

DIGITAL PROCESSING OF SPEECH SIGNALS

Lawrence R. Rabiner

*Acoustics Research Laboratory
Bell Telephone Laboratories
Murray Hill, New Jersey*

Ronald W. Schafer

*School of Electrical Engineering
Georgia Institute of Technology
Atlanta, Georgia*

Prentice-Hall, Inc., Englewood Cliffs, New Jersey 07632

ЦИФРОВАЯ ОБРАБОТКА РЕЧЕВЫХ СИГНАЛОВ

Л.Р.Рабинер, Р.В.Шафер

Перевод с английского под редакцией
М. В. Назарова и Ю. Н. Прохорова

Москва «Радио и связь» 1981

Рабинер Л. Р., Шафер Р. В.

P12 Цифровая обработка речевых сигналов: Пер. с англ./Под ред. М. В. Назарова и Ю. Н. Прохорова. — М.: Радио и связь, 1981. — 496 с., ил.

В пер. 2 р. 50 к.

Рассматриваются вопросы цифровой обработки речевых сигналов в системах передачи информации и управления ЭВМ голосом. Излагаются проблемы цифрового представления речевых сигналов: временная дискретизация, интерполяция, квантование, проектирование цифровых фильтров. Обсуждаются способы построения цифровых систем передачи, систем идентификации и верификации диктора.

Предназначена для инженеров, специализирующихся в данной области, а также для студентов вузов соответствующих специальностей.

Р 30602—185
046(01)—81 11—81 (С.) 2402040000

ББК 32.87
6Ф1

Редакция литературы по электросвязи

© 1978 by Bell Laboratories, Incorporated
© Перевод на русский язык, предисловие, примечания, предметный указатель, издательство «Радио и связь», 1981.

ПРЕДИСЛОВИЕ К РУССКОМУ ИЗДАНИЮ

Методы цифровой обработки и передачи речевых сигналов в настоящее время интенсивно развиваются. Это прежде всего обусловлено прогрессом в области цифровой микросхемотехники, благодаря которому появилась реальная возможность изготовления сложной цифровой аппаратуры передачи сообщений, а также цифровых устройств распознавания речи, синтеза речи и др. Первые образцы таких устройств, уже освоенные промышленностью, вызвали повышенный интерес разработчиков к открывающимся возможностям и привлекли новых приверженцев этого направления исследований к изучению современных методов и алгоритмов цифровой обработки речи. Эта задача, однако, оказалась довольно трудной, так как за последнее десятилетие разработано большое количество алгоритмов разной эффективности и разного назначения, описание которых разбросано по журнальным статьям и докладам.

В предлагаемой вниманию читателей книге, по-видимому, впервые собраны различные способы цифровой обработки и передачи речевых сигналов, которые излагаются доступно и достаточно глубоко, и применение которых иллюстрируется примерами решения наиболее важных практических задач. Авторами подобран богатый иллюстративный материал, позволяющий читателю глубже изучить современные представления о свойствах сигнала, основные особенности цифровых методов обработки и определить собственное отношение к достоинствам и ограничениям последних.

Важно подчеркнуть, что основные задачи обработки и передачи речи — создание систем низкоскоростной передачи с высоким качеством восприятия сигнала, способных функционировать в реальных условиях, методов объективной оценки качества восприятия речи, систем распознавания слитной речи на фоне мешающих факторов — пока еще не решены. С этих позиций книга будет полезна и потому, что в ней приводятся итоги плодотворных, выполненных в последнее время, исследований, которые могут послужить основой для новых идей, направленных на решение указанных проблем.

Книга не лишена отдельных недостатков, связанных с некоторой непоследовательностью в терминологии, многословием, описками. Редакторы и переводчики приложили немало усилий для устранения этих недостатков, но уверенности, что удалось заметить все неточности, нет. В книге отражены не все методы цифровой обработки сигналов, в частности, не упоминаются направления, развиваемые в Советском Союзе. Редакторы сочли полезным привести список дополнительной литературы с указанием ряда работ советских авторов.

Профессор М. В. Назаров, доцент Ю. Н. Прохоров

ПРЕДИСЛОВИЕ

Эта книга является результатом совместной работы, которая началась еще в студенческие годы в МТИ¹, окрепла в период тесного сотрудничества в лаборатории Белла, продолжавшегося около шести лет, и связывает авторов в настоящее время как коллег и друзей. Поводом для начала работы над книгой послужила учебная статья по цифровому представлению речевых сигналов, написанная для специального выпуска ТИИЭР² по цифровой обработке сигналов, редактируемого профессором Аланом Оппенгеймом. Во время подготовки этой статьи стало понятно, что накопившихся результатов по цифровой обработке речи вполне достаточно для написания книги.

Авторы убедили себя, что они вполне способны написать текст такой книги и приступили к обсуждению способа ее построения. Были предложены, по крайней мере, три способа построения, и далее перед нами возник вопрос, какой из них обеспечит наиболее связное изложение предмета, если это вообще возможно. Суть каждого из способов состояла в предложении излагать содержание с одной из трех точек зрения: с позиции цифрового представления сигналов; на основе теории оценивания параметров; в соответствии с различными областями применения.

После длительной дискуссии стало ясно, что наиболее фундаментальными понятиями являются те, которые относятся к цифровому представлению речи, и что глубокое понимание подобных представлений позволит читателю не только понять, но и развить методы и способы оценивания параметров, а также разрабатывать системы цифровой обработки речи. Поэтому мы сконцентрировали содержание книги вокруг нескольких основных положений, относящихся к цифровому представлению речевых сигналов, изложив далее специальные методы оценивания параметров и способы их применения. Книга построена следующим образом. Глава 1 посвящена введению в круг задач обработки речи и содержит краткое обсуждение областей применения основных результатов. В гл. 2 содержится краткий обзор основ цифровой обработки сигналов. Предполагается, что читатель основательно подготовлен в области линейных систем и преобразований Фурье и, по крайней мере, имеет представление об обработке сигналов. Эта глава предназначена для ознакомления читателя с обозначениями и содержит краткие справочные сведения из теории цифровой обработки сигналов; в ней излагаются вопросы дискретизации, прореживания и интерполяции. Эти преобразования широко используются при разработке систем обработки речи. В гл. 3 обсуждаются физические основы образования звуков в речевом тракте и вводятся различные цифровые модели, описывающие этот процесс.

¹ Массачусетский технологический институт. (Прим. ред.)

² Журнал «IEEE Transactions». (Прим. ред.)

В гл. 4 рассмотрены методы обработки речи во временной области. Глава содержит обсуждение некоторых основных понятий, используемых при цифровой обработке речи — например, функций кратковременной энергии, среднего значения сигнала речи, среднего количества переходов через нуль, кратковременной автокорреляционной функции. В конце главы изложены принципы нелинейного сглаживания, которое наиболее эффективно при обработке результатов измерений во временной области. Глава 5 посвящена вопросам непосредственного цифрового представления речевых сигналов, т. е. его кодированию. Здесь обсуждаются вопросы равномерного и неравномерного квантования последовательности мгновенных значений, адаптивного и разностного квантования, кодирования с предсказанием. Представлены структурные схемы кодеров для обычной ИКМ и адаптивной разностной ИКМ.

Глава 6 является первой из двух глав, посвященных вопросам спектрального представления речи. Эта область традиционно является одной из тех, которым уделяется наибольшее внимание со стороны специалистов, поскольку ряд основных систем обработки речи, таких, как звуковой спектрограф и полосный вокодер, непосредственно связан с обсуждаемыми в данной главе вопросами. Так, здесь показано, как общий подход к спектральному анализу и синтезу речи открывает возможность исследования ряда систем обработки речи. Глава 7 — вторая по спектральному представлению речи — посвящена вопросам гомоморфной обработки. Идея, лежащая в основе последней, состоит в таком преобразовании речевого сигнала в частотную область, когда сигнал представляется в виде суммы отдельных составляющих, которые могут быть разделены с помощью общих методов линейной фильтрации. В данной главе обсуждаются методы выполнения этой процедуры и даются несколько примеров гомоморфной фильтрации. В гл. 8 изложены вопросы линейного предсказания сигналов. Это представление основано на аппроксимации речевого сигнала во временной области с минимальной средней квадратической ошибкой. Установлено, что этот метод является устойчивым, надежным и точным методом для представления речевых сигналов в разных ситуациях. В заключительной гл. 9 содержится обсуждение систем обработки речи, применимых при речевом общении человека с машиной. Цель данной главы: дать примеры построения некоторых систем обработки речи и показать, как идеи, развиваемые в книге, применяются в этих системах. Принципы построения систем, обсуждаемых в гл. 9, основаны на машинном синтезе речи, верификации и идентификации диктора, а также распознавании речи.

Содержание книги представляет собой курс по цифровой обработке речи, рассчитанный на семестр. Для активизации учебного процесса каждая глава содержит задачи, которые соответствуют материалу главы и предназначены для его закрепления. Успешное выполнение домашних заданий необходимо для хорошего усвоения теоретических положений. Однако, как увидит читатель, многие методы обработки речи носят эмпирический характер. Поэтому

Введение

1.0. Цель книги

Цель книги заключается в том, чтобы показать, как методы цифровой обработки могут быть использованы в задачах речевой связи¹. В данной вводной главе излагаются общие сведения о природе речевого сигнала, о том, как методы цифровой обработки могут быть использованы для изучения его свойств, обсуждается ряд основных задач, в которых применяются методы цифровой обработки.

1.1. Речевой сигнал

Речь предназначена для общения. Возможности речи с этой точки зрения можно характеризовать по-разному. Один из количественных подходов основан на теории информации, разработанной Шенноном [1]. В соответствии с этой теорией речь можно описать ее информационным содержанием или *информацией*. Другой способ описания речи заключается в представлении ее в виде *сигнала*, т. е. акустического колебания. Хотя идеи теории информации играют важную роль при построении сложных систем связи, но, как будет ясно из содержания книги, наиболее полезными на практике являются представления речи в виде колебания или в виде некоторой параметрической модели.

Речевое общение начинается с того, что в мозгу диктора возникает в абстрактной форме некоторое сообщение. В процессе речеобразования это сообщение преобразуется в акустическое речевое колебание. Информация, содержащаяся в сообщении, представлена в акустическом колебании весьма сложным образом. Сообщение сначала преобразуется в последовательности нервных импульсов, управляющих артикуляторным аппаратом (т. е. перемещением языка, губ, голосовых связок и т. д.). В результате воздействия нервных импульсов артикуляторный аппарат приходит в движение, результатом которого является акустическое речевое колебание, несущее информацию об исходном сообщении.

Сообщение, передаваемое с помощью речевого сигнала, является дискретным, т. е. может быть представлено в виде последо-

при изучении методов цифровой обработки речевых сигналов полезно проводить эксперименты. При чтении курса надо иметь в виду, что первое приближение к такому эксперименту может быть получено путем задания курсовых проектов по одному из следующих разделов: литературные обзоры и доклады; проекты по реализации технических устройств; машинное моделирование. Структура проектов и перечень тем по всем разделам приведены в конце гл. 9. Эти проекты весьма популярны среди наших студентов, поэтому мы призываем преподавателей внедрять их в учебный процесс.

Ряд лиц прямо или косвенно оказали заочительное влияние на содержание книги. Мы выражаем глубокую благодарность Дж. Л. Фланагану, руководителю акустического отдела лаборатории Белла, который сыграл и роль «инспектора», и роль наставника для обоих авторов. В течение ряда лет он служил нам образцом того, как надо проводить исследования и излагать данные о них доступным языком. Весьма велико его влияние как на содержание книги, так и на карьеры авторов.

Другими специалистами, с которыми нам посчастливилось сотрудничать, являются доктор Б. Гоулд из МТИ (лаборатория Линкольна), проф. А. Оппенгейм из МТИ, проф. К. Стивенс из МТИ. Эти специалисты — наши учителя и коллеги, и мы глубоко им признательны. Непосредственное участие в подготовке этой книги к печати принимали проф. П. Нолл из Бременского университета, который высказал ряд критических замечаний, доктор Р. Крохирэ из лаборатории Белла, просмотревший первый вариант книги, проф. Т. Барнвелл из Джорджии, который дал ценные комментарии к тексту. Г. Шоу тщательно работал над домашними заданиями. Дж. М. Триболет, Д. Длугоус, Р. П. Папамихалис, С. Гаглио, М. Ричардс и Л. Кайзер высказали ценные замечания по последней редакции книги. Наконец, мы хотим поблагодарить редакционный отдел лаборатории Белла за контроль над изданием книги и К. Патито, которая обеспечила превосходную работу по подготовке машинописного текста книги после ее многочисленных переделок. Мы благодарны П. Блайне, Дж. Ринболд, Дж. Эванс, Н. Кеннел и К. Тиллери за помощь в подготовке ранних вариантов отдельных глав. Мы благодарны также Джону и Мэри Франклин за их поддержку, оказанную одному из авторов (Р. В. Ш.). Авторы выражают благодарность отделу фотопечати лаборатории Белла, где был подготовлен полный текст книги.

Авторы

¹ В книге понятие речевой связи охватывает вопросы построения разнообразных систем обработки, хранения и передачи речевых сигналов, включая как традиционные системы телефонной связи, так и специальные информационные системы общения человека с ЭВМ. (Прим. ред.)

вательности символов из конечного их числа. Символы, из которых составлен речевой сигнал, называются *фонемами*. В каждом языке имеется присущее ему множество фонем, обычно от 30 до 50. Например, в английском языке можно выделить 42 фонемы (см. гл. 3).

Особый интерес представляет оценка скорости передачи информации, содержащейся в речевом сигнале. Грубая оценка получается из того, что физические ограничения на перемещение элементов артикуляторного аппарата позволяют человеку произносить в среднем 10 фонем в секунду. Если фонемы представить числами в двоичной системе счисления, то для всех фонем английского языка более чем достаточно шестизначного двоичного кода. Принимая среднюю скорость произнесения равной 10 фонемам в секунду и пренебрегая корреляцией между соседними фонемами, получим, что скорость передачи информации составляет 60 бит/с. Другими словами, при нормальном темпе произнесения письменный эквивалент речевого сигнала содержит 60 бит/с. Эта оценка, однако, не учитывает таких факторов, как индивидуальность и эмоциональное состояние диктора, скорость произнесения, громкость речи и т. д.

В системах речевой связи сигнал передается, хранится и обрабатывается различными способами. Задачи техники обуславливают применение различных форм представления речевого сигнала. Однако во всех случаях им присущи следующие особенности: 1) сохранение информационного содержания речевого сигнала; 2) представление речевого сигнала в форме, удобной для передачи и хранения, или в виде, позволяющем легко и достаточно гибко преобразовывать речевой сигнал без существенных информационных потерь.

Представление речевого сигнала должно быть таким, чтобы его информационное содержание легко воспринималось автоматически с помощью машины или при прослушивании человеком. Далее будет показано, что представление речевого сигнала¹ (но не его информационного содержания) может потребовать от 500 до 10^6 бит/с. При разработке способа представления речевого сигнала существенное влияние оказывают методы обработки сигнала.

1.2. Обработка сигналов

Задача обработки сигналов схематически представлена на рис. 1.1. В случае речевых сигналов источником информации является человек. Измерению или наблюдению обычно подвергается акустическое колебание. Обработка сигнала предполагает в первую очередь формирование описания² на основе некоторой модели с последующим преобразованием полученного представления

¹ В реальном масштабе времени. (Прим. ред.)

² Имеется в виду выбор совокупности физических параметров, определяющих процесс восприятия речи. (Прим. ред.)

в требуемую форму. Последним шагом в процессе обработки является выделение и использование информационного содержания сигнала. Этот шаг может осуществляться путем прослушивания сигнала человеком или его автоматической обработки. В качестве примера можно рассмотреть систему идентификации диктора из заданного ансамбля дикторов, в которой используется представление речевого сигнала в виде зависящего от времени спектра. Одним из возможных преобразований сигнала в этих условиях является усреднение спектра по всей фразе, сравнение среднего спектра с эталонами, имеющимися для каждого диктора, и затем выбор соответствующего диктора на основе полученных мер сходства спектров. Для данного примера информационным содержанием сигнала являются признаки индивидуальности диктора. Таким образом, обработка сигнала в общем случае предусматривает решение двух основных задач: получить общее представление сигнала либо в форме речевого колебания, либо в виде параметров и преобразовать полученное представление в более удобную для решаемой задачи форму.

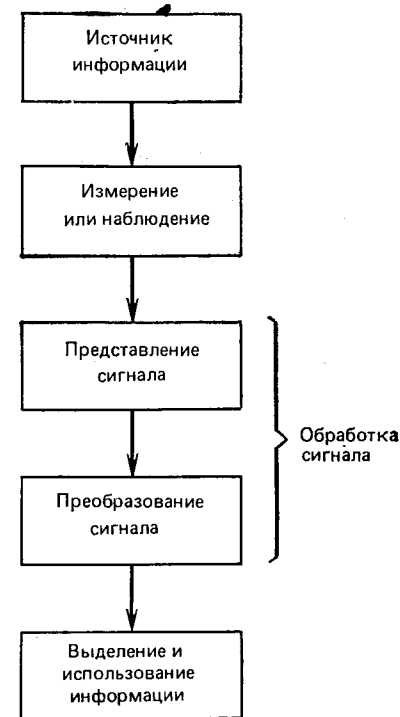


Рис. 1.1. Схема обработки информации

1.3. Цифровая обработка сигналов

Основное место в книге занимает исследование различных методов цифровой обработки речевых сигналов. Цифровая обработка включает как получение дискретных представлений сигнала, так и теорию, расчет и применение цифровых алгоритмов для преобразования полученных дискретных представлений. Конечная цель цифровой обработки сигналов такая же, как и при аналоговой обработке. Поэтому правомерно спросить, почему цифровые методы обработки требуют специального изучения в рамках общих методов обработки сигнала. Для этого имеется ряд серьезных причин. Первая, и возможно наиболее важная, заключается в том, что использование цифровых методов позволяет реализовать достаточно сложные алгоритмы обработки. В данной книге излагаются алгоритмы обработки в дискретном времени. В большинстве случаев они не могут рассматриваться как некоторые приближения аналоговой обработки, так как их в принципе нельзя реализовать

в аналоговых устройствах. Первые методы цифровой обработки речевых сигналов имитировали сложные аналоговые системы. Существовала точка зрения, согласно которой на ЭВМ следовало моделировать сложные аналоговые системы для выбора оптимальных параметров. При имитации первой аналоговой системы оказалось, что для проведения вычислений необходимо весьма большое количество машинного времени. Например, для обработки речи, прозвучавшей всего лишь несколько секунд, потребовалось более часа. В середине 1960-х гг. положение изменилось кардинальным образом. Причиной этому послужили разработка более быстрых ЭВМ и бурное развитие теории цифровой обработки сигналов. Стало ясно, что цифровые системы обработки сигналов обладают рядом достоинств, далеко превосходящих возможности простого моделирования аналоговых систем. Действительно, согласно современной точке зрения система цифровой обработки речевых сигналов, выполненная в виде программы на ЭВМ, реализует точный алгоритм обработки и может быть изготовлена в виде специализированного вычислительного устройства.

Одновременно с прогрессом в области теории успешно развивалась и технология изготовления цифровых устройств, что еще более увеличило преимущества цифровых методов обработки перед аналоговыми. Цифровые системы надежны и компактны. Технология производства интегральных схем достигла в настоящее время такого уровня, когда сложнейшая система обработки может быть реализована в виде одной микросхемы. Скорость выполнения логических операций в микросхемотехнике столь высока, что в большинстве случаев системы обработки речевых сигналов могут функционировать в реальном масштабе времени.

Существует и ряд других причин для применения цифровых методов обработки речевых сигналов в системах связи. Например, при использовании соответствующих кодов речевой сигнал может быть передан по каналу связи при наличии шума с малой вероятностью ошибки. Кроме того, если речевой сигнал представлен в цифровой форме, то он ничем не отличается от других цифровых сигналов. Поэтому один и тот же канал может быть использован как для передачи речи, так и для передачи данных, при этом отличия возникают только при декодировании. Если требуется повысить скрытность передачи, то цифровой сигнал имеет значительные преимущества перед аналоговым. Последовательность двоичных единиц, представляющая цифровой сигнал, для повышения секретности может быть перемешана известным образом и затем вновь восстановлена в приемнике. По упомянутой здесь и другим причинам цифровые методы в настоящее время широко применяются при решении задач обработки речевых сигналов [3].

1.4. Цифровая обработка речи

При рассмотрении вопросов применения цифровой обработки речевых сигналов к задачам связи полезно сконцентрировать внимание на трех основных направлениях: представлении речевых сигналов в цифровой форме, цифровой реали-

зации аналоговых методов обработки и методах, основанных исключительно на цифровой обработке.

Представление речевых сигналов в цифровой форме является, конечно, одним из центральных вопросов. Поэтому в книге рассматривается хорошо известная теорема дискретизации [4], утверждающая, что всякий ограниченный по полосе частот сигнал может быть представлен в виде последовательности равноотстоящих отсчетов, взятых с достаточно высокой частотой¹. Таким образом процедура дискретизации лежит в основе теории и приложений цифровой обработки. Существует ряд способов дискретного представления речевых сигналов. Как показано на рис. 1.2, эти способы могут быть разбиты на две большие группы —

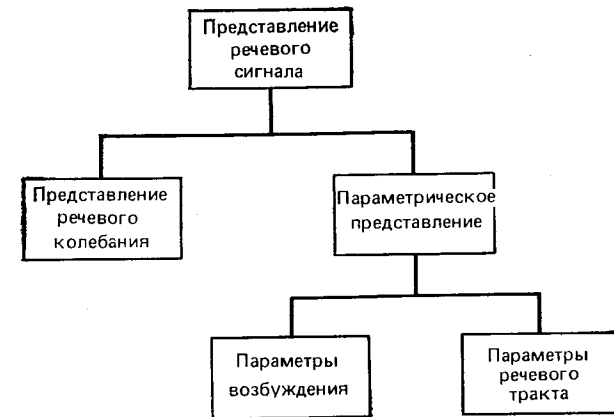


Рис. 1.2. Способы представления речевого сигнала

цифровое и параметрическое представление речевого колебания. Цифровое представление речевого колебания, как это следует из названия, основано на сохранении формы колебания в процессе дискретизации и квантования. Параметрическое представление базируется на описании речевого сигнала, как выходного отклика модели речеобразования. На первом этапе построения параметрического представления речевое колебание подвергается дискретизации и квантованию, а затем обрабатывается для получения параметров модели. Параметры модели обычно разделяются на параметры возбуждения (относящиеся к источнику звуков речи) и параметры голосового тракта (относящиеся непосредственно к отдельным звукам речи)¹.

На рис. 1.3 представлены результаты сравнительного анализа различных цифровых представлений по требуемой скорости передачи информации. Пунктирная линия, проходящая через точку 15 кбит/с, отделяет группу цифровых представлений речевого колебания (слева) от параметрических представлений (справа), которые обладают меньшим информационным объемом². Как следует из рисунка, требуемая скорость передачи изменяется от 75 бит/с (что примерно соответствует скорости передачи письменного эквивалента речи) до 200 000 бит/с и более при простейшем цифровом представлении речевого колебания. Таким образом, в зависимости от типа цифрового представления сигнала требуемая для его передачи скорость может изменяться примерно в 3000 раз. Конечно, скорость передачи далеко не единственный фактор, определяющий выбор типа цифрового представ-

¹ В отечественной литературе эта теорема известна как теорема Котельникова. (Прим. ред.)

² Детальное обсуждение параметрических моделей речевого сигнала содержится в гл. 3.

³ Понятие информационного объема сигнала введено А. А. Харкевичем. (Прим. ред.)

ления. Другими факторами являются стоимость, гибкость цифрового представления, качество восприятия речи и т. д. Обсуждение этих вопросов содержится в заключительных главах книги.

Наиболее важным фактором, определяющим выбор цифрового представления сигнала и методов цифровой обработки, является специфика решаемой приклад-

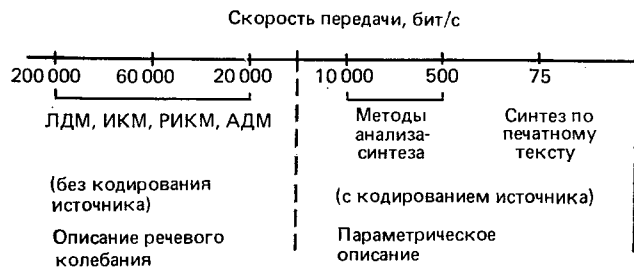


Рис. 1.3. Диапазон скоростей передачи при различном представлении речевого сигнала

ной задачи. На рис. 1.4 приведено несколько примеров из обширной области передачи и обработки речевых сигналов. Полезно кратко рассмотреть каждый из них для того, чтобы методы обработки, рассматриваемые в последующих главах, были более понятными.

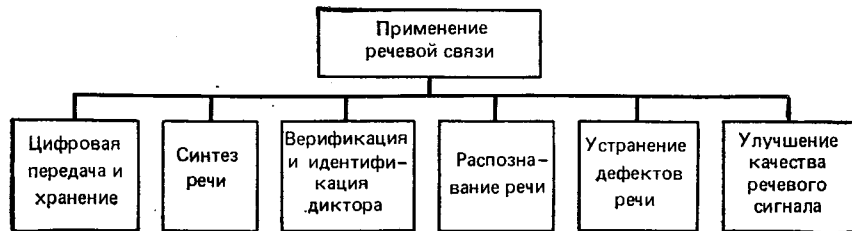


Рис. 1.4. Области применения речевой связи

1.4.1. Цифровая передача и хранение речевого сигнала

Одним из наиболее ранних и наиболее важных примеров применения обработки речевого сигнала является вокодер или кодер голоса (voice-coder), созданный Дадли в 1930-х гг. [5]. Целью разработки вокодера являлось уменьшение полосы частот, необходимой для передачи речи¹. Эта задача актуальна и в настоящее время, несмотря на наличие широкополосных спутниковых, СВЧ и оптических систем связи. Кроме того, необходимы дешевые и как можно более низкоскоростные преобразователи речи в цифровую форму для их использования в цифровых телефонных сетях связи. Одной из положительных сторон применения цифровых систем является возможность обеспечения скрытности передачи.

¹ Более точно, снижение требуемой пропускной способности канала связи при передаче речи. (Прим. ред.)

1.4.2. Системы синтеза речи

Большой интерес к системам синтеза речи объясняется необходимостью разработки способа экономичного хранения речевого сигнала в системах речевого ответа [6]. Подобная система реализует цифровой алгоритм автоматического сообщения голосом информации, которую запрашивает пользователь с клавиатуры пульта или специального терминала. Поскольку пультом может служить обычный телефонный аппарат с кнопочным набором, система речевого ответа может широко использоваться в коммутируемых телефонных сетях без установки какого-либо дополнительного оборудования [3]. Системы синтеза речи играют большую роль и при обучении правильному произношению речи [7].

1.4.3. Системы верификации и идентификации диктора

Методы верификации и идентификации диктора [8] включают установление подлинности, или идентификации, личности говорящего. Система верификации выносит решение о том, является ли говорящий тем, за кого он себя выдает. Системы такого типа применимы при управлении процессом доступа к информации или ограничении доступа, а также при проведении различного рода автоматических кредитных операций. Системы идентификации диктора должны выдать решение о том, кто из ограниченного числа дикторов произнес данную фразу. Такие системы могут применяться в области судебной экспертизы.

1.4.4. Системы распознавания речи

В самом общем виде системы распознавания должны преобразовывать речевое сообщение в эквивалентный текст. Сложность задачи распознавания определяется условиями произнесения и контекстом произносимой фразы, а также наличием или отсутствием возможности настройки на диктора. Системы распознавания речи могут применяться в различных устройствах, например, пишущих машинках, управляемых голосом или при речевом общении с ЭВМ. Совместное использование систем распознавания и синтеза речи позволяет получить систему передачи речевого сигнала с минимально возможной скоростью передачи [9].

1.4.5. Устранение дефектов речи

Здесь предполагается обработка речевого сигнала и отображение полученной информации в виде, наиболее приемлемом для обучаемого индивидуума. Например, воспроизведение сигнала, записанного на магнитофонную ленту с различной скоростью, наиболее подходит для слепых, поскольку позволяет им прослушивать текст с любого желаемого места. Разработан также ряд методов цифровой обработки сигнала для сенсорного и визуального отображения информации при обучении глухих речи [10].

1.4.6. Улучшение качества речевого сигнала

В ряде случаев речевой сигнал, поступающий в систему связи, оказывается искаженным, что снижает качество передачи. В этом случае методы цифровой обработки могут быть использованы для улучшения качества восприятия сигнала. Примерами подобных разработок являются устранение реверберации (или эха), устранение шума в речевом сигнале, восстановление речевого сигнала, записанного в гелиевокислородной среде, которая используется в качестве дыхательной смеси водолазами.

1.5. Заключение

В данной главе рассмотрены основные области применения методов цифровой обработки сигналов. Очевидно, что они весьма обширны и рассмотреть их достаточно глубоко в одной книге чрезвычайно трудно. При написании книги рассмотрено несколько вариантов ее построения, например, изложение может быть проведено в соответствии с классификацией представлений сигналов (см. рис. 1.2) или же наоборот, так, чтобы основное внимание уделялось применению методов цифровой обработки. Фактически о каждой области применения (рис. 1.4) может быть написана отдельная книга. Третья возможность, избранная здесь, состоит в построении книги в соответствии с имеющимися методами цифровой обработки сигналов. Такой подход по нашему мнению позволяет сконцентрировать внимание на наиболее важных направлениях рассматриваемой проблемы. Поэтому в последующих главах содержится обзор методов цифровой обработки сигналов (гл. 2), введение в цифровые модели речевого сигнала (гл. 3), обсуждение представлений речевого сигнала во временной области (гл. 4), кодирования речевого колебания (гл. 5), кратковременных спектральных представлений (гл. 6), гомоморфной обработки (гл. 7), линейного предсказания (гл. 8). В этих главах подробно рассмотрены вопросы теории цифровой обработки речевых сигналов. Для иллюстрации применения этой теории в различных приложениях в гл. 9 рассматриваются примеры систем речевого общения человека и машины.

2

Основы цифровой обработки сигналов

2.0. Введение

В данной книге рассматриваются цифровые методы обработки речи, поэтому читатель должен хорошо представлять основы теории цифровой обработки сигналов. В главе проводится краткий обзор основных положений этой теории, вводятся обозначения, которые используются далее. Читатели, которые незнакомы с методами описания и анализа сигналов в дискретном времени, при необходимости могут получить ряд полезных сведений в руководствах по цифровой обработке сигналов [1—3].

2.1. Сигналы и системы в дискретном времени

Изучение методов обработки или передачи информации естественно начинать с представления сигналов в виде непрерывно изменяющихся процессов. Акустическое колебание, формируемое в

речевом тракте человека, имеет именно такую природу. С математической точки зрения его можно описать функцией непрерывного времени t . Аналоговые (непрерывные во времени) сигналы будут обозначаться через $x_a(t)$. Речевой сигнал можно представить и последовательностью чисел. Такое представление и составляет, по существу, предмет данной книги. Последовательности обозначаются далее через $x(n)$. Если последовательность чисел представляет собой последовательность мгновенных значений аналогового сигнала, взятых периодически с интервалом T , то эта операция дискретизации будет иногда обозначаться через $x_a(nT)$. На рис. 2.1 показан пример речевого сигнала в аналоговой форме и в виде последовательности отсчетов, взятых с частотой дискретизации

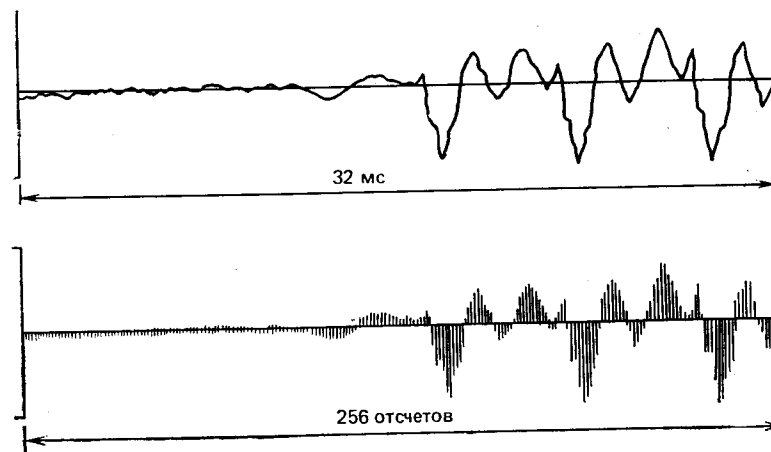


Рис. 2.1. Представление речевого сигнала

8 кГц. Для удобства даже при рассмотрении дискретных сигналов иногда на графике будет изображаться непрерывная функция, которая может рассматриваться как огибающая последовательности отсчетов. При изучении систем цифровой обработки речи нам потребуется несколько специальных последовательностей. Часть из них представлена на рис. 2.2. Единичный отсчет или последовательность, состоящая из одного единичного импульса, определяется как

$$\delta(n) = \begin{cases} 1, & n = 0, \\ 0, & \text{в других случаях.} \end{cases} \quad (2.1)$$

Последовательность единичного скачка имеет вид

$$u(n) = \begin{cases} 1, & n \geq 0, \\ 0, & n < 0, \end{cases} \quad (2.2)$$

Экспоненциальная последовательность

$$x(n) = a^n. \quad (2.3)$$