиграм корп.»).
 Рассчитана на использование в составе компьютерных систем.

его сопряженной работе с говорящим устройством LISA показан на рис. 7.3. Простота интерфейса делает возможным подключение данной говорящей системы практически к любому ведущему компьютеру, который нуждается в периферийном устройстве с речевым выходом.

В работе системы с речевым выходом LISA используется метод, который фирма «Сентиграм корпорейшн» называет параметрическим синтезом речевого сигнала. Этот тип синтеза требует скорости подачи речевых данных примерно 300—600 байт/с, что по эффективности ставит ее в один ряд с ЛПК-синтезаторами речи. Фактически эти методы синтеза речи обладают значительным сходством, но, поскольку метод параметрического кодирования речевого сигнала остается секретом фирмы, относительно его имеется очень мало информации. Известно, однако, что для обеспечения потока данных между компьютером и терминалом пользователя применяется интерфейс типа RS-232 и что скорость передачи данных между ними составляет 110—19 200 бод.

В синтезаторе LISA установлен микропроцессор Z80, оснащенный ЗУПВ (емкостью 16К) для хранения закодированных речевых сообщений. Звук с выхода этой системы, генерируемый громкоговорителем мощностью 2 МВт с диаметром диффузора 100 мм, по громкости вполне достаточен для средней комнаты. Конструкторы проявили предусмотрительность, оснастив систему LISA стандартным соединительным гнездом для подключения головного телефона (наушников). Это дает возможность пользователю компьютера прослушивать речевой выход в полном уединении, не мешая окружающим. «Сентиграм корпорейшн» выпускает не только систему LISA для обработки речи, но и систему MIKE, предназначенную для распознавания речи, и ряд других систем для осуществления речевого диалога между компьютером и пользователем.

Синтезатор Чипток

Это одна из немногих чисто программных речевых систем, разработанная Аланом Севиллом, которая предназначена специально для персональных компьютеров и практически не требует дополнительной аппаратуры. Простота речевого синтезатора Чипток объясняется использованием в нем специфического процесса кодирования речевого сигнала. Система спроектирована в расчете на работу с компьютерами серии TRS-80 фирмы «Рейдио шэк» и компьютером H-8 фирмы «Хит». Единственное, что здесь необходимо для дискретизации (преобразования аналогового сигнала в цифровой) наговоренных слов и их последующего воспроизведения, — это наличие одного двоичного порта ввода-вывода. Для дискретизации и воспроизведения речевого сигнала можно использовать недорогой звуковой усилитель и микрофон. В случае применения любого пригодного для работы с этой системой компьютера для кодирования речевого сигнала необходимо выполнить простейшую модификацию аппаратных средств. При работе с машиной H-8 фирмы «Хит» для кодирования речевого сигнала достаточно вставить одну-единственную дополнительную перемычку. Для речевого воспроизведения закодированной

речи необходимо изменить соединение всего одного вывода интегральной схемы на плате передней панели. Этот вывод следует подключить к внутреннему громкоговорителю. В то же время компьютер TRS-80 фирмы «Рейдио шэк» способен воспроизводить закодированный речевой сигнал через звуковой выход кассетного магнитофона, не требуя введения в схему каких-либо изменений. Здесь кодирование речевого выхода требует небольших изменений клавишного пульта (или добавления нескольких внешних логических вентиляей), что обеспечивает требуемое стробирование речевого сигнала. Программный вариант синтезатора Чипток будет производить дискретизацию вашего голоса (с кодированием речевого сигнала) при его подключении к соответствующему входному порту, а затем воспроизводить по команде закодированную речь с помощью звуковой выходной системы.

Использованный в устройстве Чипток процесс кодирования речевого сигнала не был описан в предыдущей главе, посвященной кодированию таких сигналов. Он, однако, имеет сходство с процессом кодирования по методу дельта-модуляции, состоящее в том, что в нем по каждой выборке из речевого сигнала также запоминается всего один бит данных. Такой же процесс используется и в некоторых других чисто программных синтезаторах речи. Он получил название метода кодирования речевого сигнала с частотной модуляцией (ЧМ). Суть процесса поясняется на рис. 7.4. Поступающий на вход синтезатора речевой сигнал изображен в верхней части рисунка. Этот сигнал дискретизируется с частотой, равной примерно 4000 выборок в секунду. Каждая выборка является бинарной величиной, которой присваивается единичное значение, если входной сигнал больше предварительно установленного порога для единицы, и нулевое, если он меньше порога, установленного для нуля. Поскольку в качестве входных каскадов пороговой схемы используются простейшие ТТЛ-вентили, пороговый уровень для логического нуля равен 0,8 В, а для логической единицы — 2,4 В. По мере того как входной речевой сигнал изменяется по амплитуде между двумя этими пороговыми уровнями цифровой код речи меняется от 0 до 1. Изображенный в нижней части рис. 7.4 поток выходных импульсов представляет одновре-

менно и последовательность однобитовых данных, хранимых в памяти, и выходные данные, направляемые на громкоговоритель для воспроизведения речи. Совершенно очевидно, что между входными и выходными сигналами очень мало сходства, однако получаемый таким путем искаженный речевой выход можно все же понять. Учитывая невысокую стоимость этой системы, можно считать, что она превосходно справ-

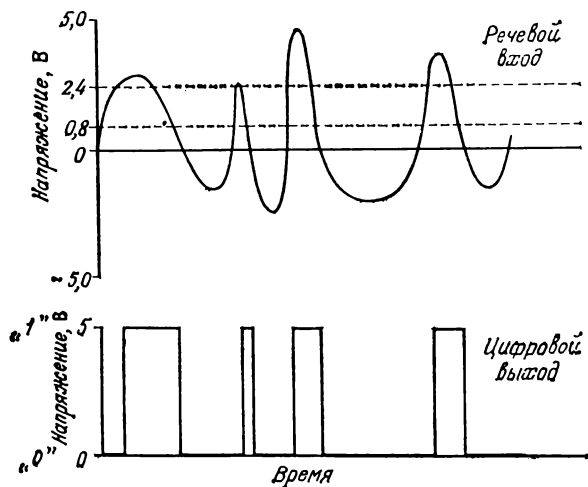


Рис. 7.4. Процесс бинарного ЧМ-кодирования.

ляется с задачей получения от компьютера самых первых слов.

Для компьютера TRS-80 система Чипток выпускается в виде магнитофонной кассеты, а для компьютера Н-8 — в виде диска Н-17. Приобретаемая программа обеспечивает возможность кодирования и восстановления речевого сигнала для каждого из этих компьютеров. Оба этих программных средства содержат демонстрационную, или говорящую, программу, которая помогает обучить компьютер речи. В программе содержатся слова, наговоренные Аланом Севиллом, которые можно вывести из памяти посредством 16 отдельных распечаток голоса (дампов), а также демонстрационная алфавитная последовательность.

Устройство Компьюкордер фирмы «Компьютокер консалтантс»

Устройство Компьюкордер фирмы «Компьютокер консалтантс» представляет собой недорогой цифровой прибор, сочетающий накопитель с выходным устройством. Речевая плата, которую, по существу, можно считать твердотельным магнитофоном, — второе важнейшее изделие этой фирмы, созданной в период появления

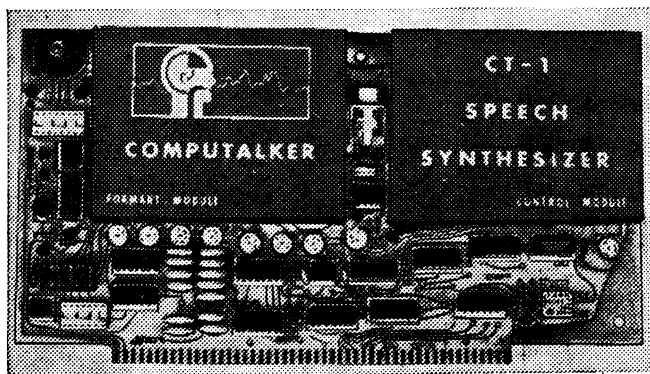


Рис. 7.5. Синтезатор речи СТ-1 фирмы «Компьютокер консалтантс». (С разрешения фирмы «Компьютокер консалтантс».)

первых микропроцессоров и специализирующейся сейчас на синтезе речи. Свой первый синтезатор речи для персонального компьютера фирма начала поставлять с ноября 1976 г. Этот синтезатор СТ-1 был спроектирован в расчете на подключение к шине S-100. Устройство работало по методу аналогового формантного синтеза.

Учитывая историческое значение синтезатора СТ-1 как одного из первых серийных синтезаторов для персональных компьютеров, мы считаем не лишним показать его внешний вид (рис. 7.5). Конструкция синтезатора по тем временам была довольно прогрессивна. Однако весьма высокая стоимость этого синтезатора ставила его в разряд предметов роскоши. Поэтому фирма «Компьютокер консалтантс» прекратила производство этого устройства и вместо него стала выпускать синтезатор с прямым кодированием-восста-

повлением речевого сигнала — Компьюкордер. В отличие от синтезатора СТ-1 прибор Компьюкордер можно приобрести и сегодня как периферийное устройство для персонального компьютера.

Синтезатор Компьюкордер (рис. 7.6) позволяет получать высококачественную речь на любом языке, наговоренную как мужским, так и женским голосом. Цифровое устройство записи обладает также способностью хранить и воспроизводить музыку и звуковые эффекты. Словарь Компьюкордера формируется самим пользователем, который просто наговаривает его в

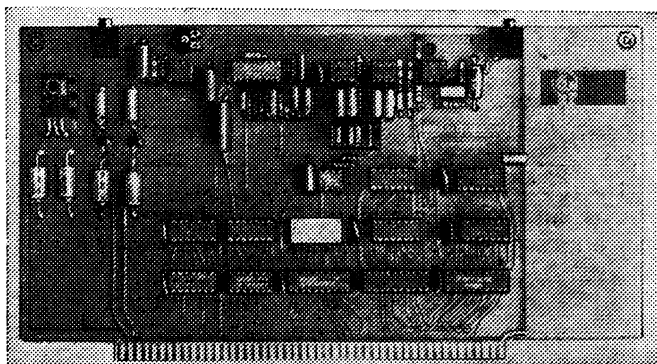


Рис. 7.6. Синтезатор речи Компьюкордер фирмы «Компьютокер консалтантс». (С разрешения фирмы «Компьютокер консалтантс».)

подключаемый к синтезатору микрофон. Затем наговоренный словарь подвергается дискретизации методом сжатия данных, причем соответствующие средства размещены на схемной плате синтезатора Компьюкордер. Получаемые речевые данные в компактной форме поступают на хранение в память компьютера, и их можно воспроизводить по команде программного управления. Помимо микрофона пользователь должен дополнить эту систему своим звуковым усилителем и громкоговорителем, которые обеспечивают громкость, достаточную для обычного помещения.

При работе с устройством Компьюкордер имеется возможность пойти на определенный компромисс, согласовав удобным образом требования к качеству речевого выхода и необходимому для запоминания

речи объему памяти. Пользователь может выбрать следующие расходы памяти в расчете на одну секунду речи: 1,25, 2, 3 или 4 кбайт. В зависимости от объема ЗУ компьютера можно остановиться на приемлемом для вас удельном расходе памяти для воспроизведения речи. При использовании для односекундной речи объема памяти 4 кбайт качество воспроизведения приближается к качеству воспроизведения речи с магнитной лентой. При этом эффективная скорость дискретизации равна 32 000 бит/с. Компьютер с памятью 64 кбайт при таком темпе воспроизведения может хранить речь длительностью примерно 16 с и обеспечивает «кристально чистый» речевой выход с абсолютно естественным звучанием. Если компьютер оснастить накопителем на гибком диске для хранения речи, длительность речевого выхода соответственно возрастет. Программы, поставляемые для синтезатора Компьюкордер, написаны на ассемблере для микропроцессора 8080; запись осуществляется под управлением дисковой операционной системы CP/M фирмы «Диджитал рисерч инкорпорейтид». Используются диски и другого формата. В силу своей невысокой стоимости синтезатор Компьюкордер идеален для пользователей, которые готовы самостоятельно формировать свои словари.

Устройство Майкромаус

Интерфейс с речевым синтезатором типа Майкромаус выпускает фирма «Майкроминт инкорпорейтид», изготовляя его по конструкции, которая была предложена Стивом Сьярсиа в июньском номере журнала *Byte Magazine* за 1981 г. В синтезаторе использована микросхема Диджитокер фирмы «Нэшнл семикондактор», созданная для синтеза речи по методу дельта-модуляции. Хотя синтезатор речи Майкромаус рассчитан на совместную работу с компьютерами Эппл II и TRS-80, он может быть сопряжен и с другими компьютерами, если воспользоваться параллельными портами для передачи данных.

Плата синтезатора речи Майкромаус показана на рис. 7.7. Его одноплатный вариант представляет собой съемное периферийное устройство, которое вставляется непосредственно в гнездо компьютера Эппл II. Другой вариант того же устройства в виде автономного

блока поставляется как интерфейс для компьютера TRS-80. В составе последнего варианта имеется интерфейсный ленточный кабель для подключения к соединителю расширительной шины модели I или III компьютера TRS-80. Словарь синтезатора речи Майкромаус размещен на двух ПЗУ емкостью по 64 кбит. Состав словаря приведен в табл. 7.2. Словарь с таким же составом слов, размещенный на двух предварительно записанных словарных ПЗУ, изготавливает и фирма «Нэшнл семикондактор».

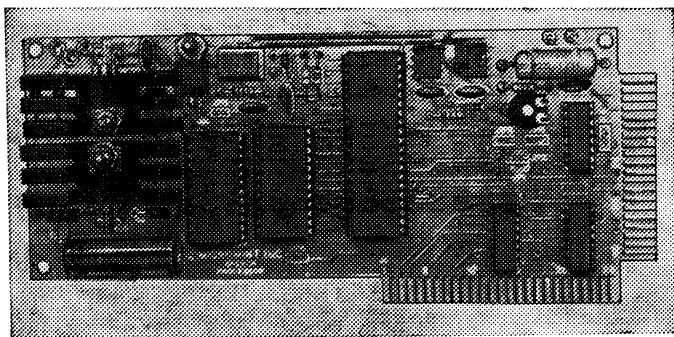


Рис. 7.7. Синтезатор речи Майкромаус фирмы «Майкроминт». (С разрешения фирмы «Майкроминт инк.».)

Для выборки слов, хранимых в памяти речевого синтезатора, компьютеры пользуются оператором роке или оператором out. Когда в устройство Майкромаус подается выходной код, соответствующий какому-то слову из словаря, синтезатор по линии сигнала готовности уведомляет компьютер, что система еще занята произнесением предыдущей фразы. В конце периода воспроизведения речи синтезатор сигнализирует компьютеру, что он готов принять следующий 8-битовый адресный байт для своего словаря. Благодаря такой работе служебное машинное время, необходимое для создания речевой программы, весьма мало. Так, например, чтобы сказать английскую фразу *At the mark, the time is 2 : 45 PM (beep)*¹,

¹ У метки время 2 ч 45 мин пополудни ... (бип.)— Прим. перев.

Таблица 7.2. Словарь речевого синтезатора Диджитокер

Слово	8-разрядный двоичный адрес SW8—SW1	Слово	8-разрядный двоичный адрес SW8—SW1	Слово	8-разрядный двоичный адрес SW8—SW1
THIS IS DIGITALKER	00000000	Q	00110000	IS	01100000
ONE	00000001	R	00110001	IT	01100001
TWO	00000010	S	00110010	KILO	01100010
THREE	00000011	T	00110011	LEFT	01100011
FOUR	00000100	U	00110100	LESS	01100100
FIVE	00000101	V	00110101	LESSER	01100101
SIX	00000110	W	00110110	LIMIT	01100110
SEVEN	00000111	X	00110111	LOW	01100111
EIGHT	00001000	Y	00111000	LOWER	01101000
NINE	00001001	Z	00111001	MARK	01101001
TEN	00001010	AGAIN	00111010	METER	01101010
ELEVEN	00001011	AMPERE	00111011	MILE	01101011
TWELVE	00001100	AND	00111100	MILLI	01101100
THIRTEEN	00001101	AT	00111101	MINUS	01101101
FOURTEEN	00001110	CANCEL	00111110	MINUTE	01101110
FIFTEEN	00001111	CASE	00111111	NEAR	01101111
SIXTEEN	00010000	CENT	01000000	NUMBER	01110000
SEVENTEEN	00010001	400 HERTZ TONE	01000001	OF	01110001
EIGHTEEN	00010010	80 HERTZ TONE	01000010	OFF	01110010
NINETEEN	00010011	20 MS SILENCE	01000011	ON	01110011
TWENTY	00010100	40 MS SILENCE	01000100	OUT	01110100
THIRTY	00010101	80 MS SILENCE	01000101	OVER	01110101
FORTY	00010110	160 MS SILENCE	01000110	PARENTHESIS	01110110
FIFTY	00010111	320 MS SILENCE	01000111	PERCENT	01110111

SIXTY	00011000	CENTI	01001000	PLUS	01111001
SEVENTY	00011001	CHECK	01001001	POINT	01111010
EIGHTY	00011010	COMMA	01001010	POUND	01111011
NINETY	00011011	CONTROL	01001011	PULSES	01111100
HUNDRED	00011100	DANGER	01001100	RATE	01111101
THOUSAND	00011101	DEGREE	01001101	RE	01111110
MILLION	00011110	DOLLAR	01001110	READY	01111111
ZERO	00011111	DOWN	01001111	RIGHT	10000000
A	00100000	EQUAL	01010000	SS *	10000001
B	00100001	ERROR	01010001	SECOND	10000010
C	00100010	FEET	01010010	SET	10000011
D	00100011	FLOW	01010011	SPACE	10000100
E	00100100	FUEL	01010100	SPEED	10000101
F	00100101	GALLON	01010101	STAR	10000110
G	00100110	GO	01010110	START	10000111
H	00100111	GRAM	01010111	STOP	10001000
I	00101000	GREAT	01011000	THAN	10001001
J	00101001	GREATER	01011001	THE	10001010
K	00101010	HAVE	01011010	TIME	10001011
L	00101011	HIGH	01011011	TRY	10001100
M	00101100	HIGHER	01011100	UP	10001101
N	00101101	HOUR	01011101	VOLT	10001110
O	00101110	IN	01011110	WEIGHT **	10001111
P	00101111	INCHES	01011111		

* SS преобразует любое слово в единственном числе во множественное число.

** Адрес 143 — последний законный адрес в данном списке слов. Использование адреса с ббльшим номером приводит к генерации неразборчивой, бессмысленной речи.

требуется всего 15 обращений к синтезатору. Каждое обращение связано с получением доступа к конкретному слову, числу или пустому интервалу, так что в процессе воспроизведения речи у компьютера имеется

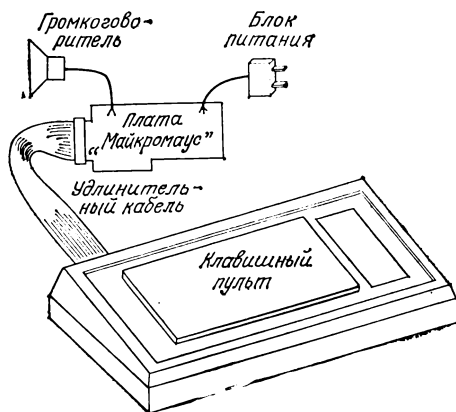


Рис. 7.8. Вариант синтезатора Майкромаус, предназначенный для работы с компьютером TRS-80. (С разрешения фирмы «Майкроминт инк.»)

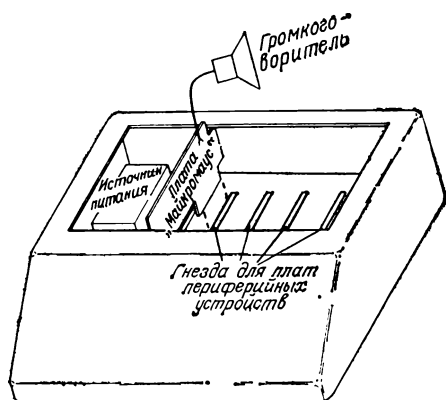


Рис. 7.9. Вариант синтезатора Майкромаус, рассчитанный на работу с компьютером Эппл II. (С разрешения фирмы «Майкроминт инк.»)

достаточно времени для выполнения других операций. Майкромаус — очень простой и вместе с тем эффективный синтезатор речи. На рис. 7.8 и 7.9 показано, как производить установку этого синтезатора для работы с компьютерами TRS-80 или Эппл II.

Синтезатор Диджитокер фирмы «Нэшнл семикондактор»

Система синтеза речи Диджитокер выпускается изготовителем в нескольких различных конфигурациях. Центральной частью системы является речевой процессор MM 54104, выполненный в виде интегральной схемы. Эта микросхема, насчитывающая 40 внешних

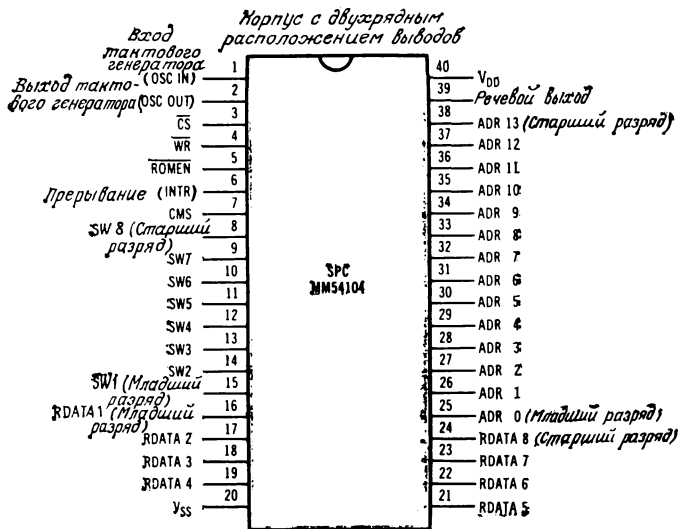


Рис. 7.10. Распределение выводов микросхемы (вид сверху) речевого процессора MM 54104, используемого в синтезаторе Диджитокер фирмы «Нэшнл семикондактор». (С разрешения фирмы «Нэшнл семикондактор корп.»)

выводов, изображена на рис. 7.10. Структурная схема речевого процессора MM 54104 показана на рис. 7.11. Сложность микросхемы синтезатора речи Диджитокер, о которой можно судить по этой структурной схеме, отражает модификации, которые были внесены фирмой «Нэшнл семикондактор» в стандартный процесс синтеза по методу дельта-модуляции. Хотя в схеме имеется блок, названный декодером демодулятора, функции дельта-демодуляции осуществляются вспомогательными регистрами и усилителями. Эти периферийные регистры снижают эффективную скорость передачи данных в синтезаторе и, следовательно, уменьшают

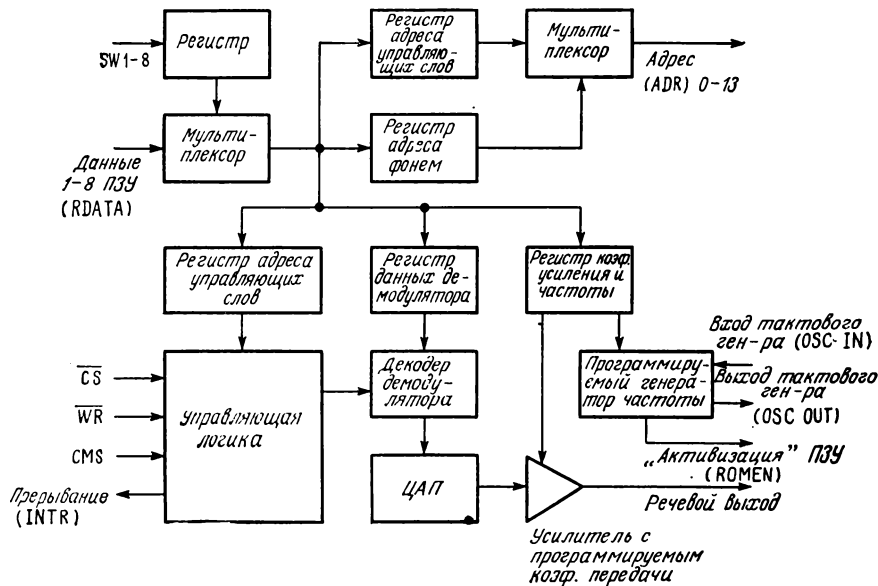


Рис. 7.11. Структурная схема речевого процессора MM 54104.

необходимую емкость памяти. Выводы микросхемы речевого процессора, обозначенные SW1—8,— это входная 8-разрядная шина стартового адреса, которая обеспечивает возможность считывания из речевого

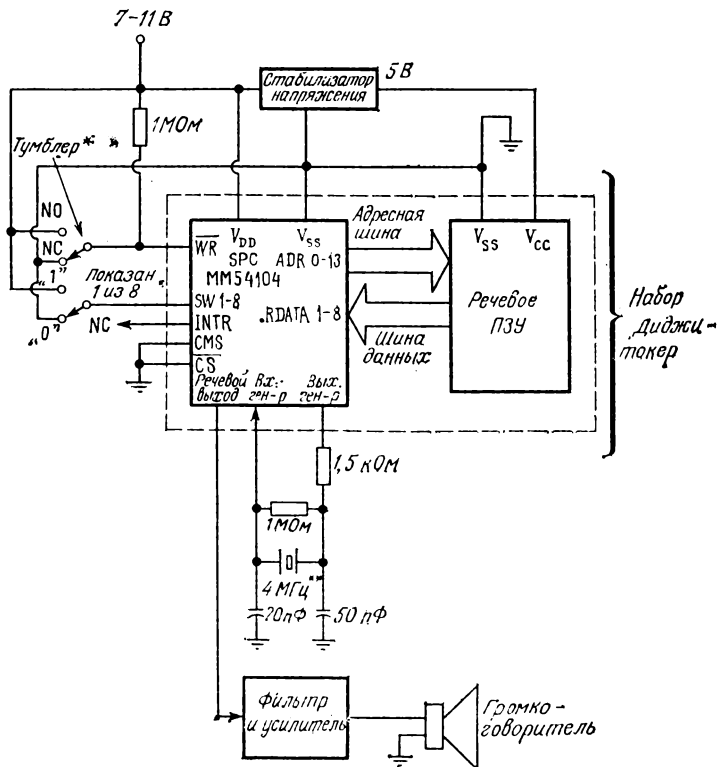


Рис. 7.12. Структурная схема речевого синтезатора Диджитокер с управлением от переключателей. (С разрешения фирмы «Нэшнл семикондактор корп.».)

* Тумблер однополюсный на два направления.

** Кварцевый резонатор (4 МГц) НС 18 фирмы «Электродайнемикс корп.» емкостью 20 пФ.

ПЗУ до 256 отдельных звуков или выражений. Путем подачи 8-разрядного слова на эти порты и установки на линии WR высокого уровня производится обращение к наговоренному слову или фразе, и начинается процесс воспроизведения речи.

D0 — D7 передают стартовые адреса слова или фразы. Линия WR эквивалентна линии чтение/запись (R/W), сигнал которой используется для стробирования данных, подлежащих вводу в регистр управления речью. Линия A0 — это линия выбора команды, которая определяет способ, используемый микросхемой речевого процессора для обработки сигнала прерывания (INTR) перед его возвратом в микропроцессор. Сигнал с вывода «выбор микросхемы» (CS) производит подготовку микросхемы речевого процессора к приему команд микропроцессора в виде декодированных адресов ПЗУ.

Помимо базовой микросхемы речевого процессора фирма «Нэшнл семикондактор» предлагает несколько словарных ПЗУ с предварительно записанными в них (в закодированной форме) словарями для различных применений. Стандартный словарный набор DT 1050 синтезатора Диджитокер включает два 64-кбитовых ПЗУ, которые содержат такой же словарь, какой приведен в табл. 7.2. Поскольку в синтезаторе Майкромаус используется тот же набор микросхем, эти словари одинаковы. Другой набор микросхем, предлагаемый фирмой «Нэшнл семикондактор», содержит базовый числовой набор DT 1052 Диджитокер. Этот ограниченный словарный набор имеет в своем составе наговоренные числа от 0 до 9, «.» и пять различных по длительности пауз, вводимых в воспроизводимую речь.

Фирма «Нэшнл семикондактор» может также поставлять речевые ПЗУ типа DT 1057 Диджитокер с другим стандартным словарем (табл. 7.3). Слова, входящие в состав второго словарного набора, служат хорошим дополнением к словам практически для любого возможного приложения. Чтобы убедиться в этом, достаточно рассмотреть конкретный пример с какой-либо специфической фразой и установить, можно ли найти в этих словарях (табл. 7.2 и 7.3) нужные слова или эквивалентную им замену. Следует помнить, что, пользуясь синтезатором Диджитокер, можно не только произносить слова из этих словарей с необходимыми паузами между ними, но и связывать (конкатенировать) слова и выражения из словарей в более длинные слова. Так, если нужно произнести, скажем, слово *megohms*, то достаточно задать, руководствуясь

Таблица 7.3. Содержание словарных ПЗУ речевого синтезатора

Слово	8-разрядный двоичный адрес SW8 -- SW1	Слово	8-разрядный двоичный адрес SW8 -- SW1	Слово	8-разрядный двоичный адрес SW8 -- SW1
ABORT	00000000	FARAD	00101100	PER	01011000
ADD	00000001	FAST	00101101	PICO	01011001
ADJUST	00000010	FASTER	00101110	PLACE	01011010
ALARM	00000011	FIFTH	00101111	PRESS	01011011
ALERT	00000100	FIRE	00110000	PRESSURE	01011100
ALL	00000101	FIRST	00110001	QUARTER	01011101
ASK	00000110	FLOOR	00110010	RANGE	01011110
ASSISTANCE	00000111	FORWARD	00110011	REACH	01011111
ATTENTION	00001000	FROM	00110100	RECEIVE	01100000
BRAKE	00001001	GAS	00110101	RECORD	01100001
BUTTON	00001010	GET	00110110	REPLACE	01100010
BUY	00001011	GOING	00110111	REVERCE	01100011
CALL	00001100	HALF	00111000	ROOM	1100100
CAUTION	00001101	HELLO	00111001	SAFE	01100101
CHANGE	00001110	HELP	00111010	SECURE	01100110
CIRCUIT	00001111	HERTZ	00111011	SELECT	01100111
CLEAR	00010000	HOLD	00111100	SEND	01101000
CLOSE	00010001	INCORRECT	00111101	SERVICE	01101001
COMPLETE	00010010	INCREASE	00111110	SIDE	01101010
CONNECT	00010011	INTRUDER	00111111	SLOW	01101011
CONTINUE	00010100	JUST	01000000	SLOWER	01101100
COPY	00010101	KEY	01000001	SMOKE	01101101
CORRECT	00010110	LEVEL	01000010	SOUTH	01101110
DATE	00010111	LOAD	01000011	STATION	01101111
DAY	00011000	LOCK	01000100	SWITCH	01110000

DECREASE	00011001	MEG	01000101	SYSTEM	01110001
DEPOSIT	00011010	MEGA	01000110	TEST	01110010
DIAL	00011011	MICRO	01000111	TH **	01110011
DIVIDE	00011100	MORE	01001000	THANK	01110100
DOOR	00011101	MOVE	01001001	THIRD	01110101
EAST	00011110	NANO	01001010	THIS	01110110
ED *	00011111	NEED	01001011	TOTAL	01110111
ED *	00100000	NEXT	01001100	TURN	01111000
ED *	00100001	NO	01001101	USE	01111001
ED *	00100010	NORMAL	01001110	UTH ***	01111010
EMERGENCY	00100011	NORTH	01001111	WAITING	01111011
END	00100100	NOT	01010000	WARNING	01111100
ENTER	00100101	NOTICE	01010001	WATER	01111101
ENTRY	00100110	OHMS	01010010	WEST	01111110
ER	00100111	ONWARD	01010011	SWITCH	01111111
EVACUATE	00101000	OPEN	01010100	WINDOW	10000000
EXIT	00101001	OPERATOR	01010101	YES	10000001
FAIL	00101010	OR	01010110	ZONE	10000010 ****
FAILURE	00101011	PASS	01010111		

* ED — суффикс, который используется для преобразования глагольной формы настоящего времени в форму прошедшего времени. Однако способ, которым произносится этот суффикс, меняется от слова к слову. По этой причине предлагается четыре различных звуковых варианта, каждый из которых следует опробовать со словом, преобразуемым в прошедшее время, и тем самым найти лучший по качеству звучания результат. Суффиксы ED, расположенные на адресах 31 и 32, следует использовать со словами, заканчивающимися на согласные T или D соответственно, например со словами *exit* и *load*. Суффикс ED с адресом 34 следует использовать со словами, заканчивающимися на мягкий звук, например как в слове *ask*. Суффикс ED с адресом 33 следует применять со всеми остальными словами.

** Суффикс TH добавляется к таким количественным числительным, как *six*, *seven*, *eight*, для получения порядковых числительных *sixth*, *seventh*, *eightth* и т. д.

*** Суффикс UTH может быть использован для прибавления к таким количественным числительным, как *twenty*, *thirty*, *forty*, для получения порядковых числительных, подобных *twentieth*, *thirtieth*, *fortieth* и т. д.

**** Адрес 130 — последний законный адрес в данном конкретном списке слов. Использование адреса с бóльшим номером приводит к генерации неразборчивой, бессмысленной речи.

табл. 7.3, последовательность из адресов 45 (шестнадцатеричный) и 52 (шестнадцатеричный), не внося между ними никакого разделительного интервала — паузы.

Для тех, кто намерен изучить синтезатор речи Диджитокер, не вникая в сложности соединений между отдельными компонентами аппаратных средств, фирма «Нэшл семикондактор» выпускает законченную плату DT 1000, которая предназначена для исследования возможностей речевого синтеза. Общий вид платы

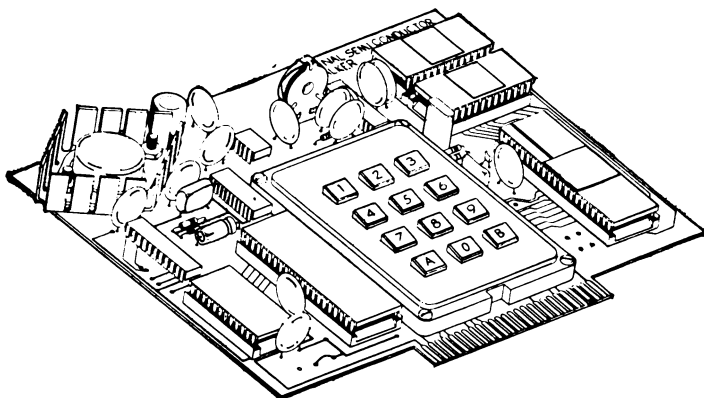


Рис. 7.14. Плата для макетирования синтезаторов речи DT1000 Диджитокер фирмы «Нэшл семикондактор корпорейшн». (С разрешения фирмы «Нэшл семикондактор корп.».)

показан на рис. 7.14. В состав платы входят: микросхема речевого процессора, два ПЗУ со словарями, состав которых приведен в табл. 7.2, выходной фильтр нижних частот и небольшой предварительно запрограммированный микропроцессор. Для работы с этой пробной платой синтезатора речи необходимы лишь источник питания с напряжением 6—11 В и недорогой громкоговоритель. Эта демонстрационная плата, отличающаяся исключительной гибкостью и простотой, очень удобна для практического использования набора микросхем Диджитокер. Система Диджитокер представляет собой одну из основных систем речевого синтеза в интегральном исполнении, имеющихся в продаже и построенных по методу прямого кодирования-восстановления речевого сигнала. Поскольку словари, необходимые для построения речи средствами

этой системы, закодированы специалистами фирмы «Нэшнл семикондактор», возможности для создания собственного словаря здесь довольно ограничены. Фирма «Нэшнл семикондактор» может выполнить кодирование заказного словаря по запросу пользователя, однако затраты на одно слово в таком случае будут весьма высокими. Следовательно, эта система наилуч-

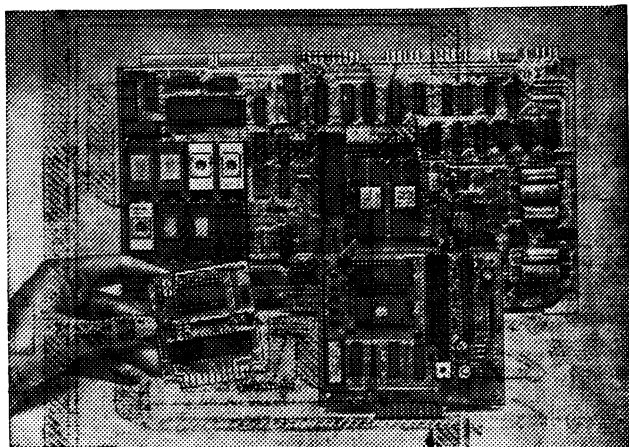


Рис. 7.15. Семейство синтезаторов речи фирмы «Телесенсориз».

шим образом подходит для таких приложений, в которых целесообразно использовать уже существующие постоянные словари.

Устройство Серисес III фирмы «Телесенсориз спич системс»

Модульный речевой синтезатор Серисес III относится к семейству речевых устройств, выпускаемых фирмой «Телесенсориз спич системс». На рис. 7.15 модуль Серисес III расположен внизу справа; он представляет собой печатную схемную плату среднего размера. Меньшая по размерам плата (показана внизу слева) — это снятая с производства система Серисес II, а большая плата — синтезаторная система LPC SPEECH 1000, описанная далее в данной главе.

Синтезаторный модуль Серис III — это схемная плата полноструктурного речевого синтезатора, предназначенного для использования в простейших прикладных системах. В состав синтезаторного модуля входят следующие компоненты: фирменный речевой синтезатор фирмы «Телесенсор», базовый словарь, насчитывающий 119 слов, речевой фильтр нижних частот и звуковой усилитель. Кроме того, в речевом

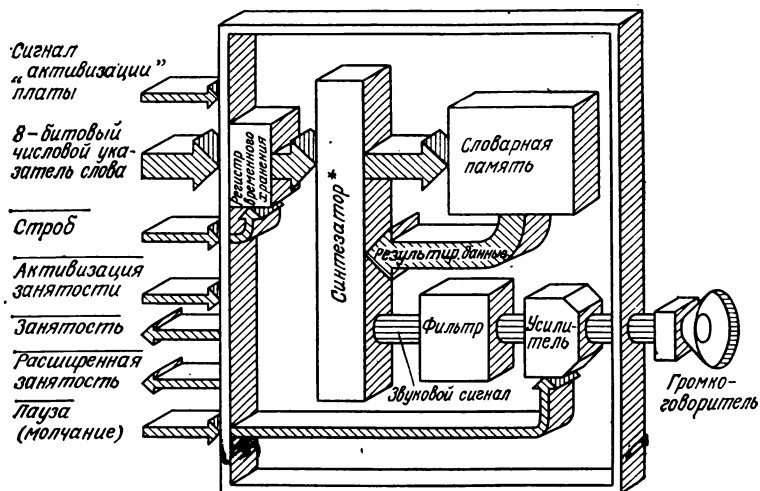


Рис. 7.16. Структурная схема синтезатора речи Серис III фирмы «Телесенсоры спич системс». (С разрешения фирмы «Телесенсоры спич системс».)

* Синтезатор на основе заказного контроллера ПЗУ фирмы «Телесенсоры спич системс».

модуле предусмотрено дополнительное гнездо для установки ПЗУ, которое позволяет вводить в систему дополнительный (заказной) словарь. Модуль работает от одного источника питания с напряжением +5 В и способен генерировать синтезированную речь общей длительностью примерно 100 с. Хотя фирма «Телесенсоры спич системс» держит в секрете используемый ею метод синтеза речи, она называет его методом речевого синтеза с помощью заказного контроллера ПЗУ, и, скорее всего, кодирование выполняется методом дельта-модуляции. Структурная схема синтезаторного модуля Серис III изображена на рис. 7.16. Поскольку

этот модуль является автономной системой для генерации речи, здесь требуется лишь командное слово и строб. Сигнал занятости, поступающий от модуля в управляющий внешний процессор, указывает последнему, что синтезатор продолжает речевой вывод. Для генерации речи систему можно использовать в сочетании с простейшим интерфейсом на переключателях (без микропроцессора). Такая конфигурация, которая может служить в качестве автономного синтезатора,

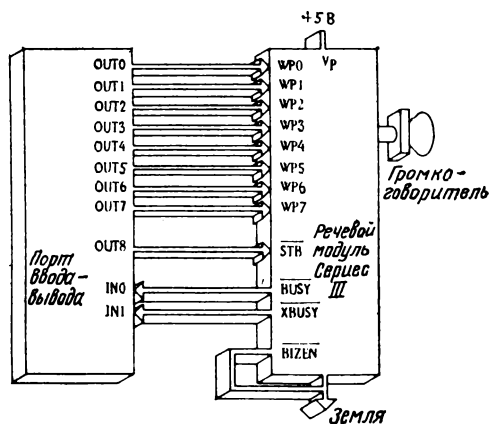


Рис. 7.17. Речевой синтезатор Серис III с управлением посредством входных переключателей. (С разрешения фирмы «Телесен-сори спич системс».)

показана на рис. 7.17. Для выбора любого из 256 возможных выражений в ней требуется в общей сложности девять полных переключателей.

Модуль Серис III может быть также подключен к параллельному микропроцессорному порту ввода-вывода. Это позволяет управлять произнесением отдельных речевых выражений с помощью управляющей программы. Соединение речевого модуля Серис III с интерфейсом микропроцессорного порта ввода-вывода изображено на рис. 7.18. Нетрудно видеть, что соединения, которые необходимо выполнить, чтобы приступить к генерации синтезированной речи под управлением компьютера, довольно просты.

Базовый словарь синтезаторного модуля Серис III содержит слова, достаточные для большей части при-

ложений, связанных с передачей информации о состояниях и численных значениях характеристик устройств и прикладных систем. Набрал соответствующий телефонный номер, вы можете услышать речь говорящих устройств фирмы «Телесенсоры», постоянно работающих в автоматическом режиме, и оценить качество их звучания. Эта демонстрация возможностей синтезаторов по телефону уникальна в том отношении, что позволяет прослушать не магнитную запись, а действительно генерируемую в процессе синтеза речь:

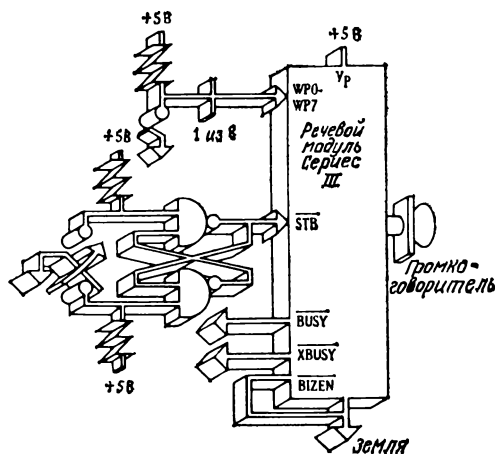


Рис. 7.18. Типичный микропроцессорный интерфейс синтезатора.

набрав номер, вы услышите, как говорит компьютер. Преимущество такого способа демонстрации качества говорящих устройств очевидно, поскольку она носит диалоговый характер. После ввода речевого синтезированного сообщения вам предлагается словесный набор кодов, которые вы можете вводить вручную при помощи тастатурного номеронабирателя. Вы можете выбрать любой из следующих кодов:

- 1 — повторить набор
- 2 или 3 — демонстрация преобразования в синтезированную речь
- 4 или 5 — демонстрация возможностей устройства LPC Speech 1000
- 6 или 7 — демонстрация возможностей модуля Серис III

Подобный способ демонстрации открывает довольно необычные возможности, позволяя в ходе диалога с компьютером определить, что вы слышите. Поскольку

телефонный номер синтезированного абонента функционирует под управлением компьютера, он отзывается почти 24 часа в сутки и вы можете воспользоваться им в любое удобное для вас время.

Синтезатор Когнивокс фирмы «Войстек»

Речевой синтезатор Когнивокс одновременно служит и для распознавания речи. Это периферийное устройство голосового ввода-вывода серии VIO предназначено для использования с большинством современных персональных компьютеров. Ниже перечислены компьютеры, с которыми может работать синтезатор Когнивокс:

Модель Когнивокс	Компьютер
VIO-1001	AIM-65 фирмы «Роквелл»
VIO-1002	ПЕТ фирмы «Коммодор»
VIO-1003	Эппл II фирмы «Эппл»
VIO-132	Сорсерер фирмы «Эксиди»
VIO-232	Системы на основе Z80:
VIO-332	TRS-80, Левел II с 16-кбайтовой памятью
VIO-432	ПЕТ фирмы «Коммодор»

Периферийные устройства голосового ввода-вывода серии VIO-1000 при их использовании с перечисленными выше компьютерами отличаются превосходными характеристиками распознавания и высоким качеством речевого вывода. Остальные периферийные устройства серии VIO представляют собой более экономичные средства ввода-вывода с более низким качеством речевого воспроизведения.

Периферийные устройства серии VIO-1000 способны распознавать до 32 слов, объединенных в короткие фразы, и обладают аналогичной способностью к синтезу речи. Генерация голоса осуществляется на основе прямой цифровой записи голоса пользователя. Такой способ кодирования речевого сигнала открывает путь к построению пользователем собственных заказных словарей. Синтезаторные системы, как правило, поставляются в высококачественном корпусе, изготовленном литьем под давлением, в котором помимо синтезатора находятся встроенный звуковой усилитель и громкоговоритель с размером диффузора 76 мм. В комплект синтезатора включены также входной микрофон,

источник питания, кассета с программой и руководство по работе с синтезатором. Эта система весьма недорога, позволяет практически познакомиться с техникой синтеза и распознавания речи.

Аналоговые формантные синтезаторы речи

Следующую группу речевых синтезаторов составляют устройства, воспроизводящие голос человека по методу синтеза формантных частот. Поскольку эти устройства формируют свой выход не путем прямого восстановления речи, они говорят, как роботы. Однако такие устройства дают возможность пользоваться неограниченным словарем — и это очень часто отодвигает на задний план все их недостатки, связанные с неестественным звучанием речи. В частности, при определенной натренированности слушателей к восприятию фонетической формантной речи синтезаторы такого типа оказываются вполне приемлемыми и могут с успехом заменить синтезаторы речи других типов.

Хотя данный метод генерации синтезированной речи используется уже довольно давно, подобные устройства по-прежнему сохраняют свою конкурентоспособность. В табл. 7.4 перечислены имеющиеся в продаже формантные синтезаторы речи. Благодаря фонемному методу синтеза возможности этих синтезаторов весьма широки. Остановимся теперь подробнее на этих средствах аналогового моделирования голосового тракта человека.

Читающая машина фирмы «Курцвайл компьютер продактс»

Хотя читающая машина КРМ фирмы «Курцвайл» не является непосредственно компьютерным периферийным устройством и при ее проектировании такое использование не предусматривалось, эту машину можно считать важной вехой в развитии средств синтеза речи. КРМ — это комбинированная интерфейсная система, состоящая из оптического читающего устройства и синтезатора речи; она предназначена в первую очередь в качестве читающего аппарата для слепых.

Читающая машина КРМ обеспечивает прямой индивидуальный доступ к печатным и машинописным

Таблица 7.4. Характеристики выпускаемых формантных синтезаторов

Название и фирма-изготовитель	Тип синтеза	Размер	Речевые возможности	Тип сопряженного компьютера
Читающая машина КРМ фирмы «Курцвайл»	Формантный с фонемным возбуждением	Настольное устройство	Выдача в речевой форме текста от оптического читающего устройства	Последовательный порт типа RS-232
Свит токер фирмы «Майкроминт»	Синтез с помощью микросхемы SC-01A фирмы «Вотракс»	Печатная плата 76×102 мм	Словарь из неограниченного числа слов, составляемых из фонем	Эппл II, параллельный интерфейс типа TTL
Микросхема SC-01A фирмы «Вотракс»	Формантный с фонемным возбуждением	Корпус типа DIP (с двухрядным расположением выводов), 22 вывода	Словарь из неограниченного числа слов, составляемых из фонем	Любой
Тайп'н ток фирмы «Вотракс»	Формантный с фонемным возбуждением	203×127×76 мм	Преобразование печатного текста в речь	Последовательный порт типа RS-232

материалам слепым и людям с плохим зрением. Система KRM модели III довольно компактна и внешне напоминает стандартное множительное устройство копировального типа. Читающая машина обеспечивает фонетический синтез речи из информации, поступающей от оптической системы распознавания печатных и машинописных знаков; это наделяет ее неограниченными в словарном отношении возможностями речевого воспроизведения текста. Рис. 7.19 иллюстрирует использование машины KRM модели III. Книга укла-



Рис. 7.19. Работа с читающей машиной KRM модели III фирмы «Курцвайл компьютер продактс». (С разрешения фирмы «Курцвайл компьютер продактс».)

дывается на читающее сканирующее устройство считываемой страницей вниз. Система оптически отслеживает строки печатного текста и декодирует их, воспроизводя содержание книги в словесной форме. Небольшой пульт управления оснащен удобной панелью, которая дает пользователю возможность осуществлять временный останов машины, возвращаться на одну или несколько строк для их повторного прослушивания, пропускать очередные строки текста, находить нужное слово и заставлять машину произносить его а также производить начальную установку машины в соответствии с форматом книги. На этой маленькой

панели управления имеются также органы регулировки громкости, скорости чтения и высоты основного тона. Управляющее устройство размером $25 \times 230 \times 51$ мм содержит встроенный громкоговоритель, откуда и звучит преобразованный в фонетическую форму текст.

Все электронные блоки читающей машины фирмы «Курцвайл» показаны на фотографии рис. 7.20. Помимо панели управления (внизу слева) здесь изобра-

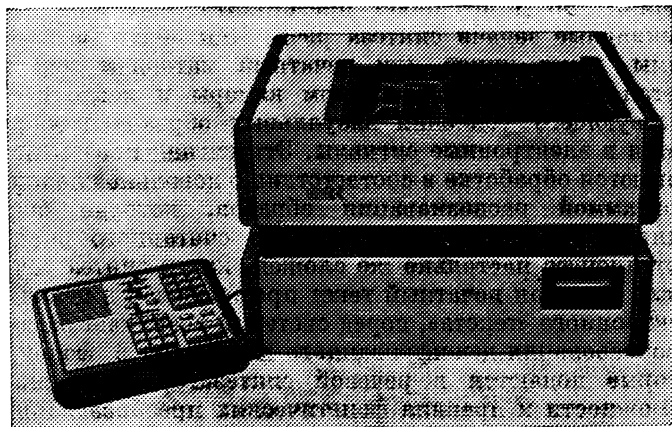


Рис. 7.20. Электронные блоки читающей машины KRM модели III. (С разрешения фирмы «Курцвайл компьютер продактс».)

жены два блока: оптическая следящая система и электронное устройство управления оптической сканирующей системой, причем первая из них установлена на втором. Для автоматического чтения документов размером 278×356 мм в оптической следящей системе используется камера в виде линейной матрицы сканирующих элементов. Электронное устройство управления для следящей системы включает схемы, необходимые для выполнения таких операций, как сканирование, распознавание знаков и преобразование текста в последовательности фонем, из которых складывается речевой выход. Чтобы придать читающей системе большую гибкость, фирма «Курцвайл» установила на передней панели электронного устройства управления цифровой кассетный накопитель на маг-

нитной ленте, что облегчает ввод в систему программных средств, необходимых для работы читающей машины.

Чтобы проиллюстрировать сложность процесса чтения, выполняемого машиной KRM, специалисты фирмы «Курцвайл компьютер продактс» составили блок-схему этого процесса, по которой можно легко проследить операции, реализуемые машиной. Последовательность действий, происходящих при чтении, отражена на рис. 7.21; эта диаграмма не только поясняет работу машины, но и позволяет представить, сколь сложны прикладные задачи синтеза речи. Как видно из блок-схемы, текст книги или печатный материал сканируется оптически посредством камеры и после соответствующего усиления визуальных образов преобразуется в электронные сигналы. Эти сигналы затем подвергаются обработке в соответствии с довольно сложной программой распознавания образов, которая обеспечивает идентификацию каждого считанного знака. Этот процесс настолько же сложен, как и синтез речи. После того как печатный текст преобразован в цепочки электронного «текста», перед системой возникает задача преобразования дешифрованного материала в фонемы, которые подаются в речевой синтезатор. Словарные возможности и правила фонетических преобразований, используемые в системе, довольно типичны для большинства синтезаторов, преобразующих печатный текст в речь. На выходе системы получается фонетически воспроизводимый текст, для восприятия которого требуется известный навык. Однако следует отметить, что способности этой системы в речевом воспроизведении текста поистине ни с чем не сравнимы. Эта система речевого синтеза, безусловно, преследует весьма благородные цели. Ее приложение показывает, насколько далеко ее возможности опережают возможности говорящих торговых автоматов.

Синтезатор Свит токер фирмы «Майкроминт»

Синтезатор Свит токер — второе периферийное устройство для синтеза речи, выпускаемое фирмой «Майкроминт инкорпорейтид». В то время как в ее первом устройстве — Майкромаус — для генерации речевого выхода используется набор микросхем Диджитокер,

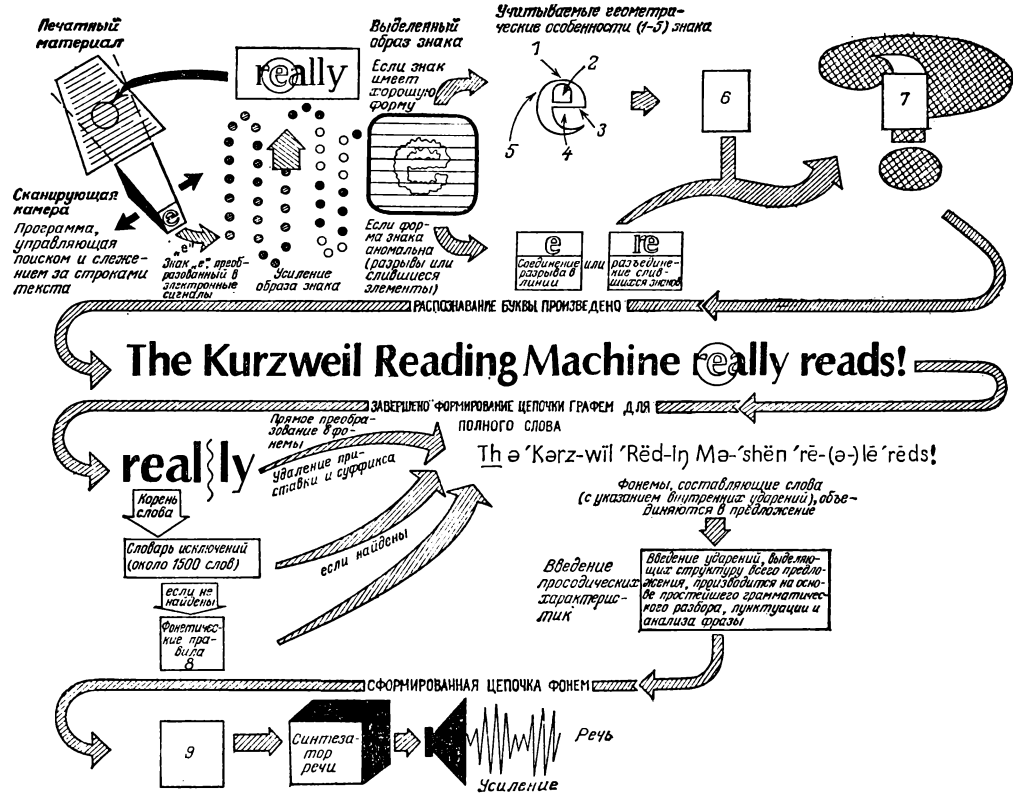
Рис. 7.21. Схема функционирования читающей машины KRM модели III. (С разрешения фирмы «Курцвайл компьютер продактс».)

1. Выпуклость вверх. 2. Замкнутый контур в верхней части. 3. Горизонтальная линия. 4. Вогнутость, обращенная вправо. 5. Выпуклость, обращенная влево. 6. Обучающийся блок. Запоминает вновь встретившиеся формы и проверяет правильность распознавания по шаблону, взятым из памяти заново выученного алфавита. 7. Блок устранения неоднозначности букв. Учитывает специфические свойства: необычные пропорции, контекст, разделение букв на слова, положение строки в тексте. 8. Фонетические правила. Разделение на слоги, контекст ударения в словах. 9. Правила речевого синтеза. Преобразования фонем в сигналы управления звуковым синтезатором (сигналы отражают такие параметры, как высота основного тона, ударение, длительность, межфонемные переходы).

АНАЛИЗ ДЛЯ РАСПОЗНАВАНИЯ ЗНАКОВ

ПРЕОБРАЗОВАТЕЛЬ В ФОНЕМЫ

ГЕНЕРАЦИЯ РЕЧИ



в устройстве Свит токер для этой цели применяется микросхема фонетического синтеза речи SC-01A Вотракс, выпускаемая отделением «Вотракс» фирмы «Федерал скрю уоркс». Синтезатор Свит токер представляет собой небольшую печатную плату, изготавливаемую в двух форматах. Синтезатор Свит токер, предназначенный для компьютера Эппл II, подключается непосредственно в гнездо для периферийного устройства на стандартном шасси компьютера Эппл II. Вто-

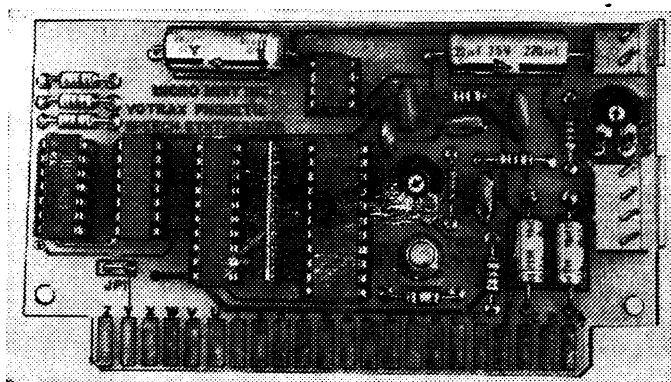


Рис. 7.22. Плата фонетического синтезатора Свит токер фирмы «Майкроминт». (С разрешения фирмы «Майкроминт инк.».)

рая плата синтезатора Свит токер спроектирована в расчете на сопряжение с любым компьютером, оснащенным параллельным TTL-портом. Поскольку в устройстве Свит токер используется микросхема фонетического синтеза Вотракс, в следующем разделе мы приведем характеристики этой схемы.

Первоначальная конструкция синтезатора Свит токер, спроектированная Стивом Сьярсиа, была описана в сентябрьском номере журнала *Byte Magazine* за 1981 г. Параллельный вариант схемной платы Свит токер изображен на рис. 7.22, который позволяет судить о размерах платы и простоте ее конструкции. Этот модуль синтезатора питается от источников с напряжениями +5 и +12 В. Обе схемные платы оснащены фильтрами звуковых частот и усилителями мощности с регуляторами громкости громкоговори-

телей. Поскольку оба варианта синтезатора Свит токер имеют фоновое возбуждение, их словари неограниченны, что связано с возможностью соединения фонем при образовании слов.

Микросхема фоновомого речевого синтезатора типа SC-01 фирмы «Вотракс»

Синтезатор речи SC-01, выпускаемый фирмой «Вотракс», представляет собой полностью автономную полупроводниковую интегральную схему. Конструкция этого

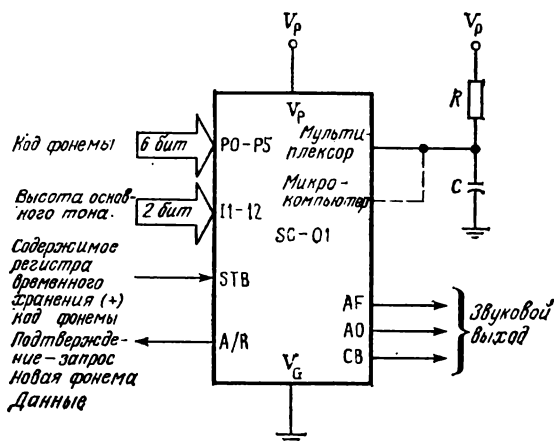


Рис. 7.23. Поток информации в процессе работы синтезатора SC-01 фирмы «Вотракс». (С разрешения фирмы «Вотракс корп.».)

построенного на одной микросхеме синтезатора такова, что позволяет ему — при поступлении на его вход цепочек фонем — фонетически синтезировать непрерывную связную речь. Входящая в речевой синтезатор SC-01 микросхема имеет встроенную память на 64 различные фонемы, доступ к которым осуществляется посредством 6-разрядного двоичного кода. Надлежащим образом выбирая последовательности фонем для ввода в синтезатор, на его выходе можно воспроизвести требуемые речевые образы.

Структурная схема синтезаторной микросхемы SC-01 с 22 выводами показана на рис. 7.23. В этой схеме отсутствуют сложные соединения с интерфейсом управляю-

щего компьютера. Шесть двоичных разрядов входа используются для адресации требуемой фонемы в памяти микросхемы, а еще два входных разряда — для выбора высоты основного тона запрашиваемой фонемы. Если управляющие данные на эти входы поступают со скоростью примерно 70 бит/с, синтезаторная схема заговорит. Поскольку интегральная схема спроектирована на основе К/МОП-технологии, полный потребляемый ею во время работы ток составляет всего 9 мА.

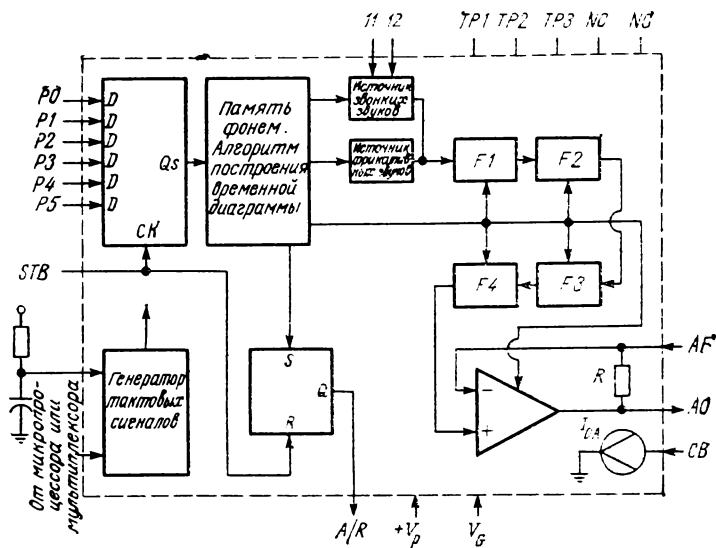


Рис. 7.24. Структурная схема внутренних блоков синтезатора SC-01 фирмы «Вотракс».

Эта особенность синтезаторной микросхемы в сочетании с совместимостью ее входов с ТТЛ-схемами делает ее идеальным устройством для сопряжения с любым персональным компьютером.

В действительности функционирование такого синтезатора в интегральном исполнении весьма сходно с работой формантных синтезаторов (см. гл. 6). Структуру этой микросхемы (рис. 7.24) полезно сравнить со схемами, рассмотренными в предыдущей главе, что дает возможность идентифицировать различные по функциональному назначению блоки. Явное различие между ними и данным синтезатором заключается в том, что в

последнем формантные фильтры F_1 , F_2 , F_3 и F_4 соединены последовательно, а не параллельно. Однако конечный результат в обоих случаях одинаков.

Кроме отдельной синтезаторной микросхемы фирма «Вотракс» выпускает также несколько различных модулей речевых синтезаторов, в которых эта микросхема используется для генерации речи. Первый из них — это модуль Спич ПЕК (контроллер доступа к фонемам), показанный на рис. 7.25. Он представляет собой относительно недорогую автономную систему с низкой скоростью передачи управляющих битов, в которой для гене-

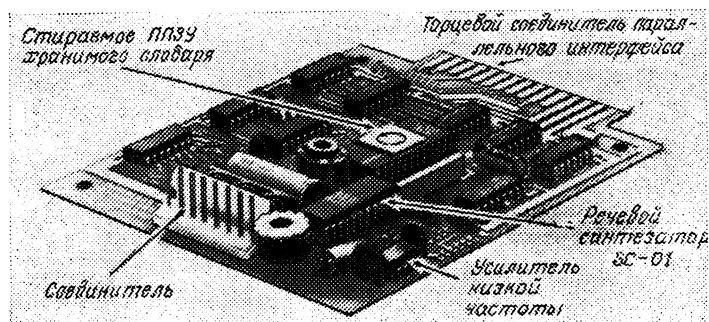


Рис. 7.25. Синтезаторный модуль Спич ПЕК фирмы «Вотракс». (С разрешения фирмы «Вотракс».)

рации речи используется встроенная память фонемных последовательностей и слов. Принятая конструкция контроллера позволяет системе хранить доступные для выборки слова в форме 8-байтовых приращений. Емкость встроенного в модуль стираемого программируемого ПЗУ типа 2716 предоставляет в распоряжение пользователя словарь из 255 слов. Кроме работы в режиме обращения к хранимому в памяти словарю, контроллер доступа к фонемам может также производить выборку отдельных фонем, вводимых для генерации фонетической речи внешним устройством. Это наделяет говорящую систему способностью работать с неограниченным словарем.

Структурная схема контроллера доступа к фонемам данного модуля (рис. 7.26) показывает схемные компоненты этой платы. Для генерации фонетической речи необходимые соединения с платой осуществляются

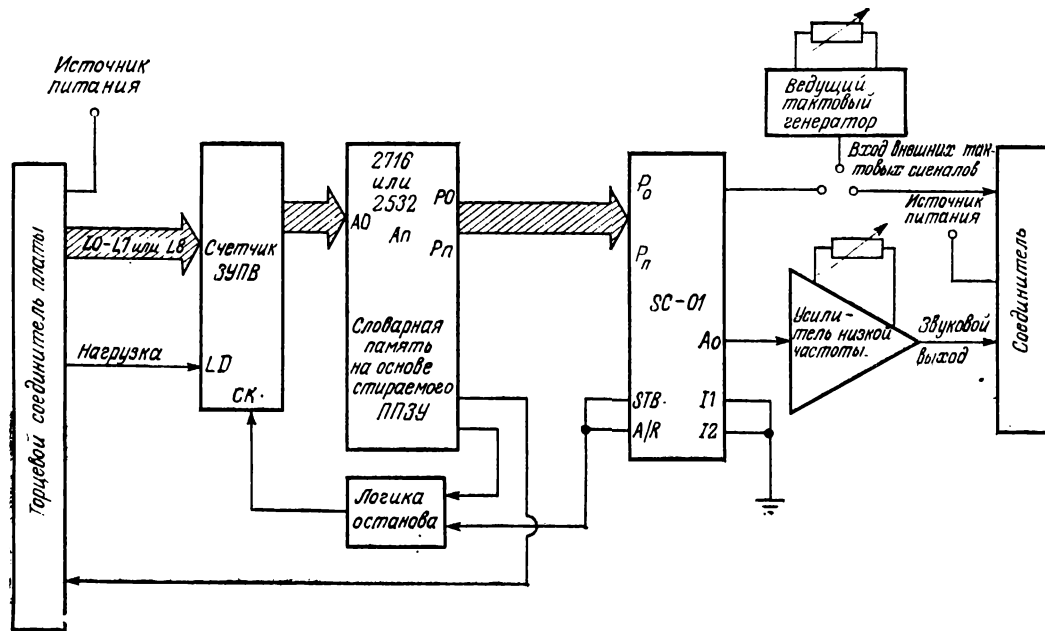


Рис. 7.26. Структурная схема межблочных соединений в синтезаторе Спич ПЕК фирмы «Вотракс». (С разрешения фирмы «Вотракс».)

непосредственно через ряд параллельных портов. В модуль Спич ПЕК входит, кроме того, звуковой усилитель, осуществляющий прямое возбуждение громкоговорителя.

Фирма «Вотракс» предлагает и другое автономное устройство — VSM (многоцелевой речевой модуль) типа речевого синтезатора, которое построено аналогично модулю Спич ПЕК, но дополнительно имеет встроенный микропроцессор. Это устройство, оснащенное

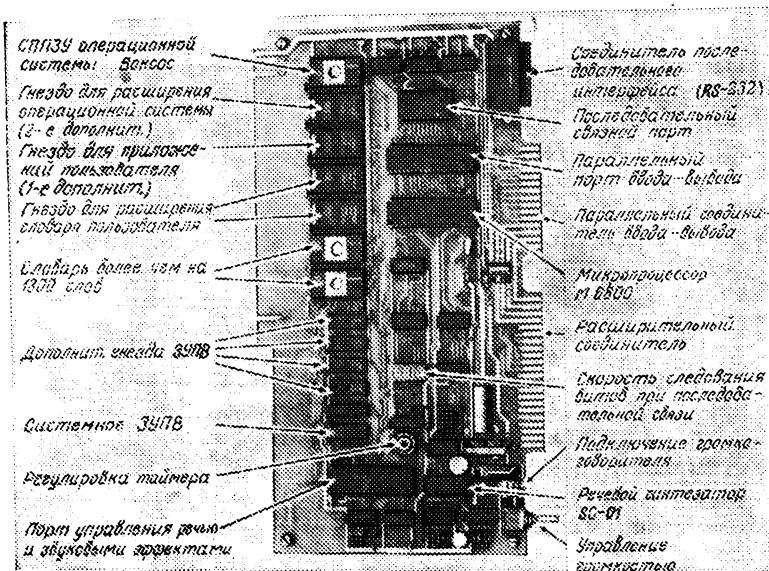


Рис. 7.27. Микрокомпьютерная плата синтезатора VSM/1 фирмы «Вотракс». (С разрешения фирмы «Вотракс».)

встроенным микропроцессором, можно использовать в качестве полноструктурного микрокомпьютера, позволяющего моделировать и разрабатывать говорящие системы. Встроенный процессор можно также запрограммировать на выполнение управляющих и контрольных функций, которые осуществляются одновременно с генерацией речи (неограниченной по словарному составу) в реальном масштабе времени.

Модуль VSM (рис. 7.27) содержит параллельный компьютерный порт — для непосредственного сопря-

жения с персональным компьютером и последовательный порт RS-232 — для непосредственного сопряжения с операторскими терминалами. Работа системы с речью организуется встроенной речевой операционной системой Воксос. На фотографии модуля VSM видны свободные гнезда, которые предусматривают возможность расширения как операционной системы, так и автономного словаря. Из сказанного ясно, что благодаря автономному микропроцессору (M 6800) возможности этого синтеза модуля речи весьма значительны.

Приведенная на рис. 7.28 структурная схема одноплатаго устройства, сочетающего компьютер с речевым синтезатором, дает представление о широком применении периферийных схем, сопряженных с базовой микросхемой синтезатора речи SC-01. Когда в распоряжение синтезатора речи выделяется определенный ресурс вычислительной мощности, его возможности становятся практически безграничными. В этом случае объем памяти для хранения программных средств и словаря, находящегося под управлением пользователя, довольно велик, благодаря чему подобную систему можно использовать для решения сложных прикладных задач. Входящая в состав операционных программных средств программа позволяет осуществлять межкомпьютерную передачу произносимых голосом команд, что, по сути дела, обеспечивает регистрацию речевого выхода. В результате ведущий компьютер может управлять периферийными синтезаторами речи во многом аналогично тому, как он управляет буферизованным принтером. Хотя цена такого устройства относительно высока в сравнении со стоимостью большинства других периферийных модулей, пригодных для синтеза речи с помощью персональных компьютеров, его применение экономически весьма выгодно.

Модуль VSM может располагать встроенным словарем на 1300 слов. Кроме того, в его памяти можно хранить записи различных звуковых эффектов и макрорекордации для воспроизведения записанных звуков. Хотя эта система и не способна преобразовывать печатный текст в речь, ее возможные приложения бесчисленны. Поскольку фирма «Вотракс» разработала автономный модуль Тайп'н ток — синтезатор речи, способный осуществлять преобразование печатного текста в речь, — далее мы рассмотрим его.

Речевой синтезатор Тайп'н ток фирмы «Вотракс»

Синтезатор Тайп'н ток фирмы «Вотракс» сочетает в себе ранее описанные устройства с алгоритмом преобразования печатного текста в речь. Как показано на рис. 7.29, устройство Тайп'н ток сконструировано в виде автономного синтезатора, который можно подключать в



Рис. 7.29. Речевой синтезатор Тайп'н ток фирмы «Вотракс». (С разрешения фирмы «Вотракс».)

качестве периферийного устройства к любому персональному компьютеру. Он оснащен встроенным источником питания, и все его рабочие связи осуществляются исключительно через последовательный порт RS-232C. Генерацию речи синтезатор производит в двух режимах: 1) при вводе английского текста и 2) при вводе фонем.

Поскольку этот синтезатор имеет собственный микропроцессор и буферную память объемом 750 знаков, для генерации речи с неограниченным словарем можно применять даже очень небольшие компьютеры. Процесс генерации речи с помощью синтезатора Тайп'н ток

столь же прост, как и действие последовательного построчно печатающего принтера. Действительно, синтезаторную систему в этом случае можно подсоединять к компьютеру непосредственно через последователь-

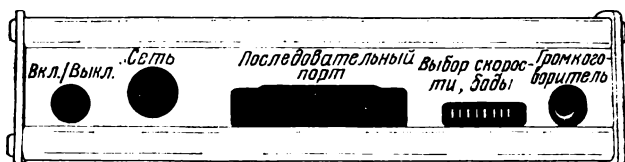


Рис. 7.30. Соединения синтезатора Тайп'н ток с компьютером. (С разрешения фирмы «Вотракс».)

ный порт принтера, генерируя речь вместо печатания текста.

Насколько просты соединения между синтезатором Тайп'н ток и персональным компьютером, можно видеть

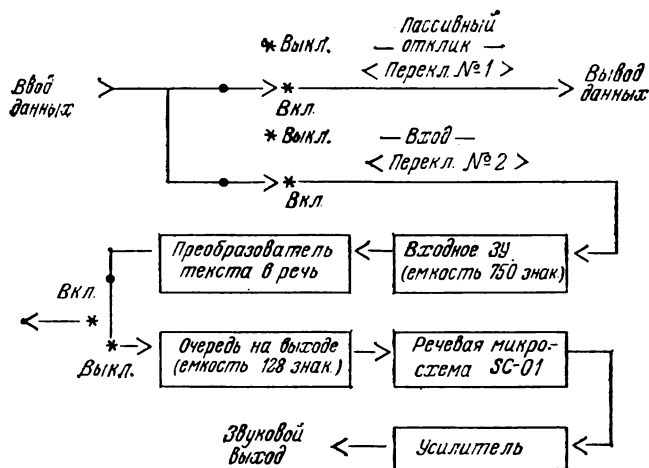


Рис. 7.31. Структурная схема синтезатора Тайп'н ток фирмы «Вотракс». (С разрешения фирмы «Вотракс».)

на рис. 7.30. На рисунке показаны соединительные разъемы и органы управления, установленные на задней панели компьютера. Слева на панели расположены выключатель питания и соединительный разъем для подключения кабеля питания. В центре задней панели

Таблица 7.5. Технические характеристики синтезатора Тайп'н ток

Физические характеристики	
Ширина	193 мм
Глубина	123 мм
Высота	65 мм
Масса	0,896 кг
Допустимые окружающие условия	
Температура	
При работе	от 4 до 38°C
При хранении	от -37 до +65°C
Влажность (без конденсации)	
При работе	от 20 до 90%
При хранении	от 5 до 95%
Электрические характеристики	
Вход трансформатора *	120 В, 60 Гц, 20 Вт
Вход TNT *	выводы 1—3: 26 В пер. тока, 180 мА вывод 2: 20 В пост. тока, 50 мА вывод 4: опорный уровень 0 В
Интерфейс	совместим с RS-232C, 75—9600 бод кадровые биты: стартовый—1; данные—8; стоповый—1
Звуковой выход	1 Вт на громкоговорителе с входн. сопротивлением 8 Ом, связь по пере- менному току

* Используется только трансформатор фирмы «Вотракс»—компонент № 01р-1224 В

С разрешения фирмы «Вотракс»

установлен соединительный разъем последовательного порта RS-232, а также переключатель скорости передачи, который позволяет изменять ее значение в пределах 75—9600 бод. На правом краю панели установлен соединительный разъем для подключения громкоговорителя, в качестве которого может использоваться любой стандартный динамик с входным сопротивлением 8 Ом.

Внутреннее устройство синтезатора Тайп'н ток фирма «Вотракс» хранит в секрете, и поэтому мало что известно о фактическом исполнении схем, применяемых в этом устройстве. На структурной схеме (рис. 7.31) показаны лишь основные рабочие функции синтезатора, отражающие принципиальный механизм его работы.

Анализ этой схемы показывает, что система может служить как периферийным устройством с пассивным откликом (прозрачным), так и прямым интерфейсом с компьютерным управлением. В любом случае поступающие через интерфейс RS-232 последовательные данные от компьютера накапливаются во входном запоминающем устройстве емкостью 750 знаков текста. По мере введения данных преобразователь печатного текста в речь преобразует английский текст в коды фонемного управления, которые направляются на микросхему речевого синтезатора SC-01. После этого фонетически генерируемая речь усиливается и выводится через громкоговоритель. Благодаря сложности используемых программных средств синтезатор Тайп'н ток может принимать фонемы и печатный текст, обеспечивая пассивный отклик вводимой информации, а также воспринимать данные из знаков верхнего и нижнего регистров в различных информационных контекстах.

Технические характеристики синтезатора Тайп'н ток приведены в табл. 7.5. В ней отражены оперативные возможности системы при работе с любым персональным компьютером, а также требования к ней. Это весьма простое и вместе с тем мощное устройство, работающее по методу формантного синтеза речи. Оно обеспечивает неограниченный словарь, но его речь, как и у всех формантных синтезаторов, по звучанию напоминает голос робота.

Синтезаторы на основе цифрового моделирования голосового тракта

Все коммерческие синтезаторы речи, описываемые в данном разделе, генерируют речь на основе цифрового моделирования голосового тракта. Основными методами речевого синтеза в системах этого типа являются методы ЛПК и Паркор. ЛПК-синтезаторы речи разрабатываются главным образом в США, тогда как метод Паркор развивается в Японии. Одно из главных препятствий, которое японцам приходится преодолевать, чтобы обеспечить сбыт своих говорящих изделий в США, связано с генерацией правильно произносимых английских слов для словаря, хранимого в ПЗУ синтезаторов. Ситуация весьма уникальна. В то время как японцы, по всей видимости, смогут научиться читать и писать

Таблица 7.6. Характеристики выпускаемых речевых синтезаторов ЛПК-типа

Фирма-изготовитель и название	Тип синтеза	Размер	Речевые возможности	Тип сопряженного компьютера
1. «Хитати» HD61885 HD38880, 38881, 38882	Паркор 10-полюсный Паркор 10-полюсный	28-выводной корпус типа DIP 42-выводной корпус типа DIP	Словарный состав речи определяется содержимым заказно- го ПЗУ	Любой
2. «Стрит электро- никс» ЭКО II и ЭКО-Джи Пи	10-полюсное ЛПК («Тексас инстру- ментс») 10-полюсное ЛПК («Тексас инстру- ментс»)	50×178×127 мм (пла- та машины Эплл II)	Фонемный ЛПК-пре- образователь «текст — речь»	Эплл II (ЭКО II) Любой (ЭКО-Джи Пи)
3. «Спич технолоджи корп.» M410, VR/S100	12-полюсное ЛПК 12-полюсное ЛПК	Печатная плата 102×102 мм Печатная плата 127×254 мм	Постоянные ЗУ со стандартным слова- рем	Любой (M410) Сопряженный с ши- ной S-100 (VRS-100)

4. «Телесенсорн спич системс» Спич 1000	12-полюсное ЛПК	Печатная плата 178×305 мм	Словарное ПЗУ	RS-232: параллельный интерфейс на ТТЛ-схемах Любой с интерфейсом RS-232 Любой с интерфейсом RS-232, шина Малтибас
SP1020	12-полюсное ЛПК	432×457×76 мм	Словарное ПЗУ	
Проуз 2000	12-полюсное ЛПК	Печатная плата 178×305 мм	Преобразователь «текст — речь»	
5. «Тексас инструментс» Спик энд спелл	10-полюсное ЛПК	203×305×50 мм	Словарь не более 200 слов Внешнее ПЗУ	Микропроцессор 6502
TMS 5200	10-полюсное ЛПК	28-выводной корпус типа DIP	Преобразователь «речь — ЛПК» реального времени	
PASS	ЛПК-кодер	Переносное устройство		

английские тексты более бегло, чем это способны делать даже сами американцы, маловероятно, что их английская речь окажется приемлемой для наших словарных систем на ПЗУ. Словарь таких систем должны «наговаривать» либо американские дикторы, либо японские дикторы, говорящие по-английски без японского акцента.

В табл. 7.6 перечислены выпускаемые сейчас устройства речевого синтеза, основанные на цифровом моделировании голосового тракта. Хотя эти устройства разработаны преимущественно по методу линейного предиктивного кодирования, в таблице упоминается и метод Паркор; тем самым мы хотим подчеркнуть, что и в этой области наблюдается значительная активность. (Метод Паркор в своих основных чертах очень сходен с методом ЛПК.)

Предполагается, что в ближайшие годы список этих говорящих изделий будет быстро расти. Синтез речи по методу линейного предиктивного кодирования успешно конкурирует с другими методами синтеза речи, что объясняется его большой эффективностью и меньшими требованиями к скорости передачи управляющих данных. Кроме того, по своему качеству ЛПК-речь приближается к записанной речи.

Синтезаторы HD 61885 и HD 38880, 38881, 38882 фирмы «Хитати»

Хотя говорящие изделия фирмы «Хитати» не просто приобрести из-за высокой стоимости, описывая их здесь, мы хотим познакомить читателя с общим положением дел в области генерации речи по методу Паркор. Фирма «Хитати» выпускает два основных типа синтезаторов речи в интегральной форме. Первый из них — речевой синтезатор HD 61885, К/МОП-микросхема на одном кристалле. В корпусе с 28 выводами размещены Паркор-синтезатор речи и 32-кбитовое ПЗУ, служащее речевым накопителем. Кроме того, в состав устройства входят цифро-аналоговый преобразователь и интерфейс клавишного пульта, используемого для ввода информации. Емкость ПЗУ, расположенного на кристалле, эквивалентна 26 секундам речи. В среднем это соответствует словарю из 63 слов. Добавление к микросхеме синтезатора внешней памяти увеличивает

возможную длительность синтезируемой речи на 100 с в расчете на одно ПЗУ. Поскольку к синтезатору можно добавить еще шестнадцать ПЗУ, он наделен очень широкими в словарном отношении речевыми возможностями.

Синтезатор второго типа (также в интегральном исполнении), выпускаемый фирмой «Хитати», — это микросхема HD 38880. Как и синтезатор HD 61885, он работает по методу Паркор, реализующему линейное предиктивное кодирование, однако построен на основе кремниевой микросхемы Р-МОП-типа. Основное различие между этой схемой и К/МОП-микросхемой состоит в том, что она потребляет большую мощность. Возможности хранения речи у синтезатора HD 38880 (в виде микросхемы) такие же, как и у К/МОП-варианта. Фактически, если не считать различий в полупроводниковой технологии микросхем, эти два синтезатора почти одинаковы. По принципу работы и степени сложности оба этих устройства очень сходны с синтезаторами, которые мы опишем дальше.

Синтезаторы ЭКО II и ЭКО-Джи Пи

Синтезаторы серии ЭКО, производимые фирмой «Стрит электроникс корпорейшн», можно считать первыми ЛПК-синтезаторами речи для персональных компьютеров, которые появились в продаже. Основной частью синтезаторов этой серии является речевой процессор TMS-5200 фирмы «Тексас инструментс». В настоящее время фирма «Стрит электроникс» изготавливает синтезаторы ЭКО двух вариантов: ЭКО II и ЭКО-Джи Пи. Модуль речевого синтезатора ЭКО II показан на рис. 7.32; это говорящее периферийное устройство, предназначенное для компьютера Эпл II, весьма компактно. Слева на фотографии схемной платы видны два свободных гнезда для ПЗУ, которые дают возможность ввести в модуль словарную память. Действуя по принципу фонетической генерации ЛПК-речи, синтезаторы серии ЭКО обладают неограниченным словарем — это отличительная черта всех формантных синтезаторов. Их речь также напоминает голос робота.

Одна из действительно привлекательных особенностей фонетического ЛПК-синтеза связана с тем, что режим построения цепочек из фонем здесь можно ком-

бинировать с режимом прямой записи ЛПК-речи. Это позволяет добиваться естественного звучания при воспроизведении распространенных слов и сохраняет возможность перехода на фонемный режим генерации речи при воспроизведении слов заказного словаря.

Кроме того, фирма «Стрит электроникс» ввела в синтезаторы серии ЭКО программу грамматического разбора текста Текстокер. Эта программа воспринимает стандартный выход в ASCII-коде, например следующий за оператором «Печатать», и преобразует его непосредственно в речевую форму. Поскольку программа Тек-

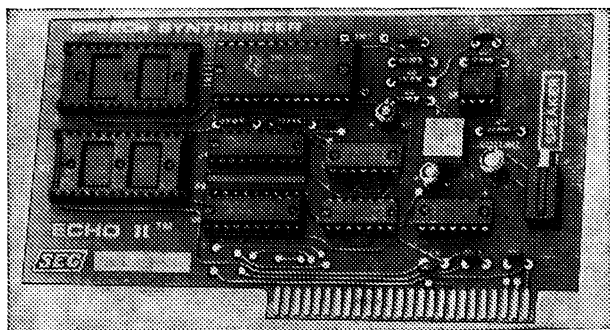


Рис. 7.32. ЛПК-синтезатор речи ЭКО фирмы «Стрит электроникс корпорейшн». (С разрешения фирмы «Стрит электроникс корп.».)

стокер содержит почти 400 правил английского произношения, она обеспечивает корректное речевое воспроизведение 96% из тысячи наиболее распространенных слов английского языка. Используя программу Текстокер, слова, представляющие собой исключения из правил английского произношения, можно корректировать двумя способами. Первый способ исправления произношения сводится к намеренно неправильному написанию слова, с тем чтобы синтезатор «клюнул» на эту уловку и дал правильный вариант, т. е. воспроизвел слово так, как оно звучит в действительности. Скажем, в тех случаях, когда в слове имеется произносимая буква, слово наиболее правильно произносится тогда, когда эта буква просто опускается из вводимого ASCII-текста. Например, чтобы правильно произнести слово *label*, написание его намеренно искажается и оно вводится как *lable*.

Помимо возможности преобразования печатного текста в речь синтезаторы серии ЭКО в состоянии также генерировать речь в режиме прямого фонемного возбуждения. Этот режим более сложен, поскольку здесь возникает необходимость в фонетической связи отдельных слов. Во многих случаях, однако, это единственный реальный путь к правильному произношению слов с необычным написанием.

Средствами для преобразования печатного текста в речь и генерации фонетической речи располагает также и синтезатор ЭКО-Джи Пи. Эта система имеет встроен-

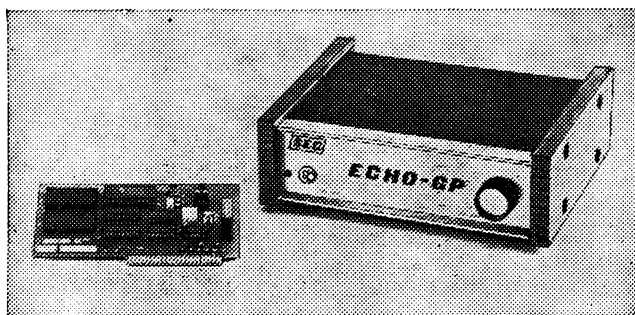


Рис. 7.33. ЛПК-синтезатор речи ЭКО-Джи Пи фирмы «Стрит электроникс корпорейшн». (С разрешения фирмы «Стрит электроникс корп.».)

ный микропроцессор и может использоваться как автономный синтезатор речи. Для сопряжения с компьютером, использующим данный синтезатор в качестве периферийного устройства, могут служить последовательный интерфейс типа RS-232 или параллельный TTL. Органы оперативного управления синтезатора ЭКО-Джи Пи (рис. 7.33) отличаются исключительной простотой и состоят всего лишь из выключателя питания и регулятора громкости. Синтезатор в основном используется точно так же, как модель ЭКО II, созданная для компьютера Эппл II. Различие между этими синтезаторами состоит только в том, что ЭКО-Джи Пи спроектирован в расчете на сопряжение через универсальный интерфейс с любым персональным компьютером. По своим возможностям в области генерации речи он не уступает ЭКО II, и в нем также использована программа Текстокер для генерации речевого вывода на основе преоб-

разования «печатный текст — речь». Введенный в модуль ЭКО-Джи Пи микропроцессор во время генерации синтезированной речи выполняет большинство служебных операций, тем самым снимая нагрузку с компьютера. Это, естественно, повышает стоимость синтезатора, так как она включает в себя и стоимость его дополнительного «интеллекта».

Устройства M410 и VR/S 100

В состав говорящих изделий, изготавливаемых фирмой «Спич технолоджи», входит микросхема ЛПК-синтезатора, выпускаемая фирмой «Дженерал инструментс». И это не удивительно, если учесть, что специалисты фирмы «Спич технолоджи» работали в тесном контакте с сотрудниками компании «Дженерал инструментс». Модуль речевого генератора M410 содержит 12-полюсную схему ЛПК-синтезатора речи. Относительно небольшая плата (рис. 7.34) рассчитана на сопряжение с параллельным TTL-портом компьютера любой микропроцессорной или микрокомпьютерной системы. Модуль разделен встроенным словарем объемом до 120 слов и автономным звуковым усилителем, рассчитанным на подключение громкоговорителя с входным сопротивлением 8 Ом. Он также оснащен микропроцессором 8039 с внутренним программированием, который предназначен для управления голосовой периферийной микросхемой. Словари для синтезатора M410 поставляет сама фирма «Спич технолоджи»; они строятся на стираемых ППЗУ типа 2716 и 2732. Помимо кодирования слов для стандартных словарей фирма «Спич технолоджи» за дополнительную плату производит также кодирование заказных словарей для любых речевых систем.

Чтобы оценить сложность схемы синтезатора M410, обратимся к рис. 7.35. Хотя большую часть данной структурной схемы занимают микропроцессор и сопряженные с ним компоненты памяти, здесь также представлена микросхема речевого синтезатора (U5), которая показана вместе с ее интерфейсом для микропроцессора и выходным звуковым усилителем. Программа для специального процессора, входящего в состав микропроцессора 8039, хранится в СППЗУ-памяти, занимая

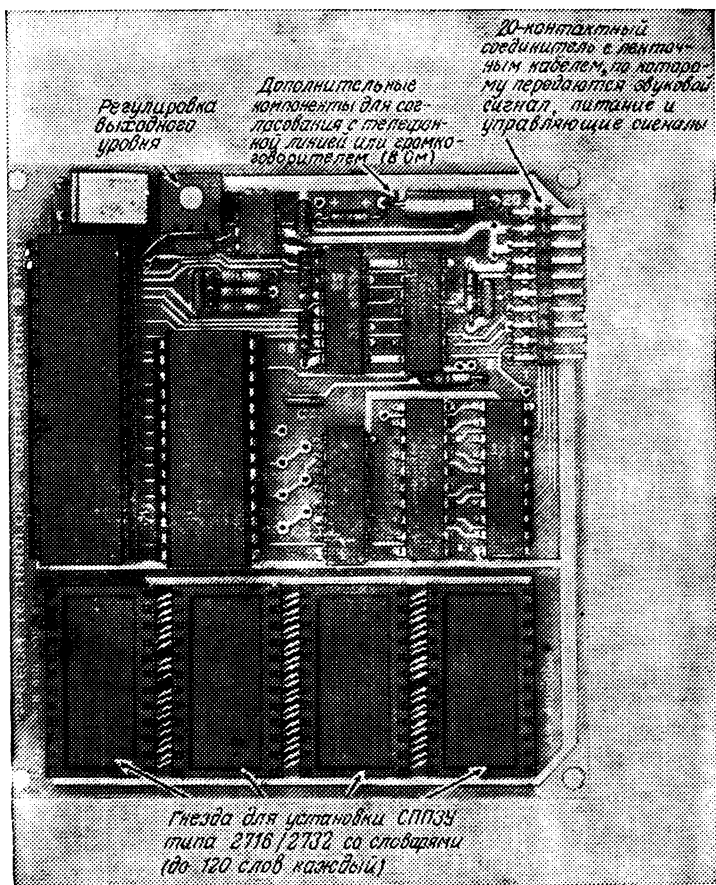


Рис. 7.34. Плата синтезатора речи М410 фирмы «Спич технолоджи корпорейшн». (С разрешения фирмы «Спич технолоджи корп.».)

ее первые примерно 900 байт. В этих адресных ячейках можно хранить данные словаря. Учитывая скромные размеры этого устройства, его можно отнести к очень мощным речевым синтезаторам.

Вторым периферийным синтезатором речи, предлагаемым фирмой «Спич технолоджи», является генератор речи типа VR/S100. Это более крупный модуль, который рассчитан на подключение к шине S-100 и обладает примерно теми же возможностями, как и синтезатор

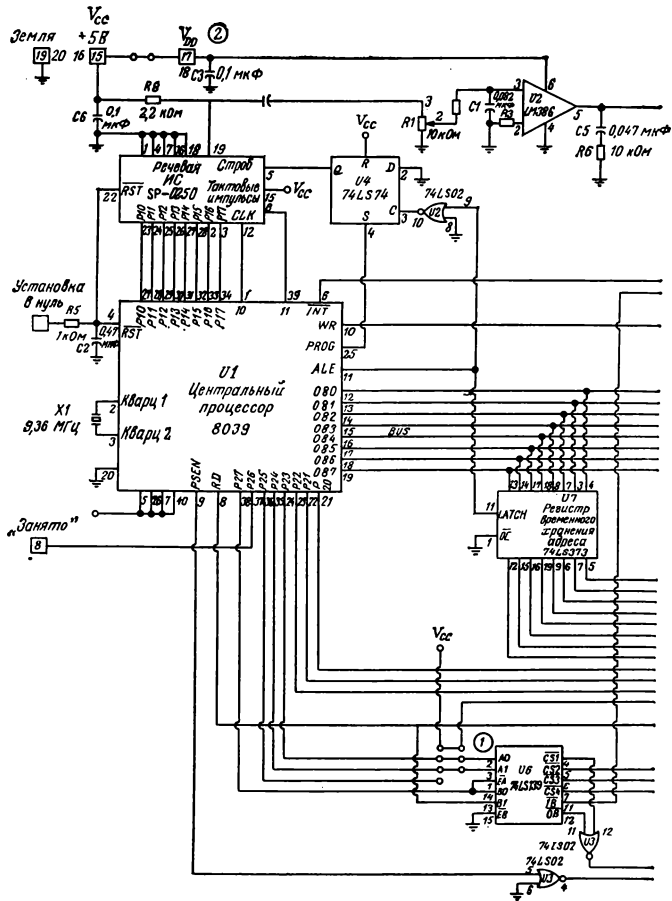
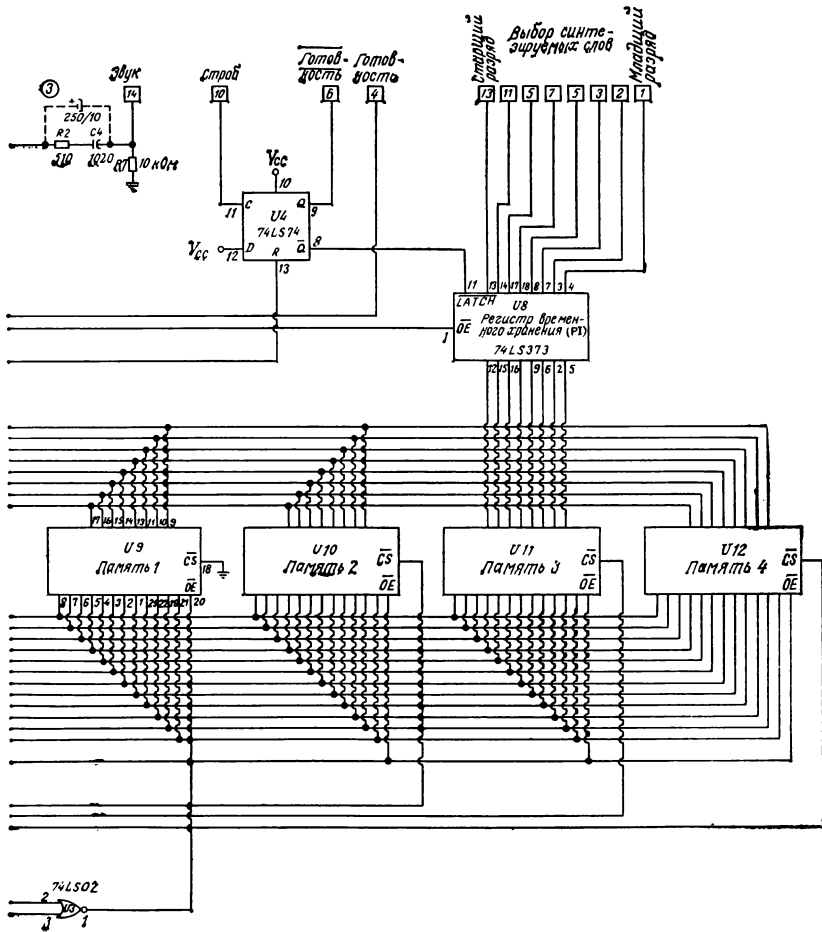


Рис. 7.35. Структурная схема синтезатора речи М410. (С раз 1. Выбор соединения для 2716/2732 (показано в положении 2716). получения независимого питания звуковой цепи (6—10 В) сее для громкоговорителя (8 Ом) устанавливается альтернативный

М410. Важнейшее различие между этими двумя устройствами связано с тем, что в синтезаторе VR/S100 может одновременно использоваться до двух микросхемных синтезаторов речи, что позволяет получать с одной платы два речевых выхода. Таким образом, здесь возможен одновременный вывод двух различных сообще-



решения фирмы «Спич технолоджи корп.».)

2. Линия напряжения V_{DD} обычно подсоединена к V_{CC} . Для дивительную перемычку следует разорвать. 3. Вместо P2 и C4 компонент.

ний, что очень удобно в таких приложениях, как многоканальные телефонные системы. Схемная печатная плата устройства VR/S100 (рис. 7.36) содержит встроенный микропроцессор и микросхему синтезатора речи. Система обеспечивает формирование сообщений длиной до 128 слов для двух одновременно работающих линий.

Словари для этого речевого генератора можно формировать на основе стандартного словарного состава или же генерировать из заказной речи, которую изготовитель кодирует за дополнительную плату.

Устройство Спич 1000, SP 1020 и Проуз 2000

Это еще одни говорящие изделия фирмы «Телесенсоры спич системс». Помимо синтезаторов речевого сигнала, который был описан ранее в данной главе, компания

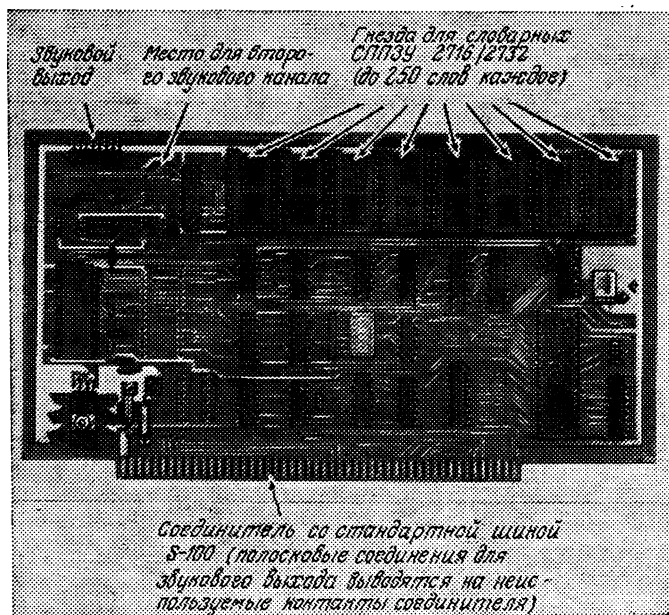


Рис. 7.36. Плата синтезатора речи VR/S 100 фирмы «Спич технологии корпорейшн». (С разрешения фирмы «Спич технологии корпорейшн».)

изготавливает речевые синтезаторы, генерирующие синтезированную речь по методу ЛПК. Первое из этих изделий, Спич 1000, представляет собой плату синтезатора речи большого размера, которая на рис. 7.15 показана на заднем плане. Хотя синтезатор Спич 1000— довольно дорогое периферийное устройство, он отличается высокой эффективностью.

Метод ЛПК-синтеза, используемый в устройстве Спич 1000, опирается на 12-полосный многозвенный фильтр, который позволяет построить цифровую модель голосового тракта человека. И поскольку схемная плата синтезатора включает микропроцессор 8085А, устройство обладает достаточным «интеллектом», чтобы реагировать на внешние команды и выполнять операции по осуществлению речевого выхода независимо от ведущего компьютера. Структурная схема платы синтеза-

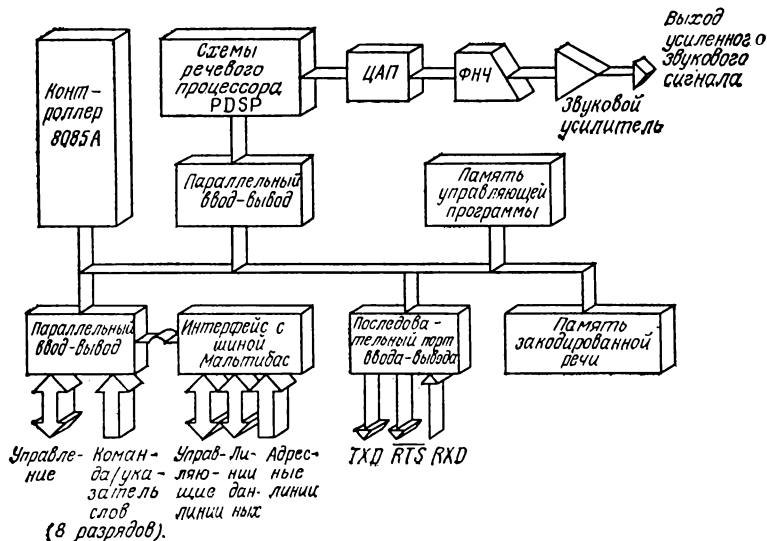


Рис. 7.37. Структурная схема устройства Спич 1000 фирмы «Телесенсорнн спич системс». (С разрешения фирмы «Телесенсорнн спич системс».)

тора Спич 1000 изображена на рис. 7.37. Анализ этой схемы показывает, что плата действительно представляет собой гибкий и многосторонний по своим возможностям речевой синтезатор, отличающийся высоким качеством воспроизведения. Он может синтезировать разные голоса: мужские и женские, говорящие как на одном, так и на многих языках.

Синтезатор Спич 1000 может работать в режиме управления запросами на прерывание или в режиме запрашивающих команд. Для получения 8-разрядной информации с ТТЛ-схем он подключается к интерфейсу

через шину Мальтибас или параллельный порт. Для вывода речи используется, кроме того, стандартный последовательный порт RS-232C. Поскольку емкость встроенной памяти составляет 458 кбит, синтезатор Спич 1000 способен говорить без повторов почти 200 с подряд. Словарь для этого синтезатора можно составить на основе стандартного словарного списка или путем заказного кодирования слов для специально формируемого словаря. Эту работу за соответствующую плату выполняет фирма «Телесенсори спич системс».

Одноплатный синтезатор речи выпускается в виде автономного устройства, включающегося по запросу. В этом варианте в его состав входят источник питания,

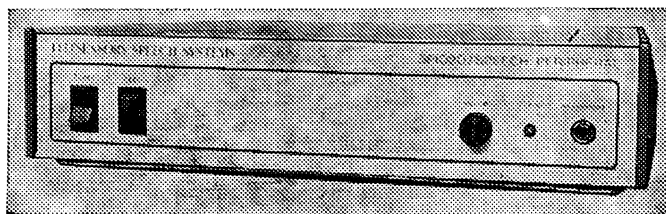


Рис. 7.38. ЛПК-синтезатор SP 1020 фирмы «Телесенсори спич системс», содержащий плату Спич 1000. (С разрешения фирмы «Телесенсори спич системс».)

кожух и интерфейс для связи с внешним электронным компьютером. В таком виде он известен под названием «периферийное говорящее устройство SP1020». Автономный ЛПК-синтезатор речи, изображенный на фотографии рис. 7.38, в принципе ничем не отличается от одноплатного синтезатора Спич 1000 — ему лишь приданы средства для самостоятельной работы. Благодаря своей гибкости эта система, снабженная последовательным интерфейсом RS-232C, безусловно, может использоваться очень широко. Чтобы познакомиться с качеством речи синтезатора Спич 1000, достаточно набрать демонстрационный телефонный номер фирмы «Телесенсори спич системс». Соединившись таким образом непосредственно с компьютерным интерфейсом, можно услышать не магнитную запись, а непосредственно «компьютерную» речь.

Помимо двух уже упомянутых изделий фирма «Теле-

сенсори спич системс» изготавливает также систему, преобразующую печатный текст в речь, — преобразователь «текст — речь» Проуз 2000. Этот синтезатор в одноплатной конфигурации изображен на рис. 7.39. В состав данного «разумного» говорящего модуля входит фирменная программа речевого синтеза по правилам, разработанным фирмой «Телесенсори спич системс», что дает пользователю возможность применять практически неограниченный словарь. Система работает по принципу преобразования текста, закодиро-

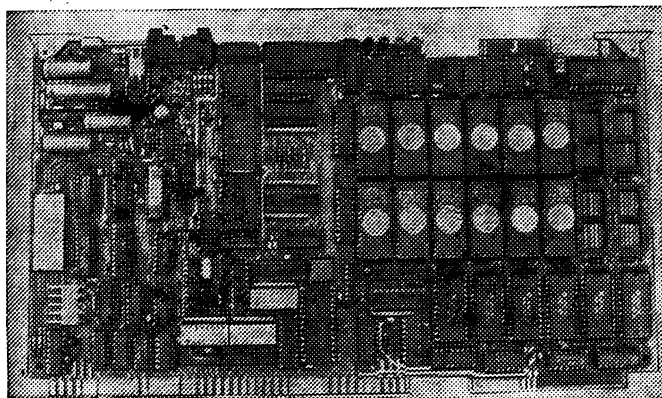


Рис. 7.39. Плата преобразователя «текст — речь» типа Проуз 2000 фирмы «Телесенсори спич системс». (С разрешения фирмы «Телесенсори спич системс».)

ванного ASCII-кодом, в аллофоны и обеспечивает речь, отличающуюся высокой степенью разборчивости. Кроме того, из текста автоматически выбираются интонационные характеристики, свойственные речи человека, которые затем вводятся в выходной речевой сигнал; это позволяет получить речь, имеющую естественное звучание.

Устройство Проуз 2000 продается в одноплатном варианте, как и устройство Спич 1000, но может выпускаться в кожухе, в котором помимо этой платы помещен источник питания и установлены необходимые соединительные разъемы. Сопряжение с ведущим компьютером может выполняться через последовательный

порт RS-232, поскольку в составе устройства имеется микропроцессор 8086. Скорость передачи данных, необходимая для обеспечения речевого выхода, составляет всего 100 бит/с. С качеством речевого выхода синтезатора Проуз 2000 при прямом преобразовании печатного текста в речь можно также ознакомиться, набрав демонстрационный телефонный номер фирмы.

Устройство Спик энд спелл фирмы «Тексас инструментс»

Возможно, покажется странным, что это устройство характеризуется как периферийное устройство речевого синтеза; однако здесь нет ошибки. Несколько компаний уже изготовили интерфейсы, обеспечивающие возможность сопряжения устройства Спик энд спелл с портом персонального компьютера. Поскольку выпущенное фирмой «Тексас инструментс» обучающее устройство было первым коммерческим изделием, в котором использовался ЛПК-метод генерации речи, оно быстро нашло применение и в качестве речевого синтезатора.

Устройство Спик энд спелл включает микросхему для речевого синтеза TMS5100 и заказной вариант 4-разрядного микропроцессора TMS1000. Так как фирма «Тексас инструментс», проявив предусмотрительность, снабдила схемную плату синтезатора Спик энд спелл портом для расширения словаря, в дальнейшем были изготовлены интерфейсы, обеспечивающие непосредственное соединение этого устройства с компьютерами нескольких типов. Этот порт изображен в правом верхнем углу структурной схемы устройства Спик энд спелл (рис. 7.40).

Один из периферийных интерфейсных преобразователей для устройства Спик энд спелл выпускает фирма «Перком дейта компани» (Гарленд, шт. Техас). Интерфейс спроектирован в расчете на сопряжение устройства Спик энд спелл через его расширительный порт с компьютером TRS-80, а также обеспечивает управление работой синтезатора от его встроенного словаря. Каждое слово словаря подвергается внешней выборке и произносится путем использования команд реек или роке. Хотя комбинации устройств Спик-2-Ми-2 фирмы «Перком» и Спик энд спелл фирмы «Тексас инструментс» представляет собой очень недорогой интерфейс для

синтеза речи по методу ЛПК, резидентный словарь синтезатора Спик энд спелл страдает отсутствием слов из повседневной практики. Некоторые из недостающих слов можно генерировать, введя с помощью интерфейса дополнительные модули, обеспечивающие расширение словаря.

Еще одним изготовителем интерфейсных модулей для устройства Спик энд спелл является фирма «Ист-коуст майкропродактс» (Одентон, шт. Мэриленд). В отличие от фирмы «Перком» эта компания не держит в секрете принцип действия своего устройства, давая пользователям возможность не только управлять выходом синтезатора Спик энд спелл, но и производить считывание его словаря в память. Интерфейс спроектирован в первую очередь для микропроцессоров 6502 и соединяется с ними через параллельный порт 6522.

Устройство TMS5200

Среди многочисленных устройств для генерации по методу ЛПК, выпускаемых фирмой «Тексас инструментс», микросхема речевого процессора типа TMS5200, вероятно, наиболее интересна. Она является центральной частью ЛПК-синтезатора, выпускаемого фирмой «Стрит электроник корпорейшн», и изготавливается фирмой «Тексас инструментс» в составе пробных наборов типа TMSK101 и TMSK102. Фирма «Тексас инструментс» держит в секрете технологию своих говорящих устройств, и поэтому об их конструктивных особенностях мало что известно.

Портативная система для анализа и синтеза (PASS)

Одной из последних разработок фирмы «Тексас инструментс» является портативный генератор-преобразователь «речь — ЛПК-код». Относительно этой системы имеются некоторые сведения, что позволяет нам привести ее описание. Это устройство, называемое портативной системой для анализа и синтеза (PASS), представляет собой прототип конструкции генератора-преобразователя «речь — ЛПК-код», работающего в реальном времени. До недавних пор формирование ЛПК-параметров речи осуществлялось только с помощью

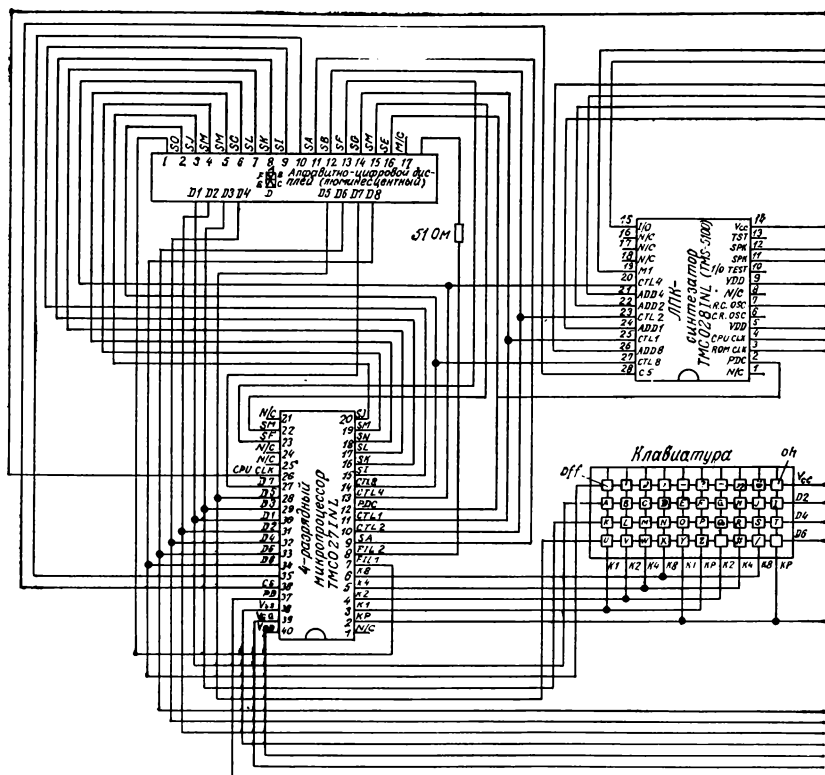


Рис. 7.40. Схема устройства Спик энд спелл фирмы «Тексас переменного тока отсутствует.)

больших универсальных компьютеров. Новая портативная система PASS (рис. 7.41), по-видимому, будет весьма дорогой. Устройства такого типа предполагается распределить по различным региональным центрам по всей стране с целью применения в качестве средств кодирования заказных словарей ЛПК-речи. Создание этой портативной высокоскоростной системы имеет огромное значение для всей области синтеза речи и генерации кодов по ЛПК-методу.

При работе с системой PASS говорящий просто на-

странение, популярность ЛПК-синтезаторов речи, безусловно, возрастет. Описанное устройство — первая доступная портативная система, предназначенная для ЛПК-кодирования речи непосредственно в реальном времени.

На этом мы завершаем наш обзор выпускаемых ныне говорящих устройств. Хотя, возможно, уже поя-

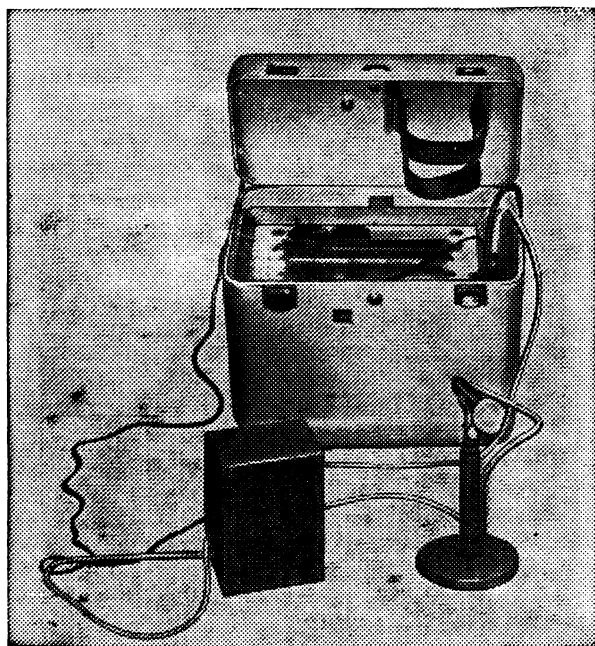


Рис. 7.41. Система ЛПК-кодирования PASS фирмы «Тексас инструментс».

вились и другие изделия подобного типа, отсутствие их в обзоре — не наша вина. Область речевого синтеза развивается столь быстро, что даже если бы все издание такой книги, как наша, было завершено в один день, вряд ли мы могли бы дать исчерпывающий список всех изделий. Тем не менее основные устройства, существующие сейчас, нами были отмечены.

Глава 8

ПРИМЕНЕНИЯ КОМПЬЮТЕРНОЙ РЕЧИ

Последняя тема нашей книги о синтезе речи — его практические приложения. Мы постараемся показать, как можно использовать синтезаторы. В данной главе мы рассмотрим несколько возможных применений говорящих систем в домах, а затем отважимся высказать ряд предложений относительно новых достаточно нетривиальных приложений говорящих систем. Мы надеемся, что некоторые высказанные здесь идеи помогут заинтересованному читателю добиться немалых результатов при практическом использовании говорящих устройств и что эта работа доставит ему удовольствие.

Данная глава разделена на три основные части. В первой из них мы обсудим простейшую речевую систему, которая состоит из персонального компьютера и говорящего периферийного устройства. Во второй части рассмотрим использование персонального компьютера, оснащенного говорящим периферийным устройством, в сочетании с различными периферийными устройствами, встречающимися в быту. И наконец, заключительная часть главы посвящена возможным коммерческим применениям систем синтеза речи. Некоторые из таких приложений уже реализованы на практике, внедрение других — не за горами, в настоящее время идет их лабораторная «доводка». Независимо от того, какое вы найдете применение для речевого синтезатора, работа с ним, безусловно, доставит вам много приятных часов. Осуществив какую-то одну из своих задумок, вы, возможно, захотите перейти к более сложным экспериментам. Ведь эта область электроники еще очень молода и постоянно развивается! И ваша работа здесь может принести вам в будущем известность.

Теперь поговорим о возможных применениях систем синтезированной речи, состоящих из базового ком-

пьютера и периферийного устройства для синтеза речи. При всей простоте такой системы возможности ее использования достаточно широки.

Простейшие системы синтеза речи

После того как ваша компьютерная система синтеза речи произнесла несколько слов, вы, несомненно, задумаетесь о возможностях разумного использования ее способностей. В настоящем разделе мы будем говорить о приложениях простейшей системы, включающей компьютер, клавишный пульт и дисплей, а также синтезатор речи. Ее блок-схема показана на рис. 8.1. Хотя

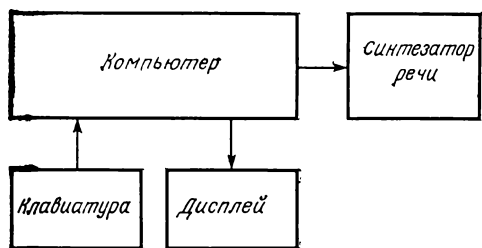


Рис. 8.1. Простейшая система синтеза речи.

на первый взгляд такую систему не назовешь мощной, существует множество проектов ее практического применения, к реализации которых вы можете приступить. И открывающиеся перед вами перспективы ограничены лишь вашим воображением. Рассмотрим некоторые возможные применения таких систем.

Говорящие часы

Это одна из простейших говорящих систем, которые вы можете сконструировать, и для практического результата вам потребуется лишь программа. Предполагается, что ваш компьютер может вести счет времени, располагая для этого электронными часами, работающими в реальном времени. Если такие часы в нем отсутствуют, то можно попытаться написать программу с временными циклами, которая введет необходимые обратные связи для счета секунд, минут и часов. В результате вы получите систему, состоящую из компьютера и синтезатора речи, которая будет говорить вер-

ное время по команде запроса или автоматически, скажем, с часовым интервалом. Для осуществления этого проекта нужна специальная программа, обеспечивающая правильный счет времени суток и запуск синтезатора речи, который по требованию объявляет время.

Программа, реализующая говорящие часы реального времени, может быть написана на ассемблере, либо на Бейсике, либо на любом другом языке программирования высокого уровня, который в конкретном случае

определяется особенностями компьютера. Естественно, одной из важнейших предпосылок для реализации данного проекта должно быть наличие периферийного синтезатора словаря, необходимого для выполнения функций говорящих часов; минимальный состав слов такого словаря приведен в табл. 8.1. Здесь перечислено в общей сложности 29 слов — их достаточно для полной характеристики времени суток при любых значениях часов и минут. Поскольку желательно иметь воз-

Таблица 8.1. Словарь говорящих часов

the	thirteen
time	fourteen
is	fifteen
(pause)	sixteen
one	seventeen
two	eighteen
three	nineteen
four	twenty
five	thirty
six	forty
seven	fifty
eight	oh
nine	clock
ten	
eleven	
twelve	

можность не только объявлять часы, но также выдавать по команде верное время в любой момент, необходимо располагать речевыми средствами для счета и произнесения всей последовательности чисел от 0 до 59. При использовании синтезатора с фонемным возбуждением эти слова должны быть заранее закодированы через фонемы. Если синтезатор имеет готовый словарь, то следует найти коды команд, которые будут обеспечивать доступ к выражениям, указывающим нужное время суток. После того как в системе создан исходный словарь, можно приступить к генерации программы, которая позволит построить систему типа говорящих часов.

Блок-схема возможной программы для говорящих часов приведена на рис. 8.2. Хотя эта программа составлена в общей форме, она тем не менее достаточно специфична, чтобы ее блоки принятия решений можно было

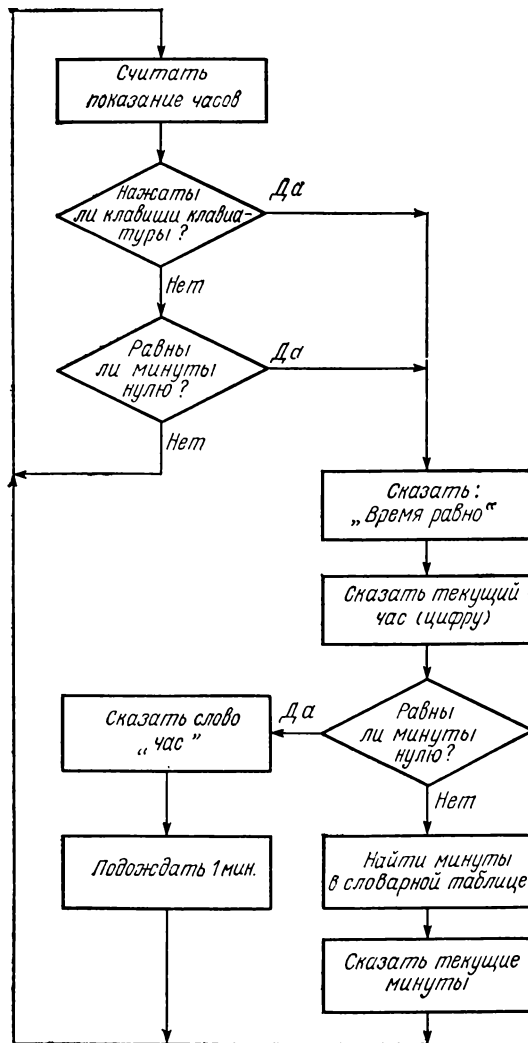


Рис. 8.2. Блок-схема программы для говорящих часов.

без особого труда превратить в программные блоки на языке программирования конкретного компьютера. Приведенная блок-схема программы сравнительно проста, однако некоторая сложность заключается в том, что время, выражаемое в часах и минутах, характеризуется в иной форме, чем время, выражаемое только в часах. Так, если часы просто объявляются автоматически, например «5 o'clock» (5 часов), то произносимое численное значение часов никогда не превышает 12. Если же с клавишного пульта вводится команда запроса времени в произвольные моменты, то речевой вывод времени между 5 и 6 часами может изменяться в пределах от 5:01 до 5:59. Простейшим способом словарного преобразования чисел в слова может быть построение таблицы с 60 числовыми входами для запроса минут в пределах от 00=o'clock (час) до 59=fifty-nine. Таким путем доступ к значениям минут может производиться непосредственно из компьютерной памяти, что позволяет получать соответствующие табличные слова для возбуждения синтезатора. Если в синтезаторе имеется программа преобразования типа «текст — речь», то о таком преобразовании чисел, вероятно, не придется заботиться. Большинство программ преобразования текста в речь обладает способностью определять, как произносятся числа от 1 до 100, исходя исключительно из числового входа. В таком случае все, что необходимо задать синтезатору речи в качестве команды для произнесения времени,— это числовое значение текущего времени дня. Алгоритм «текст — речь» самостоятельно справится с выдачей минут в форме речи. Некоторые алгоритмы преобразования «текст — речь» обладают таким «интеллектом», который позволяет установить, что полученное ими от компьютера число указывает время, поскольку первые цифры в нем отделены от двух последних двоеточием. При поступлении на вход синтезатора цепочки текста «5:22» он произнесет время как five ... twenty two (пять ... двадцать два).

Этот довольно простой эксперимент можно легко провести с любым компьютером, обладающим способностью читать показания внутренних машинных часов, которые отсчитывают время суток. В некоторых компьютерах это осуществляется посредством функции time S. В ряде других машин, оснащенных аппаратными часами, операция считывания времени сводится к чте-

нию показания микросхемы часов, которое снимается с портов этого периферийного устройства. Если компьютер не имеет ни встроенных электронных часов, ни внешних аппаратных, то создание говорящих часов можно начать с построения очень точного хронизирующего контура с длительностью цикла 1 с. Этот контур необходимо связать со схемой пересчета на 60, которая будет выдавать минутные отметки. Сигналы счета минут должны подаваться на другую схему пересчета на 60, которая отслеживает время в минутах между отсчетами часов. И наконец, следует ввести еще схему пересчета на 12, которая изменяет свое состояние каждый полный цикл 60-минутного счетчика и, таким образом, показывает часы. Числовой выход всех этих контуров говорящая система может использовать для сообщения по запросу точного времени суток. Чтобы лучше разобраться в работе программных часов, посмотрим на рис. 8.3. Приведенная здесь блок-схема программы показывает довольно простой путь построения программных часов, которые служат для запуска говорящей программы. Реализация такой программной системы в говорящем компьютере позволит весьма наглядно продемонстрировать возможности синтезированной речи. Говорящие часы будут продолжать говорить время, пока действует их управляющая программа, давая на протяжении всего дня показания текущего времени.

Если, несколько расширив пример с говорящими часами, ввести в систему с общим контуром отсчета времени некоторую программу распознавания заданного момента времени, то в таком случае мы получим не просто говорящие часы, а часы, которые можно запрограммировать таким образом, что в заданные моменты — или, возможно, даже в заданные моменты конкретных дней месяца — они будут произносить определенные фразы. Проявив еще немного творческой смекалки, можно построить говорящий будильник, который будет запрограммирован так, чтобы он будил вас утром фразой: «Вставай» или «Вставай, соня, уже поздно». Если у вас назначено свидание или вам следует о чем-то вспомнить в определенное время, то систему можно запрограммировать так, чтобы она в нужный момент выдала нужное сообщение. Например, вам захочется, чтобы в шесть часов вечера компьютер сказал: «Включи телевизор — время передачи новостей». Если у вас

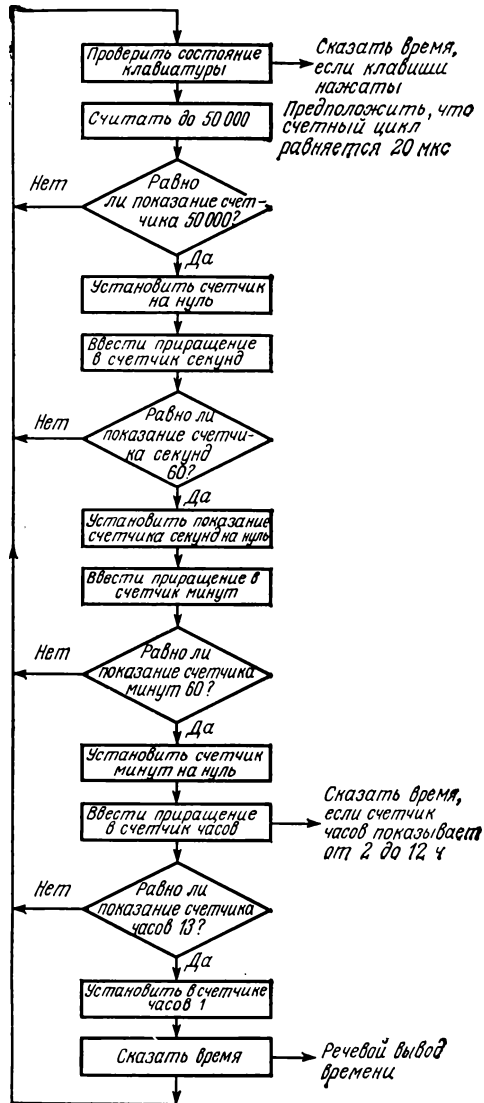


Рис. 8.3. Возможная структура программы для часов, показывающих время.

довольно напряженный распорядок дня, то вы можете составить календарь своих деловых и прочих встреч, где будет перечислено все, что необходимо сделать в течение каждого дня, и компьютер по утрам будет сообщать вам о намеченных на текущий день делах. Достоинство этой системы состоит в том, что она исключительно надежна и никогда не забудет напомнить вам о неотложных делах. Полезность создания подобной системы для занятого человека не вызывает сомнений. Утреннее словесное напоминание о намеченном распорядке дня, безусловно, поможет вам лучше организовать свою деловую жизнь.

Развлечения и игры

Мы рассмотрели лишь один пример использования простейшей системы синтеза речи. По мере того как вы будете открывать все новые возможности конкретного периферийного устройства с синтезатором речи и накапливать опыт работы с ним, у вас может зародиться мысль объединить речевое устройство с игровыми программами. Естественно, что это требует глубокого знания программ, которые вы хотите модифицировать. Так, если они написаны на Бейсике, то вы сможете легко выбрать операторы «печатать» (которые, как правило, используются для взаимодействия с игрой) и либо заменить, либо дополнить эти операторы словесными, или наговариваемыми, сообщениями. Одной из самых распространенных игр, в которой введение словесных сообщений дает значительный эффект, является игра «Звездная дорога»: ведь компьютер космического корабля начинает выглядеть куда реалистичнее, если он разговаривает.

Одно из ограничений, которое следует учитывать при введении в игры синтезированной речи, — это словарь, доступный данному периферийному устройству со средствами синтеза речи. Если синтезатор обладает способностью генерировать неограниченный словарь, то его возможности взаимодействия с компьютерной игрой будут поистине беспредельны. Если же, напротив, речевая система имеет весьма ограниченный словарь, то вам придется много потрудиться. Нередко вы сумеете найти слова, которые с успехом заменят недостающие словарю. Подумайте также об использовании омо-

нимов, которые отличаются написанием от нужных вам слов, но звучат очень сходно. Другой, довольно искусственный способ, который можно применить для создания необходимых, но отсутствующих в словаре синтезатора слов, представляет собой конкатенирование, т. е. слияние двух имеющихся в словаре слов для генерации третьего. Например, использование слова *be* и числа 4 (*four*) позволят конкатенировать слово *before*. Таким искусственным сцеплением можно образовать еще много слов, так что, прежде чем опустить руки от тщетности поисков замены недостающему слову, попытайтесь скомпоновать это слово из тех, что имеются в словаре синтезатора.

Если вы пользуетесь синтезатором с фонемным возбуждением, то можете найти удачное применение роботизированному голосу этой системы. В прошлом некоторые киностудии и студии, занимавшиеся звуковыми эффектами, не жалели средств и сил, чтобы получить речь, имеющую «нечеловеческое звучание». Располагая же системой с фонемным возбуждением, можно по желанию изменять звуки речи и фонемы. Если соответствующим образом подобрать высоту тона и звучание фонем, то фонемный синтезатор позволит создать весьма интересные звуковые эффекты. А присоединив к синтезатору громкоговоритель, вы сможете произвести ошеломляющее впечатление на слушателей. Но все это не более чем только еще одна возможность использования синтезаторной системы.

Говорящие устройства для людей с дефектами зрения и слуха

Если перейти к более серьезным задачам, то и здесь фонемному синтезатору можно найти дело по плечу. Так, генерируемые им чистые фонемные звуки могут оказать огромную помощь в обучении правильной речи людей с расстройствами речи. Располагая простейшими средствами программирования фонем и устройством, способным произносить их непрерывно, можно создать приспособление, о котором мечтают многие годы специалисты по исправлению речи. Это средство обучения, несомненно, окажет огромную помощь в исправлении дефектов речи у детей. Можно пойти и дальше. Используя фонемный синтезатор в сочетании со спектрографом

для генерации видимых изображений речевых сигналов, мы получим устройство, которое можно применить для обучения речи глухих. Произнося звуки, они должны добиваться их визуального (по спектрограмме) совпадения со звуками синтезатора. Занимаясь подобными упражнениями, люди, которые никогда не слышали живой речи, могут самостоятельно научиться говорить.

Одно из наиболее важных применений простейшей системы синтеза речи — помощь слепым и людям с большой потерей зрения. Проблема ввода информации в компьютер для слепых в основном решена посредством клавиатуры со шрифтом Брайля. Однако вывод данных из компьютера оказывается в этом случае более сложной задачей, и до недавних пор наиболее эффективным решением ее считалось применение принтеров со шрифтом Брайля. Далее, если правильно подобрать синтезатор речи, который осуществлял бы вывод всей информации компьютера, в том числе и символов, соответствующих нажатию каждой клавиши, то слепые смогут эффективно использовать все широкие возможности компьютера. Для этого, конечно, необходима обширная компьютерная программа с алгоритмом преобразования «текст — речь», которая обеспечит неограниченный по составу словарь.

Большинство имеющихся в настоящее время в продаже синтезаторов с фонемным возбуждением паделены доступным для пользователя довольно простым алгоритмом преобразования «текст — речь». При отсутствии нужной вам преобразовательной программы вы можете поставить перед собой смелую задачу создания ее собственными силами. В данной области было выполнено немало работ, и результаты многих из них опубликованы ¹.

¹ См., например, следующие работы: McIlroy M. D. Synthetic English Speech by Rule, Computer Science Technical Report 14, Bell Telephone.— N. Y.: Murray Hill, March 1974; Elowitz H. et al. Automatic Translation of English Text to Phonetics by means of Letter-to-Sound Rules, Naval Research Laboratory Report 7948, January 1976; Allen G. P. Synthesis of Speech from Unrestricted Text, Proceedings of the IEEE, April 1976, 64 (4), pp. 433—42. [Есть русский перевод: Аллен. Синтез речи из произвольного текста, ТИИЭР, 64, № 4, с. 161, 1976.]

Эти публикации помогут вам глубже понять механизм действия алгоритмов синтеза речи по правилам, используемым в фонемных синтезаторах.

Если вы интересуетесь алгоритмами преобразования текста в речь, которые используются в большинстве фонемных синтезаторов, то данные публикации познакомят вас с «секретами» механизмов их работы. Указанные программы отличаются исключительной сложностью, поскольку в правилах произношения письменного английского текста имеются тысячи исключений. Ни одну из существующих программ построения синтезированной речи по правилам нельзя считать идеальной. Поэтому любое усовершенствование существующего алгоритма подобных программ, безусловно, найдет отклик и получит признание.

Еще одно применение

Мне хотелось бы здесь предложить еще одно применение простейшей системы речевого синтеза. Это — использование говорящего компьютера для обучения речи птиц. Компьютер способен многократно и монотонно повторять одно и то же слово, а это идеальный метод для обучения разговору попугаев и других говорящих пернатых. Однако не удивляйтесь, если ваша птица будет безошибочно копировать все речевые изъяны синтезатора. Ведь она воспринимает многократное повторение не только слов, но и всех речевых дефектов устройства. Можно даже обучить птицу говорить голосом робота. Представляете себе эту картину!

Этим, конечно, не ограничиваются возможные приложения простейшей системы синтеза речи. Как я уже говорил, сфера ее применений зависит исключительно от вашего воображения. «Интеллект» такого синтезатора можно значительно развить, если управлять генерируемой им речью не с клавишного пульта, а программным путем, так что займитесь составлением программ, способных говорить. Работая над такими программами, вы лучше узнаете вашу систему и одновременно получите большое удовольствие.

Говорящий персональный компьютер с периферийными устройствами

Этот небольшой раздел адресуется тем, кто любит заниматься усовершенствованием различных устройств. Детально освоив простейшую систему синтеза речи, вы

сможете существенно расширить возможности говорящего компьютера, дополнив его несколькими периферийными устройствами. Большинство из них может быть оснащено несколькими параллельными или последовательными портами ввода-вывода. Хотя описываемые в этом разделе применения имеют преимущественно «домашний уклон», ничто не мешает вам расширить сферу их действия. По мере все более тесной интегра-

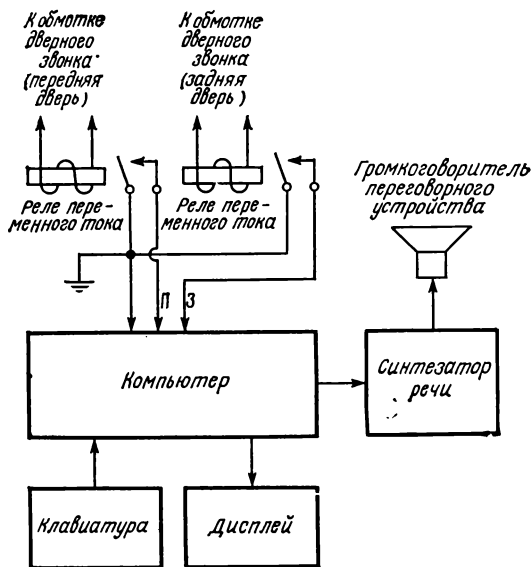


Рис. 8.4. Схема говорящего дверного звонка.

ции говорящего компьютера с системами сугубо практического назначения вы осознаете его впечатляющую эффективность. Подумайте, например, о соединении говорящего компьютера с дверным звонком.

Говорящий дверной звонок

Схему, приведенную на рис. 8.4, можно использовать для создания довольно необычного говорящего дверного звонка. Единственное, что необходимо для этого, — два небольших реле, которые следует подсоединить параллельно обмотке катушки дверного звонка. Реле, естественно, должны быть рассчитаны на то же номи-

нальное напряжение, что и звонок, так что будьте внимательны при выборе этих компонентов. Контакты реле нужно подсоединить к двоичным входам параллельного порта ввода компьютера. Если кнопки управления звонком установлены на передней и на задней входных дверях, то компьютер нужно подключить к обеим кнопкам, как показано на рисунке. Если дверная кнопка одна, то тогда, очевидно, достаточно одного реле и одного двоичного входа компьютера.

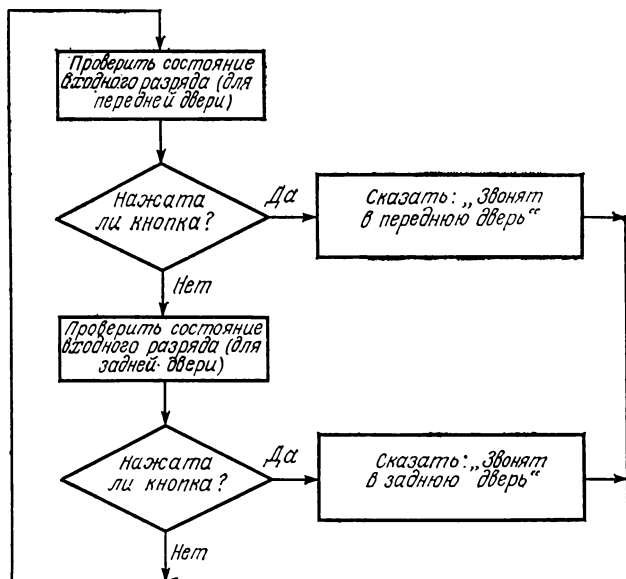


Рис. 8.5. Блок-схема программы для говорящего дверного звонка.

Выбор речевого выхода, посредством которого говорящий компьютер будет отвечать на звонок, полностью на вашем усмотрении. Можно, например, запрограммировать его говорить фразу: «Кто-то звонит в дверь». Если в вашем доме есть и передняя, и задняя двери, то вы можете запрограммировать компьютер говорить в зависимости от нажатия той или другой кнопки следующие фразы: «Кто-то звонит в переднюю дверь» или «Кто-то звонит в заднюю дверь». Вероятно, вы захотите ввести в речевую систему сигнал предупреждения (см. гл. 4), который будет предшествовать

сообщению компьютера. Возможный вариант программы реализации этого приложения показан на рис. 8.5 в виде блок-схемы. Хотя структура программы исключительно проста, два ее блока, обозначенные «Сказать (фразу)», следует программировать во многом аналогично оператору «Печатать», что позволит компьютеру надлежащим образом реагировать на команду «Говори». Чтобы упростить программирование говорящего звонка, хранимый в памяти системы словарь должен содержать обе используемые фразы.

Говорящая система охраны жилища

Еще одно достаточно простое применение говорящего компьютера в доме — его использование в качестве системы охранной сигнализации. Вообразите потрясение грабителя, когда он услышит: «Взлом окна спальни — вызывайте полицию!». Естественно, не меньший эффект даст использование и других фраз того же содержания. Аппаратная реализация этой системы показана на рис. 8.6. В качестве контрольных контактов для окон и дверей можно использовать просто магнитные переключатели. Каждый переключатель присоединяется к двоичному входу компьютерного порта ввода и остается в замкнутом состоянии до тех пор, пока провод не окажется перерезанным или не будет открыто соответствующее окно (или дверь). Если подсоединить к другому входному компьютерному порту простейший тепловой датчик, то систему можно одновременно использовать для предупреждения о пожаре или просто чрезмерном повышении температуры поблизости. То обстоятельство, что компьютер может тестировать ряды входного порта, будучи предварительно запрограммированным на то, что сигнал на одном входе формируется контактами защиты от взлома, а на другом — тепловыми контактами, позволит ему определить, какое сообщение следует выдавать в том или другом случае. Блок-схема простейшей программы для говорящей системы охраны жилища показана на рис. 8.7. По своим функциям эта программа во многом сходна с программой говорящего дверного звонка и отличается от последней лишь характером сообщений, используемых в качестве речевого вывода.

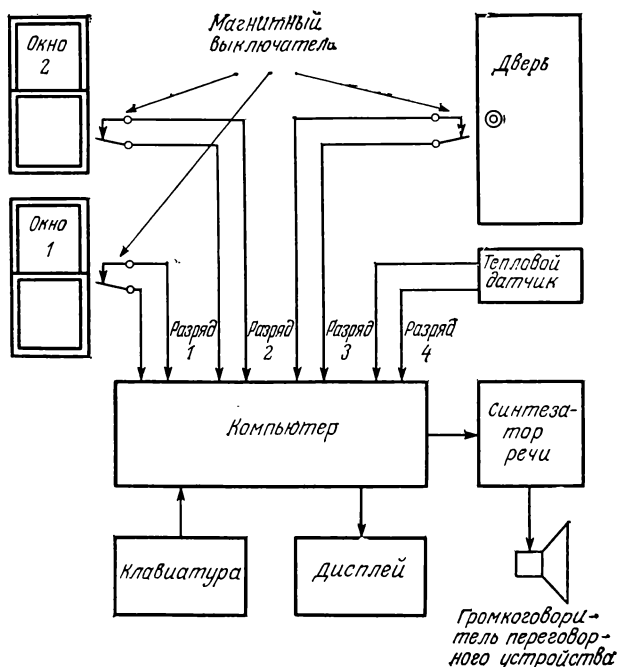


Рис. 8.6. Говорящая система охранной сигнализации для дома.

Существенное различие между данной и предыдущей программами заключается также в следующем: чтобы сильнее обескуражить грабителя, следует ввести временную задержку между поступлением сигнала о взломе и произнесением компьютером сообщения. Это, безусловно, произведет больший эффект на грабителя, если он, уже проникнув в дом, вдруг услышит предупреждение о взломе. Однако вносить входную задержку в сигнал от теплового датчика, разумеется, не следует, поскольку предупреждение о возможном пожаре должно быть дано без промедления. Говорящая система обеспечения охраны жилища может быть расширена настолько, насколько вы пожелаете. Как известно, основные трудности при создании эффективной системы безопасности связаны с подведением проводов к окнам и дверям. Точно так же система сигнализации может быть подсоединена к говорящему компьютеру. Однако в большинстве существующих стандартных

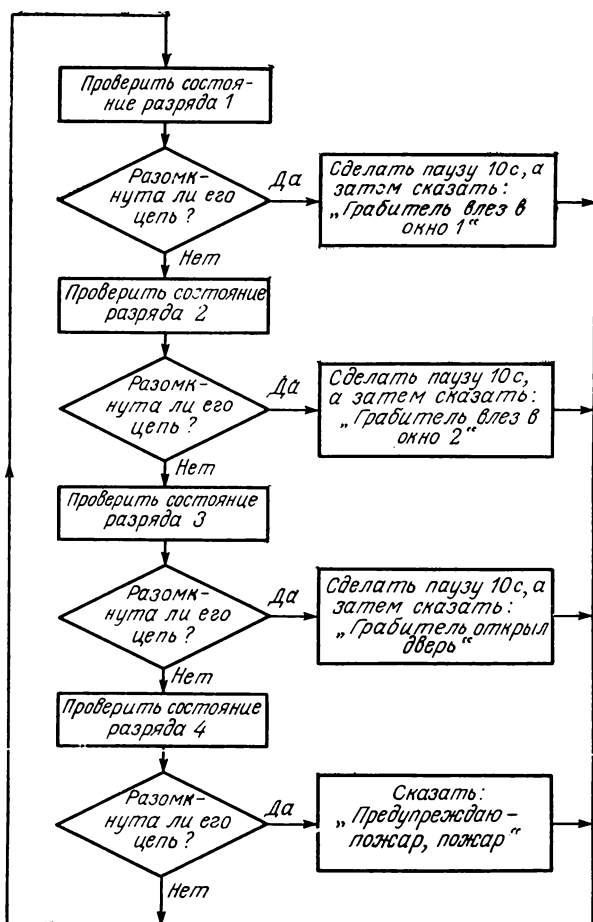


Рис. 8.7. Пример построения говорящей системы охранной сигнализации для дома.

систем охранной сигнализации все переключатели соединены последовательно от окна к окну и (или) к двери и т. д., что исключает возможность идентификации места взлома. Следовательно, при подключении к такой стандартной системе компьютер можно запрограммировать только на фразу: «В доме грабитель». Но это произведет на грабителя, пожалуй, не меньшее впечатление, чем фраза компьютера с точным указанием места взлома.

Замок, отпирающийся голосом

Говорящей компьютерной системе можно найти эффективное применение, если соединить ее с периферийным устройством, наделенным простейшими средствами распознавания голоса. Если вы дополните говорящий компьютер устройством распознавания голоса, то обнаружите, что число его возможных приложений станет практически неограниченным. Одно из таких приложений, правда, несколько искусственное (но зато лег-

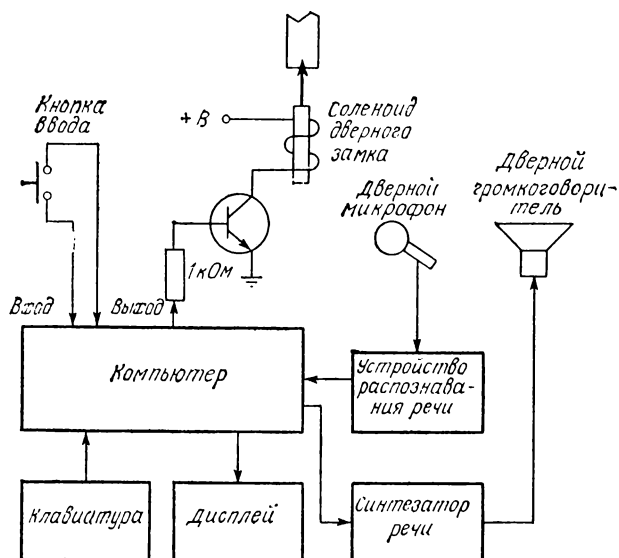


Рис. 8.8. Возможная система электронного замка, отпирающегося голосом.

ко реализуемое), — это замок, который будет открываться в ответ только на заранее запрограммированные голоса.

Поскольку большинство систем распознавания голоса обладают способностью распознавать по меньшей мере 16 различных фраз, в сочетании с синтезатором голоса их можно использовать для создания очень надежного замка с голосовым управлением. Принципиальная схема реализации такой системы изображена на рис. 8.8. Помимо периферийного устройства со средствами синтеза речи здесь понадобится еще система

распознавания речи, соединенная с компьютером, а также некая модификация дверного замка, причем замок должен быть электронным. Двойчный выход компьютера используется для управления соленоидом дверного запора через мощный транзистор, напряжение которого согласовано с номинальным напряжением соленоида. Кроме того, на входной двери устанавливается кнопочный выключатель, нажатие на который уведомляет компьютер о желании человека войти, а также микрофон и громкоговоритель, соединенные с устройством распознавания голоса и речевым синтезатором соответственно.

Работа этой системы полностью определяется программными средствами; поэтому она обладает исключительной гибкостью, что позволяет легко приспособить ее к конкретным нуждам. Рассмотрим оперативную программу — весьма фундаментальную по своей структуре, — которую можно использовать для управления замком, отпирающимся голосом. Структура этой программы показана на рис. 8.9.

Специфика использования речевого синтеза в данной системе состоит в том, что компьютер может попросить человека, желающего войти, повторить наугад любое из четырех предварительно записанных в его память слов. Устройство распознавания голоса должно пройти дополнительную тренировку в распознавании тех же четырех слов, произносимых каждым человеком, которому разрешен вход в помещение с дверью, отпираемой голосом. По команде компьютера человек, желающий войти, должен повторить слово, запрошенное компьютером.

Преимущество диалоговой системы такого рода заключается в том, что если была произведена магнитная запись вашего голоса в попытке «обмануть» замок, то произвольный выбор паролей, безусловно, отобьет у «электронного грабителя» охоту проникнуть в помещение с помощью магнитной записи нужного голоса. Хотя замок, открывающийся голосом, довольно сложен для практической реализации, мы упомянули о нем, чтобы проиллюстрировать все разнообразие возможных приложений компьютерного синтеза речи. Если вы хотите двигаться в этой области дальше и обладаете достаточными техническими способностями, то можно подумать об установке речевого синтезатора в кухне: во

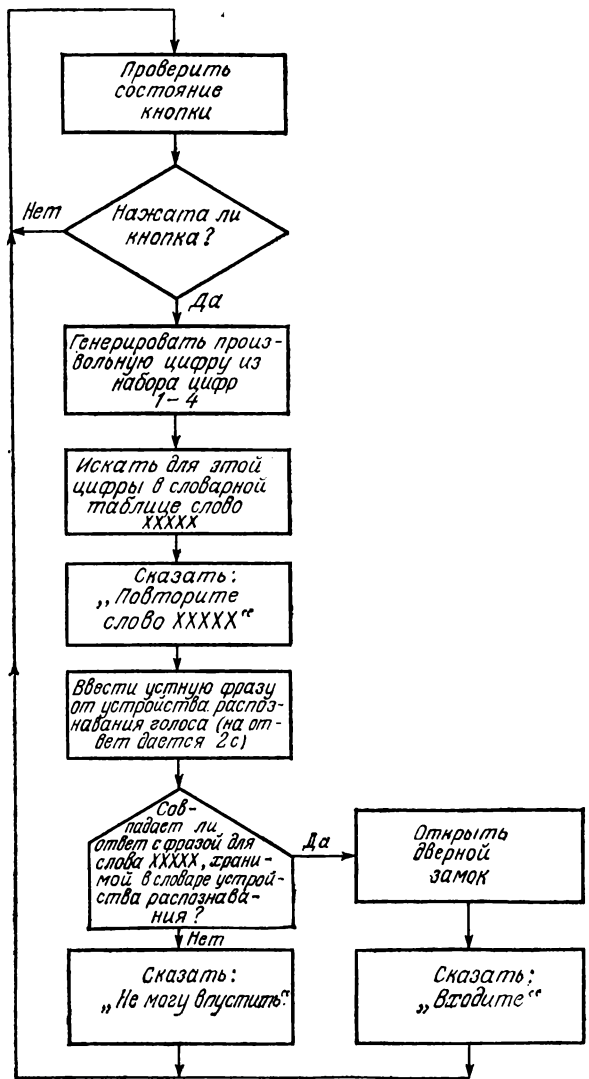


Рис. 8.9. Блок-схема программы для электронного замка, открываемого голосом.

время готовки он будет читать вам поваренные рецепты или сообщать о текущем состоянии обогревательной и охлаждающей систем в квартире с указанием общего расхода энергии.

Число подобных приложений бесконечно. Если вы не поленились подсоединить к компьютеру все ваши домашние системы, то у вас может возникнуть желание поделить компьютер способностью отвечать на телефонные звонки и говорить. В зависимости от выбранной конструкции подобной системы можно добиться самых разнообразных результатов. Так, можно построить систему, с которой вы сможете разговаривать по телефону, находясь вне дома. Введя (после соединения) несколько числовых кодов с кнопочного номеронабирателя, в ответ вы получите от компьютера полный отчет о состоянии вашего дома по таким показателям, как температура, безопасность, число бывших в ваше отсутствие телефонных вызовов и т. д.; все эти данные компьютер будет считывать с соответствующих индикаторов. Такая система может избавить вас от больших неприятностей и значительных затрат, если сообщит вам, к примеру, что за ваше двухнедельное отсутствие температура в холодильнике поднялась до 27°C. Придется срочно звонить приятелю и просить, чтобы он позаботился о сохранности ваших припасов.

Если говорящий компьютер подключить к телефону, то его возможные применения сразу многократно возрастут. Так, используемый в системе охранной сигнализации говорящий компьютер может быть запрограммирован на получение доступа к телефону непосредственно при взломе. В таком случае он может вызвать по телефону вашего друга (или полицию), описав при этом ситуацию и одновременно указав адрес. Такие системы уже имеются, однако их говорящая схема, как правило, механически воспроизводит соответствующую запись и, таким образом, зависит от надежности механического устройства. У компьютера, оснащенного периферийным устройством со средствами синтеза речи, этой проблемы нет, но для него, конечно, требуется резервный источник питания, который действует в случае неисправности сети или ее намеренном отключении.

Итак, мы перечислили ряд возможных применений системы синтеза речи. Хотя мы коснулись лишь того,

что буквально «лежит на поверхности», надеемся, что эти примеры разожгут ваше воображение и оно подтолкнет вас на дальнейшие поиски. Как только в вашем распоряжении окажется компьютерная система, способная устно взаимодействовать с внешними абонентами, чтобы реализовать очередную идею, вам достаточно написать соответствующую программу. И если система сможет говорить с вами, опираясь на вводимую в нее информацию о происходящих вокруг нее событиях, то это, безусловно, можно считать вашим замечательным достижением.

И наконец, завершая обзор приложений синтезаторов речи, коснемся их коммерческого использования. Эта область еще таит в себе много неизведанного. Пока писалась эта книга, каждый день появлялись сообщения о все новых устройствах с синтезированной речью. И недалеко то время, когда, подойдя к автомату по продаже безалкогольных напитков, вы услышите синтезированный голос: «Не хотите ли опустить 40 центов?».

Коммерческие приложения синтезированной речи

Эта область применения синтезированной речи стремительно развивается. Изготовители торговых автоматов, автомобилей, детских игрушек состязаются между собой в стремлении выпустить изделия с встроенными в них синтезаторами речи. Пожалуй, со времени изобретения транзистора в промышленных лабораториях США не вспыхивало такого ажиотажа. Каждый старается наделить свои изделия способностью говорить. Когда эта книга печаталась, синтезаторы речи уже были успешно использованы в таких устройствах: стиральная машина; СВЧ-печь; торговый автомат по продаже шипучих напитков; видеоигровой автомат; лифт; автомобиль; электронный мультиметр; некоторые настольные электронные игры для детей; телефонный справочник; автоматический информатор, дающий справки об авиарейсах; часы; телевизоры; калькуляторы; электронные «переводчики»; проигрыватели; компьютеры. Все эти устройства стали говорящими.

Список можно продолжать бесконечно. Через пять — десять лет он скорее всего займет уже страниц пять по-

добной книги. Что я говорить, способность предоставить пользователю прямой речевой вывод кажется весьма привлекательной; однако, по мере того как в США все в большей степени внедряется двуязычное образование, возникает вопрос: какой же язык применять в говорящих системах? Письменный текст можно представлять на двух языках параллельно, обеспечивая равные возможности для восприятия любого из них. При устном общении многоязычные тексты одинакового содержания придется воспроизводить последовательно, один за другим. Одна из проблем, которую необходимо будет решить (или проигнорировать) при дальнейшем распространении говорящих систем, связана с тем, как лучше произнести одну и ту же фразу на разных языках. Быть может, наступит день, когда, войдя в магазин с намерением купить говорящую СВЧ-печь, мы услышим вопрос: «Какую печь вы желаете приобрести — говорящую на английском, испанском, итальянском и т. д.?» Хотя сегодня это выглядит курьезом, очень возможно, что окончательное решение этой проблемы сведется к установке на говорящих изделиях «языкового переключателя», повернув который вы сможете услышать язык по своему выбору.

Не исключено, что наступит день, когда мы, оглянувшись вокруг, воскликнем: «Долго ли все эти окружающие нас немые устройства будут оставаться такими, как они есть?» Быть может, рано или поздно все вокруг нас заговорит. Примет ли общество говорящие вещи или отвергнет их целиком и полностью, покажет время. Первые несколько говорящих предметов в нашем доме, безусловно, будут восприниматься как забавная новинка, однако каково придется нашим внукам, если им некуда будет укрыться от электронных голосов, заполнивших все жизненное пространство.

Я вовсе не хочу сказать, что оснащение любого коммерческого изделия средствами синтеза речи само по себе следует рассматривать как дурное начинание. Мысль моя сводится к тому, что изготовителям следует хорошо подумать, прежде чем вводить синтезированную речь в свои серийные системы. Если в нашем распоряжении останутся регуляторы громкости и выключатели для управления речевым выводом бытовых устройств и мы сможем продолжать пользоваться визуальными

дисплеями, то тогда при желании мы еще сможем наслаждаться тишиной. В противном случае нас не будет покидать ощущение, что мы постоянно находимся в толпе.

Заканчивая книгу о синтезированной речи, генерируемой компьютером, я хочу напомнить вам, что все, о чем здесь рассказано, — это лишь верхушка айсберга. Наука о генерации искусственной речи пока находится в младенческом возрасте. Ведь большинство людей еще продолжают считать, что говорящий компьютер относится к области научной фантастики. И я надеюсь, что эта книга немного приоткрыла перед вами «дверь в будущее». Все это уже не научная фантастика, а отчасти — сегодняшняя реальность. До вас донесся голос из завтрашнего дня.

СЛОВАРЬ ТЕРМИНОВ

(в употребляемых в книге значениях)

Автокорреляция. Способ обработки сигнала, при котором сигнал задерживается и затем задержанный сигнал умножается на первоначальный.

Адамово яблоко. Выпуклость на передней поверхности шеи (у мужчин), образованная щитовидным хрящом гортани.

АИКДМ. Адаптивная импульсно-кодовая дельта-модуляция.

Аксон. Отросток нервной клетки, проводящий нервные импульсы от тела нейрона к иннервированным органам или другим нейронам.

Акустический. Имеющий отношение к звуку или слуховому восприятию.

Аллофон. Произносимые вариации фонем, зависящие как от самих фонем, так и от их положения в словах.

Амплитуда. Величина (громкость) сигнала.

Аналоговый. Плавно (без заметных дискретных скачков) изменяющийся сигнал.

Аппаратные средства. Электронные компоненты вычислительных машин и периферийных устройств.

Артефакт. Процесс или образование, не свойственные изучаемому объекту в норме и возникающие обычно при его исследовании искусственным путем.

Артикуляция. Работа органов речи при произнесении того или иного звука; степень отчетливости произношения.

Афония. Отсутствие голоса (при сохранности шепотной речи), наблюдающееся при заболеваниях гортани и связанное с неспособностью голосовых связок вибрировать.

Аффриката. Согласный звук, состоящий из взрывного и фрикативного элементов.

Байт. Часть машинного слова, состоящая обычно из восьми битов.

Барабанная перепонка. Упругая тонкая пластинка из соединительной ткани между наружным и средним ухом. Передает звуковые колебания через слуховые косточки во внутреннее ухо.

Белый шум. Полностью случайные сигналы, содержащие все частоты; акустически звучат как шипение.

Бит. Разряд в двоичной системе.

Бронхи. Трубчатые воздухоносные ветви трахей; часть голосового тракта человека,

Взрывной звук. Согласный звук, характеризующийся длинной паузой, следующей за резким усилением амплитуды при произнесении согласного.

Вокодер. Любое устройство электронного кодирования речи.

Восстановленная речь. Речь, восстановленная электронными средствами из записанной в памяти компьютера человеческой речи.

Выпрямитель. Электронный компонент, пропускающий ток только в одном направлении; используется для преобразования переменного тока в постоянный.

Высота основного тона. Доминирующая частота, создаваемая источником звука.

Гармоники. Искажения чисто синусоидального сигнала, заключающиеся в появлении гармонических составляющих, кратных основной частоте.

Гласный звук. Звук речи, состоящий главным образом из голосового тона.

Глотка. Часть пищеварительного тракта от пищевода до ротовой полости.

Глухой звук. Звук, произносимый без колебаний голосовых связок.

Голосовые связки. Парные эластичные тяжи в боковых стенках гортани. При их колебаниях и сужении голосовой щели образуются звуки голоса.

Гомоморфный фильтр. Фильтр, пропускающий нужный сигнал и подавляющий все нежелательные компоненты.

Гортань. Часть голосового тракта, включающая голосовые связки.

Грамматический разбор. Разделение фразы или компьютерной программы на элементарные структурные части.

Громкость. Воспринимаемая характеристика звука, связанная с его силой и высотой тона.

Губной звук. Звук, произносимый губами.

Двоичная система. Система счисления по основанию 2; все числа в ней представлены в виде различных наборов нулей и единиц.

Дельта-модуляция. Тип цифрового кодирования аналогового сигнала, основанный на использовании изменений сигнала.

Диалект. Разновидность языка, употребляемая более или менее ограниченной группой людей, связанных социальной, территориальной и пр. общностью.

Дисфония. Расстройства голоса, вызываемые заболеваниями гортани.

Дифтонг. Сочетание двух гласных звуков.

Затухание. Уменьшение амплитуды или мощности сигнала.

Звонкий звук. Звук, для произнесения которого используются голосовые связки.

ЗУПВ. Запоминающее устройство с произвольной выборкой. Временная память, сохраняющая информацию до тех пор, пока не отключается питание.

Импеданс. Эквивалентное сопротивление электрической цепи переменному току.

Интегратор. Электронная схема или система, выполняющая математическое интегрирование сигнала.

Интонация. Изменение тона голоса, позволяющее особо выделять отдельные слова или фразы.

Конкатенировать. Связывать в последовательность.

Кодирование. Преобразование сигнала из одной формы в другую.

Коэффициент отражения. Величина предикции в цифровой фильтрации при синтезе речи по методу ЛПК.

Лингвистика. Наука об общих законах строения и функционирования человеческого языка.

Линейное предиктивное кодирование (ЛПК). Способ математического моделирования речи, основанный на цифровой фильтрации голосовых и неголосовых колебаний (параметрическое кодирование).

Линейный. Изменяющийся линейно, или непрерывно, как, например, аналоговый сигнал.

Микрофон. Устройство, преобразующее звуковые колебания в электрические сигналы.

Микропроцессор. «Сердце» компьютера; центральный процессор, в котором роль центра управления выполняет монолитная интегральная схема.

Модуляция. Способ кодирования сигнала; изменение характеристик сигнала по заданному закону, медленное по сравнению с основным колебанием.

Монотонность. Интонация с надоедливой постоянной высотой тона.

Морф. Последовательность связанных фонем, образующая минимальную значимую часть словоформы.

Морфема. Минимальная значимая часть слова; совокупность морфов, имеющих одинаковое значение и ряд других общих признаков.

Нейрон. Нервная клетка, состоящая из тела и отходящих от нее отростков; основная структурная и функциональная единица нервной системы.

Нормализация. Сопоставление сигнала или некоторой величины с опорным, или эталонным, значением.

Посовой. 1) Имеющий отношение к носу; 2) группа согласных звуков.

Обертоны см. *Гармоники*.

Обратная связь. Управляемая реакция в системе, при которой выходной сигнал используется для корректировки.

Октава. Интервал частот, верхняя граница которого вдвое превышает нижнюю.

Омонимы. Слова с различным написанием, но с одинаковым произношением.

Оральный. Ротовой, относящийся ко рту.

Основная частота. Низшая частота сигнала, содержащего гармоники.

Палатальный звук. Артикуляция звуков путем касания кончиком языка твердого нёба.

Паркор. Частичная корреляция — вид линейного прогнозирования сигналов; способ синтеза речи, используемый некоторыми японскими фирмами.

Перегрузка из-за большой крутизны. Искажения сигнала, вызываемые дельта-модуляцией; возникают, когда дельта-модулятор не успевает за изменениями сигнала.

Период. Полный цикл колебания в периодическом процессе.

Полоса частот. Часть спектра сигнала, существенная для его правильной передачи.

Полосовой фильтр. Фильтр, пропускающий только ограниченную полосу частот.

Полость. Полая область, например гортань или носовая полость, обладающая резонансными свойствами.

Порог. Фиксированная величина, или значение амплитуды, используемая для отбора сигналов выше или ниже ее.

Порт. Соединение между компьютером и его периферийными устройствами; ввод-вывод.

Последовательное включение. Включение, при котором каждый каскад схемы соединен своим выходом с входом следующего каскада.

Постоянное запоминающее устройство (ПЗУ). Постоянная память ЭВМ, обычно используемая для хранения данных или программ. Содержащаяся в ней информация сохраняется и при выключенном питании.

Принцип Найквиста (теорема отсчетов). Утверждает, что для сохранения верности передачи частота выборки сигнала должна по крайней мере вдвое превышать самую высокую частотную составляющую сигнала.

Промежуточный звук. Согласный звук речи, относящийся к категории полугласных.

Псевдослучайный. Имеющий вид случайного на протяжении ограниченного отрезка времени.

Резонанс. Усиление колебаний определенной частоты при отсутствии такового на других частотах.

Резонатор. Устройство, усиливающее (выделяющее) некоторые частоты по сравнению с другими.

Рекурсивный фильтр. Многозвенный лестничный фильтр с многими обратными связями.

Рецепторные клетки. Специализированные нервные клетки, преобразующие раздражения, воспринимаемые извне или из организма, в нервное возбуждение, передаваемое в центральную нервную систему.

Семантика. Раздел языкознания, изучающий значения слов.

Синтаксис. Способы соединения слов (и их форм) в словосочетания и предложения, а также соединения предложений в сложные предложения.

Синтезатор речи. Устройство для генерации искусственной речи.

Синусоидальное колебание. Периодический аналоговый сигнал, описываемый формулой $\sin 2\pi ft$, где f — частота, t — время, $\pi = 3,14159$.

Слог. Сегмент речи или письма, более длинный, чем фонема, но короче слова (за исключением односложных слов).

Слуховой нерв. Группа нейронов, передающих сигналы от ушей к головному мозгу.

Случайный. Характеризующийся равной вероятностью любого события.

Согласный звук. Звук речи, относящийся к категориям взрывных, фрикативных, носовых, промежуточных и полугласных; негласный звук.

Спектр. Совокупность всех частот, содержащихся в сигнале.

Спектрграмма (звука). Видимая запись речи с помощью спектрографа, показывает частотный спектр речи.

Трахея. Часть дыхательного пути между гортанью и бронхами, орган дыхания.

Улитка. Спиралевидная трубка во внутреннем ухе; преобразует звуковые колебания в импульсы, возбуждающие слуховой нерв.

Ушная раковина. Внешняя часть уха.

Фильтр нижних частот. Электрическая схема, ослабляющая высокие частоты и пропускающая низкие.

Фонема. Основная звуковая единица речи.

Фонетика. Раздел языкознания, изучающий способы образования звуков речи и их акустические характеристики.

Форманта. Частотная полоса, выделяющаяся в спектре речевого сигнала.

Формантный синтез. Синтез речи, основанный на восстановлении формантных частотных полос.

Фрикативный звук. «Статический» звук; при произнесении остаются неподвижными все части голосового тракта, кроме голосовых связок и (или) легких; согласный звук речи, имеющий широкий частотный спектр, произносимый с шумом.

Фурье-преобразование. Математический метод определения спектрального состава сигнала.

Цифровой. Относящийся к области дискретной техники — в противоположность аналоговой.

Шепот. Речь, состоящая только из глухих звуков.

Шум. Широкополосный сигнал, характеризующийся случайными изменениями амплитуды, частоты и пр., не несущий никакой информации.

Электрический аналог. Электронная модель физического явления.

Язык (анат.). Мышечный вырост на дне ротовой полости; принимает участие в акте речи, Центр артикуляции,

ПРИЛОЖЕНИЕ

Мы предлагаем читателю набор схем, которые можно использовать в работе с синтезатором речи; их можно применять как автономно, так и совместно с другими схемами для синтеза речи.

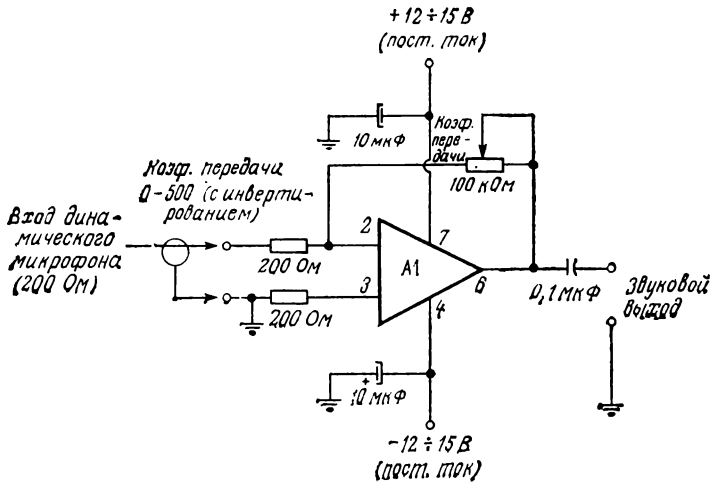


Рис. I. Усилитель динамического микрофона. [A1 — операционный усилитель типа 741.]

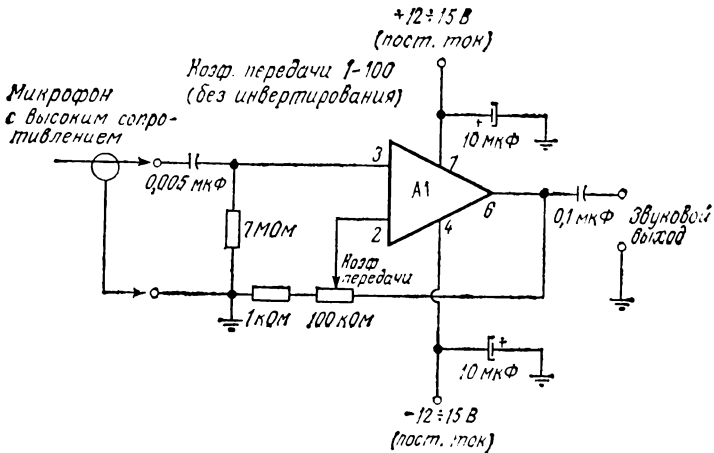


Рис. II. Усилитель микрофона с высоким входным сопротивлением. [A1 — то же, что на схеме рис. I.]

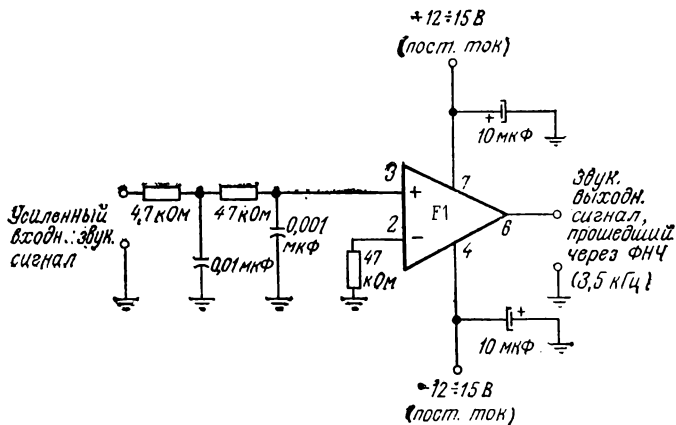


Рис. III. Фильтр нижних частот (3,5 кГц). [F1 — операционный усилитель 741.]

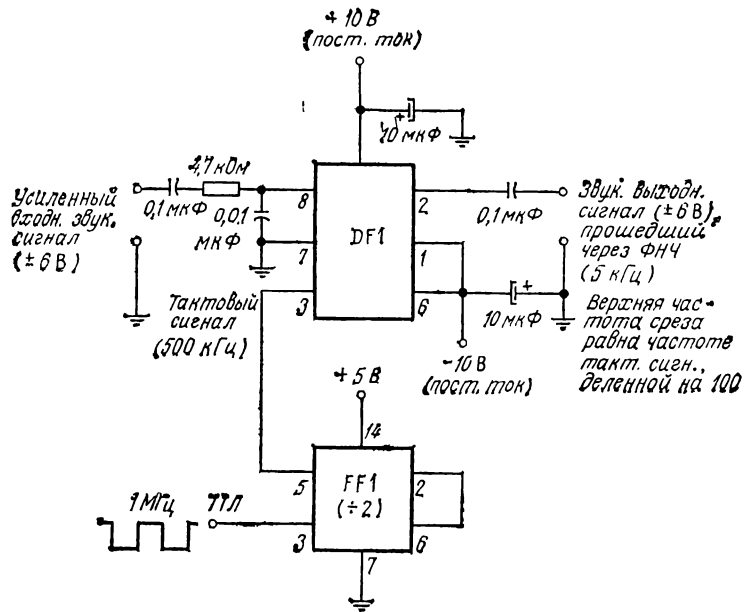


Рис. IV. Фильтр нижних частот (15 кГц) с крутыми спадами, использующий фильтр на переключаемых конденсаторах. [DF1 — коммутируемый ФНЧ типа Р 5609 фирмы «Ретикон» (Саннивейл, шт. Калифорния).]

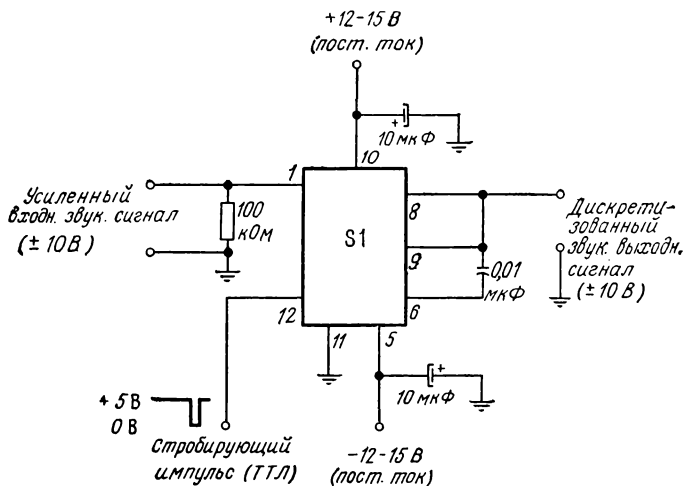


Рис. V. Дискретизатор аналогового сигнала. [S1 — недорогая интегральная схема выборки и запоминания типа AD582KD фирмы «Аналог дивайсес» (Норвуд, шт. Массачусетс).]

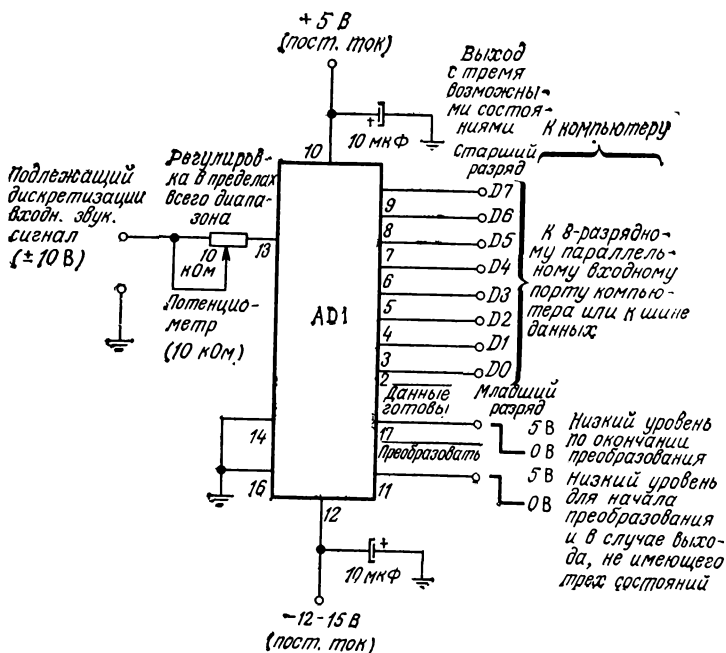


Рис. VI. Схема дискретизатора речевого сигнала. [AD1 — аналого-цифровой преобразователь типа AD570JN фирмы «Аналог дивайсес» (время преобразования примерно 40 мкс).]

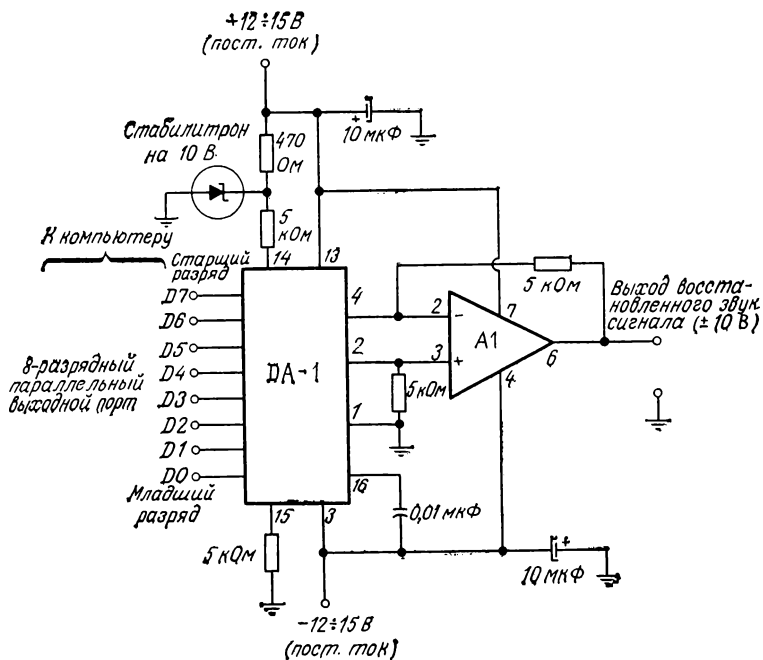


Рис. VII. Схема восстановления речевого сигнала. [DA1 — микросхема цифро-аналогового преобразователя типа DAC-08CQ фирмы «Пресижн монолитикс» (Санта-Клара, шт. Калифорния); A1 — то же, что на схеме рис. I.]

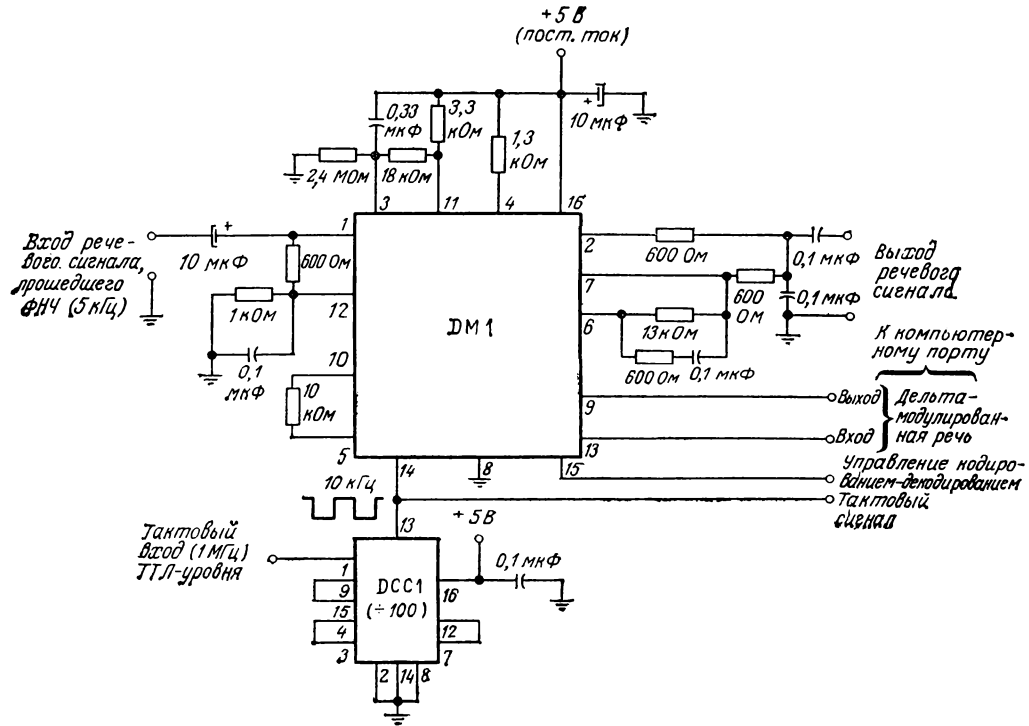
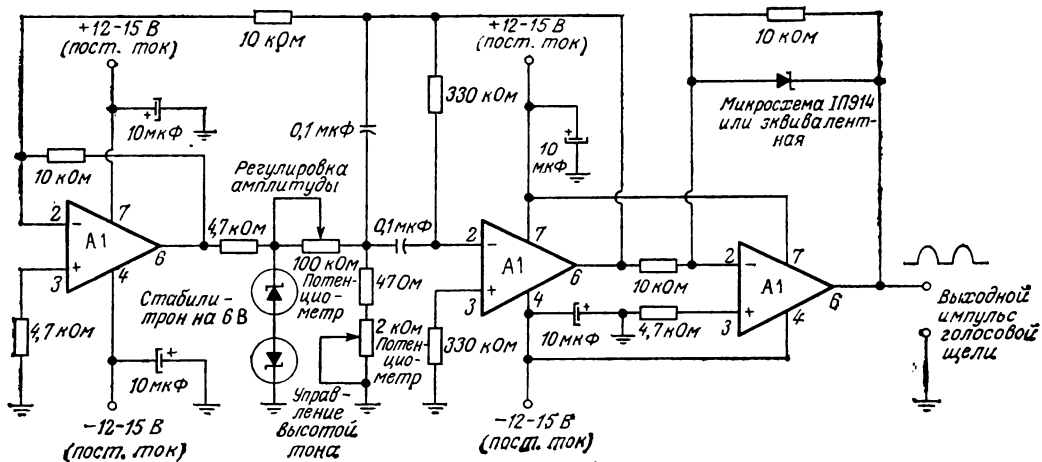


Рис. VIII. Схема кодирования-восстановления по методу дельта-модуляции. [DM1 — микросхема MC 3417 дельта-модулятора/демодулятора фирмы «Моторола»; DCC1 — двояный счетчик типа 74 LS 390.]



Изменение высоты тона
голоса с 50 до 400 Гц

Рис. 1X. Генератор импульса голосовой щели с регулируемой высотой основного тона, [A1 — см. схему на рис. 1.]

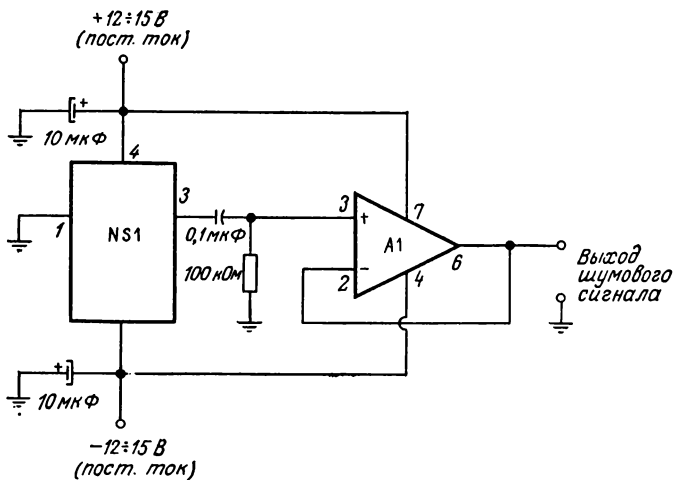


Рис. X. Цифровой источник фрикативного шума. [NS1 — микросхема MM5837N цифрового источника шума фирмы «Нэшнл семикондактор»; A1 — см. схему рис. I.]

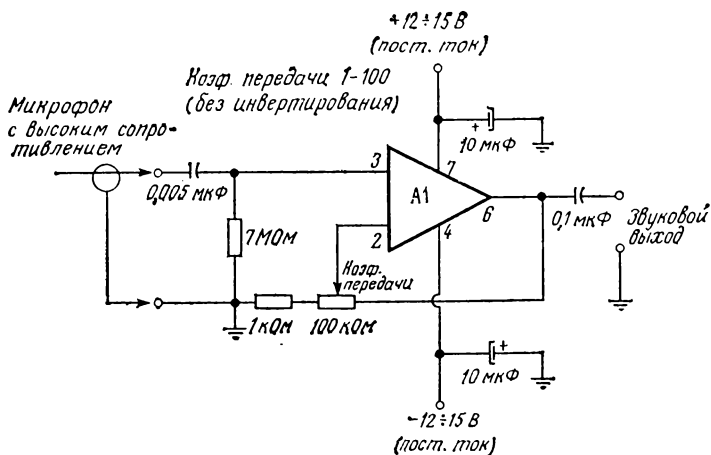


Рис. XI. Перестраиваемый полосовой фильтр формантных частот (средняя частота перестраивается в диапазоне 200—4000 Гц). [A1 — операционный усилитель типа 741.]

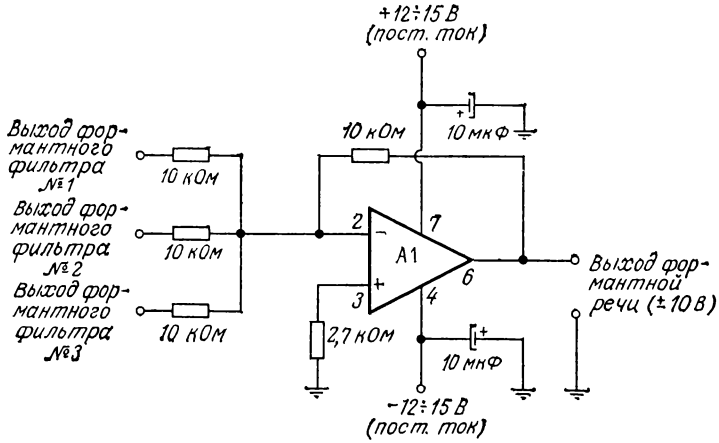


Рис. XII. Схема суммирования формантных частот. [A1 — см. схему на рис. I.]

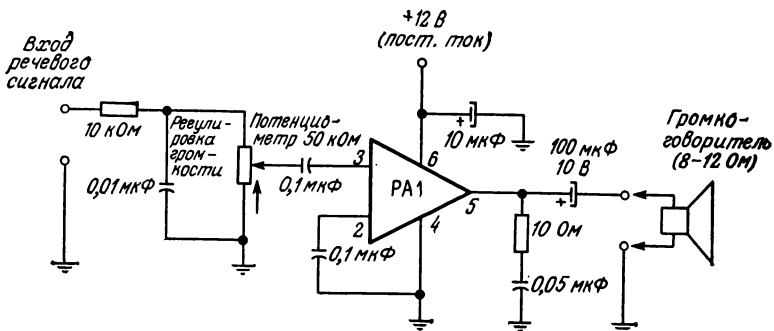


Рис. XIII. Усилитель мощности для синтезатора речи. [PA1 — усилитель мощности типа LM-386N фирмы «Нэшнл семикондактор».]

ЛИТЕРАТУРА

ФИЗИОЛОГИЯ РЕЧИ

- Curry R.** The Mechanism of the Human Voice.— New York: David McKay Co., Inc., 1940.
- Denes P. B., Pinson E. N.** The Speech Chain.— Baltimore, MD: Waverly Press, Inc., 1963.
- Fletcher H.** Speech and Hearing in Communication.— Princeton, NJ: Van Nostrand Co., Inc., 1953.
- Gray G. W., Wise C. M.** The Bases of Speech.— New York: Harper & Brothers, 1946.
- Kaplan H. M.** Anatomy and Physiology of Speech.— New York: McGraw-Hill, 1960.
- Zemlin W. R.** Speech and Hearing Science: Anatomy and Physiology.— Englewood Cliffs, NJ: Prentice-Hall, Inc., 1968.

ЛИНГВИСТИКА

- Chomsky N., Halle M.** The Sound Pattern of English.— New York: Harper and Row, Publishers, 1968.
- Fairbanks G.** Voice and Articulation Drillbook.— New York: Harper and Row, Publishers, 1960.
- Harris Z. S.** Methods of Structural Linguistics.— Chicago: University Press, 1951.
- Heffner R.** General Phonetics.— Madison: University of Wisconsin Press, 1950.
- Hill A. A.** Introduction to Linguistic Structures.— New York: Harcourt, Brace and World, 1958.
- Jones D.** The Phoneme: Its Nature and Use.— Cambridge, U.K.: Heffer, 1950.
- Lyons J.** Introduction to Theoretical Linguistics.— Cambridge, U.K.: Cambridge University Press, 1968.
- Shuy R. W.** Discovering American Dialects.— Urbana, IL.: National Council of Teachers of English, 1967.

РЕЧЬ И ЭЛЕКТРОННЫЙ СИНТЕЗ

- Flanagan J. L.** Speech Analysis, Synthesis and Perception, 2nd Edition.— New York: Springer-Verlag, 1972.
- Markel J. D., Gray A. H.** Linear Prediction of Speech.— New York: Springer-Verlag, 1976.
- Oppenheim A. V., Schaffer R. W.** Digital Signal Processing.— Englewood Cliffs, NJ: Prentice-Hall, Inc., 1975.
- Peled A., Liu B.** Digital Signal Processing, Theory, Design and Implementation.— New York: John Wiley and Sons, 1976.

- Potter R. K., Kopp G. A., Kopp H. G.** Visible Speech.— New York: Dover Publications, 1966.
- Rabiner L. R., Gold B.** Theory and Application of Digital Signal Processing.— Englewood Cliffs, NJ: Prentice-Hall, Inc., 1975.
- Steele R.** Delta Modulation Systems.— London: Halsted Press, 1975.
- Teja E. R.** Teaching Your Computer to Talk.— Blue Ridge Summit, Pa: Tab Books, Inc., 1981.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

- Кондратов А. М.** Звуки и знаки.— М.: Знание, 1966.
- Кюннап Э. Ю.** Синтезаторы речевых сигналов.— Институт кибернетики АН СССР, Таллин: Валгус, 1975.
- Цазаров М. В., Кувшинов Б. И., Попов О. В.** Теория передачи сигналов.— М.: Связь, 1970.
- Сапожков М. А.** Речевой сигнал в кибернетике и связи.— М.: Связь, 1963.
- Синтезаторы (сб. статей).— М.: Знание, 1983.
- Фант Г.** Акустическая теория речеобразования.— М.: Наука, 1964.
- Фланнаган Дж.** Анализ, синтез и восприятие речи.— М.: Связь, 1968.
- Черри К.** Человек и информация.— М.: Связь, 1972.
- Гаргальяно Т., Фояз К.** Система для проектирования синтезаторов речи.— Электроника, 1981, т. 54, № 3, с. 24—29.
- Шаплин К., Франц Дж., Гуди К.** Программа синтеза речи для персонального компьютера.— Электроника, 1981, т. 54, № 3, с. 30—34.
- Уиггинс Р., Брентинхэм Л.** Синтезатор речи на трех ИС.— Электроника, 1979, т. 52, № 18, с. 25—34.
- Вейнрич Д.** Синтезатор речи с естественным звучанием.— Электроника, 1980, т. 53, № 8, с. 36—43.

ПРЕДМЕТНО-ИМЕННОЙ УКАЗАТЕЛЬ

- Адаптивная импульсно-кодовая дельта-модуляция 102
- Акустические резонаторы 73
- Акустическое взаимодействие 64
схема 65—67
- Алломорфы 50
- Аллофон 45, 216
- Амплитудный детектор и пороговая схема 64
- Аналоговый интегратор 102
— компаратор 99
— синтез формантных частот 15—17, 104, 136
— формантный синтезатор речи 156—173
- Аналого-цифровой преобразователь 90, 91, 116
- Артикуляция, упражнения 55—57, 216
- Бибиджа* машина 73
- Бейсик 45
- Белл А. Г. 75
- Бод, скорость передачи 132, 172
переключатель 172
- Варьные звуки 27, 41, 48—49, 217
- Водер 78—79
- Восприятие компьютерной речи 60
- Встречающиеся наиболее часто слова 51, 52
— звуки 51, 53
- Выборки (отсчеты)
абсолютные 97
аналогового сигнала 90
механизм 87
система 90
теорема 86, 90
частота 86, 88, 96
- Высота основного тона 217
генератора 119
голоса 35, 117
- Генератор псевдослучайного шума 119
- Генератор-преобразователь «речь — ЛПК-код» 189
- Глотка 32, 33, 217
- Говорящая система охраны жилища 206—208
- Говорящие меха 74
— механизмы 73—76
— устройства для людей с дефектами зрения и слуха 201—203
— часы 194—200
— читающие устройства 20, 156—161
- Говорящий дверной звонок 204—206

- Голосовой тракт 22, 33, 35, 73, 120
положение элементов 46, 47
электронные модели 14, 18
Голосовые связки 32, 35, 48, 217
колебания 35—36
— органы
механические 77
электрические 76
— фильтры 35—41
Гортань 31—34, 217
- Дельта-модуляция 101—103, 143—145, 152, 217
Диалект 58, 217
общеамериканский 44
Дифтонг 45—46, 50, 108, 111, 217
- Звуки
гласные 33, 46—47, 108, 217
мощность 54
носовые 107
согласные 48
фрикативные 33, 48, 107, 119
Звуковые эффекты 96
Зрительный нерв 24
- Импульсно-кодовая модуляция 91—92
Интегратор 102—103
Интерфейсный преобразователь 188
— модуль 189
Искусственная речь 23
Искусственный голос 20
- Кадр данных 110
для глухих звуков 122
для звонких звуков 122
ЛПК 116, 122
типы 122—123
Кемпелен В. 73
К/МОП-технология 164
Кодирование 218
последовательных приращений 99
Кодирование-восстановление аналоговых сигналов 87
речи 135
Компьютер «вежливый» 65
Компьютер персональный 12, 18, 22, 92, 124, 136, 155—156, 164, 168, 193
возможности памяти 83
интерфейс универсальный 180
ЛПК-синтезатор речи 178
Компьютерная речь
восприятие 60
правила 60—64
синтезаторы 22
Компьютерный язык 45
Конкатенирование 147, 218
фонемы 163
Котельников В. А. 86
Кратценштейн 73
- Лазерный принтер 14
Лингвистика 42, 43—59
Линейное предиктивное кодирование 123, 218
- Микропроцессор 127, 146—147, 180—182, 188, 218
автономный 168
встроенный 179, 183
порт ввода-вывода 153
6502 189
Микросхема для речевого синтеза 188

- синтезатора 163—164, 179
 Морфемы 50, 218
- Нейроны 25, 218
 Нервная система 25
 Новые приложения 21
 Псовые согласные 48, 49
- Обновляющий управляющий кадр 120
 Обратная связь 219
 слуховая 27
 речевая 27
 Обучающая детская игра 19
 Оптическая следящая система 158
 Оптическое распознавание знаков 156
 — читающее устройство и синтезатор речи 156
 Основной тон голоса
 высота 106, 107
 импульс 35, 37, 119
 Относительные изменения амплитуды 97
 «Отпечаток» голоса 39
 Отсчетов теорема 85, 91
- Параллельный порт 186
 данных 138
 компьютерный 167
 Параметрическое кодирование речевого сигнала 130
 синтез 132
 Паркор-синтез 173, 175—177, 219
 Паскаль 45
 Передача данных, скорость 132, 176
 ПЗУ на кристалле 176
 Полугласные звуки 48
- Портативная система для анализа и синтеза 189—192
 Предикторные коэффициенты 124
 Преобразования текста в речь, алгоритм 170
 Преобразователь аналого-цифровой 90, 91, 116
 — цифро-аналоговый 119
 Программные средства 133, 137, 168
 Программы на ассемблере 95—96
 Промежуточные звуки 54, 219
 согласные 48, 49
- Резонансная теория слуха 29
 Резонансы со свистом и шипением 41—42, 49
 Резонатор 36—37, 219
 Рецепторные клетки 25
 Риш Р. 77, 78
 Роботы 12
 Речевые данные в компактной форме 137
- Речь
 восстановление 80—82, 84—85
 выборки 116
 генератор 180—181, 182—184
 генератор-преобразователь «речь — ЛПК-код» 189—191
 звуки, частота появления 53
 запоминание и хранение 83—84, 138
 искусственная 23
 кодированная 116
 линейное предиктивное кодирование 176
 механизм вывода речи ЛПК-генерация 188

- у человека 30—35
 - модуль 167
 - прямое кодирование-восстановление 87—103
 - общеамериканский диалект 46
 - распознавание 68, 135, 155, 156
 - расход памяти 138
 - скорость 158
 - спектральный состав 116
 - терминал 130—132
 - физиологический процесс 23
 - фонетический синтез 83
- Сигнал** предупредительный 63
— посторонний 91
- Синтез речи**
аналоговый, формантных частот 104—115
ЛПК 115—126
на основе функций Уолша 126
непосредственное кодирование-восстановление 87—103
основные методы 80—81
специфический процесс кодирования 135
тахометр 81—83
фонетический 80
Фурье 126
- Синтезатор речи**
автономный 167
аллофонный 113, 114
аналоговый формантный 156, 157
артикуляции упражнения 55—57
в виде микросхемы 163
интерфейс 139
завтрашнего дня 126
ЛПК 80, 178
- на основе цифрового моделирования голосового тракта 173—180
 - оптическое читающее устройство 156
 - преобразующий текст в речь 160
 - с прямым кодированием-восстановлением 136
 - с фонемным возбуждением 17—18, 103, 108—113
- Синтезированная речь** 13
выпускаемые устройства 127
коммерческие приложения 213—215
определение 84—86
по методу дельта-модуляции 138
применения 193
простейшая система 194
- Словарь** 138, 140—141
автономный 168
встроенный 131
заказной 125
записанный в ПЗУ 147—150, 176
предварительно наговоренный 125
стандартный 184
- Слух** 28—30
субъективная оценка слушателя 55
- Слуховая обратная связь** 26
- Спектрограмма** 39—41, 220
- Стираемое ППЗУ** 165
- Структура предложения** 59
- Текст в речь, преобразователь** 115, 178, 187—188
программа грамматического разбора 178
- Телефонная теория слуха** 29
- Теория места** 29—30

- Терминал компьютера 130—
133
ТТЛ-вентили 134
- Улитка (анат.) 29, 220
Устное общение 13
Ухо 26—29
- Фильтры**
голосовые 35—41
нижних частот 94—95
полосовые 17
управляемые напряжени-
ем 40
- Фонемы 17—18, 44—46, 47,
49, 54—56, 105, 220
«нестатические» 111
основные 56
«статические» 49, 111
управляющая таблица 111
- Фонология 43
- Форманта 220
в спектре речи 38
частоты 17—18, 37, 41, 73,
104—105
- Формантный синтезатор 106—
115
аналоговый 80
- Фортран 45
- Фрикативные звуки 27, 33, 41,
74
— фонемы 48, 49
- Фурье-синтез 126, 221
- Цифровое моделирование голо-
сового тракта 15, 19
- Цифровой процессор речи 117
- Частичная автокорреляция
(паркор) 18
- Хранение в памяти, временное
64, 168

ОГЛАВЛЕНИЕ

Предисловие редактора перевода	5
Вступление	8
Глава 1. Основы компьютерного синтеза речи	12
Глава 2. Как мы говорим?	23
Глава 3. Немного о лингвистике	43
Глава 4. Этика «поведения» компьютера — синтезатора речи	60
Глава 5. Немного истории	72
Глава 6. Современные методы синтеза речи	80
Глава 7. Серийные системы синтеза речи	127
Глава 8. Применения компьютерной речи	193
Словарь терминов	216
Приложение	222
Литература	230
Предметно-именной указатель	232

Джон Кейтер

КОМПЬЮТЕРЫ — СИНТЕЗАТОРЫ РЕЧИ

Научный редактор А. Н. Кондрашова

Мл. редактор И. Б. Ильченко

Художник С. А. Бычков

Художественный редактор Н. М. Иванов

Технический редактор Л. В. Рыбалко

Корректор В. И. Постнова

ИБ № 5301

Сдано в набор 25.03.85. Подписано к печати 22.10.85.

Формат 84×108^{1/32}. Бумага типографская № 1. Печать высокая.

Гарнитура обыкновенная. Объем 3,75 бум. л. Усл. печ. л. 12,6.

Усл. кр.-отт. 12,92. Уч.-изд. л. 11,80. Иад. № 9/3804

Тираж 80000 экз. Заказ № 989. Цена 65 коп.

ИЗДАТЕЛЬСТВО «МИР»

129820, ГСП, Москва, И-110, 1-й Рижский пер., 2

Ордена Октябрьской Революции и ордена Трудового Красного Знамени

МПО «Первая Образцовая типография» имени А. А. Жданова

Союзполиграфпрома при Государственном комитете СССР

по делам издательств, полиграфии и книжной торговли.

113054, Москва, Валуев, 28

В СЕРИИ «В МИРЕ НАУКИ И ТЕХНИКИ»

В 1985 ГОДУ ИЗДАТЕЛЬСТВО «МИР» ВЫПУ-
СТИЛО КНИГУ

А. Эндрю «Искусственный интеллект».
Перевод с английского.

Книга известного английского ученого, специалиста в области кибернетики Алекса Эндрю в популярной форме рассказывает о поисках и достижениях исследователей, работающих на одном из передовых рубежей современной науки — в области искусственного интеллекта.

Предназначена для широкого круга читателей, интересующихся достижениями современной науки и техники.

В СЕРИИ «В МИРЕ НАУКИ И ТЕХНИКИ»

В 1985 ГОДУ ИЗДАТЕЛЬСТВО «МИР»
ВЫПУСТИТ КНИГУ

А. Азимов «Язык науки». Перевод с
английского.

Книга известного американского ученого и популяризатора науки Айзека Азимова рассказывает о происхождении различных научных терминов и связанных с ними научных открытиях.

Предназначена для широкого круга читателей.